The Effects of Pro-Innovation Policies: Theory and Evidence*

Edwin L.-C. Lai† Pang Paul Wang‡ Isabel K. M. Yan§

July 5, 2025.

Abstract

We analyze patent promotion policies (PPP) in China by constructing a general equilibrium growth model with endogenous productivity improvement and carrying out empirical analyses. We examine how provincial PPP stimulates R&D and patent generation and analyze how it affects national welfare. We derive a necessary condition for provincial PPP to be welfare-improving for the nation. We then carry out an elaborate empirical welfare analysis by calibrating the model and carrying out counterfactual exercises to find out the impacts of provincial PPP on national welfare in China. We found that there is a non-trivial probability of 26% that provincial PPP adopted in 1995-2009 had been welfare-improving in China. However, the expected value of the welfare change caused by PPP during this period was negative in every year, and it was increasing over time.

JEL Classification codes: O31, O34, O38

Keywords: Pro-innovation policies, Chinese patents, innovation, patent subsidy

*We thank Yao Amber Li for her immense help in the initial stage of this project.

†Corresponding author. CESifo Research Network Fellow. Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Phone: (852) 2358-7611; Email: elai@ust.hk or EDWIN.L.LAI@gmail.com . He acknowledges support by the General Research Fund of the Research Grants Council of Hong Kong (Project No. 16506820).

‡School of Accounting and Finance, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. Email: paul.wang@polyu.edu.hk

§Department of Economics and Finance, City University of Hong Kong, Kowloon Tong, Hong Kong. Email: efyan@cityu.edu.hk. She acknowledges support from the General Research Fund (project no. 9043088).

1 Introduction

Innovation is widely considered as a key engine that drives a nation's long-run growth and a major determining factor of a country's competitiveness. Many countries have resorted to the provision of subsidies to promote innovation (Dimos and Pugh, 2016). Extensive empirical studies have been conducted in the literature to examine the relationship between subsidy policies and innovation in a wide range of countries, including the United States (Smith, 2020), Germany (Hud and Hussinger, 2016), Switzerland (Beck, Lopes-Bento and Schenker-Wicki, 2016), Italy (Bronzini and Piselli, 2016), Turkey & Poland (Szczygielski and Grabowski and Pamukcu and Mehmet and Tandogan, 2017), pan-Europe (Hünermund and Czarnitzki, 2019; Szücs, 2020), South Korea (Doh and Kim, 2014; Choi and Lee, 2017), and India (Ivus and Jose and Sharma, 2021). The effects of pro-innovation policies can vary according to different countries' degree of trade liberalization (Geng and Saggi, 2022) and other economic policies such as climate change policies (Gans, 2012) and fiscal policies (Montmartin and Herrera, 2015; Stojčić and Srhoj and Coad, 2017).

In recent decades, China has also emphasized the use of pro-innovation policies aiming at stimulating inventions and sustaining higher economic growth. According to World Intellectual Property Organization (WIPO), China had the fastest growth in patent filings among all major countries during the last decade, with a growth of over 174% from 561,472 in 2012 to 1,538,549 in 2021.[1] The growth in the US, Japan, Germany, France, UK, and Switzerland in the corresponding period were 7.5%, $-15.8\%, -9.8\%, -5.9\%, 3.6\%$ and 14%, respectively. Furthermore, in 2011, the China National Intellectual Property Administration (CNIPA) became the busiest patent office in the world when it overtook the United States Patent and Trademark Office (USPTO) and Japan Patent Office (JPO) in receiving patent applications. CNIPA reached another milestone in 2015 when it surpassed the USPTO and JPO in the number of patents issued to become the patent office that awards the largest number of patents in the world.[2] Hu, Zhang, and Zhao (2017) has provided a nice documentation of the patent growth in China in 2007-2011, and the surge in patents was found to be associated with increased productivity (Fang and He and Li, 2020).

One factor widely considered as important in contributing to the stark contrast between China and other countries in the growth of patent output is the extensive R&D subsidy policies implemented in China, which have led to a remarkable increase in R&D expenditure as a share of GDP (Boeing, 2016; Guo and Guo and Jiang, 2016; Guo and Guo and Jiang, 2017; Gao and Hu and Liu and Zhang, 2021).[3] Among various pro-innovation policies, one important type is patent promo-

[1] A common measure of countries' innovative capability is the number of patent applications granted.

[2] Data taken from the World Intellectual Property Office (WIPO).

[3] In fact, R&D subsidies are also found to generate other economic effects, such as raising employment (Lanahan

tion policies (PPP), which were widely adopted by provincial governments across China. PPP may take the form of reward or subsidy policies that directly impact the benefits and costs of patent production.[4]

This paper aims at examining the effectiveness of these PPP in stimulating provincial R&D expenditure, innovation, and their impacts on national welfare. The model considers two types of PPP. The first type is provincial patent reward policy that incentivizes successful patent applications via tax deductions, tax rebates, and rewards for holders of granted patents. The second type is provincial patent subsidy policy that provides subsidies to inventors for paying patent application or maintenance fees.[5] To carry out the analysis, we develop a tractable general equilibrium growth model with endogenous productivity improvement through innovation. Based on our model, we identify four sources of externalities, namely knowledge spillovers (positive), business stealing (negative), appropriability (positive) and congestion externalities (negative). This implies that the decentralized provincial market equilibrium gives rise to a sub-optimal national allocation of labor to R&D activity that is strictly lower than the national social planner's allocation if the net externality is positive. We derive a necessary and sufficient condition for this to be true, and use it to guide our empirical welfare analysis. If this condition holds, it justifies the use of provincial patent-promotion policy to close the wedge.

We go on to calibrate the model and carry out counterfactual exercises to estimate the impacts of provincial PPP on national welfare. We employ patent-level data from the China National Intellectual Property Administration (CNIPA) and PPP data from 31 Chinese provinces in 1985-2009. This is an elaborate empirical welfare analysis. The most important result of this exercise is that the probability that provincial PPP in China was welfare-improving during 1995-2009 is 25.9%, which is non-trivial. Thus, we conclude that PPP in China during this period was potentially welfare-improving. This stands in contrast to the findings of the prevailing literature.

Overall, our contribution to the literature is twofold. First, we develop a novel general equilibrium endogenous growth model with growth being propelled by productivity improvement through innovation. The model captures the mechanism through which provincial PPP affects provincial and national quantity and quality of patents generated through its stimulative effect on provincial R&D. We derive the condition under which provincial PPP will improve national welfare. We derive a necessary condition for the socially optimal level of R&D to be higher than the equilibrium level,

and Joshi and Johnson, 2021), signaling R&D quality (Bianchi and Murtinu and Scelera, 2019), enhancing university cooperation (Szücs, 2018), extending collaboration networks (Broekel and Fornahl and Morrison, 2015; Chapman and Lucena and Afcha, 2018), and increasing access to other forms of financing (Meuleman and De Maeseneire, 2012).

[4] Other countries have also adopted various country-specific patent policies, such as compulsory licensing or tax reduction on patent income, to promote innovation production. However, the effectiveness of these policies was mixed across countries (Baten and Bianchi and Moser, 2017; Gaessler and Hall and Harhoff, 2021).

[5] These two policies had very similar innovation-stimulating effects, and welfare effects. So, we pool them together when we carry out our empircial analysis.

and explain why it is plausible that PPP are potentially welfare-improving in China and other low- and middle-income countries.

Second, we carry out an elaborate empirical welfare analysis that is closely linked to the theoretical model. We estimate the probability that PPP was welfare-improving during 1995-2009 as well as the expected value of the welfare change in China caused by PPP during this period. We find that while there is a non-trivial probability that PPP was welfare-improving, the expected value of the welfare change caused by PPP was negative. Although our result that PPP was welfare-reducing is in line with the literature, we arrive at our conclusion based on a different mechanism. Most papers in the literature argue that PPP were welfare-reducing because there is a limited amount of resources that can be used for R&D, and PPP caused more resources to be allocated to develop a large quanity of low-quality patents that had limited commercial value and less resources to be allocated to develop high-quality patents that had high commercial value. Thus, there is a quality-quantity tradeoff, leading to lower welfare. By contrast, in our model, there is a limited amount of resources that can be used for either R&D or production, and PPP caused more resources to be allocated to R&D and less resources allocated to production. Thus, PPP would be welfare-reducing if the market-determined amount of resources allocated to R&D was already too high.

The rest of this paper is organized as follows. Section 2 reviews the relevant literature, describes the patenting growth in China, the various theories trying to explain it and highlights our contributions. Section 3 describes the theoretical framework used to analyze the effects and implications of PPP. Section 4 presents the empirical welfare analysis. The empirical approach, the data used and the results of each set of work are discussed. Finally, Section 5 concludes the paper.

2 Analyses of Patenting Growth and Patent Promotion in China

Since the 1980s, China has experienced a significant increase in patent applications. There are several theories in the existing literature that explain the driving forces behind this explosive growth.

One important theory emphasizes the role of government support policies. Li (2012) uses Chinese provincial data and patent data to show that patent applications increased after a province launched a patent application fees subsidy system. Dang and Motohashi (2015) examine how patent promotion policies affect the number of high- and low-quality patents granted. They use the number of claims to proxy for the quality of each patent and find that policies that provide monetary rewards increase the number of high-quality patents. Zhang and Nie (2021) found that the public health insurance program for rural residents increases the number of pharmaceutical patents filed, with a modest increase in quality. Wei, Xu, Yin and Zhang (2021) study the effects of the pro-innovation programs in China. They focus on a patent subsidy policy (the InnoCom program) in China.

It is found that the policy encourages firms to hold more patents, but disregards the quality of patents. The subsidy overall reduces the value of patents either through lowering the quality of newly granted patents or through the inefficient trade of patents. The outcome was that aggregate welfare declined substantially. Cao, Jiang, Li and Zhu (2023) adopt a Schumpeterian growth model featuring innovating firms' quantity–quality trade-off between radical and incremental innovations under the "quantity-based" patent-subsidy schemes of China. Their model highlights the existence of a scarce R&D resource, which induces a firm-level trade-off between quality and quantity. They found that, from 2001–2007 to 2008–2014, patent subsidies reduced the aggregate balanced-growth welfare by 3.31% because of the decline in the quality of patents.

In contrast, this paper presents a perspective different from the notion of a quantity-quality trade-off. In our model, the quality (citation count per year) of a patent generated by research is a random variable with a fixed probability density function in each year. Thus, there is a fixed probability of generating a patent at the top kth percentile of quality which is considered commercializable and also defined as high-quality. Under PPP, the market allocates more resources to research and less to production. Consequently, this leads to an increase in the number of both low- and high-quality patents awarded. It is the latter effect that fosters growth. This, however, does not necessarily increase welfare, as the realized rate of innovation may be too high compared with the social optimum. Thus, whether the rate of innovation is too high is an empirical question. In our empirical welfare analysis, we find that there is a non-trivial probability of 25.9% that PPP had been welfare-improving in China during the sample period of 1995-2009. However, we find that the expected welfare change to China as a whole caused by the provincial PPP was negative, indicating that the rate of innovation before the implementation of PPP might have already been too high compared with the social optimum.

3 Theory

This section presents a tractable general equilibrium model to examine the effects of pro-innovation policies on research activities and innovations.[6] The key equations are presented in this section, and the details of the model are relegated to Online Appendix A.

3.1 Model Setup

Consider an economy (a country with many regions) in which there is a continuum of differentiated varieties indexed by $j \in [0, \mu]$ that yield utility to consumers. Labor is the only factor input, which

[6] This paper extends the general equilibrium model of Chor and Lai (2016) that analyzes the relationship between cumulative innovation, growth and welfare-improving patent policy.

can be interpreted broadly as the totality of all resources that can be used for production or R&D. Thus, "labor" or "worker" in this model stands for a composite factor input consisting of labor, capital, land, and other resources that can be used for production or research.[7] The "wage" w is therefore the value of one unit of this composite factor input.[8] The words "labor", "worker", "labor force", "factor input", and "composite factor input" refer to the same thing and can be used interchangeably. The "wage" and "the unit value of the composite factor input" refer to the same thing. We assume that the composite factor input is mobile across regions, there is free trade of goods within the country and there exists a national capital market. Free mobility of the composite factor implies that the returns to the composite factor in all regions are equal.

Each unit of labor (or "worker") can engage in one of two activities, namely, in the production of differentiated varieties or in R&D. With regard to the former, the technologies used in production all feature constant returns-to-scale. $Z_\tau(j)$ is the labor productivity associated with the best available idea for producing variety j at time τ. Thus, $1/Z_\tau(j)$ units of labor are required to produce each unit of output, so that the unit cost faced by the firm that produces this variety is $w_\tau/Z_\tau(j)$.

In each region there is a research technology that can be used to generate commercializable patents. The objective of R&D activity is to generate commercializable patents embodying ideas that improve existing labor productivity. Each idea entails improving labor productivity for producing a specific differentiated variety.[9] The generation of these ideas follows a Poisson process, with the aggregate arrival rate being determined by the total quantity of R&D labor input.

In region i, the research technology yields a Poisson arrival rate of getting a patent equal to $f^i(R_i)$ where R_i is the input into R&D, and conditional on getting a patent, there is a probability δ that the patent is commercializable (i.e. high-quality). Conditional on commmercializability, there is a probability distribution of the productivity improvement ζ, which is a random variable. We assume that there is a fixed measure of the variety of goods given by μ.

By getting a commercializable patent, the firm can prevent others from using the idea. Thus, the patent-owning firm is the only one who can produce and market the variety (say, variety j) with this new technology and will indeed have the entire market for j to itself as it is now the most productive manufacturer of j. This monopoly power only expires when the next idea that improves upon j arrives.[10] Ideas that are not patented but are marketed can be imitated immediately by

[7] Formally, we can assume that human capital, labor, physical capital, land, and other resources are employed in fixed proportions to each other in a Leontief fashion for carrying out R&D and they are also employed in (different) fixed proportions for carrying out production.

[8] We shall eventually set $w = 1$ for normalization.

[9] Alternatively, the innovation process can be viewed as one that brings improvements along a quality dimension. Each new idea brings greater benefits to consumers without increasing the production cost of the good (and thus the market price).

[10] Given the continuous measure of varieties, the probability of the same actor successively obtaining two ideas for producing the same variety is zero

other firms that would compete away the inventor's profits. It follows that firms will immediately patent any new idea they receive. In equilibrium, then, no goods are marketed without first being patented.[11]

To encourage innovation, the government of region i offers two incentives to firms. The first type of policy rewards firms with a certain amount of money once a patent is acquired. We call this a reward policy, and the financial reward per patent is equal to S_i. The second type of policy directly lowers the cost of factor input (i.e., labor) used in R&D that pursue patents. This type of incentive policy will be called subsidy policy, and the monetary subsidy per R&D worker is equal to a fraction s_i of the wage.

We now describe the productivity level associated with each arriving high-quality (i.e. commercializable) idea. To initialize the innovation process, we assume that at the start of time ($\tau = 0$), there is a baseline technology for each variety that is freely available to all firms. We normalize the productivity of this baseline technology to be 1, so that $Z^{(0)}(j) = 1$ for all varieties $j \in [0, \mu]$. Next, we define $Z^{(k)}(j)$ to be the productivity associated with the k-th idea to arrive for variety j, where k is a non-negative integer. Thus, $\{Z^{(0)}(j), Z^{(1)}(j), Z^{(2)}(j), \ldots\}$ form a sequence of the successive best technologies for producing this variety. To describe how this frontier technology evolves, define $\zeta^{(k+1)}(j) \equiv Z^{(k+1)}(j)/Z^{(k)}(j)$ to be the productivity improvement associated with the next idea to arrive. We adopt a specification where the size of each productivity improvement is stochastic so that $\zeta^{(k+1)}(j)$ is the realization of an independent draw from an underlying stationary random variable, denoted by ζ.

We shall begin our analysis with a general distribution for ζ. Later in the paper, we will specialize into a Pareto distribution. Suppose the random variable on productivity improvement (or inventive step) ζ is drawn from the following standardized Pareto distribution with shape parameter $\theta > 1$:

$$\Pr(\zeta < z) = 1 - z^{-\theta}, \text{ where } z \in [1, \infty). \tag{1}$$

Recall that a lower θ implies a more fat-tailed distribution which places greater weight on drawing relatively large productivity improvements. This Pareto distribution will prove to make a very tractable model. It is straight-forward to show that:

$$\mathbb{E}[\zeta^{-1}] = \frac{\theta}{1+\theta} \quad \text{where } \mathbb{E} \text{ is the expectation operator,} \tag{2}$$

which we will use later in our analysis of the model.

[11] We first assume that patents do not expire, i.e., that patent duration is infinite, so that incumbent patent holders lose their monopoly power only if they are displaced by a new idea. It can be shown that the main results remain the same when there is a finite patent length.

3.2 General Equilibrium

We now embed the above cumulative innovation process in a general equilibrium setting.

3.2.1 Utility

There are n regions in this country. The utility function of the representative consumer in each region as of time 0 is given by:

$$W_0 = \int_0^\infty e^{-\rho\tau} \ln u_\tau \, d\tau. \tag{3}$$

Here, ρ is the rate of time preference, while u_τ aggregates the instantaneous utility from the consumption of differentiated varieties at time τ. Denoting the quantity of variety j consumed at time τ by $x_\tau(j)$, we assume that u_τ is given by:

$$u_\tau = \exp\left\{\int_0^\mu \ln x_\tau(j) \; dj\right\}. \tag{4}$$

The representative consumer chooses $\{x_\tau(j)\}_{\tau=0}^\infty$ to maximize (3), subject to the intertemporal budget constraint:

$$\int_0^\infty e^{-r\tau} X_\tau \, d\tau \leq b(0),$$

where $X_\tau = \int_0^\mu p_\tau(j)\, x_\tau(j)\, dj$ denotes the flow of consumption spending at time τ, with $p_\tau(j)$ being the corresponding price of variety j at that time (which the consumer takes as given). The prevailing interest rate is r, and this is determined by the rate of return for owning a patent, which is the only asset in our economy. Finally, $b(0)$ denotes the present value of the future stream of wage income that will be earned by each consumer plus the value of their initial asset holdings at date 0.

It is well known (see, e.g., Grossman and Helpman, 1991) that solving this dynamic optimization problem yields the following:

$$r = \rho + \frac{\dot{X}}{X},$$

where $\dot{X}$ is the time derivative of X. Let $E_\tau \equiv LX_\tau$ denote aggregate expenditure in this country at time τ. Thus, the above equation becomes

$$r = \rho + \frac{\dot{E}}{E}.$$

3.2.2 Market structure and profits generated by high-quality patents

Let Π_τ be the total expected profits of all firms at time τ and $(\Pi_s)_\tau$ be the expected profits per variety at time τ. Thus, $\Pi_\tau = \mu \cdot (\Pi_s)_\tau$. In Online Appendix A, we show that

$$\Pi_\tau = \left(1 - \mathbb{E}[\zeta^{-1}]\right) E = \frac{E}{(1+\theta)}$$

if we assume the Pareto distribution as stated in equation (1). Since E_τ is constant over time and ζ is drawn from a stationary distribution, the above profits are also constant over time.

3.2.3 Research Productivity of the research technology

Let the contribution to the incremental Poisson arrival rate of new patents in region i by the marginal R&D worker in region i be equal to $\lambda^i(R_i)$, where R_i is the allocation of labor to R&D in region i. We assume that $\lambda^i(R_i)$ is a convex and decreasing function that converges to zero: $\left(\lambda^i\right)'(R_i) < 0$, $\left(\lambda^i\right)''(R_i) > 0$, $\lambda^i(0) > 0$, $\lambda^i(\infty) = 0$. This is shown in Figure 1. These assumptions imply that the success rate of obtaining a patent by a marginal worker is relatively high when fewer labor resources are devoted to R&D. As more and more resources are devoted to R&D, the additional success rate contributed by the marginal worker decreases. The intuition behind this assumption is the diminishing returns to R&D effort. The earlier efforts yield a higher marginal Poisson arrival rate (because they reach for the lower-hanging fruits). The Poisson arrival rate as a function of R_i is defined as $f^i(R_i) \equiv \int_0^{R_i} \lambda^i(x)\,dx$. Note that we assume that all R&D workers are equally productive. Thus, the Poisson arrival rate of any R&D worker is a random variable with an expected value equal to the mean $\phi^i(R_i) \equiv \left(\int_0^{R_i} \lambda^i(x)\,dx\right)/R_i = f^i(R_i)/R_i$. It can be easily shown from Figure 1 that $\left(\phi^i\right)'(R_i) < 0$ as $\left(\lambda^i\right)'(R_i) < 0$. Intuitively, since the incremental Poisson arrival rate of the marginal R&D worker is lower than the average Poisson arrival rate of the existing R&D workers, the allocation of an additional unit of labor to R&D lowers the average Poisson arrival rate.

<Figure 1 about here>

In any region, conditional on the research yielding a patent, there is a probability δ that the patent is commercializable. Conditional on commmercializability, there is a probability distribution of the inventive step ζ. Refer to Appendix B for the assumed probability density function of the quality of patents (measured by the citation per year (CPY) to be defined later) generated by the research technology. The empirical facts align with the assumption of constant δ, and a constant share ($k\%$) of commercializable patents in the context of our model. We define the expected number of high-quality patents awarded in region i as $h_i(R_i) \equiv \delta f^i(R_i)$.

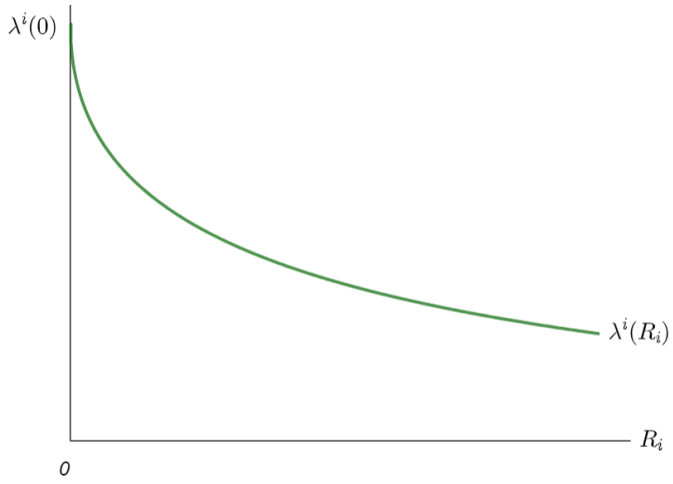


Figure 1: $\lambda^i(R_i)$ is the incremental Poisson arrival rates of patents in region i, where R_i is the allocation of labor to R&D in i.

3.2.4 Savings, Investments, and the Structure of the Economy

Since R_i is the quantity of labor in region i that are hired by firms to engage in R&D activity, the remaining quantity of labor in region i, $l_i - R_i$, engages in the production of goods, where l_i is the total amount of labor employed in region i. Note that l_i is endogenous as each region hires an amount of labor from the national labor market to satisfy the need for R&D and production of goods. In equilibrium, the demand for labor is equal to the supply of labor in the national labor market: $\sum_i l_i = L$. Firms need to obtain financing in order to hire R&D workers as their profits accrue in the future. This spending on R&D is the only form of investment in our model as the innovation is undertaken to generate ideas that yield an expected future stream of profits. There is a national capital market (financial intermediary). To obtain financing, firms issue claims (in the national capital market) on the flow of profits from their patents. These claims are the only assets in the economy (the country), and the total value of these assets at time τ is denoted by A_τ. Note that A_τ is equal to the expected value of each high-quality patent (call it $(A_s)_\tau$) multiplied by the number of varieties μ, given that we have only one active commercializable (i.e. high-quality) patent for each variety.[12] The above investment in R&D activity is financed through the savings of workers. This could occur with savers directly owning the equity of firms or with the financing channeled through an intermediary such as a bank.

Ex-post (i.e. after it is revealed whether a patent is commercializable), only high-quality (i.e. commercializable) patents have commercial value. Therefore, high-quality patents have positive asset value, $A_s > 0$, and low-quality patents have no asset value.

[12] $A_\tau = \int_0^\mu (A_s)_\tau (j) dj = \mu \cdot (A_s)_\tau$ since $(A_s)_\tau (j) = (A_s)_\tau$ for all j.

3.2.5 Factor Market Equilibrium in Each Region

The composite factor input will be allocated, through the market, to work in either (i) R&D or (ii) production. In equilibrium, the returns to the composite factor from engaging in production and R&D are equal due to free mobility of the factor and free entry into the industries. Free mobility of the composite factor implies that the returns to the composite factor, called w, are equal in all regions.

Assume that the job of an R&D worker is not only to create a new patentable idea but also prepare a patent application for the idea. The Poisson arrival rate of a patentable idea covers the successful completion of both tasks. Different provinces offer different levels of awards for approved patents and subsidies for application of patents. A monetary reward S_i is given to any firm once a patent is granted, regardless of whether it is commercializable. The government of region i provides any qualified firm with a subsidy of a fraction s_i of the wage of an R&D worker during the process of applying for a patent (regardless of whether the application is successful). We can think of ws_i as a monetary subsidy on the patent application fee. There is an important distinction between these two channels of pro-innovation policies (i.e. reward policy and subsidy policy). The patent reward is granted only after a patent is issued, while the patenting subsidy is granted to subsidize patent application.

In a competitive labor market, to obtain an interior solution of labor allocation (i.e. positive amount of labor allocated to R&D and production), the wage of an R&D worker needs to be equal to their expected "value of marginal product" (VMP), which is equal to the expected value of a patent before its commercializability is known (i.e., $\delta A_s = \delta A/\mu$ in the absence of the patent reward policy) multiplied by the average Poisson arrival rate faced by the R&D worker. Moreover, the wage of a production worker w must be equal to that of an R&D worker so that workers are indifferent between being employed in R&D or production. Recall that $\phi^i(R_i) \equiv \frac{f^i(R_i)}{R_i}$ is the expected Poisson arrival rate of an R&D worker. Therefore, in the absence of pro-innovation policies, the equilibrium labor allocation condition is given by

$$\frac{A}{\mu}\delta \cdot \frac{f^i(R_i)}{R_i} = w \quad \text{for } i = 1, 2, ..., n \tag{5}$$

where A, R_i (assuming that there is an interior solution), and w are endogenous. (We shall set $w = 1$ later as normalization.)[13]

In the presence of pro-innovation policies s_i and S_i, the expected value of a patent whose commercializability is not yet known is $\delta A_s + S_i = \delta A/\mu + S_i$, while the cost of one unit of the composite factor is $w(1 - s_i)$. Thus, the equilibrium condition that determines the amount of R&D

[13] The sufficient condition for the existence of an interior solution to (5) is $\frac{A}{\mu}\delta \cdot \lim_{R_i \to 0} \frac{f^i(R_i)}{R_i} > w$.

carried out by firms in i is given by

$$\left(\frac{A}{\mu}\delta + b_i S_i\right) \cdot \frac{f^i\left(R_i\right)}{R_i} = w\left(1 - a_i s_i\right) \quad \text{for } i = 1, 2, ..., n \tag{6}$$

where a_i is a dummy variable equal to one if there is subsidy policy in province i and zero otherwise; b_i is a dummy variable equal to one if there is reward policy in province i and zero otherwise.

3.3 Results and Implications

We are now ready to solve the model and discuss its results and implications.

3.3.1 Steady State

First, we set $w = 1$ as normalization. We shall demonstrate below that this economy will jump immediately to a steady state with $\dot{A} = \dot{X} = \dot{E} = 0$. The equations in the Online Appendix (21), (22), (23), (24), (26), and the normalization $w = 1$ together yield $4+n$ equations in $4+n$ unknowns r, Π, E, A, $\{R_i^*\}_{i=1}^n$, where a variable with an asterisk denotes the equilibrium value. In this subsection 3.3.1, we assume a standardized Pareto distribution for the productivity improvement variable to make the model more tractable.

We define the equilibrium Poisson arrival rate of high-quality patents in region i as $h_i^* \equiv \delta f^i\left(R_i^*\right)$. Note that $\frac{d}{dA}\left[h_i^*\right] > 0$ for all i, as $\frac{d\left(R_i^*\right)}{dA} > 0$ according to (6). Thus, we can write h_i^* as $h_i^*\left[\underset{(+)}{R_i^*}\left(\underset{(+)}{A}, \underset{(+)}{s_i}, \underset{(+)}{S_i}\right)\right] = h_i^*\left(\underset{(+)}{A}, \underset{(+)}{s_i}, \underset{(+)}{S_i}\right)$ for short.

With the derivation presented in section A.3.1 of the Online Appendix A, the steady state condition is given by

$$A\theta\rho + A\left(1+\theta\right) \cdot \frac{\sum_{i=1}^n h_i^*\left(\underset{(+)}{A}, \underset{(+)}{s_i}, \underset{(+)}{S_i}\right)}{\mu} = L. \tag{7}$$

Equation (7) together with (6) can be used to solve for the expected value of a high-quality patent (A/μ) as well as the equilibrium amount of R&D labor input (R_i^*). The equations will be used to carry out our comparative statics exercise in subsection 3.3.2.

3.3.2 Comparative Statics and Propositions

We carry out a number comparative statics exercises. From (7), we obtain

$$\frac{dA}{ds_i} < 0; \qquad \frac{dA}{dS_i} < 0;$$

and

$$\frac{dR_i^*}{ds_i} > 0; \qquad \frac{dR_i^*}{dS_i} > 0;$$

and

$$\frac{dR_j^*}{ds_i} < 0; \qquad \frac{dR_j^*}{dS_i} < 0 \text{ for } j \neq i;$$

and

$$\frac{d\left[\sum_{k=1}^{n} f^i\left(R_k^*\right)\right]}{ds_i} > 0; \qquad \frac{d\left[\sum_{k=1}^{n} f^i\left(R_k^*\right)\right]}{dS_i} > 0;$$

Refer to Appendix A for the proof of the above sets of inequalities. The first two sets of inequalities above indicate that home province patent subsidy policy, or home province patent reward policy will increase the allocation of resources to R&D in the home province. They have effects on A through (7), which in turn impacts R_i^* through (6). Although home subsidy and reward policies reduce the value of a high-quality patent (A/μ), which tends to lower the allocation of resources to R&D (R_i^*), such effect is more than compensated for by increased incentives to carry out R&D resulting from the home patent subsidy or patent reward policies. Thus, the effects through A are second-order. On the other hand, the third set of inequalties above indicates that home province patent subsidy or reward policy will reduce the allocation of resources to R&D in other provinces where there is no PPP, as the value of each commercializable patent falls while there is no government subsidies to offset this negative impact on the incentives to carry out R&D. Moreover, the last set of inequalities above indicates that PPP in any province will increase the total number of patents approved at the national level, despite the fact that it reduces the numbers of patents approved in provinces with no PPP. In other words, the positive effect in the province with PPP outweighs the negative effect in the provinces with no PPP. This is a subtle point. The intuition can be found in (24) in Online Appendix A. In that equation, as A falls, the profit rate per period, Π/A, increases. Thus, the hazard rate at which an active patent will be superseded, $\sum_{k=1}^{n} f^i\left(R_k^*\right)/\mu$, must increase so as to keep the expected rate of return of each asset equal to the rate of time preference ρ. This induces more entry into R&D nationally. Note that an increase in R_i^* imply three things. First, it means that the amount of R&D expenditure increases. Second, it implies that there is an increase in the Poisson arrival rate of patents, denoted by $f^i\left(R_i^*\right)$. Third, this also means increases in the numbers of high- and low-quality patents awarded per period, given by $\delta f^i\left(R_i^*\right)$ and $f^i\left(R_i^*\right)(1-\delta)$, respectively.[14] The first is about the input measure of R&D, while the third is about the output measure of R&D.

For each individual province i and at the national level, the expected share of high-quality patents is equal to δ, which is independent of PPP, as the probability of each patent being high-quality (commercializable) is assumed to be a constant δ.

Thus, we have the following two propositions.

[14] As footnote 7 indicates, "worker" in this model stands for a collection of human capital, unskilled labor, physical capital, and so on, employed in fixed proportions to each other in a Leontief manner. Thus, the R&D expenditure is the sum of the costs of all these factor inputs used in R&D.

Table 1: Comparative Statics Results

	A	R_i^*	$f^i(R_i^*)$	$R_j^*\,(j \neq i)$	$f^i(R_j^*)\,(j \neq i)$	$\sum_{k=1}^n f^i(R_k^*)$	δ
$S_i \uparrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$	$\uparrow$	No effect
$s_i \uparrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$	$\uparrow$	No effect

$S_i \equiv$ Monetary reward per patent offered in region i
$s_i \equiv$ Monetary subsidy per R&D worker as a fraction of the wage offered in region i
$A \equiv$ Economy-wide total value of high-quality (i.e. commercializable) patents
$R_i^* \equiv$ R&D expenditure in region i
$f^i(R_i^*) \equiv$ No. of patents approved in region i
$\sum_{k=1}^n f^i(R_k^*) \equiv$ No. of patents approved at the national level
$\delta =$ "Expected no. of high-quality patents approved in region i" /"Total no. of patents approved in region i"
= "No. of high-quality patents approved in the country" /"Total no. of patents approved in the country"

Proposition 1 *PPP in a province increases the R&D expenditure and the numbers of high- and low-quality patents awarded in the home province but reduces the R&D expenditure and the numbers of high- and low-quality patents awarded in other provinces where there is no PPP.*

Proposition 2 *PPP in a province increases the R&D expenditure and the numbers of high- and low-quality patents awarded at the national level.*

We summarize the comparative statics results by Table 1.

<Table 1 about here>

3.4 Welfare Analysis: Market Equilibrium and Social Optimum

In this section, we derive the condition under which the market without government intervention delivers less than the optimal amount of R&D. If that condition is satisfied, government pro-innovation policies that increases R_i will be welfare-improving if the increases are not excessive, and if the policies are financed by lump sum taxes. The result is summarized by Proposition 3. Note that even when the socially optimal level of R&D is higher than the equilibrium level, an excessive increase in R&D induced by PPP can be welfare-reducing as it lowers current consumption too much. Therefore, the dosage of PPP must strike the right balance between dynamic gains and static losses to ensure welfare improvement. Thus, there are two critical conditions for PPP to be welfare-improving. First, the socially optimal level of total R&D expenditure in China exceeds the equilibrium level (Inequality (16)). Second, the dosage of PPP cannot be too high. In this subsection, we focus on deriving the first condition.

3.4.1 Equilibrium condition

To evaluate the welfare impact of patent promotion policies, we start by analyzing the market equilibrium without PPP. Equations in the Online Appendix A, (21), (22), (23), (24), (25), and the normalization $w = 1$ together yield $4 + n$ equations in $4 + n$ unknowns r, Π, E, A, $\{R_i^*\}_{i=1}^n$.

Solving the above system of equations, we obtain the market equilibrium condition:

$$\begin{aligned} \frac{\sum_{i=1}^n R_i}{L} &= \left(1 - \mathbb{E}[\zeta^{-1}]\right) - \mathbb{E}[\zeta^{-1}]\frac{\rho\mu R_i}{Lh_i} \text{ for } i = 1, 2, ..., n \\ &= \left(\frac{1}{1+\theta}\right) - \left(\frac{\theta}{1+\theta}\right)\frac{\rho\mu R_i}{Lh_i} \text{ under Pareto distribution of } \zeta \end{aligned} \tag{8}$$

where $h_i \equiv \delta f^i(R_i)$. The derivation is explained in more detail in Online Appendix A.

3.4.2 Contrast with the social optimum

To understand the efficiency properties of the steady state, we compare the above market equilibrium with the outcomes under a benign social planner for the country as a whole. Conceptually, the social planner's problem can be formulated as a labor allocation decision over the amount of labor to employ in R&D in each province i at each time point τ, as well as the amount of labor to assign to the production of each variety $j \in [0, \mu]$ at each time point τ (which we denote by $L_\tau^p(j)$). Formally, the planner sets out to maximize the welfare of the representative consumer:

$$\max_{\{R_{i\tau}\}_{i=1}^n, \{L_\tau^p(j)\}_{j=0}^\mu} \quad W_0$$

$$\text{s.t.} \quad \int_0^\mu L_\tau^p(j)\, dj = L_\tau - \sum_{i=1}^n R_{i\tau} \quad \text{for all } \tau \geq 0, \tag{9}$$

$$\text{and} \quad Lx_\tau(j) = L_\tau^p(j) Z_\tau(j) \quad \text{for all } j \in [0, \mu] \text{ and } \tau \geq 0. \tag{10}$$

The first constraint (9) is a labor market-clearing condition that states that all labor not engaged in R&D must be employed in production. In addition, the second constraint (10) clears the market for each variety j, setting the quantity demanded equal to the quantity produced at each point in time.

As we show in Appendix section C, the solution to this social planner's problem features an equal allocation of labor to the production of each variety. In other words, given the choice of $\{R_{i\tau}\}_{i=1}^n$, we have equal allocation of production labor to all varieties: $\mu L_\tau^p(j) = L_\tau - \sum_{i=1}^n R_{i\tau}$ for all $j \in [0, \mu]$, bearing in mind that we have a measure of varieties equal to μ. Further, if we assume Pareto distribution for ζ, we have

$$W_0 = \mu \left\{ \frac{1}{\rho} \ln \left[L - \sum_{i=1}^{n} R_i \right] + \frac{\sum_{i=1}^{n} \delta f^i (R_i)}{\mu \rho^2 \theta} - \frac{\ln L}{\rho} - \frac{\ln \mu}{\rho} \right\}.$$

Under market equilibrium, since all varieties are ex ante symmetric, the law of large numbers implies that the expected price of the state-of-the-art product of all varieties are equal. Therefore, the expected output of all varieties are equal as well. Thus, we can use the above formula to calculate the welfare under market equilibrium as well.

Based on the derivation in Online Appendix A, the socially optimal value of $R_{i\tau}$ in each region i obeys the following condition.

$$\frac{\sum_{i=1}^{n} R_{i\tau}}{L_\tau} = 1 - \frac{\rho\mu}{\delta L_\tau \lambda^i (R_{i\tau}) \mathbb{E} [\ln \zeta]} \text{ for } i = 1, 2, ..., n \tag{11}$$

3.4.3 Four sources of externalities

In this subsection, we drop the time subscript for simplicity. Under market equilibrium, equation (8) implies that

$$\begin{aligned} \frac{\sum_{i=1}^{n} R_i}{\left(\sum_{i=1}^{n} h_i\right)/\mu} &= \frac{\left(1 - \mathbb{E}[\zeta^{-1}]\right) \cdot \left(L - \sum_{i=1}^{n} R_i\right)}{\mathbb{E}[\zeta^{-1}] \cdot \left[\rho + \left(\sum_{i=1}^{n} h_i\right)/\mu\right]} \quad \text{in the general case} & (12) \\ &= \frac{\left(\frac{1}{1+\theta}\right) \cdot \left(L - \sum_{i=1}^{n} R_i\right)}{\left(\frac{\theta}{1+\theta}\right) \cdot \left[\rho + \left(\sum_{i=1}^{n} h_i\right)/\mu\right]} \quad \text{under Pareto distribution of } \zeta & (13) \end{aligned}$$

On the other hand, under the social optimum for the country as a whole, (11) yields

$$\begin{aligned} \frac{\sum_{i=1}^{n} R_i}{L} &= 1 - \frac{\rho\mu R_i}{L R_i \lambda^i (R_i) \delta} \cdot \frac{1}{\mathbb{E}(\ln \zeta)} \\ &= 1 - \frac{\rho\mu R_i}{L g_i} \cdot \frac{1}{\mathbb{E}(\ln \zeta)} \end{aligned}$$

where $g_i \equiv R_i \lambda^i (R_i) \delta$. This further implies that under social optimum for the country as a whole

$$\begin{aligned} \frac{\sum_{i=1}^{n} R_i}{\left(\sum_{i=1}^{n} g_i\right)/\mu} &= \left[\mathbb{E}(\ln \zeta)\right] \frac{\left(L - \sum_{i=1}^{n} R_i\right)}{\rho} \quad \text{in the general case} & (14) \\ &= \frac{1}{\theta} \cdot \frac{\left(L - \sum_{i=1}^{n} R_i\right)}{\rho} \quad \text{under Pareto distribution of } \zeta & (15) \end{aligned}$$

Comparing the conditions for market equilibrium and the social optimum at the national level in (12) and (14) respectively, we identify four sources of externalities:

1. intertemporal spillover (positive externality) — The term under social optimum: ρ vs. the term under equilibrium: $\rho + \left(\sum_{i=1}^{n} h_i\right)/\mu$ in the denominator on the RHS of (12) and (14).

The effective discount rate that matters for the market equilibrium is $\rho + \left(\sum_{i=1}^{n} h_i\right)/\mu$, whereas that which drives the planner's problem is ρ. This difference is due to the "intertemporal spillover" effect that arises from the cumulative nature of innovation: In the market equilibrium, firms apply a higher effective discount rate when evaluating the value of a successful innovation because they do not internalize the positive knowledge spillovers that they generate for future productivity improvements. Instead, they discount the future at a higher rate that incorporates the hazard rate of being leapfrogged, $\left(\sum_{i=1}^{n} h_i\right)/\mu$. *Ceteris paribus*, this intertemporal spillover effect would push the market equilibrium towards featuring under-investment in R&D relative to the planner's outcome.

2. appropriability effect (positive externality) — The term under social optimum: $[\mathbb{E}(\ln \zeta)]$ vs. the term under equilibrium: $\left(1 - \mathbb{E}[\zeta^{-1}]\right)$ in the numerator on the RHS of (12) and (14).

The term $1 - \mathbb{E}[\zeta^{-1}]$ appears in the numerator on the right-hand side of (13), whereas we instead have a $\mathbb{E}[\ln \zeta]$ term in the corresponding equation (15) for the social planner. This contrast captures the "appropriability" effect. With the log utility specification in (3), a successful innovation drawn from ζ yields a (flow) welfare gain per unit of the good consumed equal to $\mathbb{E}[\ln \zeta]$. However, the profit share that accrues to the innovating firm is only $1 - \mathbb{E}[\zeta^{-1}]$ (c.f., equation (22)). As $\mathbb{E}[\ln \zeta] > 1 - \mathbb{E}[\zeta^{-1}]$ except in the special case where ζ is a point mass at 1, firms are indeed unable to appropriate the full surplus enjoyed by society from their innovation.[15] Thus, this effect also pushes the market equilibrium towards a lower allocation of labor to research than the planner would choose.

3. business-stealing effect (negative externality) — The term under social optimum: 1 vs. the term under equilibrium: $\frac{1}{\mathbb{E}[\zeta^{-1}]}$ on the RHS of (12) and (14).

Since innovation effort erodes the profits of preceding innovators in a way that the social planner would want to fully internalize. We associate this effect with the $\frac{1}{\mathbb{E}[\zeta^{-1}]}$ term that appears in (12), but which is absent in (14). The latter reflects the fact that aggregate profits in the economy are still constant, despite the new innovation occurring, whereas the firm that owns this innovation would value the private profit flows that it generates. As the support of ζ is $(1, \infty)$, we clearly

[15] To see this, straightforward calculus will show that the inequality $\ln x \geq (x-1)/x$ holds for all x, with equality if and only if $x = 1$. Taking expectations with respect to the probability density function of ζ yields the desired inequality. This inequality will moreover be strict except in the special case where ζ is a point mass at 1.

have $\frac{1}{\mathbb{E}[\zeta^{-1}]} > 1$, so that this business-stealing effect pushes the private actors in the direction of over-investing in research relative to the planner's problem.[16]

4. congestion effect (negative externality) — The term under social optimum: $\left(\sum_{i=1}^{n} g_i\right)/\mu$ in the denominator of LHS of (12) vs. the term under equilibrium: $\left(\sum_{i=1}^{n} h_i\right)/\mu$ in the denominator on the LHS of (12).

The addition of one more researcher reduces the average arrival rate of ideas over all researchers (See, for example, p.100 of Jones (2002)). This is analogous to the effect that the addition of one more car in a highway slows down the average speed over all cars in the highway. The negative externality is reflected in the fact that $g_i(R_i) < h_i(R_i)$ for all i if we plug the equilibrium value of R_i into $g_i(R_i)$ and $h_i(R_i)$.[17]

As is shown in the comparison between (13) and (15), under Pareto distribution, effects 2 and 3 cancel each other, leaving only effects 1 and 4. Under what condition will the net externality be positive so that the social optimum features a higher value of $\sum_{i=1}^{n} R_i$ than under equilibrium? It is straightforward to show that (13) implies that, under equilibrium,

$$\left\{\sum_{i=1}^{n} R_i\right\}^{eqm} = \frac{L\left(\sum_{i=1}^{n} h_i\right)}{\left(\rho\mu + \sum_{i=1}^{n} h_i\right)\theta + \sum_{i=1}^{n} h_i},$$

whereas (15) implies that, under social optimum,

$$\left\{\sum_{i=1}^{n} R_i\right\}^{SO} = \frac{L\left(\sum_{i=1}^{n} g_i\right)}{\rho\theta\mu + \sum_{i=1}^{n} g_i}.$$

Comparing the above two equations, it can be easily shown that

$$\begin{aligned}
&\left\{\sum_{i=1}^{n} R_i\right\}^{SO} > \left\{\sum_{i=1}^{n} R_i\right\}^{eqm} \\
\iff\ &\frac{L\left(\sum_{i=1}^{n} g_i\right)}{\rho\theta\mu + \sum_{i=1}^{n} g_i} > \frac{L\left(\sum_{i=1}^{n} h_i\right)}{\left(\rho\mu + \sum_{i=1}^{n} h_i\right)\theta + \sum_{i=1}^{n} h_i} \text{ when } R_i \text{ is the equilibrium value} \\
\iff\ &\frac{1}{\sum_{i=1}^{n} g_i} - \frac{1}{\sum_{i=1}^{n} h_i} < \frac{1}{\rho\mu} \text{ when } R_i \text{ is the equilibrium value} \qquad (16)
\end{aligned}$$

In order to gain insights into the condition under which (16) is more likely to hold, we introduce more structure to the function $\lambda^i(R_i)$ by assuming that $\lambda^i(R_i) = \beta^i(R_i + 1)^{-\gamma}$, where β^i, $\gamma > 0$.[18]

[16] The presence of monopoly-pricing power *per se* does not distort labor allocations in our model. The reason is that all firms charge the same markup in expectation (drawn from the standardized Pareto distribution, $\eta(m)$), so that the allocation of production labor across varieties cannot be improved upon *ex ante*. See the related discussion in Grossman and Helpman (1991), p.70.

[17] $\frac{f^i(R_i)}{R_i} > \lambda^i(R_i)$ for all $R_i > 0$ as long as $\lambda^i(R_i)$ is downward sloping.

[18] This functional form guarantees that $\lambda^i(0) \neq \infty$.

In Online Appendix section A.4.3, we show that a sufficient condition for inequality (16) is

$$\left(\frac{\gamma}{1-\gamma}\right)\frac{1}{\text{hazard rate}} < \frac{1}{\rho}. \tag{42}$$

When (16) holds, small increases in R&D expenditure induced by patent promotion policies financed by lump sum taxes will be welfare improving. Thus, we have Proposition 3 below.

Proposition 3 *Inequality (16) gives a sufficient condition for the market-determined level of R&D expenditure in a country to be less than the socially optimal level. When that condition is satisfied, sufficiently small increases in R&D expenditures by firms induced by pro-innovation policies financed by lump sum taxes are welfare improving.*

Inequality (42) implies that it is more likely to hold if the elasticity of incremental Poisson arrival rate with respect to R&D input, γ, is smaller, the country is more patient (has lower discount for the future, ρ), and has higher rate of replacement of old technology by new technology (higher hazard rate). In that case, beyond the establishment of an effective patent system, additional subsidy and reward policies to promote patenting will be welfare-improving.

To summarize, there are four sources of externalities, two positive and two negative. Under the assumption of Pareto distribution of the inventive step of commercializable patents, one of the positive externalities (approriability effect) exactly offset one of the negative externalities (business-stealing effect), leaving behind just one positive (intertemporal spillovers effect) and one negative externalities (congestion effect). When the congestion externality is sufficiently weak, as indicated in Inequality (16), the socially optimal level of national R&D is higher than the non-intervention market-determined level, making patent subsidies and rewards potentially welfare-improving if the magnitude of subsidies are not too high.

4 Empirical Analysis

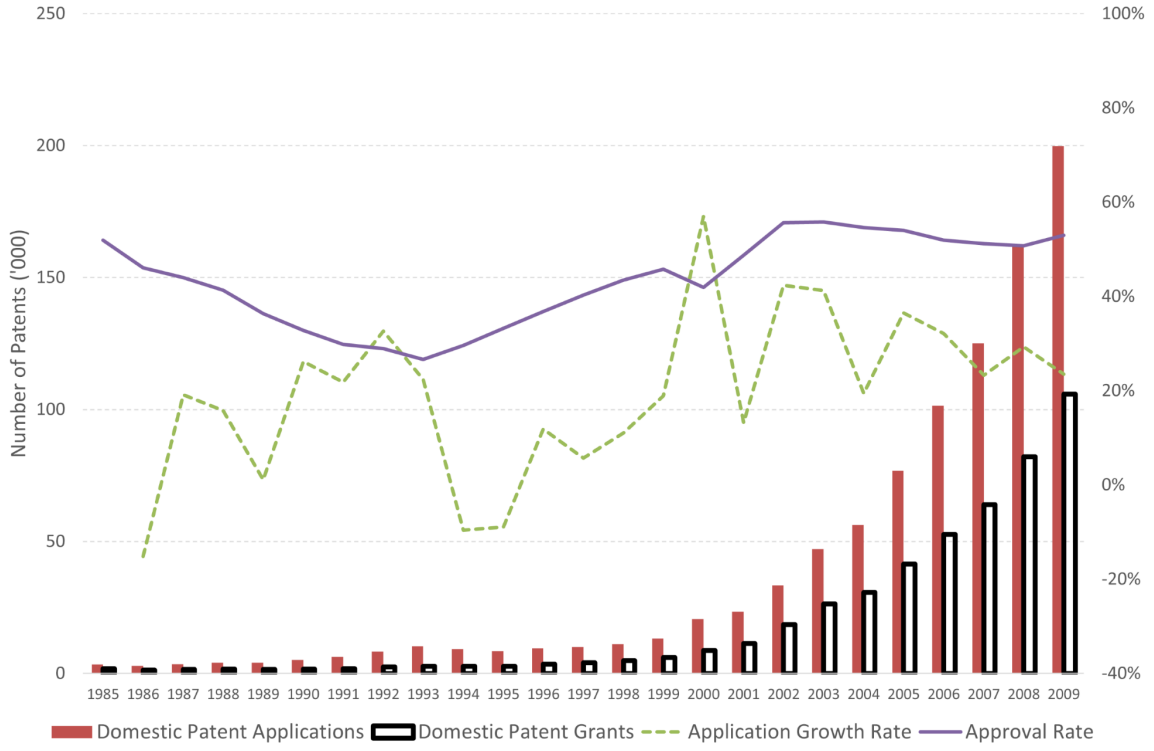
4.1 Data Sources and Description

Three data sets are used in this study. The first dataset contains information on invention patents filed by Chinese entities between 1985 and 2009 obtained from the China National Intellectual Property Administration (CNIPA). For each patent, the dataset contains 44 dimensions, such as patent application number, filing date, publication date, filing location, applicant name, industry classification, and grant and termination dates if the patent was granted or terminated. In this paper, we only focus on invention patents because they are more innovation-intensive and more technologically important compared with utility model patents and design patents. By focusing on

invention patents, we aim to capture a subset of patents that represents a more concentrated source of technological advancement and innovation. Furthermore, our choice to analyze only invention patents aligns with previous research in this field. Notably, studies such as Long and Wang (2016), Dang and Motohashi (2015), and Li (2012) have focused on invention patents in their investigations. Similarly, when examining patents globally, researchers such as Smith (2020) and Moser and Voena (2012) have also primarily concentrated on invention patents. By following this established practice, we ensure consistency and comparability with existing literature while building upon prior findings.

In this paper, we use the average number of citations per year (CPY) to measure patent quality. CPY is calculated by dividing the total number of forward citations accumulated from the publication year of the patent to 2023 by the number of years. As forward citations refer to the number of times a particular patent is cited by subsequent patents, they tend to increase over the age of a patent. In other words, ceteris paribus, older patents tend to have more citations than newer ones. It is thus appropriate to take the average annual citations per year since the patent publication. In terms of data source, we obtained the CPY count for each patent by extracting the citations from Google Patents[19]. The data was downloaded in August 2023. The data set consists of nearly 1.98 million patents filed between 1985 and 2009 worldwide.

Figure 2: Chinese invention patent applications and approvals by CNIPA from 1985 to 2009



Since 1985, the number of applications filed by Chinese entities has been steadily growing each

[19] https://patents.google.com/. The data was downloaded in August 2023

Table 2: Costs of Applying for Patents with CNIPA. Source: CNIPA

Description	**Fee Amount (RMB)**
Application Fee	900
Claim surcharges for over 11 claims	150 per claim
Documentation surcharge	50 over 31 pages and 100 over 301 pages
Early Publication surcharge	80
Substantive Examination Fee	2500
Re-examination Fee	1000
Patent Certificate Fee	200

Table 3: Annual Maintenance Costs for a Patent with CNIPA. Source: CNIPA

Year After Patent Grant	1-3	4-6	7-9	10-12	13-15	16-20
Annual Maintenance Fee (RMB)	900	1200	2000	4000	6000	8000

year, with an average annual increase of approximately 16.8%. The annual approval rate has ranged between 40.4% and 65.6% throughout this time period. Of these patents, we mainly focus on the dynamics of the 0.96 million patents filed within China because China's pro-innovation policies only affect these patents. Figure 2 gives a summary of the patents filed within China. Their average annual growth rate of applications is slightly higher at approximately 20% per year, but the annual approval rates are slightly lower, ranging between 26.6% and 55.8%. The costs of applying for a patent with CNIPA are listed in Table 2, and the annual maintenance cost is given in Table 3.

<Figure 2, Table 2, Table 3 about here>

The second data set contains information on provincial pro-innovation policies in China. To identify these policies, we did an extensive search over various local governments' websites and the Annals of Chinese Intellectual Property Rights (2000-2008).[20] These data sources provide a comprehensive overview of the various legislations in the area of patent policy. Tables 4 and 5 list the policies adopted by various provinces, their respective adoption years, and the coverage.[21] Table 4 presents the information related to the adoption of reward policies[22], which are denoted by S_i in our theoretical model. From now on, "province" shall mean "province/municipality" for simplicity, unless otherwise stated. From 1995 to 2015, 10 provinces and three municipalities implemented 13 qualifying reward policies, mainly in the form of tax refunds and deductions. Both coastal regions and inland provinces were represented. Table 5 lists the subsidy policies of various provinces, which are represented by s_i in our theoretical model[23]. These policies were implemented from 1999 to

[20] These data sources have been used in other papers like Dang and Motohashi (2015).

[21] These tables are consistent with the policy information presented in Li (2012) and Dang and Motohashi (2015).

[22] Our reward policies correspond to "grant-contingent rewards" in Dang and Motohashi (2015).

[23] Our subsidy policies correspond to the policies in Li (2012) and a combination of "filing fee subsidies" and "examination fee subsidies" in Dang and Motohashi (2015).

Table 4: Provincial patent reward policy (RP_i)

Start Year	**Province**	**Specific Policy**	**Patent type coverage**	**Applicant Type**
1995	Guangdong	Tax refund	Invention and utility	Firms
1997	Liaoning	Tax refund	Invention and utility	Firms
1998	Hubei	Tax refund	Invention and utility	Firms
1999	Jilin	Tax refund	Invention and utility	Firms
2000	Shanghai	Rewards	All patents	All applicants
2005	Beijing	Tax incentives	All patents	All applicants
2005	Anhui	Tax deduction	All patents	All applicants
2007	Chongqing	Tax incentives	All patents	Firms
2009	Jiangxi	Tax incentives	All patents	All applicants
2009	Jiangsu	Tax deduction	All patents	Firms
2009	Qinghai	Tax deduction	All patents	Firms
2011	Tianjin	Tax incentives	All patents	All applicants
2015	Gansu	Rewards	Invention and utility	All applicants

2007 by 28 different provinces and municipalities, both coastal and inland. The policies of each province remained in place each year after their implementation, although minor changes in policy details could occur over time.[24]

<Tables 4 and 5 about here>

The third data source is the China Statistical Yearbook on Science and Technology, where we obtained data for R&D Expenditures and Science and Technology (S&T) Expenditures in each province. R&D Expenditures consist of spending on Basic Research, Applied Research, and Experimental Development by Institutions, Large & Medium-sized Enterprises, and Higher Education

[24] We appreciate an anonymous reviewer's comments and it would be ideal to construct a treatment variable using reward and subsidy data on specific patent. However, to our best knowledge, publicly available data of this nature is not yet available for studies in the extant literature. Refinement can be made upon future availability of data.

Table 5: Provincial patent subsidy policy (SP_i)

Start Year	**Province**	**Specific Policy**	**Patent Type Coverage**
1999	Shanghai	Application and maintenance fees	Cover all
2000	Beijing	Application fees	Cover all
2000	Tianjin	Application fees	Invention, utility
2000	Jiangsu	Application fees	Cover all
2000	Chongqing	Application fees	Cover all
2001	Zhejiang	Application fees	Invention
2001	Heilongjiang	Application fees	Cover all
2001	Guangxi	Application fees	Cover all
2001	Hainan	Application fees	Invention
2001	Sichuan	Application fees	Invention, utility
2002	Fujian	Application fees	Cover all
2002	Jiangxi	Application fees	Invention, utility
2002	Henan	Application fees	Cover all
2002	Guizhou	Application fees	Invention
2002	Inner Mongolia	Application fees	Cover all
2002	Xinjiang	Application fees	Cover all
2003	Shaanxi	Application fees	Cover all
2003	Anhui	Application fees	Cover all
2003	Shandong	Application fees	Invention
2003	Shanxi	Application fees	Cover all
2003	Yunan	Application and maintenance fees	Cover all
2003	Tibet	Application fees	Cover all
2004	Jilin	Application fees	Invention
2004	Hunan	Application and maintenance fees	Cover all
2005	Hebei	Application fees	Invention
2005	Qinghai	Application fees	Cover all
2006	Liaoning	Application fees	Invention
2007	Hubei	Application fees	Cover all

facilities. The real R&D Expenditures and real S&T Expenditures are obtained by deflating the nominal values with the 1985 prices. The data for R&D Expenditures are available from 1998 onwards, while the data for S&T Expenditures spans over a longer time frame from 1991 to 2008. Therefore, the two datasets overlap during 1998-2008. An extrapolation approach is used to extrapolate the values of R&D Expenditures backward to earlier years in 1991-1997 based on its strong correlation with the S&T Expenditures. Details are provided in Appendix D.

Findings from World Intellectual Property Organization (WIPO (1996)) reported that the percentage of patents that were ultimately commercialized was about 5-7 percent. Data from the US Patents and Trademark Office (USPTO) also shows that the distribution of citations per patent is severely right skewed. Therefore, patents in the highest percentiles have significantly higher citations per patent than the rest of the sample. A cutoff at the top 10th percentile was thus selected to distinguish high-quality patents (or 'highly cited' patents) from the rest of the sample (Mudambi & Swift, 2014). In this paper, patents falling into the top ten percent average CPY among all patents awarded in China in that year are defined as high-quality patents. They spread across different provinces, ranging from 0.02 percent of all patents granted in Tibet to 23.4 percent in Beijing.

4.2 Empirical Welfare Analysis

We go on to calibrate the model and carry out counterfactual exercises. In this section, we analyze the probability that PPP had been welfare-improving in China during 1995-2009.

First, we try to find out how likely inequality (16) holds in China. Following section 3.4.3, we assume that

$$\begin{aligned} \lambda_{it}(R_{it}) &= \beta_{it}(R_{it}+1)^{-\gamma} \\ \text{which implies that} \qquad f_{it}(R_{it}) &= \frac{\beta_{it}}{1-\gamma}\left[(R_{it}+1)^{1-\gamma}-1\right]. \end{aligned}$$

Recall Equation (16), a sufficient condition for the equilibrium level of $\sum_i R_{it}$ to be lower than the socially optimal level:

$$\frac{1}{\sum_{i=1}^n g_{it}} - \frac{1}{\sum_{i=1}^n h_{it}} < \frac{1}{\rho\mu_t} \qquad \text{where } h_{it} \equiv \delta f_{it}(R_{it}); \qquad g_{it} \equiv \delta R_{it}\lambda_{it}(R_{it})$$

In Online Appendix A, we show that a sufficient condition of (16) is (42):

$$\left(\frac{\gamma}{1-\gamma}\right)\frac{1}{\text{hazard rate}} < \frac{1}{\rho}$$

where the hazard rate is the rate of complete capital loss of an asset (which is a patent) due to its being replaced by a more productive new patented technology.[25]

[25] By definition,

$$\text{hazard rate} = \frac{\delta\sum_{i=1}^n \beta_{it}\left[(R_{it}+1)^{1-\gamma}-1\right]}{(1-\gamma)\mu_t}.$$

We use (42) to guide us to find the combination of parameters that gives rise to the result that the equilibrium level of $\sum_i R_{it}$ is lower than the socially optimal level. We note that (42) will be more likely satisfied when θ is large (which makes hazard rate/ρ large) and γ is small. In addition to (42), PPP is welfare-impoving only when s (the fraction by which PPP reduces the cost of R&D) is small. Furthermore, the values of the exogenous variables and the parameters used have to be plausible. They have to be validated by the data, external sources or other studies in the literature.

We assume different combinations of the values of s, γ, δ, θ and ρ, and compute the probability that PPP was welfare-improving in China in different years. To do that, we first calibrate the model year by year and then carry out counterfactual exercises year by year. The values of s, δ, ρ, γ and θ are assumed to be constant over time. In addition, the set of plausible values of s, δ, ρ, and γ are invariant over time. In order to fit the model with the data, we allow the set of plausible values of θ as well as the values of β_i and μ to vary from year to year. So, β_i should be denoted by β_{it} and μ denoted by μ_t. The parameter μ_t is estimated based on the data and our model. R_{it} is from the data. The parameter β_{it} is chosen so as to make $f_{it}(R_{it})$ match the number of patents awarded in province i at year t. In Online Appendix C, we explain how we choose the range of plausible or permissible values of each of these parameters. After some careful analyses, we determine that the plausible values are: $s = 0.1$ and 0.2; $\rho = 0.03$, 0.05 and 0.07; $\delta = 0.03$, 0.05 and 0.1; $\gamma = 0.6$, 0.7, or 0.8; $\theta =$ slightly greater than one (it cannot be less than or equal to 1; sometimes we use 1.05 to indicate it) to a maximum of no more than 20. Our sample includes the years 1995-2009, and all 31 provinces of China.

For each year, we carry out counterfactual exercises by counterfactually removing the PPP in the provinces where PPP were implemented. We calculated the impacts on the values of R_{it} in those provinces. We find that condition (16) holds for a non-trivial share of the plausible parameter space. We then calculate the impact on the welfare of China from the following equation, which is also given in Appendix section C.

$$W_0 = \mu_t \left\{ \frac{1}{\rho} \ln \left[L_t - \sum_{i=1}^{n} R_{it} \right] + \frac{\sum_{i=1}^{n} \delta f^i(R_{it})}{\mu \rho^2 \theta} - \frac{\ln L_t}{\rho} - \frac{\ln \mu_t}{\rho} \right\}$$

In Online Appendix B, we describe how we calibrate our model using the data on R&D in the knowledge- and technology-intensive (KTI) sectors in each province in each year R_{it}, the number of patents approved in each province in each year M_{it}, the value-added in the KTI sectors in China in each year L_t, and the values of the parameters s, γ, δ, θ and ρ. Then we describe how we carry out the counterfactual exercise of removing PPP in the provinces in which PPP have been implemented in each year.

Below we explain our findings. First, we note that if we assume that PPP reduces the cost of R&D by 10% (i.e. $s = 0.1$), whenever condition (16) holds, PPP is almost always ($> 90\%$)

welfare-improving. This is simply because when (16) holds any small subsidy to R&D will be welfare-improving. In this case, $s = 10\%$ seems to be a sufficiently small subsidy. If we assume that $s = 0.2$, it is still true that PPP is welfare-improving when (16) holds for a vast majority of cases ($> 84\%$). This indicates that $s = 20\%$ seems to be a sufficiently small subsidy as well. Now, let us refer to Figures 3 to 5. Note that all of Figures 3 to 5 are based on $\rho = 0.05$ and $\delta = 0.1$, though the values of ρ and δ do not affect the charts at all. Please refer to Online Appendix B for a detailed explanation.

In Figure 3 (a), the green area indicates combinations of γ and θ for which PPP leads to welfare improvement in the year 2009 when $s = 0.1$, while the red area indicates that PPP leads to welfare deterioration. As explained in Online Appendix section C, we assume that it is equally likely that γ is equal to 0.6, 0.7, or 0.8, and these values consist of all the plausible values of γ. Likewise, we assume that all the values of θ in the plausible parameter space in the chart are equally likely and these are exhaustive of all the plausible values. Therefore, the area of the green zone divided by the total area of the parameter space (i.e. the green area plus the red area) is the probability that PPP is welfare improving in 2009 when $s = 0.1$. This number turns out to be 33.5%.

<Figures 3 (a) and (b) about here.>

Figure 3 (b) is similar to Figure 3 (a) but for the year 2002 and $s = 0.1$. Applying the same principle, we calculate that the probability that PPP is welfare improving in 2002 when $s = 0.1$ is 28.4%.

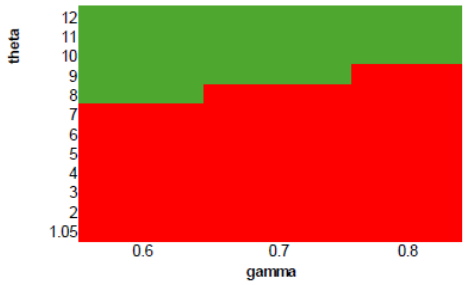
For each of 2002 and 2009, the corresponding chart for $s = 0.2$ will show the same pattern as for $s = 0.1$ given by Figures 3 (a) and 3 (b), but the share of the green area will be smaller, as it is less likely for the subsidy to be welfare improving when it is larger. It turns out that the probability that PPP is welfare improving in 2009 when $s = 0.2$ is 28.7%, while that probability in 2002 when $s = 0.2$ is 16.7%.

In Figure 4 (a), the green area indicates all the year-θ combinations for which PPP is welfare improving when $\gamma = 0.6$ and $s = 0.1$, while the red area indicates that PPP leads to welfare deterioration. For any given year, the area of the green zone divided by the total area of the green and red zones is the probability that PPP is welfare improving in that year when $\gamma = 0.6$ and $s = 0.1$. For example, for the year 2002, the probability is $7/20 = 35.1\%$, and for the year 2009, the probability is $5/12 = 41.8\%$.

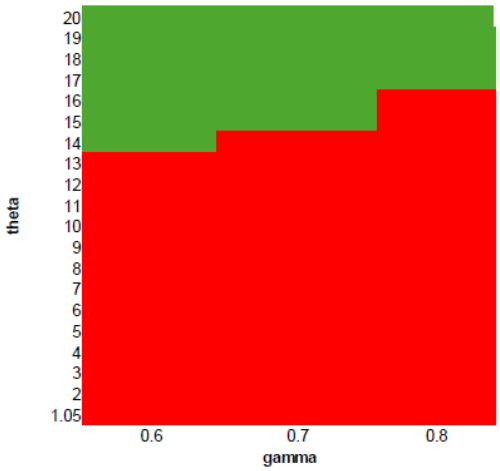
<Figures 4 (a) and (b) about here.>

Figure 4 (b) is similar to Figure 4 (a), but for $\gamma = 0.8$ and $s = 0.1$. As expected, for any given year, the probability that PPP is welfare improving is smaller than that when $\gamma = 0.6$ and $s = 0.1$, as a larger γ makes it less likely that the socially optimal level of national R&D is larger than the equilibrium level, as indicated in inequality (42). For example, the probability that PPP

Figure 3: The probability that PPP was welfare-improving for (a) Year 2009 and s=0.1 and (b) Year 2002 and s=0.1.The green area indicates that PPP was welfare-improving, while the red area indicates that PPP was welfare-reducing. The probability that PPP is welfare-improving is equal to the green area divided by the sum of the green area and red area. The values of rho and delta do not affect the chart at all.

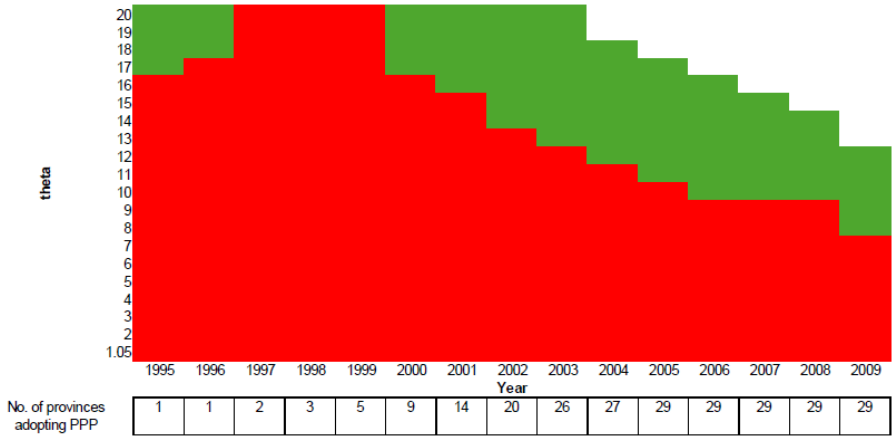


(a) Year 2009 and $s = 0.1$.

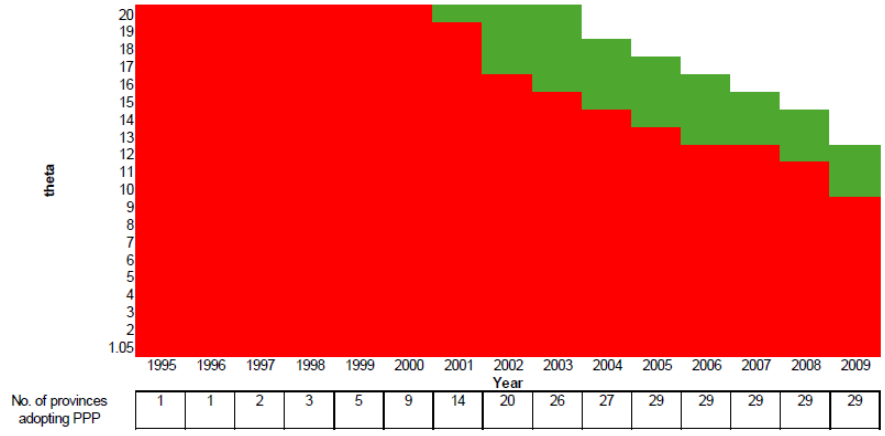


(b) Year 2002 and $s = 0.1$.

Figure 4: The probability that PPP was welfare-improving for (a) gamma=0.6 and s=0.1 and (b) gamma=0.8 and s=0.1. The green area indicates that PPP was welfare-improving, while the red area indicates that PPP was welfare-reducing.The probability that PPP is welfare-improving during 1995-2009 is the weighted average of the probabilities across all years, with the weight in each year being the number of provinces adopting PPP in that year. The probability that PPP is welfare-improving in a year is equal to the green area divided by the sum of the green area and the red area for that year. The values of rho and delta do not affect the chart at all.



(a) $\gamma = 0.6$ and $s = 0.1$



(b) $\gamma = 0.8$ and $s = 0.1$

is welfare improving in 2002 when $\gamma = 0.8$ and $s = 0.1$ is 4/20=20%, and the probability for 2009 is $3/12 = 25\%$.

For either $\gamma = 0.6$ or $\gamma = 0.8$, the corresponding chart for $s = 0.2$ will show the same pattern as for $s = 0.1$ given by Figures 4 (a) and (b), but the share of the green area will be smaller, as it is less likely for the subsidy to be welfare improving when it is larger. For example, the probability that PPP is welfare improving in 2002 when $\gamma = 0.8$ and $s = 0.2$ is $2/20 = 10\%$, and for 2009 the probability is $3/14 = 21.4\%$.

Figure 5 is a 3-D plot in which the space sandwiched between the blue surface and the green surface indicates all combinations of year-γ-θ for which PPP is welfare-improving, while the space sandwiched between the green surface and the red horizontal surface at the bottom indicates the combinations of year-γ-θ for which PPP is welfare-reducing. Note that Figures 3 (a) and 3 (b) are in fact slices of Figure 5 at years 2009 and 2002 respectively while Figures 4 (a) and (b) are slices at $\gamma = 0.6$ and $\gamma = 0.8$, respectively.

<Figure 5 about here.>

Table 6 gives more details about the results of our probability estimations. Note that the probability that PPP was welfare-improving is different for different combinations of year, γ and s. This is because L_t, $\{M_{it}\}$, $\{R_{it}\}$ and the set of provinces that adopted PPP were different for different years. Thus, for the same γ and s, the plausible range of θ for which the model is consistent with the data is different for different years. The minimum plausible value of θ is always a value slightly larger than 1 regardless the year-γ-s combination, as any value of θ less than or equal to 1 implies that the mean of the inventive step ζ is equal to infinity and the expected cost markup is infinity.[26] On the other hand, the maximum value of θ varies from year to year. Conditional on L_t, $\{M_{it}\}$, $\{R_{it}\}$, the maximum plausible θ corresponds to a hazard rate equal to infinity. Furthermore, out of the plausible range of θ, the range of θ for which PPP is welfare-improving is also different for different years, as condition (42) stipulates that, for PPP to be welfare-improving in that year, there is a restriction on the hazard rate. This in turn poses a restriction on θ. Therefore, the probability that PPP is welfare-improving is different for different years, even if we assume the same plausible ranges of γ and s for all years.

<Table 6 about here.>

In Table 6, the probability that PPP is welfare-improving for the period 1995-2009 is calculated as the weighted average of the probabilities across all the years, with the weight for each year being the number of provinces adopting PPP in that year. Note that PPP became more widespread in the later years: the years 2002-2009 accounted for 86% of province-years adopting PPP (218 province-years of PPP out of a total of 253) during the sample period. In the interest of space, we

[26] Recall that ζ is a random variable with Pareto distribution and shape parameter θ.

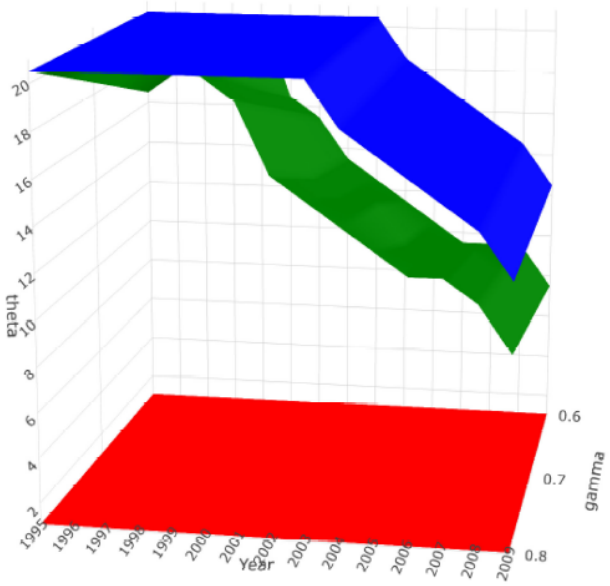


Figure 5: 3-D Plot of θ against γ and year when $s = 0.1$. The blue surface is the maximum permissible value of θ corresponding to each year-gamma combination. The space between the blue surface and the green surface is all the combinations of year-γ-θ for which PPP is welfare-improving. The space between the green surface and the red surface is all the combinations of year-γ-θ for which PPP is welfare-reducing. The values of ρ and δ do not affect the chart at all. Figures 3 (a) and (b) are slices of this 3-D figure at Years 2009 and 2002 respectively while Figures 4 (a) and (b) are slices at $\gamma = 0.6$ and $\gamma = 0.8$, respectively.

Table 6: Probability that PPP is Welfare-Improving in Various Years.

Year	$\gamma = 0.6$	$\gamma = 0.6$ or 0.7	$\gamma = 0.6$ to 0.8	Number of Year-Province Adopting PPP
1995-2001	12.82%	9.17%	6.44%	35
2002	30.08%	26.32%	22.56%	20
2003	35.09%	31.33%	27.57%	26
2004	37.04%	31.75%	28.22%	27
2005	39.12%	36.32%	31.67%	29
2006	41.44%	35.52%	31.57%	29
2007	39.57%	33.44%	29.35%	29
2008	36.73%	31.80%	27.92%	29
2009	38.84%	34.96%	31.07%	29
1995 to 2009	34.14%	29.68%	25.90%	253

Note: The table shows the probability that PPP is welfare-improving in each year as well as aggregated over all years during 1995-2009, assuming that s=0.1 and s=0.2 are equally probable. The probability that PPP is welfare-improving during 1995-2009 is the weighted average of the probabilities across all years, with the weight in each year being the number of provinces adopting PPP in that year. The probability that PPP is welfare-improving in a year is equal to the green area divided by the sum of the green area and the red area for that year. The values of ρ and δ do not affect the probabilities at all.

group together 1995-2001 as one single "year". In Online Appendix section E, we show a table with more detailed breakdown of the probabilities that PPP is welfare-improving in different years and under different assumptions about s and γ. Let us now focus on the most important result: During the sample years 1995-2009, assuming that it is equally likely that $s = 0.1$ or 0.2, and equally likely that γ is 0.6, 0.7 or 0.8, the probability that PPP is welfare-improving is 25.9%, which is non-trivial.[27] Thus, we have the following proposition.

Proposition 4 *There is a non-trivial probability that PPP had been welfare-improving in China during 1995-2009.*

Despite Proposition 4, however, we found that the expected value of the welfare change caused by PPP is negative in all years. The simplest way to demonstrate this is to assume that $\rho = 0.03$, $\delta = 0.1$, $s = 0.1$ and $\gamma = 0.6$ and calculate the expected value of the welfare impact of PPP in each year. This combination of $\{\rho,\ \delta,\ s,\ \gamma\}$ gives rise to the highest probability that PPP is welfare-improving and therefore it is the scenario most favorable to delivering the result that the expected welfare impact of PPP was positive. Even with this combination, however, Table 7 shows that the expected value of the welfare change caused by PPP was negative in all years. The negative

[27] If we assume that $s = 0.2$ only, the probability that PPP is welfare-improving is 23.5%, which is still non-trivial.

Table 7: Expected welfare impact per province adopting PPP in Various Years

	$s = 0.1$	
Year	$\gamma = 0.6$	Number of year-province adopting PPP
1995	-852	1
1996	-1117	1
1997	-1820	2
1998	-2544	3
1999	-3497	5
2000	-9115	9
2001	-13626	14
2002	-22980	20
2003	-31574	26
2004	-43255	27
2005	-61809	29
2006	-82192	29
2007	-107228	29
2008	-137177	29
2009	-177101	29
1995 to 2009	-46393	253

Note: This table shows the expected welfare impact per province for provinces that adopted PPP in each year as well as the average over all years during 1995-2009. It is assumed that $\rho = 0.03$ and $\delta = 0.1$.

expected value of the welfare impact arises from the fact that it is dominated by the negative welfare impact of PPP when θ is small. In other words, it is likely that the congestion externality dominates the knowledge spillovers externality so that the net externality is negative. That is, there were already too much resources allocated to R&D before the implementation of PPP. As shown in Table 7, the magnitude of the negative welfare-reduction of PPP per province increased over time. This is due to the increasing values of L_t, $\{M_{it}\}$, $\{R_{it}\}$ over time. As the real value-added in the KTI sector, the number of patents awarded and the R&D expenditure in the KTI sector all increased over time, they caused the real absolute amount of patent promoation subsidies to increase over time, and so the welfare loss caused by the subsidies increased over time as well.

<Table 7 about here>

Thus, we have the most important proposition of this paper:

Proposition 5 *The expected value of the welfare change caused by provincial patent promotion policies during 1995-2009 was negative. This negative welfare change occurred every year and it was increasing over time.*

5 Conclusion

China's rapid ascendancy to a nation with the largest number of annual patent applications and approvals has received much international attention. This paper provides a structural framework to study, both theoretically and empirically, the effectiveness of provincial patent promoting policies (PPP) in stimulating provincial R&D expenditure and patent production, and their impacts on national welfare.of China during 1995-2009.

We construct a tractable general equilibrium endogenous growth model that features the presence of two positive and two negative externalities. This implies that the decentralized provincial market equilibrium delivers a sub-optimal national allocation of resources to R&D activity that is strictly lower than the national social planner's benchmark if the net externality is positive, which is true under certain condition. If this condition holds, it justifies the use of a sufficiently small dosage of patent-subsidy and reward policy to increase the rate of innovation and close the wedge. Assuming a Pareto distribution for the inventive step of commercializable patents, one positive externality (the appropriability effect) exactly offsets one negative externality (the business-stealing effect), leaving behind a single positive externality (the intertemporal spillovers effect) and a single negative externality (the congestion effect). If the congestion externality is sufficiently weak and the intertemporal externality is sufficiently strong, the socially optimal level of national R&D input exceeds the market-determined level without government intervention. We derive a condition under which this is true. Consequently, patent subsidies and rewards have the potential to improve welfare if the magnitudes of subsidies are sufficiently moderate.

According to the above condition, it is plausible that it is more likely for patent promotion policies to be welfare-improving in fast-growing low- and middle-income countries like China than in the high-income countries, given that the former is more likely to have more low-hanging fruits available (the elasticity of incremental Poisson arrival rate with respect to R&D input, γ, is smaller), more patience (lower discount for the future, ρ) which puts higher weight on rate of innovation and lower weight on current consumption, and higher hazard rate faced by each patent (i.e. rate of replacement of old technology by new technology). Thus, beyond the establishment of an effective patent system, additional patent subsidy and patent reward policies to promote patenting is likely to be welfare-improving in these countries.[28]

We calibrate the model with the data and external sources, and then carry out counterfactual exercises to evaluate the welfare impact of PPP in China during 1995-2009. This empirical welfare

[28] Even under such circumstance, an excessive increase in R&D induced by PPP can be welfare-reducing as it lowers current consumption too much. Therefore, the dosage of PPP must strike the right balance between dynamic gains and static losses to ensure welfare improvement. Thus, there are two critical conditions for PPP to be welfare-improving. First, the socially optimal level of total R&D expenditure in China exceeds the equilibrium level (Inequality (16)). Second, the dosage of PPP cannot be too high.

analysis is guided by the condition under which the net externality is positive, which we derive from the theory. We found that there is a non-trivial probability of 26% that PPP in China implemented during the sample period had been welfare-improving. Thus, PPP was protentially welfare-improving in China.

Our paper presents a perspective different from the prevailing notion in the literature that patent-promotion policies (PPP) in China caused an increase in the quantity of patents but a deterioration of the average quality of patents, and this led to a decrease in welfare. Although we obtain the same conclusion as the literature that PPP in China have been welfare-reducing, our model offers a different mechanism behind the result. The literature typically assumes that there is a limited amount of R&D resources. When more resources are allocated to develop low-quality patents, less resources are allocated to develop high-quality ones. By contrast, in our model, there is a limited amount of resources that can be used for either R&D or production, and PPP caused more resources to be allocated to R&D and less resources allocated to production. Thus, PPP would be welfare-reducing if the market-determined amount of resources allocated to R&D was already too high before the implementation of PPP. Our empircal analysis indicates that this was probably the case.

Appendix

A Proofs of comparative statics results in subsection 3.3.2

Recall Equation (7)

$$A\theta\rho + A\left(1+\theta\right)\cdot\frac{\delta\sum_{j=1}^{n} f^{j}\left[\underset{(+)}{R_{j}^{*}}\left(\underset{(+)}{A},\underset{(+)}{s_{j}},\underset{(+)}{S_{j}}\right)\right]}{\mu} = L. \tag{7}$$

From (7), we obtain 2 results:

Result	
1	$A \uparrow \longrightarrow LHS_7 \uparrow$;
2	s_i , $S_i \uparrow \longrightarrow LHS_7 \uparrow$ for all i;

Thus, we have the following comparative statics results.

- From (7), we can conclude that

Line	
1	s_i , $S_i \uparrow$
2	$\longrightarrow A \downarrow$
3	$\longrightarrow \sum_{j=1}^{n} f^{j}\left[\underset{(+)}{R_{j}^{*}}\left(\underset{(+)}{A},\underset{(+)}{s_{j}},\underset{(+)}{S_{j}}\right)\right] \uparrow$
4	and $\sum_{j\neq i} f^{j}\left[\underset{(+)}{R_{j}^{*}}\left(\underset{(+)}{A},\underset{(+)}{s_{j}},\underset{(+)}{S_{j}}\right)\right] \downarrow$
5	$\longrightarrow f^{i}\left(\underset{(+)}{R_{i}^{*}}\left(\underset{(+)}{A},\underset{(+)}{S_{i}},\underset{(+)}{s_{i}}\right)\right) \uparrow$
6	$\longrightarrow R_i^* \uparrow$
7	$\longrightarrow R_j^* \downarrow$ for $j \neq i$.

Rationale for the second line: This directly arises from results (1) and (2) of the previous table.

Rationale for the third line: Since RHS_7 is unchanged as L stays constant, LHS_7 must stay unchanged as A falls, which means that $\sum_{j=1}^{n} f^{j}\left[\underset{(+)}{R_{j}^{*}}\left(\underset{(+)}{A},\underset{(+)}{s_{j}},\underset{(+)}{S_{j}}\right)\right]$ must rise.

Rationale for the fourth line: Since $s_j = S_j = 0$ for $j \neq i$ but A falls, $\sum_{j\neq i} f^{j}\left[\underset{(+)}{R_{j}^{*}}\left(\underset{(+)}{A},\underset{(+)}{s_{j}},\underset{(+)}{S_{j}}\right)\right]$ must fall.

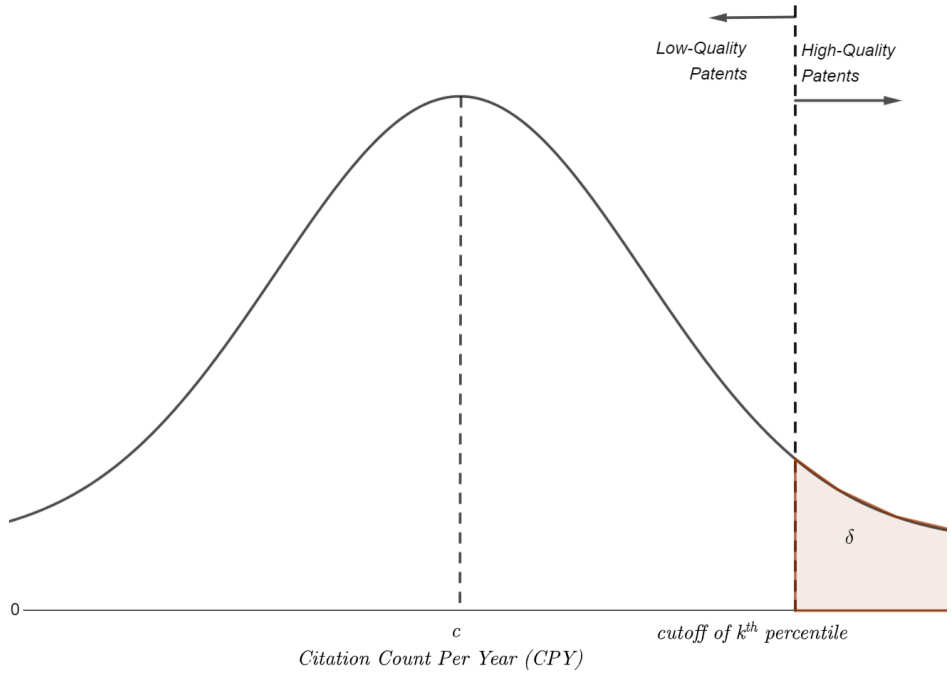


Figure 6: The probability density function of the citation count per year (a measure of patent quality) of patents generated by the research technology.

Rationale for the fifth line: Since $\sum_{j=1}^{n} f^j \left[\underset{(+)}{R_j^*} \left(\underset{(+)}{A}, \underset{(+)}{s_j}, \underset{(+)}{S_j} \right) \right]$ rises and $\sum_{j \neq i} f^j \left[\underset{(+)}{R_j^*} \left(\underset{(+)}{A}, \underset{(+)}{s_j}, \underset{(+)}{S_j} \right) \right]$ falls, $f^i \left(\underset{(+)}{R_i^*} \left(\underset{(+)}{A}, \underset{(+)}{S_i}, \underset{(+)}{s_i} \right) \right)$ must rise.

Rationale for the sixth line: R_i^* varies positively with $f^i \left(\underset{(+)}{R_i^*} \left(\underset{(+)}{A}, \underset{(+)}{S_i}, \underset{(+)}{s_i} \right) \right)$ by definition.

Rationale for the seventh line: Since $s_j = S_j = 0$ for $j \neq i$ but A falls, $f^j \left[\underset{(+)}{R_j^*} \left(\underset{(+)}{A}, \underset{(+)}{s_j}, \underset{(+)}{S_j} \right) \right]$ and R_j^* must both fall.

B The assumed probability density function of citation count per year

Refer to Figure 6, which shows the probability density function of the citation count per year (a measure of patent quality) of patents generated by the research technology.

<Figure 6 about here>

- The probability density function is assumed to follow a log-normal distribution, denoted by $\log N \sim(c,\, \sigma^2)$.
- The area under the curve is equal to one.
- The value of c, which denotes the mean of the log-normal distribution, is exogenous and may vary over time. σ is assumed to be an exogenous constant.
- High-quality (i.e. commercializable) patents are defined as those falling into the top kth percentile nationally in terms of citation count per year (CPY), while the remaining patents are classified as low-quality.
- The probability of a patent being high-quality is equal to δ.
- During the period of 1995-2009, we observed noteworthy patterns in China's patent landscape. First, the average citation per year (CPY) for all granted patents exhibited a consistent increase over time. (Figure 7) Second, the ratio of CPY between high-quality and low-quality patents (i.e. the ratio of CPY between the top 10% and the bottom 90%, as we assume $k = 10$ to be the central case) remained relatively stable. (Figure 8) Within the framework of our model, these observations are consistent with the following assumptions: First, the share of high-quality patents in China as a whole, k%, remained stable throughout the period. Second, c experienced an upward trend. In essence, the entire curve shifted rightward over time. Third, the probability δ remained stable over time. In sum, c increased over time but σ remained unchanged over time. Thus, we maintain these assumptions in our model.

<Figure 7 and Figure 8 about here>

C Solving the social planner's problem

Observe from (34) that the quantity of each variety is given by: $\ln x_\tau(j) = \ln L^p_\tau(j) + \ln Z_\tau(j) - \ln L$. Substituting this expression for $\ln x_\tau(j)$ into the utility function (19), the maximand of the social planner's problem in section 3.4.2 in the main text (also in Online Appendix section A.4.2) can be rewritten as:

$$W_0 = \int_0^\infty e^{-\rho\tau} \left[\int_0^\mu \ln L^p_\tau(j)\; dj + \int_0^\mu \ln Z_\tau(j)\; dj - \mu \ln L \right]\; d\tau. \tag{17}$$

First, note that the expected log productivity of a variety at time τ can be evaluated as: $E\left[\ln Z_\tau\right] = \sum_{k=0}^{\infty} \frac{(\tau H/\mu)^k}{k!} \left(e^{-\tau H/\mu}\right) E[\ln Z^{(k)}]$, where $Z^{(k)}$ denotes the productivity level if exactly k ideas have arrived between times 0 and τ. It can be shown that $\ln Z^{(k)}$ is the sum of k independent

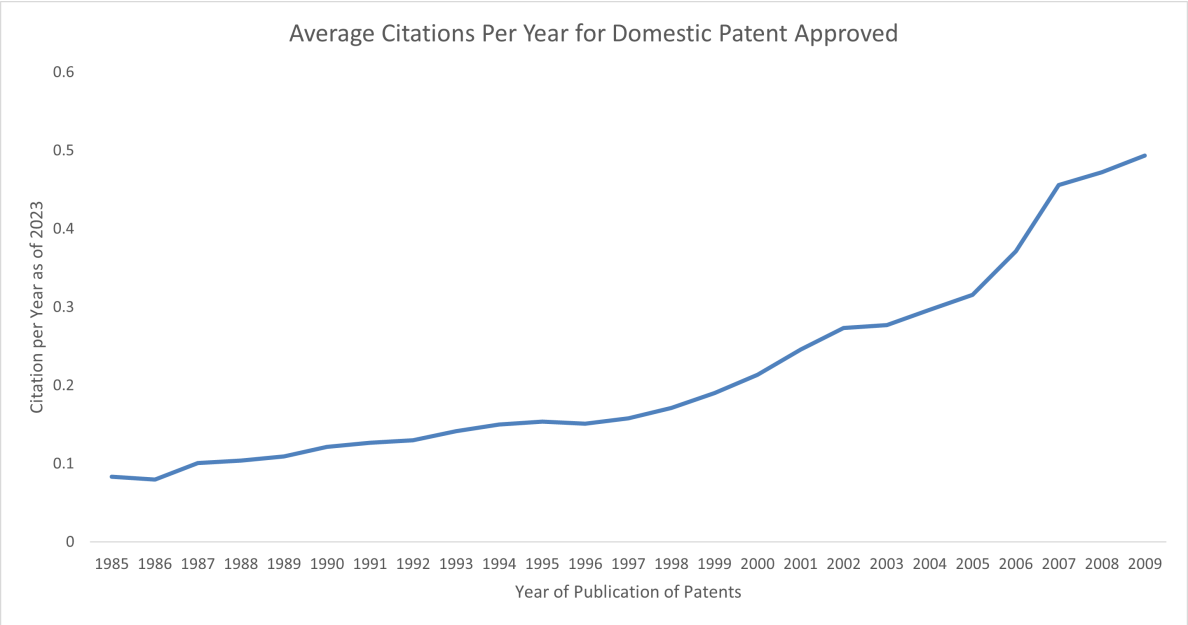


Figure 7: Average Citations per Year as of 2023 plotted against Year of Publication of the Patents

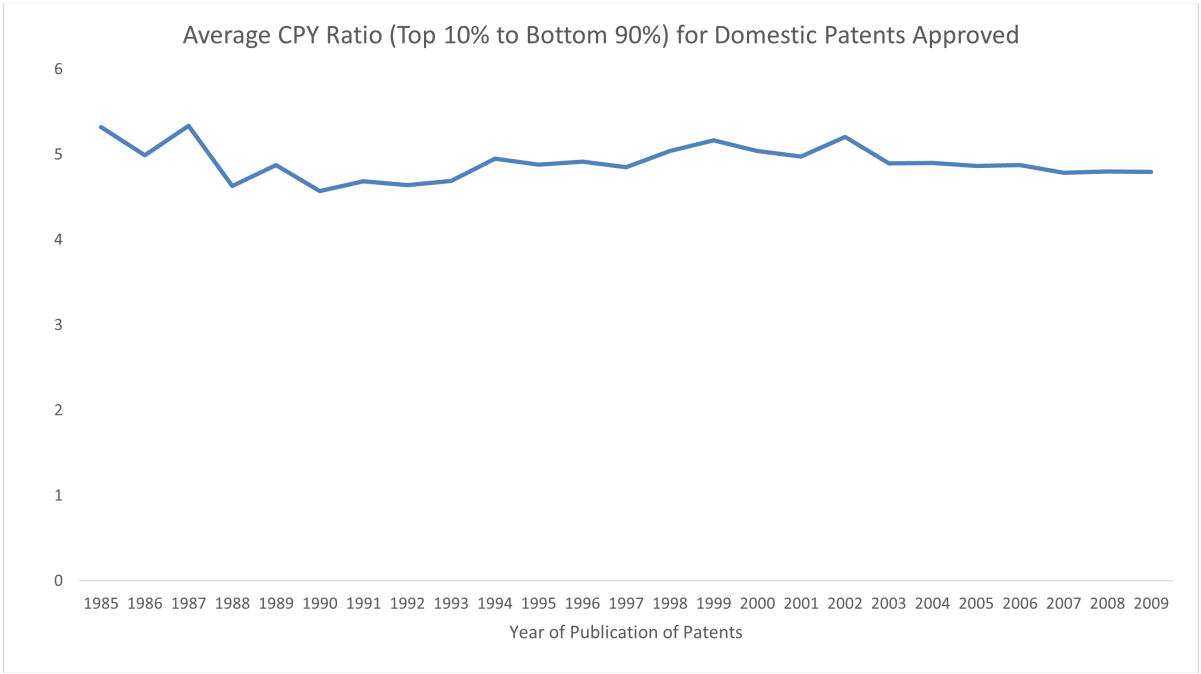


Figure 8: Ratio of the average citation per year as of 2023 of high-quality patents to the average citation per year as of 2023 of low-quality patents plotted against Year of Publication of the Patents.

realizations of the random variable $\ln\zeta$, so that: $E\left[\ln Z^{(k)}\right] = kE\left[\ln\zeta\right]$.[29] We thus have: $E\left[\ln Z_\tau\right] = E\left[\ln\zeta\right]\sum_{k=0}^{\infty}\frac{(\tau H/\mu)^k}{k!}\left(e^{-\tau H/\mu}\right)k = \frac{\tau H}{\mu}E\left[\ln\zeta\right]$ (recall that $H = \sum_{i=1}^{n} h_i = \sum_{i=1}^{n}\delta f^i\left(R_i\right)$).

Next, conditional on her choice of $\{R_i\}_{i=1}^n$, the social planner would solve the following problem for allocating production labor to each variety in order to maximize the expression in (17):

$$\begin{aligned}\max_{\{L_\tau^p(j)\}_{j=0}^{\mu}} \quad & \int_0^{\mu}\ln L_\tau^p(j)\ dj \\ \text{s.t.} \quad & \int_0^{\mu} L_\tau^p(j)\ dj = L - \sum_{i=1}^{n} R_i \quad \text{for all } \tau \geq 0.\end{aligned}$$

It is straightforward to see that the solution to this is to have equal allocation of production labor to all varieties: $\mu L_\tau^p(j) = L - \sum_{i=1}^{n} R_i$ for all varieties $j \in [0,\ \mu]$ and all $\tau \geq 0$. This arises from the fact that $\ln L_\tau^p(j)$ is concave in $L_\tau^p(j)$, so that the integral will be maximized by spreading production labor evenly across each variety.

Substituting these expressions for $\int_0^{\mu}\ln Z_\tau(j)\ dj$ and $\ln L_\tau^p(j)$ into (17), we have:

$$\begin{aligned} W_0 &= \mu\int_0^{\infty} e^{-\rho\tau}\left[\ln\left(L - \sum_{i=1}^{n} R_i\right) + \frac{\tau\cdot\sum_{i=1}^{n} h_i}{\mu}\mathbb{E}\left[\ln\zeta\right] - \ln L - \ln\mu\right]\ d\tau \\ &= \mu\left\{\frac{\ln\left[L - \sum_{i=1}^{n} R_i\right]}{\rho} + \frac{\sum_{i=1}^{n}\delta f^i\left(R_i\right)}{\rho^2\mu}\mathbb{E}\left[\ln\zeta\right] - \frac{\ln L}{\rho} - \frac{\ln\mu}{\rho}\right\}.\end{aligned}$$

If we assume Pareto distribution for ζ, we have

$$W_0 = \mu\left\{\frac{1}{\rho}\ln\left[L - \sum_{i=1}^{n} R_i\right] + \frac{\sum_{i=1}^{n}\delta f^i\left(R_i\right)}{\mu\rho^2\theta} - \frac{\ln L}{\rho} - \frac{\ln\mu}{\rho}\right\}$$

■

D Correlation between Science and Technology Expenditure and R&D Expenditure

An extrapolation approach is used to extrapolate the values of R&D Expenditures backward to earlier years in 1991-1997 based on its strong correlation with the S&T Expenditures. Details are as follows.

For each year, we calculate the correlation coefficient between provincial R&D expenditures and provincial S&T expenditures. We find that the correlation coefficient between provincial R&D

[29] The key step is to recognize that the random variable $Z^{(k-1)} = Z^{(k-1)}/Z^{(0)}$ is the product of $k-1$ independent realizations of ζ. (Recall that $\ln Z^{(0)} = 0$.) Thus, $\ln Z^{(k-1)}$ is the sum of $k-1$ independent realizations of the random variable $\ln\zeta$, so that: $\mathbb{E}[\ln Z^{(k-1)}] = (k-1)\mathbb{E}[\ln\zeta]$ for $k \geq 1$.The expected log productivity of the k-th idea to arrive thus increases linearly in k, while also increasing in the expected log size of each iid productivity improvement.

Table 8: Correlation Coefficient (Corr.) and Regression Results (i.e. Intercept (Inter.) and Slope) between S&T Expenditures and R&D Expenditures

Province	**Corr.**	**Inter.**	**Slope**	**Province**	**Corr.**	**Inter.**	**Slope**
Beijing	0.9958	47208	0.51	Hubei	0.9957	14846	0.39
Tianjin	0.9989	-3353	0.50	Hunan	0.9857	8779	0.34
Hebei	0.9970	-2314	0.54	Guangdong	0.9972	13610	0.49
Shanxi	0.9951	10665	0.25	Guangxi	0.9739	-1406	0.34
Inner Mongolia	0.9864	-243	0.34	Hainan	0.8855	1850	0.06
Liaoning	0.9946	-16412	0.61	Chongqing	0.9962	-5039	0.45
Jilin	0.9867	-3785	0.51	Sichuan	0.9852	38232	0.33
Heilongjiang	0.9872	5964	0.47	Guizhou	0.9875	-1050	0.48
Shanghai	0.9907	-69127	0.60	Yunnan	0.9875	-2548	0.48
Jiangsu	0.9979	-15829	0.48	Tibei	0.9207	792	-0.03
Zhejiang	0.9976	-12162	0.52	Shaanxi	0.9967	21355	0.48
Anhui	0.9948	9161	0.31	Gansu	0.9827	2991	0.39
Fujian	0.9985	-365	0.51	Qinghai	0.9442	-147	0.33
Jiangxi	0.9991	-5182	0.65	Ningxia	0.9919	-675	0.39
Shandong	0.9958	-36729	0.59	Xinjiang	0.9817	2815	0.16
Henan	0.9934	6540	0.38				

Expenditures and provincial S&T Expenditures is over 88%, which is very high. Thus, we go ahead and regress provincial R&D Expenditures on provincial S&T Expenditures during this overlapping period, then use the estimated coefficients of the regressions, together with the actual provincial S&T Expenditures from 1991 to 1997, to estimate the R&D Expenditures from 1991 to 1997. The correlation analysis and regression results (i.e. intercept and slope estimates) of each province are shown in Table 8. With extrapolation, we have the data of R&D Expenditures from 1991 onwards.[30]

<Table 8 about here>

[30] In this exercise, we estimate the S&T expenditure of Chongqing City and that of Sichuan Province (ex-Chongqing) from 1990 to 1996 by extrapolation. To do so, we first calculate the average ratio of S&T Expenditures between Sichuan Province and Chongqing City from 1998 to 2008. We find that Chongqing accounts for 23.6% of the S&T Expenditures in the combined entity of Sichuan and Chongqing between 1998 and 2008. We then apply this ratio to the actual Sichuan S&T Expenditures from 1990 to 1996 to estimate the Chongqing S&T Expenditures from 1990 to 1996.

References

[1] Baten, Joerg, Nicola Bianchi, and Petra Moser. 2017. "Compulsory Licensing and Innovation–Historical Evidence from German Patents after WWI." Journal of Development Economics 126: 231-242

[2] Beck, Mathias, Cindy Lopes-Bento, and Andrea Schenker-Wicki. 2016. "Radical or Incremental: Where does R&D Policy Hit?" Research Policy 45(4): 869-883.

[3] Bianchi, Mattia, Samuele Murtinu, and Vittoria Scalera. 2019. "R&D Subsidies as Dual Signals in Technological Collaborations." Research Policy 48(9): 103821.

[4] Boeing, Philipp. 2016. "The Allocation and Effectiveness of China's R&D Subsidies-Evidence from Listed Firms." Research Policy 45(9): 1774-1789.

[5] Broekel, Tom, Dirk Fornahl, and Andrea Morrison. 2015. "Another Cluster Premium: Innovation Subsidies and R&D Collaboration Networks." Research Policy 44(8): 1431-1444.

[6] Bronzini, Raffaello and Paolo Piselli. 2016. "The Impact of R&D Subsidies on Firm Innovation." Research Policy 45(2): 442-457.

[7] Callaway, Brantly and Petro Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." Journal of Econometrics 225: 200-230.

[8] Cao, Linyi, Helu Jiang, Guangwei Li, and Lijun Zhu. 2024. "Haste makes waste? Quantity-based subsidies under heterogeneous innovations." Journal of Monetary Economics 142: 103517

[9] Chapman, Gary, Abel Lucena, and Sergio Afcha. 2018. "R&D Subsidies & External Collaborative Breadth: Differential Gains and the Role of Collaboration Experience." Research Policy 47(3): 623-636

[10] Choi, Joonhwan and Jaegul Lee. 2017. "Repairing the R&D Market Failure: Public R&D Subsidy and the Composition of Private R&D." Research Policy 46(1): 25-44.

[11] Chor, Davin and Edwin L.-C. Lai. 2023. "Cumulative Innovation, Growth and Protection of Ideas." Manuscript.

[12] Dang, Jianwei and Kazuyuki Motohashi. 2015. "Patent statistics: A good indicator for innovation in China? Patent subsidy program impacts on patent quality." China Economic Review 35: 137-155.

[13] Doh, Soogwan and Byungkyu Kim. 2014. "Government support for SME Innovations in the Regional Industries: The Case of Government Financial Support Program in South Korea." Research Policy 43(9): 1557-1569.

[14] Dimos, Christos and Geoff Pugh. 2016. "The Effectiveness of R&D Subsidies: A Meta-Regression Analysis of the Evaluation Literature." Research Policy 45(4): 797-815.

[15] Fang, Jing, Hui He, and Nan Li. 2020. "China's Rising IQ (Innovation Quotient) and Growth: Firm-Level Evidence." Journal of Development Economics 147: 102561.

[16] Gaessler, Fabian, Bronwyn Hall, and Dietmar Harhoff. 2021. "Should There be Lower Taxes on Patent Income?" Research Policy 50(1): 104129.

[17] Gao, Yuchen, Yimei Hu, Xielin Liu, and Huanren Zhang. 2021. "Can Public R&D Subsidy Facilitate Firms'Exploratory Innovation? The Heterogeneous effects Between Central and Local Subsidy Programs." Research Policy 50(4): 104221.

[18] Gans, Joshua. 2012. "Innovation and Climate Change Policy." American Economic Journal: Economic Policy 4(4): 125-145.

[19] Geng, Difei and Kamal Saggi. 2022. "Tariff Barriers and the Protection of Intellectual Property in the Global Economy." European Economic Review 104074.

[20] Grossman, Gene M. and Elhanan Helpman. 1991. "Trade, Knowledge Spillovers, and Growth." European Economic Review 35(2-3): 517-526.

[21] Guo, Di, Yan Guo and Kun Jiang. 2017. "Funding Forms, Market Conditions, and Dynamic Effects of Government R&D Subsidies: Evidence from China." Economic Inquiry 55(2): 825-842.

[22] Guo, Di, Yan Guo and Kun Jiang. 2016. "Government-Subsidized R&D and Firm Innovation: Evidence from China." Research Policy 45(6): 1129-1144.

[23] Huang, Jiashun, Weiping Li, Xijie Huang, Yuandi Wang and Lijia Guo (2020), Technology and Innovation in China: A Patent Citation-based Analysis, Science, Technology & Society 26: 2 (2020): 344–365.

[24] Hud, Martin and Katrin Hussinger. 2015. "The Impact of R&D Subsidies During the Crisis." Research Policy 44(10): 1844-1855.

[25] Hünermund, Paul and Dirk Czarnitzki. 2019. "Estimating the Causal Effect of R&D Subsidies in a Pan-European Program." Research Policy 48(1): 115-124.

[26] Ivus, Olena, Manu Jose, and Ruchi Sharma. 2021. "R&D Tax Credit and Innovation: Evidence from Private Firms in India." Research Policy 50(1): 104128.

[27] Jin, Zhenji, Yue Shang and Jian Xu. 2018. "The Impact of Government Subsidies on Private R&D and Firm Performance: Does Ownership Matter in China's Manufacturing Industry?" *Sustainability*.

[28] Jones, Charles I. 2002. Introduction to Economic Growth, W. W. Norton, second edition.

[29] Kortum, Samuel S. 1997. "Research, Patenting, and Technological Change." Econometrica 65(6): 1389-1419.

[30] Lanahan, Lauren, Amol Joshi, and Evan Johnson. 2021. "Do Public R&D Subsidies Produce Jobs? Evidence from the SBIR/STTR Program." Research Policy 50(7): 104286.

[31] Lanjouw, J. O. 1998. "Patent Protection in the Shadow of Infringement: Simulation Estimations of Patent Value." The Review of Economic Studies 65 (4), 671-710.

[32] Long, Cheryl Xiaoning and Jun Wang. 2016. "Evaluating Patent Promotion Policies in China: Consequences for Patent Quantity and Quality." in Prud'homme D., Song H. (eds) Economic Impacts of Intellectual Property-Conditioned Government Incentives, Springer Science: Singapore.

[33] Li, Xibao. 2012. "Behind the Recent Surge of Chinese Patenting: An Institutional View." Research Policy 41(1): 236-249.

[34] Meuleman, Miguel and Wouter De Maeseneire. 2012. "Do R&D Subsidies Affect SMEs'Access to External Financing?" Research Policy 41(3): 580-591.

[35] Montmartin, Benjamin and Marcos Herrera. 2015. "Internal and External Effects of R&D Subsidies and Fiscal Incentives: Empirical Evidence using Spatial Dynamic Panel Models." Research Policy 44(5): 1065-1079.

[36] Mudambi, R., & Swift, T. (2014). Knowing when to leap: Transitioning between exploitative and explorative R&D. Strategic Management Journal, 35(1), 126–145.

[37] National Science Board. 2022. "Production and Trade of Knowledge- and Technology-Intensive Industries", NSB-2022-6, National Science Board - Science & Engineering Indicators, April 19, 2022.

[38] National Science Board. 2020. "Production and Trade of Knowledge- and Technology-Intensive Industries", NSB-2020-5, National Science Board - Science & Engineering Indicators, January 23, 2020.

[39] Silva, J.M.C. Santons and Silvana Tenreyro. 2006. "The Log of Gravity." The Review of Economies and Statistics 88(4): 641-658

[40] Smith, Daniel. 2020. "The Effects of Federal Research and Development Subsidies on Firm Commercialization Behavior." Research Policy 49(7): 104003

[41] Serrano, Carlos J. 2018. "Estimating the Gains from Trade in the Market for Patent Rights." International Economic Review 59(4) 1877-1904.

[42] Stojčić, Nebojša, Stjepan Srhoj, and Alex Coad. 2020. "Innovation Procurement as Capability-Building: Evaluating Innovation Policies in Eight Central and Eastern European Countries." European Economic Review 121:103330.

[43] Sun, Lixin (2024), Time Preference and Economic Growth: The Case for China and International Comparisons, Economics of Transition and Institutional Changes, 32, p.683–704.

[44] Szczygielski, Krzysztof, Wojciech Grabowski, Mehmet Teoman Pamukcu, and Vedat Sinan Tandogan. 2017. "Does Government Support for Private Innovation Matter? Firm-level Evidence from Two Catching-up Countries." Research Policy 46(1): 219-237.

[45] Szücs, Florian. 2020. "Do Research Subsidies Crowd Out Private R&D of Large Firms? Evidence from European Framework Programmes." Research Policy 49(3): 103923.

[46] Szücs, Florian. 2018. "Research Subsidies, Industry-University Cooperation and Innovation." Research Policy 47(7): 103923.

[47] Wang, Mei, Marc Oliver Rieger, Thorsten Hens (2016), How Time Preferences Differ: Evidence from 53 countries, Journal of Economic Psychology 52, p.115–135.

[48] WIPO. 1996. "WIPO-IFIA International Symposium on the Commercialization of Patented Inventions." WIPO-IFIA-MINDS. Kuala Lumpur

[49] Wei, Shang-Jin, Jianhuan Xu, Ge Yin, and Xiaobo Zhang. 2021. "Bureaucratic Bean Counting and Patent Subsidies: A Welfare Evaluation of China's Pro-innovation Program." Presentation at the 8th Annual Conference on Trade, Growth and Development.

[50] Yan, Kitming Isabel, Eline Korenromp, and Eran Bendavid. 2015. "Mortality Changes after Grants from the Global Fund to Fight AIDS, Tuberculosis and Malaria: an Econometric Analysis from 1995 to 2010." BMC Public Health 15:977.

[51] Zhang, Gupeng, & Chen, Xiangdong. 2012. "The Value of Invention Patents in China: Country Origin and Technology Field Differences". China Economic Review 23(2), 357–370

[52] Zhang, Xuan and Huihua Nie. 2021. "Public Health Insurance and Pharmaceutical Innovation: Evidence from China". Journal of Development Economics 148: 102578

Online Appendix: To be posted online only.

A Theory

A.1 Model Setup

This section gives the details of the theoretical model presented in Section 3. Consider an economy (a country with many regions) in which there is a continuum of differentiated varieties indexed by $j \in [0, \mu]$ that yield utility to consumers. Although we assume that labor is the only factor input, we would like to interpret it broadly as the totality of all resources that can be used for production or R&D. Thus, "labor" or "worker" in this model stands for a composite factor input consisting of labor, capital, land, and other resources that can be used for production or research.[31] The "wage" w is therefore the value of one unit of this composite factor input. We shall eventually set $w = 1$ for normalization. The words "labor", "worker", "labor force", "factor input", and "composite factor input" refer to the same thing and can be used interchangeably. The "wage" and "the unit value of the composite factor input" refer to the same thing and can be used interchangeably. When we carry out the empirical work, the unit value of the composite factor input will be measured in a certain quantity of real yuan. We assume that the composite factor input is mobile across regions, there is free trade of goods within the country and a national capital market; firms in any region are free to hire the composte factor input in all regions. Free mobility of the composite factor implies that the returns to the composite factor in all regions are equal. Firms in the economy are small, in the sense that each firm produces only one variety while taking the prevailing real unit cost of the composite factor as given.

Each unit of labor (or "worker") can engage in one of two activities, namely, in the production of differentiated varieties or in R&D. With regard to the former, the technologies used in production all feature constant returns-to-scale. $Z_\tau(j)$ is the labor productivity associated with the best available idea for producing variety j at time τ. Thus, $1/Z_\tau(j)$ units of labor are required to produce each unit of output, so that the unit cost faced by the firm that produces this variety is $w_\tau/Z_\tau(j)$.

In each region there is a research technology that can be used to generate commercializable patents. The objective of R&D activity is to generate commercializable patents embodying ideas that improve existing labor productivity. Each idea entails improving labor productivity for producing a specific differentiated variety.[32] The generation of these ideas follows a Poisson process,

[31] Formally, we can assume that human capital, labor, physical capital, land, and other resources are employed in fixed proportions to each other in a Leontief fashion for carrying out R&D and they are also employed in (different) fixed proportions for carrying out production.

[32] Alternatively, the innovation process can be viewed as one that brings improvements along a quality dimension.

with the aggregate arrival rate being determined by the total quantity of R&D labor input. The relationship between the aggregate arrival rate and the total quantity of R&D labor input will be explained later.

In region i, the research technology yields a Poisson arrival rate of getting a patent equal to $f^i(R_i)$ where R_i is the input into R&D, and conditional on getting a patent, there is a probability δ that the patent is commercializable (i.e. high-quality). Conditional on commmercializability, there is a probability distribution of the productivity improvement ζ, which is a random variable. We assume that there is a fixed measure of the variety of goods given by μ.

We assume that there is a fixed measure of the variety of goods given by μ. Following Kortum (1997) and Klette and Kortum (2004), when a new idea is received, the identity of the variety to which the idea applies is determined by a random draw from a uniform distribution on the interval $[0, \mu]$. As we will see in a moment, in the innovation process we specify, the expected gains are the same regardless of the identity of the variety, so that innovators would in fact be indifferent between varieties even if they could direct their efforts to a particular variety j.

We consider a situation where innovations are strictly cumulative. In particular, knowledge about production technologies spreads immediately once the good in question is commercialized, so that the underlying know-how is available to all economic agents. This could be, for example, because it is easy to reverse engineer the technology after observing a physical sample of the good. As the current best technology for producing each marketed good is widely known, each subsequent innovation strictly builds upon this knowledge to generate productivity improvements. In equilibrium, the best patented technology for each variety will indeed be used in production with the good being marketed. Therefore, any new idea will always improve the patented frontier technology for the variety to which it is applied.[33]

In the absence of intellectual property rights (IPR) protection, the dispersive nature of knowledge would provide little incentive for private agents to undertake R&D. We, therefore, require that an IPR regime is in place that allows any new idea to be patented at a cost. By getting a commercializable patent, the firm can prevent others from using the idea. Thus, the patent-owning firm is the only one who can produce and market the variety (say, variety j) with this new technology and will indeed have the entire market for j to itself as it is now the most productive manufacturer of j. This monopoly power only expires when the next idea that improves upon j arrives.[34]. Ideas that

Each new idea brings greater benefits to consumers without increasing the production cost of the good (and thus the market price).

[33] Alternatively, the innovation process can be viewed as one that brings improvements along a quality dimension. Each new idea brings greater benefits to consumers without increasing the production cost of the good (and thus the market price).

[34] Given the continuous measure of varieties, the probability of the same actor successively obtaining two ideas for producing the same variety is zero

are not patented but are marketed can be imitated immediately by other firms that would dispute the inventor's profits. It follows that firms will immediately patent any new idea they receive. In equilibrium, then, no goods are marketed without first being patented. (We first assume that patents do not expire, i.e., that patent duration is infinite, so that incumbent patent holders lose their monopoly power only if they are displaced by a new idea. It can be shown that the main results remain the same when there is a finite patent length.)

To encourage innovation, the government of region i offers two incentives to firms. The first type of policy rewards firms with a certain amount of money once a patent is acquired. We call this a reward policy, and the financial reward per patent is equal to S_i. The second type of policy directly lowers the cost of factor input (i.e., labor) used in R&D that pursue patents. This type of incentive policy will be called subsidy policy, and the monetary subsidy per R&D worker is equal to a fraction s_i of the wage.

We now describe the productivity level associated with each arriving high-quality (i.e. commercializable) idea. To initialize the innovation process, we assume that at the start of time ($\tau = 0$), there is a baseline technology for each variety that is freely available to all firms. We normalize the productivity of this baseline technology to be 1, so that $Z^{(0)}(j) = 1$ for all varieties $j \in [0, \mu]$. Next, we define $Z^{(k)}(j)$ to be the productivity associated with the k-th idea to arrive for variety j, where k is a non-negative integer. Thus, $\{Z^{(0)}(j), Z^{(1)}(j), Z^{(2)}(j), \ldots\}$ form a sequence of the successive best technologies for producing this variety. To describe how this frontier technology evolves, define $\zeta^{(k+1)}(j) \equiv Z^{(k+1)}(j)/Z^{(k)}(j)$ to be the productivity improvement associated with the next idea to arrive. We adopt a specification where the size of each productivity improvement is stochastic so that $\zeta^{(k+1)}(j)$ is the realization of an independent draw from an underlying stationary random variable, denoted by ζ. This means in particular, that the distribution of the productivity improvements does not depend on the identity of the variety (j), how many ideas have already arrived (k), or the productivity level of the last drawn idea ($Z^{(k)}(j)$), so that the innovation process is symmetric across varieties.

Later in the paper, we will sometimes assume that the random variable ζ is drawn from the following standardized Pareto distribution with shape parameter $\theta > 1$:

$$\Pr(\zeta < z) = 1 - z^{-\theta}, \text{ where } z \in [1, \infty). \tag{18}$$

Recall that a lower θ implies a more fat-tailed distribution which places greater weight on drawing relatively large productivity improvements. This Pareto distribution will prove to make the model more tractable.

A.2 General Equilibrium

We now embed the above cumulative innovation process in a general equilibrium setting.

A.2.1 Utility

There are n regions in this country. The utility function of the representative consumer in each region as of time 0 is given by:

$$W_0 = \int_0^\infty e^{-\rho\tau} \ln u_\tau \, d\tau. \tag{19}$$

Here, ρ is the rate of time preference, while u_τ aggregates the instantaneous utility from the consumption of differentiated varieties at time τ. Denoting the quantity of variety j consumed at time τ by $x_\tau(j)$, we assume that u_τ is given by:

$$u_\tau = \exp\left\{\int_0^\mu \ln x_\tau(j) \; dj\right\}. \tag{20}$$

Since there is free trade of goods in the country and a national capital market and the wages of all regions are equal due to free mobility of labor, we can assume that there is one single representative consumer for all regions of the country. The representative consumer chooses $\{x_\tau(j)\}_{\tau=0}^\infty$ to maximize (19), subject to the intertemporal budget constraint:

$$\int_0^\infty e^{-r\tau} X_\tau \, d\tau \leq b(0),$$

where $X_\tau = \int_0^\mu p_\tau(j)\, x_\tau(j)\, dj$ denotes the flow of consumption spending at time τ, with $p_\tau(j)$ being the corresponding price of variety j at that time (which the consumer takes as given). The prevailing interest rate is r, and this is determined by the rate of return for owning a patent, which is the only asset in our economy. Finally, $b(0)$ denotes the present value of the future stream of wage income that will be earned by each consumer plus the value of their initial asset holdings at date 0.

It is well known (see, e.g., Grossman and Helpman, 1991) that solving this dynamic optimization problem yields the following:

$$r = \rho + \frac{\dot{X}}{X},$$

where $\dot{X}$ is the time derivative of X. (We will omit the time subscript for equations that hold for all $\tau \geq 0$.) Thus, the rate of growth of consumption spending should equal the difference between the market interest rate and one's private rate of time preference. The variable l_i denotes the total amount of labor employed in region i. $L = \sum_i l_i$ is total supply of labor in the country, which is assumed to be constant. Let $E_\tau \equiv LX_\tau$ denote aggregate expenditure in this country at time τ. Thus, the above equation becomes

$$r = \rho + \frac{\dot{E}}{E}. \tag{21}$$

We will show later in this section that E_τ is, in fact, constant in a steady state so that we can write $E_\tau = E$. With L also constant over time, it follows from (21) that $r = \rho$. Note that E/μ is the

expenditure per variety at each date since we have a fixed measure μ of symmetric varieties. As the expenditure on each j is invariant to its price, the price elasticity of demand for each variety is -1.

A.2.2 Market structure and profits generated by high-quality patents

Firms owning commercializable patents compete by setting prices. If no commercializable ideas have yet arrived for variety j by time τ, then that variety is priced at marginal cost (i.e., w_τ) since the baseline technology is freely accessible to all potential producers. On the other hand, if at least one idea has arrived for a variety j, firms compete under Bertrand competition. The firm with the most productive idea will set a marginal price just high enough to keep the second most productive firm (and thus to keep all other firms) out of the market. Therefore, the equilibrium price for variety j at time τ when precisely k ideas have arrived is equal to: $p_\tau(j) = w_\tau / Z^{(k-1)}(j)$. The price markup that the most productive firm sets are simply: $Z^{(k)}(j)/Z^{(k-1)}(j)$, for $k \geq 1$. If we denote by $\eta(m)$ the cumulative density function of the price markup m (defined as price/unit cost), it follows that $\eta(m)$ inherits the distribution of ζ.[35]

Consider now a variety for which at least one idea has arrived. We compute the expected flow profits that accrue to the most productive firm for this variety as follows. Observe that the share of expenditures E_τ/μ that accrue to this firm as profits is $(m-1)/m$. Let Π_τ be the total expected profits of all firms at time τ and $(\Pi_s)_\tau$ be the expected profits per variety at time τ. Thus, $\Pi_\tau = \mu \cdot (\Pi_s)_\tau$. Therefore,

$$\begin{aligned} \Pi_\tau &= \mu \cdot (\Pi_s)_\tau = \mu \left\{ \int_{m=1}^{\infty} \left(\frac{m-1}{m} \right) \frac{E_\tau}{\mu} \, d\eta(m) \right\} = \left(1 - \mathbb{E}[\zeta^{-1}]\right) E \\ &= \frac{E}{(1+\theta)} \text{ if we assume the Pareto distribution as stated in equation (18)} \\ &\quad \text{as } \mathbb{E}[\zeta^{-1}] = \frac{\theta}{(1+\theta)} \end{aligned} \tag{22}$$

Since E_τ is constant over time and ζ is drawn from a stationary distribution, the above profits are also constant over time.

A.2.3 Research Productivity of the research technology

Suppose in each region there is a research technology that uses R&D to generate commercializable patents.

[35] As is common in this literature (e.g., O'Donoghue 1998), we assume that there is regulatory authority prohibiting collusion or anticompetitive behavior in which the owner of the best patented idea would buy out the firm with the second best patented idea to allow for a higher price premium.

Let the contribution to the incremental Poisson arrival rate of new patents in region i by the marginal R&D worker in region i be equal to $\lambda^i(R_i)$, where R_i is the allocation of labor to R&D in region i. We assume that $\lambda^i(R_i)$ is a convex and decreasing function that converges to zero: $\left(\lambda^i\right)'(R_i) < 0$, $\left(\lambda^i\right)''(R_i) > 0$, $\lambda^i(0) > 0$, $\lambda^i(\infty) = 0$. This is shown in Figure 1. These assumptions imply that the success rate of obtaining a patent by a marginal worker is relatively high when fewer labor resources are devoted to R&D. As more and more resources are devoted to R&D, the additional success rate contributed by the marginal worker decreases. The intuition behind this assumption is the diminishing returns to R&D effort. The earlier efforts yield a higher marginal Poisson arrival rate (because they reach for the lower-hanging fruits). The Poisson arrival rate as a function of R_i is defined as $f^i(R_i) \equiv \int_0^{R_i} \lambda^i(x)\,dx$. Note that we assume that all R&D workers are equally productive. Thus, the Poisson arrival rate of any R&D worker is a random variable with an expected value equal to the mean $\phi^i(R_i) \equiv \left(\int_0^{R_i} \lambda^i(x)\,dx\right)/R_i = f^i(R_i)/R_i$. It can be easily shown from Figure 1 that $\left(\phi^i\right)'(R_i) < 0$ as $\left(\lambda^i\right)'(R_i) < 0$. Intuitively, since the incremental Poisson arrival rate of the marginal R&D worker is lower than the average Poisson arrival rate of the existing R&D workers, the allocation of an additional unit of labor to R&D lowers the average Poisson arrival rate.

In any region, conditional on the research yielding a patent, there is a probability δ that the patent is commercializable. Conditional on commmercializability, there is a probability distribution of the inventive step ζ as has been explained in equation (1). Refer to Appendix B for the assumed probability density function of the quality of patents generated by the research technology. The empirical facts align with the assumption of constant δ, and a constant share ($k\%$) of commercializable patents in the context of our model. We define the expected number of high-quality patents awarded in region i as $h_i(R_i) \equiv \delta f^i(R_i)$.

A.2.4 Savings, Investments, and the Structure of the Economy

Since R_i is the quantity of labor in region i that are hired by firms to engage in R&D activity, the remaining quantity of labor in region i, $l_i - R_i$, engages in the production of goods, where l_i is the total amount of labor employed in region i. Note that l_i is endogenous as each region hires an amount of labor from the national labor market to satisfy the need for R&D and production of goods. In equilibrium, the demand for labor is equal to the supply of labor in the national labor market: $\sum_i l_i = L$. Firms need to obtain financing in order to hire R&D workers as their profits accrue in the future. This spending on R&D is the only form of investment in our model as the innovation is undertaken to generate ideas that yield an expected future stream of profits. There is a national capital market (financial intermediary). To obtain financing, firms issue claims (in the national capital market) on the flow of profits from their patents. These claims are the only assets in this closed economy (the country), and the total value of these assets at time τ is denoted by A_τ.

Note that A_τ is equal to the expected value of each patent (call it $(A_s)_\tau$) multiplied by the number of varieties μ, given that we have only one active patent for each variety.[36] The above investment in R&D activity is financed through the savings of workers. This could occur with savers directly owning the equity of firms or with the financing channeled through an intermediary such as a bank.

Since only high-quality patents have commercial value, high-quality patents have positive asset value, $A > 0$, and low-quality patents have no asset value. Ex-post, all asset returns come from firms that hold high-quality patents. Hereinafter, we shall use the terms "commercializable patents" and "high-quality patents" interchangeably, and "non-commercializable patents" and "low-quality patents" interchangeably.

Recall that workers have two sources of income, their labor income and asset returns. Thus, total saving in the economy is equal to total income minus spending, $L + rA - E$. Equating this with aggregate net investment $(\dot{A})$ at each point in time, we have:

$$L + rA - E = \dot{A}. \tag{23}$$

Households (consisting of workers) own businesses by holding the debt issued by businesses in the national capital market and paying taxes to the national government.

The prevailing interest rate r must be equal to the rate of return for owning an asset (namely, a patent), which in turn must equal the flow profit rate plus the rate of capital gains, less the probability (called "hazard rate") of suffering a complete capital loss due to the arrival of a new idea that supersedes the existing patent. This yields:

$$\begin{aligned} & r = \frac{\Pi_s}{A_s} + \frac{\dot{A}_s}{A_s} - \frac{\sum_{i=1}^{n} \delta f^i(R_i)}{\mu} \\ \Longrightarrow \qquad & r = \frac{\Pi}{A} + \frac{\dot{A}}{A} - \frac{\sum_{i=1}^{n} h_i(R_i)}{\mu} \quad \text{since } A = \mu A_s \end{aligned} \tag{24}$$

A.2.5 Factor Market Equilibrium in Each Region

The composite factor input will be allocated, through the market, to work in either (i) R&D or (ii) production. In equilibrium, the returns to the composite factor from engaging in production and R&D are equal due to free mobility of the factor and free entry into the industries.

As discussed earlier, we assume that the composite factor input is mobile across regions, there is free trade of goods within the country and a national capital market; firms in any region are free to hire the composite factor input in all regions. Free mobility of the composite factor implies that the returns to the composite factor, called w, in all regions are equal.

[36] $A_\tau = \int_0^\mu (A_s)_\tau (j) dj = \mu \cdot (A_s)_\tau$ since $(A_s)_\tau (j) = (A_s)_\tau$ for all j.

Assume that the job of an R&D worker is not only to create a new patentable idea but also prepare a patent application for the idea. The Poisson arrival rate of a patentable idea covers the successful completion of both tasks. Different provinces offer different levels of patenting rewards and subsidies. A monetary reward S_i is given to any firm once a patent is granted, regardless of its commercializability. For the subsidy policy, the government of region i provides any qualified firm with a subsidy of a fraction s_i of the wage of an R&D worker during the process of applying for a patent (regardless of whether the application is successful). We can think of ws_i as a monetary subsidy on the patent application fee. There is an important distinction between these two channels of pro-innovation policies (i.e. reward policy and subsidy policy). The patenting reward is granted only after a patent is issued, while the patenting subsidy is granted to subsidize patent application.

In a competitive labor market, to obtain an interior solution of labor allocation (i.e. positive amount of labor allocated to R&D and production), the wage of an R&D worker needs to be equal to their expected "value of the marginal product" (VMP), which is equal to the expected value of a patent whose commercializability is not yet known (i.e., $\delta A_s = \delta A/\mu$ in the absence of the patent reward policy) multiplied by the average Poisson arrival rate faced by the R&D worker. Moreover, the wage of a production worker w must be equal to that of an R&D worker so that workers are indifferent between being employed in R&D or production. Recall that $\phi^i(R_i) \equiv \frac{f^i(R_i)}{R_i}$ is the expected Poisson arrival rate of an R&D worker. Therefore, in the absence of pro-innovation policies, the equilibrium labor allocation condition is given by

$$\frac{A}{\mu}\delta \cdot \frac{f^i(R_i)}{R_i} = w \quad \text{for } i = 1, 2, ..., n \tag{25}$$

where A, R_i (assuming that there is an interior solution), and w are endogenous. (We shall set $w = 1$ later as normalization.)

In the presence of pro-innovation policies s_i and S_i, the expected value of a patent whose commercializability is not yet known is $\delta A_s + S_i = \delta A/\mu + S_i$, while the cost of one unit of the composite factor is $w(1 - s_i)$. Thus, the equilibrium condition for firms carrying out R&D is given by

$$\left(\frac{A}{\mu}\delta + b_i S_i\right) \cdot \frac{f^i(R_i)}{R_i} = w(1 - a_i s_i) \quad \text{for } i = 1, 2, ..., n \tag{26}$$

where a_i is a dummy variable equal to one if there is subsidy policy in province i and zero otherwise; b_i is a dummy variable equal to one if there is reward policy in province i and zero otherwise.

A.3 Results and Implications

We are now ready to solve the model and discuss its results and implications.

A.3.1 Steady State

First, we set $w = 1$ as normalization. We shall demonstrate below that this economy will jump immediately to a steady state with $\dot{A} = \dot{X} = \dot{E} = 0$. (21), (22), (23), (24), and (26), and the normalization $w = 1$ together yield $4+n$ equations in $4+n$ unknowns r, Π, E, A, $\{R_i^*\}_{i=1}^n$, where a variable with an asterisk denotes the equilibrium value. We define the equilibrium Poisson arrival rate of high-quality patents in region i as $\delta f^i\left(R_i^*\right) \equiv h_i^*$. Note that $\frac{d}{dA}\left[h_i^*\right] > 0$ for all i, as $\frac{d\left(R_i^*\right)}{dA} > 0$ according to (26). Thus, we can write h_i^* as $h_i^*\left[\underset{(+)}{R_i^*}\left(\underset{(+)}{A}, \underset{(+)}{s_i}, \underset{(+)}{S_i}\right)\right] = h_i^*\left(\underset{(+)}{A}, \underset{(+)}{s_i}, \underset{(+)}{S_i}\right)$ for short.

Eliminating E, r, and Π from (21), (22), (23), and (24), we obtain

$$\left(\rho + \frac{\dot{E}}{E}\right) + \left(\frac{1+\theta}{\theta}\right) \cdot \frac{\sum_{i=1}^n h_i^*\left(\underset{(+)}{A}, \underset{(+)}{s_i}, \underset{(+)}{S_i}\right)}{\mu} - \frac{L}{\theta A} = \frac{\dot{A}}{A} \tag{27}$$

On the other hand, (21), (22), and (24) together imply

$$\frac{\dot{E}}{E} - \frac{\dot{A}}{A} = \left(\frac{1}{1+\theta}\right)\frac{E}{A} - \frac{\sum_{i=1}^n h_i^*\left(\underset{(+)}{A}, \underset{(+)}{s_i}, \underset{(+)}{S_i}\right)}{\mu} - \rho \tag{28}$$

Below, we shall show that the economy must jump immediately to $\dot{A} = 0$ starting from any initial condition.

First, suppose $LHS_{27} < 0$ so that $\dot{A}/A < 0$. Suppose also that $\frac{\dot{E}}{E} > \frac{\dot{A}}{A}$. Then $LHS_{28} > 0$. Moreover, it implies that $\frac{E}{A}$ increases and $-\sum_{i=1}^n h_i^*\left(\underset{(+)}{A}, \underset{(+)}{s_i}, \underset{(+)}{S_i}\right)$ increases (as A falls), which implies that RHS_{28} increases, which implies that LHS_{28} is even more positive. Repeated application of this logic implies that $\frac{E}{A}$ will explode to infinity. This will not be consistent with a sustainable steady state. In other words, $\dot{A}/A < 0$ must be associated with $\frac{\dot{E}}{E} \leq \frac{\dot{A}}{A}$, which implies that $\dot{A}/A < 0$ must be associated with $\dot{E}/E < 0$.

Second, suppose $LHS_{27} > 0$ so that $\dot{A}/A > 0$. Suppose also that $\frac{\dot{E}}{E} < \frac{\dot{A}}{A}$. Then $LHS_{28} < 0$. Moreover, it implies that $\frac{E}{A}$ decreases and $-\sum_{i=1}^n h_i^*\left(\underset{(+)}{A}, \underset{(+)}{s_i}, \underset{(+)}{S_i}\right)$ decreases (as A rises), which implies that RHS_{28} decreases, which implies that LHS_{28} is even more negative. Repeated application of this logic implies that $\frac{E}{A}$ will degenerate to zero. This will not be consistent with a sustainable steady state. In other words, $\dot{A}/A > 0$ must be associated with $\frac{\dot{E}}{E} \geq \frac{\dot{A}}{A}$, which implies that $\dot{A}/A > 0$ must be associated with $\dot{E}/E > 0$.

(27) and (28) give rise to

$$\left(\frac{1}{1+\theta}\right)\frac{E}{A}+\frac{1}{\theta}\cdot\frac{\sum_{i=1}^{n}h_i^*\left(\underset{(+)}{A},\underset{(+)}{s_i},\underset{(+)}{S_i}\right)}{\mu}=\frac{L}{\theta A} \tag{29}$$

which must hold all the time.

If $\dot{A}<0$, then $\frac{\dot{E}}{E}-\frac{\dot{A}}{A}\leq 0$, which implies that E/A falls or is unchanged over time, but this cannot be consistent with (29), as LHS_{29} falls and RHS_{29} rises over time and so the equation cannot hold. Conversely, if $\dot{A}>0$, then $\frac{\dot{E}}{E}-\frac{\dot{A}}{A}\geq 0$, which implies that E/A rises or is unchanged over time, but this cannot be consistent with (29), as LHS_{29} rises and RHS_{29} falls over time and so the equation cannot hold. Thus, it must be the case that $\dot{A}/A=0$ all the time, which means that E/A is constant over time and that $\dot{E}/E=0$ all the time.

Thus, the economy must jump immediately to $\dot{A}=\dot{X}=\dot{E}=0$ starting from any initial condition. Thus, (21) implies that $r=\rho$ all the time.

As $\dot{A}=\dot{E}=0$, (28) becomes

$$0=\left(\frac{1}{1+\theta}\right)\frac{E}{A}-\frac{\sum_{i=1}^{n}h_i^*\left(\underset{(+)}{A},\underset{(+)}{s_i},\underset{(+)}{S_i}\right)}{\mu}-\rho. \tag{30}$$

Eliminating $\left(\frac{1}{1+\theta}\right)\frac{E}{A}$ between (29) and (30) yields

$$A\theta\rho+A\left(1+\theta\right)\cdot\frac{\sum_{i=1}^{n}h_i^*\left(\underset{(+)}{A},\underset{(+)}{s_i},\underset{(+)}{S_i}\right)}{\mu}=L. \tag{31}$$

Equation (31) together with (26) can be used to solve for A and R_i^*. The equations will be used to carry out our comparative statics exercise in subsection A.3.2.

A.3.2 Comparative Statics and Propositions

We carry out a number comparative statics exercises. From (7), we obtain

$$\frac{dA}{ds_i}<0; \qquad \frac{dA}{dS_i}<0;$$

and

$$\frac{dR_i^*}{ds_i}>0; \qquad \frac{dR_i^*}{dS_i}>0;$$

and

$$\frac{dR_j^*}{ds_i}<0; \qquad \frac{dR_j^*}{dS_i}<0 \qquad \text{for } j\neq i;$$

and

$$\frac{d\left[\sum_{k=1}^{n} f^i\left(R_k^*\right)\right]}{ds_i} > 0; \qquad \frac{d\left[\sum_{k=1}^{n} f^i\left(R_k^*\right)\right]}{dS_i} > 0.$$

Refer to Appendix A for the proof of the above sets of inequalities. The inequalities above indicate that home province patent subsidy policy, or home province patent reward policy will increase the allocation of resources to R&D in the home province. They have effects on A through (7), which in turn impacts R_i^* through (6). Although home subsidy and reward policies reduce the value of a high-quality patent (A/μ), which tends to lower the allocation of resources to R&D (R_i^*), such effect is more than compensated for by increased incentives to carry out R&D resulting from the home patent subsidy or patent reward policies. Thus, the effects through A are second-order. On the other hand, home province patent subsidy policy, or home province patent reward policy will reduce the allocation of resources to R&D in other provinces where there is no PPP, as the value of each commercializable patent falls while there is no government subsidies to offset this negative impact on the reward to R&D.

Proposition 1. *PPP in a province increases the R&D expenditure and the numbers of high- and low-quality patents awarded in the home province but reduces the R&D expenditure and the numbers of high- and low-quality patents awarded in other provinces where there is no PPP.*

Moreover, the last set of inequalities above indicates that PPP in any province will increase the total number of patents approved at the national level, despite the fact that it reduces the numbers of patents approved in provinces with no PPP. In other words, the positive effect in the province with PPP outweighs the negative effect in the provinces with no PPP. This is a subtle point. The intuition is that, as A falls, the profit rate per period, Π/A, increases, and this induces more entry into R&D nationally. Thus, $\sum_{k=1}^{n} f^i\left(R_k^*\right)$ increases. This makes sense as (24) indicates that when the profit rate Π/A increases the hazard rate at which an active patent will be superseded, $\sum_{k=1}^{n} f^i\left(R_k^*\right)/\mu$, must increase so as to keep the expected rate of return of each asset equal to the rate of time preference ρ.

Proposition 2. *PPP in a province increases the R&D expenditure and the numbers of high- and low-quality patents awarded at the national level.*

Note that an increase in R_i^* imply three things. First, it means that the amount of R&D expenditure increases. Second, it implies an increase in the Poisson arrival rate of patents, denoted by $f^i\left(R_i^*\right)$. Third, this also means increases in the numbers of high- and low-quality patents awarded per period, given by $\delta f^i\left(R_i^*\right)$ and $f^i\left(R_i^*\right)(1-\delta)$, respectively.[37] The first is about the input measure of R&D, while the third is about the output measure of R&D.

For each individual province i and at the national level, the expected share of high-quality

[37] As footnote 7 indicates, "worker" in this model stands for a collection of human capital, unskilled labor, physical capital, and so on, employed in fixed proportions to each other in a Leontief manner. Thus, the R&D expenditure is the sum of the costs of all these factor inputs used in R&D.

Table 9: Comparative Statics Results

	A	R_i^*	$f^i(R_i^*)$	$R_j^*\,(j \neq i)$	$f^i(R_j^*)\,(j \neq i)$	$\sum_{k=1}^n f^i(R_k^*)$	δ
$S_i \uparrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$	$\uparrow$	No effect
$s_i \uparrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$	$\uparrow$	No effect

$S_i \equiv$ Monetary reward per patent offered in region i

$s_i \equiv$ Monetary subsidy per R&D worker as a fraction of the wage offered in region i

$A \equiv$ Economy-wide total value of high-quality (i.e. commercializable) patents

$R_i^* \equiv$ R&D expenditure allocated to R&D in region i

$f^i(R_i^*) \equiv$ No. of patents approved in region i

$\sum_{k=1}^n f^i(R_k^*) \equiv$ No. of patents approved at the national level

$\delta \equiv$ "Expected no. of high-quality patents" /"Total no. of patents" approved in region i and nationally

patents is equal to δ, which is independent of PPP, as the probability of each patent being high-quality (commercializable) is assumed to be a constant δ.

We summarize the comparative statics results by Table 9.

<Table 9 about here>

A.4 Welfare Analysis: Market Equilibrium and Social Optimum

In this section, we derive the condition under which the market without government intervention delivers less than the optimal amount of R&D. If that condition is satisfied, government pro-innovation policies that increases R_i will be welfare-improving if the increases are not excessive, and if the policies are financed by lump sum taxes. The result is summarized by Proposition 3.

A.4.1 Equilibrium condition

In this subsection, we focus on the case without pro-innovation policies. Equations (21), (22), (23), (24), (25), and the normalization $w = 1$ together yield $4 + n$ equations in $4 + n$ unknowns r, Π, E, A, $\{R_i^*\}_{i=1}^n$.

Plugging Π from (22), E from (23), and $\delta f^i(R_i)$ from (25), into equation (24), we obtain

$$
\begin{aligned}
\rho &= (L+\rho A)\left(1-\mathbb{E}[\zeta^{-1}]\right)\frac{1}{A}-\frac{\sum_i (R_i\mu)/A}{\mu} \\
\implies \quad \rho A &= (L+\rho A)\left(1-\mathbb{E}[\zeta^{-1}]\right)-\sum_{i=1}^{n} R_i \\
\implies \quad A &= \frac{L\left(1-\mathbb{E}[\zeta^{-1}]\right)-\sum_i R_i}{\rho\cdot\mathbb{E}[\zeta^{-1}]} \\
\implies \quad A &= \frac{L-(1+\theta)\sum_i R_i}{\rho\theta} \quad \text{assuming Pareto distribution for } \zeta
\end{aligned}
$$

Equation (25) implies that the market equilibrium level of R_i obeys

$$
\begin{aligned}
\frac{A}{\mu}h_i &= R_i \quad \text{where} \quad h_i \equiv \delta f^i(R_i) \text{ for } i=1,2,...,n \\
\implies \quad \frac{R_i}{h_i} &= \frac{L-(1+\theta)\sum_i R_i}{\rho\theta\mu} \quad \text{when substituting for } A \text{ from the above expression} \\
\implies \quad \frac{\rho\theta\mu R_i}{Lh_i} &= 1-(1+\theta)\frac{\sum_i R_i}{L}
\end{aligned}
$$

$$
\begin{aligned}
\implies \quad \frac{\sum_i R_i}{L} &= \left(\frac{1}{1+\theta}\right)-\left(\frac{\theta}{1+\theta}\right)\frac{\rho\mu R_i}{Lh_i} \quad \text{under Pareto distribution of } \zeta \qquad (32) \\
&= \left(1-\mathbb{E}[\zeta^{-1}]\right)-\mathbb{E}[\zeta^{-1}]\frac{\rho\mu R_i}{Lh_i} \quad \text{under general distribution}
\end{aligned}
$$

A.4.2 Welfare Calculation and Contrast with the social optimum

To understand the efficiency properties of the steady state, it is instructive to compare the above market equilibrium with the outcomes under a benign social planner for the country as a whole. Conceptually, the social planner's problem can be formulated as a labor allocation decision over the share of labor to employ in R&D, as well as the amount of labor to assign to the production of each variety $j \in [0,\mu]$ at each time point τ (which we denote by $L^p_\tau(j)$). Formally, the planner sets out to maximize the welfare of the representative consumer:

$$
\begin{aligned}
\max_{\{R_{i\tau}\}_{i=1}^n, \{L^p_\tau(j)\}_{j\in[0,\ \mu]}} \quad & W_0 \\
\text{s.t.} \quad & \int_0^\mu L^p_\tau(j)\, dj = L_\tau - \sum_{i=1}^{n} R_{i\tau} \quad \text{for all } \tau \geq 0, \qquad (33) \\
\text{and} \quad & Lx_\tau(j) = L^p_\tau(j) Z_\tau(j) \quad \text{for all } j\in[0,\mu] \text{ and } \tau \geq 0. \qquad (34)
\end{aligned}
$$

The first constraint (33) is a labor market-clearing condition that states that all labor not engaged in R&D must be employed in production. In addition, the second constraint (34) clears the market for each variety j, setting the quantity demanded equal to the quantity produced at each point in time. We drop the subscript τ in this subsection from now on for simplicity.

As we show in Appendix section C, the solution to this social planner's problem features an equal allocation of labor to the production of each variety. In other words, given the choice of $\{R_{i\tau}\}_{i=1}^{n}$, we have equal allocation of production labor to all varieties: $\mu L_{\tau}^{p}(j) = L_{\tau} - \sum_{i=1}^{n} R_{i\tau}$ for all $j \in [0, \mu]$, bearing in mind that we have a measure of varieties equal to μ.

From Appendix section C, assuming Pareto distribution of ζ, so that $\mathbb{E}[\ln \zeta] = \frac{1}{\theta}$, we have

$$\begin{aligned} W_0 &= \mu \int_0^{\infty} e^{-\rho\tau} \left\{ \ln \left[L - \sum_{i=1}^{n} R_i \right] + \left[\frac{\sum_{i=1}^{n} \delta f^i(R_i)}{\mu} \mathbb{E}[\ln \zeta] \right] \tau - \ln L - \ln \mu \right\} d\tau \\ &= \mu \left\{ \frac{\ln [L - \sum_{i=1}^{n} R_i]}{\rho} + \frac{\sum_{i=1}^{n} \delta f^i(R_i)}{\mu \rho^2 \theta} - \frac{\ln L}{\rho} - \frac{\ln \mu}{\rho} \right\} \quad \text{under Pareto dist'n of } \zeta \end{aligned} \tag{35}$$

Thus, the first order condition that determines the socially optimal level of R_i in each region i is given by

$$\begin{aligned} \frac{dW_0}{dR_i} &= \frac{1}{\rho} \cdot \frac{-1}{L - \sum_{i=1}^{n} R_i} + \mathbb{E}[\ln \zeta] \frac{1}{\rho^2} \cdot \frac{\lambda^i(R_i)\,\delta}{\mu} = 0 \qquad \text{for } i = 1, 2, ..., n \\ \Longrightarrow \quad \frac{\sum_{i=1}^{n} R_i}{L} &= 1 - \frac{\rho\mu}{L \lambda^i(R_i)\,\delta \mathbb{E}[\ln \zeta]} \end{aligned} \tag{36}$$

Under market equilibrium, since all varieties are ex ante symmetric, the law of large numbers implies that the expected price of the state-of-the-art product of all varieties are equal. Therefore, the expected output of all varieties are equal as well. This means that there is equal allocation of labor to the production of each variety. In other words, out of the total supply of factor input L_t, the market allocates $\{R_i\}_{i=1}^{n}$ to R&D in each province i, and allocates the rest of the resources equally to all varieties $j \in [0, \mu]$ for production. This procedure is just like that followed by the social planner. Thus, we can use (35) to calculate the welfare under market equilibrium as well.

A.4.3 Four sources of externalities

Under equilibrium, equation (32) implies that

$$\frac{\sum_{i=1}^{n} R_i}{(\sum_{i=1}^{n} h_i)/\mu} = \frac{\left(1 - \mathbb{E}[\zeta^{-1}]\right) \cdot (L - \sum_{i=1}^{n} R_i)}{\mathbb{E}[\zeta^{-1}] \cdot [\rho + (\sum_{i=1}^{n} h_i)/\mu]} \quad \text{in the general case} \tag{37}$$

$$= \frac{\left(\frac{1}{1+\theta}\right) \cdot (L - \sum_{i=1}^{n} R_i)}{\left(\frac{\theta}{1+\theta}\right) \cdot [\rho + (\sum_{i=1}^{n} h_i)/\mu]} \quad \text{under Pareto distribution of } \zeta \tag{38}$$

Under social optimum at the national level, (36) implies that

$$\begin{aligned}\frac{\sum_{i=1}^{n} R_i}{L} &= 1 - \frac{\rho\mu R_i}{LR_i\lambda^i(R_i)\,\delta} \cdot \frac{1}{\mathbb{E}(\ln\zeta)} \\ &= 1 - \frac{\rho\mu R_i}{Lg_i} \cdot \frac{1}{\mathbb{E}(\ln\zeta)}\end{aligned}$$

where $g_i \equiv R_i\lambda^i(R_i)\,\delta$. This further implies that under social optimum,

$$\frac{\sum_{i=1}^{n} R_i}{\left(\sum_{i=1}^{n} g_i\right)/\mu} = [\mathbb{E}(\ln\zeta)]\,\frac{\left(L - \sum_{i=1}^{n} R_i\right)}{\rho} \quad \text{in the general case} \tag{39}$$

$$= \frac{1}{\theta} \cdot \frac{\left(L - \sum_{i=1}^{n} R_i\right)}{\rho} \quad \text{under Pareto distribution} \tag{40}$$

Comparing (37) and (39), we can identify four sources of externalities:

1. intertemporal spillover (positive externality) — The term under social optimum: ρ vs. the term under equilibrium: $\rho + \left(\sum_{i=1}^{n} h_i\right)/\mu$ in the denominator on the RHS.

The effective discount rate that matters for the market equilibrium is $\rho + \left(\sum_{i=1}^{n} h_i\right)/\mu$, whereas that which drives the planner's problem is ρ. This difference is due to the "intertemporal spillover" effect that arises from the cumulative nature of innovation: In the market equilibrium, firms apply a higher effective discount rate when evaluating the value of a successful innovation because they do not internalize the positive knowledge spillovers that they generate for future productivity improvements. Instead, they discount the future at a higher rate that incorporates the hazard rate of being leapfrogged, $\left(\sum_{i=1}^{n} h_i\right)/\mu$. *Ceteris paribus*, this intertemporal spillover effect would push the market equilibrium towards featuring under-investment in R&D relative to the planner's outcome.

2. appropriability effect (positive externality) — The term under social optimum: $[\mathbb{E}(\ln\zeta)]$ vs. the term under equilibrium: $\left(1 - \mathbb{E}[\zeta^{-1}]\right)$ in the numerator on the RHS.

The term $1 - \mathbb{E}[\zeta^{-1}]$ appears in the numerator on the right-hand side of (37), whereas we instead have a $\mathbb{E}[\ln\zeta]$ term in the corresponding equation (39) for the social planner. This contrast captures the "appropriability" effect. With the log utility specification in (19), a successful innovation drawn from ζ yields a (flow) welfare gain per unit of the good consumed equal to $\mathbb{E}[\ln\zeta]$. However, the profit share that accrues to the innovating firm is only $1 - \mathbb{E}[\zeta^{-1}]$ (c.f., equation (22)). As $\mathbb{E}[\ln\zeta] > 1 - \mathbb{E}[\zeta^{-1}]$ except in the special case where ζ is a point mass at 1, firms are indeed unable to appropriate the full surplus enjoyed by society from their innovation.[38] Thus, this effect also

[38] To see this, straightforward calculus will show that the inequality $\ln x \geq (x-1)/x$ holds for all x, with equality if and only if $x = 1$. Taking expectations with respect to the probability density function of ζ yields the desired inequality. This inequality will moreover be strict except in the special case where ζ is a point mass at 1.

pushes the market equilibrium towards a lower allocation of labor to research than the planner would choose.

3. business-stealing effect (negative externality) — The term under social optimum: 1 vs. the term under equilibrium: $\frac{1}{\mathbb{E}[\zeta^{-1}]}$ on the RHS.

Innovation effort erodes the profits of preceding innovators in a way that the social planner would want to fully internalize. We associate this effect with the $\frac{1}{\mathbb{E}[\zeta^{-1}]}$ term that appears in (37), but which is absent in (39). The latter reflects the fact that aggregate profits in the economy are still constant, despite the new innovation occurring, whereas the firm that owns this innovation would value the private profit flows that it generates. As the support of ζ is $(1, \infty)$, we clearly have $\frac{1}{\mathbb{E}[\zeta^{-1}]} > 1$, so that this business-stealing effect pushes the private actors in the direction of over-investing in research relative to the planner's problem.[39]

4. congestion effect (negative externality) — The term under social optimum: $\left(\sum_{i=1}^{n} g_i\right)/\mu$ in the denominator of LHS of (37) vs. the term under equilibrium: $\left(\sum_{i=1}^{n} h_i\right)/\mu$ in the denominator on the LHS of (37).

The addition of one more researcher reduces the average arrival rate of ideas over all researchers (See, for example, p.100 of Jones 2002). This is analogous to the effect that the addition of one more car in a highway slows down the average speed over all cars in the highway. The negative externality is reflected in the fact that $g_i(R_i) < h_i(R_i)$ for all i if we plug the equilibrium value of R_i into $g_i(R_i)$ and $h_i(R_i)$.[40]

Now, let us focus on the special case of Pareto distribution for the inventive step ζ. As is shown in the comparison between (38) and (40), under Pareto distribution, effects 2 and 3 cancel each other, leaving only effects 1 and 4. Under what condition will the net externality be positive so that the social optimum features a higher value of $\sum_{i=1}^{n} R_i$ than under equilibrium? It is straightforward to show that (38) implies that, under equilibrium,

$$\left\{\sum_{i=1}^{n} R_i\right\}^{eqm} = \frac{L\left(\sum_{i=1}^{n} h_i\right)}{\left(\rho\mu + \sum_{i=1}^{n} h_i\right)\theta + \sum_{i=1}^{n} h_i},$$

whereas (40) implies that, under social optimum,

$$\left\{\sum_{i=1}^{n} R_i\right\}^{SO} = \frac{L\left(\sum_{i=1}^{n} g_i\right)}{\rho\theta\mu + \sum_{i=1}^{n} g_i}.$$

[39] The presence of monopoly-pricing power *per se* does not distort labor allocations in our model. The reason is that all firms charge the same markup in expectation (drawn from the standardized Pareto distribution, $\eta(m)$), so that the allocation of production labor across varieties cannot be improved upon *ex ante*. See the related discussion in Grossman and Helpman (1991), p.70.

[40] $\frac{f^i(R_i)}{R_i} > \lambda^i(R_i)$ for all $R_i > 0$ as long as $\lambda^i(R_i)$ downward sloping. Therefore, $g_i(R_i) < h_i(R_i)$ for all $R_i > 0$.

Comparing the above two equations, it can be easily shown that

$$\begin{aligned}
&\left\{\sum_{i=1}^{n} R_i\right\}^{SO} > \left\{\sum_{i=1}^{n} R_i\right\}^{eqm} \\
\iff \quad &\frac{L\left(\sum_{i=1}^{n} g_i\right)}{\rho\theta\mu + \sum_{i=1}^{n} g_i} > \frac{L\left(\sum_{i=1}^{n} h_i\right)}{\left(\rho\mu + \sum_{i=1}^{n} h_i\right)\theta + \sum_{i=1}^{n} h_i} \text{ when } R_i \text{ is the equilibrium value} \\
\iff \quad &\frac{1}{\sum_{i=1}^{n} g_i} - \frac{1}{\sum_{i=1}^{n} h_i} < \frac{1}{\rho\mu} \text{ when } R_i \text{ is the equilibrium value} \qquad (41)
\end{aligned}$$

In order to further analyze under what condition (41) [same as (16)] is more likely to hold, we introduce more structure to the function $\lambda^i(R_i)$ by assuming that $\lambda^i(R_i) = \beta^i(R_i+1)^{-\gamma}$ (this ensures that $\lambda^i(0) \neq \infty$), where β^i, $\gamma > 0$. This implies that

$$f^i(R_i) = \frac{\beta^i}{1-\gamma}\left[(R_i+1)^{1-\gamma} - 1\right].$$

So,

$$\sum_{i=1}^{n} g_{it} = \delta \sum_{i=1}^{n} R_{it}\beta_{it}(R_{it}+1)^{-\gamma}$$

and

$$\sum_{i=1}^{n} h_{it} = \frac{\delta \sum_{i=1}^{n} \beta_{it}\left[(R_{it}+1)^{1-\gamma} - 1\right]}{1-\gamma}.$$

Therefore, inequality (41) [same as (16)] becomes

$$\frac{\mu_t}{\delta \sum_{i=1}^{n} R_{it}\beta_{it}(R_{it}+1)^{-\gamma}} - \frac{(1-\gamma)\mu_t}{\delta \sum_{i=1}^{n} \beta_{it}\left[(R_{it}+1)^{1-\gamma} - 1\right]} < \frac{1}{\rho}$$

Note that the second term on the left hand side,

$$\frac{(1-\gamma)\mu_t}{\delta \sum_{i=1}^{n} \beta_{it}\left[(R_{it}+1)^{1-\gamma} - 1\right]} = \frac{1}{\text{hazard rate}}$$

where the hazard rate is the rate of complete capital loss of an asset (which is a patent) due to its being replaced by a more productive new patented technology.

Furthermore, it can be shown that $R_{it}\beta_{it}(R_{it}+1)^{-\gamma}$ is greater than $\beta_{it}\left[(R_{it}+1)^{1-\gamma} - 1\right]$ for $R_{it} \geq 1$ and $\gamma > 0$. Therefore, the first term in the above inequality is smaller than $1/(1-\gamma)$ multiplied by the second term. It follows that a sufficient condition for the inequality (41) [same as (16)] to hold is

$$\left(\frac{\gamma}{1-\gamma}\right)\frac{1}{\text{hazard rate}} < \frac{1}{\rho}. \qquad (42)$$

This is therefore a sufficient condition for the socially optimal value of $\sum_{i=1}^{n} R_i$ to be greater than the equilibrium value. Below, we will argue that fast-growing low- and middle-income countries

such as China tend to have smaller γ, higher hazard rate, and lower ρ, so that condition (42) is more likely to be satisfied.

There are three observations:

1. As $\sum_{i=1}^{n} g_i < \sum_{i=1}^{n} h_i$, the $LHS_{41} > 0$ as long as $\lambda^i(R_i)$ is downward sloping. In the limiting case when the slope of $\lambda^i(R_i)$ approach zero, $\sum_{i=1}^{n} g_i \longrightarrow \sum_{i=1}^{n} h_i$, and $LHS_{41} \longrightarrow 0$, and (41) [same as (16)] is automatically satisfied. Therefore, the flatter is $\lambda^i(R_i)$, the more likely (41) [same as (16)] is satisfied. Let us define the the elasticity of the curve $\lambda^i(R_i)$ as "congestion elasticity". Then, the flatter is the curve $\lambda^i(R_i)$, the smaller is the congestion elasticity, and the smaller is the congestion externality, and the more likely the intertemporal spillovers externality will dominate the congestion externality. This means that the socially optimal level is greater than the equilibrium level of national expenditure in R&D. When $\lambda^i(R_i) = \beta^i(R_i + 1)^{-\gamma}$, the elasticity of congestion is $-R_i \cdot \gamma / (R_i + 1) \approx -\gamma$ in each province. The smaller is γ the more likely that (42) is satisfied.

2. Inequality (42) is more likely to hold if ρ is smaller, i.e. consumers are more patient. The economic intuition is that, for a given hazard rate, a smaller ρ implies that the positive intertemporal spillovers externality is stronger (i.e. the fraction (ρ + hazard rate)/ρ is larger), making it more likely that the net externality is positive.

3. For countries undergoing faster technological change, the hazard rate faced by each patented technology is higher. This reduces the left hand side of (42), making the inequality more likely to hold. The economic intuition is similar to point (2) above: For a given ρ, a higher hazard rate implies that the intertemporal spillovers externality is stronger (i.e. the fraction (ρ + hazard rate)/ρ is larger), making it more likely that the net externality is positive.

When (41) [same as (16)] holds, sufficiently small increases in provincial R&D expenditures induced by patent promotion policies financed by lump sum taxes will be welfare improving. Thus, we have Proposition 3 below.

Proposition 3. *Inequality (41) [same as (16)] gives a sufficient condition for the market-determined level of R&D expenditure in a country to be less than the socially optimal level. When that condition is satisfied, sufficiently small increases in R&D expenditures by firms induced by pro-innovation policies financed by lump sum taxes will be welfare improving.*

Inequality (42) implies that it is more likely to hold if the elasticity of incremental Poisson arrival rate with respect to R&D input, γ, is smaller, the country is more patient (has lower discount for the future, ρ), and has higher rate of replacement of old technology by new technology (higher

hazard rate). In that case, beyond the establishment of an effective patent system, additional subsidy and reward policies to promote patenting will be welfare-improving.

To summarize, there are four sources of externalities, two positive and two negative. Under the assumption of Pareto distribution of the inventive step of commercializable patents, one of the positive externalities (approriability effect) exactly offset one of the negative externalities (business-stealing effect), leaving behind just one positive (intertemporal spillovers effect) and one negative externalities (congestion effect). When the congestion externality is sufficiently weak, as indicated in Inequality (41) [same as (16)], the socially optimal level of national R&D is higher than the non-intervention market-determined level, making patent subsidies and rewards potentially welfare-improving if the magnitude of subsidies are not too high.

B Counterfactual exercise of removing PPP from provinces where PPP has been implemented

Since the reward policy has the same effect as a subsidy policy (both stimulate R&D in the province where PPP is implemented), we simplify our analysis by assuming PPP to be just a subsidy to the cost of R&D. We then assume that the effect of PPP is to reduce the R&D cost of firms by a fraction s.

We assume different combinations of the values of s, γ, δ, θ and ρ, and compute how PPP affect the welfare of China in different years. We try $s = 0.1$ and 0.2, $\rho = 0.03$, 0.05 and 0.07; $\delta = 0.03$, 0.05 and 0.1; $\gamma = 0.6$, 0.7 and 0.8; $\theta = 1.05$ to 20.0 (the set of plausible values varies with t). We justify why we consider these values to be the only plausible ones in Online Appendix C. We focus on all 31 provinces over the years 1995-2009. We calibrate the model year by year and carry out counterfactual exercises year by year. The values of s, δ, ρ, γ and θ are assumed to be constant over time. In addition, the set of plausible values of s, δ, ρ, and γ are invariant over time. In order to fit the model with the data, we allow the set of plausible values of θ as well as the values of β_i and μ to vary from year to year. So, β_i should be denoted by β_{it} and μ denoted by μ_t.

Assume that the incremental Poisson arrival rate when the amount of R&D (measured in real yuan) is R_{it} is given by

$$\lambda_{it}(R_{it}) = \beta_{it}(R_{it} + 1)^{-\gamma}.$$

where β_{it} captures the R&D technology specific to province i at t as well as the policies other than PPP that promotes the generation of patents in i at t.

- From the equation for the incremental Poisson arrival rate, the number of patents generated in i at t is given by

$$M_t^i = \frac{\beta_{it}}{1-\gamma}\left[(R_{it}+1)^{1-\gamma}-1\right]$$

Let R_{it}^{act} be defined as the actual value of R_{it}, and $\left(M_t^i\right)^{act}$ the actual value of M_t^i. Then, we have

$$\left(M_t^i\right)^{act} = \frac{\beta_{it}}{1-\gamma}\left[\left(R_{it}^{act}+1\right)^{1-\gamma}-1\right] \tag{43}$$

For a given value of γ, we can calculate β_{it}.

- **Estimation of $\frac{A_t}{\mu_t}$.**

Let a_{it} be a dummy variable which is equal to one if PPP has been implemented in i at t, and zero otherwise.

Recall that the equilibrium condition (6) as given in the main text is

$$\left(\frac{A_t}{\mu_t}\delta + bS_{it}\right)\cdot\frac{f^i\left(R_{it}\right)}{R_{it}} = w_t\left(1-a_{it}s\right) \quad \text{for } i=1,2,...,n$$

where $f^i\left(R_i\right) = M_t^i$. Since the reward policy has the same effect as a subsidy policy (both stimulate R&D in the province where PPP is implemented), we proxy PPP by a subsidy to the cost of R&D. Setting $w_t = 1$ as normalization, (6) is simplified as

$$\frac{A_t}{\mu_t}\delta\cdot M_t^i = R_{it}\left(1-a_{it}s\right) \tag{44}$$

$$\Longrightarrow \qquad R_{it}^{act} = \frac{\left(A_t\right)^{act}}{\mu_t}\delta\cdot\frac{\left(M_t^i\right)^{act}}{1-a_{it}s} + \varepsilon_{it} \tag{45}$$

where ε_{it} is an iid error term that has zero mean across all i for any given t, so that $\sum_i \varepsilon_{it} = 0$ for any given t. (45) implies that

$$\frac{\left(A_t\right)^{act}}{\mu_t} = \frac{\sum_i R_{it}^{act}}{\delta\sum_i\left(\frac{\left(M_t^i\right)^{act}}{1-a_{it}s}\right)} \tag{46}$$

This is how $\frac{A_t}{\mu_t}$ is calculated, conditional on δ and s.

- Calculation of μ_t.

According to the theory (see equation (7)),

$$A_t = \frac{L_t}{\rho\theta + (1+\theta)\cdot\frac{\delta\sum_i M_t^i}{\mu_t}} \tag{47}$$

where L_t is the value-added (VA) of knowledge- and technology-intensive (KTI) sectors in China at t in real yuan. The data for L_t can be found from the National Science Board report of the US. (47) implies that

$$(A_t)^{act} = \frac{L_t}{\rho\theta + (1+\theta)\cdot\frac{\delta\sum_i \left(M_t^i\right)^{act}}{\mu_t}} \tag{48}$$

$$\Longrightarrow \qquad \frac{(A_t)^{act}}{\mu_t} = \frac{L_t}{\mu_t\rho\theta + (1+\theta)\cdot\delta\sum_i \left(M_t^i\right)^{act}}. \tag{49}$$

(46) and (49) together imply that

$$\begin{aligned} \frac{\sum_i R_{it}^{act}}{\delta\sum_i\left[\frac{\left(M_t^i\right)^{act}}{1-a_{it}s}\right]} &= \frac{L_t}{\mu_t\rho\theta + (1+\theta)\cdot\delta\sum_i\left(M_t^i\right)^{act}} \\ \Longrightarrow \qquad \frac{\mu_t\rho\theta}{\delta} &= \frac{L_t}{\left(\sum_i R_{it}^{act}\right)/\sum_i\left[\frac{\left(M_t^i\right)^{act}}{1-a_{it}s}\right]} - (1+\theta)\sum_i\left(M_t^i\right)^{act} \end{aligned} \tag{50}$$

from which we can calculate μ_t, conditional on s, δ, θ and ρ.

- How does the assumed values of θ, ρ and δ affect the likelihood that PPP is welfare-improving in year t?

  Assume that θ is year-specific. Note from (50) that

  $$\theta \uparrow \quad \Longrightarrow \frac{\mu_t\rho\theta}{\delta}\downarrow \Longrightarrow \frac{\mu_t\rho}{\delta}\downarrow \Longrightarrow \frac{\mu_t\rho}{\delta\sum_i\left(M_t^i\right)^{act}}\downarrow \text{ (as } \left(M_t^i\right)^{act} \text{ is data)}$$

  $$\Longrightarrow \frac{\rho}{\text{hazard rate}}\downarrow \text{ (as hazard rate} = \frac{\delta\sum_i\left(M_t^i\right)^{act}}{\mu_t}) \Longrightarrow \frac{\text{hazard rate}}{\rho}\uparrow$$

  This, according to (42), will make it more likely that PPP is welfare-improving. Since only θ, but not ρ or δ, can affect the RHS of (50), only θ, but not ρ or δ, can affect whether PPP is more likely to be welfare improving.

- (43) and (45) together imply that

  $$R_{it}^{act} = \frac{A_t}{\mu_t}\delta\cdot\left(\frac{1}{1-a_{it}s}\right)\frac{\beta_{it}}{1-\gamma}\left[\left(R_{it}^{act}+1\right)^{1-\gamma}-1\right]+\varepsilon_{it}$$

Let R_{it}^{pa} be defined by

$$R_{it}^{pa} = \frac{A_t}{\mu_t}\delta\cdot\left(\frac{1}{1-a_{it}s}\right)\frac{\beta_{it}}{1-\gamma}\left[\left(R_{it}^{pa}+1\right)^{1-\gamma}-1\right]$$

It turns out that, on average, R_{it}^{pa} will be sufficiently close to R_{it}^{act} when γ is sufficiently large. We will choose values of γ that are sufficiently large so that R_{it}^{pa} will be sufficiently close to R_{it}^{act} on average. Once we have done that, we can use R_{it}^{pa} to proxy for the model-predicted value of R_{it}^{act} for calculating welfare changes. We also define

$$\left(M_t^i\right)^{pa} = \frac{\beta_{it}}{1-\gamma}\left[\left(R_{it}^{pa}+1\right)^{1-\gamma}-1\right].$$

- Calculate the counterfactual values of R_{it} when PPP is removed

To evaluate the impact of PPP on R_{it}, we carry out a counterfactual exercise of removing PPP from province i where PPP has been implemented and see how it impacts R_{it}, treating β_{it}, μ_t, s, γ, δ, θ and ρ as given. But A_t will have to be re-calculated and is denoted by $(A_t)^{CF}$.

Let R_{it}^{CF} be the counterfactual value of R_{it} when PPP is removed in provinces where PPP has been implemented. Therefore, at a given t,

$$R_{it}^{CF} = \frac{(A_t)^{CF}}{\mu_t}\delta\cdot\frac{\beta_{it}}{1-\gamma}\left[\left(R_{it}^{CF}+1\right)^{1-\gamma}-1\right] \quad \text{for all i}$$

where

$$\frac{(A_t)^{CF}}{\mu_t} = \frac{L_t}{\mu_t\rho\theta+(1+\theta)\cdot\delta\sum_i\left(M_t^i\right)^{CF}}$$

and

$$\left(M_t^i\right)^{CF} = \frac{\beta_{it}}{1-\gamma}\left[\left(R_{it}^{CF}+1\right)^{1-\gamma}-1\right] \quad \text{for all i}$$

The above equations can be condensed into

$$R_{it}^{CF} = \frac{L_t\cdot\delta\cdot\frac{\beta_{it}}{1-\gamma}\left[\left(R_{it}^{CF}+1\right)^{1-\gamma}-1\right]}{\mu_t\rho\theta+(1+\theta)\cdot\delta\sum_i\frac{\beta_{it}}{1-\gamma}\left[\left(R_{it}^{CF}+1\right)^{1-\gamma}-1\right]} \quad \text{for all } i \tag{51}$$

Thus, (51) defines a system of n equations and n unknowns $\left\{R_{it}^{CF} \text{ for all i}\right\}$, where n is the number of provinces used for our analysis.

From the proof in Appendix A, PPP in province i reduces A_t, increases R_{it}, reduces $\{R_{jt}\}_{j\neq i}$ and increases $\sum_i M_t^i$. Thus, we conclude that when we counterfactually remove PPP from the provinces where PPP were implemented, the counterfactual values are such that (i) $R_{it}^{CF} > R_{it}^{act}$ when $a_{it}=0$; (ii) $R_{it}^{CF} < R_{it}^{act}$ when $a_{it}=1$; (iii) $\sum_i\left(M_t^i\right)^{CF} < \sum_i\left(M_t^i\right)^{pa}$; (iv) $(A_t)^{CF} > (A_t)^{act}$.

- To evaluate the impact of PPP on welfare, we calculate the difference between the following two expressions:

$$(W_0)^{pa} = \frac{\mu_t}{\rho}\left\{\ln\left[L_t - \sum_{i=1}^{n} R_{it}^{pa}\right] + \frac{\delta\sum_{i=1}^{n} f_{it}\left(R_{it}^{pa}\right)}{\mu_t\rho\theta} - \ln L_t - \ln\mu_t\right\}$$

and

$$(W_0)^{CF} = \frac{\mu_t}{\rho}\left\{\ln\left[L_t - \sum_{i=1}^{n} R_{it}^{CF}\right] + \frac{\delta\sum_{i=1}^{n} f_{it}\left(R_{it}^{CF}\right)}{\mu_t\rho\theta} - \ln L_t - \ln\mu_t\right\}$$

where

$$f_{it}\left(R_{it}\right) = \frac{\beta_{it}}{1-\gamma}\left[\left(R_{it}+1\right)^{1-\gamma} - 1\right] \text{ for } R_{it} = R_{it}^{pa} \text{ or } R_{it}^{CF}.$$

C More about the Data and Parameters

In this section, we describe how L_t, R_{it}, as well as other parameters are determined. We do not want to commit to a single value for each parameter, as this will reduce the credibility of our result. Instead, for each parameter, we identify a set of plausible (discrete) values. Since we do not have reason to believe that some values within the set of plausible ones are more likely to be true than others, we assume that each value in the set of plausible values is equally likely. During 1995-2009, there were 253 year-province in which PPP was implemented. For different years, the data L_t, $\{M_{it}\}$, $\{R_{it}\}$ and the set of provinces that adopted PPP are different. For any given year t, for each combination of γ, s, θ, δ and ρ, we can determine whether PPP is welfare-improving for that year based on our counterfactual exercises described in Online appendix section B. (For any given year, for some combinations of γ, s, θ, δ and ρ, PPP is welfare-reducing, while for other combinations of γ, s, θ, δ and ρ, PPP is welfare-improving.) Based on the assumption that each value in the set of plausible values of each of the parameters is equally likely, we can calculate the probability that PPP is welfare-improving for China in year t.

For example, for the year 2008, 29 provinces implemented PPP. As shown in the rest of this section, we determine that the plausible values of the parameters are: $\rho \in \{0.03, 0.05, 0.07\}$; $\gamma \in \{0.6, 0.7, 0.8\}$; $\delta = \{0.03, 0.05, 0.1\}$; s = $\{0.1, 0.2\}$. It turns out that the set of plausible values of θ is a function of t, s and γ, but independent of ρ and δ. For example, when $t = 2008$, $s = 0.1$, $\gamma = 0.6$, the set of plausible values of θ is $\{1, 2, 3, ..., 14\}$, and out of these 14 values, PPP is welfare-improving only when $\theta \in \{10, 11, 12, 13, 14\}$. So, in 2008, for s=0.1 and $\gamma = 0.6$, the probability that the PPP implemented by the 29 provinces is welfare-improving for China is $5/14 = 35.7\%$.

- The variable L_t is taken as the real value-added (in million real yuan) in the knowledge- and technology-intensive sectors (KTI sectors) in China in year t. It is calculated as follows.

The value-added in the KTI sector is the amount contributed by a country to the value of a good or service and excludes purchases of domestic and imported materials and inputs. KTI industries

include high R&D intensive and medium-high R&D intensive industries based on a classification by the Organisation for Economic Co-operation and Development (OECD).[41] To obtain this value for each year, we used the KTI value added as a share of real domestic GDP for China for the years 2002, 2007, 2011, 2015, and 2019 from the National Science Board (2022), as shown in the following table.[42] We denote this percentage by k_t.

Year	k_t
2002	14.0%
2007	14.5%
2011	14.9%
2015	15.0%
2019	15.5%

Since k_t as stated in National Science Board (2022) increased with time, we calculated the fitted line using the values in the above table and used that fitted line to interpolate and extrapolate the value of k_t for the other years in our sample. Once we calculated k_t for each year, we multipled it by the real GDP of China to obtain L_t. The following table shows the values of k_t and L_t (in million real yuan) for each year t.

[41] According to National Science Board (2020, p.11), KTI industries includes five high R&D intensive industries and eight medium-high R&D intensive industries (Table 6-1). The five high R&D intensive industries are aircraft; computer, electronic, and optical products; pharmaceuticals; scientific R&D services; and software publishing. The eight medium-high R&D intensive industries are chemicals excluding pharmaceuticals; electrical equipment; information technology (IT) services; machinery and equipment; medical and dental instruments; motor vehicles; railroad and other transportation; and weapons.

[42] Refer to Figure KTI-7 on p.30 of "Productive and Trade of Knowledge- and Technology-Intensive Industries", NSB-2022-6, National Science Board - Science & Engineering Indicators, April 19, 2022.

Year	k_t	L_t (in million real yuan)
1995	12.2%	371,530
1996	12.2%	393,908
1997	12.8%	419,445
1998	12.8%	410,787
1999	13.4%	419,970
2000	14.0%	444,697
2001	14.0%	448,172
2002	14.0%	451,329
2003	14.1%	467,056
2004	14.2%	498,142
2005	14.3%	526,585
2006	14.4%	544,217
2007	14.5%	574,988
2008	14.6%	617,137
2009	14.7%	609,136

- The variable R_{it} is taken as the R&D expenditure in the KTI sectors in province i at year t. We calculate R_{it} as a fraction r_t of the total R&D expenditure in province i at year t. The fraction is assumed to be the same for all provinces. There are no data for r_t in China. We can only speculate about it. We know that in 2017, the KTI sectors accounted for about 77% of all business funded R&D in the US.[43] We use this as a benchmark. We speculate that the value of r_t in China was increasing over time during our sample period as China gradually put more recources into the KTI sectors. Thus, we assume r_t in China to be increasing over time, starting at 0.5 in 1990 and gradually increasing to 0.7 in 2009. Extrapolation along this trend will give $r_t = 0.77$ at 2017, the same value as that of the US in 2017. Even if we change the starting value slightly (say to $r_t = 0.6$ in 1990), the main results of the paper will be about the same.

- The plausible range of the rate of time preference, ρ, is determined as follows. Kortum (1997) used $\rho = 0.07$, based on matching ρ to the real return observed on stock markets in the US. We believe that it is reasonable to assume that the value of ρ for China should not be higher than that for the US. So, we try $\rho = 0.03$, 0.05 and 0.07. We assume that each of the plausible values is equally probable.

- Assuming that $\lambda^i(R_i) = \beta^i (R_i + 1)^{-\gamma}$, the plausible range of congestion elasticity γ is determined as follows.

[43] See Figure 6-5 in National Science Board (2020).

First, we find that as γ decreases, the model's fit in predicting R_{it} falls. When $\gamma = 0.5$ or less, the average deviation of the predicted value R_{it}^{pa} from the actual value R_{it} is greater than 15%. We consider this to be unacceptable. So, we rule out $\gamma \leq 0.5$.

Second, we calculate how γ affects the model's fit in predicting the elasticity of M_{it} with respect to R_{it}, as explained below. It turns out that the fit decreases with γ, as shown below.

The model says that

$$M_t^i = \frac{\beta_{it}}{1-\gamma}\left[(R_{it}+1)^{1-\gamma} - 1\right]$$

which implies that $(1-\gamma)\frac{R_{it}(R_{it}+1)^{-\gamma}}{(R_{it}+1)^{1-\gamma}-1}$ is the elasticity of M_{it}/β_{it} with respect to R_{it}. The model-predicted elasticity at $R = \frac{R_{it}+R_{i,t+1}}{2}$ is therefore

$$\begin{aligned}
\varepsilon_{\text{model}}^{t,i}(\gamma) &= (1-\gamma)\frac{\left(\frac{R_{it}+R_{i,t+1}}{2}\right)\left(\frac{R_{it}+R_{i,t+1}}{2}+1\right)^{-\gamma}}{\left(\frac{R_{it}+R_{i,t+1}}{2}+1\right)^{1-\gamma}-1} \\
&= \frac{1-\gamma}{1-\left[\left(\frac{R_{it}+R_{i,t+1}}{2}\right)^{1-\gamma}\right]^{-1}} \text{ as } R_{it}+1 \approx R_{it}
\end{aligned}$$

The estimated elasticity at $R = \frac{R_{it}+R_{i,t+1}}{2}$ from the data is

$$\begin{aligned}
\varepsilon_{\text{actual}}^{t,i} &= \frac{M_{i,t+1}/\beta_{i,t+1} - M_{it}/\beta_{it}}{R_{i,t+1}-R_{it}} \cdot \frac{R_{i,t+1}+R_{it}}{M_{i,t+1}/\beta_{i,t+1} + M_{it}/\beta_{it}} \\
&\approx \frac{M_{i,t+1}/1.2 - M_{it}}{R_{i,t+1}-R_{it}} \cdot \frac{R_{i,t+1}+R_{it}}{M_{i,t+1}/1.2 + M_{it}}
\end{aligned}$$

since 1.2 is the average value of $\beta_{i,t+1}/\beta_{it}$ over t = 2005, 2006, 2007, 2008 and

i = {16 provinces with the highest number of patents and adopted PPP in 2005-2009}. That is why we use 1.2 to normalize $M_{i,t+1}$ to make it comparable to M_{it}.

For t = 2005, 2006, 2007, 2008, and

i = {16 provinces with the highest number of patents and adopted PPP in 2005-2009}, we calculate

$$\Phi(\gamma) = \frac{1}{TN}\sum_{t=2005}^{2008}\sum_{i=1}^{16}\frac{\frac{M_{i,t+1}/1.2-M_{it}}{R_{i,t+1}-R_{it}} \cdot \frac{R_{i,t+1}+R_{it}}{M_{i,t+1}/1.2+M_{it}}}{\frac{1-\gamma}{1-\left[\left(\frac{R_{it}+R_{i,t+1}}{2}\right)^{1-\gamma}\right]^{-1}}}$$

where $T = 4$ and $N = 16$. For the model to be considered a good fit for the elasticity, $\Phi(\gamma)$ should not significantly differ from 1.

We obtain the following table.

γ	0.6	0.7	0.8	0.9	1.1
$\Phi(\gamma)$	1.02	1.31	1.76	2.53	6.42

Thus, when $\gamma = 0.6$, the fit of the model for the elasticity of M_{it}/β_{it} with respect to R_{it} is very good. As γ increases, the fit for the elasticity falls. The fit of the model is considered unacceptable at $\gamma = 0.9$, when $\Phi(\gamma) > 2$. Thus, the plausible values of γ is the outcome of the compromise between (i) the fit of the model in predicting R_{it}, which increases with γ, and (ii) the fit of the model in predicting the elasticity of M_{it} with respect of R_{it}, which decreases with γ. When $\gamma < 0.6$, the model's prediction of R_{it} is unacceptable; when $\gamma > 0.8$, the model's prediction of the elasticity of M_{it} with respect of R_{it} is unacceptable. Therefore, we conclude that the plausible values of γ are $\{0.6, 0.7, 0.8\}$. In addition, we assume that each of the plausible values $\{0.6, 0.7, 0.8\}$ is equally probable.

- The plausible range of the fraction by which R&D cost is reduced by PPP, s, is determined as follows.

  Based on data for 2011-2015, Jin, Shang and Xu (2018) find that the elasticity of R&D with respect to R&D subsidy is equal to 0.442.[44] According to (43) and (45) (omitting the error term),

$$R_{it} = \frac{A_t}{\mu_t}\delta \cdot \left(\frac{1}{1-s}\right)\frac{\beta_{it}}{1-\gamma}\left[(R_{it}+1)^{1-\gamma}-1\right] \tag{52}$$

$$\begin{aligned}
\Longrightarrow \quad \frac{dR_{it}}{ds} &= \frac{R_{it}}{(1-s)-\frac{A_t}{\mu_t}\delta\beta_{it}(R_{it}+1)^{-\gamma}} \\
&= \frac{R_{it}}{(1-s)-(1-s)(1-\gamma)R_{it}(R_{it}+1)^{-\gamma}\left[(R_{it}+1)^{1-\gamma}-1\right]^{-1}} \\
&\quad \text{by substituting for } \frac{A_t}{\mu_t}\delta\beta_{it} \text{ from (52)}
\end{aligned}$$

$$\Longrightarrow \quad \frac{s}{R_{it}}\frac{dR_{it}}{ds} = \frac{s}{1-s}\left[\frac{(R_{it}+1)^{1-\gamma}-1}{(R_{it}+1)^{-\gamma}(R_{it}\gamma+1)-1}\right] = 0.44$$

$$\begin{aligned}
\Longrightarrow \quad s &= \left\{\frac{1}{0.44}\cdot\frac{(R_{it}+1)^{1-\gamma}-1}{\left[(R_{it}+1)^{-\gamma}(R_{it}\gamma+1)-1\right]}+1\right\}^{-1} \\
&= \frac{0.44\left[(R_{it}+1)^{-\gamma}(R_{it}\gamma+1)-1\right]}{(R_{it}+1)^{1-\gamma}+0.44(R_{it}+1)^{-\gamma}(R_{it}\gamma+1)-1.44}
\end{aligned}$$

[44] On p.9 of the paper, they state: "This result means that if the government increases subsidies by 1%, manufacturing enterprises are likely to increase their R&D investment by 0.442%."

For $\gamma = 0.6$, 0.7, 0.8, respectively, we calculate

$$s = \frac{1}{T}\sum_t \frac{1}{N}\sum_i \frac{0.44\left[(R_{it}+1)^{-\gamma}(R_{it}\gamma+1)-1\right]}{(R_{it}+1)^{1-\gamma}+0.44\,(R_{it}+1)^{-\gamma}(R_{it}\gamma+1)-1.44}$$

for $t \in \{2005\text{-}2009\}$; $T = 5$;

$i \in \{16$ provinces with the highest number of patents and adopted PPP in 2005-2009$\}$; $N = 16$. The value of s that we assume in our analysis should not deviate too much from the "plausible" value of s calculated above. It turns out that we obtain the following ressult.

γ	0.6	0.7	0.8
s	0.20	0.23	0.25

If we restrict ourselves to considering only $s \in \{0.1, 0.2, 0.3\}$, then $s = 0.2$ (which in fact covers the range 0.15 to 0.25) is the closest value to the above estimates of s. However, we believe the point estimate of Jin, Shang and Xu (2018) is probably imprecise, so we include $s = 0.1$ as another plausible value. We consider $s = 0.3$ to be too large to be plausible. In addition, we assume that it is equally probable that $s = 0.1$ or $s = 0.2$.

- The shape parameter of the Pareto distribution, θ. The plausible range of θ is determined as follows.

  The data L_t, $\{M_{it}\}$, $\{R_{it}\}$ and the set of provinces that adopted PPP are different for different years. Thus, for the same γ and s, the plausible range of θ for which the model is consistent with the data is different for different years. The minimum plausible value of θ is a value infinitesimally larger than 1 (which we denote by 1^+), as any value of θ less than or equal to 1 implies that the mean of the inventive step ζ is equal to infinity. It also implies that the expected price markup (price/unit cost) is infinity. When θ is small (close to 1), the implied average inventive step is large, and the hazard rate is small. As θ increases, the implied average inventive step gets smaller, and the hazard rate increases. For each year t, conditional on L_t, $\{M_{it}\}$, $\{R_{it}\}$ and the set of provinces that adopted PPP, the maximum plausible θ (call it $\theta_t^{\max}$) corresponds to a hazard rate equal to infinity. The model cannot be consistent with the data for any value of θ greater than $\theta_t^{\max}$.

  To put a lower bound on the expected price markup, we do not consider any θ greater than 20 for any t.[45] Thus, in some years, $\theta_t^{\max}$ is set at 20 even though the value of θ corresponding to a hazard rate of infinity is greater than 20. It turns out that these are the years during which the total number of year-province adopting PPP is less than 1/3 the total number of year-province

[45] This is equivalent to restricting the expected price markup (i.e. price/unit cost ratio) to no less than 1.05, which is quite low. A lower expected price markup means that the market is more competitive and the hazard rate is higher, as more minor innovations are allowed to be patentable.

Table 10: Probability that PPP is Welfare-Improving in Various Years

	$s = 0.1$ or 0.2			$s = 0.1$			$s = 0.2$			
Year	$\gamma = 0.6$	$\gamma = 0.6$ or 0.7	$\gamma = 0.6$, 0.7, 0.8	$\gamma = 0.6$	$\gamma = 0.6$ or 0.7	$\gamma = 0.6$, 0.7, 0.8	$\gamma = 0.6$	$\gamma = 0.6$ or 0.7	$\gamma = 0.6$, 0.7, 0.8	Number of year-province adopting PPP
1995	20.05%	13.78%	9.19%	20.05%	15.04%	10.03%	20.05%	12.53%	8.35%	1
1996	12.53%	6.27%	4.18%	15.04%	7.52%	5.01%	10.03%	5.01%	3.34%	1
1997	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2
1998	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3
1999	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	5
2000	15.04%	10.03%	6.68%	20.05%	15.04%	10.03%	10.03%	5.01%	3.34%	9
2001	20.05%	15.04%	10.86%	25.06%	20.05%	15.04%	15.04%	10.03%	6.68%	14
2002	30.08%	26.32%	22.56%	35.09%	32.58%	28.40%	25.06%	20.05%	16.71%	20
2003	35.09%	31.33%	27.57%	40.10%	37.59%	33.42%	30.08%	25.06%	21.72%	26
2004	37.04%	31.75%	28.22%	39.00%	33.43%	29.71%	35.09%	30.08%	26.73%	27
2005	39.12%	36.32%	31.67%	41.30%	38.35%	33.43%	36.94%	34.30%	29.90%	29
2006	41.44%	35.52%	31.57%	43.89%	37.62%	33.44%	39.00%	33.43%	29.71%	29
2007	39.57%	33.44%	29.35%	40.13%	33.44%	28.99%	39.00%	33.43%	29.71%	29
2008	36.73%	31.80%	27.92%	35.84%	32.26%	28.67%	37.62%	31.35%	27.17%	29
2009	38.84%	34.96%	31.07%	41.84%	37.66%	33.47%	35.84%	32.26%	28.67%	29
1995 to 2009	34.14%	29.68%	25.90%	36.56%	32.29%	28.21%	31.72%	27.06%	23.58%	253

Note: The table shows the probability that PPP is welfare-improving at different years for different combinations of γ and s. The values of ρ and δ do not affect the probabilities at all. The probabilities in 1997, 1998 and 1999 are zero as we capped the value of θ at 20 for all years to simplify computation.

with PPP during the sample period.[46] The probability that PPP is welfare-improving even if we do not put a cap on θ would not be large for these years. So, the probability that PPP is welfare-improving during the entire sample period will be underestimated, but the percentage underestimation will not be too large.

In sum, the set of plausible values of θ in year t is $\{1^+, 2, 3, ..., \theta_t^{\max}\}$. In addition, we assume that each of the plausible values is equally probable, as there is no reason to believe that any value is more probable than the others.

D Detailed breakdown of the probabilities that PPP is welfare-improving for different t, γ and s

<Table 10 about here>

[46] These years are 1995-2003, which together account for 32% of a total of 253 year-province adopting PPP.