From imitation to innovation: Where is all that Chinese R&D going?*

Michael König† Kjetil Storesletten‡ Zheng Song§ Fabrizio Zilibotti¶

April 5, 2018

Preliminary Draft

Abstract

We construct a dynamic model where firms are heterogenous in productivity and are subject to distortions. The productivity distribution evolves endogenously as the result of the decisions of individual firms that seek to upgrade over time their productivity. Firms can adopt two strategy to improve their productivity: imitation and innovation. The theory bears predictions about the behavior of firms and the aggregate equilibrium. We perform the structural estimation of the stationary state of the dynamic model using a Simulated Method of Moments approach which targets moments of the empirical distribution of R&D and productivity growth. We estimate the model using data from Taiwan and mainland China. The estimation highlights some interesting findings. The model predictions align well with the data, and the estimated model also yields a good quantitative fit to the data. There are some important differences between Taiwan and China. R&D investments are significantly less productive in China. The evidence is consistent with a significant extent of fudging and overreporting of R&D in China. After correcting for this behavior, R&D investments are estimated to be as productive in China as in Taiwan.

JEL Codes:

Keywords: China, Imitation, Innovation, Misallocation, Moral Hazard, Productivity, R&D, Subsidies, Taiwan, Traveling Wave.

1 Introduction

This paper analyzes from a theoretical and empirical perspective the determinants of productivity growth. We construct a dynamic model where firms are heterogenous in productivity and are subject to distortions in the fashion of Hsieh and Klenow (2009). The productivity distribution evolves endogenously as the result of the decisions of individual firms that seek to upgrade over time their productivity. Firms can adopt two strategies to improve their productivity: they can either attempt to imitate the technology used by other firms or make an investment that allows them to explore new avenues to increase their productivity. A key factor for the imitation-vs-innovation decision is the state

*We would like to thank Chang-tai Hsieh, Jesse Perla, Torsten Persson, Daniel Xu, and seminar participants at Cambridge University, CIFAR, National University of Singapore, Tsinghua University, University College London, University of St. Andrews, Yale University, 2017 Asian Meeting of the Econometric Society.

†University of Zurich, Department of Economics, michael.koenig@econ.uzh.ch.

‡University of Oslo, Department of Economics, kjetil.storesletten@econ.uio.no.

§Chinese University of Hong Kong, Department of Economics, zheng.michael.song@gmail.com.

¶Yale University, Department of Economics, fabrizio.zilibotti@yale.edu.

of productivity, which determines the comparative advantage of the two strategies. The farther behind frontier the firm is, the more likely it is to succeed in imitating technologies from other firms. Conversely, for firms close to the frontier the scope for learning from other firms is more limited, and they must try to break new grounds in order to improve. The investment decision is also affected by other firm characteristics, most notably, labor and capital market distortions (wedges) that can be retrieved from the data using the methodology proposed by Hsieh and Klenow (2009). The main contribution of the paper is the empirical implementation of the theory: we propose a structural estimation of the stationary equilibrium of the dynamic model using a Simulated Method of Moments approach which targets moments of the empirical distribution of R&D and productivity growth.

We estimate the model using data from Taiwan and China. We are motivated from the observation that in recent years the stellar economic performance of China has been accompanied by a growing emphasis on innovation. China's economic development has been driven for long by high investment rates, the reallocation of resources between sectors and firms, and the adoption technologies that were invented abroad (see, e.g., Zilibotti 2017). Yet, with increasing development, these engines of growth are losing power. The aspiration to a transition towards innovation-based growth has become a central part of the discourse of the political elite. For example, according to the National Innovation-driven Development Strategy Outline issued in June 2016, China should transform into an innovation-oriented economy by 2020 and into a technological innovation powerhouse by 2050.[1] Consistent with the new emphasis, there has a been a boom in Chinese R&D investments, a standard measure of innovative investments. If in the 1990s China invested barely 1% of its GDP, in 2017 China's R&D-GDP ratio has increased to 2.12%. Yet, there it is questionable whether this boom in R&D expenditure reflects a real change of gear in innovation. The skeptical view is that it stems from a muscular state intervention and that the high R&D expenditure need not result in a major push on productivity growth, if R&D expenditure is either misallocated (i.e., it is done by firm that are not well suited to innovate) or, worse, if it simply reflects firms' creative accounting to benefit from fiscal incentives to R&D.

Our paper aims to cast some light on these important questions with the aid of a theory and the data. In order to make progress, we use the micro-level evidence from Taiwan as a benchmark for mainland China. Taiwan offers good census data about R&D investments. While we could alternatively use a Western economy, Taiwan is a more natural comparison for mainland China, not only because of its geographic and cultural proximity, but also because of the structural similarities between the two regions. In both cases, economic development has been led by the manufacturing sector, first through the growth of low value added industries and later through a transition to a higher level of technological sophistication (e.g., electronics and semiconductors). Both economies have still today a relatively large manufacturing sector accounting for about 30% of the respective GDP.

The estimation highlights some interesting findings. The model predictions align well with the data, and the estimated model yields a good quantitative fit to the data. There are some important differences between Taiwan and China. In mainland China, the firm-level productivity of R&D investments appears to be significantly lower than in Taiwan. In addition, while in both economies high-productivity firms are more likely to invest in R&D than low-productivity one (an observation that is in line with a key prediction of the theory), the correlation is significantly higher in Taiwan than in China. In contrast, policy distortions are more important for China. Overall, the model fits better the data of Taiwan than

[1] The 13th Five-Year Plan (2016–2020) emphasizes the importance of promoting research in strategic and frontier fields. In his speech at the G20 Summit in September 2016, Xi Jinping emphasized that China will pursue innovation-driven growth and called for the creation of an innovative world economy. Likewise, Premier Li Keqiang argued that innovation is the most important driver for development and, thereby, necessarily has to take on a central role in China's development strategy.

the data of China.

We explore then the possibility that the less sharp results for China be driven by firms fudging and overreporting R&D expenditure. According to this hypothesis, the existence of generous public subsidies (documented below) induces many Chinese firms to relabel part of their expenditure as R&D in order to qualify for them. We formalize this hypothesis by introducing moral hazard into the theory in a way that allows us to identify empirically which proportion of firms pretend to spend on R&D without really doing so at the different levels of the productivity distribution. We find that moral hazard has no impact on the estimates for Taiwan (suggesting that this is not an important issue in Taiwan) while it improves significantly the fit of the model for mainland China (suggesting that this is an important issue in China). Moreover, if one corrects for fake R&D, the key technology parameters (except for R&D costs) are estimated to be very similar for Taiwan and mainland China. According to our structural estimates, on the one hand a large share of Chinese firms report opportunistically R&D investments they do not perform. On the other hand, the real R&D investments increase productivity significantly. To check the plausibility of the results, we match the data for R&D investments to the data for patents. We find that R&D investments have a sizeable effect on future productivity growth when one restricts attention to Chinese firms that both invest in R&D and patent new inventions.

The contribution of the paper is twofold. On the one hand, the application provide interesting lessons for understanding the economic development of China. On the other hand, the insights of the paper go beyond the specific (if important) case of China. Over the last few years, there has been a surge of theoretical studies of the evolution of the distribution of firm size and productivity. However, there has been no attempt, to the best of our knowledge, to test the empirical predictions of these theories. In contrast, the literature on misallocation has proven powerful in guiding empirical studies of the extent and effects of misallocation of productive factors. However, this literature is limited by the static nature of the analysis, especially the fact that differences in total factor productivity are assumed to be exogenous. Our paper combines the insights of the two literatures, and study the effect of misallocation on the evolution of productivity growth.

1.1 Literature Review (very preliminary and incomplete)

Our paper is related to various streams of the growth and development literature. First, it broadly relates to literature on the determinants of success and failure in technological convergence (e.g., Hall and Jones 1999, Klenow and Rodriguez-Clare 1997, Acemoglu and Zilibotti 2001, Hsieh and Klenow 2010). The importance of technology diffusion stretches back to the seminal work of Griliches (1957). The importance of R&D investments and spillovers for growth and technology diffusion is a core element of the neo-Schumpeterian theory à la Aghion and Howitt (1992), see also Griliches (1998). While many of these papers emphasize a process of creative destructions whereby new firms are carriers of innovation, recent research by Garcia Macia, Hsieh and Klenow (2016) finds that a larger share of the productivity improvements originate from incumbent firms. This evidence is consistent with the tenets of our theory.

The dichotomy between innovation and imitation in the process of development is emphasized by Acemoglu, Aghion and Zilibotti (2006). The important role of misallocation as a determinant of aggregate productivity differences is related to the influential work of Hsieh and Klenow (2009). Our study builds on their methodology, although it attempts to endogenize the distribution of productivity across firms, which is instead exogenous in their work. The importance of misallocation in China is also emphasized, among others, by Song, Storesletten, and Zilibotti (2011), Hsieh and Song (2016), Cheremukhin, Golosov, Gurev, and Tsyvinski (2016), Tombe and Zhu (2016), Brandt, Kambourov, and Storesletten (2017).

Our paper is also closely related to the recent theories describing the endogenous evolution of the distribution of firm size and productivity. These studies include, among others, Jovanovic and Rob (1989), Luttmer (2007 and 2012), Ghiglino (2012), Perla and Tonetti (2014), Acemoglu and Cao (2015), Lucas and Moll (2015), Koenig, Lorenz, and Zilibotti (2016), Benhabib, Perla, and Tonetti (2017), Akcigit *et al.* (2018).

Finally, our paper is related to the literature studying the effect of R&D on growth and the effect of policy on R&D investments. These studies include, among others, Klette and Kortum (2004), Lentz & Mortensen (2008), Acemoglu, Hsieh and Klenow (2015), Akcigit and Kerr (2017), Acemoglu, Akcigit, Alp, Bloom, and Kerr (2017).

1.2 Literature on China's policy to stimulate innovation

Our paper is also related to the empirical literature studying R&D policy in China. Ding and Li (2015) provide a comprehensive overview of the instruments adopted by the Chinese government intervention to foster R&D. The systematic policy intervention to stimulate innovation had its first impetus in 1999 and accelerated in 2006 with the adoption of the Medium and Long-term National Plan for Science and Technology Development. The policy instruments are manifold. The first is direct government funding of research through the establishment of tech parks, research centers, and a series of mission-oriented programs. The most important among such programs is Torch, a program aimed to kick-start innovation and start-ups through the creation of innovation clusters, technology business incubators, and the promotion of venture capital. Next, an important part of the government strategy is the provision of tax incentives for innovation. This takes the form of tax deductions applicable to wages, bonus and allowances of R&D personnel, corporate tax rate cuts, and R&D subsidies. For instance, firms are granted a 150% tax allowance against taxable profits on the level of R&D expenditure and 100% tax allowance against taxable profits on donations to R&D foundations. In addition, firms that qualify as innovative can obtain exemption from import duties and VAT on imported items for R&D purposes. Firms that are invited to join science and technology parks are often exempted from property taxes and urban land use. Finally, "innovative firms" receive subsidies on investments. Unused tax allowance can be carried forward to offset future taxes.

The policy interventions leave ample margins for discretion. For instance, central and local governments can decide which firms to invite to be part of science and technology parks, which firms receive priority in High-Tech Special Economic Zones, etc. In short, incentives can be very heterogenous across provinces, local communities, sectors, and even at the firm level, often as a function of political connections (Bai, Hsieh and Song 2016).

Some empirical studies attempt to evaluate the effects R&D investments and R&D policy in China. Hu and Jefferson (2009) use a data set that spans the population of China's large and medium size enterprises for the period from 1995 to 2001. In spite of not being a representative sample, these enterprises performed nearly 40% of China's R&D in 2001. The authors estimate a patents–R&D elasticity is 0.3 when evaluated at the sample mean of the real R&D expenditure (and even lower at the median). This is much smaller than similar estimates for the U.S. and European firms for which the result of earlier studies find elasticities in the range 0.6–1. The study is based on data in the 1990s. However, Dang and Motoyashi (2013) find similar results using data for the period 1998–2012 based on matching the NBS to patent data.

Jia and Ma (2017) use a panel dataset of Chinese listed companies covering 2007 to 2013 to assess the effects of tax incentives on firm R&D expenditures and analyze how institutional conditions shape these effects. They show that tax incentives have significant effects on the R&D expenditure reported by firms. A 10% reduction in R&D user costs leads firms to increase R&D expenditures by 4% in the short run. They also document considerable effect heterogeneity: Tax incentives significantly stimulate R&D in private firms but have a lower influence on state-owned enterprises' R&D expenditures.

The paper is that is closest to our research is Chen, Liu. Suarez Serrato and Xu (2017). They analyze the effects of the InnoCom program, a large scale incentive for R&D investment in the form of a corporate income tax cut. They exploit variation over time in discontinuous tax incentives to R&D. Before 2008, firms with an R&D investment over revenue above 5% qualified for a large tax reduction, from 33% to 15%. After 2008, a new legislation was enacted implying three thresholds of 3%, 4%, and 6% for firms of different size. By combining administrative tax data and survey data, they analyze whether and how firms respond to the tax incentive. They find that firms are highly responsive to the tax incentives in the InnoCom program and they show that there is significant bunching at the various R&D policy notches. Moreover, the response of firms suggests a significant amount of fudging, in particular, a large fraction of the firms appear to respond to the tax incentive by relabeling non-R&D expenditures as R&D expenses.

The paper is structured as follows: Section 2 presents the theory. Section 3 discussed the data and some descriptive evidence for Taiwan and mainland China. Section 4 presents the methodology and results of the structural estimation. Section 5 concludes. An appendix contains technical results and additional tables and figures.

2 Theory

Consider a dynamic economy populated by a unit measure of monopolistically competitive long-lived firms.

2.1 Static firm problem

Consider first the firm's static problem. The production technology is Cobb Douglas in capital and labor and features constant returns to scale. The production $Y_i(t)$ of firm i at time t is,

$$Y_i(t) = A_i(t) K_i(t)^{\alpha} L_i(t)^{1-\alpha},$$

where $\alpha \in (0,1)$, $K_i(t)$ is capital input, $L_i(t)$ is labor input, and $A_i(t)$ is firm-specific TFP. Following Hsieh and Klenow (2009), firms have heterogeneous TFP and rent capital and labor at competitive markets, subject to distortions. We summarize the distortions in a single output wedge, denoted by τ_i. Note that $\tau_i < 0$ indicates a negative wedge, which one can interpret as an output subsidy. In our analysis, τ_i can be regarded either as combination of capital and labor wedges, or as an output wedge.[2] Firms produce differentiated goods that are combined into a homogenous final good by a Dixit-Stiglitz aggregator with a constant elasticity of substitution $\eta > 1$.[3]

[2] More formally, let τ_{Ki} and τ_{Li} denote, respectively, a firm-specific "tax" on capital and labor, respectively. Then, the output wedge τ_i is defined by the following equation:

$$1 - \tau_i \equiv (1+\tau_{Ki})^{-\alpha} (1+\tau_{Li})^{-(1-\alpha)}.$$

[3] More formally, we assume a standard CES aggregation where $Y = \left(\int_0^1 Y_i^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}$ where $\eta > 1$ is the elasticity of substitution.

We present the full microfoundation of the static model and characterize equilibrium in the appendix. Here, we simply discuss the two equilibrium conditions that will be used in the empirical investigation and that are sufficient to derive the dynamic equilibrium. Firm i's current (period t) profits are given by

$$\pi_i(t) = \Xi \times (A_i(t)(1-\tau_i(t)))^{\eta-1}, \tag{1}$$

where Ξ is a constant that depends on the parameters of the model, τ_i is the output wedge, and A_i denotes total factor productivity (TFP). Intuitively, profits increase in TFP and decrease in the wedge. Moreover, firm i's revenue satisfies

$$P_i(t)Y_i(t) = \Psi \times (A_i(t)(1-\tau_i(t)))^{\eta-1}, \tag{2}$$

where Ψ is a constant. Intuitively, revenue is higher for firms with high TFP and a low wedge.

We assume $\tau \in \{\tau_h, \tau_l\}$ where $\tau_l < \tau_h$. The realization of τ follows a persistent Markov process with parameters $p_{\tau_h} > 0.5$ and $p_{\tau_l} > 0.5$. This implies that a high-wedge (low-wedge) firm in the current period is more likely be a high-wedge (low-wedge) in the next period. In the empirical analysis, we allow a more flexible characterization of τ.

2.2 Productivity dynamics

The focal point of the analysis is the endogenous evolution of the productivity distribution and the strategy firms adopt to increase their productivity over time. Let $\hat{a}_i \equiv \log(A_i)$. Advancement occurs over a productivity ladder where each successful attempt to move up the ladder results in a constant log-productivity accrual: $\hat{a}_{i,t+1} = \hat{a}_{i,t} + \tilde{a}$, where $\tilde{a} > 0$ is a constant (thus, $\hat{a} \in \{\tilde{a}, 2\tilde{a}, ...\}$). We define $a \equiv \hat{a}/\tilde{a}$ and denote by $a \in \mathcal{N}^+$ the ranking in the productivity ladder. We denote by P the productivity distribution, by $P_1, P_2, ...$ the proportion of firms at each rung of the ladder, and by $F_j = \sum_{b=1}^{j} P_b$ the associated cumulative distribution. We abstract from entry and exit, and assume a constant population of firms. Finally, we model innovation as a step-by-step process. In previous research, we allowed for the possibility of heterogenous productivity increases, including large steps, from which we abstract here from simplicity.

Firms can increase their productivity via two alternative strategies: *innovation* and *imitation.* Imitation is modelled as an attempt to improve productivity through learning from other firms (e.g., by adopting better managerial practices). More formally, a firm pursuing the imitation strategy is randomly matched with another firm in the empirical distribution. If a firm is matched with a more productive firm, it adopts its higher productivity with probability $q > 0$. If the firm is matched with a less productive firm (or if it fails an adoption attempt), it retains its initial productivity.[4] Because of random matching, the probability that an imitating firm with log-productivity a moves up the ladder equals $q\sum_{k=1}^{\infty} P_{a+k} = q(1-F_a)$.

A firm pursuing innovation can improve its productivity via two channels. First, by discovering something genuinely new that is unrelated to the knowledge set of other firms. We label this process *in-house R&D,* and denote by $p_i > 0$ the probability of success through in-house R&D. We assume that p is drawn from an i.i.d. distribution with density function $g : [0, \overline{p}] \rightarrow \Gamma \subset R^+$ and c.d.f. $G : [0, \overline{p}] \rightarrow [0, 1]$ where $\overline{p} < 1$. The realization of p is observed at the beginning of each period t, before firms choose whether to innovate or imitate. If in-house R&D fails, the firm can still improve via imitation. In this case, the probability of success is $\delta q(1-F_a)$. Thus, the total probability of success of a firm pursuing innovation is $p_i + (1-p_i)\delta q(1-F_a)$. In the basic model, an interesting trade-off

[4] Matching is directed, in the sense that there is no effect for the targeted firm.

arises only if $\delta < 1$ (otherwise all firms prefer to innovate). However, δ need not be less than unity when we introduce a cost a doing in-house R&D below.

In the model outlined thus far, maximizing TFP growth is equivalent to maximizing the value of the firm (i.e., the discounted value of profits). In particular, it is optimal for firm i to choose the innovation strategy if and only if

$$p_i \geq Q(a, \tau; P) \equiv \frac{q(1-\delta)(1-F_a)}{1-\delta q(1-F_a)}, \tag{3}$$

where P denotes the productivity distribution. Note that, since $Q_a < 0$, the proportion of innovating firms will be non-decreasing in the initial productivity. The intuitive reason is that imitation is less effective for high-productivity firms, because these are less likely to be matched with a more productive firm. Note that we specify τ as an argument of the function Q for future reference, although τ has for the moment no bearing on the innovation-imitation decision of the firm.

It is useful to introduce some new notation. The function

$$\chi^{\text{im}}(a, p, \tau; P) = 1 - \chi^{\text{in}}(a, p; P) = \begin{cases} 1 & \text{if} \quad p \leq Q(a, \tau; P) \\ \\ 0 & \text{if} \quad p > Q(a, \tau; P) \end{cases} \tag{4}$$

is an indicator that takes the unit value when a firm finds it optimal to imitate. Conversely, $\chi^{\text{in}}(a, p, \tau; P)$ indicates the case in which the firm conducts innovation.

A salient feature of the theory is that the productivity growth gap between innovating and imitating firms is increasing in the TFP level. To see why, consider a range of productivities where the imitation/innovation choice depends on the realization of p (at low a's, all firms find it optimal to imitate, irrespective of the realization of p). In particular, all firms drawing $p \in [Q(a, \tau; P), \bar{p}]$ will pursue the innovation strategy, while the other firms will try to imitate. The productivity growth gap is $(1 - \delta q(1-F_a)) \times \left(\int_{Q(a,\tau;P)}^{\bar{p}} p dG(p) / (1 - G(Q(a, \tau; P))) - Q(a, \tau; P) \right)$, which is an increasing function of a. Intuitively, the gap increases in a because (i) the productivity of imitation also declines, and (ii) the mass of inframarginal firms that strictly prefer to innovate increases.[5]

We can now write the law of motion of the distribution of log-productivity, $P_a(t)$. This is given by the following system of integro-difference equations:

$$\begin{aligned} & P_a(t+\Delta t) - P_a(t) \\ &= \int_{\underline{p}}^{\bar{p}} \left[\begin{array}{c} \chi^{\text{in}}(a-1, p, \tau; P) \times (p + (1-p)\delta q(1-F_{a-1}(t))) P_{a-1}(t) + \\ +\chi^{\text{im}}(a-1, p, \tau; P) \times q(1-F_{a-1}(t)) P_{a-1}(t) \\ -\chi^{\text{in}}(a, p, \tau; P) \times (p + (1-p)\delta q(1-F_a(t))) P_a(t) \\ -\chi^{\text{im}}(a, p, \tau; P) \times q(1-F_a(t)) P_a(t) \end{array} \right] dG(p) \end{aligned} \tag{5}$$

The first and second line inside the integral sign capture, respectively, successful innovating and imitating firms that had productivity $a-1$ in period t and that managed to move up the ladder. The third and fourth line inside the integral capture, respectively, successful innovating and imitating firms that had a productivity a in period t and that move to $a+1$.

The next proposition characterizes the stationary distribution associated with this system of difference equation. To prove our results, we let $\Delta t \to 0$, which allows us to approximate the solution by

[5] Note, though, that the productivity of in-house R&D is on average higher for low-TFP firms. This is due to a selection effect: only low-productivity firms facing a very high realization of p choose R&D.

a system of ordinary differential equations that is more tractable. However, numerical analysis shows that the solution has the same qualitative properties (i.e., traveling wave with Pareto tails), when Δt is not infinitesimal.

Proposition 1 *Consider the model of innovation-imitation described in the text whose equilibrium law of motion satisfies Equation (5), where each firm draws p from a distribution $G : [0, \overline{p}] \to [0, 1]$. Let $\Delta t \to 0$, $\hat{p} \equiv \int_0^{\overline{p}} p \; dG(p)$, and assume that $q > \hat{p}$. Then, there exists a travelling wave solution of the form $P_a(t) = f(a - \nu t)$ with velocity $\nu = \nu(q, \delta, g(p)) > 0$, with left and right Pareto tails. For a given t, P_a is characterized as follows: (i) for a sufficiently large, $P_a(t) = O\left(e^{-\rho(a-\nu t)}\right)$, where the exponent ρ is the solution to the transcendental equation $\rho\nu = \hat{p}(e^{\rho} - 1)$; (ii) for a sufficiently small, $P_a(t) = O\left(e^{\lambda(a-\nu t)}\right)$, where the exponent λ is the solution to the transcendental equation $\lambda\nu = q(1 - e^{-\lambda})$.*

Intuitively, a traveling wave is a productivity distribution that is stationary after removing the (endogenous) constant growth trend. Our proposition exploits (and generalizes) the result that random growth with a lower reflecting barrier generates a Pareto tail — a result formalized by Kesten (1973), and applied in economics by Gabaix (1999; 2009). Although our model features no reflecting barrier strictly speaking, for each p and at a given t there is a threshold $a^*(p, t)$ such that all firms below $a^*(p, t)$ imitate.[6] Among imitators, the probability of progress is higher for the less productive than for the more productive firms, since imitators improve sample randomly from the distribution of firms and improve their productivity with probability $q(1 - F_a) \geq p_i$.[7] Thus, the sub-distribution of imitating firms catches up preventing the upper end of the distribution from blowing up.

Interestingly, and different from earlier studies, our model also feature a Pareto tail of low-productivity firms. The left tail originates from the fact, on the one hand, low-productivity firms grow faster on average, while on the other hand the threshold $a^*(p, t)$ increases over time. This prevents the convergence of the sub-distribution of imitating firms to a mass point. Figure 1 illustrates the mechanism.[8] Panel A shows the threshold and the force implying convergence. Panel B illustrates the traveling wave.

FIGURE 1 (Panels A and B) HERE

In Proposition 1, there is no algebraic representation of the velocity of the travelling wave, and ν can only be defined implicitly and solved for numerically. In the appendix we provide the analytical characterization of an approximation for a particular case in which there is no heterogeneity in p and in-house R&D precludes altogether any possibility of learning through imitation ($\delta = 0$), see Remark 1. In this case, all firms above a certain productivity threshold imitate and all firms below that threshold innovate, as in König, Lorenz, and Zilibotti (2016). Although this special case is too stark to match the empirical data (for instance, the assumption that all firms are equally productive at doing R&D is difficult to reconcile with the data), it provides a useful illustration of the mechanics of the model.

[6] Note that the stationary nature of the traveling wave implies that $a^*(p, t + \nu t) = a^*(p, t) + \nu t$.

[7] When p is stochastic, there exists $a(\bar{p})$ such that all firms with $a \leq a(\bar{p})$ imitate, irrespective of the realization of p. All firms in this part of the distribution have a higher expected productivity growth than firms with $a > a(\bar{p})$, irrespective of the realization of p.

[8] The solutions for ρ and λ are in an implicit form and involve transcendental equations. Standard methods allow one to show that the equations $\rho\nu = q\left(1 - e^{-\rho}\right)$ and $\lambda\nu = \bar{p}\left(1 - e^{-\lambda}\right)$ admit closed-form solutions for ρ and λ if, respectively, $\frac{q}{\nu}e^{-\frac{q}{\nu}}/\nu \leq e^{-1}$ and $\frac{\bar{p}}{\nu}e^{-\frac{q}{\nu}} \leq e^{-1}$. In particular, $\lambda = W\left(-\frac{q}{\nu}e^{-\frac{q}{\nu}}\right) + \frac{q}{\nu}$ and $\rho = W\left(-\frac{\bar{p}}{2\nu}e^{-\frac{\bar{p}}{2\nu}}\right) + \frac{\bar{p}}{2\nu}$ where W denotes the Lambert-W function.

2.3 Productivity Dynamics with Costly Innovation

Next, we extend the analysis to an environment in which innovation requires a costly investment while maintaining the assumption that imitation does not entail any cost. We assume that the innovation investment is entirely on the extensive margin. Namely the R&D investment is a fixed cost. The current profits are then

$$\pi_i(t) = \Xi(1-\tau_i(t))^{\eta-1}A_i(t)^{\eta-1} - c_i(t),$$

where

$$c_i(t) = \begin{cases} \bar{c}\bar{A}(t)^{\eta-1} & \text{if} \quad i \text{ innovates} \\ 0 & \text{if} \quad i \text{ imitates} \end{cases} . \tag{6}$$

Here, $\bar{A}(t)$ denotes the average productivity at time t and $\bar{c} > 0$ is a parameter. Note that, for simplicity, we set $\Xi = R^{-1}$, where R denotes the gross interest rate. This normalization entails no loss of generality since the equilibrium is pinned down by the ratio $\bar{c}/\Xi$, allowing one to normalize either of the two constants.

The choice of the optimal strategy to upgrade productivity is now more involved, because the investments in R&D affects the future productivity of the firm and its future strategy, which in turn depends on the future productivity distribution. The distribution P is in turn endogenous. While the problem can be written down using dynamic programming techniques and solved numerically, it does not allow any analytical characterization. Moreover, the stationary distribution is typically not independent of the initial conditions. Since our main goal is to estimate the model, these are all severe complications that hinder the econometric implementation of the theory. Therefore, we make an important simplifying assumption that renders the current model similar to the model of Proposition 1. We assume that firms are owned by overlapping generations of two-period lived manager-entrepreneurs.[9] Following Song, Storesletten and Zilibotti (2011), we assume that in each period the firm is owned by an old entrepreneur who is residual claimant on the firms' profits, and run by a young manager.[10] We assume that the manager makes the R&D investment decision affecting next-period productivity. The manager pays for the R&D investment, and finance this investment having access to perfect capital markets.

Under these conditions, the function Q is now defined as

$$Q(a,\tau;P) = \frac{q(1-\delta)(1-F_a)(1-E(\tau'|\tau))^{\eta}\left(e^{(\eta-1)\tilde{a}}-1\right) + \bar{c}e^{(\eta-1)(\bar{a}-a)}}{(1-E(\tau'|\tau))^{\eta}\left(e^{(\eta-1)\tilde{a}}-1\right)(1-q\delta(1-F_a))},$$

where $E(\tau'|\tau)$ denotes the conditional expectation of next prior wedge. The definitions of $\chi^{\text{im}}(a,p;P)$ and $\chi^{\text{in}}(a,p;P)$ are unaffected.

The law of motion of productivity is now described by a version of Equation 5 that takes into

[9] Note that one can make the same assumption in the model with $\bar{c} = 0$.

[10] How the manager is paid is unimportant in our model. In Song, Storesletten and Zilibotti (2011), she receives a share of the current profit to satisfy an incentive-compatibility constraint. Here, we could make the same assumption, or alternatively assume that managers are simply paid a wage. Since the focus is the intertemporal decision, and there are by assumption no financial frictions, we abstract from information frictions between the owner and the manager, and let the compensation simply be that the manager inherits the firm.

account of the heterogeneity in wedges:

$$
\begin{aligned}
&P_a(t+\Delta t) - P_a(t) \\
&= \sum_{j\in\{l,h\}} \mu_{\tau_j}(t) \times \int_{\underline{p}}^{\overline{p}} \left[\begin{array}{c} \chi^{\mathrm{in}}(a-1,p,\tau_j;P) \times (p+(1-p)\,\delta q\,(1-F_{a-1}(t)))\,P_{a-1}(t)+ \\ +\chi^{\mathrm{im}}(a-1,p,\tau_j;P) \times q\,(1-F_{a-1}(t))\,P_{a-1}(t) \\ -\chi^{\mathrm{in}}(a,p,\tau_j;P) \times (p+(1-p)\,\delta q\,(1-F_a(t)))\,P_a(t) \\ -\chi^{\mathrm{im}}(a,p,\tau_j;P) \times q(1-F_a(t))P_a(t) \end{array} \right] dG(p), \quad (7)
\end{aligned}
$$

where $\mu_{\tau_l}, \mu_{\tau_h}$ denote the proportion of for low- and high-wedge firms. The model is closed by the law of motion for μ_{τ_l} :

$$
\mu_{\tau_l}(t+\Delta t) - \mu_{\tau_l}(t) = (1-p_{\tau_h})\left(1-\mu_{\tau_l}(t)\right) - (1-p_{\tau_l})\,\mu_{\tau_l}(t),
$$

where μ_{τ_l} converges in the long run to $\mu_{\tau_l} = (1-p_{\tau_h})\,/\,((1-p_{\tau_h})+(1-p_{\tau_l})) \equiv \bar{\mu}_{\tau_l}$.

The next proposition extends Proposition 1 to the model with costly R&D investments.

Proposition 2 *The characterization of Proposition 1 can be extended to a model with costly R&D investments where $\bar{c} > 0$. In this case, the solution is affected by the wedges and their distribution. More formally, there exists a travelling wave solution of the form $P_a(t) = f(a-\nu t)$ with velocity $\nu = \nu\left(q,\delta,g(p),\bar{c},\tau_h,\tau_l,\bar{\mu}_{\tau_l}\right) > 0$, with left and right Pareto tails. For a given t, P_a is characterized as follows: (i) for a sufficiently large $P_a(t) = O\left(e^{-\rho(a-\nu t)}\right)$, where the exponent ρ is the solution to the transcendental equation $\rho\nu = \hat{p}\,(e^{\rho}-1)$; (ii) for a sufficiently small $P_a(t) = O\left(e^{\lambda(a-\nu t)}\right)$, where the exponent λ is the solution to the transcendental equation $\lambda\nu = q(1-e^{-\lambda})$.*

FIGURE 2 (Panels A, B, C, and D) HERE

Consider Figure 2. The four panels plot the expected TFP growth as a function of the log-TFP level (a) conditional on innovation and imitation. Each plot shows firms with a given wedge facing a particular realization of p. Both schedules are downward sloping because imitation is more productive for laggards. Pursuing innovation flattens the profile, because part of what firms try to achieve is independent of the productivity distribution. Panel A refers to the case in which $\bar{c} = 0$, corresponding to Proposition 1. In this case, firms maximize expected TFP growth, and the wedge τ does not affect their choice. The threshold a^* is such that all firms with an initial log-productivity above a^* pursue the innovation strategy, whereas all firms with an initial log-productivity below a^* pursue the imitation strategy. The other three panels display a case in which $\bar{c} > 0$, as in Proposition 2. In panel B, we hold constant p and τ at the same level as in panel A ($\tau = \tau_l$). Now, the threshold a^* is shifted to the right relative to panel A, because innovation entails a cost. In this case, the expected TFP growth for firms at the threshold $a = a^*$ is higher for firms investing in in-house R&D than for firms pursuing imitation – higher expected growth is required to compensate for the R&D investment cost. Panel C shows a case in which $\tau = \tau_h$. The higher wedge shifts a^* further to the right. When innovation is costly, the expected growth rate is a decreasing function of the wedge when the wedge is positive. In general, the relationship between the wedge and expected growth is non-monotonic, since a sufficiently large subsidy may induce firms to do R&D even though imitation yields a higher expected growth rate. Finally, panel D shows the case of firms facing a higher realization of p (while $\tau = \tau_l$). A higher draw of p makes innovation more attractive and reduces a^*.

In summary, the model has four testable implications:

1. *Ceteris paribus*, the proportion of firms investing in R&D is increasing in TFP;
2. *Ceteris paribus*, firms with higher wedges are less likely to do R&D. It then follows from equation (2) that firms with higher sales are more likely to do R&D, even after conditioning on TFP;
3. TFP growth is inversely related to the current productivity, especially so for firms that do no invest in R&D;
4. The gap in TFP growth between firms investing and not investing in R&D increases with productivity a.

3 Data and Descriptive Evidence

We consider micro-level data for the economies of Taiwan and mainland China. The Chinese data is from the Annual Survey of Industries conducted by China's National Bureau of Statistics for the years between 1998 and 2007 and 2011-13. This survey is a census of all state-owned firms and the private firms with more than five million RMB in revenues in the industrial sector.[11] We have data for the R&D expenditure at the firm level in the periods 2001-2003 and 2005-2007. To estimate firm-level productivity growth, we focus on a balanced panel for all manufacturing firms in the period 2001-2007, focusing on firms that are in our sample for both in 2001 and 2007. Although this is a firm-level survey, most of the Chinese firms were single-plant firms during this period. For robustness we also study the sample of the balanced panel between 2007 and 2012.[12]

The Taiwanese data is at the plant level, collected by Taiwan's Ministry of Economic Affairs, for the years between 1999 and 2004. Like China, more than 90 percent of Taiwanese manufacturing plants are owned by single-plant firms in the period. Following Aw, Roberts and Xu (2011), we ignore the distinction between plant and firm for the Taiwanese data. With some abuse of notation, we will use the word "firm" to refer to the observations in the Chinese and Taiwanese samples. To make the Taiwanese sample more comparable to its Chinese counterpart, we drop the firms with annual sales below 18 million Taiwan dollars.[13]

TABLE 1 HERE

Table 1 reports summary statistics for the Chinese and Taiwanese balanced panels. Chinese firms are on average larger than Taiwanese firms. The difference in size is in large part accounted for by the Chinese state-owned enterprises. The fraction of R&D firms (i.e., those reporting positive R&D expenditure) is stable for the Chinese firms: 17 and 18 percent in 2001 and 2007, respectively. The fraction of R&D firms is lower in the Taiwanese sample: 13 and 10 percent in 1999 and 2004, respectively. Our empirical analysis will primarily focus on the extensive margin of R&D decision. We later examine R&D intensity as a robustness check. The first motivation for focusing on the extensive margin is that in our model R&D investment as a discrete choice, making it natural to confront the

[11] The selection criteria changed to sales above 20 million Yuan for all firms in 2010.

[12] We do not use the 2013 firm data because China's National Bureau of Statistics adjusted the definition of firm employment in 2013, making the 2013 employment data inconsistent with those in the earlier years.

[13] Our main facts are robust to firm size thresholds (see Appendix XXX). The exchange rate between Chinese Yuan (Taiwan dollars) and US dollars was about 8 Yuan/US$ (30 NT$/US$) in the sample periods.

theoretical implications with the empirical data along the extensive margin. The second motivation is that the intensive margin is subject to more severe measurement errors. This issue has been noted in the literature that has studied firm-level R&D expenditure in Western countries (see, e.g., Lichtenberg 1992; Acemoglu, Aghion, Griffith, and Zilibotti 2010).[14] In spite of this, it is useful to note that Taiwanese R&D firms spend on average significantly more on R&D: their mean R&D-revenue ratio is more than three times larger in Taiwan than in mainland China.

The figures that follows show the distribution of R&D and TFP growth broken down by the percentiles of TFP (in level, in the initial year) and value added. TFP is estimated using our model following the methodology proposed by Hsieh and Klenow (2009), which we discuss in more detail in Section 4 below and in Appendix **??**. In all graphs, TFP and value added are normalized by the median value in the two-digit manufacturing industry to which each firm belongs. This disaggregation yields 23 and 30 industries for Taiwan and China, respectively.[15] We exclude from our analysis the bottom 10% of the distribution because the survivor bias is especially severe for low-productivity firms. We refer to firms reporting and not reporting a positive level of R&D expenditure as R&D and non-R&D firms, respectively.

Figures 3 and 4 are for Taiwan and mainland China, respectively. Each figure comprises four panels.

FIGURE 3 (Panels A, B, C, and D) HERE
FIGURE 4 (Panels A, B, C, and D) HERE

Panel A in each figure shows that the share of R&D firms is increasing in TFP. This is in line with our innovation-imitation theory which postulates that imitation is more attractive for less productive firms. The pattern is more pronounced in Taiwan than in mainland China. The figure shows that in Taiwan the propensity to perform R&D increases steeply in TFP after the 6th decile. The percentage of R&D firms goes up from 10% to over 35% in the four top deciles. The increase is less steep in mainland China. There, the percentage of R&D firms at low TFP levels is higher than in Taiwan (15%), and goes up to only 22% for the top percentiles.

Panel B in Figures 3–4 shows that firms' revenue (value added) is positively correlated with the share of R&D firms, i.e., the propensity to engage in R&D increases in size. The relationship between the share of R&D firms and revenue is steeper than that between the share of R&D firms and TFP. For instance, 60% of firms in Taiwan and 50% of firms in mainland China which belong to the top revenue decile conduct R&D. If the correlation was entirely driven by high-TFP firms being larger, we should see a similar relation in panels A and B. Therefore, the steeper profile in panel B suggests that there are other factors, in addition to TFP, that drive the relationship between revenue and R&D propensity. We show below that the pattern is driven by the output wedge in a way that is consistent with our theory (i.e., the wedge is associated negatively with the propensity to invest in R&D).

Panels C and D in Figures 3–4 show the relationships between TFP growth and the distribution of initial TFP. Panel C focuses on non-R&D firms and shows that the TFP growth rate is decreasing in TFP both in Taiwan and mainland China. In other words, there is strong convergence in productivity

[14] R&D expenditure is vaguely defined in China's industrial survey. For instance, it does not distinguish between R&D performed and R&D paid for by the firm. There are also policy incentives for firms to mis-report R&D expenditure (Chen et al., 2017).

[15] The results are robust to controlling for industry fixed effects at a higher level of disaggregation, although some industries must then be dropped since the small number of firms make the estimation of the production functions susceptible to outliers.

levels across non-R&D firms. This is consistent with our theory insofar as learning through random interactions and imitation is easier for less productive firms. Taiwan and mainland China exhibit approximately the same decline in TFP growth as the intial firm TFP increases. We acknowledge that part of this decline may be attributed to survivor's bias: low-performing firms are more likely to exit causing the TFP growth of the surviving firms to be higher. We believe this problem to be especially important for low-TFP firms. This is why we trim the lower tail of the distribution, as mentioned above.

Panel D compares the TFP growth for R&D firms and non-R&D firms holding constant the initial TFP level. The figure shows that TFP growth is larger for R&D firms than for non-R&D firms in both economies. In Taiwan, this is true at all percentiles with a gap significantly increasing in the initial TFP level. In mainland China, there is no significant difference in TFP growth for the bottom 40% of the distribution, while R&D firms outperform non-R&D firms at high TFP levels. Moreover, the differences in productivity growth between R&D and non-R&D firms is significantly larger in Taiwan than in mainland China, especially at high TFP levels. This is an important observation to which we return below.

To gain a better understanding of the properties of the data, it is useful to move from simple to partial correlations. To this aim, Table 2 and 3 report the result of a set of regressions for Taiwan and mainland China, respectively. Panel A of Tables 2–3 is related to Panels A–B of Figures 2–3, while panel B of Tables 2–3 is related to panels C–D of Figures 2–3.

TABLE 2 (Panels A, B) HERE
TABLE 3 (Panels A, B) HERE

Columns (1)–(2)–(3) of panel A show the estimated coefficient (and the standard errors) of a linear probability model whose left-hand side variable is a dummy for R&D firms. All regressions use annual data and include industry fixed effects and year dummies, with standard errors clustered at the firm level. The table shows that the fraction of R&D firms is robustly correlated with TFP.[16] The positive association becomes stronger in both Taiwan and mainland China and if one controls for the estimated output wedge.[17] Both the positive correlation with TFP and the negative correlation with the output wedge are in line with the prediction of the theory. A large output wedge discourages firms from investing in R&D because the tax reduces the future expected profit. Column (3) shows that the opposite-sign property of the partial correlations is robust to controlling for a dummy for exporting firms. Since both economies are highly export-oriented, one might worry about composition effects – for instance, exporting firms being both more productive and having a high propensity to R&D. While in both economies exporting firms are significantly more likely to engage in R&D, the coefficients of interest are robust to the inclusion of export dummies. Column (4) and (5) show that the results are also robust to the inclusion of firms' fixed effects. As a firm becomes more (less) productive over time, the likelihood that it performs R&D increases (decreases).

Panel B reports the coefficient of regressions whose left-hand side variable is the TFP growth over the sample period, 1998–2004 for Taiwan and 2001-07 for mainland China (note that in this case we

[16] The regressions in panel A use annual data, and always include time dummies.

[17] The way in which we estimate output wedges is discussed in Section 4 below. In this descriptive analysis, we ignore potentially important concerns with measurement error. This is an important caveat, especially because of a potential division bias. Because of this concerns, we do not emphasize quantitative implications here. We address measurement error problems in the structural estimation.

do not use annual observations, so the number of observations is lower). TFP growth is regressed on the initial log-TFP level and on the R&D dummy, both measured in the initial period (1998 for Taiwan and 2001 for mainland China). The tables show a robust pattern by which TFP growth is correlated negatively with the initial TFP (consistent with panel C of Figures 3–4) and positively with a R&D dummy (consistent with panel D of Figures 3–4). Quantitatively, the effect of R&D on TFP growth is significantly larger in Taiwan and than in mainland China. Column (3) breaks down the R&D dummy into three dummies, one per each tercile of R&D expenditure along the intensive margin. The results of this breakdown is somewhat different in Taiwan and in mainland China. In Taiwan, all three dummies are significant and the coefficients are of similar magnitudes. Thus, the intensive margin of R&D expenditure does not appear to matter. In contrast, in China only the dummy for the top tercile is significant. Namely, only R&D firms which spend a high share of their revenue on R&D experience significant effects on their future productivity growth.

In the case of China, it also interesting to control for the ownership structure, given the importance of State-Owned Enterprises (SOE) in the Chinese economy. Controlling for SOE status does not affect any of the results in Table 3. SOE have a significantly higher propensity to engage in R&D, and a significantly lower TFP growth than other firms. The results are also robust to controlling for provincial dummies. The detailed regression results are provided in Appendix Table A.1.

The descriptive evidence presented in this section is robust. One concern is that the results may be driven by a subset of industries (e.g., semiconductors) for which R&D is especially salient. However, we find that the patterns do not change significantly if we exclude the five most R&D-intensive industries. Next, although we already showed that the results hold up when we include exporter dummies, one might still wonder whether the results hold separately for exporting and non-exporting firms. We find that they do. Finally we break down the data by regions in China.

NB: ROBUSTNESS RESULTS NOT INCLUDED IN THIS VERSION.

4 Structural Estimation

In this section, we structurally estimate a version of the model laid out in Section 2. Before describing the estimation procedure, we discuss measurement issues and the way in which we generalize the model to make its estimation viable.

We follow the method proposed by Hsieh and Klenow (2009) to retrieve the empirical distribution of TFP and of the wedges. In particular, for each firm we have information about value added, capital stock, and total labor costs. The model implies that

$$A_i \propto \frac{(P_i Y_i)^{\frac{1}{\eta-1}}}{(rK_i)^{\alpha}(wL_i)^{1-\alpha}}, \tag{8}$$

$$(1-\tau_i) \propto \frac{(P_i Y_i)^{\frac{1}{\eta-1}}}{A_i}. \tag{9}$$

We set $\eta = 5$, which corresponds to a 20% mark-up.[18] The labor output elasticity in an industry is calibrated to the average labor share in that industry over the sample period scaled by $1 - 1/\eta$. Labor

[18] This is in line with Brandt et al. (2017) and De Loecker and Warzynski (2012) who estimate firm-level markups for Chinese manufacturing firms. MMM The median value of the industry median markups is MMM 21% and 23% in 2001 and 2007, respectively. The results are robust to setting $\eta = 3$ (as in Hsieh and Klenow, 2009) and $\eta = 7$ (as in Hsieh and Song, 2015).

income is severely under-reported in China. Following Hsieh and Klenow (2009), we increase the labor share in each industry proportionally so that the aggregate labor share equals 50 percent. Likewise, we adjust Taiwanese labor shares to match the aggregate labor share of 70 percent. As a robustness check, we will also directly estimate the labor output elasticity using Olley and Pakes (1996). Capital stock is constructed by perpetual inventory method, where the depreciation rate is set equal to 5 percent, following Hsieh and Song (2015).

We assume that τ_i remains constant during the period in which we observe the firm. We then measure τ_i by the time average over the available observations. This is done in order to reduce the potential measurement error in τ_i. We generalize the the stylized model of Section 2 in two important dimensions. First, we depart from the assumption that τ_i is orthogonal to A_i, since this assumption is rejected by the data. Instead, we allow τ_i to be correlated with A_i and retrieve the correlation from the data.[19] Second, the theoretical model assumes no heterogeneity in R&D costs. However, allowing heterogeneity in R&D costs improves the fit of the model. Thus, in some specifications, we let $\bar{c}_i$ be drawn from a normal distribution assumed to be i.i.d. across firms. Finally, we assume p to be uniformly distributed in the interval $[0, \bar{p}]$.

We proceed as follows. First, we select a set of moments from the data. In particular, we consider four intervals of the distribution of TFP and value added in each of the four panels in Figures 3–4 above: the 11th-to-50th percentile, the 51th–to–80th percentile, the 81th–to–95th percentile and the 96th percentile and above.[20] This yields sixteen target moments from the data. Figure A.1 in the appendix plots the sixteen moments and their confidence intervals. Next, we estimate the parameters of the models using Simulated Method of Moments. In the simplest model, we have four parameters $\bar{c}$, $\bar{p}$, δ and q. In the model with heterogenous R&D cost, we also estimate the standard deviation of costs σ_c.

We estimate the stationary distribution of the model. Namely, we search for the parameters that minimize the distance between the targeted moments and the stationary distribution. Note that the tractability of the model is crucial for our approach. In particular, our SMM approach requires simulating the model under some parameter configuration, calculating the distance from the targeted moments. One could in principle simulate the distribution of a large number of firms for every trial of a parameter configuration. However, this would be very demanding in terms of computing time. Our system of difference equations allows us to attain the stationary distribution very quickly. Although we have no general proof of ergodicity, the numerical simulations have always converged to a unique stationary distribution irrespective of initial conditions.

We estimate parameters using the Simulated Method of Moments (see, e.g., Bloom 2009). The sample is randomly generated by bootstrapping for K times, where K is set to 500. Denote by $g_{m,k}$ the mth moment in the kth sample. We mininize the weighted sum of the distance between the empirical and simulated moments:

$$\hat{\theta} = \arg\min_{\theta} h(\theta)' W^{-1} h(\theta),$$

where $h_m(\theta) = g_m(\theta) - \frac{1}{K}\sum_k^K g_{m,k}$ and $\frac{1}{K}\sum_k^K g_{m,k}$ is the moment averaged across K samples. Denote Ω the variance-covariance matrix of the bootstrapped moments. Under the null hypothesis, Ω is

[19] In the model, it is the future wedge that determines the R&D decision. Since we estimate the model using only two observation (e.g., for Taiwan we measure TFP in 1999, and TFP growth over the period 1999-2004), we assume that firms base their R&D investment decision on the observed τ_i in the first period. The results are not sensitive to estimating the wedges using the final period. However, we prefer not to do so because of potential endogenity problems.

[20] The total employment of Taiwanese firms in each TFP interval is roughly the same. As mentioned above, we abstract from the first decile to limit the extent of the survivors' bias problem.

proportional to the variance-covariance matrix of the simulated moments. We use the identity matrix as the benchmark weighting matrix to avoid the potential small-sample bias (see, e.g., Altonji and Segal 1996). We also obtain results from using the optimal weighting matrix as a robustness check. The difference between the true and estimated parameter follows asymptotically a normal distribution with mean zero and the variance-covariance matrix of V, where $V = \left(DW^{-1}D'\right)^{-1}$ and $D = \frac{\partial h(\theta)}{\partial \theta}|_{\theta=\hat{\theta}}$. The variance of the estimated parameters are on the diagonal of V.

4.1 Results

We start by estimating the model using the Taiwanese data. This estimation will serve as a benchmark. Consider, first, the *parsimonious* model where all firms have the same cost parameter $\bar{c}_i = \bar{c}$ (see column 1 of Table 2). Figure 5 shows that the model is fairly successful in replicating the data, although both the TFP-R&D profile in panel A and the revenue-R&D profile in panel B are steeper in the model than in the data.

FIGURE 5 (Panels A, B, C, and D) HERE
FIGURE 6 (Panels A, B, C, and D) HERE
FIGURE 7 (Panels A, B, C, and D) HERE
FIGURE 8 (Panels A, B, C, and D) HERE

The fit of the model improves if we allow heterogeneous R&D costs (henceforth, the *flexible* model), as a visual inspection of Figure 6 suggests. The sum of distances is more than 80% lower (from 0.28 to 0.05). The fit of both the TFP-R&D profile in panel A and the revenue-R&D profile in panel B are visibly improved. The model with homogenous $\bar{c}$ in Figure 5 predicts too large a share of R&D firms among the high-productivity (and high-revenue) firms, and conversely too small a share of non-R&D firms among the low-productivity (and low-revenue) firms. Heterogenous R&D costs make the model consistent with less steep R&D-TFP and R&D-revenue profiles by increasing randomness in the R&D decision. The flexible model also improves slightly the fit of panel D. The estimated parameters for the parsimonious and flexible model are shown in columns (1)–(2) of Table 4. Relative to the parsimonious model, the flexible model yields a lower estimate of δ (namely, a larger opportunity cost of innovation in terms of foregone imitation) and a higher estimate of $\bar{p}$.

TABLE 4 HERE

Next, we estimate the model using the data from mainland China. Figures 7–8 and columns (3)–(4) of Table 4 report the results of the parsimonious and flexible model, respectively. Neither of the models produces nearly as good a fit as in the Taiwanese case. The estimates of the flexible model imply a very low productivity of R&D ($\bar{p} = 0.03$, while in Taiwan $\bar{p} = 0.22$), and altogether very little difference between R&D and non-R&D firms. Note, in particular that the estimated probability of passive imitation (conditional on doing R&D) is $\delta = 0.94$, indicating that there is very little difference between doing R&D and not doing R&D: if $\delta = 1$ and $\bar{p} = 0$, then R&D would have no effect whatsoever on the performance of firms. As a visual inspection of Figures 7–8 shows, even the estimated model that minimize the distance from the data misses important qualitative features. First, it fails to reproduce the positive TFP-R&D observed in the data. Second, it predicts hardly any difference in TFP growth between R&D and non-R&D firms, even at high TFP levels where in reality Chinese R&D firms outperform non-R&D firms. Another puzzling feature is that the point estimates of the technology parameters are very different between mainland China and Taiwan.

We now explore an alternative hypothesis that drastically improves the fit of the model for mainland China.

4.2 Moral hazard

A possible explanation for the low average productivity of R&D investments in China is that some firms report R&D expenditure either to conform to guidelines (e.g., in the case of SOE) or in response to fiscal incentives for R&D spending but do not actually try to innovate.

To explore this alternative hypothesis, we augment our theory with a simple model of moral hazard. We assume that firms have an incentive to misreport R&D expenditure. In particular, we assume that some firms can collect R&D subsidies and simply relabel some of their operational expenditure as R&D. These R&D subsidies are assumed to be firm specific. Although there exist statutory rules determining which firms qualify for R&D subsidies, it is well documented that (at least in the period considered by our study) Chinese firms were subject to a great deal of preferential and discretionary treatments. For instance, some firms could be invited to be part technology parks, or receive support from local authorities, etc. (see Bai, Hsieh and Song 2016).

We assume that firms can relabel non-R&D expenditures as R&D expenditures without any costs (or, alternatively, it is incorporated in the subsidy that the firm receive). Namely, firms do not face any risk of being caught if they engage in creative accounting to claim benefits and subsidies are altogether ineffective. While this is clearly an extreme assumption – some firms might face a larger probability of being audited and might therefore not find it in their interest to fake R&D expenditures – it would be difficult to identify the share of the total expenditure that each firm actually spends on R&D in response to subsidies.

We allow subsidies to be correlated with observable firm characteristics, in particular, TFP and output wedges. To this aim, we let the subsidy be denoted by

$$\sigma_i = \xi_{0i} + \xi_1 \log\left(A_i - \bar{A}\right) + \xi_2 \log\left(1 - \tau_i\right) \qquad (10)$$

where ξ_{0i} is an i.i.d. shock with mean ξ_0, while ξ_1 and ξ_2 are parameters that will be estimated from the data. The correlation is intended to capture the possibility for the government (or local authorities) to target firms with particular characteristics.

Before turning to the structural estimation, we consider again some descriptive statistics. We do not have good measurements of actual R&D subsidies at the firm level. However, we have self reported measures of whether firms do receive public subsidies. We construct a dummy that switches on for firms reporting that they receive subsidies. Table 5 reports some descriptive correlation results. Column (1)–(2) show that the subsidy dummy is a predictor of R&D investments even after controlling for all other variables that have been shown to correlate with propensity to R&D. Columns (3)–(4) regress instead the subsidy dummy on a set of variables, showing in particular that subsidies are positive correlated with TFP and negatively correlated with the output wedge (i.e., positively correlated with $1-\tau_i$). While the subsidy measure is problematic (and for this reason we do not use it in the structural estimation), we find these correlations interesting and worth reporting.

TABLE 5 HERE

Consider, next, the structural model. Moral hazard drives a wedge between observed and actual R&D expenditure: firms' self-reported R&D does not correspond to an actual investment in R&D when the model predicts that the firm prefers to imitate rather than innovate. In line with the model, we assume that the data contain information about reported R&D, whereas actual R&D is not observed.

TABLE 6 HERE

Table 6 reports the estimates of the model with fake R&D, given the process for firm-specific subsidies (10). Columns (1)–(2) are for Taiwan, while columns (3)–(4)–(5) are for mainland China. Columns (1) and (3) reproduce the results of the flexible model for comparison. In the case of Taiwan, the extended model with fake R&D adds no explanatory power relative to the model without fake R&D. All the coefficients of the moral hazard models are insignificantly different from zero. In contrast, when estimated on Chinese data, the model with fake R&D improves the fit significantly. The improvement can be seen in three dimensions. First, the sum of distances decreases from 0.21 to 0.16. Second, both ξ_1 and ξ_2 are estimated to be negative, in line with the discussion of the descriptive evidence above. Namely, government subsidies appear to target high-productivity firms and well-connected firms (i.e., firms with low or negative output wedges). Last, but not least important, the estimated parameters q, δ, $\bar{p}$ are now very similar to those estimated for Taiwan. In column (5), we constrain these three parameters to take on the same value as the point estimate for Taiwan. This implies that the model has now about the same number of degrees of freedom as the flexible model of column (3). The sum of distances in column (5) is still 19% smaller than in the flexible model without fake R&D in column (3).

FIGURE 9 (PANEL A, B, C, D) HERE

Figure 9 shows visually how the model fits the target moments with the aid of the usual four graphs. The fit is very good for panels A, C, and D. Note that each panel now displays three schedules: (i) the moments in the data, (ii) the fit of the model (which refers to measured R&D), and (iii) the results restricted to the firm which, according to the model, do real R&D expenditure (as opposed to report the results of creative accounting). This takes advantage of the fact that the estimated model allows us to identify the firms that really do R&D. According to our estimates a mere 1% of the Chinese firms engage in true R&D. However, the productivity of R&D investments is very high. In fact, the productivity of R&D investment is as high as in Taiwan, although the average cost of R&D is higher. Figure 9 provides the breakdown between firms that do and firms that fake R&D. Only firms with very high TFP (and very high output wedges) do real R&D. For high-TFP firms, the productivity growth difference relative to non-R&D firms is 10%, which is comparable to what is observed in Taiwan. This is consistent with the casual observation that a China has a small number of very productive firms (e.g., Huawei) that are highly innovative and successful internationally.

We acknowledge that our analysis is subject to an important *caveat*. The extreme assumption that subsidies are totally ineffective (and that R&D is a binary decision at the firm level) may lead to overestimate the extent of fake R&D. In reality, firms may both relabel expenditure and do some real R&D investments. Therefore, our estimates provide an upper bound to fake R&D. The reality is likely to lie in between the predictions of the models in column (3) (no fake R&D, many firms doing R&D, and a low productivity of R&D investment) and in columns (4)–(5) of Table 6 (some fake R&D, only few firms doing real R&D, and productivity of the true R&D investments being high). Either way, our analysis provides a warning against taking at face value the boom in aggregate R&D expenditure in China.

4.3 Patents

Our analysis focuses on investment in R&D. A common measure of the success of innovative activity is patents. If some firms fudge R&D to cash subsidies, it is plausible to expect that most such firms will not patent innovations. If this assumption is correct, we should expect that R&D firms which are granted one or more patents experience on average higher productivity growth than the rest of R&D firms.

To test this prediction, we collected data for all the patents approved by the State Intellectual Property Office (SIPO) in 2001-07. The approval rate for domestic applicants is about 40%. We match the SIPO data with the 2007 NBS data. We find that 7,558 NBS firms (out of a total of 300,000) and 3,060 firms in the balanced panel have one or more patents for which they applied in the period 2001-2007. Interestingly, the firms with a positive number of patents that belong to the balanced panel have applied on average for fifteen patents during the sample period. This is a fairly large number.

FIGURE 10 HERE

Consider Figure 10. Panel A shows that the propensity to patenting innovation is sharply increasing in the initial TFP level. Panel B plots the average number of patents as a function of TFP broken down by R&D and non-R&D firms. Not surprisingly, almost all patents come from firms reporting R&D activities. The number of patents increase sharply from the 81st to the 96th percentile. The same pattern emerges from panel C which plots the proportion of R&D firms and the proportion for non-R&D firms with one or more patent. The gap is increasing in the TFP level.

Panel D displays the most interesting finding. Firms with a positive number of patents exhibit a significant growth difference (ca. 5 percentage points) relative to firms that do not perform R&D. For high-TFP firms, the gap is even larger. Non-patenting R&D firms outperform non-R&D firms by a much smaller difference, about 2%. This evidence is consistent with the hypothesis that part of the R&D firms that do not patent innovations are in reality R&D fudgers. This might explain why on average R&D firms experience a smaller productivity growth than Taiwanese firms. However, a *caveat* is in order. Some of the non-patenting R&D firms may have actually truthfully invested in R&D, and the failure to patent any innovation may reflect bad luck of low skill. So, identifying fudgers with non-patenting R&D firms may exaggerate the effect we aim to identify.

MORE ROBUSTNESS ANALYSIS TO BE DONE

5 Conclusions

[VERY PRELIMINARY AND INCOMPLETE] We constructed and estimated a dynamic model where firms are heterogenous in productivity and are subject to a variety distortions. We estimated the model using data from Taiwan and mainland China. The data conform with the main predictions of the theory. Using Taiwan as a benchmark to asses China's R&D investments, we conclude that the evidence is consistent with a significant extent of fudging and overreporting of R&D in China.

6 References

Acemoglu, Daron, Philippe Aghion, Rachel Griffith, and Fabrizio Zilibotti (2010). "Vertical integration and technology: theory and evidence," *Journal of the European Economic Association* 8(5), 989-1033.

Acemoglu, Daron,Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr (2017). "Innovation, Reallocation, and Growth," NBER Working Paper No. 18993.

Acemoglu, Daron, and Dan Cao (2015). "Innovation by entrants and incumbents." *Journal of Economic Theory* 157, 255-294.

Acemoglu, Daron and Fabrizio Zilibotti (2001). "Productivity differences," *The Quarterly Journal of Economics* 116(2), 563-606.

Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti (2006). "Distance to frontier, selection, and economic growth," *Journal of the European Economic Association* 4(1), 37-74

Akcigit, Ufuk, and William R. Kerr (2017). "Growth through Heterogeneous Innovations," *mimeo, University of Chicago.*

Akcigit, Ufuk, Santiago Caicedo Soler, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi (2018). "Dancing with the Stars: Innovation through Interactions." NBER Working Paper 24466.

Aghion, Philippe, and Peter Howitt (1992). "A Model of growth through creative destruction ," *Econometrica* 60(2), 323-351.

Altonji, Joseph G., and Lewis M. Segal (1996). "Small-sample bias in GMM estimation of covariance structures," *Journal of Business and Economic Statistics* 14(3), 353-366.

Aw, Bee Yan, Mark J. Roberts, and Daniel Yi Xu (2011). "R&D investment, exporting, and productivity dynamics," *American Economic Review* 101(4), 1312-44.

Bai, Chong-En, Chang-Tai Hsieh, and Zheng Michael Song (2016). "The Long Shadow of China's Fiscal Expansion," *Brookings Papers on Economic Activity* 2016(2), 129-181.

Benhabib, Jess, Jesse Perla, and Christopher Tonetti (2014). "Catch-up and fall-back through innovation and imitation," *Journal of Economic Growth* 19(1), 1-35.

Benhabib, Jess, Jesse Perla, and Christopher Tonetti (2017). "Reconciling models of diffusion and innovation: a theory of the productivity distribution and technology frontier," *NBER Working Paper* No. w23095.

Bloom, Nicholas (2009). "The impact of uncertainty shocks," *Econometrica* 77(3), 623-685.

Bloom, Nicholas, Rachel Griffith and John Van Reenen (2002). "Do R&D tax credits work?Evidence from a panel of countries 1979-1997," *Journal of Public Economics* 85, 1-31.

Brandt, Loren, Gueorgui Kambourov, and Kjetil Storesletten (2016). "Firm Entry and Regional Growth Disparities: the Effect of SOEs in China ," *mimeo, University of Toronto.*

Chen, Zhao, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi Xu (2017). "Notching R&D Investment with Corporate Income Tax Cuts in China," *mimeo, Duke University.*

Cheremukhin, Anton, Mikhail Golosov, Sergei Guriev, and Aleh Tsyvinski. (2017). "The economy of People[redacted]s Republic of China from 1953," *mimeo, Yale University.*

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

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status," *American Economic Review* 102 (6), 2437-71.

Ding, Xuedong and Jun Li (2015). "Incentives for innovation in China: building an innovative economy," *Routledge.*

Gabaix, Xavier (2009). "Power laws in economics and finance," *Annual Review of Economics* 1(1),255-294.

Garcia-Macia, Daniel, Chang-Tai Hsieh, and Peter J. Klenow (2016). "How destructive is innovation?," *NBER Working Paper* No. w22953.

Ghiglino, Christian (2012). "Random walk to innovation: Why productivity follows a power law," *Journal of Economic Theory* 147(2), 713-737.

Griliches, Zvi (1957). "Hybrid corn: An exploration in the economics of technological change," *Econometrica* 501-522.

Griliches, Zvi (1998). "Patent statistics as economic indicators: a survey," *R&D and productivity: the econometric evidence, University of Chicago Press* 287-343.

Hall, Robert E., and Charles I. Jones (1999). "Why do some countries produce so much more output per worker than others?" *Quarterly Journal of Economics* 114(1), 83-116.

Hall, Robert E., and Dale W. Jorgenson (1967). "Tax Policy and Investment Behavior," *American Economic Review* 57(3), 391-414.

Hsieh, Chang-tai and Peter J. Klenow (2009). "Misallocation and manufacturing TFP in China and India," *Quarterly Journal of Economics* 124(4), 1403-1448.

Hsieh, Chang-Tai, and Peter J. Klenow (2010). "Development accounting," *American Economic Journal: Macroeconomics* 2(1),207-23.

Hsieh, Chang-Tai, and Zheng Michael Song (2015). "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China," *rookings Papers on Economic Activity* 2015(1),295-366.

Hu, Albert Guangzhou, and Gary H. Jefferson (2009). "A great wall of patents: What is behind China's recent patent explosion?" *Journal of Development Economics* 90(1),57-68.

Jia, Junxue, and Guangrong Ma (2017). "Do R&D tax incentives work? Firm-level evidence from China," *China Economic Review* 46, 50-66.

Jovanovic, Boyan, and Rafael Rob (1989). "The growth and diffusion of knowledge," *Review of Economic Studies* 56(4),569-582.

Kesten, Harry (1973). "Random difference equations and renewal theory for products of random matrices," *Acta Mathematica* 131(1), 207-248.

Klenow, Peter J. and Andrés Rodríguez-Clare (1997). "The neoclassical revival in growth economics: Has it gone too far? " *NBER Macroeconomics Annual* 12, 73-103.

Klette, Tor Jakob, and Samuel Kortum (2004). "Innovating firms and aggregate innovation," *Journal of Political Economy* 112(5), 986-1018.

Konig, Michael D., Jan Lorenz, and Fabrizio Zilibotti (2016). "Innovation vs. imitation and the evolution of productivity distributions," *Theoretical Economics* 11(3), 1053-1102.

Lentz, Rasmus, and Dale T. Mortensen (2008). "An empirical model of growth through product innovation," *Econometrica* 76(6), 1317-1373.

Lichtenberg, Frank R (1992). “R&D investment and international productivity differences,” *NBER Working Paper* No. w4161.

Lucas Jr, Robert E., and Benjamin Moll (2014). “Knowledge growth and the allocation of time,” *Journal of Political Economy* 122(1), 1-51.

Luttmer, Erzo GJ (2007). “Selection, growth, and the size distribution of firms,” *Quarterly Journal of Economics* 122(3), 1103-1144.

Luttmer, Erzo GJ (2012). “Technology diffusion and growth,” *Journal of Economic Theory* 147(2), 602-622.

Olley, Steven, and Ariel Pakes (1996). “The dynamics of productivity in the telecomunications equipment industry,” *Econometrica* 64, 1263-97.

Perla, Jesse, and Christopher Tonetti (2014). “Equilibrium imitation and growth,” *Journal of Political Economy* 122(1), 52-76.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti (2011). “Growing like China,” *American Economic Review* 101(1), 196-233.

Tombe, Trevor, and Xiaodong Zhu (2017). “Trade, migration and productivity: A quantitative analysis of China,” *mimeo, University of Toronto.*

Gabaix, Xavier (1999).“Zipf’s law for cities: an explanation,” *Quarterly Journal of Economics* 114(3), 739-767.

Zilibotti, Fabrizio (2017). “Growing and Slowing Down Like China,” *Journal of the European Economic Association* 15(5), 943-988.

Technical Appendix

This section is preliminary and incomplete.

Proof of Proposition 1.

From the monotonicity of $Q(a,\tau;P)$ it follows that there exists a threshold function $a^*(\tau;P)$ such that

$$\begin{array}{c} Q(a,\tau;P) \geq \overline{p} \text{ if } a \leq a^*(\tau;P), \\ Q(a,\tau;P) < \overline{p} \text{ if } a > a^*(\tau;P). \end{array} \tag{11}$$

To economize in notation, in this proof we write $a^*(t) = a^*(\tau;P)$ and $p(a(t)) = Q(a,\tau;P)$ when this is no source of confusion (recall that τ plays no role in this proposition).

The difference equation for the updates of the log-productivity distribution can then be broken down as follows:

$$\begin{aligned} \dot{P}_a(t) &= \lim_{\Delta t \to 0} P_a(t+\Delta t) - P_a(t) \\ &= \begin{cases} q\left[(1-F_{a-1}(t))\,P_{a-1}(t) - (1-F_a(t))P_a(t)\right] & \text{if} \quad a < a^*(t), \\ \begin{pmatrix} q\,(1-F_{a-1}(t))\,P_{a-1}(t) \\ -G(p(a))\left[q(1-F_a(t))P_a(t)\right] \\ -\int_{p(a(t))}^{\overline{p}}\left[(p+(1-p)\,\delta q\,(1-F_a(t)))\,P_a(t)\right]dG(p) \end{pmatrix} & \text{if} \quad a = a^*(t)+1, \\ \begin{pmatrix} G(p(a-1)) \times q\,(1-F_{a-1}(t))\,P_{a-1}(t) \\ +\int_{p(a(t)-1)}^{\overline{p}}\left[(p+(1-p)\,\delta q\,(1-F_{a-1}(t)))\,P_{a-1}(t)\right]dG(p) \\ -G(p(a)) \times q\,(1-F_a(t))\,P_a(t) \\ -\int_{p(a(t))}^{\overline{p}}\left[(p+(1-p)\,\delta q\,(1-F_a(t)))\,P_a(t)\right]dG(p) \end{pmatrix} & \text{if} \quad a > a^*(t)+1. \end{cases} \end{aligned} \tag{12}$$

Intuitively, if $a < a^*(t)$, all firms with productivity a and $a-1$ imitate; if $a > a^*(t)+1$, all firms with productivity a facing a realization $p > p(a(t))$ and all firms with productivity $a-1$ facing a realization $p > p(a(t)-1)$ innovate, while all other firms with productivity a and $a-1$ imitate; if $a = a^*(t)+1$, all firms with productivity a facing a realization $p > p(a(t))$ innovate, and all other firms with productivity a and $a-1$ imitate.

Going from the p.d.f to the corresponding c.d.f. yields:

$$\begin{aligned} \dot{F}_a(t) &= \sum_{j=1}^{a} \dot{P}_j(t) \\ &= \begin{cases} -q\,(1-F_a(t))\,(F_a(t)-F_{a-1}(t)) & \text{if} \quad a \leq a^*(t) \\ \begin{pmatrix} -G(p(a(t)))\,q\,(1-F_a(t))\,(F_a(t)-F_{a-1}(t)) \\ -\int_{p(a(t))}^{\overline{p}}\begin{bmatrix} (p+(1-p)\,\delta q\,(1-F_a(t)))\times \\ (F_a(t)-F_{a-1}(t)) \end{bmatrix} dG(p) \end{pmatrix} & \text{if} \quad a > a^*(t) \end{cases} \end{aligned} \tag{13}$$

For future reference, we define the complementary cumulative distribution function $H_a(t) = 1 -$

$F_a(t)$. Equation (13) can alternatively be written as:

$$\dot{H}_a(t) = \sum_{j=1}^{a}\left(1-\dot{P}_j\right)(t)$$
$$= \begin{cases} qH_a(t)\left(H_a(t)-H_{a-1}(t)\right) & \text{if} \quad a \le a^*(t) \\ \left(\begin{array}{c} G\left(p\left(a\left(t\right)\right)\right)qH_a\left(t\right)\left(H_a\left(t\right)-H_{a-1}\left(t\right)\right) \\ +\int_{p(a(t))}^{\overline{p}} \left[\begin{array}{c} \left(p+\left(1-p\right)\delta qH_a\left(t\right)\right)\times \\ \left(H_a\left(t\right)-H_{a-1}\left(t\right)\right) \end{array} \right] dG\left(p\right) \end{array} \right) & \text{if} \quad a > a^*(t) \end{cases} \tag{14}$$

Note that $\dot{F}_a(t) \le 0$ and $\dot{H}_a(t) \ge 0$. Then, it follows from the argument in Bramson (1983) and König, Lorenz, and Zilibotti (2016) that there exists a travelling wave solution of the form $F_a(t) = \tilde{f}(a-\nu t)$ (or, equivalently $H_a(t) = \tilde{h}(a-\nu t)$) with velocity $\nu > 0$. Differentiating yields $\dot{F}_a(t) = -\nu\tilde{f}'$. Therefore, having defined $x = a - \nu t$, we can rewrite (13) as:

$$-\nu\tilde{f}'(x)$$
$$= \begin{cases} -q\left(1-\tilde{f}(x)\right)\left(\tilde{f}(x)-\tilde{f}(x-1)\right) & \text{if} \quad x \le x^* \\ \left(\begin{array}{c} -G\left(p\left(x\right)\right)\left[q(1-\tilde{f}(x))\left(\tilde{f}(x)-\tilde{f}(x-1)\right)\right] \\ -\int_{p(x)}^{\overline{p}} \left[\begin{array}{c} \left(p+\left(1-p\right)\delta q\left(1-\tilde{f}(x)\right)\right)\times \\ \left(\tilde{f}(x)-\tilde{f}(x-1)\right) \end{array} \right] dG\left(p\right) \end{array} \right) & \text{if} \quad x > x^* \end{cases} \tag{15}$$

or, identically,

$$-\nu\tilde{h}'(x)$$
$$= \begin{cases} q\tilde{h}(x)\left(\tilde{h}(x)-\tilde{h}(x-1)\right) & \text{if} \quad x \le x^* \\ \left(\begin{array}{c} G\left(p\left(x\right)\right)\left[q\tilde{h}(x)\left(\tilde{h}(x)-\tilde{h}(x-1)\right)\right] \\ +\int_{p(x)}^{\overline{p}} \left[\begin{array}{c} \left(p+\left(1-p\right)\delta q\tilde{h}(x)\right)\times \\ \left(\tilde{h}(x)-\tilde{h}(x-1)\right) \end{array} \right] dG\left(p\right) \end{array} \right) & \text{if} \quad x > x^* \end{cases}, \tag{16}$$

Consider, first, the range $x \le x^*$. Using the upper part of (15) yields the following Delay Differential Equation (DDE):[21]

$$-\nu\tilde{f}'(x) = -q\left(1-\tilde{f}(x)\right)\left(\tilde{f}(x)-\tilde{f}(x-1)\right). \tag{17}$$

This equation allows us to characterize the (asymptotic) left tail of the distribution. Taking the limit for $x \to -\infty$, we can make the following linear approximation that ignores second order terms:

$$-\nu\tilde{f}'(x) \simeq q\left(\tilde{f}(x)-\tilde{f}(x-1)\right).$$

We guess that this linear DDE has a solution of the form $\tilde{f}(x) = c_1 e^{\lambda x}$ for for $x \to -\infty$. Replacing $\tilde{f}(x)$ by its guess and $\tilde{f}'(x)$ by its derivative, and simplifying terms, allows us to verify the guessed

[21] See also Asl and Ulsoy (2003); Bellman and Cooke (1963); Smith (2010).

solution as long as the following transcendental equation in λ is satisfied:

$$\lambda\nu \simeq q(1-e^{-\lambda}). \tag{18}$$

The solution to this transcendental equation is given by

$$\lambda = \frac{\nu W\left(-\frac{qe^{-\frac{q}{\nu}}}{\nu}\right)+q}{\nu},$$

where W denotes the Lambert W-function, and we require that $\frac{qe^{-\frac{q}{\nu}}}{\nu} \leq \frac{1}{e}$.

Consider, next, the range of large x where the solution for $x > x^*$ applies in (16). Then, we can write the following DDE

$$-\nu\tilde{h}'(x) = \begin{pmatrix} G(p(x))\left[q\tilde{h}(x)\left(\tilde{h}(x)-\tilde{h}(x-1)\right)\right] \\ +\int_{p(x)}^{\overline{p}}\begin{bmatrix}\left(p+(1-p)\,\delta q\tilde{h}(x)\right)\times \\ \left(\tilde{h}(x)-\tilde{h}(x-1)\right)\end{bmatrix} dG(p)\end{pmatrix}. \tag{19}$$

We use this DDE to characterize the right tail of the distribution as $x \to +\infty$. In this limit, we can make a linear approximation ignoring second-order terms:

$$-\nu\tilde{h}'(x) \simeq \hat{p}\left(\tilde{h}(x)-\tilde{h}(x-1)\right),$$

where $\hat{p} = \int_0^{\overline{p}} p\ dG(p)$. Note that $\lim_{x\to\infty} p(x) = 0$ since for x arbitrarily large imitation becomes totally ineffective and firms choose almost surely to innovate. We guess a solution of the DDE of the form $\tilde{h}(x) = c_2 e^{-\rho x}$ for $x \to +\infty$. The guess is verified as long as the following transcendental equation holds:

$$\rho\nu \simeq \hat{p}(1-e^{-\rho}).$$

The solution to the transcendental equation satisfies

$$\rho = \frac{-\nu W\left(-\frac{\hat{p}e^{-\frac{\hat{p}}{\nu}}}{\nu}\right)-\hat{p}}{\nu},$$

where W denotes the Lambert W-function, and we require that $\frac{\hat{p}e^{-\frac{\rho}{\nu}}}{\nu} \leq \frac{1}{e}$.

This concludes the proof. ■

Remark 1 *Consider the model of innovation-imitation of Proposition 1. Assume $p_i = p$ for all i (no ex-ante heterogeneity in in-house R&D capability) and $\delta = 0$ (firms pursuing innovation cannot imitate). Let $\Delta t \to 0$ and assume that $q > p$. Then there exists a travelling wave solution of the form $P_a(t) = f(a-\nu t)$ with velocity $\nu > 0$, with left and right Pareto tails, characterized as follows: (i) The right power tail exponent ρ can be approximated by the root of the equation $(q-p)e^{\frac{q\rho}{p-q}} + pe^{\rho} - q = 0$ (where a real root is guaranteed to exists if $q > p$), while the left tail exponent λ can be approximated by $\lambda \simeq q\rho/(q-p)$; and (iii) the traveling wave velocity is approximately given by $\nu \simeq p(e^{\rho}-1)/\rho$.*

Proof of Remark 1. When $\delta = 0$ and $p_i = p$ for all i, we can write the DDEs (17) and (19) in the following compact form:

$$-\nu \tilde{h}'(x) = \begin{cases} q\tilde{h}(x)(\tilde{h}(x-1) - \tilde{h}(x)), & \text{if } x \leq x^*, \\ p(\tilde{h}(x-1) - \tilde{h}(x)), & \text{if } x > x^*. \end{cases}$$

Consider, first, the case $x \leq x^*$. The definition of $\tilde{h}$ as a complementary c.d.f implies the following boundary condition: $\lim_{x \to -\infty} \tilde{h}(x) = 1$. The solution to this DDE can be expressed as (cf. Asl and Ulsoy 2003):

$$\tilde{h}(x) = -\sum_{k=-\infty}^{\infty} \alpha_k e^{\lambda_k x}.$$

The boundary condition implies that $\lambda_0 = 0$ and $\alpha_0 = -1$, so that we can write

$$\tilde{h}(x) = 1 - \sum_{k \neq 0} \alpha_k e^{\lambda_k x}.$$

Taking only the dominant term (and denoting it by $\lambda_1 = \lambda$; we also set $\alpha_1 = \alpha$), we can write the following approximation

$$\tilde{h}(x) \sim 1 - \alpha e^{\lambda x},$$

were the exponent λ is given by Equation (18). Denote by $\tilde{\pi}$ the associated probability mass function (pmf). Then,

$$\tilde{\pi}(x) = \tilde{h}(x-1) - \tilde{h}(x) \sim 1 - \alpha e^{\lambda(x-1)} - (1 - \alpha e^{\lambda x}) = \alpha e^{\lambda x}(1 - e^{-\lambda}).$$

Next, consider the range $x > x^*$. There,

$$\nu \tilde{h}'((x) = p(\tilde{h}(x) - \tilde{h}(x-1)),$$

with the solution

$$\tilde{h}(x) = \sum_{k=-\infty}^{\infty} \beta_k e^{-\rho_k x}.$$

The definition of $\tilde{h}$ as a complementary c.d.f implies the following boundary condition: $\lim_{x \to \infty} \tilde{h}(x) = 0$. By the same procedure as in the other case, we can write the following approximation:

$$\tilde{h}(x) \sim \beta e^{-\rho x},$$

where the exponent ρ is given by Equation (**??**). The corresponding pmf $\tilde{\pi}$ is given by

$$\tilde{\pi}(x) = \tilde{h}(x-1) - \tilde{h}(x) \sim \beta e^{-\rho(x-1)} - \beta e^{-\rho x} = \beta e^{-\rho x}(e^{\rho} - 1).$$

Next, requiring continuity of the pmf at the threshold $x = x^*$ yields

$$\alpha(1 - e^{-\lambda}) = \beta(e^{\rho} - 1).$$

Solving for β yields $\beta = \frac{1-e^{-\lambda}}{e^{\rho}-1}\alpha$. Inserting it into the equation for π yields $\tilde{\pi}(x) = (1 - e^{-\lambda})\alpha e^{-\rho x}$ for $x > x^*$.

At the threshold $x = x^*$, the expected profits from innovation and imitation must be the same. Thus setting $\chi^{\text{im}}(a, P) = \chi^{\text{in}}(a, P)$ in Equation (4) implies that

$$p = q\tilde{h}(0) = q(1-\alpha).$$

Solving for α yields

$$\alpha = \frac{q-p}{q}.$$

Hence, we can write

$$\tilde{h}(x) \sim \begin{cases} 1 - \frac{q-p}{q} e^{\lambda x}, & \text{if } x \leq x^*, \\ \frac{q-p}{q} \frac{1-e^{-\lambda}}{e^{\rho}-1} e^{-\rho x}, & \text{if } x > x^*, \end{cases}$$

and

$$\tilde{\pi}(x) \sim \begin{cases} \frac{q-p}{q}(1-e^{-\lambda}) e^{\lambda x}, & \text{if } x \leq x^*, \\ \frac{q-p}{q}(1-e^{-\lambda}) e^{-\rho x}, & \text{if } x > x^*. \end{cases}$$

Furthermore, the properties of the pmf function $\tilde{\pi}$ require that

$$\begin{aligned} 1 = \sum_{x=-\infty}^{\infty} \tilde{\pi}(x) &= \frac{q-p}{q}(1-e^{-\lambda}) \left(\sum_{x=-\infty}^{0} e^{\lambda x} + \sum_{x=1}^{\infty} e^{-\rho x} \right) \\ &= \frac{q-p}{q} \frac{\left(e^{\rho} - e^{-\lambda}\right)}{e^{\rho} - 1}. \end{aligned}$$

Next, we know that the right-tail exponent satisfies $\nu\rho = p(e^{\rho} - 1)$, while the left-tail exponent satisfies $\lambda\nu = q(1 - e^{-\lambda})$. We then can write

$$\begin{aligned} 1 = \frac{q-p}{q} \frac{\left(e^{\rho} - 1 + 1 - e^{-\lambda}\right)}{e^{\rho} - 1} &= \frac{q-p}{q} \frac{\left(\frac{\nu\rho}{p} + \frac{\lambda\nu}{q}\right)}{\frac{\nu\rho}{p}} \\ &= \frac{q-p}{q} \left(1 + \frac{p\lambda}{q\rho}\right). \end{aligned}$$

Further, we have that

$$\frac{p(e^{\rho} - 1)}{\rho} = \frac{q(1 - e^{-\lambda})}{\lambda}.$$

Hence, we have to solve a system of two unknowns, λ, ρ, given by

$$\begin{aligned} \frac{\lambda p}{q\rho} &= \frac{1 - e^{-\lambda}}{e^{\rho} - 1}, \\ 1 + \frac{\lambda p}{q\rho} &= \frac{q}{q-p}. \end{aligned}$$

Solving the second equation for λ gives

$$\lambda = \frac{q\rho}{q-p},$$

and inserting into the first yields a nonlinear equation for ρ given by

$$(q-p)e^{\frac{q\rho}{p-q}} = q - pe^{\rho}. \tag{20}$$

■

Proof of Proposition 2. TO BE ADDED ■

Additional References for the Proof Section

Asl, Farshid Maghami, and Galip Ulsoy (2003) “Analysis of a system of linear delay differential equations,” *Journal Economic Dynamics and Control* 125(2), 215–223.

Bellman, Richard, and Kenneth Cooke (1963). *Differential-difference equations.*

Bramson, Maury (1983). “Convergence of solutions of the Kolmogorov equation to traveling waves. *Memoirs of the American Mathematical Society* 44, 285.

Smith, Hal (2010). *An introduction to delay differential equations with applications to the life sciences*. Springer Science & Business Media.

Table 1: Summary Statistics

	Number of Firms	Number of R&D Firms	Median Revenue (million USD)	Mean Revenue (million USD)	Median R&D Intensity (%)	Mean R&D Intensity (%)
	Balanced Panel of Chinese Firms					
2001	70340	11916	2.39	9.75	0.31	0.90
2007	70340	12508	5.96	30.45	0.45	1.46
	Private Chinese Firms in the Balanced Panel					
2001	60022	8595	2.27	7.19	0.32	0.89
2007	63178	9893	5.82	22.45	0.41	1.35
	Balanced Panel of Taiwanese Firms					
1999	11229	1487	1.66	7.82	1.14	2.80
2004	11229	1144	2.01	11.59	1.07	2.22

Note: R&D intensity is the ratio of R&D expenditure to revenue. Median and mean R&D intensity is the median and mean value of R&D intensity among R&D firms.

Table 2: Balanced Panel of Taiwanese Firms, 1999-2004

Panel A: Correlations between Firm Characteristics and R&D Decision

	(1)	(2)	(3)	(4)	(5)
	rd_d	rd_d	rd_d	rd_d	rd_d
log_tfp	0.0933***	0.921***	0.854***	0.00458***	0.0941***
	(0.00341)	(0.0207)	(0.0205)	(0.00135)	(0.0161)
wedge		-1.024***	-0.951***		-0.103***
		(0.0245)	(0.0242)		(0.0184)
ex_d			0.0916***		
			(0.00705)		
Year effects	+	+	+	+	+
Industry effects	+	+	+	-	-
Firm effects	-	-	-	+	+
Observations	44,326	44,326	44,326	44,326	44,326
R-squared	0.200	0.386	0.395	0.009	0.010
Number of id				11,118	11,118

Panel B: Correlations between Firm Initial Characteristics and TFP Growth

	(1)	(2)	(3)
	tfp_g	tfp_g	tfp_g
log_tfp	-0.0659***	-0.0680***	-0.0697***
	(0.00487)	(0.00491)	(0.00581)
rd_d	0.116***	0.105***	
	(0.0225)	(0.0230)	
rd_s_h			0.0701**
			(0.0294)
rd_s_m			0.122***
			(0.0356)
rd_s_l			0.117***
			(0.0172)
ex_d		0.0397***	0.0396***
		(0.00563)	(0.00543)
Year effects	+	+	+
Industry effects	+	+	+
Observations	9,996	9,996	9,996
R-squared	0.066	0.069	0.070

Note: We drop firms with TFP in the bottom 10 percentiles. (1) to (3) in Panel A are pooled regressions, while (4) and (5) are the regressions with firm fixed effect. The independent variable is rd_d, a dummy variable for R&D that equals one if firm R&D expenditure is positive and zero otherwise. log_tfp and log_va are log TFP and log value added, respectively. Wedge refers to the calibrated firm output wedge (see Section XXX for details). ex_d is a dummy variable for exports. rd_s_h is a dummy variable for high R&D intensity, which equals one if firm R&D expenditure over sales is in the 67th percentile and above (among all R&D firms). rd_s_m is a dummy variable for medium R&D intensity, which equals one if firm R&D expenditure over sales is between the 33th and 66th percentiles. rd_s_l is a dummy variable for low R&D intensity, which equals one if firm R&D expenditure over sales is positive and below the 33th percentile. The dependent variable in Panel B is tfp_g, the annualized firm TFP growth over the period of the balanced panel. We use the initial-period value for all the right-hand-side variables. Standard errors are reported in parenthesis. For the regressions without firm fixed effects, observations are weighted by employment and standard errors are clustered by firm (Panel A) or industry (Panel B).

Table 3: Balanced Panel of Chinese Firms, 2001-2007

Panel A: Correlations between Firm Characteristics and R&D Decision

	(1)	(2)	(3)	(4)	(5)
	rd_d	rd_d	rd_d	rd_d	rd_d
log_tfp	0.0347***	0.525***	0.499***	0.00136*	0.0294***
	(0.00153)	(0.00724)	(0.00731)	(0.000764)	(0.00371)
wedge		-0.587***	-0.557***		-0.0331***
		(0.00822)	(0.00830)		(0.00430)
ex_d			0.0437***		
			(0.00260)		
Year effects	+	+	+	+	+
Industry effects	+	+	+	-	-
Firm effects	-	-	-	+	+
Observations	441,064	441,064	441,064	441,064	441,064
R-squared	0.364	0.429	0.430	0.582	0.582
Number of id				70,268	70,268

Panel B: Correlations between Firm Initial Characteristics and TFP Growth

	(1)	(2)	(3)
	tfp_g	tfp_g	tfp_g
log_tfp	-0.0794***	-0.0792***	-0.0795***
	(0.00173)	(0.00175)	(0.00180)
rd_d	0.00925**	0.0112**	
	(0.00432)	(0.00431)	
rd_s_h			0.0235***
			(0.00622)
rd_s_m			0.00878
			(0.00611)
rd_s_l			-0.000787
			(0.00612)
ex_d		-0.0117**	-0.0124**
		(0.00483)	(0.00481)
Year effects	+	+	+
Industry effects	+	+	+
Observations	62,598	62,598	62,598
R-squared	0.113	0.114	0.115

Note: See Table 2 for the variable definitions and regression specifications.

Table 4: Structural Estimation (Simple and Flexible Models)

	(1)	(2)	(3)	(4)
	Taiwan		China	
	Simple Model	Flexible Model	Simple Model	Flexible Model
q	0.366	0.374	0.198	0.259
	(0.011)	(0.011)	(0.003)	(0.005)
delta	0.979	0.788	1.121	0.944
	(0.026)	(0.035)	(0.002)	(0.022)
p_bar	0.199	0.219	0.013	0.031
	(0.006)	(0.010)	(0.000)	(0.006)
c_bar	0.701	1.452	0.844	6.083
	(0.105)	(0.092)	(0.013)	(1.095)
sigma_c		0.695		1.596
		(0.008)		(0.046)
Sum of Distances	0.280	0.049	0.495	0.209

Note: Standard errors are reported in the parenthesis. The flexible model refers to the model that allows heterogeneous R&D costs.

Table 5: Correlations between Subsidy, Firm Characteristics and R&D Decision

VARIABLES	(1) rd_d	(2) rd_d	(3) subsidy_d	(4) subsidy_d
log_tfp	0.0332***	0.0618***	0.00896***	0.0480***
	(0.00143)	(0.00158)	(0.00171)	(0.00191)
subsidy_d	0.167***	0.121***		
	(0.00346)	(0.00294)		
wedge_ave		-0.0370***		-0.0611***
		(0.00222)		(0.00263)
ex_d		0.0922***		0.118***
		(0.00267)		(0.00320)
soe		0.225***		0.139***
		(0.00449)		(0.00564)
Year effects	+	+	+	+
Industry effects	+	+	+	+
Observations	441,064	441,064	441,064	441,064
R-squared	0.385	0.424	0.014	0.056

Note: The dependent variable in (1) and (2) is the dummy variable for R&D. subsidy_d is a dummy variable that equals one for firms reporting positive subsidy and zero otherwise. wedge_ave is the average log TFPR (see the text for details). soe is a dummy variable that equals one for state-owned enterprises and zero otherwise. The dependent variable in (3) and (4) is subsidy_d. See Table 2 for definitions for the other variables. Standard errors clustered by firm are reported in parenthesis. Observations are weighted by employment.

Table 6: Structural Estimation (Fake R&D)

	(1)	(2)	(3)	(4)	(5)
	Taiwan		China		
	Flexible Model	Fake R&D	Flexible Model	Fake R&D	Fake R&D with Taiwan's q, delta and p_bar
q	0.374	0.382	0.259	0.370	0.374
	(0.011)	(0.012)	(0.005)	(0.008)	
delta	0.788	0.786	0.944	0.840	0.788
	(0.035)	(0.098)	(0.022)	(0.574)	
p_bar	0.219	0.224	0.031	0.200	0.219
	(0.010)	(0.013)	(0.006)	(0.024)	
c_bar	1.452	1.513	6.083	15.805	19.511
	(0.092)	(0.254)	(1.095)	(1.634)	(0.859)
sigma_c	0.695	0.696	1.596	0.159	0.182
	(0.008)	(0.061)	(0.046)	(0.161)	(0.201)
xi_0		0.000		0.920	0.910
		(0.141)		(0.203)	(0.007)
xi_1		-0.000		-0.177	-0.152
		(0.100)		(0.122)	(0.003)
xi_2		-0.000		-0.277	-0.239
		(0.146)		(0.159)	(0.256)
sigma_xi		0.000		0.829	0.805
		(0.074)		(0.101)	(0.006)
Sum of Distances	0.049	0.048	0.209	0.159	0.176

Note: Standard errors are reported in the parenthesis. The flexible model refers to the model that allows heterogeneous R&D costs. Column (5) uses the estimated q, delta and p_bar from Taiwan in Column (1).

Tables in Appendix

Table A1: Balanced Panel of Chinese Firms, 2001-2007

Panel A: Correlations between Firm Characteristics and R&D Decision

	(1)	(2)
VARIABLES	rd_d	rd_d
log_tfp	0.0470***	0.435***
	(0.00141)	(0.00724)
wedge		-0.470***
		(0.00820)
ex_d		0.0531***
		(0.00254)
soe	0.266***	0.160***
	(0.00504)	(0.00423)
Year effects	+	+
Industry effects	+	+
Province effects	+	+
Observations	441,039	441,039
R-squared	0.405	0.449

Panel B: Correlations between Firm Initial Characteristics and TFP Growth

	(1)	(2)	(3)
VARIABLES	tfp_g	tfp_g	tfp_g
log_tfp	-0.0821***	-0.0821***	-0.0825***
	(0.00188)	(0.00188)	(0.00189)
rd_d	0.0163***	0.0168***	
	(0.00306)	(0.00301)	
rd_s_h			0.0299***
			(0.00467)
rd_s_m			0.0150***
			(0.00462)
rd_s_l			0.00455
			(0.00587)
ex_d		-0.00313	-0.00395
		(0.00458)	(0.00451)
soe	-0.0328***	-0.0324***	-0.0331***
	(0.00687)	(0.00670)	(0.00652)
Year effects	+	+	+
Industry effects	+	+	+
Province effects	+	+	+

Observations	62,598	62,598	62,598
R-squared	0.153	0.153	0.154

Note: This table reports robustness of the results in Table 3 with respect to ownership and province dummies.

FIGURE 1. Panel A

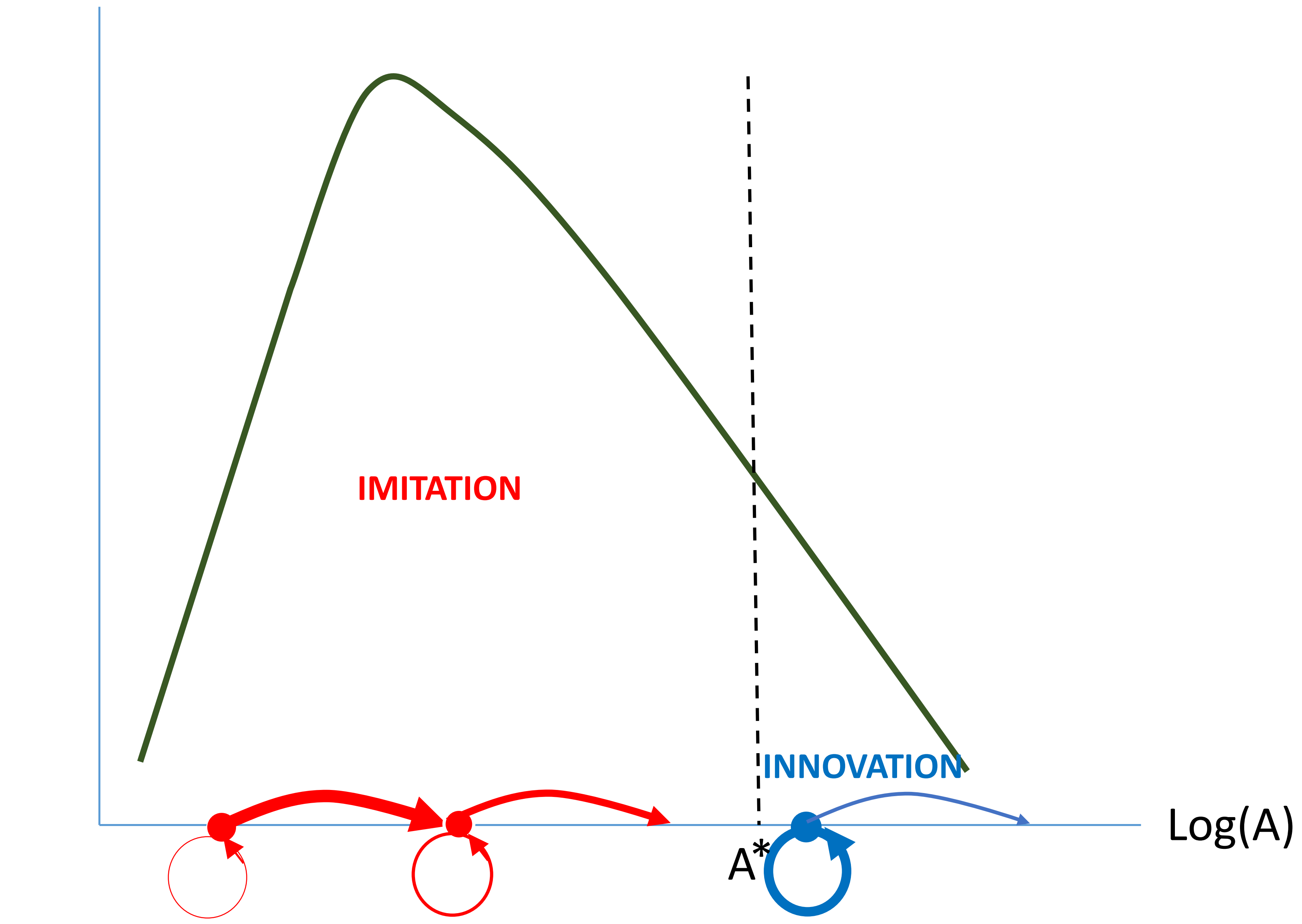


FIGURE 1. Panel B

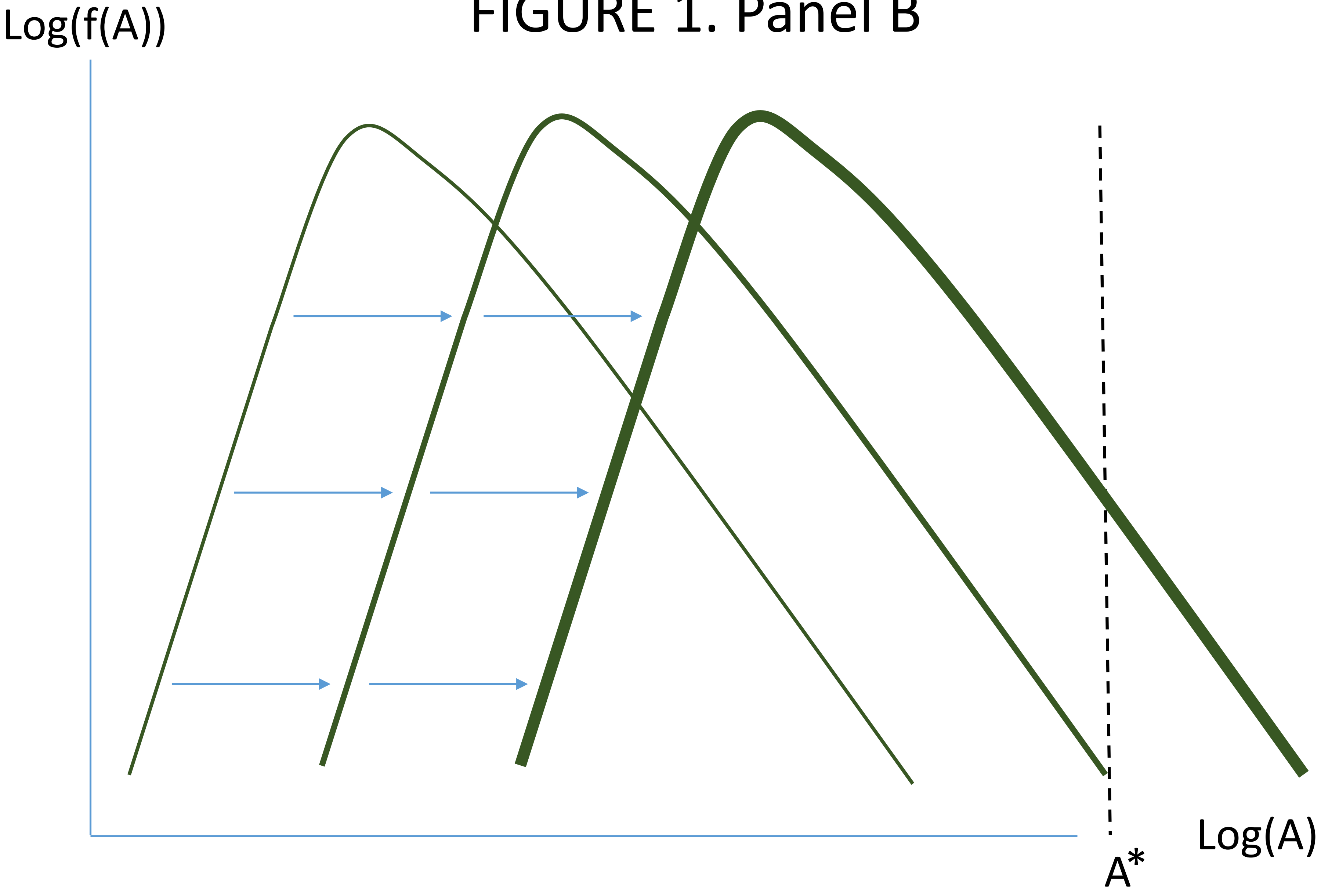


FIGURE 2. Panel A

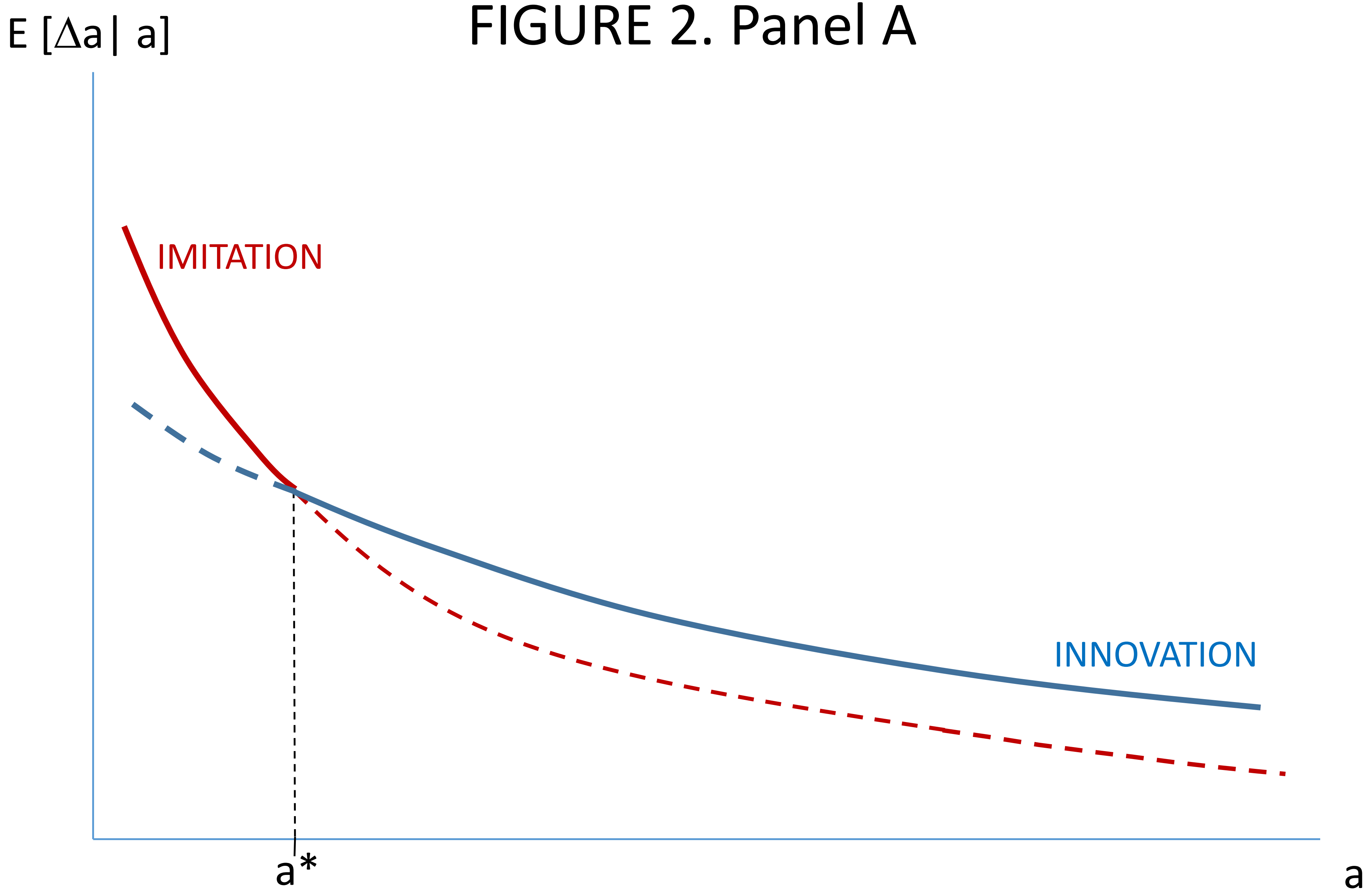


FIGURE 2. Panel B

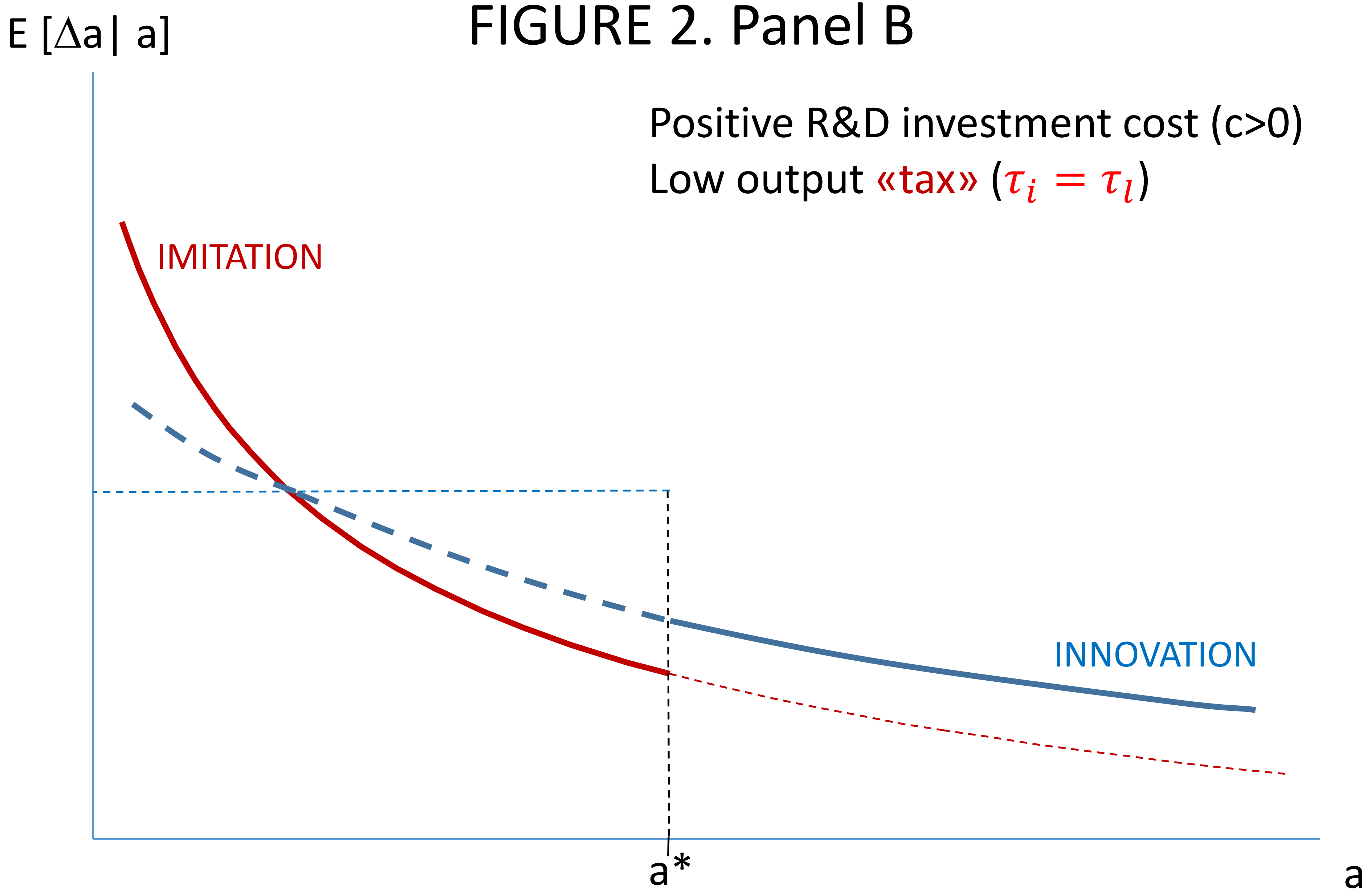


FIGURE 2. Panel C

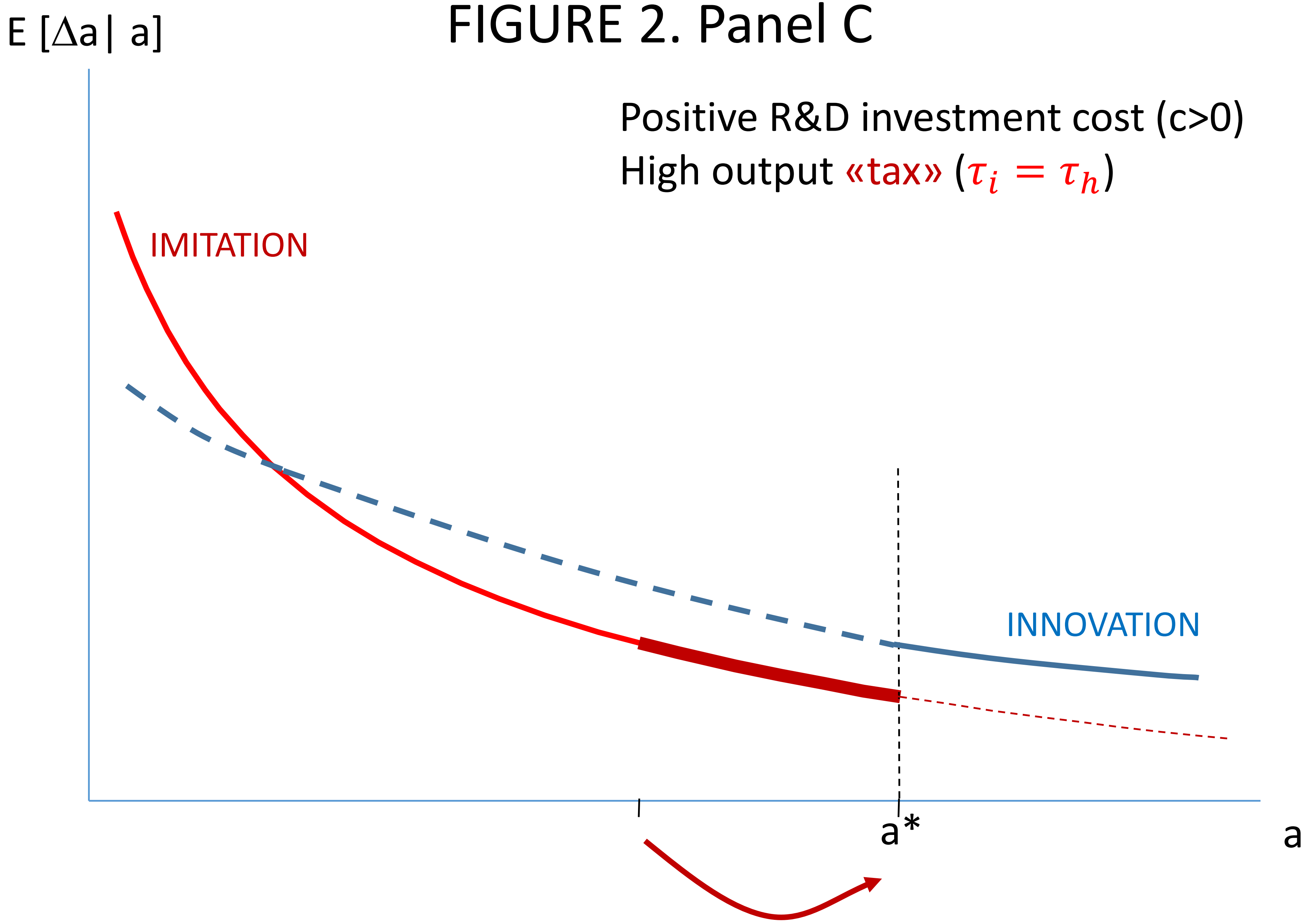


FIGURE 2. Panel D

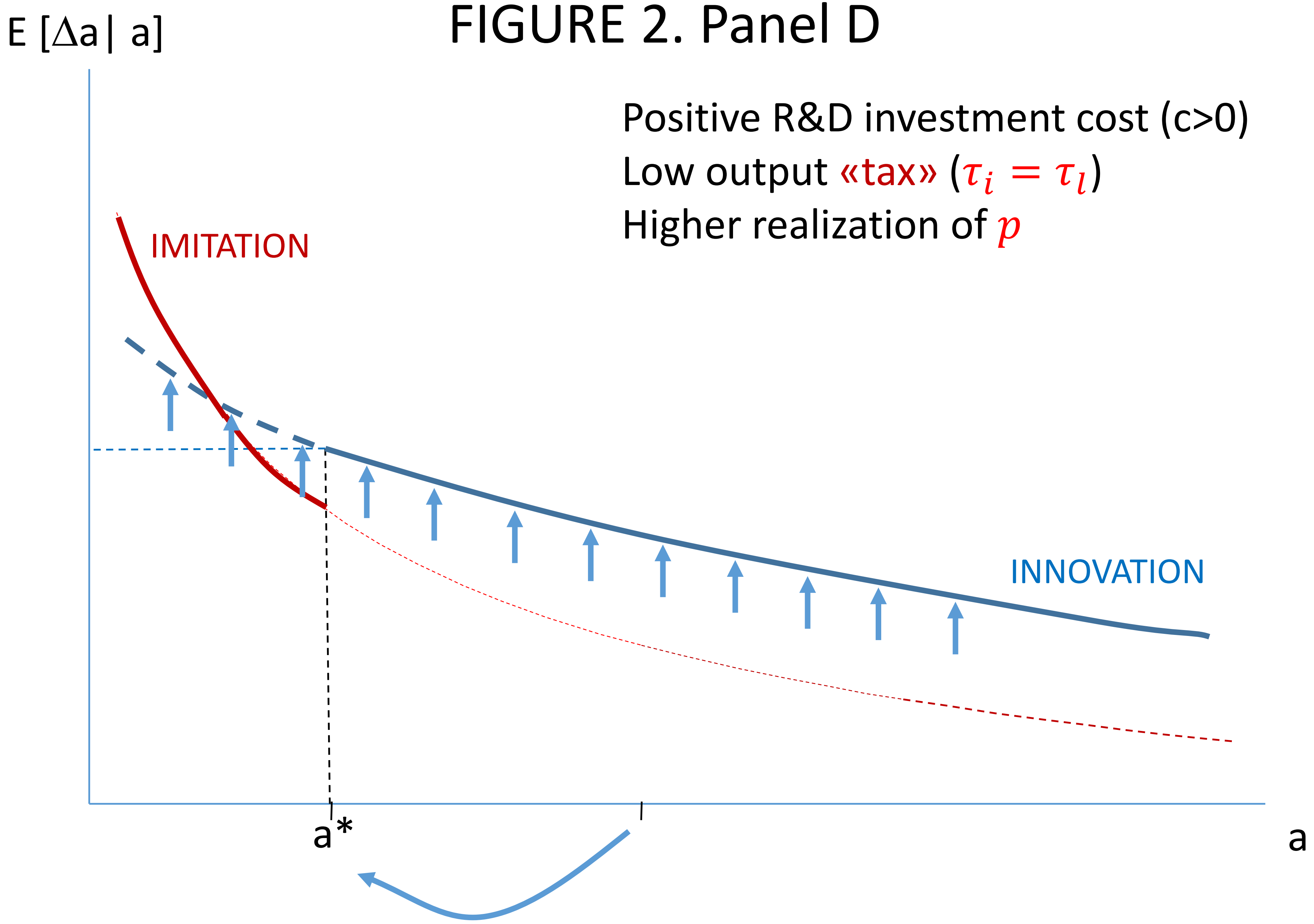
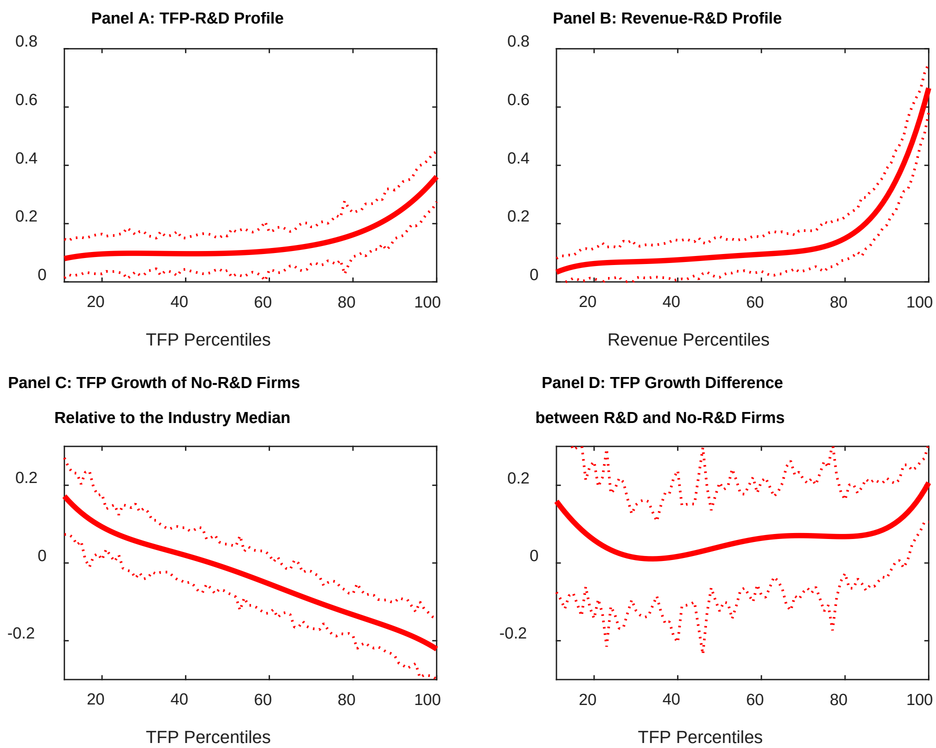
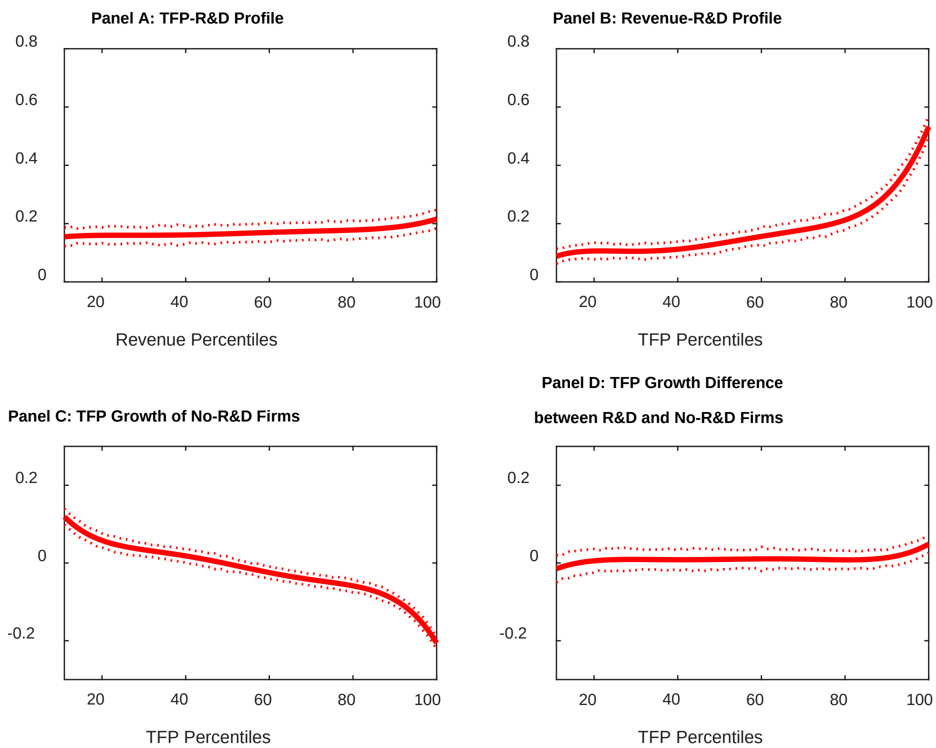


Figure 3: Taiwanese Firms in the Balanced Panel 1999-2004



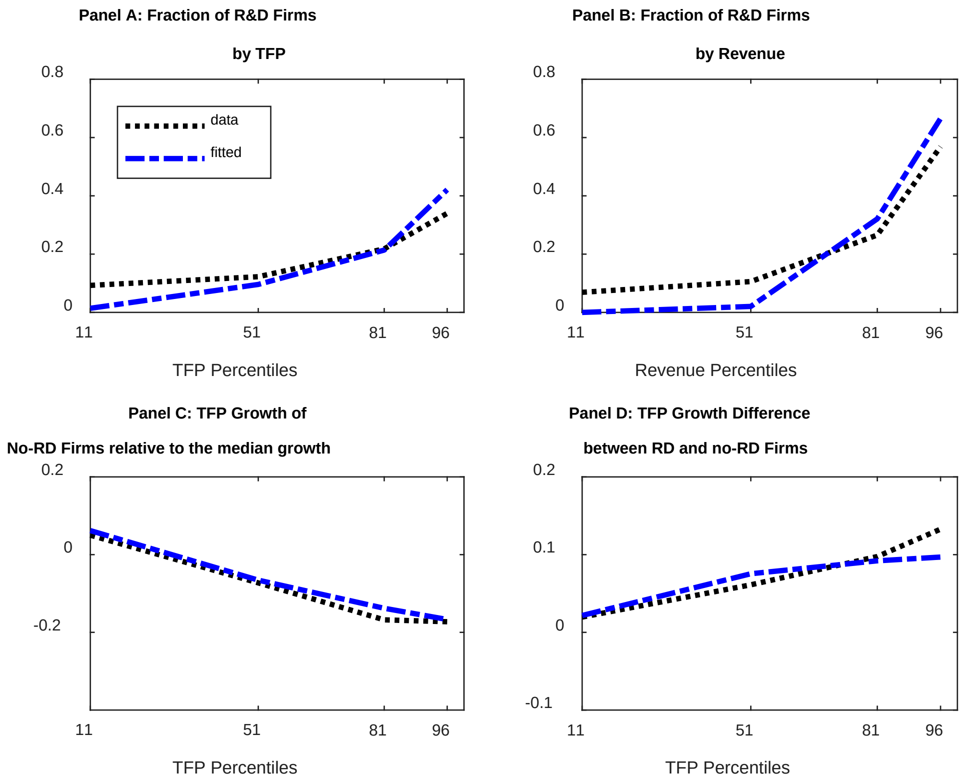
Note: X-axis in Panel A, C and D is the first-period TFP percentiles. X-axis in Panel B is the first-period value added percentiles. The solid lines in Panel A and B plot the first-period fraction of R&D firms in each TFP and value added percentile, respectively. The solid line in Panel C plots the median annualized TFP growth among no-R&D firms in each TFP percentile. The solid line in Panel D plots the difference of the median TFP growth between R&D and no-R&D firms. A firm's TFP growth is its TFP growth relative to the median TFP growth in the industry. All the solid lines are smoothed by fifth-order polynomial. The dotted lines plot the 95-percent confidence intervals by bootstrap.

Figure 4: Chinese Firms in the Balanced Panel 2001-2007



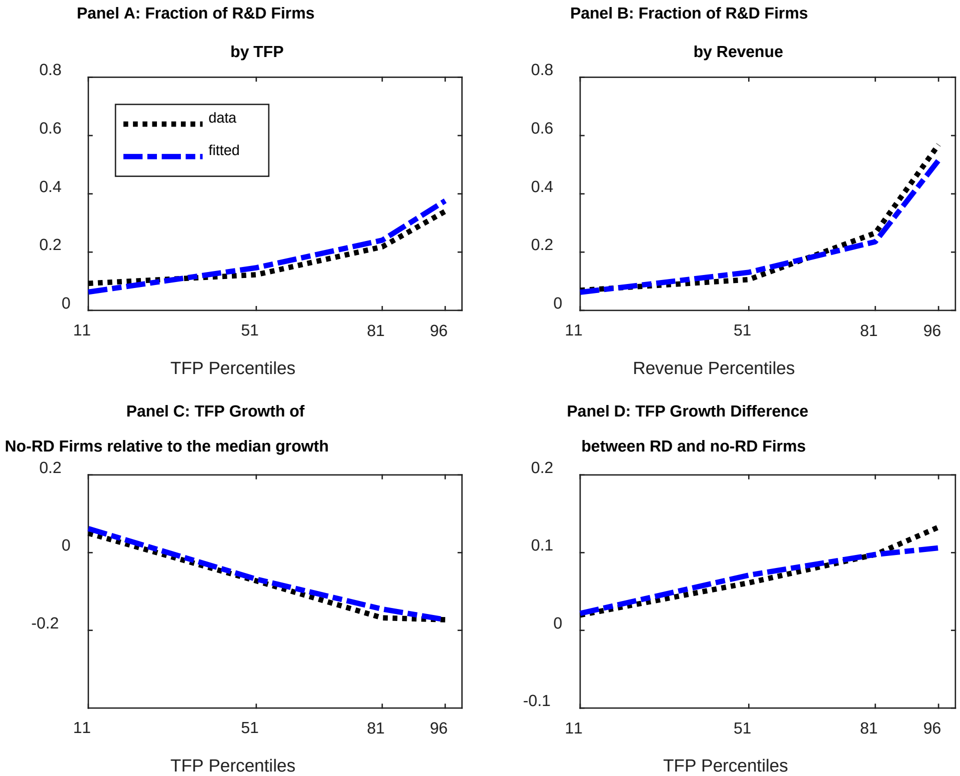
Note: See Figure 1.

Figure 5: Structural Estimation for Taiwan (Parsimonious Model)



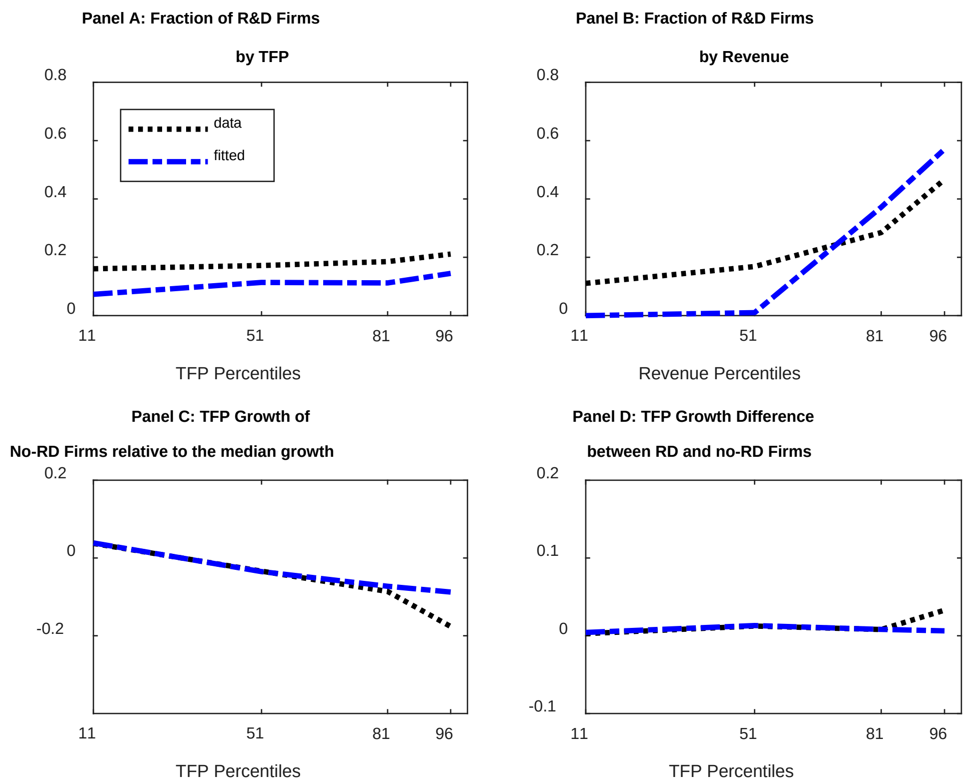
Note: The dotted and dashed lines plot the empirical moments from the balanced panel of Taiwanese firms and the simulated moments from the simple model, respectively. We plot the mean values in four intervals: the 11th percentile to median, the 51th to 80th percentile, the 81th and 95th percentile and the 96th percentile and above.

Figure 6: Structural Estimation for Taiwan (Flexible Model)



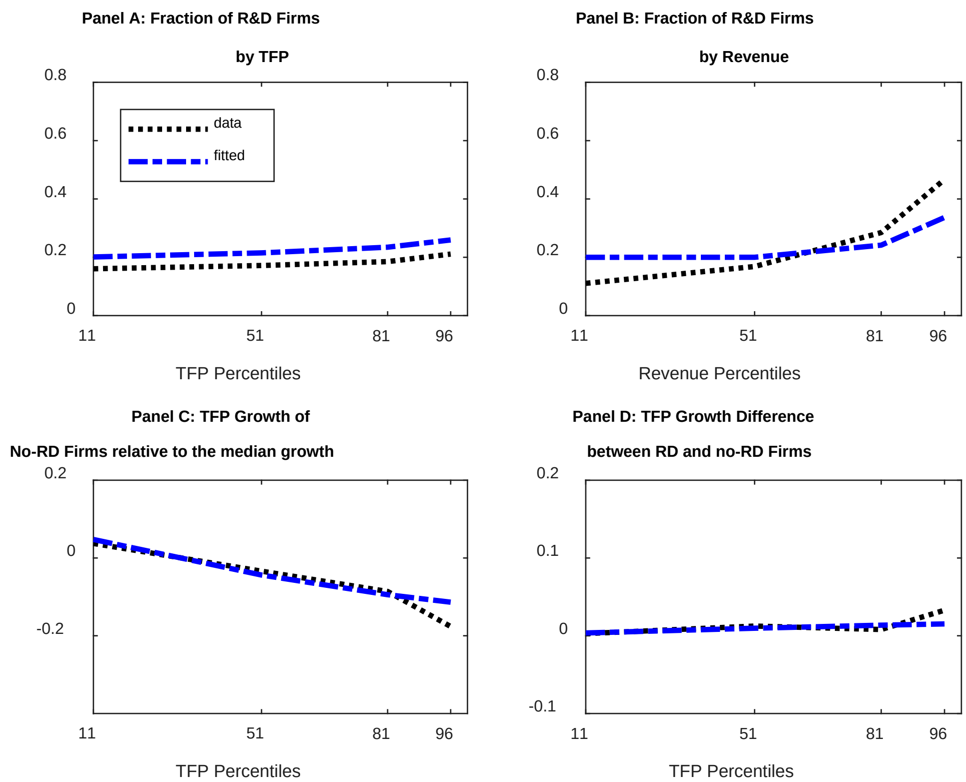
Note: The dotted and dashed lines plot the empirical moments from the balanced panel of Taiwanese firms and the simulated moments from the flexible model that allows heterogeneous R&D costs, respectively.

Figure 7: Structural Estimation for China (Parsimonious Model)



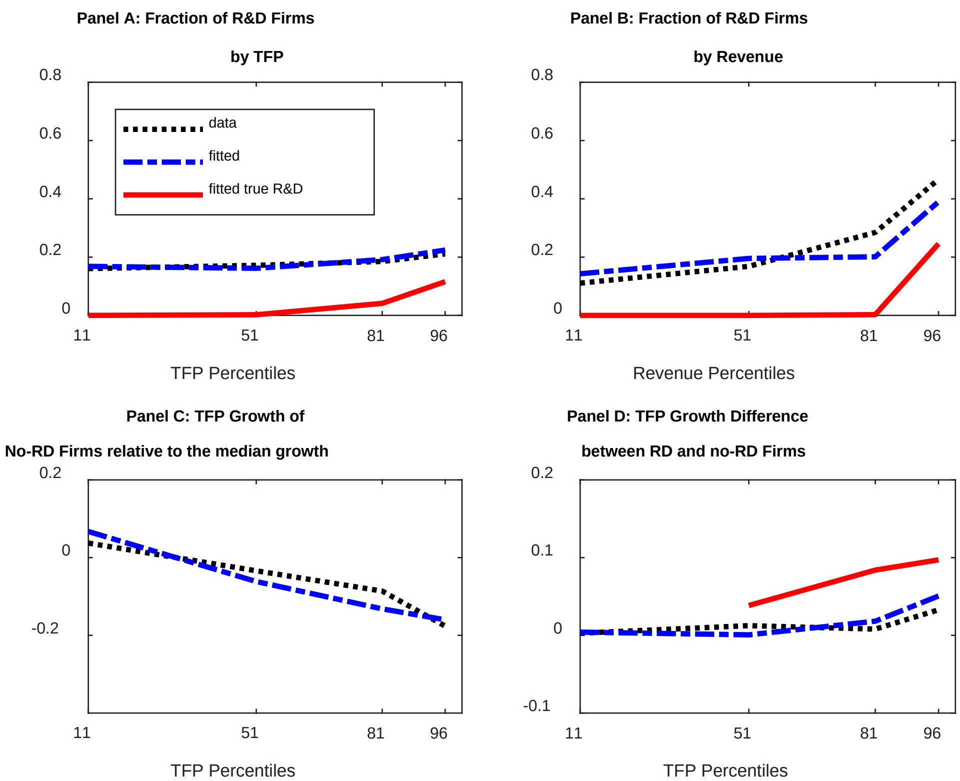
Note: The dotted and dashed lines plot the empirical moments from the balanced panel of Chinese firms and the simulated moments from the simple model, respectively.

Figure 8: Structural Estimation for China (Flexible Model)



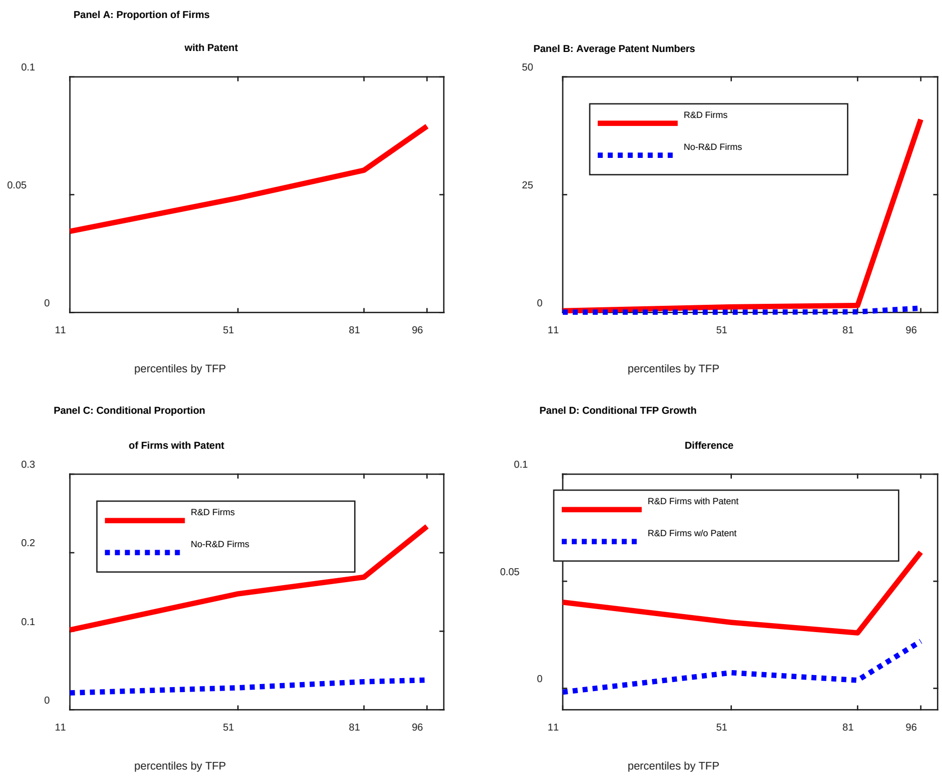
Note: The dotted and dashed lines plot the empirical moments from the balanced panel of Chinese firms and the simulated moments from the flexible model that allows heterogeneous R&D costs, respectively.

Figure 9: Structural Estimation for Chinese Firms (Fake R&D)



Note: The dotted and dashed lines plot the empirical moments from the balanced panel of Chinese firms and the simulated moments from the model that allows fake R&D, respectively. The solid line plots the simulated moments for true R&D.

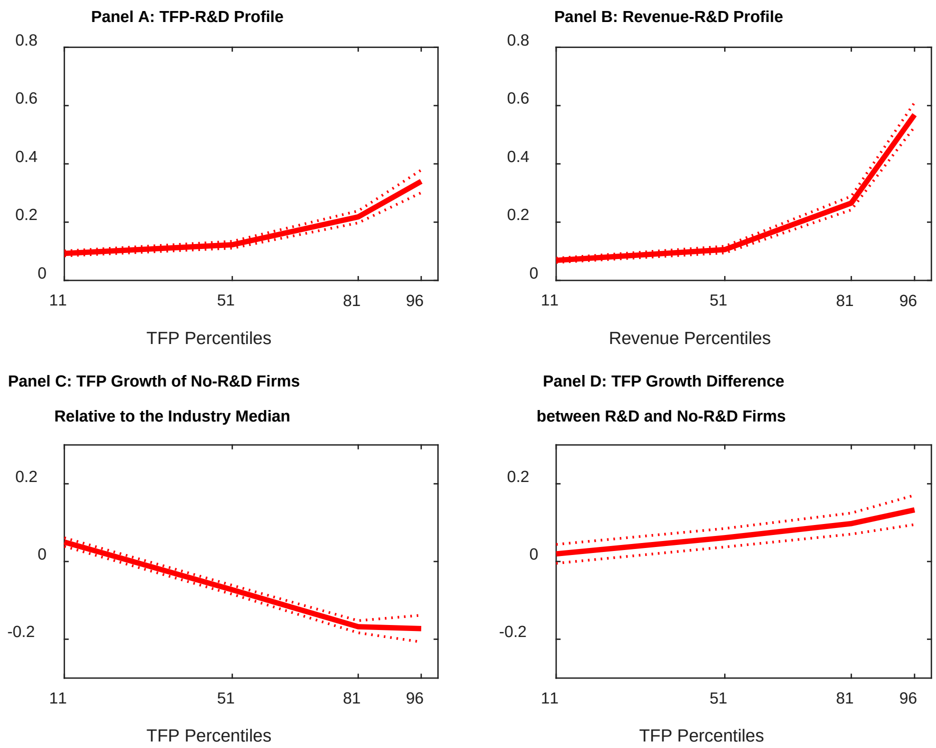
Figure 10: Patent Data for Chinese Firms



Note: Panel A plots the proportion of firms with patents in each TFP interval. Panel B plots the average number of patents among R&D firms (the solid line) and among non-R&D firms (the dotted line). Panel C plots the proportion of R&D firms with one or more patents (the solid line) and the proportion for non-R&D firms with one or more patents (the solid line). The solid line in Panel D plots the TFP growth difference between R&D firms with patents and non-R&D firms. The dotted line plots the TFP growth difference between R&D firms without patents and non-R&D firms.

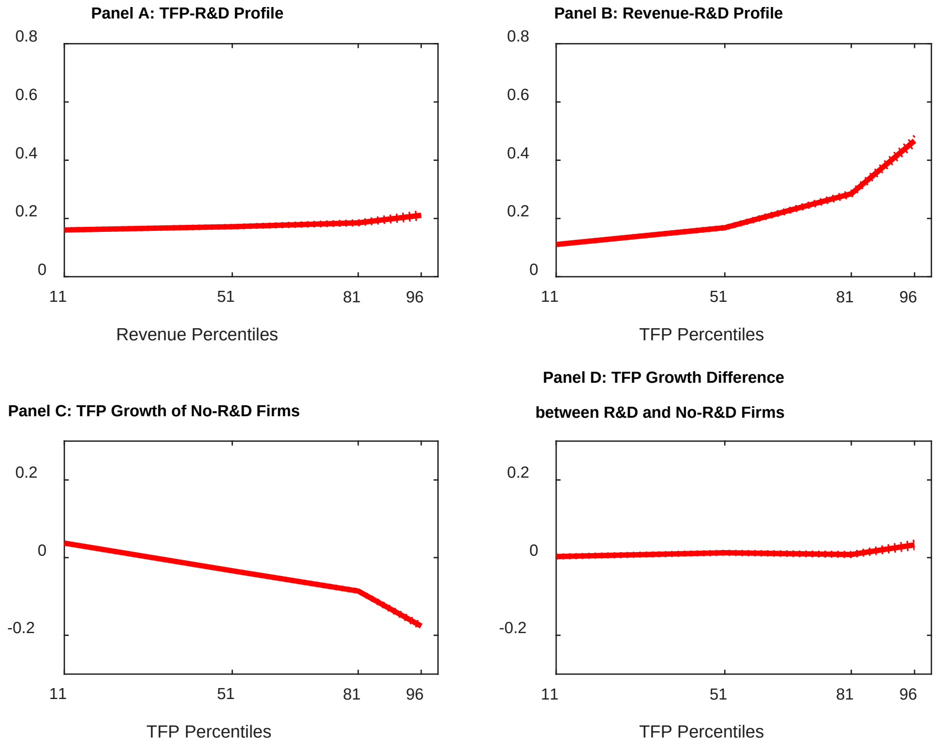
Figures in Appendix

Figure A.1: Taiwanese Firms in the Balanced Panel 1999-2004



Note: We plot the mean values in four intervals: the 11th percentile to median, the 51th to 80th percentile, the 81th and 95th percentile and the 96th percentile and above. All the others follow the specifications in Figure 1.

Figure A.2: Chinese Firms in the Balanced Panel 2001-2007



Note: See the explanations for Figure A.1.

APPENDIX

ADDITIONAL FIGURES

Figure: Non-Exporting Taiwanese Firms

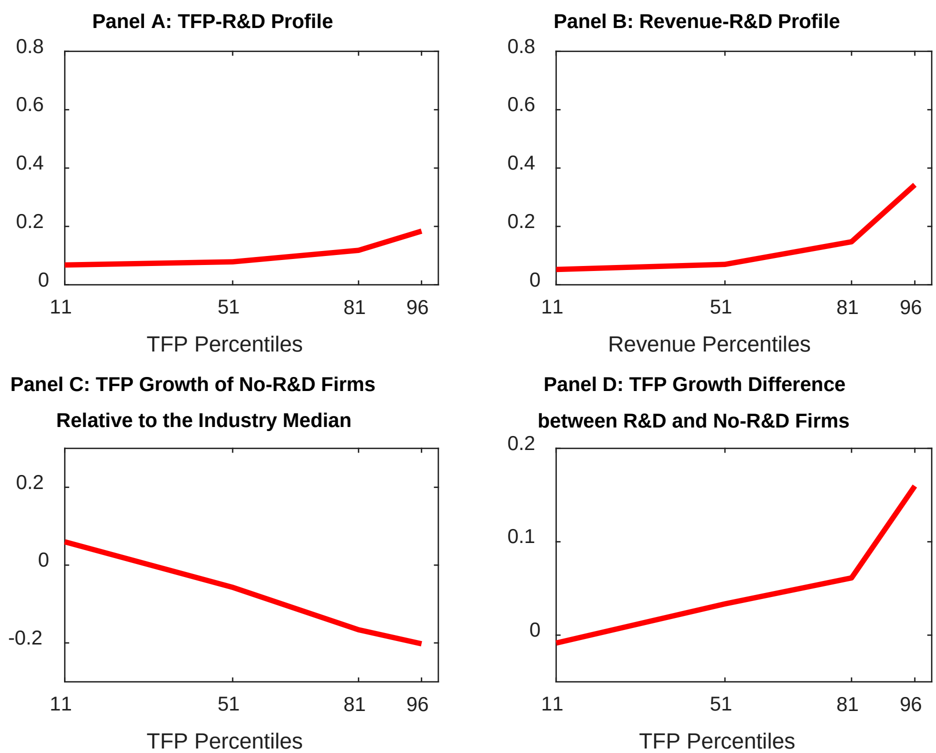


Figure: Exporting Taiwanese Firms

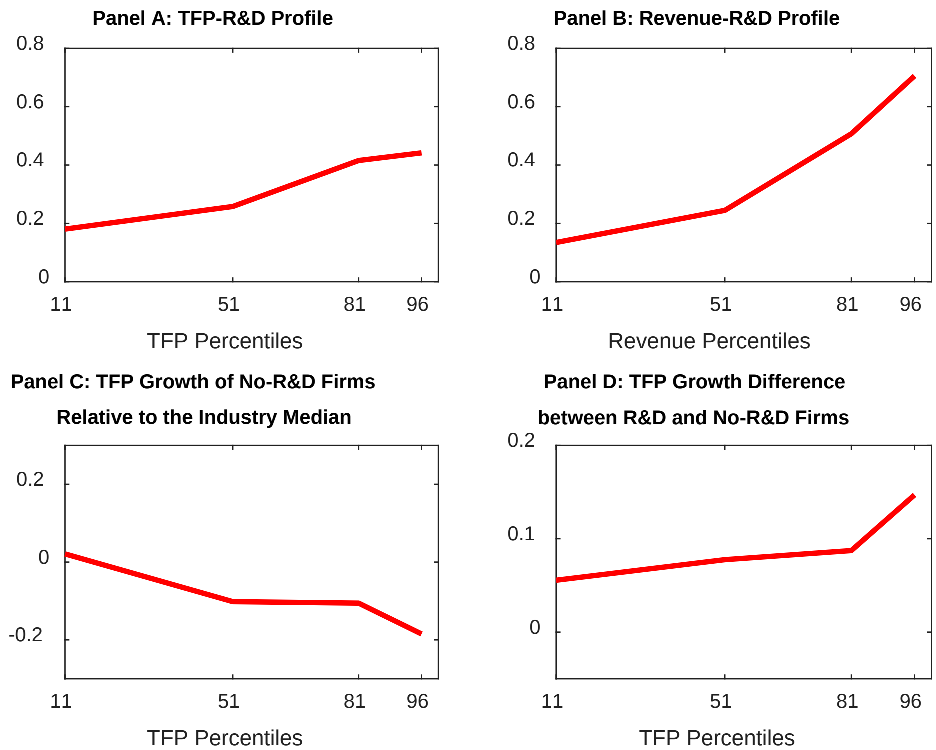


Figure: Non-Exporting Chinese Firms

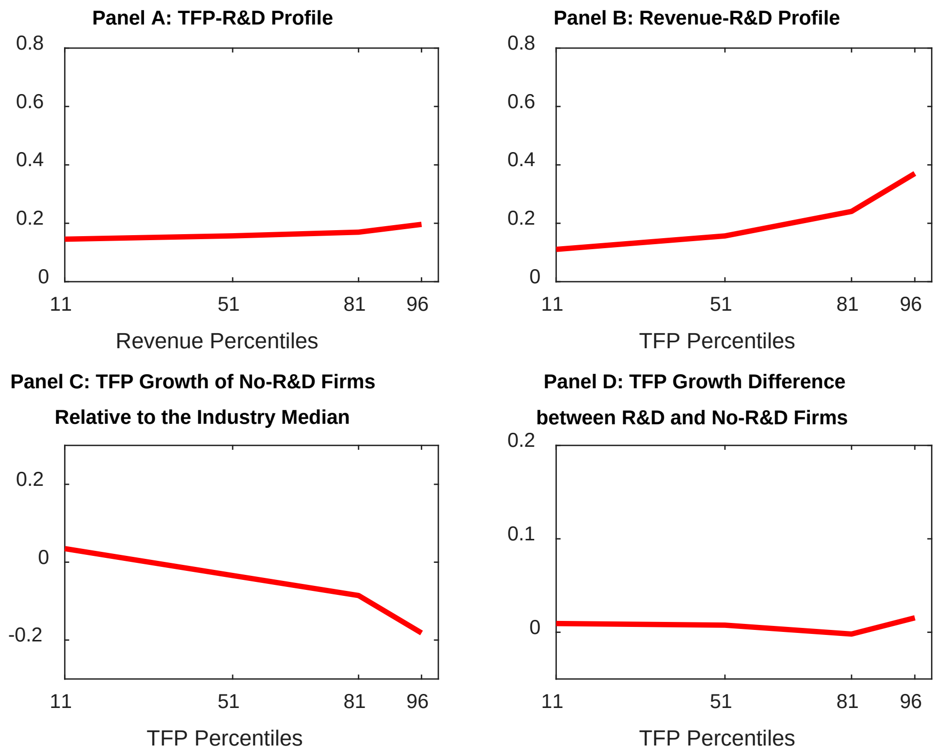


Figure: Exporting Chinese Firms

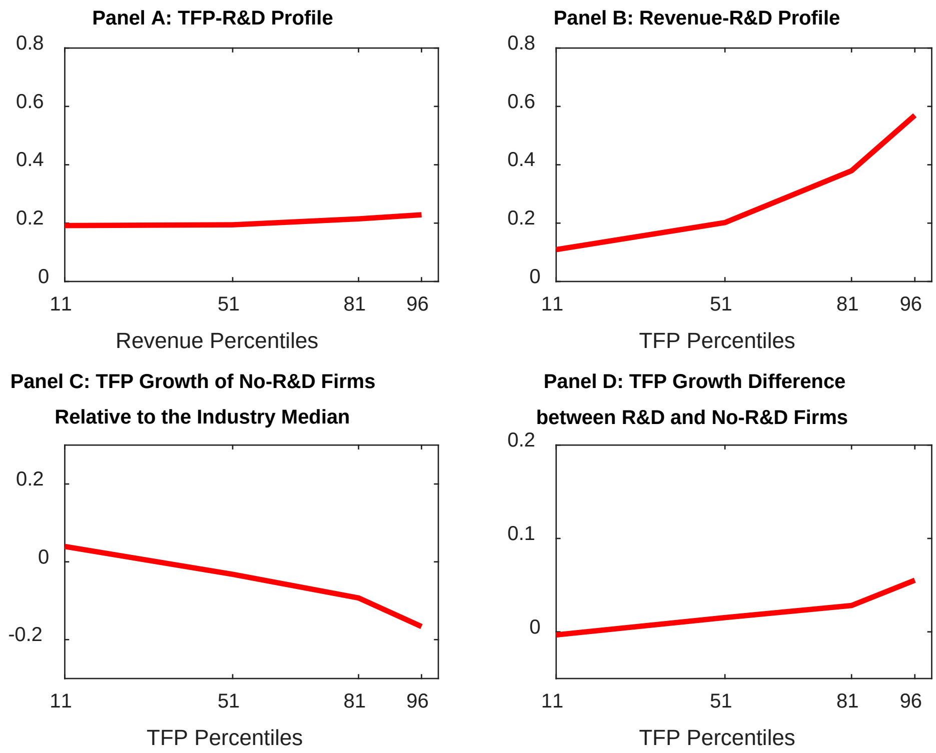


Figure: North China

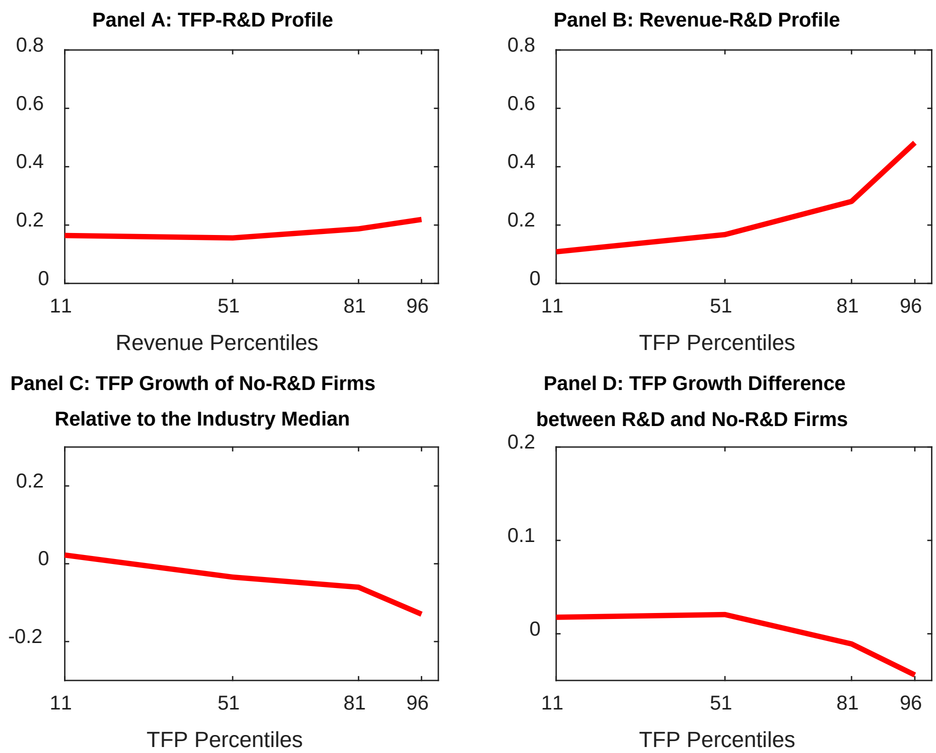


Figure: Northeast China

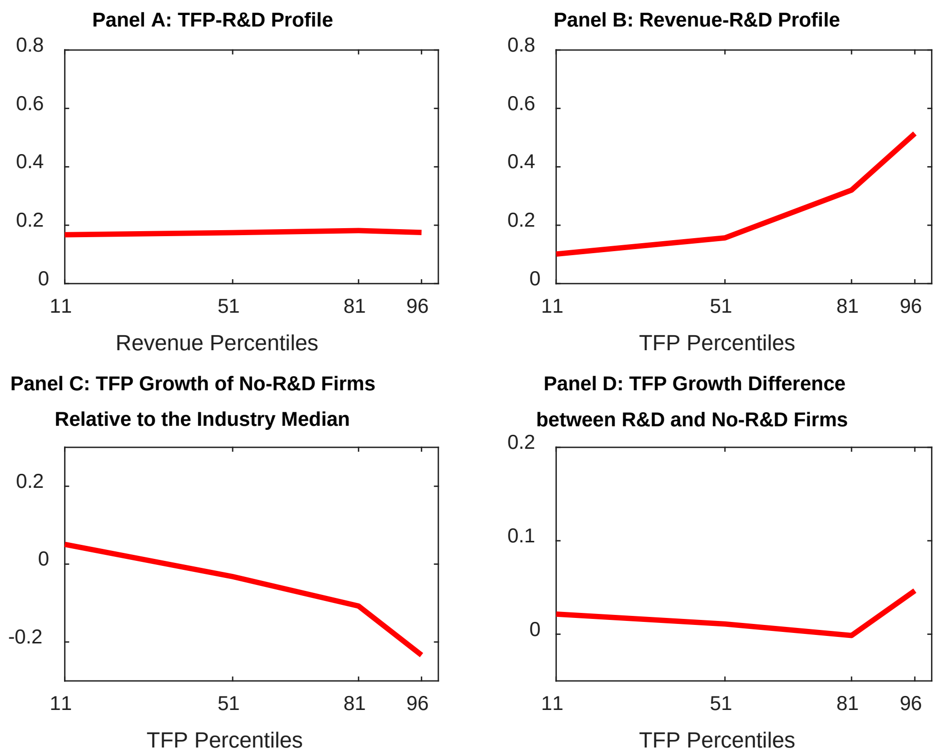


Figure: East China

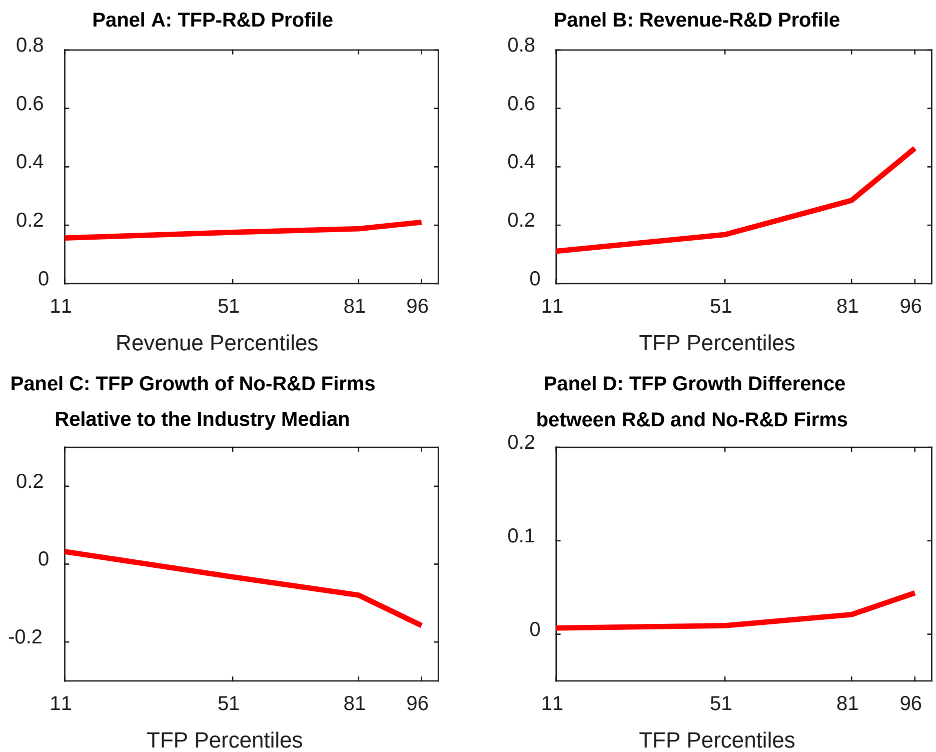


Figure: South China

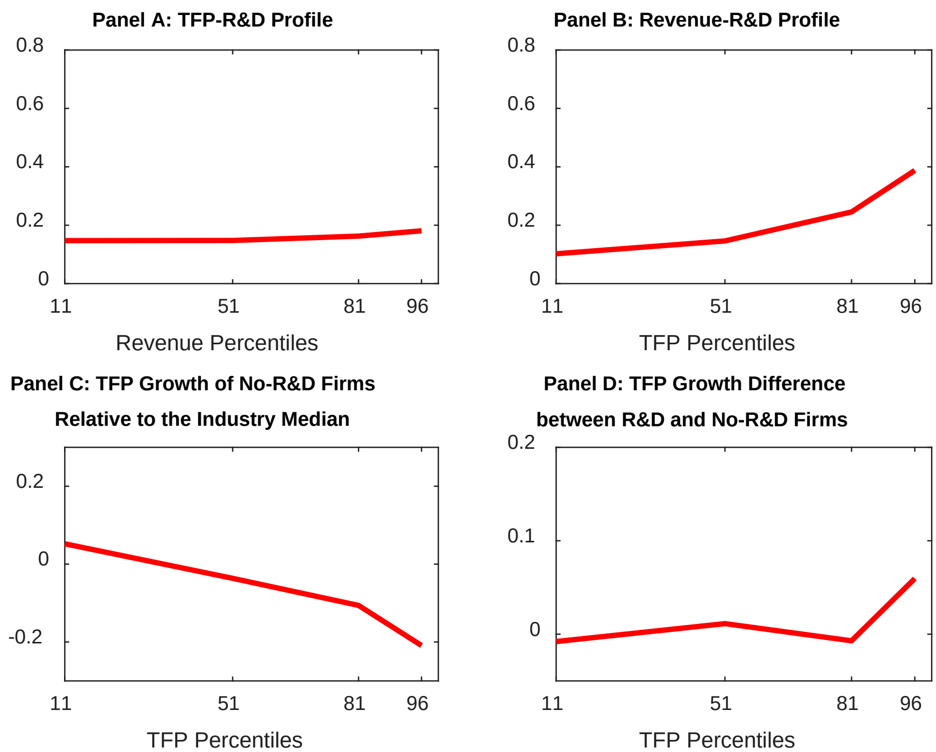


Figure: Southwest China

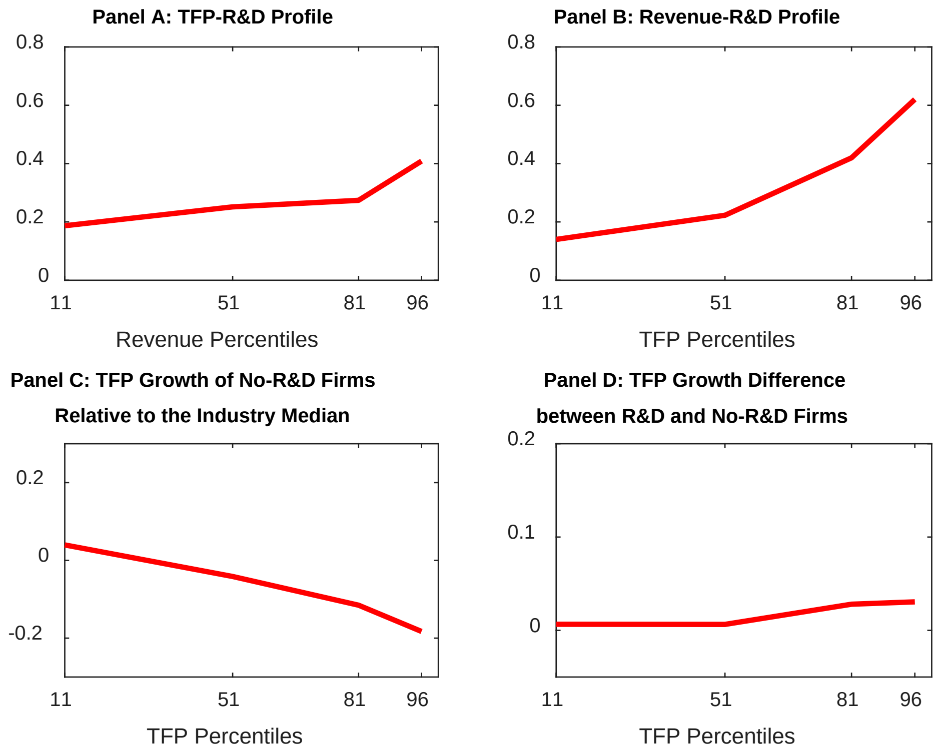


Figure: Northwest China