Discussion Papers In Economics And Business

Technology Diffusion, Pareto Distribution, and Patent Policy

Keiichi Kishi

Discussion Paper 16-31

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

Technology Diffusion, Pareto Distribution, and Patent Policy

Keiichi Kishi

Discussion Paper 16-31

December 2016

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

Technology Diffusion, Pareto Distribution, and Patent Policy*

Keiichi Kishi†

December 16, 2016

Abstract

We develop a Schumpeterian growth model based on technology diffusion. Each firm has a different productivity level. New entrants enter into the targeted industries by learning the existing technologies owned by the other firms. Some of the new entrants succeed to adopt the frontier technology. The other new entrants may adopt the non-frontier technologies. We show that if it is extremely difficult to adopt the frontier technology, the technology diffusion generates the Pareto distributions of firm size, productivity, and innovation size. Further, we introduce the minimum innovation size required for a patent into the model. That is, the patent office grants the patents only for superior inventions. We show that an increase in minimum innovation size may reduce the average patentable innovation size because of an endogenous response of the distribution of innovation size. This implies that if the patent office requires the superior innovations for the patents, it may cause innovators to produce a larger amount of inferior patentable innovations.

JEL-Classification: O30, O33, O34
Keywords: Technology diffusion, Innovation, Pareto distribution

*The author acknowledges the financial support from JSPS Research Fellowships for Young Scientists. I am responsible for any remaining errors.

†Graduate School of Economics, Osaka University, 1-7 Machikaneyama, Toyonaka, Osaka 560-0043, Japan, E-mail address: ba000069@gmail.com

1 Introduction

A Pareto (Pareto-tailed) distribution often emerges in the real economy, for example, firm size (sales and employees), firm's productivity level (total factor productivity (TFP)), innovation size (citations of patents and scientific papers, and financial returns from licensing fees of patents), income, wealth, consumption, consumption growth, population of cities, etc.[1] A Pareto distribution has the following counter-cumulative distribution function:

$$\Pr\{X \geq x\} = \left(\frac{x_{min}}{x}\right)^{\zeta} \text{ for } x \geq x_{min},$$

where scale parameter $x_{min} > 0$, shape parameter $\zeta > 0$, and X is a random variable.[2] Why do some variables follow the Pareto distribution? If there exists a unique mechanism to generate a Pareto distribution, it would explain the evolution of any variable that follow a Pareto distribution. However, the previous studies found the several mechanisms to generate a Pareto distribution. Therefore, we must develop several models to match the evolution of each variable. This paper mainly focuses on the firm size, firm's productivity level, and innovation size.

To generate the long-run Pareto distributions, we develop the dynamic general equilibrium model, which combines the elements of Aghion and Howitt (1998, Ch.3), Lucas and Moll (2014), and Perla and Tonetti (2014). Each firm tries to improve its productivity levels (technological level, product management, product quality, etc.) by gaining the existing knowledge, which is owned by the other firms. Some of the firms may succeed at adopting the frontier knowledge (productivity) level into their productions. On the other hand, the other firms may adopt the non-frontier level of knowledge into their productions. The existing knowledge diffuses across many firms over time. These knowledge diffusions (knowledge infections) improve the overall productivity level in the economy over time. We show that if it is extremely difficult to adopt the frontier technology, these diffusion processes generate the Pareto-tailed distributions of firm size, productivity level, and innovation size. The results are consistent with the empirical facts (see, e.g., Fujiwara et al. 2004, Luttmer 2007, Reed 2001 for the firm size, König et al. 2016 for the firm's TFP, and Silverberg and Verspagen 2007 for the innovation size). More precisely, each firm can adopt the frontier technology with a probability, as in Aghion and Howitt (1998, Ch.3). If the firms fail to adopt the frontier technology, they try to adopt the non-frontier technology, which is drawn from the endogenous productivity distribution across all industries, as in Lucas and Moll (2014) and Perla and Tonetti (2014). We show that if the probability for adopting a frontier technology is sufficiently small, the

[1]See, for example, Gabaix (2009), König et al. (2016), Luttmer (2007), Newman (2005), Reed (2001), Silverberg and Verspagen (2007), and Toda and Walsh (2015).

[2]The distribution has a Pareto tail (fat tail), implying that there are $c > 0$ and $\zeta > 0$ such that $\lim_{x\to\infty} \Pr\{X \geq x\}/x^{-\zeta} = c$, where X is a random variable. When the literature refers to the Pareto distribution, it usually means that the distribution has a Pareto tail, meaning that it takes a power-law form for large x.

distributions of firm size, productivity, and innovation size become the Pareto distributions in the balanced-growth equilibrium. Furthermore, we yield the *endogenous* shape parameter of the long-run Pareto distributions. These results crucially differ from the results of Lucas and Moll (2014) and Perla and Tonetti (2014). See section 2 for the differences between this paper and the previous studies of Aghion and Howitt (1998, Ch.3), Lucas and Moll (2014), and Perla and Tonetti (2014).

To demonstrate the importance of the endogenous shape parameter, we apply the model to a patent policy. In many countries, patent laws have designated letters for the minimum innovation sizes required for patentability, that is, the patent office grants a patent only for a sufficiently superior invention compared to prior inventions. For example, the "inventive step" requirement in Europe corresponds to the minimum innovation size, while it is synonymous with the "nonobviousness" requirement of the United States. The Federal Trade Commission (FTC) (2003) and the National Academy of Sciences (NAS) (2004) had recommended strengthening the minimum innovation size in order to raise the overall quality of patents by reducing the number of improperly issued patents. Thereafter, in the decision of the United States Supreme Court for KSR International Co. v. Teleflex Inc. (2007), the minimum innovation size substantially increased in the United States.[3]

Inspired by these developments, we analyze the relationship between the minimum innovation size required for a patent and the average patentable innovation size. We show that a Pareto distribution describes the long-run distribution of patentable innovation size. If the shape parameter of a Pareto distribution is exogenous, an increase in minimum innovation size (scale parameter) always raises the average patentable innovation size. However, an increase in minimum innovation size reduces the average patentable innovation size only if it simultaneously raises the endogenous shape parameter. This is because a higher shape parameter has a negative effect on the average value of a Pareto distribution. More precisely, a higher shape parameter reduces the probability of large innovation sizes, thus reducing the average patentable innovation size. Since we show that there exists a negative relationship between minimum innovation size and average patentable innovation size under some assumptions, the average patent quality in the United States may decrease after the United States Supreme Court decision in 2007, which increased minimum innovation size.

The remainder of the paper is organized as follows. In Section 2, we survey the related literature. In Section 3, we outline the dynamic general equilibrium model, which incorporates the elements of Aghion and Howitt (1998, Ch.3), Lucas and Moll (2014), and Perla and Tonetti (2014). Further, we introduce the minimum innovation size into the model. In Section 4, we derive the productivity distributions. In Section 5, we consider the functional form of economic growth rate. In Section 6, we show the existence of a unique balanced-growth

[3]The purpose of this paper for the patent policy is the same as that of Kishi (2014). See Kishi (2014) for more detailed discussion of minimum innovation size.

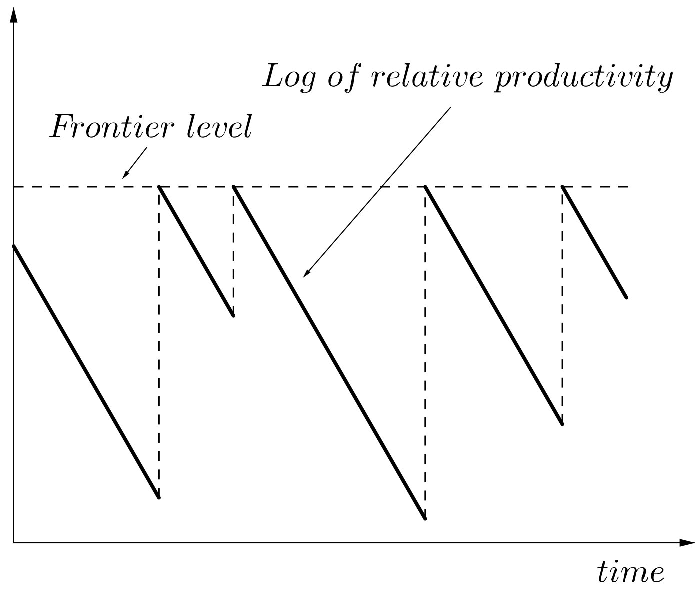

Figure 1: The exponential negative growth and the Poisson jump to the deterministic value generate a power-law distribution.

equilibrium. Section 7 analyzes the effects of minimum innovation size on the average patentable innovation size and growth. Section 8 concludes.

2 Related literature

There are some mechanisms generating a Pareto distribution. Steindl (1965) is a simple model to generate a Pareto distribution. The first application of Steindl (1965) to an endogenous growth model is in Aghion and Howitt (1998, Ch.3). In Aghion and Howitt's model (1998, Ch.3), the firm's relative productivity (the productivity divided by the frontier technology) decreases at a constant rate because of the growth of the frontier technology (see Fig.1). This implies the obsolescence of the existing knowledge. However, the firm's productivity can jump to the frontier level if the firm succeeds in an innovation with a constant Poisson arrival rate. This mechanism generates a power-law distribution of the firm's productivity with a bounded support. Furthermore, the innovation size (the frontier technology divided by the productivity level) obeys a Pareto distribution with no upper bound of the support (see also Kishi 2014).

Lucas and Moll (2014) and Parla and Tonetti (2014) generalize Aghion and Howitt's model (1998, Ch.3) in the sense that the innovation size is a random variable, which is drawn from the endogenous productivity distribution across all industries. For example, in Lucas and Moll's (2014) model, the firms can draw the existing knowledge (productivity) from the productivity distribution with a Poisson arrival rate. They decide to adopt the existing knowledge if the drawn productivity is superior to the firm's current productivity. In Parla and Tonetti's (2014) model, a firm conducts the research and development (R&D) if the firm's productivity reaches a lower threshold level because of the obsolescence of the productivity. This is because the low-productivity firm has an advantage to improve the productivity by adopting the other firm's superior productivity. Then, a sufficiently low-productivity firm adopts the existing knowledge, which is drawn from the productivity distribution. In sum, in contrast to Aghion and Howitt (1998, Ch.3), the models of Lucas and Moll (2014) and Perla and Tonetti (2014) admit to jump to the non-frontier technological level because of innovation. Lucas and Moll (2014) and Perla and Tonetti (2014) show that if the initial productivity distribution has a Pareto tail (fat tail), the economy experiences the sustained growth in the balanced-growth equilibrium. Further, Perla and Tonetti (2014) show that if the initial distribution is a Pareto distribution, the long-run distribution is also a Pareto distribution, whose shape parameter corresponds to that of initial distribution (see also Perla et al. 2015). Lucas and Moll (2014) also show a similar result under a constant Poisson arrival rate. The assumption of initial Pareto distribution implies that many firms have very large productivities at the initial period, that is, we require an unbounded support and a fat tail of the initial productivity distribution. Further, note that the initial distribution is historically determined, that is, it describes the initial state of the economy. Since the Pareto distribution emerges in many countries, it is unreasonable to apply the results of Lucas and Moll (2014) and Perla and Tonetti (2014). For example, see Fujiwara et al. (2004) and Luttmer (2007) for the Pareto-tailed distributions of firm size in the United States and European countries. Therefore, we develop the sustained-growth model generating a Pareto distribution in the balanced-growth equilibrium, irrespective of initial distribution.

Our stochastic process of the industry's (firm's) productivity also relates to that of Luttmer (2007, 2012). In Luttmer's (2007, 2012) models, the incumbent firm's productivity follows a geometric Brownian motion. This implies that the expected growth rate of the productivity is constant over time. If the firm's productivity reaches a lower threshold level, the firm exits from the market since the sufficiently low-productivity firm anticipates the long-time negative profits stream in the future due to the payment of fixed cost for the production. New entrants (R&D firms) enter into the economy, equipped with a deterministic level of productivity. Luttmer (2007, 2012) shows that this stochastic process generates a Pareto tail in the stationary productivity distribution. By contrast, in our model, the growth rate of the industry's (relative) productivity level is simply deterministic and constant over time when the productivity improvement does not occur.[4] The incumbents exit from the market if the productivity level

[4]See Perla et al. (2015), who introduced the geometric Brownian motion into Perla and Tonetti's (2014) model.

reaches a lower threshold level, because of the entry of new firms equipped with superior productivity. That is, Schumpeter's creative destruction works as the force to exit the low-productivity firms. Furthermore, the new entrant's productivity is a random variable, which takes either a frontier level or the random value drawn from the endogenous productivity distribution.

There are many recent studies about Pareto distribution. Acemoglu and Cao (2015) and König et al. (2016) introduce the technology adoption (technology diffusion) into the standard quality ladder models of Acemoglu (2009, Ch.14), and Grossman and Helpman (1991, Ch.4). In Acemoglu and Cao's (2015) model, the imitation firms can enter into the targeted industry, equipped with a deterministic productivity level. König et al. (2016) combine the models of Aghion and Howitt (1998, Ch.3) and Grossman and Helpman (1991, Ch.4). Then, they show a Pareto tail of the stationary productivity (firm size) distribution. Further, as is well known, the geometric Brownian motion with a Poisson jump to a deterministic value generates a Pareto tail in the stationary distribution (see Gabaix 2009). This can be interpreted as a generalization of Steindl (1965), in which the growth rate of the variable is constant over time when the variable does not jump. This generalized version of Steindl (1965) is usually applied to the economic models (see, e.g., Benhabib et al. 2016, Jones and Kim 2015, Toda and Walsh 2015). The Kesten (1973) process is known as the discrete-time stochastic process generating a Pareto tail. Nirei and Aoki (2016) apply the Kesten process to generate the Pareto distributions of income and wealth. Further, see Gabaix et al. (2016), who summarize the previous findings of the continuous-time stochastic processes generating a Pareto tail. Then, Gabaix et al. (2016) analytically show the transitional dynamics of the distributions and the speeds of convergence.

As part of the literature on patent policy, this paper relates to Kishi (2014), Koléda (2008), and O'Donoghue and Zweimüller (2004), who introduce the minimum innovation size required for a patent into the R&D-based growth models. Kishi (2014) introduces the minimum innovation size into the model of Aghion and Howitt (1998, Ch.3). Therefore, the stationary distribution of innovation size is an endogenous Pareto distribution. Then, Kishi (2014) shows an increase in minimum innovation size and reduces the average innovation size. This is because a higher minimum innovation size causes smaller accumulation of the low-productivity firms. Then, it reduces the average innovation size, since the low-productivity firms tend to attain the large innovation size by adopting the frontier technology. By contrast, this paper provides the opposite results. An increase in minimum innovation size causes larger accumulation of the low-productivity firms. Then, the firms tend to adopt lower productivity levels by learning the existing lower levels of technologies. This leads to lower average innovation size. Further, in Kishi's (2014) model, the distributions of productivity and firm size do not have a Pareto tail, which is inconsistent with empirical facts. In contrast, this paper shows that the distributions of productivity, firm size, and innovation size have a Pareto tail. Koléda (2008) considers the exogenous Pareto distribution of innovation size. Therefore, higher minimum innovation size always raises the mean value of average patentable innovation size. In the model of O'Donoghue and Zweimüller (2004), each firm optimally chooses the endogenous innovation size, which corresponds to the minimum innovation size. That is, the distribution of innovation size is degenerate, which is inconsistent with empirical facts (see, e.g., Silverberg and Verspagen 2007). Then, higher minimum innovation size always raises the mean value.

3 Theoretical framework

To examine the stationary distributions of firm size, productivity, and innovation size, we develop a Schumpeterian growth model based on the technology diffusion in which time is continuous. Further, we examine the effect of the minimum innovation size on the quality of innovations. We focus on the balanced-growth equilibrium, in which all endogenous variables grow at constant rates.

3.1 Households

There is a representative household with CRRA (constant relative risk aversion) preferences given by

$$U = \int_0^{\infty} e^{-\rho t} \frac{C(t)^{1-\theta} - 1}{1-\theta} dt \tag{1}$$

where $\rho > 0$ and $\theta \geq 1$ are the subjective discount rate and the inverse of the intertemporal elasticity of substitution, respectively.[5] $C(t)$ denotes the consumption per capita at date t. There is no population growth. The representative household's optimization problem implies the well-known Euler equation for consumption:

$$\frac{\dot{C}(t)}{C(t)} = \frac{1}{\theta}(r - \rho) \tag{2}$$

and the transversality condition (TVC)

$$\lim_{t \to \infty} e^{-\rho t} C(t)^{-\theta} V(t) = 0 \tag{3}$$

where r is the interest rate, which is constant over time in the balanced-growth equilibrium, and $V(t)$ is the asset value per capita at date t.

3.2 Final goods

The final good, which we take as the numéraire, is produced under perfect competition, according to the production function

$$Y(t) = L^{1-\alpha} \int_0^1 A(i,t)^{1-\alpha} x(i,t)^{\alpha} di, \tag{4}$$

[5] We restrict our attention to the case $\theta \geq 1$, which is usually satisfied in the real economy. See, for example, Havranek et al. (2015), which reports $\theta = 10/6$ in the United States by using 1429 estimates of published studies.

where $\alpha \in (0,1)$, $Y(t)$ is gross output at date t, L is the population size, $x(i,t)$ is the amount of intermediate product $i \in [0,1]$ at date t, and $A(i,t)$ is a productivity level for the latest version of intermediate product i at date t. The final good can be used interchangeably for consumption as an intermediate input or as a R&D input. The inverse demand function for $x(i,t)$ derived from Eq. (4) is $p(i,t) = \partial Y(t)/\partial x(i,t) = \alpha L^{1-\alpha} A(i,t)^{1-\alpha} x(i,t)^{\alpha-1}$

3.3 Intermediate goods

One unit of intermediate product is produced using one unit of the final good. The latest innovator in industry i chooses price $p_j(i,t)$ to maximize profit $\Pi_j(i,t) \equiv p_j(i,t)x_j(i,t) - x_j(i,t)$ where $j = M, C$. There are two states $j = M, C$ in each industry. The state $j = M$ $(j = C)$ represents the industry where the latest innovator's technology is (not) protected by the patent. Suppose that the latest patented innovator can set the monopoly price $p_M(i,t) = 1/\alpha$ under the strong patent protection[6] and thus can earn the monopoly profit $\Pi_M(i,t) = \tilde{\pi}_M A(i,t)$, where $\tilde{\pi}_M \equiv [(1-\alpha)/\alpha] L \tilde{x}_M$ and $\tilde{x}_M \equiv \alpha^{\frac{2}{1-\alpha}}$. The amount of production in the monopoly industry is $x_M(i,t) = \tilde{x}_M L A(i,t)$. Suppose that unpatented technology can be imitated immediately at no cost, and thus, the industry becomes perfectly competitive. Then, the price is $p_C(i,t) = 1$, and thus, the profit for unpatented technology is $\Pi_C(i,t) = 0$. The amount of the production for the competitive industry is $x_C(i,t) = \tilde{x}_C L A(i,t)$, where $\tilde{x}_C \equiv \alpha^{\frac{1}{1-\alpha}}$. Note that $\tilde{x}_C > \tilde{x}_M$ holds. Further, note that the firm size (sales), that is, $p_j(i,t)x_j(i,t)$, is proportional to the productivity $A(i,t)$. Therefore, if the distribution of $A(i,t)$ has a Pareto tail, then the firm size distribution also has a Pareto tail.

To simplify the analysis, we define the log of relative productivity $a(i,t) \equiv \ln[A(i,t)/B(t)]$, where $B(t) \equiv \min\{A(i,t) | i \in [0,1]\}$ is the minimum productivity level in the economy. From these definitions, the monopoly profit can be rewritten as $\Pi_M(i,t) = \tilde{\pi}_M A(i,t) = \tilde{\pi}_M e^{a(i,t)} B(t)$. Then, define the minimum productivity $B(t)$-adjusted profit $\pi_M(i,t) \equiv \Pi_M(i,t)/B(t) = \tilde{\pi}_M e^{a(i,t)}$. Since each $B(t)$-adjusted monopoly profit depends only on the log of relative productivity $a(i,t)$, omitting the industry index i and time t, we simply restate the $B(t)$-adjusted monopoly profit as

$$\pi_M(a) = \tilde{\pi}_M e^a \tag{5}$$

for the monopoly industry, whose log of relative productivity level is $a \equiv \ln A/B$. Later, we may simply state the variable a as either relative productivity or productivity.

3.4 Frontier technology

Following Aghion and Howitt (1998, Ch.3), suppose that there exists the public knowledge $\bar{A}(t)$, whose initial value corresponds to the frontier technology at

[6] For example, consider the sufficiently strong patent breadth, which ensures the monopoly price (see Li 2001).

initial period, that is, $\bar{A}(0) = \max\{A(i,0)|i \in [0,1]\}$. Define g as the growth rate of the public knowledge, that is, $g \equiv \dot{\bar{A}}(t)/\bar{A}(t)$, which can be endogenized later. We can show that the public knowledge coincides with frontier technology for all time, that is, $\bar{A}(t) = \max\{A(i,t)|i \in [0,1]\}$ for all $t \geq 0$, since we later show that there exist some industries whose technological level is $\bar{A}(t)$ for all t, and each productivity level $A(i,t)$ cannot exceed $\bar{A}(t)$ for all t.

Define $f(a)$ as the stationary probability density function for the relative productivity a. Note that the support of the distribution is $a \in [0, \bar{a}]$, since the minimum support is $\ln[B(t)/B(t)] = 0$ and the maximum support is $\bar{a} \equiv \ln[\bar{A}(t)/B(t)]$. If the stationary distribution $f(a)$ exists, the support of the distribution must be constant over time. Therefore, the growth rate of $B(t)$ must be equal to that of $\bar{A}(t)$ in the balanced-growth equilibrium, that is, $g = \dot{B}(t)/B(t)$.

3.5 R&D and technology diffusion

Innovations result from R&D activity. Each R&D firm can freely target any industry $i \in [0,1]$ as a candidate industry for the entry. We consider the following R&D activities, which combine the elements of Aghion and Howitt (1998, Ch.3), Lucas and Moll (2014), and Perla and Tonetti (2014). Using the fixed amount $RB(t)$ of final goods, where the parameter $R > 0$ represents the $B(t)$-adjusted fixed cost, the R&D firms (potential new entrants) can attain the frontier technology $\bar{A}(t)$ with exogenous probability $p \in (0,1)$. This setup follows Aghion and Howitt (1998, Ch.3). On the other hand, with probability $1-p$, the R&D firms can draw the productivity from the endogenous productivity distribution $f(a)$. If the new productivity a caused by R&D is higher than the productivity a' in the targeted industry, that is, $a > a'$, the R&D firm enters into the targeted industry as a new incumbent firm equipped with new superior productivity a. This setup follows Lucas and Moll (2014) and Perla and Tonetti (2014).

In summary, each R&D firm tries to improve the productivities (technological level, product management, product quality, etc.) of the existing production processes by learning the existing knowledge owned by the existing firms. These knowledge adoptions (diffusions) create new firms equipped with frontier knowledge level with probability p. On the other hand, the knowledge adoptions are incomplete with probability $1-p$, in the sense that they create new firms whose individual productivities are less than the frontier level.

3.6 Patent value

Define the value of the firm as $V_j(a)B(t)$, where $V_j(a)$ is the $B(t)$-adjusted value of the firm for the relative productivity a and the state $j = M, C$. Since the state $j = C$ is perfectly competitive, the value of the firm must be zero:

$$V_C(a) = 0 \text{ for all } a \geq 0. \tag{6}$$

We later show that when the productivity a reaches the minimum level $a = 0$, the incumbent firm exits from the industry due to the entry of the R&D firm.

That is, the creative destruction occurs only at the industries whose productivity is the minimum in the economy. Then, the $B(t)$-adjusted patent value $V_M(a)$ must follow the following Hamilton-Jacobi-Bellman (HJB) equation

$$(r-g)V_M(a) = \pi_M(a) - gV_M'(a) \text{ for all } a > 0 \tag{7}$$

and the boundary condition

$$V_M(0) = 0. \tag{8}$$

According to the HJB equation (7), the boundary condition (8), and the monopoly profit (5), we obtain the $B(t)$-adjusted patent value

$$V_M(a) = \left(\frac{\tilde{\pi}_M}{r}\right) e^a[1 - e^{-rT(a)}] \tag{9}$$

where $T(a) \equiv a/g$. Note that $T(a)$ represents the time to reach the minimum productivity for the monopolist whose current productivity is a. More precisely, $a = (a-0)$ of the numerator of $T(a)$ is the distance from the current productivity a to the minimum productivity $a = 0$. On the other hand, g in the denominator of $T(a)$ represents the speed to run the given distance $(a-0)$. Therefore, higher a and lower g cause longer time $T(a)$ until the incumbent firm exits from the industry, which has a positive effect on the patent value $V_M(a)$. This implies that $T(a)$ represents *the effective patent length*, which is the waiting time until creative destruction occurs for the monopolist whose current productivity is a.

3.7 Minimum innovation size and optimal industry choice

The patent office grants the patent if and only if the innovation size (the ratio between the new entrant's absolute productivity A and the incumbent's absolute productivity A') exceeds $\chi \geq 1$, where the policy parameter χ represents the minimum innovation size. That is, the patent office grants a patent only for a superior innovation. As shown below, the R&D firms target only the industries whose productivity level is minimum in the economy. Then, $A/B = e^a$ represents the innovation size. In summary, the patent office grants the patent if and only if $e^a > \chi \iff a > \ln \chi$.

So far, we have developed the model under the following guess. The R&D firms target only the industries whose productivity is minimum in the economy. Now, we verify this guess, that is, we show that any infinitesimal R&D firms do not have an incentive to target the other industries. First, it is obvious that R&D firms do not target the industry a' where $a' \geq \bar{a} - \ln \chi$ holds. This is because R&D firms cannot obtain the patent with probability one, and thus, they yield the negative R&D profit $-RB(t) < 0$. Next, consider the case of $a' < \bar{a} - \ln \chi$. R&D firms conduct the R&D investment in the industries where the expected patent value is largest in the economy. Targeting industry a', the expected $B(t)$-adjusted patent value (R&D benefit) is given by

$$V_R(a') \equiv (1-p)\int_{a'+\ln\chi}^{\bar{a}} V_M(a)dF(a) + pV_M(\bar{a}) \text{ for all } a' < \bar{a} - \ln \chi \tag{10}$$

where $F(a)$ represents the cumulative distribution function for a. Note that the infinitesimal R&D firms take the value function $V_M(a)$, the productivity distribution $F(a)$, and the maximum productivity $\bar{a}$ as given. Differentiating Eq. (10) with respect to a' yields

$$\frac{\partial V_R(a')}{\partial a'} = -(1-p)f(a'+\ln\chi)V_M(a'+\ln\chi) < 0 \text{ for all } a' < \bar{a} - \ln\chi,$$

since both $f(a'+\ln\chi) > 0$, which can be shown later, and $V_M(a'+\ln\chi) > 0$ hold for all $a' < \bar{a} - \ln\chi$. Therefore, R&D firms optimally target the industries whose productivity is minimum $a' = 0$ in the economy. In this case, R&D firms maximize the probability for the patentable innovation for a given distribution $f(a)$. This is because the probability $\mu(a')$ for the patentable innovation in the industry $a' < \bar{a} - \ln\chi$ is $\mu(a') \equiv \int_{a'+\ln\chi}^{\bar{a}} f(a)da$, and thus, $\mu(a')$ is maximized by targeting $a' = 0$. Further, note that any R&D firm obtains the higher productivity a than the incumbent's productivity a' with probability one, regardless of whether the innovations are patentable. This is because R&D firms target the minimum productivity industries.

Since the R&D firms enter only into the industries where the productivity is minimum, the free-entry (zero-profit) condition must hold at the industries:

$$R = V_R \tag{11}$$

where

$$V_R \equiv V_R(0) = (1-p)\int_{\ln\chi}^{\bar{a}} V_M(a)dF(a) + pV_M(\bar{a}). \tag{12}$$

Then, $R > V_R(a')$ holds for all $a' > 0$.

The above result relates to Luttmer (2007, 2012) and Perla and Tonetti (2014). In Luttmer's (2007, 2012) models, the firms optimally exit from the industries if the productivity levels reach a lower threshold level, which becomes the minimum productivity level in the balanced-growth equilibrium. This is because the sufficiently low-productivity firm forecasts the long-term negative profits because of the payment of the production fixed cost. In Perla and Tonetti's (2014) model, the incumbent firms conduct the R&D investments if the productivity levels reach a lower threshold level, which becomes the minimum productivity level in the balanced-growth equilibrium. This is because, as in our model, the low-productivity firms have the advantage of improving the productivities, since the low-productivity firms tend to draw superior productivities from the endogenous productivity distribution.[7] In our model, Schumpeter's creative destruction (the entry of new firms equipped with superior productivities) occurs at the industries whose productivities are minimum in the economy, since the new entrants tend to draw the superior productivities if they target the minimum productivity industries.

[7] In our model, the incumbents equipped with productivity $a > 0$ do not conduct R&D by paying the fixed cost $RB(t)$, since they have additional (opportunity) cost of losing their current firm values. That is, Arrow (1962) replacement effect works in our model.

The existence of the positive minimum support $B(t)$ for the distribution is known as one of the important forces to generate a Pareto-tailed distribution.[8] However, note that there exist mechanisms generating a Pareto tail without the positive minimum support (see, e.g., Gabaix 2009 and Newman 2005).

4 The productivity distributions

In this section, we derive the productivity distributions for entire industries, monopoly industries, and competitive industries.

4.1 The productivity distribution for entire industries

We derive the stationary distribution $f(a)$ for the log of relative productivity a. Note that $f(a)$ also represents *the log of innovation size distribution conditional on drawing probability* $1-p$. This is because innovations occur only at the minimum productivity industries, and thus, $a \equiv \ln A/B$ is the log of innovation size from the perspective of R&D firms. Fig.2 represents the sample path of the productivity a. The productivity level a decreases at rate g because of the growth of $B(t)$, which implies the obsolescence of existing productivity. If the productivity a reaches the minimum productivity $a=0$, the industry's productivity jumps to a higher value because of innovation. The productivity jumps to the frontier level $\bar{a}$ with probability $p \in (0,1)$. Moreover, with probability $1-p$, the productivity jumps to the non-frontier level, which is drawn from the distribution $f(a)$. We now derive the stationary distribution $f(a)$, which is generated by these dynamics of the productivity a.

Divide time into short intervals of duration $\Delta t > 0$, and the a space into short segments, each of length $\Delta h \equiv g\Delta t$. Note that the productivity a falls by $\dot{a}\Delta t = -g\Delta t \equiv -\Delta h$ during time interval Δt. Define $\phi\Delta t$ as the number (share) of the industries that reaches the minimum productivity $a=0$ during interval Δt.

Now, consider the segment centered at $a \in (0,\bar{a})$, which starts out with $f(a)\Delta h$ industries. In the next unit time period Δt, all these industries move to the left because of the obsolescence (reduction) of relative productivity level a. New entrants, as well as industries from the right, arrive to take their places. In the stationary distribution, the inflow and the outflow of the industries must be equal:

$$f(a)\Delta h = f(a+\Delta h)\Delta h + (1-p)\phi\Delta t f(a)\Delta h \text{ for all } a \in (0,\bar{a}) \tag{13}$$

where the left-hand side of the equation represents the outflow of the industries from the segment a due to the obsolescence of the relative productivity, the first term on the right-hand side is the inflow of the industries to the segment a due to the obsolescence of the relative productivity, and the second term on the

[8] If a random variable follows a geometric Brownian motion with reflection barrier, which creates a positive minimum level of the variable, the stationary Pareto distribution emerges (see Gabaix 2009).

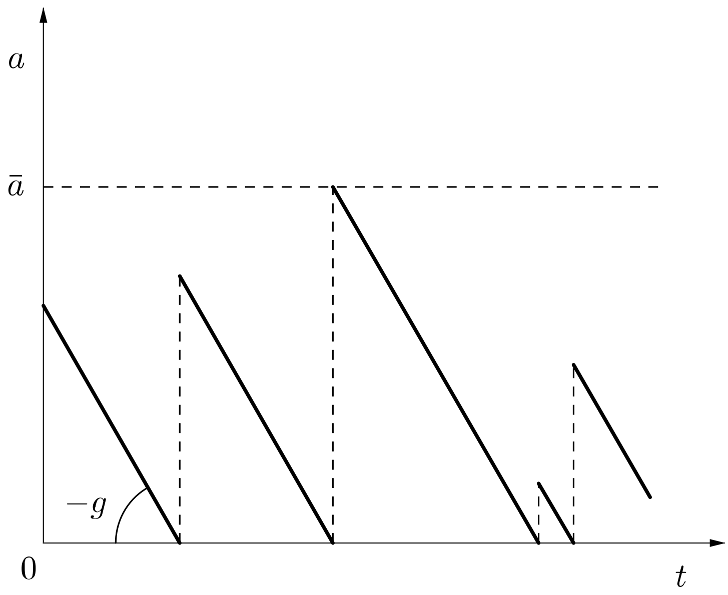

Figure 2: The evolution of the productivity a.

right-hand side is the inflow of the industries due to innovations. Expanding $f(a+\Delta h)$ around a by Taylor's theorem yields

$$f(a+\Delta h) = f(a) + f'(a)\Delta h \tag{14}$$

Cancelling the common factor Δh of Eq.(13), substituting Eq.(14) into the equation, and simplifying this yields the stationary form of the Kolmogorov forward equation (KFE):

$$0 = gf'(a) + (1-p)\phi f(a) \text{ for all } a \in (0, \bar{a}) \tag{15}$$

Next, consider the segment at $\bar{a}$. The outflow of the industries is $f(\bar{a})\Delta h$ because of the obsolescence of the relative productivity. The inflow of the industries is $p\phi\Delta t + (1-p)\phi\Delta t f(\bar{a})\Delta h$ because of innovations. In the stationary distribution, the inflow and the outflow must be equal: $f(\bar{a})\Delta h = p\phi\Delta t + (1-p)\phi\Delta t f(\bar{a})\Delta h \iff gf(\bar{a}) = p\phi + (1-p)\phi g f(\bar{a})\Delta t$. Then, as $\Delta t \to 0$, we yield

$$f(\bar{a}) = \frac{p\phi}{g}. \tag{16}$$

Solving the differential equation (15) by imposing the conditions $1 = \int_0^{\bar{a}} f(a)da$ and $f(\bar{a}) = \lim_{a \uparrow \bar{a}} f(a)$ for the continuity yields the following lemma.

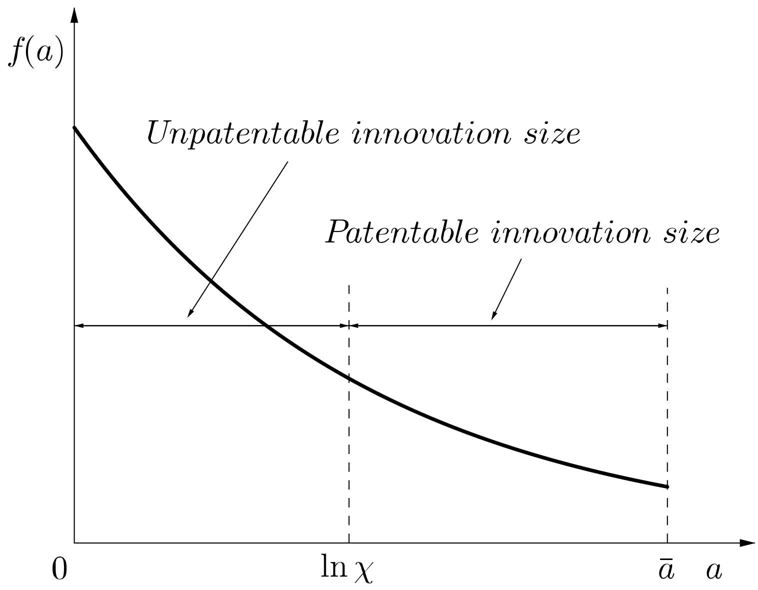

Figure 3: The productivity distribution $f(a)$.

Lemma 1. *The following equations describe the stationary distributions of both the log of relative productivity and the log of innovation size conditional on drawing the probability 1-p:*

$$f(a) = \left(\frac{\eta}{1-p}\right) e^{-\eta a} \text{ for } a \in (0, \bar{a}] \tag{17}$$

where

$$\eta \equiv \frac{(1-p)\phi}{g} > 0 \text{ and} \tag{18}$$

$$\bar{a} = \left(\frac{1}{\eta}\right) \ln\left(\frac{1}{p}\right) > 0. \tag{19}$$

Fig.3 represents the density function $f(a)$. Note that $\bar{a} > \ln\chi$ must hold in the equilibrium to ensure the possibility of patentable innovations. Drawing the productivity $a > \ln\chi$, the R&D firm can obtain a patent, and then, the firm enters into the industries as the new monopolist. On the other hand, drawing the productivity $a \leq \ln\chi$, the R&D firm fails to obtain a patent, and then, the targeted industry becomes perfectly competitive.

Lemma 1 implies that the distribution of the relative productivity $\hat{a} \equiv A/B = e^a$ is a power law. Now, define the probability density function $f_{\hat{a}}(\hat{a})$ for $\hat{a}$. Conducting the transformation of variables, we yield the power-law distribution with bounded support:

$$f_{\hat{a}}(\hat{a}) = \left(\frac{\eta}{1-p}\right)\hat{a}^{-(\eta+1)} \text{ for } \hat{a} \in (1, e^{\bar{a}}]. \tag{20}$$

Later, we show that the distribution (20) is inconsistent with empirical facts. Then, to reconcile the empirical facts and the distribution (20), we later show that it requires $p \to 0$.

4.2 The economic intuitions behind the productivity distribution

We consider the economic intuitions behind the lemma 1. According to Eq.(17), the slope of the density $f(a)$ is always negative. By Eq.(13), we can explain the reason for this as follows. First, Eq.(13) has an alternative interpretation. The right-hand side of Eq.(13) represents the composition of the number $f(a)\Delta h$ of industries equipped with the relative productivity around a. The first term on the right-hand side of Eq.(13) represents the number of industries around a, which comes from the right-hand neighboring segment $a + \Delta h$ due to the obsolescence (reduction) of relative productivity. The second term of Eq.(13) represents the number of industries around a, which comes from the innovations. The sum of these industries corresponds to the total number $f(a)\Delta h$ of the industries equipped with the relative productivity around a. At the maximum productivity $\bar{a}$, there is no inflow of the industries from above due to obsolescence (reduction) of relative productivity. On the other hand, there is the inflow due to innovation, which attains $\bar{a}$. In the neighboring segment $\bar{a} - \Delta h$, there is an inflow of industries from $\bar{a}$ due to the obsolescence, and the inflow of new entrants because of innovation. Then, the total number of industries in the segment $\bar{a} - \Delta h$ is the sum of the number of new entrants in the segment $\bar{a} - \Delta h$ and the number of industries in the segment $\bar{a}$. Therefore, the number of industries in the segment $\bar{a} - \Delta h$ is larger than that in $\bar{a}$. Similarly, consider the next neighboring segment $\bar{a} - 2\Delta h$. There is an inflow of industries from $\bar{a} - \Delta h$ due to the obsolescence, which is equal to the number of industries in the segment $\bar{a} - \Delta h$. Furthermore, there is an inflow of new entrants because of innovation in the segment $\bar{a} - 2\Delta h$. Therefore, the total number of industries in the segment $\bar{a} - 2\Delta h$ is the sum of the number of new entrants in the segment $\bar{a} - 2\Delta h$ and the number of industries in the segment $\bar{a} - \Delta h$. Therefore, the number of industries in the segment $\bar{a} - 2\Delta h$ is larger than that in the segment $\bar{a} - \Delta h$. Maintaining these considerations until the minimum productivity $a = 0$, we can explain the reason why the density $f(a)$ has a negative slope for all a.

Next, consider the economic intuitions behind η. According to lemma 1, higher η leads to larger amount of industries around the minimum productivity $a = 0$ (see Fig.4). According to Eq.(18), higher $(1-p)\phi$ raises η for a given g. On the other hand, higher g reduces η for a given $(1-p)\phi$. We consider the reasons behind these results. First, note that $(1-p)\phi\Delta t$ represents the aggregate number of R&D firms that draw the probability $1-p$ during the interval Δt. As

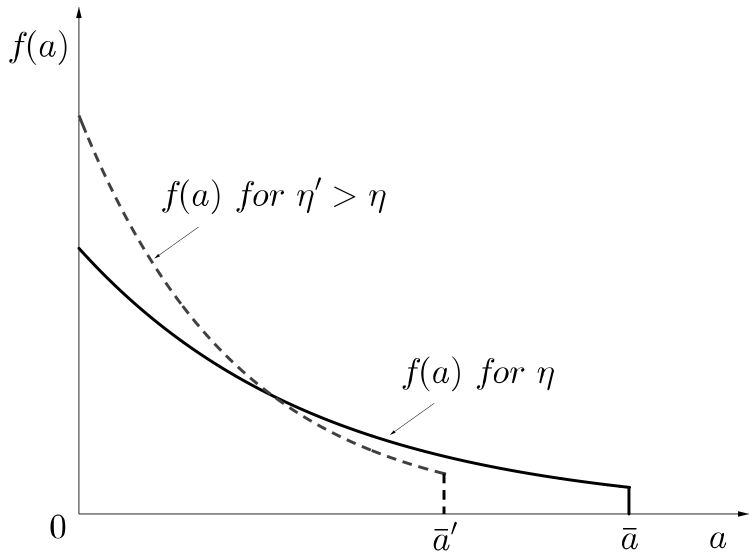

Figure 4: The productivity distributions for η and $\eta'(> \eta)$.

explained in the above paragraph, the industries are accumulated in the lower a region because of the repeated summation of the number of new entrants and the number of industries in the right-hand neighboring segment. Therefore, R&D firms tend to draw the lower productivity from $f(a)$. Then, higher $(1-p)\phi$ causes larger inflow of new entrants into the lower productivity region. This leads to a larger amount of the industries around $a = 0$. Second, higher g implies a faster obsolescence of relative productivity a for a given $(1-p)\phi$, thus causing a smaller accumulation of the industries in each segment a. This leads to the flatter slope of the density $f(a)$.

We consider the economic intuitions behind Eq.(16). Higher $p\phi$ or lower g leads to higher $f(\bar{a})$. This is because higher $p\phi$ implies that a larger amount of R&D firms adopt the frontier technology $\bar{a}$, which causes a larger amount of the industries equipped with $\bar{a}$. Lower g implies a slower obsolescence of the relative productivity. Then, it causes larger amount of the industries equipped with $\bar{a}$.

According to Eq.(19), higher η reduces the maximum support $\bar{a}$ of the distribution $f(a)$. Higher η causes a larger amount of the industries around minimum productivity $a = 0$ (see Fig.4). Then, it requires lower value of $\bar{a}$ to ensure $1 = \int_0^{\bar{a}} f(a)da$.

4.3 The productivity distributions for monopoly industries and competitive industries

Derive the productivity distribution for both the monopoly and competitive industries. Define the density function $f_M(a)$ for the monopoly industries, which satisfies $n = \int_0^{\bar{a}} f_M(a)da$, where $n \in [0, 1]$ represents the total number of monopoly industries. That is, $f_M(a)$ is not the probability density function. Similarly, define the density function $f_C(a)$ for the competitive industries, which satisfies $1 - n = \int_0^{\bar{a}} f_C(a)da$. Therefore, it requires that the sum of the number of monopoly industries equipped with productivity a and that of competitive industries equipped with a must be equal to the total number of industries equipped with a:

$$f(a) = f_M(a) + f_C(a) \text{ for all } a. \tag{21}$$

Note that there exist only monopoly industries in the region $a \in (\ln \chi, \bar{a}]$. This is because any new competitive industry enters into the unpatentable region $a \in (0, \ln \chi]$, and each competitive industry cannot enter into the region $a \in (\ln \chi, \bar{a}]$ because of the deterministic obsolescence (reduction) of relative productivity a. Then, we obtain

$$f_M(a) = f(a) \text{ for all } a \in (\ln \chi, \bar{a}]. \tag{22}$$

Next, consider the derivation of the stationary form of KFE for $a \in (0, \ln \chi]$. The outflow of the monopoly industries from $a \in (0, \ln \chi]$ is $f_M(a)\Delta h$ during the interval Δt. The inflow of monopoly industries into $a \in (0, \ln \chi]$ is $f_M(a + \Delta h)\Delta h$. The inflow and outflow must be equal in the stationary distribution:

$$f_M(a)\Delta h = f_M(a + \Delta h)\Delta h \text{ for all } a \in (0, \ln \chi]. \tag{23}$$

Conducting the Taylor expansion for $f_M(a+\Delta h)$ around a yields $f_M(a+\Delta h) = f_M(a) + f'_M(a)\Delta h$. Substituting this equation into Eq. (23) and simplifying it yields $0 = f'_M(a)g\Delta t$. Then, we obtain the following stationary form of KFE:

$$0 = f'_M(a) \text{ for all } a \in (0, \ln \chi]. \tag{24}$$

Eq.(24) implies the uniform distribution in the region $a \in (0, \ln \chi]$. Therefore, according to Eqs.(17) and (22) and the condition $\lim_{a \uparrow \ln \chi} f_M(a) = \lim_{a \downarrow \ln \chi} f_M(a)$ for the continuity, we can obtain the distribution $f_M(a)$ for all a. Then, we have $f_C(a)$ from Eq.(21). The results are summarized by the following lemma.

Lemma 2. *The following equations describe the stationary productivity distri-*

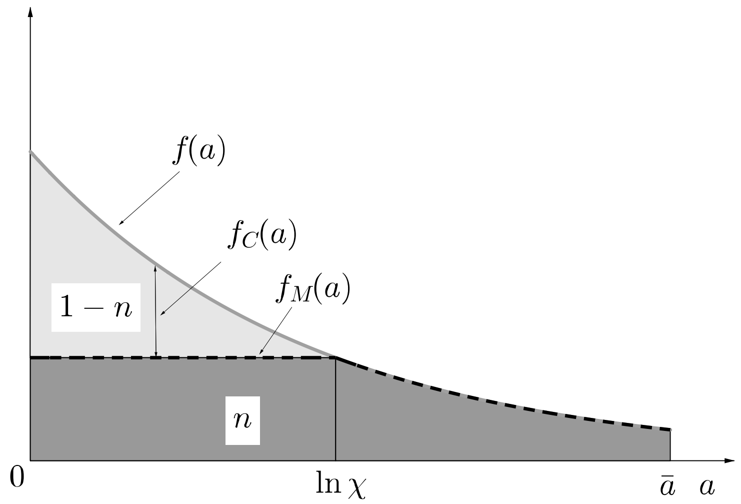

Figure 5: The productivity distributions for monopoly industries and competitive industries.

butions for monopoly industries and competitive industries:

$$f_M(a) = \begin{cases} \left(\frac{\eta}{1-p}\right) e^{-\eta \ln \chi} & \text{for } a \in (0, \ln \chi] \\ \\ \left(\frac{\eta}{1-p}\right) e^{-\eta a} & \text{for } a \in (\ln \chi, \bar{a}] \end{cases} \tag{25}$$

$$f_C(a) = \begin{cases} \left(\frac{\eta}{1-p}\right) \left(e^{-\eta a} - e^{-\eta \ln \chi}\right) & \text{for } a \in (0, \ln \chi] \\ \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

Fig.5 illustrates the distributions $f(a)$, $f_M(a)$, and $f_C(a)$. The solid gray curve represents $f(a)$. The dotted curve represents $f_M(a)$. Then, the dark gray area is equal to the number of monopoly industries, that is, $n = \int_0^{\bar{a}} f_M(a)da$. The distance between $f(a)$ and $f_M(a)$ represents $f_C(a)$, since Eq.(21) holds. Therefore, the bright gray area is equal to the number of competitive industries, that is, $1 - n = \int_0^{\ln \chi} f_C(a)da$.

The reason behind the result such that $f_M(a)$ is the uniform distribution in the region $a \in (0, \ln \chi]$ is as follows. There is no inflow of the new monopolists in the region $a \in (0, \ln \chi]$, because R&D firms cannot satisfy the requirement of minimum innovation size for a patent if they draw the productivity $a \in (0, \ln \chi]$. Then, the number of monopolists for each $a \in (0, \ln \chi]$ must be equal to that of $\lim_{a \downarrow \ln \chi} f_M(a)$. This implies uniform distribution.

5 Economic growth rate

In this section, we endogenize the growth rate g of public knowledge (frontier technology) $\bar{A}(t)$, which coincides with the growth rate of final output $Y(t)$ in the balanced-growth equilibrium, as shown in Appendix C. Note that ϕ represents the aggregate innovation rate, since the number $\phi\Delta t$ of the industries are targeted by R&D during the interval Δt. Then, we suppose that the growth rate g is proportional to the aggregate innovation rate ϕ. In particular, we suppose that

$$g = \sigma_M[\ p + (1-p)\mu\ ]\phi + \sigma_C(1-p)(1-\mu)\phi \tag{27}$$

where $\sigma_M > 0$ and $\sigma_C > 0$ are the parameters, and

$$\mu \equiv \int_{\ln \chi}^{\bar{a}} f(a)da = \left(\frac{1}{1-p}\right)\left(e^{-\eta \ln \chi} - p\right) > 0. \tag{28}$$

μ represents the probability for the patentable innovation when R&D firms draw the probability $1-p$. Therefore, the right-hand side of Eq.(27) represents the weighted sum of the aggregate patentable innovation $[\ p + (1-p)\mu\ ]\phi$ and the aggregate unpatentable innovation $(1-p)(1-\mu)\phi$.

In actual economies (such as the United States, Europe, and Japan), patent applicants must disclose information about their inventions, with this information made public 18 months after the application is filed and in sufficient detail to enable a person with ordinary skill in the art to replicate the invention. Hence, as implied by Eq.(27), the more frequently the aggregate patentable innovations $[\ p + (1-p)\mu\ ]\phi$ are created (and filed for patent), the greater is the growth g of public knowledge $\bar{A}(t)$. Furthermore, we suppose that the unpatentable innovations have a positive effect on the growth g. In this model, the information about unpatentable innovations is disclosed immediately in sufficient detail to ensure that any imitators can replicate the unpatentable inventions at no cost. Even though we consider the real economy, some of unpatented technologies would be revealed by the imitation activities. Hence, as implied by Eq.(27), the more frequently the aggregate unpatentable innovations $(1-p)(1-\mu)\phi$ are created, the greater is the growth g of public knowledge. The parameters σ_M and σ_C represent the strength of the spillover effects of aggregate patentable innovations and aggregate unpatentable innovations on g, respectively.

6 Balanced-growth equilibrium

In this section, we determine the equilibrium values of η and g and show the existence of the balanced-growth equilibrium. Substituting Eqs. (27) and (28) into Eq.(18), we yield

$$\eta = \frac{1-p}{(\sigma_M - \sigma_C)e^{-\eta \ln \chi} + \sigma_C}. \tag{29}$$

Eq.(29) determines the equilibrium value of η, since Eq.(29) contains only the endogenous variable η.

Lemma 3. *Suppose that the following inequality holds:*

$$\left(\frac{\sigma_M p + \sigma_C(1-p)}{1-p}\right)\ln\left(\frac{1}{p}\right) > \ln\chi \tag{30}$$

$$\frac{\sigma_M}{\sigma_C} \leq 2. \tag{31}$$

Then, there exists a unique equilibrium value of η*.*

Proof. See Appendix A. □

Note that Eq.(30) ensures $\bar{a} > \ln\chi$ in the balanced-growth equilibrium. In particular, as $p \to 0$, the left-hand side of Eq.(30) diverges to ∞. Then, we do not have to suppose a sufficiently small χ when $p \to 0$. The remainder of the paper assumes that Eqs.(30) and (31) hold.

Lemma 4. *An increase in* $\ln\chi$ *strictly raises the equilibrium* η *if and only if* $\sigma_M > \sigma_C$*. Otherwise, an increase in* $\ln\chi$ *reduces* η*.*

Proof. See Appendix A. □

The lemma 4 is a key to analyze the relationship between minimum innovation size and average innovation size. Therefore, we now consider the economic intuitions behind the lemma 4. According to Eq.(28), an increase in minimum innovation size $\ln\chi$ has a power to reduce the probability μ of patentable innovation for a given η. When $\sigma_M > (<)\sigma_C$ holds, the spillover effect of aggregate patentable innovations $[\ p + (1-p)\mu\ ]\phi$ on growth g is larger (smaller) than that of aggregate unpatentable innovations $(1-p)(1-\mu)\phi$. Then, higher $\ln\chi$ has a negative (positive) effect on g because of the power to reduce μ. This causes a slower (faster) speed g of obsolescence of relative productivity a, then it accumulates a larger (smaller) amount of the industries around the minimum productivity $a = 0$. This implies a higher (lower) equilibrium value of η (see Fig.4).

Later, we show that there exists a negative relationship between minimum innovation size and average patentable innovation size for some small $\ln\chi$, only if $\sigma_M > \sigma_C$ holds. This is because according to Fig.4, higher $\ln\chi$ has a negative effect on the average innovation size when $\sigma_M > \sigma_C$, since it reduces the probability for larger innovation size.

Consider the effects of $\ln\chi$ on μ and n. According to Eq. (28), $d\mu/d\ln\chi \gtreqless 0 \iff d(\eta\ln\chi)/d\ln\chi \lesseqgtr 0$. The number n of the monopoly industries is

$$n \equiv \int_0^{\bar{a}} f_M(a)da = \left(\frac{\eta}{1-p}\right)e^{-\eta\ln\chi}\ln\chi + \left(\frac{1}{1-p}\right)(e^{-\eta\ln\chi} - p). \tag{32}$$

Differentiating Eq.(32) yields $dn/d\ln\chi \gtreqless 0 \iff d(\eta\ln\chi)/d\ln\chi \lesseqgtr 0$. Of course, for a given η, an increase in minimum innovation size $\ln\chi$ has a negative effect on both the probability μ for patentable innovation size and the number

n of the monopoly industries. However, from lemma 4, if $\sigma_M < \sigma_C$ holds, an increase in $\ln \chi$ reduces η, which has a positive effect on both μ and n. This is because the lower η increases the probability for large innovation size (see Fig.4). We can show that this positive effect via endogenous response of the distribution is sufficiently small. That is, the negative effect of $\ln \chi$ on μ and n outweighs the positive effect via η, then the overall effect is always negative, that is, $d(\eta \ln \chi)/d \ln \chi < 0$ for $\sigma_M < \sigma_C$ (see Appendix B). Therefore, we yield $d\mu/d \ln \chi < 0$ and $dn/d \ln \chi < 0$.

Proposition 1. *As $p \to 0$, $f(a)$ converges to the exponential distribution*

$$f(a) = \eta e^{-\eta a} \text{ for } a \in (0, \infty). \tag{33}$$

Proof. As $p \to 0$, according to Eq.(29), we yield the equation for the determination of η:

$$\eta = \frac{1}{(\sigma_M - \sigma_C)e^{-\eta \ln \chi} + \sigma_C}. \tag{34}$$

In Appendix A, we show that the equilibrium η from Eq.(34) has a finite value. Then, from Eq.(19), $\bar{a} \to \infty$ as $p \to 0$. Therefore, we yield the result (33) from Eq.(17). □

The proposition 1 implies that the distribution of relative productivity $\hat{a} \equiv e^a = A/B$ converges to a Pareto distribution as $p \to 0$, according to Eq.(20):

$$f_{\hat{a}}(\hat{a}) = \eta \hat{a}^{-(\eta+1)} \text{ for } \hat{a} \in (1, \infty), \tag{35}$$

where η is given by Eq.(34). More specifically, Eq.(35) is the Pareto distribution with shape parameter η and scale parameter 1.

Note that the counter-cumulative distribution function $1 - F_{\hat{a}}(\hat{a}) \equiv \Pr\{\hat{a}' \geq \hat{a}\}$ of Eq.(20) is $1 - F_{\hat{a}}(\hat{a}) = 1 - \int_0^{\hat{a}} f_{\hat{a}}(\hat{a}')d\hat{a}' = -p/(1-p) + \hat{a}^{-\eta}/(1-p)$. This equation does not provide the linear relationship between $\ln[1 - F_{\hat{a}}(\hat{a})]$ and $\ln \hat{a}$, that is, $\ln[1 - F_{\hat{a}}(\hat{a})] = \ln[-p/(1-p) + \hat{a}^{-\eta}/(1-p)]$. This is inconsistent with the empirical facts of firm size distribution (see Fujiwara et al. 2004 and Newman 2005) and innovation size distribution (see Newman 2005 for the citation of scientific papers and Silverberg and Verspagen 2007 for the citation of patents and the financial returns from the patents). To reconcile the empirical facts and the theory, the maximum support of the distribution must take a sufficiently large value. That is, we require $p \to 0$, and then, we yield the linear relationship: $\ln[1 - F_{\hat{a}}(\hat{a})] = -\eta \ln \hat{a}$, which is consistent with the empirical facts. Then, the long-run distributions of firm size (which is proportional to absolute productivity A), the absolute productivity A, and innovation size $\hat{a}$ have a Pareto tail when $p \to 0$. These results crucially differ from Lucas and Moll (2014) and Perla and Tonetti (2014), in which the long-run distribution of productivity A has a Pareto tail if the initial distribution of productivity A has a Pareto tail. By contrast, in our model, the long-run distribution always has a Pareto tail, irrespective of the initial distribution.

Note that the result is discontinuous at $p = 0$. In the case of $p = 0$, the result is consistent with that of Lucas and Moll (2014), Perla and Tonetti (2014), and Perla et al. (2015). In particular, when $p = 0$, the dynamics of the distribution are equivalent to that of Perla et al. (2015). Thus, as shown in Perla et al. (2015), if the initial distribution of productivity A is a Pareto, then the long-run distribution of productivity A is also a Pareto with a shape parameter that is consistent with that of initial Pareto distribution. In contrast, in the case of $p \to 0$, the shape parameter η is an endogenous variable, which is given by Eq.(34). Furthermore, in the case of $p \to 0$, we may suppose the bounded support of the initial distribution due to the existence of initial finite frontier technological level $\bar{A}(0)$, which rejects the assumption of initial Pareto distribution in our model.

Next, consider the existence of the balanced-growth equilibrium, where the growth rate g of frontier technology $\bar{A}(t)$ corresponds to that of final good $Y(t)$, that is, g is economic growth rate, and thus, g must be equal to the growth rate of consumption $C(t)$ from the final good market-clearing condition (see Appendix C). Therefore, from the Euler equation (2), we yield

$$r = \theta g + \rho. \tag{36}$$

We consider the requirement of TVC (3). According to Eqs.(9) and (25), the aggregate asset (patent) value is $V(t)L = B(t)\int_0^{\bar{a}} f_M(a)V_M(a)da$. Then, the growth rate of asset value per capita $V(t)$ is equal to g. To satisfy TVC (3), the growth rate of the term $e^{-\rho t}C(t)^{-\theta}V(t)$ must be negative. We can satisfy this requirement for all $g \geq 0$, since $\theta \geq 1$:

$$-\rho - \theta\frac{\dot{C}(t)}{C(t)} + \frac{\dot{V}(t)}{V(t)} = -\rho - (\theta - 1)g < 0.$$

Substituting the patent value (9) and the distribution (17) into the expected patent value (12) yields the following equation

$$\begin{aligned} V_R = &\left(\frac{\tilde{\pi}_M}{r}\right)\left(\frac{\eta}{\eta - 1}\right)\left[e^{-(\eta-1)\ln\chi} - pe^{\bar{a}}\right] \\ &- \left(\frac{\tilde{\pi}_M}{r}\right)\left(\frac{\eta}{\eta + k}\right)\left[e^{-(\eta+k)\ln\chi} - pe^{-k\bar{a}}\right] + \left(\frac{\tilde{\pi}_M}{r}\right)p\left(e^{\bar{a}} - e^{-k\bar{a}}\right) \end{aligned} \tag{37}$$

where $k \equiv (r - g)/g = [(\theta - 1)g + \rho]/g > 0$, since $\theta \geq 1$ ensures $k > 0$. Note that the first term on the right-hand side of Eq.(37) is always positive, since the term represents $(1-p)\int_{\ln\chi}^{\bar{a}}(\tilde{\pi}_M/r)e^a f(a)da > 0$. The second term is always negative, since it represents $-(1-p)\int_{\ln\chi}^{\bar{a}}(\tilde{\pi}_M/r)e^{-ka}f(a)da < 0$. The third term is always positive, since it represents $pV_M(\bar{a}) > 0$. We can determine the equilibrium value of economic growth rate g from the Euler equation (36), the free-entry condition (11), and the expected patent value (37).

Proposition 2. *If the fixed cost R is sufficiently small, there exists a unique balanced-growth equilibrium.*

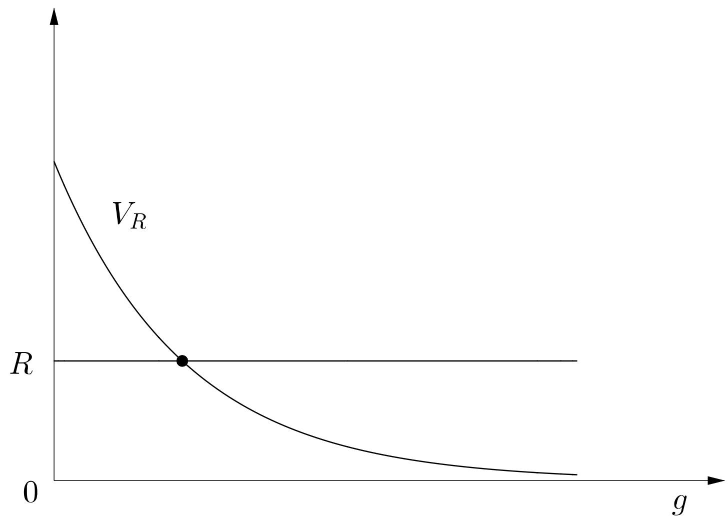

Figure 6: The existence of balanced-growth equilibrium.

Proof. First, we show that $\partial V_R/\partial g < 0$. Note that the equilibrium η is determined independently from g, as shown in Eq.(29). Then, according to Eqs.(17), (18), and (19), the productivity distribution $f(a)$ and the support $a \in (0, \bar{a}]$ are independent of g. Therefore, if the patent value $V_M(a)$ is a decreasing function with respect to g, the expected patent value (R&D benefit) $V_R \equiv (1-p)\int_{\ln\chi}^{\bar{a}} V_M(a)dF(a) + pV_M(\bar{a})$ is also a decreasing function with respect to g, that is, $\partial V_M(a)/\partial g < 0$ for any $a \Rightarrow \partial V_R/\partial g < 0$. To show that $\partial V_M(a)/\partial g < 0$ for any a, according to Eq.(9) and the definition of k, we restate the patent value for a given a:

$$V_M(a) = \left(\frac{\tilde{\pi}_M}{r}\right)\left(e^a - e^{-ka}\right). \tag{38}$$

Since $\partial k/\partial g < 0$, according to Eq.(38), higher g reduces $V_M(a)$ via the variable k for a given r. From Eq.(36), we have $\partial r/\partial g > 0$. Then, according to Eq.(38), higher g reduces $V_M(a)$ via the variable r for a given k. Summing these effects, we yield $\partial V_M(a)/\partial g < 0$ for any a, and thus, $\partial V_R/\partial g < 0$ holds.

Noting that $\lim_{g\to\infty} r = \infty$ and $\lim_{g\to\infty} k = \theta - 1$, we yield $\lim_{g\to\infty} V_R = 0$ from Eq.(37). In addition, noting that $\lim_{g\downarrow 0} r = \rho$ and $\lim_{g\downarrow 0} k = \infty$, we obtain

$$\lim_{g\downarrow 0} V_R = \left(\frac{\tilde{\pi}_M}{\rho}\right)\left(\frac{\eta}{\eta-1}\right)\left[e^{-(\eta-1)\ln\chi} - pe^{\bar{a}}\right] + p\left(\frac{\tilde{\pi}_M}{\rho}\right)e^{\bar{a}} > 0. \tag{39}$$

Therefore, if the following inequality holds,

$$\lim_{g\downarrow 0} V_R > R, \tag{40}$$

then, there exists a unique equilibrium value of g (see Fig.6). Note that since η and $\bar{a}$ are independent of R, if we take a sufficiently small R, it satisfies the condition $\lim_{g\downarrow 0} V_R > R$. If we obtain the unique equilibrium values of η and g, we can ensure a unique initial consumption level $C(0)$ from the final good market-clearing condition (see Appendix C). This implies a unique balanced-growth equilibrium. □

As shown in Fig.6, the expected patent value $V_R \equiv (1-p)\int_{\ln\chi}^{\bar{a}} V_M(a)dF(a) + pV_M(\bar{a})$ is a decreasing function with respect to g. We have two channels of negative effects of g on V_R. First, higher g reduces the expected effective patent length, which has a negative effect on V_R. Recall that the effective patent length $T(a) \equiv a/g$ represents the waiting time until the creative destruction occurs for the monopolist equipped with the relative productivity a. Higher g implies faster obsolescence (reduction) of relative productivity a over time, which reduces the expected effective patent length. Therefore, higher g has a negative effect on the expected patent value V_R. Second, we have the general equilibrium effect via the interest rate r. Higher g increases the discount rate r to evaluate the future profits stream. Thus, it reduces the expected patent value V_R.

The proposition 2 crucially differs from Lucas and Moll (2014) and Perla and Tonetti (2014), in which there exists a continuum of the balanced-growth equilibria. Each equilibrium can be attained for each shape parameter of initial Pareto distribution of productivity A. In contrast, our model has a unique balanced-growth equilibrium, irrespective of initial productivity distribution. Even though $p \to 0$, we can ensure a unique balanced-growth equilibrium, if R is sufficiently small and $(\sigma_M - \sigma_C) + \chi(\sigma_C - 1) < 0$ holds (see Appendix D). The requirement $(\sigma_M - \sigma_C) + \chi(\sigma_C - 1) < 0$ ensures $\eta > 1$ in a balanced-growth equilibrium. Under $\eta > 1$, we yield the finite mean of the Pareto distribution (35), and thus, the finite expected patent value V_R when $p \to 0$. Then, we can ensure the existence of a unique equilibrium g under $p \to 0$, if R is sufficiently small and $\eta > 1$ holds. However, the result is discontinuous at $p = 0$. In the case of $p = 0$, the result is consistent with that of Lucas and Moll (2014) and Perla and Tonetti (2014), that is, the initial productivity A distribution must be a Pareto to ensure the balanced-growth equilibrium with sustained growth (R&D investments). This requirement implies that many industries must already have very large productivities at initial date, that is, there is no upper bound of the support of the initial distribution of the productivity A and the initial distribution of the productivity A must have a fat tail. Therefore, the R&D firms cannot exhaust the resources (possibilities) of the productivity improvement in the long run, which ensures the sustained growth.

Let us consider why the initial distribution of productivity A does not need to have a Pareto tail in our model. The frontier technology $\bar{A}(t)$ continues to grow at rate g. Therefore, even though the maximum support of the initial distribution $\bar{A}(0)$ has a finite value, each R&D firm cannot exhaust the resources (possibilities) of the productivity improvement. This ensures the sustained growth. By contrast, in the models of Lucas and Moll (2014) and Perla and Tonetti (2014), if the initial productivity distribution has a finite maximum support (frontier technology), any firms' productivities coincide with the frontier level of initial productivity distribution in the long-run as long as the firms continue to draw the new superior productivity from the productivity distribution. That is, degenerate distribution emerges in the long-run. This is because the frontier technology is constant over time in the models of Lucas and Moll (2014) and Perla and Tonetti (2014). Then, any firms do not have an incentive to invest R&D in the long run, since they have already stood at the frontier level. In other words, the firms exhaust the possibilities of productivity improvement.

7 Patentable innovation size and growth

In this section, we show the effects of minimum innovation size on the average patentable innovation size and growth.

7.1 Minimum and average innovation sizes

We show the relationship between the minimum innovation size and the average patentable innovation size. To show this, we derive $f(a|a > \ln\chi)$, which is the probability density function conditional on the patentable innovation:

$$f(a|a > \ln\chi) = \frac{f(a)}{\Pr\{a > \ln\chi\}} = \left(\frac{\eta}{e^{-\eta\ln\chi} - p}\right) e^{-\eta a} \text{ for } a \in (\ln\chi, \bar{a}]. \tag{41}$$

Using Eq.(41), we yield the average patentable innovation size for the R&D firms-drawn probability $1 - p$:

$$e^{a_\chi} \equiv \int_{\ln\chi}^{\bar{a}} e^a f(a|a > \ln\chi) = \left(\frac{\eta}{\eta - 1}\right) \left[\frac{e^{-(\eta-1)\ln\chi} - pe^{\bar{a}}}{e^{-\eta\ln\chi} - p}\right]. \tag{42}$$

Then, according to Eq.(42), the average patentable innovation size is

$$e^{a_\chi^*} \equiv pe^{\bar{a}} + (1 - p)e^{a_\chi}. \tag{43}$$

Proposition 3. *Under the assumptions of* $p \to 0$ *and* $\sigma_M < 1$, *an increase in minimum innovation size* $\ln\chi$ *strictly reduces the average patentable innovation size* $e^{a_\chi^*}$ *around* $\ln\chi = 0$ *if and only if* $\sigma_C < (\sigma_M)^2$. *Otherwise, an increase in* $\ln\chi$ *raises* $e^{a_\chi^*}$ *around* $\ln\chi = 0$.

Proof. See Appendix E. □

Note that the distribution $f_{\hat{a}}(\hat{a}|\hat{a} > \chi)$ of the relative productivity $\hat{a} \equiv e^a = A/B$ converges to the Pareto distribution with scale parameter χ and shape parameter η when $p \to 0$. Further, the assumption $\sigma_M < 1$ ensures that $\eta = 1/\sigma_M > 1$ when $p \to 0$ and $\chi = 1$ in order to yield a finite mean value $e^{a_\chi^*}$ of the Pareto distribution $f_{\hat{a}}(\hat{a}|\hat{a} > \chi)$. Proposition 3 investigates the relationship between minimum innovation size and average patentable innovation

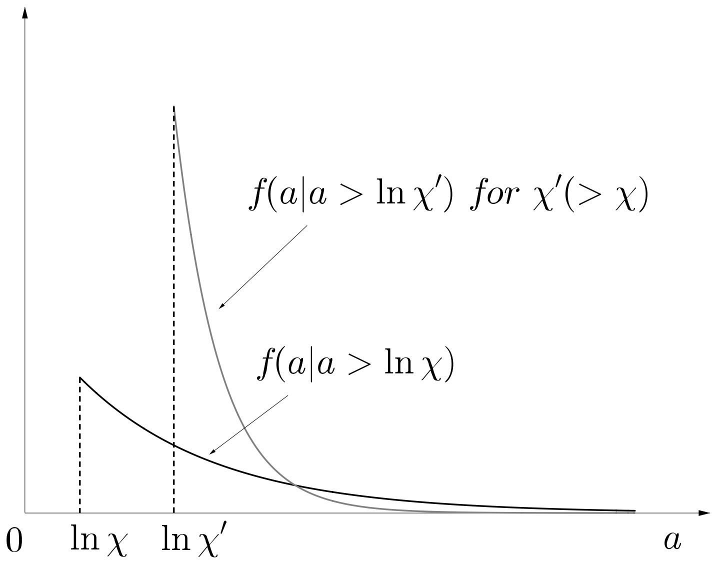

Figure 7: The comparative statics of $f(a|a > \ln\chi)$ when $\sigma_M > \sigma_C$ holds.

size in these cases. An increase in minimum innovation size $\ln\chi$ required for a patent, of course, has a positive effect on the average patentable innovation size $e^{a^*_\chi}$ via scale parameter χ. This is because a higher scale parameter χ shifts the distribution to the right (see Fig.7). However, if higher χ raises the shape parameter η, we have an additional negative effect on the average value. This is because higher η reduces the average value of Pareto distribution. Therefore, $\sigma_M > \sigma_C$ is the necessary condition for the negative relationship between minimum and average innovation size (see lemma 4). More precisely, when $\sigma_M > \sigma_C$ holds, an increase in $\ln\chi$ shifts the weigth in the distribution from the tail area to the area around the minimum innovation size because of the higher shape parameter η (see Fig.7). An increase in $\ln\chi$ causes the reduction of probability μ for patentable innovation, which has a negative effect on the speed g of obsolescence of the relative productivity a when $\sigma_M > \sigma_C$. Then, it causes a larger accumulation of the industries in each segment a, which leads to a steeper slope of the distribution $f(a)$. Therefore, R&D firms tend to draw lower productivity, which reduces the average patentable innovation size $e^{a^*_\chi}$. Proposition 3 shows that the total effect of higher $\ln\chi$ on $e^{a^*_\chi}$ is negative around $\ln\chi = 0$ if and only if $\sigma_C < (\sigma_M)^2$.

Figs.8 and 9 report the relationship between minimum innovation size and average innovation sizes of the simulations. We employ the following calibration strategy. We consider the case of $\chi = 1$ as the basis model, that is, the model without minimum innovation size, which is the standard Schumpeterian growth model. The pre-selected parameter and normalization are the subjective

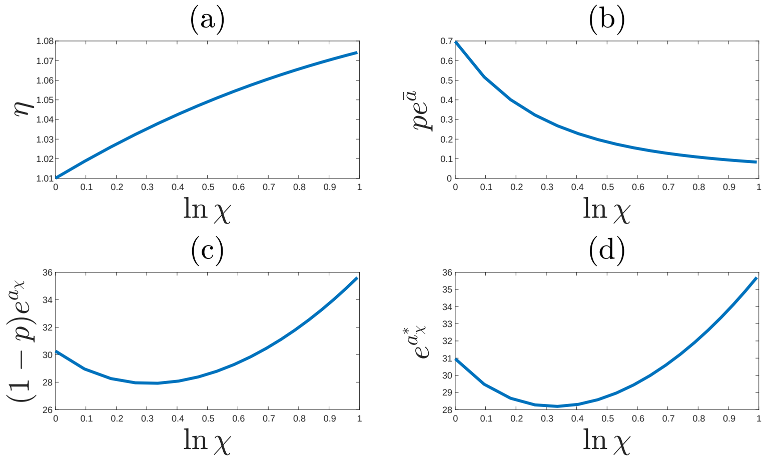

Figure 8: The minimum and average innovation sizes when $\sigma_M = 0.99 > \sigma_C = 0.9$.

discount rate $\rho = 0.02$ and the population size $L = 1$, respectively.

Recall that the very small probability p leads to the Pareto distributions of the firm size, productivity, and innovation size, which is consistent with the empirical facts. Then, we set $p = 2.2204e - 16$, which is the machine epsilon in MATLAB, that is, the smallest value $\epsilon > 0$ satisfying $1 + \epsilon > 1$. Using 1429 estimates of published studies, Havranek et al. (2015) report that the mean value of the elasticity of intertemporal substitution in the United States is about 0.6. This suggests that $1/\theta = 0.6$. In the United States, the labor share is about 2/3. This implies that $\alpha = 1/3$. The variable η is interpreted as a shape parameter of Pareto distribution when $p \to 0$. Luttmer (2007) and Gabaix (2009) estimate that η is slightly above unity (Zipf's law), which is based on a shape parameter of the size distribution of firms. To achieve this under $\chi = 1$, we set the spillover coefficient $\sigma_M = 0.99$ for aggregate patentable innovations. The growth rate of the gross domestic product (GDP) per capita in the United States has fluctuated around 2 percent at least since World War II. This fact suggests the value $g = 0.02$. To achieve this value under $\chi = 1$, we set the fixed R&D cost R, that is, $R = 42.5$.

The only remaining parameter is the spillover coefficient σ_C of unpatentable aggregate innovations. We conduct several simulations by setting different values of σ_C. For example, we set $\sigma_C = 0.9$ in Fig.8 as a case of $\sigma_M > \sigma_C$. In the case of Fig.9, we set $\sigma_C = 1.1$ as a case of $\sigma_M < \sigma_C$. Note that these choices of σ_C satisfy Eq.(31). We restrict the value χ to satisfy the conditions (30) and (40). Then, we can ensure the unique balanced-growth equilibrium. Further, note that only the parameters σ_M, σ_C, p, and χ affect the average innovation sizes. The calibration of the other parameters can be used in the

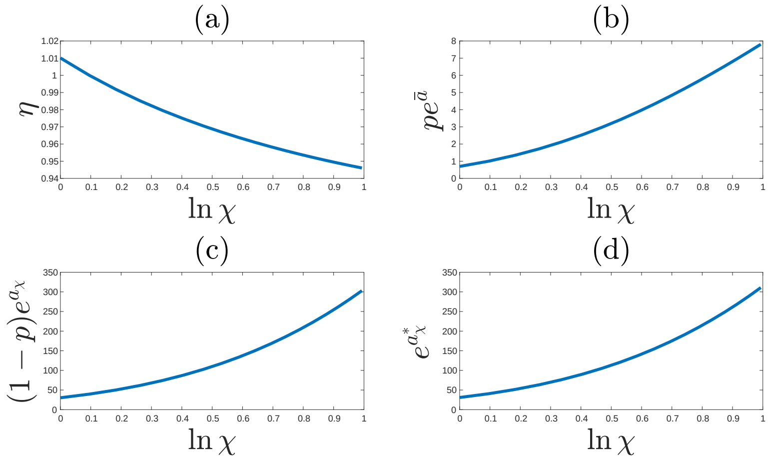

Figure 9: The minimum and average innovation sizes when $\sigma_M = 0.99 < \sigma_C = 1.1$.

analysis described in the other section.

Fig.8 reports the case of $\sigma_M > \sigma_C$. As shown in Fig.8(a), an increase in minimum innovation size $\ln\chi$ raises the shape parameter η, which is consistent with lemma 4. Therefore, as shown in Fig.8(b), according to Eq.(19), it reduces the maximum innovation size $\bar{a}$, which has a negative effect on the average patentable innovation size $e^{a^*_\chi}$. Fig.8(c) reports a U-shaped relationship between $\ln\chi$ and $(1-p)e^{a_\chi}$. There exist two competing forces that affect $(1-p)e^{a_\chi}$, that is, scale parameter χ and shape parameter η of the Pareto distribution. As shown in Fig.7, an increase in $\ln\chi$ shifts the distribution to the right, which has a positive effect on $(1-p)e^{a_\chi}$. However, we have the additional negative effect on $(1-p)e^{a_\chi}$ when $\sigma_M > \sigma_C$. An increase in $\ln\chi$ raises η, which causes smaller probability for a large innovation size (see Fig.7). Then, it reduces $(1-p)e^{a_\chi}$. The overall relationship between $\ln\chi$ and $(1-p)e^{a_\chi}$ can be nonmonotonic as shown in Fig.8(c), which causes the nonmonotonic relationship between $\ln\chi$ and average patentable innovation size $e^{a^*_\chi}$ as shown in Fig.8(d).

Fig.9 reports the case of $\sigma_M < \sigma_C$. As shown in Fig.9(a)-(b), according to lemma 4 and Eq.(19), an increase in $\ln\chi$ reduces η when $\sigma_M < \sigma_C$ holds, which raises maximum innovation size $\bar{a}$. Then, it has a positive effect on the average patentable innovation size $e^{a^*_\chi}$. Note that lower η does not produce a negative effect on $(1-p)e^{a_\chi}$. This is because it increases the probability for large patentable innovation size. Then, an increase in $\ln\chi$ always raises $(1-p)e^{a_\chi}$, which causes the positive relationship between $\ln\chi$ and average patentable innovation size $e^{a^*_\chi}$, as shown in Fig.9(c)-(d).

7.2 Minimum innovation size and growth

We demonstrate the relationship between minimum innovation size $\ln \chi$ and economic growth rate g. Note that according to Fig.6, $\partial V_R / \partial \ln \chi \gtreqless 0 \iff dg/d \ln \chi \gtreqless 0$. That is, an increase in minimum innovation size $\ln \chi$ leads to faster (slower) growth g if and only if it causes higher (lower) expected patent value V_R for a given g. Figs.10, 11, and 12 report the cases of $\sigma_C = 0.9, 1.1,$ and 1, respectively.

Let us consider the case of $\sigma_M > \sigma_C$. Higher $\ln \chi$ has a negative effect on $V_R \equiv (1-p) \int_{\ln \chi}^{\bar{a}} V_M(a) dF(a) + p V_M(\bar{a})$ for a given g because of the reduction of probability for a patentable innovation size. Furthermore, according to lemma 4, higher $\ln \chi$ raises η when $\sigma_M > \sigma_C$. Then, it reduces the average innovation size, as shown in Fig.4, which causes the reduction of V_R for a given g. In summary, we do not have a positive effect of $\ln \chi$ on V_R in the case of $\sigma_M > \sigma_C$. Therefore, we have a negative relationship between $\ln \chi$ and g, as demonstrated in Fig.10.

Next, consider the case of $\sigma_M \leq \sigma_C$. As in the case of $\sigma_M > \sigma_C$, higher $\ln \chi$ reduces the probability for a patentable innovation size, which has a negative effect on V_R for a given g. However, we have a positive effect on V_R via η. According to lemma 4, higher $\ln \chi$ reduces η when $\sigma_M \leq \sigma_C$. It increases the average innovation size (see Fig.4), which leads to higher V_R for a given g. In summary, we have two competing effects of $\ln \chi$ on V_R. Then, an increase in $\ln \chi$ may have an ambiguous effect on g. Fig.11 reports the positive relationship between $\ln \chi$ and g. However, as shown in Fig.12, where $\sigma_M = 0.99$ and $\sigma_C = 1$, if σ_C is slightly larger than σ_M, we yield the nonmonotonic relationship between $\ln \chi$ and g.

8 Conclusion

We developed a Schumpeterian growth model based on technology diffusion. Some of the firms may succeed to adopt the frontier technology into the production processes. The other firms may adopt the non-frontier level of technology into the production processes. The diffusion of technology improves the overall productivity level in the economy. Then, the economy achieves sustained growth with the aid of the evolution of the frontier technology. These technology diffusion processes generate the Pareto distributions of firm size, productivity, and innovation size if it is extremely difficult to adopt the frontier technology.

Further, we showed that an increase in minimum innovation size required for a patent may reduce the average patentable innovation size due to the endogenous response of the innovation size distribution. In particular, strengthening the minimum innovation size reduces the probability for the patentable innovations, which reduces the knowledge spillover of the existing technologies. This is because the information about the patented technologies is disclosed in the economy under the patent law. The reduction of the amount of knowledge spillover causes a larger amount of inferior existing technologies. Then, each firm tends

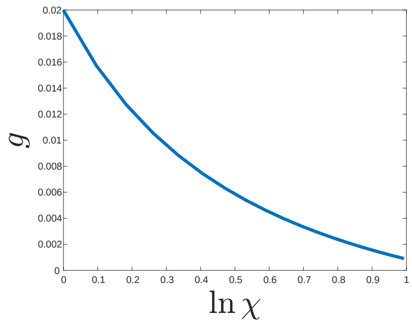

Figure 10: The minimum innovation size and growth when $\sigma_M = 0.99 > \sigma_C = 0.9$.

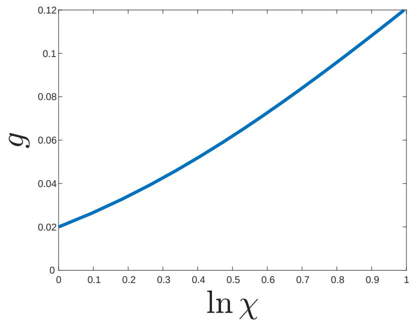

Figure 11: The minimum innovation size and growth when $\sigma_M = 0.99 < \sigma_C = 1.1$.

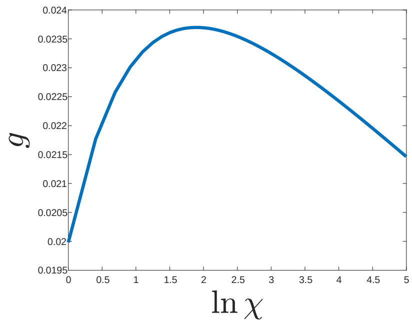

Figure 12: The minimum innovation size and growth when $\sigma_C = 1$ is slightly larger than $\sigma_M = 0.99$.

to adopt less-effective technologies by learning the existing inferior knowledge, which may reduce the average patentable innovation size. Therefore, the United States may experience a lower average patent quality after the decision of the United States Supreme Court for KSR International Co. v. Teleflex Inc. (2007), which substantially increased the minimum innovation size. This is counter to the predictions of the FTC (2003) and the NAS (2004).

A Equilibrium value of η

First, we show the existence of a unique η. For convenience, we restate Eq.(29):

$$\eta = \frac{1-p}{(\sigma_M - \sigma_C)e^{-\eta \ln \chi} + \sigma_C}. \tag{44}$$

Rewriting Eq.(44), we have

$$y(\eta) \equiv (\sigma_M - \sigma_C)\eta e^{-\eta \ln \chi} + \sigma_C \eta - (1-p) = 0 \tag{45}$$

Differentiating the left-hand side of Eq.(45) obtains

$$\frac{\partial y(\eta)}{\partial \eta} = (\sigma_M - \sigma_C)e^{-\eta \ln \chi}(1 - \eta \ln \chi) + \sigma_C \tag{46}$$

From Eq.(46), we obtain

$$\frac{\partial y(\eta)}{\partial \eta} \gtreqless 0 \iff y_1(\eta) \gtreqless y_2(\eta), \tag{47}$$

where

$$y_1(\eta) \equiv \left(\frac{\sigma_M - \sigma_C}{\sigma_C}\right) + e^{\eta \ln \chi} \text{ and} \tag{48}$$

$$y_2(\eta) \equiv \left(\frac{\sigma_M - \sigma_C}{\sigma_C}\right) \eta \ln \chi. \tag{49}$$

Note that $y_1(0) > y_2(0)$ holds. Then, we have $\partial y(\eta)/\partial \eta|_{\eta=0} > 0$.

If $(\sigma_M - \sigma_C)/\sigma_C \leq 1 \iff \sigma_M/\sigma_C \in (0, 2]$ holds, according to Eqs.(48) and (49), we yield $\partial y_1(\eta)/\partial \eta = e^{\eta \ln \chi} \ln \chi \geq \partial y_2(\eta)/\partial \eta = [(\sigma_M - \sigma_C)/\sigma_C] \ln \chi$. According to Eq.(47), this implies that $\partial y(\eta)/\partial \eta \geq 0$, since $\partial^2 y_1(\eta)/\partial \eta^2 \geq 0$ and $\partial^2 y_2(\eta)/\partial \eta^2 = 0$. Note that the equilibrium η must satisfy

$$\bar{a} > \ln \chi \iff \eta < \frac{\ln(1/p)}{\ln \chi}. \tag{50}$$

To ensure that Eq. (50), the following equation must hold

$$\begin{aligned} y\left(\frac{\ln(1/p)}{\ln \chi}\right) &= (\sigma_M - \sigma_C)p\left(\frac{\ln(1/p)}{\ln \chi}\right) + \sigma_C\left(\frac{\ln(1/p)}{\ln \chi}\right) - (1-p) > 0 \\ &\iff \left(\frac{\sigma_M p + \sigma_C(1-p)}{1-p}\right) \ln\left(\frac{1}{p}\right) > \ln \chi. \end{aligned} \tag{51}$$

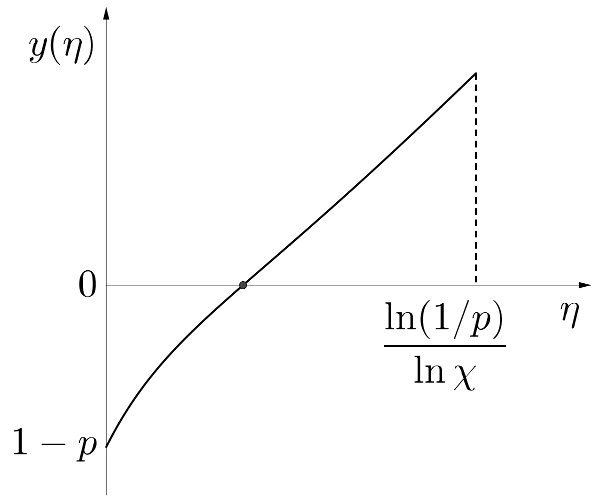

Figure 13: The determination of η.

Then, we have a unique η according to Fig.13. Note that $\partial y(\eta)/\partial \ln\chi = -\eta^2(\sigma_M - \sigma_C) \gtreqless 0 \iff \sigma_M \lesseqgtr \sigma_C$. That is, an increase in $\ln\chi$ shifts down (up) $y(\eta)$ if and only if $\sigma_M > (<)\sigma_C$. Then, we have $d\eta/d\ln\chi \gtreqless 0 \iff \sigma_M \gtreqless \sigma_C$.

Note that the above-described results hold even though $p \to 0$. In particular, when $p \to 0$, we do not have to assume a sufficiently small χ to ensure $\bar{a} > \ln\chi$. This is because the left-hand side of Eq.(51) diverges to ∞ as $p \to 0$. However, $\eta > 1$ must hold when $p \to 0$ to ensure the existence of a balanced-growth equilibrium (see Appendix D). Then, we derive the requirement to satisfy $\eta > 1$. When $p \to 0$, according to Eq.(45), the following equation determines η:

$$y(\eta) \equiv (\sigma_M - \sigma_C)\eta e^{-\eta\ln\chi} + \sigma_C\eta - 1 = 0.$$

To yield $\eta > 1$ in equilibrium, $y(1) < 0$ must hold, since $\partial y(\eta)/\partial\eta > 0$ under the assumption $\sigma_M/\sigma_C \leq 2$. Then, the following equation must hold:

$$y(1) < 0 \iff (\sigma_M - \sigma_C) + \chi(\sigma_C - 1) < 0. \tag{52}$$

For example, when $\chi = 1$, Eq.(52) requires $\sigma_M < 1$. This is because, according to Eq.(44), $\eta = 1/\sigma_M$ when $\chi = 1$ and $p \to 0$.

B The effects of $\ln\chi$ on μ and n

Now, we consider the effects of $\ln\chi$ on μ and n. For convenience, we restate Eqs.(28) and (32):

$$\mu \equiv \int_{\ln\chi}^{\bar{a}} f(a)da = \left(\frac{1}{1-p}\right)\left(e^{-\eta\ln\chi} - p\right) \tag{53}$$

$$n \equiv \int_{0}^{\bar{a}} f_M(a)da = \left(\frac{\eta}{1-p}\right) e^{-\eta\ln\chi}\ln\chi + \left(\frac{1}{1-p}\right)\left(e^{-\eta\ln\chi} - p\right). \tag{54}$$

According to Eq.(53), we have $d\mu/d\ln\chi \gtreqless 0 \iff d(\eta\ln\chi)/d\ln\chi \lesseqgtr 0$. Differentiating Eq.(54) obtains

$$\frac{dn}{d\ln\chi} = -\left(\frac{1}{1-p}\right)(\eta\ln\chi)e^{-\eta\ln\chi}\frac{d(\eta\ln\chi)}{d\ln\chi}. \tag{55}$$

Then, from Eq.(55), we yield $dn/d\ln\chi \gtreqless 0 \iff d(\eta\ln\chi)/d\ln\chi \lesseqgtr 0$.

Next, derive $d\eta/d\ln\chi$. We restate Eq.(29):

$$\eta = \frac{1-p}{(\sigma_M - \sigma_C)e^{-\eta\ln\chi} + \sigma_C}. \tag{56}$$

Differentiating Eq.(56), we yield

$$\frac{d\eta}{d\ln\chi} = \frac{(\sigma_M - \sigma_C)\eta^2 e^{-\eta\ln\chi}}{(\sigma_M - \sigma_C)e^{-\eta\ln\chi}(1-\eta\ln\chi) + \sigma_C}. \tag{57}$$

Note that lemma 4 shows that $d\eta/d\ln\chi \gtreqless 0 \iff \sigma_M \gtreqless \sigma_C$. This implies that the denominator of Eq.(57) is always positive. Then, we have

$$\frac{d(\eta\ln\chi)}{d\ln\chi} = \frac{\eta\left[\sigma_M e^{-\eta\ln\chi} + \sigma_C\left(1 - e^{-\eta\ln\chi}\right)\right]}{(\sigma_M - \sigma_C)e^{-\eta\ln\chi}(1-\eta\ln\chi) + \sigma_C} > 0, \tag{58}$$

since both numerator and denominator of Eq. (58) are always positive. Then, we obtain the result: $d\mu/d\ln\chi < 0$ and $dn/d\ln\chi < 0$.

C Final good market-clearing condition

This appendix shows the final good market-clearing condition. Substituting the amount of production for monopoly and competitive firm into the production function of final good (4), we obtain

$$Y(t) = (\tilde{x}_M)^\alpha L A_M^*(t) + (\tilde{x}_C)^\alpha L A_C^*(t) \tag{59}$$

where $A_M^*(t) \equiv B(t)\int_0^{\bar{a}} e^a f_M(a)da$ and $A_C^*(t) \equiv B(t)\int_0^{\ln\chi} e^a f_C(a)da$ represent the aggregate productivity level in the monopoly industries and that in the competitive industries, respectively. Note that we have the following relationship:

$A^*(t) = A_M^*(t) + A_C^*(t)$, where $A^*(t) \equiv B(t)\int_0^{\bar{a}} e^a f(a)da$ represents the aggregate productivity in the entire economy because Eq.(21) holds. Eq.(59) implies that $g = \dot{Y}(t)/Y(t)$. From Eqs.(25) and (26), we yield

$$\begin{aligned}\frac{A_M^*(t)}{B(t)} = &\left(\frac{\eta}{1-p}\right) e^{-\eta \ln\chi}\left(e^{\ln\chi} - 1\right) \\ &+ \left(\frac{1}{1-p}\right)\left(\frac{\eta}{\eta-1}\right)\left[e^{-(\eta-1)\ln\chi} - pe^{\bar{a}}\right]\end{aligned} \tag{60}$$

$$\begin{aligned}\frac{A_C^*(t)}{B(t)} = &-\left(\frac{\eta}{1-p}\right) e^{-\eta \ln\chi}\left(e^{\ln\chi} - 1\right) \\ &+ \left(\frac{1}{1-p}\right)\left(\frac{\eta}{\eta-1}\right)\left[1 - e^{-(\eta-1)\ln\chi}\right].\end{aligned} \tag{61}$$

The final good market-clearing condition is $Y(t) = C(t)L + \tilde{x}_M L A_M^*(t) + \tilde{x}_C L A_C^*(t) + \phi R B(t)$. Substituting Eq.(59) into this equation and simplifying it yields

$$\left(\frac{C(t)}{B(t)}\right)L = \left[(\tilde{x}_M)^\alpha - \tilde{x}_M\right]L\left(\frac{A_M^*(t)}{B(t)}\right) + \left[(\tilde{x}_C)^\alpha - \tilde{x}_C\right]L\left(\frac{A_C^*(t)}{B(t)}\right) - \phi R. \tag{62}$$

Note that the sum of the first and second terms on the right-hand side of Eq.(62) represents $B(t)$-adjusted GDP. We yield $g = \dot{C}(t)/C(t)$, because the right-hand side of Eq.(62) is constant over time. Noting that $\phi = (g\eta)/(1-p)$ from Eq.(18) and using Eqs. (60) and (61), the right-hand side of Eq.(62) can be rewritten as a function of η and g. Then, if we can ensure the unique equilibrium values of η and g, we obtain a unique value of $C(0)$ for a given initial state variable $B(0)$.

D The existence of a balanced-growth equilibrium when $p \to 0$

This appendix shows the existence of a balanced-growth equilibrium when $p \to 0$. Even though $p \to 0$, $\partial V_R/\partial g < 0$ holds because $\partial V_M(a)/\partial g < 0$ and η is independent of g. As shown in Appendix A, if $(\sigma_M - \sigma_C) + \chi(\sigma_C - 1) < 0$ holds, we yield $\eta > 1$ when $p \to 0$. Note that $\eta > 1$ ensures the finite mean of the Pareto distribution (35), and thus, we yield the finite expected patent value V_R. To show this, noting that $pe^{\bar{a}} = p^{(\eta-1)/\eta} \to 0$ as $p \to 0$, under $\eta > 1$, according to Eq.(37), the expected patent value under $p \to 0$ is

$$V_R = \left(\frac{\tilde{\pi}_M}{r}\right)\left(\frac{\eta}{\eta-1}\right)e^{-(\eta-1)\ln\chi} - \left(\frac{\tilde{\pi}_M}{r}\right)\left(\frac{\eta}{\eta+k}\right)e^{-(\eta+k)\ln\chi}. \tag{63}$$

Recall that $r = \theta g + \rho$ and $k \equiv (r-g)/g = [(\theta-1)g + \rho]/g > 0$. Since $\lim_{g\to\infty} r = \infty$ and $\lim_{g\to\infty} k = \theta - 1$, according to Eq.(63), we have $\lim_{g\to\infty} V_R = 0$.

Further, since $\lim_{g\downarrow 0} r = \rho$ and $\lim_{g\downarrow 0} k = \infty$, according to Eq.(63), we have $\lim_{g\downarrow 0} V_R = (\tilde{\pi}_M/\rho)\,[\eta/(\eta-1)]\,e^{-(\eta-1)\ln\chi}$. Then, if $\lim_{g\downarrow 0} V_R > R$ and $(\sigma_M-\sigma_C)+\chi(\sigma_C-1) < 0$ hold, we obtain a unique equilibrium g, which satisfies the free-entry condition $R = V_R$.

E Proof of proposition 3

We prove the proposition 3. According to Eq. (34), when $p \to 0$, the equilibrium η is given by

$$\eta = \frac{1}{(\sigma_M - \sigma_C)e^{-\eta\ln\chi} + \sigma_C}. \tag{64}$$

Note that $\eta > 1$ must hold to ensure a finite mean $e^{a^*_\chi}$. To ensure $\eta > 1$, we suppose that $(\sigma_M - \sigma_C) + \chi(\sigma_C - 1) < 0$ (see Appendix A). According to Eqs.(42) and (43), when $p \to 0$, the average patentable innovation size is

$$e^{a^*_\chi} = e^{a_\chi} = \left(\frac{\eta}{\eta-1}\right)e^{\ln\chi}. \tag{65}$$

Differentiating Eq.(64), evaluating at $\ln\chi = 0$, and simplifying it yields

$$\left.\frac{d\eta}{d\ln\chi}\right|_{\ln\chi=0} = \left(\frac{\sigma_M - \sigma_C}{\sigma_M}\right)\eta^2. \tag{66}$$

Differentiating Eq.(65) yields

$$\frac{de^{a^*_\chi}}{d\ln\chi} \gtreqless 0 \iff \eta(\eta-1) \gtreqless \frac{d\eta}{d\ln\chi}. \tag{67}$$

Noting that $\eta = 1/\sigma_M > 1$ when $\ln\chi = 0$, according to Eqs.(66) and (67), we yield

$$\left.\frac{de^{a^*_\chi}}{d\ln\chi}\right|_{\ln\chi=0} \gtreqless 0 \iff \sigma_C \gtreqless (\sigma_M)^2.$$

Then, we yield the proposition 3.

References

[1] Acemoglu, D. 2009. Introduction to modern economic growth. Princeton University Press.

[2] Acemoglu, D., Cao, D. 2015. Innovation by entrants and incumbents. J Econ Theory, 157, 255-294.

[3] Aghion, P., Howitt, P. 1998. Endogenous growth theory. MIT Press, Cambridge, MA.

[4] Arrow, K.J. 1962. Economic welfare and the allocation of resources for invention. In: Nelson RR (ed) The rate and direction of inventive activity: economic and social factors. Princeton University Press, Princeton, 609-626.

[5] Benhabib, J., Bisin, A., Zhu, S. 2016. The distribution of wealth in the Blanchard-Yaari model. Macroecon Dyn, 20(2), 466-481.

[6] Federal Trade Commission (FTC). 2003. To promote innovation: A proper balance of competition and patent law and policy. Available at http://www.ftc.gov/reports/promote-innovation-proper-balance-competition-patent-law-policy.

[7] Fujiwara, Y. Guilmi, C.D., Aoyama, H., Gallegati, M., Souma, W. 2004. Do Pareto-Zipf and Gibrat laws hold ture? An analysis with European firms. Physica A, 335(1), 197-216.

[8] Gabaix, X. 2009. Power laws in economics and finance. Annu Rev Econ, 1, 255-293.

[9] Gabaix, X., Lasry, J.M., Lions, P.L., Moll, B. 2016. The dynamics of inequality. Econometrica, 84(6), 2071-2111.

[10] Grossman, G.M., Helpman, E. 1991. Innovation and growth in the global economy. MIT Press, Cambridge, MA.

[11] Havranek, T. Horvath, R., Irsova, Z., Rusnak, M. 2015. Cross-country heterogeneity in intertemporal substitution. J Int Econ, 96(1), 100-118.

[12] Jones, C.I., Kim, J. 2015. A Schumpeterian model of top income inequality. Mimeo.

[13] Kesten, H. 1973. Random difference equations and renewal theory for products of random matrices. Acta Mathematica, 131(1), 207-248.

[14] Kishi, K. 2014. A patentability requirement and industries targeted by R&D. Osaka University, Graduate School of Economics and Osaka School of International Public Policy (OSIPP) Discussion Papers in Economics and Business, 14-27-Rev.

[15] Koléda, G. 2008. Promoting innovation and competition with patent policy. J Evol Econ, 18(3-4), 433-453.

[16] König, M.D., Lorenz, J., Zilibotti, F. 2016. Innovation vs. imitation and the evolution of productivity distributions. Theor Econ, 11(3), 1053-1102.

[17] Li, C.W. 2001. On the policy implications of endogenous technological progress. Econ J, 111(471), C164-C179.

[18] Lucas, R.E., Jr., Moll, B. 2014. Knowledge growth and the allocation of time. J Polit Econ, 122(1), 1-51.

[19] Luttmer, E.G.J. 2007. Selection, growth, and the size distribution of firms. Q J Econ, 122(3), 1103-1144.

[20] Luttmer, E.G.J. 2012. Technology diffusion and growth. J Econ Theory, 147(2), 602-622.

[21] National Academy of Science (NAS). 2004. A patent system for the 21st century. Available at http://www.nap.edu/openbook.php?isbn=0309089107.

[22] Newman, M.E.J. 2005. Power laws, Pareto distribution and Zipf ' s law. Contemp Phys, 46(5), 323-351.

[23] Nirei, M., Aoki, S. 2016. Pareto distribution of income in neoclassical growth models. Rev Econ Dyn, 20, 25-42.

[24] O'Donoghue, T., Zweimüller, J. 2004. Patents in a model of endogenous growth. J Econ Growth, 9(1), 81-123.

[25] Perla, J., Tonetti, C. 2014. Equilibrium imitation and growth. J Polit Econ, 122(1), 52-76.

[26] Perla, J., Tonetti, C., Waugh, M.E. 2015. Equilibrium technology diffusion, trade, and growth. Mimeo.

[27] Reed, W.J. 2001. The Pareto, Zipf and other power laws. Econ Lett, 74(1), 15-19.

[28] Silverberg, G., Verspagen, B. 2007. The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance. J Econometrics, 139(2), 318-339.

[29] Steindl, J. 1965. Random processes and the growth of firms: a study of the Pareto law, economic theory and applied statistics. Hafner publishing company.

[30] Toda, A.A., Walsh, K. 2015. The double power law in consumption and implications for testing Euler equations. J Polit Econ, 123(5), 1177-1200.