From Imitation to Innovation: Where Is all that Chinese R&D Going?*

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

September 21, 2021

Abstract

We construct an endogenous growth model with random interactions where firms are subject to distortions. The productivity distribution evolves endogenously as firms seek to upgrade their technology over time either by innovating or by imitating other firms. We use the model to quantify the effects of misallocation on TFP growth in emerging economies. We structurally estimate the stationary state of the dynamic model targeting moments of the empirical distribution of R&D and TFP growth in China during the period 2007–12. The estimated model fits the Chinese data well. We compare the estimates with those obtained using data for Taiwan. We perform counterfactuals to study the effect of alternative policies. R&D misallocation has a large effect on TFP growth.

JEL Codes: O31, O33, O47

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

*We thank three referees, Daron Acemoglu, Philippe Aghion, Joe Altonji, Ufuk Akcigit, Manuel Amador, Marco Bassetto, Christopher Flynn, Chang-Tai Hsieh, Chad Jones, Patrick Kehoe, Juan Pablo Nicolini, Jesse Perla, Torsten Persson, Chris Tonetti, Daniel Xu, Gianluca Violante, conference participants at the ASSA Annual Meeting 2019, the Bank of Canada-IMF-University of Toronto Meeting on the Chinese Economy, the 2021 China Econometric Society Meeting, CIFAR, The Economics of Creative Destruction (Paris, June 9–12 2021), and seminar participants at the Arizona State University, Cambridge University, Columbia University, Collegio Carlo Alberto, the Federal Reserve of Minneapolis, the National University of Singapore, New York University, Ohio State University, Princeton University, Stanford University, Tsinghua University, the University of California at Berkeley, the University of Chicago, University College London, University of Houston, the University of Notre Dame, University of St. Andrews, and Yale University.

[†]Vrije Universiteit Amsterdam, School of Business and Economics, m.d.konig@vu.nl.

[‡]University of Minnesota, Department of Economics, kjetil.storesletten@gmail.com.

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

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

1 Introduction

In this paper, we construct and estimate a model of endogenous technical change with random interactions where firms are subject to distortions. The goal of the paper is to quantify the dynamic effects of misallocation on the investments firms make to improve their productivity growth.

In the theory, the evolution of the total factor productivity (TFP) distribution hinges on profit-maximizing firms seeking to upgrade their technology. To this end, firms face a binary choice: they can either adopt better technologies used by other firms (*imitation*) or break new ground and search for new technologies (*innovation*). Focusing on innovation requires an investment and entails some opportunity cost of foregoing learning through random interactions. The firms' relative productivity determines the comparative advantage of the two alternative strategies. Firms that are farther from the technology frontier can gain more from random interactions. Conversely, for firms closer to the technology frontier, the scope for imitating other firms is limited, and they must innovate in order to improve their technology. The investment decision is affected by firm-specific labor and capital market distortions (wedges). These wedges affect the investments in innovation because a positive wedge reduces the gains associated with a future TFP increase.

We structurally estimate the theory exploiting the stationary equilibrium of the dynamic model. We use the Simulated Method of Moments (SMM), targeting moments of the empirical distribution of R&D and productivity growth that are salient in the theory. We use data from manufacturing firms in mainland China (henceforth, China) in the period 2007–12. We are motivated by the observation that in recent years, the rapid economic growth in China has been accompanied by a boom in R&D expenditure and growing emphasis by the government on innovation (see, e.g., Ding and Li (2015), Zilibotti (2017)).[1] However, where is all this R&D going? A common concern is that these investment decisions are distorted by policies and frictions (e.g., credit constraints) that are pervasive in China. Our methodology allows us to assess the contribution of these investments for aggregate growth.

We proxy the choice between imitation and innovation by the firms' R&D investment behavior on the extensive margin. We classify firms making R&D investments as innovators and firms not making R&D investments as imitators. We study the robustness of the results to the choice of the proxy. We measure distortions using the methodology proposed by Hsieh and Klenow (2009). In our theory, the presence of heterogeneous output wedges lowers the correlation between TFP and propensity to pursue innovation—when the decision to invest in R&D is distorted, the firm's size matters more than its TFP. We document that in our data the propensity of firms to invest in R&D is positively correlated with TFP and size—the latter correlation being stronger. Moreover, conditional on TFP, TFP growth is higher for R&D than for nonR&D firms. All these observations are in line with the predictions of the theory. We estimate the model. The estimated model matches well the target moments from a quantitative perspective. The benchmark model predicts an annual aggregate TFP growth rate of 3.6%, which is close to the empirical counterpart for China for 2007–12 (3% annual growth in our sample). This moment is not targeted in the estimation.

Next, we extend the model to allow for heterogeneous R&D costs across firms. To this aim, we introduce "innovation wedges" (that are distinct distortions from the standard output wedges) that we allow to be correlated with firm-level TFP. The estimated pattern is suggestive of an active industrial

[1] The transition toward innovation-based growth is a central theme in the government's strategy. The 13th Five-Year Plan (2016–20) emphasizes the promotion of research in strategic and frontier fields. The National Innovation-Driven Development Strategy Outline issued in June 2016 states that China should become an innovation-oriented economy by 2020 and a technological innovation powerhouse by 2050. While China invested barely 1% of its GDP in the 1990s, R&D investments increased to 2.4% of GDP by 2020.

policy, which is arguably a salient feature of China. Finally, motivated by the findings of Chen et al. (2021), we explore an extension in which some Chinese firms may respond to fiscal incentives by fudging R&D expenditure, that is, relabeling part of their operational expenditure as R&D, in order to cash in on public subsidies.

For contrast, we estimate the model using plant-level data from Taiwan, for which census data on R&D investments are also available. Taiwan is a natural comparison for China, not only for its geographic and cultural proximity, but also for the structural similarities between the two economies in which the manufacturing sector plays a central role. The results for Taiwan are qualitatively similar to those obtained for China, although both the TFP growth of R&D firms and the rate of technology diffusion are larger for Taiwanese firms. Interestingly, the estimates of the Taiwanese technology parameters are similar to the model in which we allow for some Chinese firms to misreport R&D. Finally, we perform a set of counterfactual policy experiments. In one of them, we counterfactually reduce the variance of wedges by 50%. The reduction in misallocation triggers a dynamic adjustment towards a new stationary equilibrium with higher growth. Because the reduction in misallocation strengthens the comparative advantage of high-TFP firms, the transition is associated with an increase in productivity dispersion across firms and an acceleration of growth.

The basic model abstracts from international spillovers. Arguably, China's stellar growth is partly driven by the adoption of technologies from the international frontier and is bound to slow down as China approaches the world technology frontier. In an extension, we incorporate this channel and embed the analysis into a model where misallocation affects growth only throughout the transition to the world balanced growth equilibrium. We find that even in this model an exogenous change in misallocation has a large effect on both transitional growth and the long-run GDP level relative to the frontier. These effects are largely driven by the endogenous change in the TFP distribution which impacts on the imitation-innovation decisions of firms.

Another set of counterfactuals studies the effect of a reduction in innovation costs through non-targeted R&D subsidies. We find a nonmonotonic effect on growth: a subsidy inducing a moderate increase in R&D investments speeds up TFP growth. However, a sufficiently large subsidy slows down productivity growth. The reason is that in our model the average TFP growth hinges on both innovation and imitation. Inducing too many firms to innovate has a very large opportunity cost in terms of foregone technology diffusion. Therefore, our theory provides a novel insight to the debate on innovation policy: it is important to induce the "right firms" to pursue an innovation strategy.

Related literature: Our study is related to various streams of the growth and development literature. First, it contributes to the debate 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). R&D investments and spillovers are core elements of the neo-Schumpeterian theory à la Aghion and Howitt (1992); see also Griliches (1998). While this literature highlights a process of creative destruction where new firms are carriers of innovation, recent research by Garcia-Macia et al. (2019) finds that the lion's share of aggregate growth stems from productivity growth by 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 et al. (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 et al. (2011), Hsieh and Song (2015), Cheremukhin et al. (2017), Brandt et al. (2016), and Tombe and Zhu (2019).

Our paper also contributes to the recent literature describing the endogenous evolution of the distribution of firm size and productivity. This includes, among others, Jovanovic and Rob (1989), Luttmer (2007, 2012), Ghiglino (2012), Perla and Tonetti (2014), Acemoglu and Cao (2015), Lucas and Moll (2014), König et al. (2016), Benhabib et al. (2014, 2021), Akcigit et al. (2018). While a number of these studies emphasize random interactions, the theoretical paper by König et al. (2016) is the only one highlighting a trade-off between innovation and imitation. Our theoretical model builds on that paper.

Finally, our paper is related to the burgeoning theoretical and empirical literature on firms dynamics with R&D investments and creative distruction. These studies include, among others, Bloom et al. (2002), Klette and Kortum (2004), Lentz and Mortensen (2008), Acemoglu and Cao (2015), Akcigit and Kerr (2018), Acemoglu et al. (2018), and Akcigit et al. (2021).

Related literature on R&D in China: 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 intended to kick-start innovation and start-ups through the creation of innovation clusters, technology business incubators, and the promotion of venture capital. Another important part of the government strategy is tax incentives for innovation. This takes the form of tax bonuses 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 allowances 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 heterogenous across provinces, local communities, sectors, and even at the firm level, often as a function of political connections (Bai et al., 2016).

Some empirical studies evaluate the effects of R&D investment 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. Despite not being a representative sample, these enterprises engaged in nearly 40% of China's R&D in 2001. The authors estimate the 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 smaller than similar estimates for the U.S. and European firms for which the result of earlier studies find elasticities in the range of 0.6–1. The study is based on data from the 1990s. However, Dang and Motohashi (2015) find similar results using data for the period 1998–2012 based on matching the NBS data to patent data.

Jia and Ma (2017) use a panel data set of Chinese listed companies covering 2007–13 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 less influence on state-owned enterprises' R&D expenditures.

Chen et al. (2021) 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 and find 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 nonR&D expenditures as R&D expenses.

The paper is structured as follows; Section 2 presents the theory. Section 3 discusses the data and some descriptive evidence. Section 4 presents the econometric methodology. Sections 5–7 discuss the estimation results, robustness analysis, and counterfactual experiments. Section 8 concludes. The supplementary appendix contains technical results and additional tables and figures. Further technical details are deferred to a web appendix.

2 Theory

Consider a dynamic economy populated by a unit measure of monopolistically competitive firms. Firms produce differentiated goods that are combined into a homogeneous final good by a Dixit-Stiglitz aggregator with a constant elasticity of substitution $\eta > 1$ between goods, implying $Y = \left(\int_0^1 Y_i^{(\eta-1)/\eta} di\right)^{\eta/(\eta-1)}$.

Firms are owned by overlapping generations of two-period–lived manager-entrepreneurs as in Song et al. (2011). In each period, the firm is owned by an old entrepreneur who is residual claimant on the firms' profits, but run by a young manager. In the first period of her life, the manager decides the strategy to improve the firm's productivity in the next period having access to frictionless credit markets. In the following period, she turns into an old entrepreneur, hires a young manager to run the firm, and appropriates and consumes the firm's profits.

In this environment, we can break down the firm's problem into two steps. First, there is a static maximization problem: the firm's manager hires capital and labor to maximize profits given the current state of productivity. Second, there is an intertemporal investment problem: the firm makes an investment decision that affects the next period's productivity and profits. The OLG structure simplifies the dynamic problem by turning it into a sequence of two-period decisions. This allows us to retain analytical tractability and avoid complications that would make the structural estimation problem infeasible.

Static production efficiency: The firm's technology is represented by a constant returns to scale Cobb Douglas production function:

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

where $\alpha \in (0,1)$, $K_i(t)$ is capital, $L_i(t)$ is labor, and $A_i(t)$ is total factor productivity (TFP). Like in Hsieh and Klenow (2009), firms have heterogeneous $A_i(t)$ and rent capital and labor from competitive markets subject to distortions. We summarize all distortions into a single output wedge that we view as a catch-all for a variety of firms-specific distortions on labor and credit markets–the latter being especially important in China as documented by Song et al. (2011) and Hsieh and Song (2015) . More formally, firms maximize profits taking factor prices as given, but their decisions are distorted by a set of output wedges τ_i. Note that $\tau_i < 0$ indicates a negative wedge, or an implicit output subsidy.[2]

Because the characterization of the static equilibrium is as in Hsieh and Klenow (2009), we omit details. Here, we summarize the two equilibrium conditions that are sufficient to derive the dynamic equilibrium and that we use in the empirical analysis. Firm i's current (period t) profits are given by

$$\pi_i(t) \propto (A_i(t)(1-\tau_i(t)))^{\eta-1}. \tag{1}$$

Profits increase in TFP and decrease in the wedge. Moreover, the firm's value added satisfies

$$P_i(t)Y_i(t) \propto (A_i(t)(1-\tau_i(t)))^{\eta-1}. \tag{2}$$

Intuitively, the firm's value added—or its size—is increasing in TFP and decreasing in the wedge. Equations (1)–(2) then implies that TFP satisfies

$$A_i \propto \frac{[Y_i(t)P_i(t)]^{\frac{\eta}{\eta-1}}}{[K_i(t)]^{\alpha}[N_i(t)]^{1-\alpha}}. \tag{3}$$

In the theoretical section, we assume that τ takes on only two values, $\tau \in \{\tau_h, \tau_l\}$ where $\tau_l < \tau_h$. The stochastic realizations of τ follow a persistent Markov process. Namely, the probability that τ remains constant exceeds 50% in each state. In the empirical analysis, τ has a continuous support.

Productivity dynamics: The endogenous evolution of the productivity distribution is determined by the strategy firms adopt to increase their productivity. To analyze this decision, it is useful to introduce some notation. We assume that advancements occur over a productivity ladder where each successful attempt to move up the ladder results in a constant log-productivity accrual: $\log(A_{i,t+1}) = \log(A_{i,t}) + \tilde{a}$, where $\tilde{a} > 0$ is a constant (thus, $\log(A) \in \{\tilde{a}, 2\tilde{a}, ...\}$). We define $a \equiv \log(A)/\tilde{a}$ and denote the ranking in the productivity ladder by $a \in \mathbb{N}^+$. Moreover, Π denotes the productivity distribution, $\Pi_1, \Pi_2, ...$ denotes the proportion of firms at each rung of the ladder, and $F_a = \sum_{j=1}^{a} \Pi_j$ denotes the associated cumulative distribution. We model innovation as a step-by-step process: in each period, productivity can either increase by one step or stay constant.[3] We abstract from entry and exit—a limitation to which we return below.

Firms can increase their productivity through either *innovation* or *imitation.* We model imitation as an attempt to acquire knowledge through random interactions with other firms (e.g., by adopting better managerial practices). This strategy hinges on the existing productivity distribution because firms only learn when they meet more productive firms. Innovation is modeled as an exploration of new avenues and is independent of other firms' productivity. Although both strategies could in principle require investments, the crux of the choice is the cost difference. Therefore, we normalize the cost of imitation to zero, and let the innovation cost be nonnegative.

Imitation: A firm pursuing the imitation strategy is randomly matched with another firm in the empirical distribution. If the firm meets a more productive firm, its productivity increases by one notch with probability $q > 0$. If the firm meets a less productive firm, it retains its initial productivity.

[2] The wedge τ_i can alternatively be interpreted as a geometric average of capital and labor wedges. More formally, let τ_{Ki} and τ_{Li} denote firm-specific "taxes" on capital and labor, respectively. The output wedge τ_i is then defined by the following equation: $1-\tau_i \equiv (1+\tau_{Ki})^{-\alpha}(1+\tau_{Li})^{-(1-\alpha)}$.

[3] König et al. (2016), allow for more general stochastic processes, where a successful firm can make improvements of different magnitudes. For simplicity, we abstract from this possibility.

Innovation: An innovating firm can improve its productivity via two channels. First, it can make a discovery unrelated to the knowledge set of other firms. The probability of success through this channel is p, where p is drawn from an i.i.d. distribution with cumulative distribution function $G : [0, \overline{p}] \to [0, 1]$ where $\overline{p} \leq 1$. Firms observe the realization of p before deciding whether to innovate or imitate. The heterogeneity in p avoids the stark implication that the position of the firm in the productivity distribution fully determines the innovation-imitation choice that would be rejected by the data.

If innovation fails, the firm gets a second chance to improve its technology via (passive) imitation. However, in this case the probability of success is different from that of a firm actively pursuing imitation, being equal to $\delta q (1 - F_a) \geq 0$. Thus, the total probability of success of a firm pursuing innovation is $p_i + (1 - p_i) \delta q (1 - F_a)$. We impose no restriction on the second-chance parameter δ. If $\delta > 1$, the innovation investment facilitates the absorption of new ideas through random interactions, whereas if $\delta < 1$, focusing on innovation reduces the imitation potential.

2.1 Equilibrium dynamics with costless innovation

Consider first the case studied by König et al. (2016) in which innovation entails no investment cost and $\delta < 1$. Then, the manager chooses the strategy that maximizes TFP growth, as this also maximizes expected profit. In particular, firm i chooses the innovation strategy if and only if

$$p_i \geq Q(a, \tau; \Pi) \equiv \frac{q (1 - \delta)(1 - F_a)}{1 - \delta q (1 - F_a)}, \tag{4}$$

where Π denotes the productivity distribution. Since $\partial Q / \partial a < 0$, the proportion of innovating firms will be nondecreasing in the initial productivity. Intuitively, imitation is less effective for high-productivity firms because they are less likely to meet a more productive firm. Although thus far τ has no bearing on the innovation-imitation decision, we specify it as an argument of the function to prepare the analysis of the more general case where τ matters. Note that the ex-post productivity growth gap between innovating and imitating firms is increasing in the TFP level.[4]

The Law of Motion of Productivity: We can now write the law of motion of the distribution of log-productivity, $\Pi_a(t)$. Define the indicator function

$$\chi^{\text{im}}(a, p, \tau; \Pi) = 1 - \chi^{\text{in}}(a, p, \tau; \Pi) = \begin{cases} 1 & \text{if} \quad p \leq Q(a, \tau; \Pi), \\ 0 & \text{if} \quad p > Q(a, \tau; \Pi). \end{cases} \tag{5}$$

In plain words, χ^{im} is unity when the firm finds it optimal to imitate, while $\chi^{\text{in}}(a, p, \tau; \Pi)$ is unity when it finds it optimal to innovate. The law of motion for the productivity distribution is characterized by the following system of integro-difference equations:

[4] A larger a has two opposite effects on next period's (expected) productivity gap between innovating and imitating firms. On the one hand, it reduces the potential growth through imitation, thereby increasing the gap. On the other hand, it lowers $Q(a, \tau; \Pi)$, inducing firms with lower p to innovate. This negative selection is a second-order effect which is always dominated by the former effect.

$$
\begin{aligned}
&\Pi_a(t+1) - \Pi_a(t) \\
&= \int_0^{\overline{p}} \left[\begin{array}{c} \chi^{\text{in}}\left(a-1,p,\tau;\Pi\right) \times \left(p + (1-p)\,\delta q\left(1-F_{a-1}(t)\right)\right)\Pi_{a-1}(t) + \\ +\chi^{\text{im}}\left(a-1,p,\tau;\Pi\right) \times q\left(1-F_{a-1}(t)\right)\Pi_{a-1}(t) \\ -\chi^{\text{in}}\left(a,p,\tau;\Pi\right) \times \left(p+(1-p)\,\delta q\left(1-F_a(t)\right)\right)\Pi_a\left(t\right) \\ -\chi^{\text{im}}\left(a,p,\tau;\Pi\right) \times q(1-F_a(t))\Pi_a\left(t\right) \end{array} \right] dG\left(p\right) \qquad (6) \\
&= \int_{Q(a-1,\tau;\Pi)}^{\overline{p}} \left(p+(1-p)\,\delta q\left(1-F_{a-1}(t)\right)\right)\Pi_{a-1}(t)\;dG\left(p\right) \\
&+ G\left(\min\{Q\left(a-1,\tau;\Pi\right),\overline{p}\}\right) \times q\left(1-F_{a-1}(t)\right)\Pi_{a-1}(t) \\
&- \int_{Q(a,\tau;\Pi)}^{\overline{p}} \times \left(p+(1-p)\,\delta q\left(1-F_a(t)\right)\right)\Pi_a\left(t\right)dG\left(p\right) \\
&- G\left(\min\{Q\left(a,\tau;\Pi\right),\overline{p}\}\right) \times q(1-F_a(t))\Pi_a\left(t\right)
\end{aligned}
$$

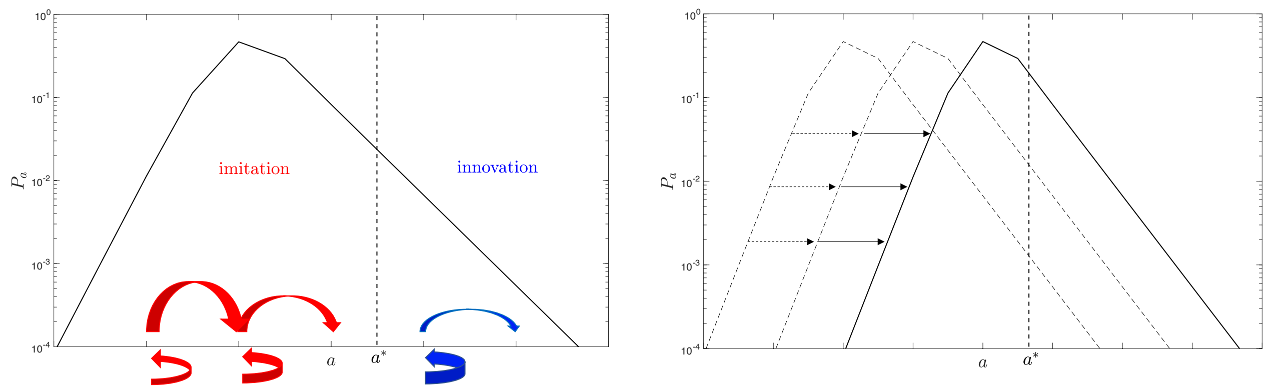
The first and second lines inside the first integral sign capture the inflow into productivity a of, respectively, successful innovating and imitating firms whose productivity was $a-1$ in period t. The third and fourth lines inside the integral capture the outflow of productivity a of, respectively, successful innovating and imitating firms whose productivity was a in period t. Note that for sufficiently low a, all firms imitate. In that case, $G = 1$ and the integrals in the expression vanish. Conversely, the share of imitating firms vanishes as $a \to \infty$.

Stationary distribution: Next, we characterize the stationary distribution associated with the system of difference equations. For ease of exposition, we first consider the special case of zero innovation cost for which a sharper analytical characterization is available.

Proposition 1 *Consider the model of innovation-imitation described in the text whose equilibrium law of motion satisfies Equation (6), where each firm draws p from a distribution $G : [0,\overline{p}] \to [0,1]$. Assume that $q > \hat{p}$, where $\hat{p} \equiv \int_0^{\overline{p}} p\; dG\left(p\right)$. Assume the cost of both imitation and innovation is equal to zero. Then, there exists a traveling wave solution of the form $\Pi_a(t) = f(a-\nu t)$ with velocity $\nu = \nu\left(q,\delta,g\left(p\right)\right) > 0$, with left and right Pareto tails. For a given t, Π_a is characterized as follows: (i) for a sufficiently large, $\Pi_a(t) = O\left(e^{-\rho(a-\nu t)}\right)$, where the exponent ρ is the solution to the transcendental equation $\rho\nu = \hat{p}\left(e^{\rho}-1\right)$; (ii) for a sufficiently small, $\Pi_a(t) = O\left(e^{\lambda(a-\nu t)}\right)$, 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. In particular, if we denote by $a^*(p,t)$ the threshold productivity such that, at time t and conditional on drawing p, all firms with productivity $a \leq a^*(p,t)$ imitate and all firms with productivity $a > a^*(p,t)$ innovate, then, $a^*\left(p,t+\Delta t\right) = a^*\left(p,t\right) + \nu\Delta t$. Thus, the function a^* is the inverse of the threshold function Q along the balanced growth path, $a^*(p,t) \equiv Q^{-1}\left(p,\tau,\Pi(t)\right)$ where $\Pi(t)$ is the stationary distribution at time t. The proof in the appendix 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 and 2009). Although our model features no reflecting barrier strictly speaking, on the one hand low productivity firms have a comparative advantage in imitation—in fact, firms with very low productivity imitate irrespective of the realization of p. On the other hand less productive firms are more likely to successfully imitate because they meet a more

Figure 1: Equilibrium Dynamics with a Stationary Distribution



Panel A Panel B

Note: Panel A displays the threshold a^* and the stationary TFP distribution. Panel B plots a traveling wave.

productive firm with a higher probability. Thus, the subdistribution of imitating firms catches up, which prevents the upper end of the distribution from diverging provided that q is sufficiently large.

A distinctive feature that sets aside our model from earlier is that the stationary distribution features a Pareto tail of low-productivity firms. The left tail originates from the fact that, although the probability of matching a better firm tends to unity at very low productivity levels, the probability of successful adoption is $q < 1$. This prevents the convergence of the subdistribution of imitating firms to a mass point. Figure 1 illustrates the equilibrium dynamics of the stationary distribution in a simplified version of the model in which all firms draw the same p. Panel A displays the threshold and the force implying convergence.[5] Panel B illustrates the traveling wave.

Proposition 1 yields no algebraic representation of the velocity of the traveling wave. In fact, ν can only be defined implicitly and solved for numerically. Numerical analysis shows that the growth rate is increasing in the parameters q, δ, and $\bar{p}$, implying that both the productivity of innovation and the rate at which ideas diffuse affect aggregate TFP growth.[6]

2.2 Equilibrium dynamics with costly innovation

Next, we generalize the analysis to an environment in which innovation requires a costly investment, which we label the R&D cost. The discounted value of profits is given by

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

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

[6] Analytical comparative statics can be obtained under a triangular approximation of the stationary distribution in a simpler version of the model with no heterogeneity in p. The analytical results under this approximation, which are consistent with the numerical results of the full model, are in the web appendix.

where

$$
c_i\left(t\right)=\begin{cases}\bar{c}\left(A_i\left(t\right)^{\theta}\bar{A}(t)^{1-\theta}\right)^{\eta-1} & \text{if} \quad i \text{ innovates,}\\ \\ 0 & \text{if} \quad i \text{ imitates.}\end{cases} \tag{7}
$$

Here, $\bar{A}(t)$ denotes the average productivity at time t and $\bar{c}>0$ and $\theta\in[0,1]$ are parameters. For simplicity, we set the discount factor to $\Xi=R^{-1}$, where R denotes the gross interest rate. This normalization entails no loss of generality since the equilibrium is pinned down by the ratio $\bar{c}/\Xi$.

We assume the innovation cost to be proportional to a geometric combination of A_i and $\bar{A}$. If $\theta=0$, the R&D investment is a fixed (overhead) labor cost (note that wages are proportional to $\bar{A}$) independent of firm-specific TFP. In the polar opposite case of $\theta=1$, the R&D investment is a variable cost in terms of managerial time whose opportunity cost is proportional to the firm's productivity. The general case captures in a flexible way a combination of fixed and variable costs. That flexibility is important in the empirical analysis because it improves the model's ability to match the heterogeneity in R&D costs relative to value added that we observe in the data. In this case, the function Q is modified as follows:

$$
Q\left(a,\tau_j;\Pi\right)=\frac{q\left(1-\delta\right)\left(1-F_a\right)}{1-\delta q\left(1-F_a\right)}+\frac{e^{(1-\theta)(\eta-1)(\overline{a}-a)}}{\left(e^{\eta-1}-1\right)\mathbb{E}\left[\left(1-\tau'\right)^{\eta-1}|\tau_j\right]}\frac{\bar{c}}{1-\delta q\left(1-F_a\right)}, \tag{8}
$$

where $\overline{a}\equiv\log\bar{A}$ and $\mathbb{E}\left[\tau'|\tau\right]$ denotes the conditional expectation of next-period wedge. The expression in Equation (8) is the same as that in Equation (4) except for the new second term. Note that although $\bar{a}$ trends over time, the second term in the right-hand side expression depends on $(\overline{a}-a)$, which is consistent with a stationary equilibrium.

There are two key differences relative to Equation (4). First, the R&D cost makes imitation more attractive *ceteris paribus*. Therefore, conditional on the realization of p, the threshold Q will be larger than in Equation (4). Second, the wedge affects the choice: a larger wedge τ_j deters innovation by reducing the future profit proportionally to TFP without affecting the R&D cost. More formally, $\partial Q/\partial\tau>0$: firms with higher wedges are less likely to engage in R&D. Note that the cost of innovation depends on the elasticity of substitution between varieties η, implying that the opportunity cost of R&D is larger when profits are higher. This assumption is necessary for balanced growth properties.[7]

The law of motion of productivity (cf. Equation (6)) must then take into account the heterogeneity in wedges:

$$
\begin{aligned}
&\Pi_a(t+1)-\Pi_a(t)\\
&=\sum_{j\in\{l,h\}}\omega_{\tau_j}\left(t\right)\times\int_0^{\overline{p}}\left[\begin{array}{c}\chi^{\mathrm{in}}\left(a-1,p,\tau_j;\Pi\right)\times\left(p+\left(1-p\right)\delta q\left(1-F_{a-1}(t)\right)\right)\Pi_{a-1}(t)+\\ +\chi^{\mathrm{im}}\left(a-1,p,\tau_j;\Pi\right)\times q\left(1-F_{a-1}(t)\right)\Pi_{a-1}(t)\\ -\chi^{\mathrm{in}}\left(a,p,\tau_j;\Pi\right)\times\left(p+\left(1-p\right)\delta q\left(1-F_a(t)\right)\right)\Pi_a\left(t\right)\\ -\chi^{\mathrm{im}}\left(a,p,\tau_j;\Pi\right)\times q(1-F_a(t))\Pi_a\left(t\right)\end{array}\right]dG\left(p\right),
\end{aligned} \tag{9}
$$

where $\omega_{\tau_l},\omega_{\tau_h}$ denote the proportion of low- and high-wedge firms, respectively. The model is closed by the law of motion for ω_{τ_l}. Let ρ_h and ρ_l denote the arrival rate of movements to τ_h and τ_l, respectively. The law of motion is then given by $\omega_{\tau_l}(t+1)-\omega_{\tau_l}(t)=\left(1-\rho_h\right)\left(1-\omega_{\tau_l}\left(t\right)\right)-\left(1-\rho_l\right)\omega_{\tau_l}\left(t\right)$, where ω_{τ_l} converges in the long run to $\bar{\omega}_{\tau_l}\equiv\left(1-\rho_h\right)/\left(\left(1-\rho_h\right)+\left(1-\rho_l\right)\right)$.

[7]Profits are proportional to $A_i^{\eta-1}$. Our specification in Equation (7) guarantees that the same is true for the innovation cost. Otherwise, secular changes in A_i would affect the incentives to do R&D making the model nonstationary. Our specification ensures that the relative level of A_i—but not its absolute value—matters for the R&D decision.

The next proposition characterizes the stationary equilibrium. The proof is an extension of the proof of Proposition 1 and is available from the web appendix.

Proposition 2 *The characterization of Proposition 1 carries over to a model with costly R&D investments where $\bar{c} > 0$. More formally, there exists a traveling wave solution of the form $\Pi_a(t) = f(a - \nu t)$ with velocity $\nu = \nu(q, \delta, G(p), \bar{c}, \tau_h, \tau_l, \bar{\omega}_{\tau_l}) > 0$, with left and right Pareto tails. Conditional on ν, the characterization of the tails is the same as in Proposition 1.*

Predictions of the theory: In summary, the model has four testable implications:

1. *ceteris paribus*, the proportion of firms engaged in R&D is increasing in TFP;
2. *ceteris paribus*, firms with higher wedges are less likely to engage in R&D. Then, Equation (2) implies that firms with higher sales are more likely to engage in R&D, even after conditioning on TFP;
3. expected TFP growth is falling in current productivity, especially so for nonR&D firms;
4. the gap in average TFP growth between R&D firms and nonR&D firms increases in TFP.

3 Data and descriptive evidence

We consider firm-level data for China and, in an extension, Taiwan. The Chinese data are from the Annual Survey of Industries conducted by China's National Bureau of Statistics for 1998–2007 and 2011–13. This survey is a census of all state-owned firms and the private firms with more than five million RMB (20 million RMB since 2010) in revenue in the industrial sector. To estimate firm-level productivity growth, we focus on a balanced panel for all manufacturing firms in 2007–12 including all firms that are in our sample in both 2007 and 2012. The data for R&D expenditure at the firm level are for the year 2007.[8] Although this is a firm-level survey, most of the Chinese firms were single-plant firms during this period. The Taiwanese data is at the plant level, collected by Taiwan's Ministry of Economic Affairs, for the years 1999–2004.[9] To make the Taiwanese sample more comparable to its Chinese counterpart, we drop the firms with annual sales below 18 million Taiwan dollars.

Table 1 reports summary statistics for the Chinese and Taiwanese balanced panels. Chinese firms are on average larger than Taiwanese firms. Part of the difference is accounted for by the Chinese state-owned enterprises (SOE). The fraction of firms reporting positive R&D expenditure in 2007 is 15% (data are not available after 2007). The corresponding fraction of R&D firms in the Taiwanese sample is 13% in 1999 and 10% in 2004.

We take investment in R&D as a proxy for the pursuit of an innovation strategy. We classify firms reporting positive R&D expenditure as innovators and all other firms as imitators. We test the robustness of the results to alternative classifications. We focus on the extensive margin of R&D for three reasons. First, it is consistent with the discrete-choice model we estimate. Second, there are important fixed costs of setting up an R&D lab, and only a small fraction of firms perform any R&D.

[8] 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.

[9] More than 90% of Taiwanese and Chinese manufacturing plants are owned by single-plant firms in the time periods we study. Following Aw et al. (2011), we ignore the distinction between plants and firms.

Table 1: Summary Statistics

Year	Number of Firms	Number of R&D Firms	Median Value Added (million USD)	Mean Value Added (million USD)	Median R&D Intensity (%)	Aggregate R&D Intensity (%)
			Balanced Panel of Chinese Firms			
2007	123368	18140	1.48	5.81	1.73	1.86
2012	123368	N.A.	3.33	11.45	N.A.	N.A.
			Private Chinese Firms in the Balanced Panel			
2007	117983	15828	1.43	4.67	1.65	1.54
2012	117771	N.A.	3.26	9.57	N.A.	N.A.
			Balanced Panel of Taiwanese Firms			
1999	11229	1487	0.16	2.91	8.50	3.14
2004	11229	1144	0.17	4.78	6.42	2.93

Note: R&D intensity is the ratio of R&D expenditure to value added. Median R&D intensity is the median R&D intensity among firms performing some R&D. Aggregate R&D intensity is the ratio of aggregate R&D expenditure to aggregate value added for all firms. Missing information is due to the lack of R&D data in 2012.

Third, the intensive margin is subject to a more severe measurement error.[10] We also discuss how the main insights of our analysis carry over to considering an intensive margin of the R&D choice.

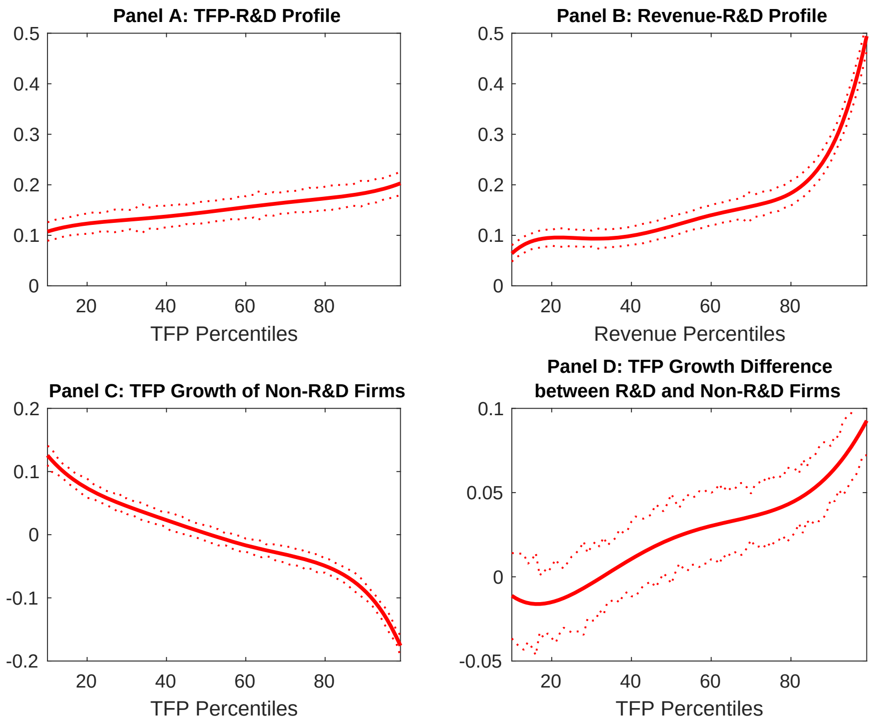
Figure 2 shows the distribution of R&D and TFP growth conditional on TFP in the initial year and conditional on firm size (measured by value added). We estimate TFP following the methodology proposed by Hsieh and Klenow (2009). This requires a calibration of the production parameter α and the demand elasticity η. We allow α to vary across industries and set α_j in (two-digit) industry j equal to the measured industry-specific labor income share. Following Hsieh and Song (2015), we set $\eta = 5$. We have also estimated TFP using the methodology of Ackerberg et al. (2015) following the implementation proposed by Brandt et al. (2017) who estimate production functions for Chinese manufacturing industries. The empirical moments we target are very similar when we use this alternative procedure—see Appendix Figure A1. To control for observable sources of heterogeneity, we regress TFP on province, industry, and age dummies and take the residual as the measure of firm TFP.

The industry classification refers to 30 two-digit manufacturing industries. We normalize firm-level value added by the median value in the industry to which each firm belongs. We do not explicitly separate R&D expenditure when estimating TFP. This could potentially bias the TFP estimates for R&D firms. The problem has no perfect solution because we do not have R&D data after 2007. To assess its importance, we adjusted TFP in the earlier period 2001–07 by subtracting R&D expenditure

[10] This issue has been noted in the literature that studies firm-level R&D expenditure in Western countries (see, e.g., Lichtenberg 1992, Acemoglu et al., 2010). In China, it is especially severe because R&D expenditure is vaguely defined in China's industrial survey. For instance, the survey does not distinguish between R&D performed and R&D paid for by the firm.

from labor costs. Then, we plotted a version of Figure 2 based on the adjusted data. The empirical moments are almost indistinguishable from the original figure. We conclude that the problem is likely quantitatively small.

Figure 2: Chinese Firms in the Balanced Panel 2007–12

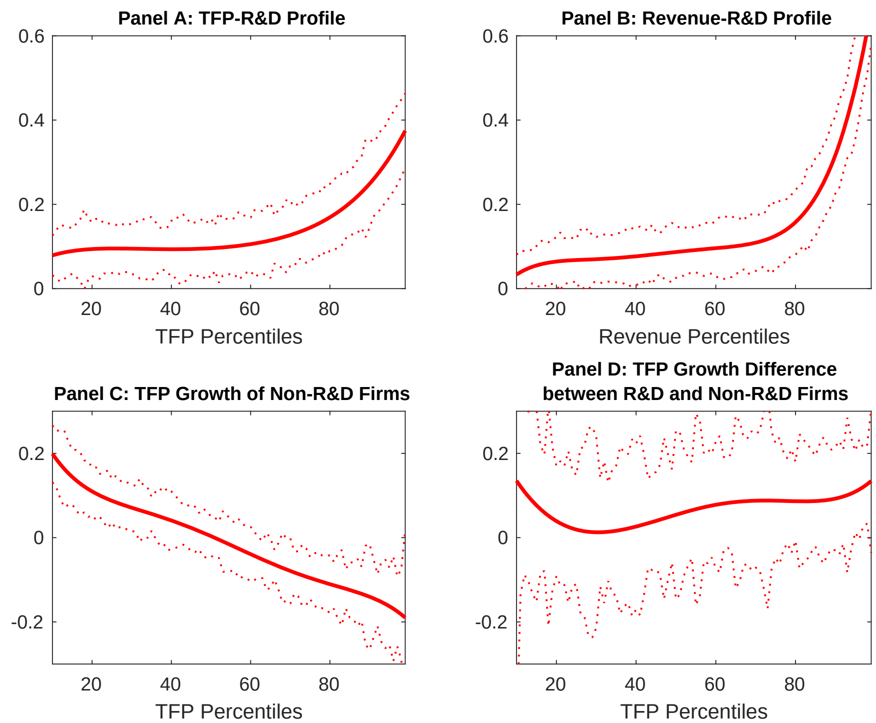


Note: The X-axis in Panels A, C, and D is the 2007 TFP percentile. The X-axis in Panel B is the 2007 value added percentile. The solid lines in Panel A and B plot the 2007 fraction of R&D firms in each TFP and value added percentile, respectively. The solid line in Panel C plots the median annualized 2007–12 TFP growth among nonR&D firms in each TFP percentile. The solid line in Panel D plots the difference between the median 2007–12 TFP growth R&D and nonR&D firms within each percentile. A firm's TFP growth is the residual of the regression of TFP growth on industry, age, and province fixed effects. All the solid lines are smoothed by a fifth-order polynomial. The dotted lines plot the 95% confidence intervals by bootstrap.

Panel A shows the share of R&D firms by TFP percentile. The positive correlation is in line with the prediction of our theory that more productive firms do more R&D. The share of R&D firms increases from 11.6% in the lowest decile to 20% in the top percentile of the TFP distribution. Panel B shows that firm size is also positively correlated with the share of R&D firms. The relationship is significantly steeper than in Panel A: almost 50% of the firms in the top percentile of the size distribution invest in R&D. Since larger firms are on average more productive, TFP is a driver of both panels. However, the steeper profile in Panel B indicates that factors other than TFP must matter. In our model, a firms' size is determined by the product of its TFP and (one minus its) wedge. Thus, firms subject to positive (negative) wedges are smaller (larger) than what their TFP alone would predict. The wedge also affects the profit and, hence, the incentive to pursue an innovation strategy. Note that, in the absence of wedges, Panels A and B would be identical. In the presence of wedges, we expect Panel B to be steeper than Panel A which is consistent with the empirical observation.

Panels C and D show relationships between TFP growth and the distribution of initial TFP. Panel C shows that the TFP growth rate is decreasing in TFP among nonR&D firms. In other words, there

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



Note: This figure is the analogue of Figure 2 for the sample of Taiwanese firms.

is strong convergence in productivity across nonR&D firms. This is consistent with the main tenet of our theory that learning through random interactions and imitation is easier for less productive firms. A concern is that the negative relationship might partly be due to a survivor's bias: low-performing firms are more likely to exit causing the TFP growth of the surviving firms to be higher. Since this problem is most important for low-TFP firms, we trim the lower tail of the distribution. Another concern is measurement error in TFP. If measurement error is classic, firms with a negative (positive) measurement error at t are overrepresented among low- (high-)TFP firms at t. Reversion to the mean would then exaggerate the convergence pattern. In the estimation section below, we model measurement error explicitly and allow it to influence the graph in Panel C.

Finally, Panel D compares the TFP growth for R&D firms and nonR&D firms at different percentiles of the TFP distribution. In line with the prediction of our theory, TFP growth is higher for R&D firms than for nonR&D firms at most percentiles. The observation that the pattern is reversed for firms with very low TFPs can be rationalized by extending our theory with heterogeneity in R&D subsidies, as we discuss below.

The same patterns emerge from a set of multiple regressions whose results are reported in Table 2. Panels A and B of Table 2 are related to Panels A–B and Panels C–D of Figure 2, respectively. Panel A shows the results for a linear probability model whose dependent 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 industry level. We also include provincial dummies and, in column (4), dummies for SOEs. The table shows that the fraction of R&D firms is robustly correlated with the log of TFP. The estimated coefficient increases significantly when we include an estimated output

Table 2: Balanced Panel of Chinese firms, 2007–2012.

PANEL A

Dependent variable: R&D decision in 2007.

	(1) $\text{R\&D}_d$	(2) $\text{R\&D}_d$	(3) $\text{R\&D}_d$	(4) $\text{R\&D}_d$
log(TFP)	0.062***	0.368***	0.343***	0.305***
	(0.0074)	(0.0284)	(0.0259)	(0.0232)
wedge		-0.410***	-0.378***	-0.332***
		(0.0357)	(0.0323)	(0.0296)
export_d			0.053***	0.054***
			(0.0134)	(0.0132)
SOE_d				0.205***
				(0.0232)
R-squared	0.143	0.208	0.211	0.224

PANEL B

Dependent variable: TFP growth.

	(1) TFP growth	(2) TFP growth	(3) TFP growth	(4) TFP growth	(5) TFP growth
log(TFP)	-0.062***	-0.062***	-0.062***	-0.062***	-0.062***
	(0.0035)	(0.0036)	(0.0035)	(0.0036)	(0.0036)
$\text{R\&D}_d$	0.036***	0.037***	0.034***		
	(0.0042)	(0.0040)	(0.0035)		
export_d		-0.006	-0.006*	-0.006	-0.007*
		(0.0038)	(0.0035)	(0.0037)	(0.0035)
SOE_d			0.029**		0.029**
			(0.0115)		(0.0113)
R&D intensity_h				0.044***	0.041***
				(0.0060)	(0.0058)
R&D intensity_m				0.042***	0.038***
				(0.0069)	(0.0056)
R&D intensity_l				0.025***	0.023***
				(0.0035)	(0.0033)
R-squared	0.122	0.122	0.123	0.122	0.123

Note: Panel A: the dependent variable $\text{R\&D}_d$ is a dummy variable switching on for firms that report positive R&D expenditure. Panel B: the dependent variable is annualized TFP growth 2007–12. Explanatory variables: log(TFP) is the logarithm of TFP; Wedge indicates the firm's output wedge $-\log(1-\tau_i)$ and is calculated as described in Section 2; export_d is a dummy variable for exporters; SOE_d is a dummy variable for state-owned firms; R&D intensity_h is a dummy variable for high R&D intensity switching on if the firm R&D expenditure over sales is in the 67th percentile and above (among all R&D firms); R&D intensity_m is the analogue dummy for medium R&D intensity (between 33rd and 66th percentiles); R&D intensity_l is the analogue dummy for low R&D intensity (below the 33rd percentile.) All the explanatory variables are from 2007. Standard errors are reported in parenthesis. The number of observations is 109,799. Observations are weighted by employment and standard errors are clustered by industry. All regressions include industry, age, and province fixed effects. We drop firms with TFP in the bottom 10 percentiles.

wedge among the regressors.[11] Both the positive correlation with TFP and the negative correlation with the output wedge line up with the predictions of the theory. A large output wedge discourages firms from investing in R&D by reducing profits. Columns (3)–(4) show that the results are not driven by exporting firms nor SOEs.

We cannot include firm fixed effects in this regression analysis because we have data on R&D investments only for 2007. However, we have performed the same analysis on an earlier sample (2001–07) for which R&D information is available in both the initial and final year. Including firm fixed effect implies that the effect of R&D are identified by firms adopting (or dropping) R&D over time. The results for the 2001–07 panel—which we report in Appendix Table A1—are qualitatively similar to those reported in Table 2: as firms become more productive over time they become more likely to perform R&D.

Panel B reports the results of regressions with average TFP growth during 2007–12 as the dependent variable. TFP growth is regressed on the initial log-TFP level and on an R&D dummy in 2007. The tables show a robust negative correlation between TFP growth and initial TFP (consistent with Panel C of Figure 2) and a robust positive correlation between TFP growth and an R&D dummy (consistent with Panel D of Figure 2). The results are robust to controlling for SOE and export firm dummies.

Columns (4) and (5) focus on the intensive margin of R&D by breaking the R&D dummy into three separate dummies, one per each tercile of R&D expenditure. All three dummies are both statistically and economically significant. Reassuringly, a higher (lower) investment in R&D is associated with a higher (lower) future TFP growth. The growth difference between the upper and lower terciles is statistically significant.[12]

Appendix Figure A2 shows that the patterns in Figure 2 are robust to a more stringent classification counting only those firms with R&D-to-value added ratio exceeding 1.73% (median among R&D firms) as innovative. While this criterion by construction reduces the share of innovative firms, the four panels are qualitatively similar. The same is true for a version of the multiple regressions in Table 2 where we apply the more stringent classification of R&D firms. The sign of the coefficients of interest is the same as in Table 2 and all coefficients are highly significant. Details are available upon request.

Another potential concern is that the results might 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 top five R&D-intensive industries. We also find that the results are similar when partitioning the sample into subgroups: exporting versus nonexporting firms, SOEs versus non-SOEs, and sorting firm by regions.

The empirical patterns are similar in Taiwan for both the intensive and the extensive margin—see Figure 3 and Appendix Table A2—with two noteworthy quantitative differences. First, the R&D-TFP profile in Panel A is steeper in Taiwan than in China. Moving from the 60th to the 99th TFP percentile the share of Taiwanese R&D firms increases from 10% to over 35%. Second, Panel D has very large standard errors. However, the regression analysis in Table A2 establishes that there is a robust and highly significant positive correlation between R&D and future productivity growth similar to China.

[11] The measurement of output wedges—which follows Hsieh and Klenow (2009)—is discussed in Section 4. Here, we note that measurement error might exaggerate the negative correlation between the estimated TFP and the output wedge, driving part of the strong opposite-sign pattern for the estimated coefficients in Table 2. We address this issue in our structural estimation below.

[12] We also run separate regressions like in Panel A to detect whether the selection into high versus low R&D intensity is driven by the TFP level. However, the data show no clear pattern.

4 Estimation

We estimate the model assuming a closed-economy environment. We classify firms performing R&D as pursuing an innovation strategy. The other firms are imitators. International knowledge spillovers are not part of the model, although they are partially incorporated through the activity of multinational firms producing in China which are part of the sample (cf. Holmes et al. (2015)). In Section 7 we study an extension explicitly incorporating international spillovers.

The estimation targets two sets of moments: moments that are informative about the economic mechanism of the model and moments that are informative about measurement error. For the former, we focus on a set of selected percentiles in Figure 2. We consider four intervals of the distribution of TFP and size in each of the four panels: the 10th to 49th percentile, the 50th to 79th percentile, the 80th to 94th percentile, and the top five percentiles.[13] This choice yields sixteen empirical target moments. Appendix Figure A3 plots the confidence intervals around these moments.

We calibrate the parameters α, η, θ, and $\tilde{a}$, and structurally estimate the remaining parameters using simulated method of moments. The calibration of α and η, discussed above, is external to the model. We return to the calibration of θ and $\tilde{a}$ when discussing the technology of innovation below.

4.1 Measurement error and wedges

Measurement error: Measurement error (m.e.) is a common concern in models of misallocation *à la* Hsieh and Klenow (2009) because it potentially affects both the measured moments of TFP and the imputed wedges. In our model, m.e. affects the target moments in Appendix Figure A3. On the one hand, it generates an attenuation bias in the relationships between the propensity to engage in R&D and both TFP (Panel A) and size (Panel B), flattening both profiles. On the other hand, it exaggerates TFP convergence in Panel C, steepening the profile. We now propose an explicit model of m.e. and discuss its estimation.

We assume that value added and inputs (capital and labor) are measured with classical m.e.:

$$\begin{aligned} \ln\left(\widehat{P_i Y_i}\right) &= \ln\left(P_i Y_i\right) + \mu_y, \\ \ln\left(\widehat{K_i^{\alpha} N_i^{1-\alpha}}\right) &= \ln\left(K_i^{\alpha} N_i^{1-\alpha}\right) + \mu_I, \end{aligned}$$

where μ_y and μ_I are i.i.d. measurement errors with variances $\hat{v}_y$ and $\hat{v}_I$. The notation with hats denotes observed variables, while no hat denotes true variables.

We make the key identifying assumption that the firm-specific wedges τ_i are constant over the unit of time we consider, that is, the 2007–12 period.[14] Under this assumption, the time series variation (2007–12) in value added and input measures at the firm level can be used to infer the extent of m.e. Equations (2) and (3) imply that inputs are proportional to value added times the output wedge, i.e., $(1-\tau_i)(P_{it}Y_{it}) \propto (r_t)^{\alpha}(w_t)^{1-\alpha}(K_{it})^{\alpha}(L_{it})^{1-\alpha}$. Then, a constant τ_i implies that

$$\Delta \ln\left[(K_{it})^{\alpha}(L_{it})^{1-\alpha}\right] = \Delta \ln\left(P_{it}Y_{it}\right) - \Delta \ln\left[(r_t)^{\alpha}(w_t)^{1-\alpha}\right],$$

[13] We zoom on the upper tail of the distribution because large and high-productivity firms are more likely to engage in R&D. We abstract from the first decile to mitigate concerns about survivors' bias.

[14] In the data, the variance of $\ln(1-\tau_i)$ is approximately constant, increasing slightly from 0.724 in 2007 to 0.769 in 2012 in the balanced sample. In the full sample, the dispersion declines slightly, from 0.806 to 0.794.

where $\Delta \ln X_t \equiv \ln X_t - \ln X_{t-1}$. The variance of (true) value added growth can then be identified as follows:

$$\begin{aligned} & cov\left(\Delta \ln\left(\widehat{P_{it}Y_{it}}\right), \Delta \ln\left(\widehat{(K_{it})^{\alpha}(L_{it})^{1-\alpha}}\right)\right) \\ = \; & cov\left(\Delta \ln(P_{it}Y_{it}) + \Delta\mu_{yt}, \Delta \ln(P_{it}Y_{it}) - \Delta \ln\left[(r_t)^{\alpha}(w_t)^{1-\alpha}\right] + \Delta\mu_{It}\right) \\ = \; & var\left(\Delta \ln(P_{it}Y_{it})\right). \end{aligned} \tag{10}$$

The second equality follows from the assumptions that m.e. is classical—implying that $\Delta\mu_{yt}$ and $\Delta\mu_{It}$ are white noise—and that the input price $(r_t)^{\alpha}(w_t)^{1-\alpha}$ is identical across firms. Therefore, the cross-sectional covariance is not affected by the aggregate input price growth. The variance of m.e. in value added and inputs can then be identified as

$$var\left(\Delta \ln\left(\widehat{P_{it}Y_{it}}\right)\right) - cov\left(\Delta \ln\left(\widehat{P_{it}Y_{it}}\right), \Delta \ln\left(\widehat{(K_{it})^{\alpha}(L_{it})^{1-\alpha}}\right)\right) = 2\hat{v}_{\mu y} \tag{11}$$

$$var\left(\Delta \ln\left(\widehat{(K_{it})^{\alpha}(L_{it})^{1-\alpha}}\right)\right) - cov\left(\Delta \ln\left(\widehat{P_{it}Y_{it}}\right), \Delta \ln\left(\widehat{(K_{it})^{\alpha}(L_{it})^{1-\alpha}}\right)\right) = 2\hat{v}_{\mu I}. \tag{12}$$

We measure the empirical covariances in Equations (11) and (12) using data on growth rates in revenue and inputs between 2007 and 2012—see Appendix Table A3. M.e. in revenue and inputs translates into m.e. in TFP. Equation (3) implies that m.e. in TFP is $\hat{a} - a = [\eta/(\eta-1)]\,\mu_y - \mu_I$, where $a \equiv \log A$. The variance of m.e. in TFP is, then, $\hat{v}_{\mu a} = [\eta/(\eta-1)]^2\,\hat{v}_{\mu y} + \hat{v}_{\mu I}$. We set $\eta = 5$ and use the empirical $\hat{v}_{\mu a}$ as a target moment in the joint estimation of the parameters of the model. In the appendix we characterize analytically the effect of m.e. on the moments of the model.[15] This analytical mapping speeds up computations significantly so that estimation becomes feasible. Adding m.e. by way of simulations would be time-consuming and would be a computational curse for the estimation.

Distribution of output wedges: The estimated output wedges and TFP are correlated. To keep consistency between the simulated model and the data, we assume a distribution of output wedges that has the same correlation between output wedges and TFP as in the empirical distribution. This correlation affects the size of aggregate distortions as discussed by Restuccia and Rogerson (2008). More formally, we assume the following relationship:

$$-\ln(1-\tau_i) = b \cdot (a_{it} - a_t) + \varepsilon_{it}^{\tau}, \tag{13}$$

where a_t is the mean of a_{it}, and we assume that $\varepsilon_{it}^{\tau} \sim \mathcal{N}(0, var(\varepsilon_{it}^{\tau}))$. We are interested in estimating the coefficient b in Equation (13) and $var(\varepsilon_{it}^{\tau})$. Estimating b by OLS yields a biased estimate because of m.e. However, the m.e. model above implies the following unbiased estimate of b,

$$b = \frac{cov(a_{it}, -\ln(1-\tau_i))}{var(a_{it})} = \frac{cov(\hat{a}_{it}, -\ln(1-\hat{\tau}_i)) - \left(\frac{\eta}{\eta-1}\right)\cdot \hat{v}_{\mu y} - \hat{v}_{\mu I}}{var(\hat{a}_{it}) - \left(\frac{\eta}{\eta-1}\right)^2 \hat{v}_{\mu y} - \hat{v}_{\mu I}}. \tag{14}$$

[15] In principle, we could estimate two of three empirical variances $\hat{v}_{\mu a}$, $\hat{v}_{\mu y}$, and $\hat{v}_{\mu I}$, (the third being a combination of the other two) in the estimation. However, the variances must be constrained to be non-negative. In the estimation of the benchmark model the constraint $\hat{v}_{\mu I} \geq 0$ turns out to be binding. This holds true for all data samples we consider. To keep the number of estimated parameters low, we impose $\hat{v}_{\mu I} = 0$ and set $\hat{v}_{\mu y} = [(\eta-1)/\eta]^2\,\hat{v}_{\mu a}$ when adding m.e. to the empirical moments.

Then, equation (13) implies $var\left(\varepsilon_{it}^{\tau}\right) = var\left(\ln\left(1-\tau_i\right)\right) - b^2 var\left(a_{it}\right)$, where $var\left(\ln\left(1-\tau_i\right)\right) = var\left(\log\left(1-\hat{\tau}_i\right)\right) - \hat{v}_{\mu y} - \hat{v}_{\mu I}$ and, by construction, $var\left(\hat{a}_{it}\right) = var(a_{it}) + \hat{v}_{\mu a}$. The resulting unbiased estimates are $b = 0.779$ and $var\left(\varepsilon_{it}^{\tau}\right) = 0.042$ (compared to biased OLS estimates of 0.802 and 0.047, respectively).[16] Moreover, one third of the variance of measured wedges is due to m.e.

4.2 The technology of R&D

Cost function. The parameter θ in Equation (7) captures the elasticity of a firm's innovation cost to its TFP. We calibrate this parameter by targeting the *relative* cross-sectional distribution of the R&D cost-to-value added ratio. Formally, we target the ratio $\mathbb{E}\left[\psi_j|a_j\right]/\mathbb{E}\left[\psi_i|a_i\right]$, where ψ denotes the ratio of innovation costs to value added and a_i and a_j denote TFP in the i'th and j'th percentile. This ratio can be expressed analytically as

$$\frac{\mathbb{E}\left[\psi_j|a_j\right]}{\mathbb{E}\left[\psi_i|a_i\right]} = \exp\left(\frac{1-\eta}{\eta}\left(1-\theta+b\right)\left(a_i-a_j\right)\right). \tag{15}$$

We use data on actual R&D costs and consider the R&D-to-value added ratio for firms in the top five percent of the TFP distribution relative to firms in the 10th-49th percentile. Empirically, this ratio declines with TFP, being 37% higher for low-TFP R&D firms (10-49th percentile) than for high-TFP (hence, larger) firms (top five percentiles). The parameter b is adjusted for m.e. in line with Equation (14). Equation (15) then implies an elasticity of $\theta \approx 0.25$. Appendix Figure A4 shows that this model fits the slope of the relationship between R&D intensity and TFP percentiles in the data well.

Step size. In the model, firms that are successful in either innovating or imitating increase their log productivity by a step size $\tilde{a}$. The choice of $\tilde{a}$ has no appreciable effect on the model's ability to fit the cross-sectional data of Figure 2. However, it affects the estimate of the parameter $\bar{c}$. The reason is that, in the theory, profits are an increasing convex function of $\tilde{a}$—see Equation (1). Because the fraction of R&D firms is a target of the estimation, a larger $\tilde{a}$ induces a larger estimated value of $\bar{c}$ (together with a lower estimate of $\bar{p}$).

Ideally, one could estimate $\tilde{a}$ jointly with the other parameters of the model. However, this is computationally infeasible. Instead, we set $\tilde{a}$ so as to target a realistic average cost of innovation as a share of value added. In particular, we set $\tilde{a} = 0.78$ which implies—conditional on the estimated parameters—an average cost of innovation of 3.7% of the industrial value added in the benchmark economy.[17] This ratio is about twice the aggregate R&D-to-value added ratio in the Chinese data. We view this as a realistic target in light of the innovation management literature documenting that formal R&D is only a part of the costs incurred by firms pursuing innovation. The purchase of new equipment often reflects the introduction of new technologies although it is recorded as capital investment. Hiring STEM workers is another facet of an innovation-oriented strategy. Finally, in an innovative firm, managers divert more of their attention to the introduction of new products or more efficient processes.[18]

[16] Note that the distribution of τ_i is by construction consistent with Equation (13). Hence the variance of τ_i depends on both b, ε_i^{τ}, and the variance of A—something we return to in Section 7. When we simulate the model, the firm-specific wedge τ is drawn each period in line with (13) and with an i.i.d. draw of ε_{it}^{τ}. Since firm-specific TFP is highly autocorrelated and $b \neq 0$, the output wedges are positively autocorrelated.

[17] This calculation assumes a risk-adjusted discount rate of 10% which we view as reasonable given pervasive financial and contractual imperfections in China and the high systematic risk of innovative activities.

[18] A coauthor of a study of 40 companies over 25 years (Colarelli O'Connor et al. (2018)) summarizes the findings as follows: "Innovation is much bigger than R&D. It involves three distinct capabilities: Discovery, Incubation, and Acceleration (DIA). R&D is just one part of the Discovery capability – invention." The other activities "often require[s] as much time and resources and deserves as much emphasis, as inventing the technologies themselves" (Colarelli O'Connor (2019)).

Productivity of innovation. We assume that firms draw p from an i.i.d. uniform probability distribution with support $[0, \bar{p}]$, where $\bar{p}$ is structurally estimated.

4.3 Simulated method of moments (SMM)

We estimate the remaining parameters using SMM (McFadden 1989). The estimated parameters minimize the distance between the target moments and the stationary distribution of the model. Analytically tractability is key for our procedure. We simulate the model under some parameter configuration, add m.e. to the moments, and calculate the distance from the targeted empirical moments. Then, we iterate on the parameter vector. The system of ordinary difference equations allows us to calculate the stationary distribution efficiently. We could in principle have simulated the distribution of a large number of firms for every trial of a parameter configuration. However, such alternative approach would be computationally too demanding.

In all our trials, the numerical simulations converge to a unique stationary distribution irrespective of initial conditions, provided that the learning parameter q is not too small. When this parameter is sufficiently close to zero, there exists no ergodic distribution.

The sample is randomly generated by bootstrapping for $K = 500$ times. Denote by $g_{m,k}$ the mth moment in the kth sample and by $g_m(\phi)$ the vector of the corresponding moments in the model, where ϕ is the vector of parameters that we estimate. We minimize the weighted sum of the distance between the empirical and simulated moments:

$$\hat{\phi} = \arg\min_{\phi} h(\phi)' \ W \ h(\phi),$$

where W is the moment weighting matrix and $h_m(\phi) = \left[g_m(\phi) - \frac{1}{K}\sum_k^K g_{m,k}\right] / g_m(\phi)$ is the percentage deviation between the theoretical and empirical moments, averaged across K samples. We use the identity matrix as the benchmark weighting matrix to avoid the potential small-sample bias—see Altonji and Segal (1996).

5 Results

We first estimate a benchmark parsimonious model that reproduces the theoretical model without any added new feature. Then, we consider some extensions. We keep the parameters θ and $\tilde{a}$ constant across specifications at the level calibrated to the benchmark economy. We focus on China and study Taiwan for comparison.

5.1 Parsimonious model

In the benchmark model—which we label the *Parsimonious* model, henceforth, PAM—we estimate five parameters: q, $\bar{c}$, $\bar{p}$, δ, and $v_{\mu a}$. The innovation cost is $c_i(t) = \bar{c} \cdot \left(A_i(t)^{\theta} \bar{A}(t)^{1-\theta}\right)^{\eta-1}$. Before turning to the results, we summarize the sources of identification of the structural parameters.

Identification. The parameter q is the probability of successful imitation conditional on meeting a more productive firm. It is mostly pinned down by the TFP convergence rate across imitating firms (Panel C of Figure 2) conditional on m.e. The parameter δ (passive imitation) is identified by the TFP convergence rate across both imitating and innovating firms (Panels C and D of Figure 2). Lack of convergence within the set of innovating firms implies that δ is close to zero, while strong convergence implies that δ is large. The parameter $\bar{p}$ is pinned down by the extent to which, conditional on TFP, innovating firms grow faster than imitating firms (Panel D of Figure 2). Given the other parameters, the innovation cost parameter $\bar{c}$ is disciplined by the total share of innovating firms. Finally, the standard deviation of m.e. affects Panels A, B, and C in Figure 2. Measurement error flattens the schedules in Panels A and B while steepening the schedule in Panel C. In other words, m.e. creates the impression of a stronger convergence in the data than there is in reality. Thus, a larger estimate of $v_{\mu a}$ implies a lower estimate of q.

Results. The estimation results are displayed in column (1) of Table 3. The estimated coefficient $q = 0.175$ implies that there is significant convergence in TFP even after removing m.e. The estimated δ is close to zero indicating that R&D has a high opportunity cost in terms of foregoing learning through random interactions. The estimated average probability of p is about 4.8% (i.e., $\bar{p}/2 = 0.096/2$.) Given the costs and benefits of the two strategies, the model predicts that ca. 12% of the firms invest in R&D which compares with 15% in the data. The estimated variance of m.e. in TFP is $\sigma_{\mu a}^2 = 0.3$. This implies that m.e. accounts for 30% of the variance of log TFP and 92% of the variance of TFP growth.

Figure 4 shows that the PAM fits the data fairly accurately, matching well the convergence pattern in Panel C and the differential growth between R&D and nonR&D firms in Panel D.[19] However, the model predicts too steep profiles for R&D-TFP (Panel A) and for revenue-TFP (Panel B).

5.2 Heterogeneous innovation costs

In this section, we allow for heterogeneity in the innovation cost parameter $\bar{c}$. Heterogeneity could arise from technological factors or from additional wedges that directly distort the imitation-innovation decision. These range from R&D subsidies to government investments in local infrastructure, science parks, to credit constraints—particularly severe for R&D investments.

Figure 4 shows that the models with heterogeneous innovation fit more accurately the target moments, especially Panel B. Intuitively, the imitation-innovation decision now depends also on the realization of c_i, thereby reducing the importance of TFP and size. This flattens the schedules in Panels A and B that were too steep in the PAM.

We consider three specifications. In the first—that we label *Flexible* model, henceforth, FLM—innovation wedges are i.i.d. across firms. In the second, we allow the wedges to be correlated with observable firm characteristics, capturing the idea that local and central governments may target their support to firms with certain characteristics. We label this specification the *Industrial Policy* model, henceforth, IPM. Finally, we consider a specification where some firms can strategically misreport innovation expenditures in order to attract subsidies without actually distorting their optimal imitation-innovation decisions. We label this model the *Fake R&D* model, henceforth, FRM.

Flexible model: In the FLM, the effective innovation cost is given by

$$c_i(t) = \left[\bar{c} - \exp\left(\xi_i(t) - \frac{\sigma_c}{2}\right) + 1\right] \cdot \left(A_i(t)^{\theta} \bar{A}(t)^{1-\theta}\right)^{\eta-1},$$

where $\xi_i \sim \mathcal{N}\left(0, \sigma_c^2\right)$. Note that $\mathbb{E}\left[\bar{c} - \exp\left(\xi_i(t) - \sigma_c/2\right) + 1\right] = \bar{c}$, so σ_c is a mean-preserving spread.

Figure 5 shows that the fit of the FLM improves upon that of the PAM. This is reflected in a lower residual sum of squares, mostly attained in Panel B. The estimated parameters are in the ballpark

[19] M.e. has a significant effect on the estimates. Ignoring it would increase the estimates for q and δ to $q = 0.702$ and $\delta = 0.500$, implying a faster convergence. The reason is that in the absence of m.e., Panel C—TFP growth conditional on TFP for nonR&D firms—dictates a fast catching up rate for low-TFP firms, as discussed above.

Table 3: Estimation, Chinese Firm Balanced Panel 2007–2012.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					Higher R&D cutoff		High & Low R&D	
	PAM	FLM	IPM	FRD	PAM	IPM	PAM	IPM
Imitation prob. q	0.175	0.271	0.361	0.275	0.294	0.546	0.223	0.359
	(0.031)	(0.019)	(0.019)	(0.051)	(0.058)	(0.052)	(0.026)	(0.052)
Second chance δ	0.008	0.020	0.001	0.019	0.141	0.001	0.058	0.019
	(0.011)	(0.021)	(0.027)	(0.024)	(0.106)	(0.080)	(0.027)	(0.026)
Innov. prod. $\bar{p}$	0.096	0.114	0.113	0.237	0.107	0.111	0.103	0.107
	(0.008)	(0.006)	(0.006)	(0.016)	(0.013)	(0.010)	(0.007)	(0.012)
$\bar{p}_l$							0.076	0.115
							(0.009)	(0.010)
Innov. cost $\bar{c}$	1.627	3.374	1.318	9.480	3.601	8.393	3.015	2.075
	(0.136)	(0.174)	(0.177)	(1.363)	(0.448)	(2.403)	(0.208)	(0.791)
$\bar{c}_l$							0.094	0.612
							(0.135)	(0.979)
Std.dev. m.e. $\sigma_{\mu a}$	0.549	0.472	0.431	0.459	0.476	0.391	0.531	0.433
	(0.014)	(0.008)	(0.005)	(0.025)	(0.022)	(0.011)	(0.015)	(0.024)
Std.dev.innov. subs. σ_c		1.243	1.092	0.011		1.969		1.206
		(0.038)	(0.037)	(0.036)		(0.213)		(0.113)
Policy inter. c_a			1.888	-10.46		2.499		2.015
			(0.159)	(1.042)		(0.356)		(0.316)
Fake share Υ				0.099				
				(0.005)				
High $\bar{p}$ share							0.872	0.682
							(0.029)	(0.047)
J-Statistic	1.518	0.507	0.368	0.362	2.690	0.516	3.310	0.882

Note: The table shows the estimated parameters for the different models discussed in the text. Columns (1)–(4) are for the Parsimonious (PAM), Flexible (FLM), Industrial Policy (IPM), and Fake R&D (FRD) model, respectively. Columns (5)–(8) presents the results for extensions discussed in the text. Bootstrapped standard errors in parentheses.

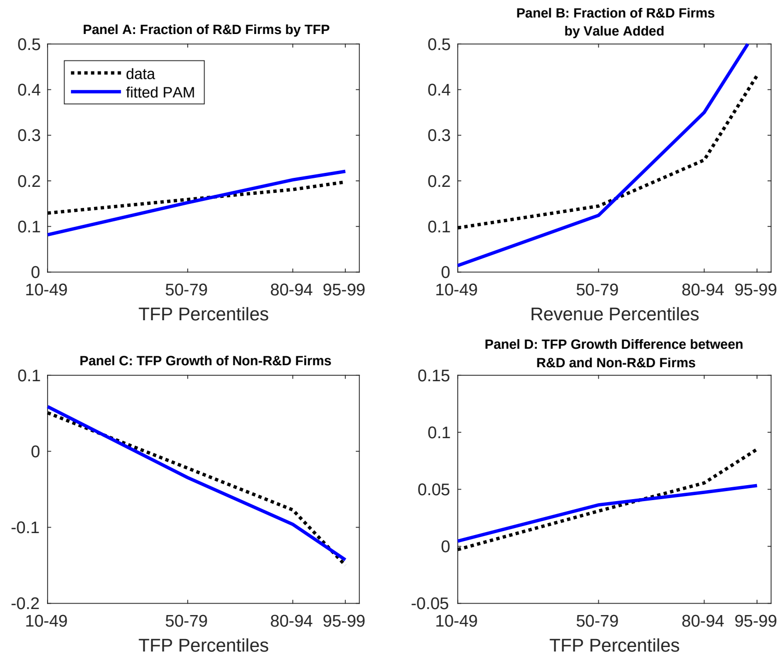
of the PAM estimates with two noteworthy differences. First, the estimate of $\bar{c}$ is larger than in the PAM. The reason is selection: *ceteris paribus*, subsidized firms (some of them facing a negative effective innovation cost) have a stronger incentive to pursue an innovation strategy. With and unchanged $\bar{c}$ too many firms want to do R&D. To match the empirical share of R&D firms, the model requires a larger average innovation cost. Second, the estimate of m.e. is now lower because the R&D cost dispersion flattens the TFP-size profile in Panel D, mitigating the need for large m.e. The implied lower observed TFP convergence is offset by a larger estimate of the q parameter. The ratio between innovation expenditure and value added is 1.1% in the FLM. This calculation excludes both positive and negative wedges from the cost paid by the firm.

Industrial Policy: Next, we allow the innovation wedges to be correlated with TFP. This captures the possibility that the government engages in some form of industrial policy targeting firms with particular characteristics (e.g., location) that are correlated with TFP. In the estimation, we do not impose any sign on this correlation. We assume the effective innovation cost c_i to be of the form

$$c_i\left(t\right) = \left[\bar{c} - \exp\left(\xi_i(t) - \frac{\sigma_c}{2}\right) + 1 + c_a\left(G(a_i) - 1/2\right)\right] \cdot \left(A_i\left(t\right)^{\theta}\bar{A}(t)^{1-\theta}\right)^{\eta-1},$$

where G is the cumulative density of a_i. $c_a > 0$ means the industrial policy favors low-TFP firms, while $c_a < 0$ means the opposite. This specification ensures that the dispersion in innovation costs is again

Figure 4: Parsimonious Model (PAM)



Note: The figure shows the fit of the PAM. It plots the moment predicted by the model against their empirical counterpart for China 2007–12. The X-axis in Panel A, C and D is the first-period TFP percentiles defined on four intervals: the 10th to 49th, the 50st to 79th, the 80th to 94th and the 95th to 99th. The X-axis in Panel B is the first-period value added percentiles defined on the same four intervals. The solid lines in Panel A and B plot the first-period fraction of R&D firms in each TFP and value added interval, respectively. Panel C plots the median annualized TFP growth among nonR&D firms in each TFP interval in the data against the corresponding expected growth rate for firms in the model. Panel D plots the difference between the median TFP growth between R&D and nonR&D firms in the data against the corresponding difference in expected growth in the model. A firm's TFP growth is the residual of the regression of TFP growth on industry, age and province fixed effects.

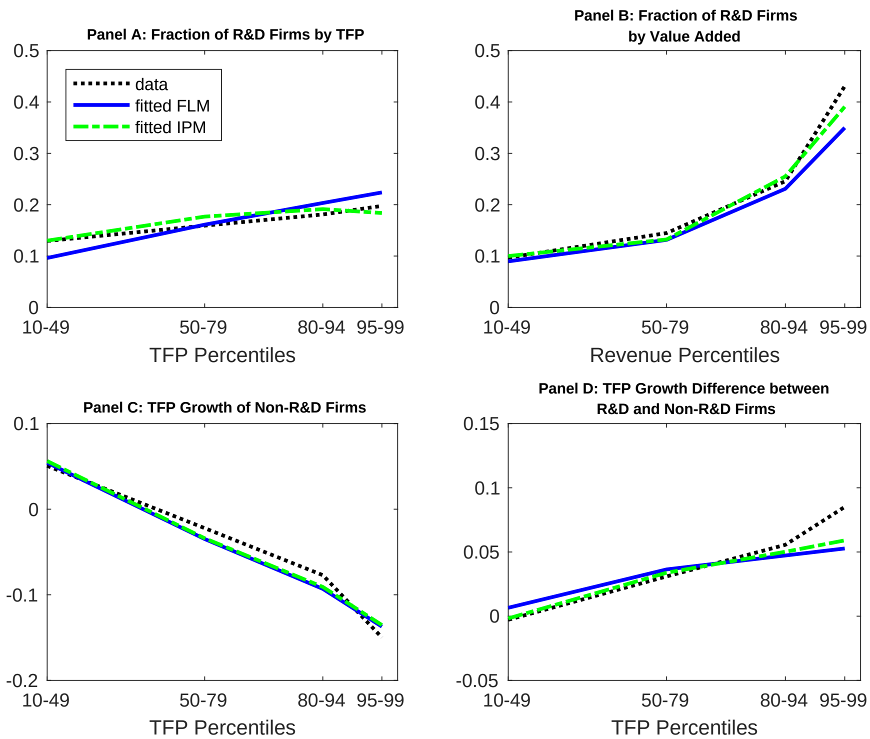
a mean-preserving spread over $\bar{c}$ so the parameter $\bar{c}$ is comparable across specifications.[20]

Column (3) in Table 3 reports the estimation result for the IPM. The fit of the model further improves relative to the FLM—the J-Statistic declines by 30%. Figure 5 shows the fit of the targeted moments for the IPM along with the FLM. The estimated value of the new parameter c_a is positive, indicating a negative correlation between TFP and innovation wedges. In other words, more productive firms are "taxed." To understand the source of identification of this parameter, compare the results for the two models in Panel A in Figure 5. In the estimated FLM, the schedule of Panel A is too steep. A positive correlation between TFP and innovation wedges reduces the propensity of high-TFP firms to innovate thereby flattening the schedule in Panel A. While increasing σ_c would attain the same result, it would also flatten the schedule in Panel B (which is already flatter in the FLM than in the data) compromising the fit of the relationship between size and TFP.

Fake R&D: Chen et al. (2021) suggests that many Chinese firms respond to subsidies by reclassifying operational expenditure as R&D. To explore this hypothesis, we augment our theory with a simple model of moral hazard. We assume that a positive proportion of firms can falsely report to be investing in R&D in order to collect subsidies without suffering any punishment. Misreporting firms

[20]To see this, note that $\int_{-\infty}^{a} G(a)g(a)da = [G(a)]^2/2$, implying that $\mathbb{E}[G(a_i) - 1/2] = 0$.

Figure 5: Flexible (FLM) and Industrial Policy (IPM) models



Note: The figure shows the fit of the FLM and IPM. See Figure 4 for additional information.

are then incorrectly classified as R&D firms. More precisely, we assume that a share Υ of firms can collect subsidies by just *claiming* to invest in R&D. After collecting the subsidy, each firm in this group decides whether it is optimal for them to actually engage in R&D. Since the econometrician cannot see which firms fudge R&D expenditure, misreporting biases the estimated productivity of innovation investments toward zero. The share Υ of privileged firms is the only additional parameter in the structural model.

Column (4) in Table 3 reports the results. According to our estimate, a share $\Upsilon = 9.9\%$ of firms can fake R&D (comprising 9.4% fudging nonR&D firms and 0.5% R&D firms.) Note that in our specification the FRM nests the IPM as a particular case when $\Upsilon = 0$. However, we obtain a very small and statistically insignificant estimate of σ_c indicating that allowing for R&D fudgers makes random heterogeneity in costs redundant. Therefore, the IPM and FRM accurately fit the data with the same number of parameters.

While the fit of the two models is very similar (with the FRM performing marginally better), the two specifications paint a somewhat different picture. On the one hand, two thirds of the R&D firms in the data are fakers. This flattens the schedules in Panels A and B. On the other hand, in the FRM the truthful R&D firms are highly positively selected as shown by the observation that the industrial policy parameter c_a switches sign relative to the IPM. In the FRM, high-TFP firms are subsidized rather than taxed. Moreover, the productivity of innovation investments is substantially higher: $\bar{p} = 0.24$ instead of $\bar{p} = 0.11$. When taking selection into account, this implies that for high-TFP firms, the productivity growth difference between nonR&D and true R&D firms is as high as 11.9%. This result is consistent with the casual observation that China has a number of highly innovative and internationally successful firms (e.g., Lenovo, Tencent). Figure A5 in the appendix illustrates the quantitative results in detail. The figure highlights that high-TFP firms have a stronger comparative advantage in pursuing innovation. Conversely, many low-TFP firms reporting spending on R&D are fudgers.

Because both the IPM and the FRM fit the target moments so accurately, it is not possible to let the data discriminate between the two models. One could possibly rely on some prior about the nature of the industrial policy in China—i.e., whether the government supports or punish the more productive firms.[21] We leave the investigation of this issue to future research.

5.3 Intensive Margin

In Section 3, we noted a positive correlation between R&D expenditure and future productivity growth among the R&D firms. Motivated by this empirical observation, we now perform robustness analysis on the R&D intensity. We pursue two alternative exercises.

Higher R&D cutoff: First, we classify as innovative only those firms with R&D intensity above the empirical median for R&D firms. We recalculate the target moments based on this alternative classification. The results for the PAM and IPM are reported in columns (5)-(6) of Table 3. Treating firms that make small investments in R&D as imitators does not alter the broad picture. The estimated cost of R&D is larger in order to match the smaller proportion of R&D firms. The estimated diffusion parameters q and δ are also larger. The findings are similar in the IPM. Appendix Figure A6 shows the fit of the targeted moments for both the PAM and IPM.

High- and low-R&D firms: Next, we introduce a distinction between high- and low-R&D firms. In the data, we assign a firm to the high-R&D group if its R&D expenditure-to-value added ratio is higher than the median 1.73% ratio. Appendix Figure A7 displays the data moments (along with the fit of the PAM and IPM.) For instance, Panels A1 and A2 are the analogues of Panel A in Figure 4 broken down by high- and low-R&D firms. Two features of the data are noteworthy: First, future productivity growth is higher for high-R&D firms. Second, the propensity to engage in R&D conditional on TFP and size are similar for the two groups of firms. Then, we augment the theory with the assumption that there exist two distinct technologies entailing different costs and success probabilities. More formally, firms are randomly assigned to either of the technologies with parameters $\{\bar{c}_l, \bar{p}_l\}$ and $\{\bar{c}, \bar{p}\}$, respectively. Each firm draws a probability of success p from the distribution to which it is assigned. The distribution of wedges is assumed to be independent of the assignment.

The targets of our estimation are now the empirical moments in Figure A7. In addition, we target the ratio of R&D expenditure to value added for high- relative to low-R&D firms, which is a factor of 8.6 in the data. The proportion of high-R&D firms is an additional parameter that we estimate.

The estimates for the PAM and IPM are reported in Columns (7)-(8) of Table 3. In the PAM, firms assigned to the $\{\bar{c}_l, \bar{p}_l\}$ group face both a lower cost and a lower average probability of success if they choose the innovation strategy.[22] In the IPM, the estimated productivities $\bar{p}_l$ and $\bar{p}$ are very similar

[21] Technically, the estimation finds two local minima. One corresponds to column (4) in Table 3. The other is a corner solution where no firms can fake R&D and the estimated parameters are as in column (3) of Table 3. In the calculation of bootstrap standard errors, some simulations yield a lower J-Statistic in the former model, while others yields the opposite result. When calculating the bootstrap standard errors of the estimates in column (4), we only consider draws where the minimum is interior.

[22] The proportion of firms drawing from the $\{\bar{c}_l, \bar{p}_l\}$ process is estimated to be 13%. Since the high-R&D firms are by construction 50% of the total R&D firms, this implies that many firms assigned to the $\{\bar{c}, \bar{p}\}$ process choose to imitate because innovation is too costly. In contrast, a large proportion of firms assigned to the $\{\bar{c}_l, \bar{p}_l\}$ process choose the

(in fact, $\bar{p}_l > \bar{p}$.) Still, selection guarantees that among the firms choosing the innovation strategy TFP growth is significantly higher for high-R&D than for low-R&D firms, consistent with the data. Intuitively, because of the high investment cost, only the very best firms (i.e., those drawing very high p's) assigned to the $\{\bar{c}, \bar{p}\}$ process choose to innovate. Appendix Figure A7 shows that both the PAM and IPM fit well the target moments.

We conclude that a simple extension of the model can capture the empirical relationship between R&D intensity and future growth. The limitation of this extension is that it does not allow firms to choose the project size. Endogenizing the intensive margin requires a more significant departure from our discrete-choice model and is left to future research.

5.4 Nontargeted moments

In this section, we discuss predictions of the model for empirical moments we do not target in the estimation.

5.4.1 Indirect inference

The estimation targets pairwise empirical correlations but not the joint correlation structure between the variables. We now use indirect inference methods to investigate whether the model is consistent with the empirical conditional correlations. To this aim, we consider a set of multiple linear regressions. In the first ones, the dependent variable is the discrete choice of pursuing an innovation strategy, while the right hand-side variables are TFP levels (in logarithms) and the wedges. Panel A of Table 4 shows the results. Column (1) restates the empirical relationship in Panel A of Table 2: the probability for a firm to invest in R&D is increasing in both TFP and output subsidies (i.e., a negative coefficient on the output wedge τ). Running the same regression on the simulated model (including m.e.) yields the same qualitative results for all models. In the PAM, the R&D decision is more elastic to TFP and wedges than in the data, while the heterogeneous cost models provide a better quantitative match to the data.[23]

Next, in Panel B of Table 4 we run linear regressions where the firms' TFP growth is the dependent variable while the initial TFP level and R&D decision (in the initial year) are the explanatory variables. Column (1) restates the empirical relationship from Table 2 that TFP growth is falling in initial TFP and that TFP growth is larger for R&D firms. Running the same regression on the simulated models replicates the empirical correlations. Moreover, all models yield elasticities that are in the ballpark of the empirical estimates. In conclusion, the indirect inference analysis shows that the model fits well the (nontargeted) joint correlation structure between TFP growth, the R&D decision, initial TFP level, and size (wedges).

5.4.2 Productivity distribution

The structural model features a *traveling wave*—a stationary productivity distribution—that evolves at a constant endogenous speed. The distribution is tent-shaped with two Pareto tails—cf. Proposition 2. In this section, we compare the stationary TFP distribution and the growth rate predicted by the theory with their empirical counterparts.

innovation strategy.

[23]The standard errors of the structurally estimated parameters are based on simulating a sample of the same size as the empirical one and estimating the regression in the same way as on the empirical sample.

Table 4: Indirect Inference, Balanced Panel of Chinese Firms, 2007–2012.

PANEL A

Dependent Variable: R&D decision in 2007

	(1) Data	(2) PAM	(3) FLM	(4) IPM	(5) FRM
log(TFP)	0.368***	0.712***	0.287***	0.298***	0.322***
	(0.0284)	(0.0145)	(0.0167)	(0.0181)	(0.0167)
wedge	-0.410***	-0.824***	-0.274***	-0.327***	-0.357***
	(0.0357)	(0.0177)	(0.0202)	(0.0214)	(0.0203)

PANEL B

Dependent Variable: TFP Growth

	(1) Data	(2) PAM	(3) FLM	(4) IPM	(5) FRM
log(TFP)	-0.062***	-0.094***	-0.104***	-0.112***	-0.081***
	(0.0035)	(0.0002)	(0.0003)	(0.0003)	(0.0004)
$R\&D_d$	0.036***	0.034***	0.033***	0.028***	0.053***
	(0.0042)	(0.0005)	(0.0005)	(0.0005)	(0.0008)

Note: The first column reports the regression results from the data. All regressions include industry, age, and province fixed effects. Columns (2)-(5) report the results from the simulated data. See Table 2 for variable definitions.

TFP dispersion: Figure 6 compares the predicted and empirical TFP distributions for the PAM, FLM, IPM, and FRM. The estimated model has a lower dispersion than TFP in the data: the PAM accounts for 72% of the empirical variance, while the three heterogeneous cost models account for between 44% and 62% of it. The inability of the model to account for the full empirical dispersion can be attributed to factors that are omitted in our stylized theory. For instance, firms could be subject to exogenous i.i.d. shocks to TFP that do not affect the cost of innovation. These shocks would increase the TFP dispersion without significantly altering firms' imitation-innovation choice. We also note that the dispersion of TFP is sensitive to the step-size parameter $\tilde{a}$. A larger $\tilde{a}$ yields a higher dispersion. In Panel A, we show that estimating the model under the assumption of $\tilde{a} = 1.12$ yields a good match of the empirical distribution. However, such a model overpredicts the empirical average R&D-to-value added ratio.[24]

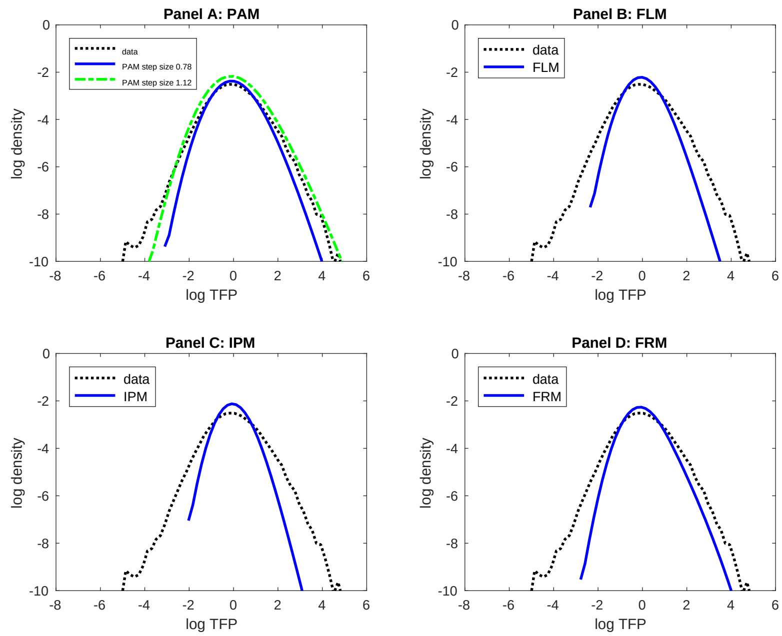
The model matches the upper tail better than the lower tail of the empirical distribution, whose measurement is notoriously noisy. Nevertheless, our model makes some progress on understanding the lower tail because most existing theories of random interactions between firms do not feature any lower Pareto tail.

TFP growth: The model yields predictions about the speed of growth of the traveling wave. The aggregate TFP growth in the data is 3%.[25] The PAM implies a steady-state annualized productivity

[24] The step size $\tilde{a}$ does not affect the fit the targeted moments. In particular, Figure 4 would look essentially identical for $\tilde{a} = 1.12$.

[25] We calculate the aggregate growth rate in the data using the methodology of Hsieh and Klenow (2009). We first

Figure 6: TFP Distribution



Note: The dotted lines represent the log density of log TFP in the data. The solid and dashed lines are the steady-state log density of log TFP for various models. Panel A shows two versions of the PAM–the benchmark model and the one with smaller step size. Panel B: FLM, Panel C: IPM, and Panel D: FRM.

growth rate of 3.6%. The corresponding figure in the IPM is 5.1%.

In our model, the aggregate growth rate is determined by both innovation and knowledge diffusion. To quantify the role of each channel we evaluate what aggregate growth would be if only one channel were operating for one period while holding fixed the firms' R&D decisions. In the PAM, innovation and imitation account for 20% and 80% of TFP growth, respectively. The share accounted for by passive imitation from innovating firms is negligible. Note that the decomposition would yield very different results if innovation were shut down permanently. Absent innovation, the long-run growth would be zero as long as the initial TFP distribution is bounded.

5.4.3 Patents

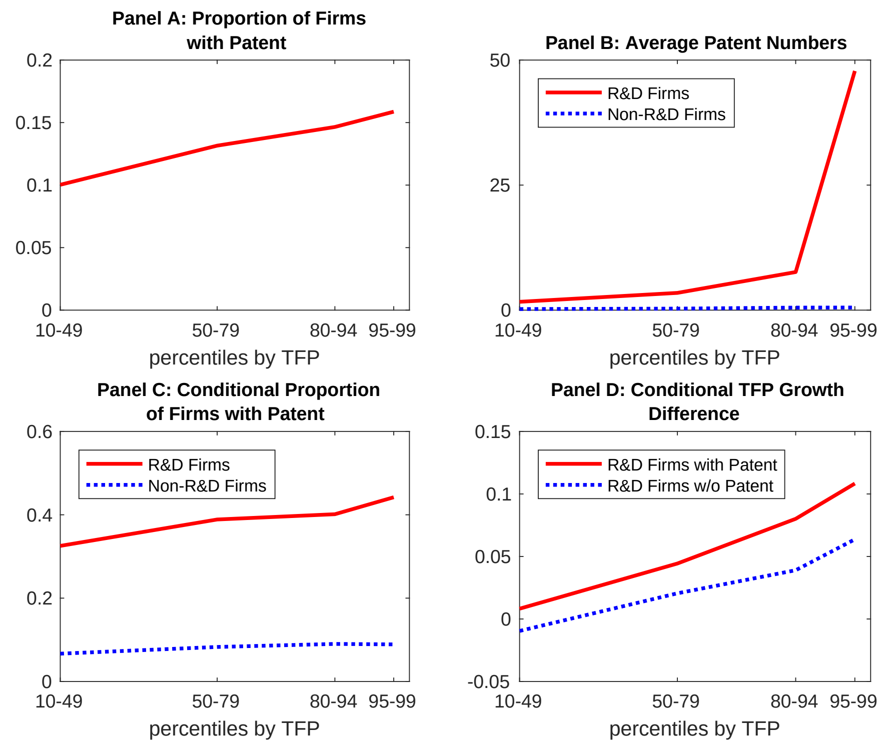
Our theory predicts that firms that are larger and more productive invest more in attempting to innovate and should therefore innovate more. Moreover, among those trying, firms that are successful at innovating should grow faster than those failing. In Section 3 we measured innovation investments by R&D expenditures. Alternatively, we could try to measure the outcome of this investment activity. A common empirical measure of successful innovation is patents. In this section, we show that the predictions of our theory are broadly consistent with data on patents.

To this end, we collect data for all the patents approved by China's State Intellectual Property Office (SIPO). We match the SIPO data with the 2012 NBS data. In the 2007–12 balanced panel of matched firms, there are 14,492 firms (out of ca. 123,000) that were granted one or more patent for

calculate TFP growth at the two-digit industry level and then aggregate up using industry deflators and value added shares.

which they applied in the period 2007–12. The total number of patents these firms applied for in that period is 146,896. This implies that on average each NBS firm with a positive number of patents applied for 10.1 patents.[26] Note that, since the average time for granting a patent is about three years, a lot of patents sought during 2007–12 have been granted after 2016. The magnitudes above are therefore a lower bound to the actual number of patents.

Figure 7: Patents of Chinese Firms, Balanced Panel 2007–12



Note: All patents in the figure refer to invention patents applied during 2007–12. We group all firms in the 2007–12 balanced panel into percentiles by their initial TFP. Panel A plots the proportion of firms with patents in each percentile. Panel B plots the average number of patents among R&D firms (solid line) and among nonR&D firms (dotted line) in each percentile. Panel C plots the proportion of R&D firms with one or more patents (solid line) and the proportion for nonR&D firms (solid line). The solid line in Panel D plots the TFP growth difference between R&D firms with patents and nonR&D firms. The dotted line plots the TFP growth difference between R&D firms without patents and nonR&D firms.

Panel A of Figure 7 shows that the propensity for patenting innovations is increasing in the TFP level, consistent with Panel A of Figure 2. Panel B plots the average number of patents as a function of TFP broken down by R&D and nonR&D firms. Essentially all patents are sourced from firms reporting some R&D activity. The same pattern emerges from Panel C, which plots the proportion of R&D firms and the proportion for nonR&D firms with one or more patent. Clearly, R&D is strongly correlated with patenting. This evidence shows that the data on R&D expenditures well captures innovation investments. Panel D displays the most interesting finding. Firms with a positive number of patents experience larger TFP growth than R&D firms without a patent. The gap increases with TFP, being

[26]In the unbalanced panel, 28,081 NBS firms (out of a total of ca. 275,000) have one or more patent for which they applied in the period 2007–12. The total number of patents these NBS firms applied for during that period is 228,634. This implies that on average each NBS firm with a positive number of patents applied for 8.1 patents.

largest for the top two deciles of the TFP distribution. This is consistent with our model, where R&D firms that are successful at innovation grow faster than those that are not able to innovate, and this difference is increasing in TFP.

This evidence is also suggestive of the hypothesis that some firms that report R&D but do not patent innovations may be fudgers–consistent with the Fake R&D model we estimated. Clearly, this evidence is only suggestive. Non-patenting R&D firms could simply be firms that invested in R&D but had bad luck. However, it is interesting to observe that the number of patents per firm increases sharply from the second-highest TFP quantile to the highest quantile (top five percent) in Panel B, precisely consistent with the prediction of the FRM that it is mostly low-TFP firms that have an incentive to opportunistically report R&D expenditure without actually engaging in it.

6 Estimating the model on different samples

In this section we re-estimate the model using data for Taiwan and for China in an earlier period (2001–07). The goal of this analysis is twofold. First, we want to assess the extent to which the estimates are robust across different samples.[27] Second, we would like to learn about the productivity of R&D investments in other economies or periods. We focus on Taiwan both because similar data are available and because it is an export-oriented economy with an important role of manufacturing industry that has commonalities with China.

China 2001–07: We first consider data for China in the earlier period of 2001–07. Table 5 shows the results.[28] Similar to the results for the 2007–12 period, our model also matches the earlier period very well: if anything, the J-Statistic is lower than in the later period. Appendix Figure A8 displays the fit of the targeted moments for the PAM and IPM.

The estimated parameters are qualitatively in line with those for 2007–12, although there are some quantitative differences. The estimated m.e. is larger, while the estimated productivity of R&D, $\bar{p}$, is about a third of the estimates for 2007–12. The technology diffusion parameter q is also substantially lower than in the benchmark sample.[29]

In summary, the results for China in the earlier period 2001–07 indicate a less important role for R&D investments as a driver of productivity growth compared to the 2007–12 period. The results are consistent with a technological interpretation: the productivity of R&D was intrinsically lower in the earlier period, and the difference between R&D firms and nonR&D firms was smaller. This interpretation is in line with the argument of Acemoglu et al. (2006) that the importance of R&D investment increases as an economy approaches the world technology frontier.

Taiwan: Consider now the data for Taiwan. The results are summarized in columns (4)–(6) of Table 5. Appendix Figure A9 displays the fit of the targeted moments for the PAM and IPM. The J-Statistic of the PAM is lower for Taiwan than in both Chinese samples.

[27] To ensure that the estimates are comparable across data samples, we estimate all models with the same step size on the TFP ladder as in the benchmark model, 0.78.

[28] We do not report the FRM because the best fit is attained by setting the proportion of fudging firms equal to zero. (For the same reason, we do not report the FRM for the case of Taiwan discussed below). This result is consistent with the observation that subsidy policy in China only became pervasive since 2007.

[29] Part of the reason why the productivity of R&D is lower may be related to the fact that a large proportion of R&D firms make very small R&D investments. In regressions analogous to columns (4) and (5) in Panel B of Table 2, we find that R&D has a significant positive effects on future TFP growth only for firms in the upper tercile of the distribution of R&D-to-value added.

Table 5: Estimation for alternative samples.

	(1)	(2)	(3)	(4)	(5)	(6)
	China: Balanced Panel 2001–2007			Taiwan: Balanced Panel 1999–2004		
	PAM	FLM	IPM	PAM	FLM	IPM
Imitation prob. q	0.036	0.090	0.093	0.286	0.501	0.371
	(0.032)	(0.046)	(0.039)	(0.086)	(0.129)	(0.120)
Second chance δ	0.033	0.164	0.069	0.027	0.002	0.001
	(0.071)	(0.084)	(0.117)	(0.023)	(0.058)	(0.047)
Innov. prod. $\bar{p}$	0.034	0.046	0.045	0.184	0.207	0.206
	(0.011)	(0.013)	(0.012)	(0.026)	(0.031)	(0.032)
Innov. cost $\bar{c}$	0.530	1.174	-0.095	3.301	4.669	4.247
	(0.159)	(0.325)	(0.198)	(0.475)	(0.717)	(0.089)
Std.dev. m.e. $\sigma_{\mu a}$	0.682	0.575	0.580	0.722	0.538	0.622
	(0.038)	(0.025)	(0.026)	(0.074)	(0.073)	(0.112)
Std.dev. innov. subs. σ_c		0.644	0.559		1.371	1.378
		(0.145)	(0.119)		(0.146)	(0.108)
Policy inter. c_a			0.513			-2.385
			(0.163)			(0.792)
J-Statistic	1.085	0.703	0.371	0.861	0.570	0.503

Note: Estimated parameters of various models using the Chinese 2001–2007 sample and the Taiwanese 1999-2004 sample. Bootstrapped standard errors in parentheses.

The estimated parameter are qualitatively in line with those for China. However, there are some noteworthy quantitative differences. First, the productivity of innovation is larger in Taiwan: the estimated average probability of success in innovation is around 9%, while it was only 5% in the 2007–12 China sample. However, the cost parameter $\bar{c}$ is also larger in the Taiwanese sample (we return to this finding in the following section.) Second, the productivity of imitation is higher, implying a faster technological convergence among Taiwanese firms than among Chinese firms. The point estimates for q are substantially larger than for China, ranging between 0.29 and 0.5. Conversely, we estimate a larger variance of the m.e. than in the Chinese data.

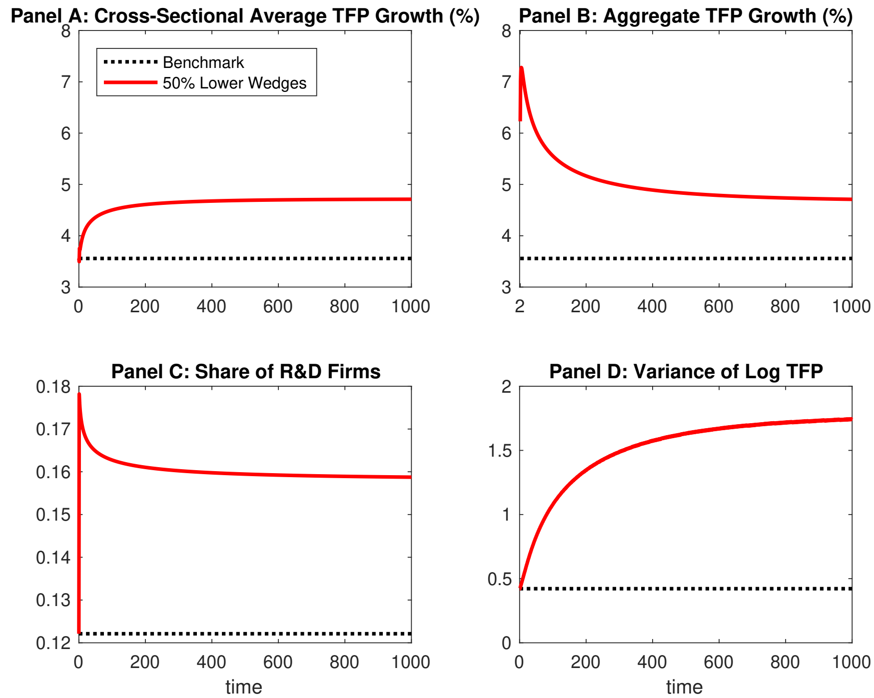
The estimated parameters are very stable across different specifications. Interestingly, the parameter estimates for the R&D technology in Taiwan are in the ballpark of the estimates of $\bar{p}$ and q in the FRM for China. One interpretation of this finding is that the FRM might best capture the true R&D technology in China.

7 Counterfactuals

In this section, we report the results of some counterfactual policy experiments we performed based on the estimated model. We discuss the PAM as the main case because its economic mechanism is very transparent. The results and figures of the IPM are similar and are deferred to the appendix.

Reducing misallocation: Our main counterfactual experiment is an exogenous reduction in misallocation. Hsieh and Klenow (2009) document large static efficiency gains from reducing misallocation in China. In our model there are additional dynamic effects through the R&D investments. Output wedges make R&D decisions depend more on firm size and less on TFP, flattening the schedule of Panel A in Figure 4 relative to Panel B in the same figure. If misallocation were removed altogether, the

Figure 8: Transition after Lowering Wedges



Note: The graph displays the transition following a 50% reduction in the variance of wedges. Panel A plots the growth rate of cross-sectional average TFP. Panel B plots the growth rate of aggregate TFP. Panel C plots the share of R&D firms, and Panel D plots the cross-sectional variance of log TFP.

correlation between size and R&D would be driven by TFP differences only, in which case the schedules in Panels A and B would be identical.

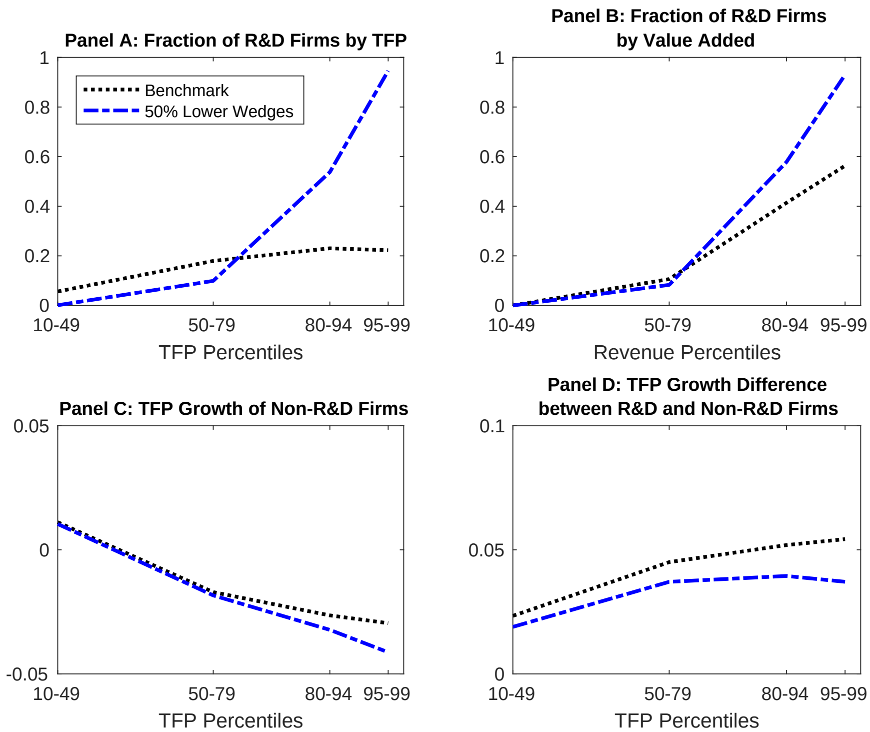
We study the dynamic effects of an unanticipated instant reduction in the variance of the logarithm of output wedge $\log(1-\tau_i)$. We engineer this reduction by halving both b^2 and $var(\varepsilon_i^\tau)$.[30] Our experiment is not directly comparable with that performed by Hsieh and Klenow (2009) because they assume wedges and TFP to follow bivariate lognormal distribution, implying that the the variance of log TFPR is a sufficient statistic for the static effect of distortions on aggregate TFP. In our model, the correlation matters and is inferred from the data. This is quantitatively important, especially for the dynamic effects: in our counterfactual, the decrease in the parameter b is actually the main source of dynamic gains. The reason is that the wedge on high-TFP firms—that efficiency considerations would require to invest in R&D—is the main distortion on R&D decisions.

Figures 8 and A10 display the transition of the economy from the initial to the counterfactual steady state for the PAM and IPM, respectively.[31] The two upper panels show the evolution of the growth rates in the cross-sectional average TFP and in aggregate TFP, respectively. Note that—to ease the visualization—Panel B does not display the initial jump in TFP arising from the static effect. In steady state, the two statistics are identical because the higher-order moments of the TFP distribution are constant over time. However, along the transition changes in higher-order moments also affect growth.

[30] Recall that $\log(1-\tau_i) = b \cdot log(A_i) + \varepsilon_i^\tau$, see eq. (13). Thus, $var(\log(1-\tau)) = b^2 \cdot var(\log(A)) + var(\varepsilon_i^\tau)$.

[31] It would be a formidable task to calculate numeric transitions in which the cost of R&D changes with the growth rate every period in line with Equation 7. For this reason, we approximate the path of the innovation cost using the constant steady-state growth.

Figure 9: Steady-State Moments in the Counterfactual Model



Note: The dashed lines display the moments in steady state for the counterfactual experiment where we reduce the variance of wedges by 50%. The dotted line displays the moments predicted by the benchmark estimated PAM.

In our experiment, the cross-sectional variance of log TFP increases over time (Panel D) because the upper tail of the TFP distribution becomes fatter.

The increase in the cross-sectional average lifts the growth rate from the initial 3.6% to 4.7% in the new PAM steady state (see column (1) in Table 6). The transition is hump shaped: the aggregate growth rate first increases sharply and then slowly declines to the new steady state. In the earlier part of the transition, the main driver for aggregate TFP growth is changes in higher-order moments. Eventually, the variance of $\log(TFP)$ stabilizes and so does the aggregate growth rate—see Panel B.

Panel C shows the evolution of the share of R&D firms. This share increases by 50% upon impact because the reduction in the dispersion of output wedges incentivates high-TFP firms to invest in R&D. This effect is dampened over time because the increase in TFP dispersion implies an endogenous increase in misallocation that partially offsets the initial decline. Still, in the new steady state, the share of R&D firms (16%) is higher than in the initial steady state. Finally, Panel D shows the increase over time in the variance of log TFP driven by high-TFP firms pulling ahead in the distribution.

Figures 9 and A11 show how the steady-state predicted target moments change after the reduction in misallocation.[32] The main changes are in Panels A and B. Panel A shows that reducing misallocation increases the correlation between R&D and TFP. In the counterfactual PAM economy, 95% of the firms in the top five TFP percentiles and 54% of the 80th–94th TFP percentiles invest in R&D—the corresponding numbers for the IPM are 66% and 30%. In contrast, hardly any firm with TFP

[32]Since these figures show simulated results, we do not add m.e. This is different from Figure 4 where we add m.e. to the simulations in order to make the results comparable with the data.

below median invests in R&D. Moreover, the size-R&D profile in Panel B is now much more similar to the TFP-R&D profile. Panel D shows that the TFP growth difference between R&D and nonR&D firms is slightly smaller in the counterfactual. This is due to a selection effect: wedges deter firms from investing in R&D, except for those drawing a very high p. This implies positive selection on p. Lowering distortions reduces the positive selection.

The main conclusion of this experiment is that misallocation has large quantitative effects on productivity growth. Moreover, it widens the dispersion in the stationary TFP distribution. In our model, firms' heterogeneity has no effect on wage inequality because the labor market is competitive. However, in a model with realistic labor market frictions (e.g., search and ex-post rent sharing) and assortative matching between workers and firms (Lindenlaub (2017)), reducing wedges and misallocation would also increase wage inequality.

International spillovers: The permanent growth effect of a reduction of misallocation (about 1.1% increase in the PAM) is arguably overly optimistic, even though it results from a large change in misallocation. Many economists believe that the fast growth of China stems, at least in part, from technological convergence due to international spillovers. According to this view, as China approaches the world economic frontier, this source of growth is going to dry up.

In this section, we embed our model in a model of technological convergence where growth arises from both random interactions and international spillovers, which we assume to increase with the distance to the world frontier as in Acemoglu et al. (2006). More formally, we assume that aggregate TFP growth in the nonfrontier economy j equals $g_{jt} = \Gamma(MIS_j, \Pi_{jt}) + \Delta(TFP_{ft}/TFP_{jt})$, where Γ is the outcome of our benchmark model, Π stands for the TFP distribution, and Δ captures learning from the frontier economy. Note that Γ is a decreasing function of the level of misallocation MIS_j. In turn, MIS_j is parameterized as above by b^2 and $var\left(\varepsilon_i^\tau\right)$.[33]

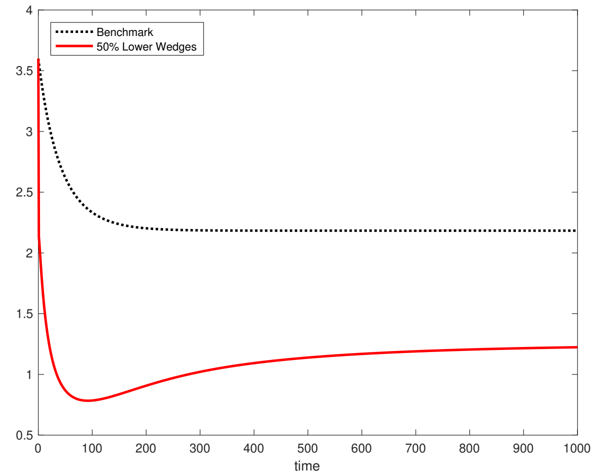
We assume $\Delta(x) = \zeta \times log(x) - d$, where ζ is a convergence parameter and d captures knowledge depreciation. In the long run, all countries grow at the same constant rate g which is set by the frontier economy, while the productivity gap between the frontier and nonfrontier economies is determined by the relative misallocation. More formally, if we denote the steady state expression of Γ_j by $\tilde{\Gamma}(MIS_j) \equiv \Gamma\left(MIS_j, \Pi(MIS_j)\right)$, the steady-state TFP difference between a generic economy k and the frontier economy f is given by: $log(TFP_f/TFP_k) = \left(\tilde{\Gamma}(MIS_f) - \tilde{\Gamma}(MIS_k)\right)/\zeta$.

We calibrate $\zeta = 2\%$ consistent with standard measures of the cross-country convergence rate. We set $d = 0.0312$ so that the model matches the aggregate growth rate of TFP in Chinese manufacturing in our sample (3%). Finally, we set the TFP growth rate for the frontier economy (i.e., $\tilde{\Gamma}(MIS_f) - x$) to 2%. Given this calibration, the current level of misallocation of China implies that its GDP per capita converges in the long run to 46% of the frontier economy's level.

Next, we counterfactually reduce misallocation by 50%. We assume an initial TFP gap between China and the frontier economy of 3.6 in line with the China-US gap estimated by Shen et al. (2015). Figure 10 shows the results for the PAM (the IPM yields similar results). The TFP gap relative to the frontier economy falls on impact and keeps shrinking thereafter, both because of the international spillover and the transitional dynamics of the random interaction model. In the long run, TFP converges to 81% of the frontier level (as opposed to 46% in the status quo). While part of this gain accrues from the static effect of reducing misallocation, Figure 10 shows that the additional dynamic gains arising from the mechanism of our model are quantitatively large: the static effect instantaneously cuts the gap from 3.68 to 2.13 while the ensuing dynamic effect further decreases it to 1.23. Throughout transition,

[33] Recall also that while during transition growth depends on the evolving TFP distribution Π_{jt}, the long-run distribution is pinned down by the level of misallocation, i.e., $\Pi_j = \Pi(MIS_j)$.

Figure 10: TFP Gap During Transition with International Spillovers



Note: Evolution over time of the ratio between the TFP of the frontier economy and that of the benchmark economy (China) in the model with international spillovers. The dotted and solid lines display simulated transitions in the estimated model and in the counterfactual where output wedges are reduced by 50%, respectively.

growth is hump-shaped, being faster than in the benchmark economy for decades. The hump results from the joint effects of the random interaction model and the declining international spillover.[34]

In conclusion, this section shows that in a model with international spillovers a reduction in misallocation has a significant effects on the speed of transition and on the long-run GDP level even in an experiment in which China does not become the permanent world leader.

The innovation technology: In this section, we perform additional counterfactuals where we change parameters of the technology of innovation. Table 6 summarizes the results for the PAM. Appendix Table A4 shows the corresponding results for the IPM. Column (1) reports the fraction of R&D firms and steady state growth rate in the estimated model. Column (2) summarizes the long-run effect of the reduction in misallocation discussed above. In columns (3), (4), and (5), we change the structural parameters q , $\bar{p}$, and $\bar{c}$ to the estimated level in Taiwan while keeping the other parameters and misallocation unchanged. In columns (6) and (7), we consider uniform taxes or subsidies on R&D that change the baseline cost of innovation $\bar{c}$ so that, respectively, 5% and 20% of the Chinese firms invest in R&D. Finally, in column (8) we consider a drastic policy inducing (e.g., through arbitrarily large R&D subsidies) all firms to pursue the innovation strategy. Although this case features no stationary distribution (namely, the variance of TFP increases perpetually) it is possible to calculate analytically the (approximate) growth rate when setting $\delta \approx 0$ (recall that the estimates of δ are always very small). Consider first setting the parameters to the Taiwanese estimated level. In the PAM, we observe a significant increase in steady-state TFP growth in all three scenarios, driven by the faster rate of both innovation and imitation. When we set only the parameter q to the Taiwanese level, the proportion of R&D firms falls. When we reset $\bar{p}$ and $\bar{c}$, it increases. Finally, when we set $q, \bar{c}$, and $\bar{p}$ to their respective Taiwanese estimated levels, the R&D share is the about the same as in the estimated model.

In columns (2)–(4) of Table A4, we report the results of the same counterfactuals in the IPM. The results of Appendix Table A4 echo the results for the PAM. For instance, simultaneously setting $q, \bar{c}$,

[34] In this experiment, China temporarily becomes the world leader during transition, although eventually it stops being the leader. The simulation takes into full account the consequences of these shifts.

and $\bar{p}$ to the respective Taiwanese levels increases TFP growth (from 5.1% to 5.8%), although it reduces the share of R&D firms (from 15% to 8%). In conclusion, the counterfactuals show that R&D is more productive in Taiwan than in China. At a constant misallocation, China would grow faster if it had access to the same innovation technology as Taiwan.

Finally, we study the effect of policies affecting the private cost of R&D. Columns (6) and (7) show that aggregate growth increases when we reduce the cost of innovation and decreases when we increase it. Therefore, fostering innovation through a moderate R&D subsidy across the board yields the standard positive growth effect as in models of endogenous technical change (e.g., Aghion and Howitt (1992)). However, the effect is nonmonotonic and turns around for sufficiently large subsidies. This is illustrated by the extreme case of column (8). There, subsidies are so large that all firms choose the innovation strategy. The long-run growth rate in the counterfactual economy is lower than the growth rate when the policy induces 20% of the firms to perform R&D. In the IPM, the growth rate of the economy where all firms pursue innovation is actually lower than in the baseline estimated economy—see column (8) of Appendix Table A4. This "too much of a good thing" result hinges on the opportunity cost of forgoing the benefit of random interactions.

Table 6: Counterfactuals, Parsimonious model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	PAM estim. model	50% lower output wedges	Taiwan's q	Taiwan's $\bar{p}$ and $\bar{c}$	Taiwan's $\bar{p}$, $\bar{c}$, and q	Increase $\bar{c}$ so share R&D firms = 5%	Decrease $\bar{c}$ so share R&D firms = 20%	All firms do R&D
Fraction of R&D Firms (%)	12.2	15.8	10.7	14.1	12.2	5	20	100
Steady State TFP Growth (%)	3.56	4.70	4.49	4.92	6.03	2.41	4.42	3.80

Note: The table reports statistics for the counterfactual experiments for the PAM discussed in the text. Column (1) reports the predicted moments of the estimated PAM for comparison.

8 Conclusion

In this paper, we construct and structurally estimate a theory of productivity growth driven by innovation and technology diffusion through random interactions. In the theory, both the TFP level and firm-specific distortions are sources of comparative advantages: firms with high TFP and firms with negative output wedges have a stronger incentive to invest in R&D. The theory bears testable predictions about the evolution of the aggregate productivity distribution. We estimate the model to earn new insights about the nature and effects of the R&D expenditure boom in China in recent years. R&D investments appear to have contributed to the aggregate productivity growth of China. However, the return to R&D investments is lower in China than in Taiwan. Moreover, pervasive output wedges often induce the wrong firms (and, conversely, deter the right ones) from investing in R&D, thereby reducing the aggregate productivity of R&D investments.

Relative to earlier Schumpeterian growth theory, our study incorporates both innovation and technology diffusion into a common framework. Moreover, it shifts the focus from the overall investment in innovation to the efficient assignment of firms to innovation and imitation activities. While innovation is the ultimate engine of growth, an excessive (or ill-targeted) policy emphasis on innovation can actually backfire. The reason is that less productive firms have a high growth potential through imitation which they forgo when they focus on innovation. Like innovation, successful imitation carries positive externalities to less productive firms. Another important message of the paper is that misallocation has significant dynamic effects. To the extent to which larger firms have stronger incentives (and fewer constraints) to pursue R&D investments, misallocation distorts the natural comparative advantage of firms in leading the innovation process, and ultimately slows down economic growth. The study has some limitations that can be addressed in future research. In our estimation, we focus on the balanced panel of firms that are in the sample both in 2007 and 2012, abstracting from entry and exit. While churning and the formation of new firms are important features of the Chinese data, we believe that a study focusing on the R&D investments of incumbent firms is no less important. In 2007, our balanced sample captures 71% of the R&D investments (which is the focal point of our study) and 63% of the value added of firms in the total sample. Because we have no data on R&D expenditure after 2007, it would be difficult to study the imitation-innovation behavior of firms entering the sample after that date. We leave it to future research to integrate entry-exit decisions and innovation and random interactions among incumbent firms into a common framework. For instance, one could add a statistical model of entry and exit accounting for the TFP growth stemming from churning. While this would affect the dynamics, we conjecture that it would likely not alter significantly the estimated parameters that are based on cross-sectional moments. Future work should also explore in more depth firm dynamics and the role of an intensive margin of R&D.

References

Acemoglu, D., P. Aghion, and F. Zilibotti (2006). Distance to frontier, selection, and economic growth. *Journal of the European Economic Association 4*(1), 37–74.

Acemoglu, D., U. Akcigit, H. Alp, N. Bloom, and W. Kerr (2018). Innovation, reallocation, and growth. *American Economic Review 108*(11), 3450–91.

Acemoglu, D. and D. Cao (2015). Innovation by entrants and incumbents. *Journal of Economic Theory 157*, 255–294.

Acemoglu, D. and F. Zilibotti (2001). Productivity differences. *The Quarterly Journal of Economics 116*(2), 563–606.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica 60*(2), 323–351.

Akcigit, U., H. Alp, and M. Peters (2021). Lack of selection and limits to delegation: Firm dynamics in developing countries. *American Economic Review 111*(1), 231–75.

Akcigit, U., S. Caicedo, E. Miguelez, S. Stantcheva, and V. Sterzi (2018). Dancing with the stars: Innovation through interactions. NBER Working Paper w24466.

Akcigit, U. and W. R. Kerr (2018). Growth through heterogeneous innovations. *Journal of Political Economy 126*(4), 1374–1443.

Altonji, J. G. and L. M. Segal (1996). Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics 14*(3), 353–366.

Asl, F. M. and A. G. Ulsoy (2003). Analysis of a system of linear delay differential equations. *Journal of Dynamic Systems, Measurement, and Control 125*(2), 215–223.

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

Bai, C.-E., C.-T. Hsieh, and Z. Song (2016). The long shadow of a fiscal expansion. *Brookings Papers on Economic Activity* (3), 129–165.

Bellman, R. and K. L. Cooke (1963). *Differential-difference equations.* Academic Press, NY.

Benhabib, J., J. Perla, and C. Tonetti (2014). Catch-up and fall-back through innovation and imitation. *Journal of Economic Growth 19*(1), 1–35.

Benhabib, J., J. Perla, and C. Tonetti (2021). Reconciling models of diffusion and innovation: A theory of the productivity distribution and technology frontier. *Econometrica Forthcoming.*

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

Bramson, M. (1983). *Convergence of solutions of the Kolmogorov equation to travelling waves*, Volume 285. American Mathematical Soc.

Brandt, L., G. Kambourov, and K. Storesletten (2016). Firm entry and regional growth disparities: The effect of soes in china. Unpublished Manuscript, University of Toronto.

Brandt, L., J. Van Biesebroeck, L. Wang, and Y. Zhang (2017). WTO accession and performance of Chinese manufacturing firms. *American Economic Review 107*(9), 2784–2820.

Chen, Z., Z. Liu, J. C. Suárez Serrato, and D. Y. Xu (2021). Notching R&D investment with corporate income tax cuts in China. *American Economic Review 111*(7), 2065–2100.

Cheremukhin, A., M. Golosov, S. Guriev, and A. Tsyvinski (2017). The economy of People's Republic of China from 1953. Unpublished Manuscript,Yale University.

Colarelli O'Connor, G. (2019). Real innovation requires more than an R&D budget. *Harvard Business Review*.

Colarelli O'Connor, G., A. C. Corbett, and L. S. Peters (2018). *Beyond the champion: Institutionalizing innovation through people.* Stanford University Press.

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

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–71.

Ding, X. and J. Li (2015). *Incentives for innovation in China: Building an innovative economy.* Routledge.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *The Quarterly Journal of Economics 114*(3), 739–767.

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

Garcia-Macia, D., C.-T. Hsieh, and P. J. Klenow (2019). How destructive is innovation? *Econometrica 87*(5), 1507–1541.

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

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

Griliches, Z. (1998). Patent statistics as economic indicators: A survey. *R&D and Productivity: The econometric evidence, University of Chicago Press*, 287–343.

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

Holmes, T. J., E. R. McGrattan, and E. C. Prescott (2015). Quid pro quo: Technology capital transfers for market access in china. *Review of Economic Studies 82*(3), 1154–1193.

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

Hsieh, C.-T. and P. J. Klenow (2010). Development accounting. *American Economic Journal: Macroeconomics 2*(1), 207–23.

Hsieh, C.-T. and Z. Song (2015). Grasp the large, let go of the small: The transformation of the state sector in China. *Brookings Papers on Economic Activity* (1), 295–366.

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

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

Jovanovic, B. and R. Rob (1989). The growth and diffusion of knowledge. *Review of Economic Studies 56*(4), 569–582.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica 131*, 207–248.

Klenow, P. J. and A. Rodriguez-Clare (1997). The neoclassical revival in growth economics: Has it gone too far? *NBER Macroeconomics Annual 12*, 73–103.

Klette, T. J. and S. Kortum (2004). Innovating firms and aggregate innovation. *Journal of Political Economy 112*(5), 986–1018.

König, M. D., J. Lorenz, and F. Zilibotti (2016). Innovation vs. imitation and the evolution of productivity distributions. *Theoretical Economics 11*(3), 1053–1102.

Lentz, R. and D. T. Mortensen (2008). An empirical model of growth through product innovation. *Econometrica 76*(6), 1317–1373.

Lindenlaub, I. (2017). Sorting multidimensional types: Theory and application. *Review of Economic Studies 84*(2), 718–789.

Lucas, R. and B. Moll (2014). Knowledge growth and the allocation of time. *Journal of Political Economy 122*(1), 1–51.

Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics 122*(3), 1103–1144.

Luttmer, E. G. (2012). Technology diffusion and growth. *Journal of Economic Theory 147*(2), 602–622.

Perla, J. and C. Tonetti (2014). Equilibrium imitation and growth. *Journal of Political Economy 122*(1), 52–76.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics 11*(4), 707–720.

Shen, K., J. Wang, and J. Whalley (2015). Measuring changes in the bilateral technology gaps between China, India and the US 1979–2008. NBER Working Paper w21657.

Smith, H. L. (2011). *An introduction to delay differential equations with applications to the life sciences*, Volume 57. Springer New York.

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

Tombe, T. and X. Zhu (2019). Trade, migration, and productivity: A quantitative analysis of China. *American Economic Review 109*(5), 1843–72.

Zilibotti, F. (2017). Growing and slowing down like China. *Journal of the European Economic Association 15*(5), 943–988.

Supplementary Appendix

A Theory

In this section, we provide the proof of Proposition 1.

Proof of Proposition 1. From the monotonicity of $Q(a,\tau;\Pi)$ in a it follows that there exists a threshold function $a^*(\tau;\Pi)$ such that[35]

$$\begin{aligned} &Q(a,\tau;\Pi) \geq \overline{p} \text{ if } a \leq a^*(\tau;\Pi), \\ &Q(a,\tau;\Pi) < \overline{p} \text{ if } a > a^*(\tau;\Pi). \end{aligned} \tag{16}$$

All firms with $a \leq a^*(\tau;\Pi)$ imitate, while some firms with $a > a^*(\tau;\Pi)$ (i.e., those with a sufficiently large p) innovate. To simplify notation, we write $a^*(t) = a^*(\tau;\Pi)$ and $Q(a) = Q(a,\tau;\Pi)$ when this is no source of confusion.

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

$$\Pi_a(t+1) - \Pi_a(t)$$
$$= \begin{cases} q\left[(1-F_{a-1}(t))\,\Pi_{a-1}(t) - (1-F_a(t))\Pi_a(t)\right] & \text{if } \quad a < a^*(t), \\ \left(\begin{array}{c} q\,(1-F_{a-1}(t))\,\Pi_{a-1}(t) \\ -G(Q(a))\left[q(1-F_a(t))\Pi_a(t)\right] \\ -\int_{Q(a)}^{\overline{p}} \left[(p+(1-p)\,\delta q\,(1-F_a(t)))\,\Pi_a(t)\right] dG(p) \end{array} \right) & \text{if } \quad a = a^*(t)+1, \\ \left(\begin{array}{c} G(Q(a-1)) \times q\,(1-F_{a-1}(t))\,\Pi_{a-1}(t) \\ +\int_{Q(a-1)}^{\overline{p}} \left[(p+(1-p)\,\delta q\,(1-F_{a-1}(t)))\,\Pi_{a-1}(t)\right] dG(p) \\ -G(Q(a)) \times q\,(1-F_a(t))\,\Pi_a(t) \\ -\int_{Q(a)}^{\overline{p}} \left[(p+(1-p)\,\delta q\,(1-F_a(t)))\,\Pi_a(t)\right] dG(p) \end{array} \right) & \text{if } \quad a > a^*(t)+1. \end{cases} \tag{17}$$

To understand this law of motion, note that (i) if $a < a^*(t)$, all firms with productivity a and $a-1$ imitate; (ii) if $a > a^*(t)+1$, all firms with productivity a facing a realization $p > Q(a)$ and all firms with productivity $a-1$ facing a realization $p > Q(a-1)$ innovate, while all other firms with productivity a and $a-1$ imitate; (iii) if $a = a^*(t)+1$, all firms with productivity a facing a realization $p > Q(a)$ innovate, and all other firms with productivity a and $a-1$ imitate. Going from the p.m.f to the corresponding c.d.f. yields:

$$\begin{aligned} F_a(t+1) - F_a(t) &= \sum_{b=1}^{a} \Pi_b(t+1) - \Pi_b(t) \\ &= \begin{cases} -q\,(1-F_a(t))\,(F_a(t) - F_{a-1}(t)) & \text{if } \quad a \leq a^*(t) \\ \left(\begin{array}{c} -G(Q(a))\,q\,(1-F_a(t))\,(F_a(t)-F_{a-1}(t)) \\ -\int_{Q(a)}^{\overline{p}} \left[\begin{array}{c} (p+(1-p)\,\delta q\,(1-F_a(t))) \times \\ (F_a(t)-F_{a-1}(t)) \end{array} \right] dG(p) \end{array} \right) & \text{if } \quad a > a^*(t) \end{cases} \end{aligned} \tag{18}$$

[35] Note that this notation involves a slight abuse of notation relative to the function a^* in the text.

Define the complementary cumulative distribution function $H_a(t) = 1 - F_a(t)$. Equation (18) can be rewritten as:

$$\begin{aligned}
H_a(t+1) - H_a(t) &= -\sum_{b=1}^{a} \left(\Pi_b(t+1) - \Pi_b(t) \right) \\
&= \begin{cases} qH_a(t)\left(H_{a-1}(t) - H_a(t)\right) & \text{if} \quad a \leq a^*(t) \\ \left(\begin{array}{c} G\left(Q\left(a\right)\right) qH_a\left(t\right)\left(H_{a-1}\left(t\right) - H_a\left(t\right)\right) \\ + \int_{Q(a)}^{\overline{p}} \left[\begin{array}{c} \left(p + (1-p)\,\delta q H_a\left(t\right)\right) \times \\ \left(H_{a-1}\left(t\right) - H_a\left(t\right)\right) \end{array} \right] dG\left(p\right) \end{array} \right) & \text{if} \quad a > a^*(t) \end{cases}
\end{aligned} \tag{19}$$

Note that $F_a(t+1) \leq F_a(t)$ (and conversely, $H_a(t+1) \geq H_a(t)$. Since the probability mass is conserved to one (and $\lim_{a\to+\infty} F_a = 1$), the fact that F_a is decreasing over time t for every a implies that the distribution must shift to the right (i.e. towards higher values of a). A distribution that is shifted in this way is called a traveling wave (Bramson, 1983). We now prove that there exists a traveling 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 $\nu > 0$. The formal argument follows Bramson (1983) and König et al. (2016) . The traveling wave solution above implies that $F_a(t+1) - F_a(t) = \tilde{f}(x - \nu) - \tilde{f}(x)$, where $x \equiv a - \nu t$. For $\nu \approx 0$, we can take the first order approximation $\tilde{f}(x-\nu) - \tilde{f}(x) \approx -\nu \tilde{f}'(x)$, and thus $F_a(t+1) - F_a(t) \approx -\nu \tilde{f}'(x)$. Therefore, we can rewrite (18) as:

$$-\nu \tilde{f}'\left(x\right) = \begin{cases} -q\left(1 - \tilde{f}(x)\right)\left(\tilde{f}(x) - \tilde{f}(x-1)\right) & \text{if} \quad x \leq x^* \\ \left(\begin{array}{c} -G\left(Q\left(x\right)\right)\left[q(1 - \tilde{f}(x))\left(\tilde{f}(x) - \tilde{f}(x-1)\right)\right] \\ - \int_{Q(x)}^{\overline{p}} \left[\begin{array}{c} \left(p + (1-p)\,\delta q\left(1 - \tilde{f}(x)\right)\right) \times \\ \left(\tilde{f}(x) - \tilde{f}(x-1)\right) \end{array} \right] dG\left(p\right) \end{array} \right) & \text{if} \quad x > x^* \end{cases} \tag{20}$$

or, identically,

$$-\nu \tilde{h}'\left(x\right) = \begin{cases} q\tilde{h}\left(x\right)\left(\tilde{h}\left(x-1\right) - \tilde{h}(x)\right) & \text{if} \quad x \leq x^* \\ \left(\begin{array}{c} G\left(Q\left(x\right)\right)\left[q\tilde{h}\left(x\right)\left(\tilde{h}\left(x-1\right) - \tilde{h}(x)\right)\right] \\ + \int_{Q(x)}^{\overline{p}} \left[\begin{array}{c} \left(p + (1-p)\,\delta q\tilde{h}\left(x\right)\right) \times \\ \left(\tilde{h}\left(x-1\right) - \tilde{h}\left(x\right)\right) \end{array} \right] dG\left(p\right) \end{array} \right) & \text{if} \quad x > x^* \end{cases} \tag{21}$$

Consider, first, the range $x \leq x^*$. Using the upper part of (20) yields the following Delay Differential Equation (DDE):[36]

$$-\nu \tilde{f}'\left(x\right) = -q\left(1 - \tilde{f}\left(x\right)\right)\left(\tilde{f}\left(x\right) - \tilde{f}\left(x-1\right)\right). \tag{22}$$

This equation allows us to characterize the (asymptotic) left tail of the distribution. Taking the limit for $x \to -\infty$, we can take the following first-order (i.e., linear) approximation:

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

[36]See also Asl and Ulsoy (2003), Bellman and Cooke (1963), and Smith (2011).

Next, we guess that this linear DDE has a solution of the form $\tilde{f}(x) = c_1 e^{\lambda x}$ 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 guess as long as the following transcendental equation in λ is satisfied:

$$\lambda\nu \simeq q(1 - e^{-\lambda}). \tag{23}$$

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 (21). Then, we can write the following DDE

$$-\nu\tilde{h}'(x) = \left(\begin{array}{c} G\left(Q\left(x\right)\right)\left[q\tilde{h}\left(x\right)\left(\tilde{h}\left(x-1\right) - \tilde{h}(x)\right)\right] \\ +\int_{Q(x)}^{\overline{p}}\left[\begin{array}{c}\left(p + \left(1-p\right)\delta q\tilde{h}\left(x\right)\right)\times \\ \left(\tilde{h}\left(x-1\right) - \tilde{h}\left(x\right)\right)\end{array}\right] dG\left(p\right)\end{array}\right). \tag{24}$$

We use this DDE to characterize the right tail of the distribution as $x \to +\infty$. Again, we take a linear approximation:

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

where $\hat{p} = \int_0^{\overline{p}} p\, dG(p)$. For the latter, note that $\lim_{x\to\infty} Q(x) = 0$ since as we take x to be arbitrarily large, imitation becomes totally ineffective and firms choose to innovate almost surely. 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:

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

The solution to the transcendental equation satisfies

$$\rho = \frac{-\nu W\left(-\frac{\hat{p}e^{-\frac{\hat{p}}{\nu}}}{\nu}\right) - \hat{p}}{\nu}, \tag{25}$$

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. ■

B Data and descriptive statics

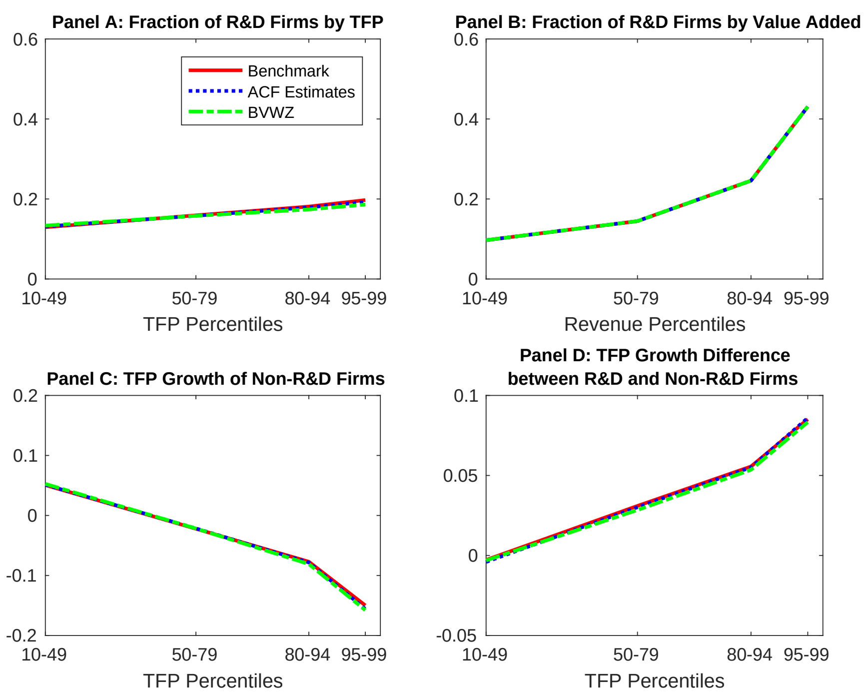
In this section we provide some details of the analysis in Section 3.

Alternative TFP estimation method based on Brandt et al. (2017)

We estimate firm-level TFPs using the methodology of Hsieh and Klenow (2009). This is consistent with our theoretical model and allows us to directly compare our results with those in the literature on misallocation. However, this approach has been criticized in the empirical industrial organization literature. If firms optimally choose the inputs in the production process to solve a dynamic maximization problem, then the estimation may suffer from an endogeneity problem. The error term of the model can contain determinants of production decisions (e.g., productivity) that are observed by the firm but not by the econometrician, leading to inconsistent estimates of TFP.

In this section, we show that the target moments of our estimation are essentially unchanged if we estimate TFP using the methodology of Ackerberg et al. (2015) that addresses an endogeneity problem in the estimation of production functions. We follow the implementation of Ackerberg et al. (2015) proposed by Brandt et al. (2017), which is also related to De Loecker and Warzynski (2012). Because Brandt et al. (2017) postulate a gross production function while we estimate TFP using a value added approach, we perform an adjustment for the two methods to be consistent. The details of the estimation are deferred to the web appendix. The results are shown in Figure A1. The data moments are indistinguishable from those used targets in our estimation. We conclude that our results are robust to using this alternative estimation method for TFP.

Figure A1: China 2007–12 Sample with Alternative TFP Measurement



Note: The figure shows the equivalent moments to Appendix Figure A3 when TFP is estimated using the methodology of Brandt et al. (2017) based on Ackerberg et al. (2015).

Regression with firm fixed effects 2001–07

In this section, we present the result of regression similar to those in Table 2 although for an earlier sample in China, 2001-2007. Since this sample has R&D data for more than one year, this sample allows to us to also run regressions with firm fixed effects. Note that the regressions in Table A1 are all on annual data, the reason being that we only have R&D data for 2001-2003 and 2005-2007. The regressions in columns (1)-(3) are pooled regressions while columns (4)-(5) are firm fixed effects (FE) regressions. All pooled regressions include dummies for year, province, industry, and age effects while the FE regressions have year and firm effects. In the FE regressions the dummies for province, industry, and age, as well as the dummies for export firms and state ownership are all subsumed in the firm fixed effects.

Table A1: Regressions with Firm Fixed Effects, 2001–07.

Dependent variable: R&D dummy

	(1)	(2)	(3)	(4)	(5)
	Pooled regressions			FE regressions	
	R&D$_d$	R&D$_d$	R&D$_d$	R&D$_d$	R&D$_d$
log(TFP)$_t$	0.059***	0.377***	0.323***	0.003***	0.186***
	(0.0049)	(0.0235)	(0.0191)	(0.0007)	(0.0054)
wedge		-0.432***	-0.361***		-0.225***
		(0.0297)	(0.0240)		(0.0066)
export$_d$			0.045***		
			(0.0118)		
SOE$_d$			0.124***		
			(0.0148)		
Firm effects	-	-	-	+	+
Year effects	+	+	+	+	+
Industry effects	+	+	+	-	-
Age effects	+	+	+	-	-
Province effects	+	+	+	-	-
R-squared	0.396	0.449	0.456	0.581	0.582

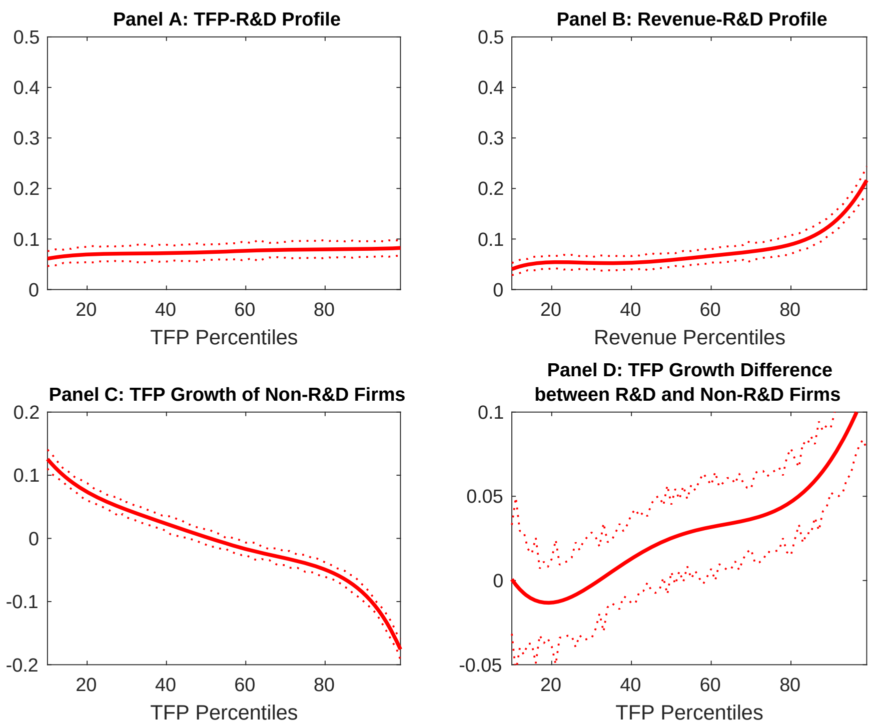
Note: The table shows regressions of an indicator of R&D on annual data for China 2001-2007. The independent variable is R&D$_d$, a dummy variable for R&D that equals one if firm R&D expenditure is positive and zero otherwise. log(TFP) is the logarithm of TFP. Wedge refers to the calibrated firm output wedge (see Section 2 for details). export$_d$ is a dummy variable for exports. SOE$_d$ is a dummy variable for state-owned firms. Standard errors are reported in parenthesis. The number of observations are 441,039 in columns (1)-(3) and 70,273 in columns (4)-(5). Observations are weighted by employment and standard errors are clustered by industry. Regressions in columns (4)-(5) include firm and year fixed effects. Regressions in columns (1)-(3) include year, industry, age, and province fixed effects. We drop firms with TFP in the bottom 10 percentiles.

The results in Table A1 show that, first, all the results from our main sample (2007-2012) in Table 2 hold up for the earlier sample, see the pooled regressions in Table A1. Second, the qualitative results hold up (significantly so) even when controlling for firm fixed effects. When comparing columns (2) and (5), we observe that the coefficients in the FE regressions are about half the size in magnitude but still highly significant. Moreover, the coefficients have always the same sign as in the pooled regressions. We conclude that our main empirical findings on the drivers of firms doing R&D—namely that R&D is positively associated with TFP and negatively associated with output wedges—holds true both in the cross section and within firms over time.

Alternative classification of innovative firms

In our main analysis, we classify all firms that report doing some R&D as innovative. However, many firms invest a very small amount of resources in R&D raising questions to whether innovation is truly a salient strategy for these firms. In this appendix, we propose an alternative classification where firms are deemed innovative only if they invest more than 1.73% of their value added. This threshold is the median R&D intensity among R&D firms in our balanced sample. Conversely, firms investing less than 1.73% are regarded as imitators. Figure A2 shows the data moments corresponding to Figure 2 when applying this more stringent definition of innovators. As one would expect, the percentage of R&D firms is now lower. Moreover, the elasticity of R&D to TFP and size is lower than for the main sample. However, the qualitative patterns are the same in Figures 2 and A2.

Figure A2: More stringent classification of innovative firms.



Note: The figure shows the equivalent moments to Figure 2 if only firms with R&D expenditure above 1.73% are classified as R&D firms.

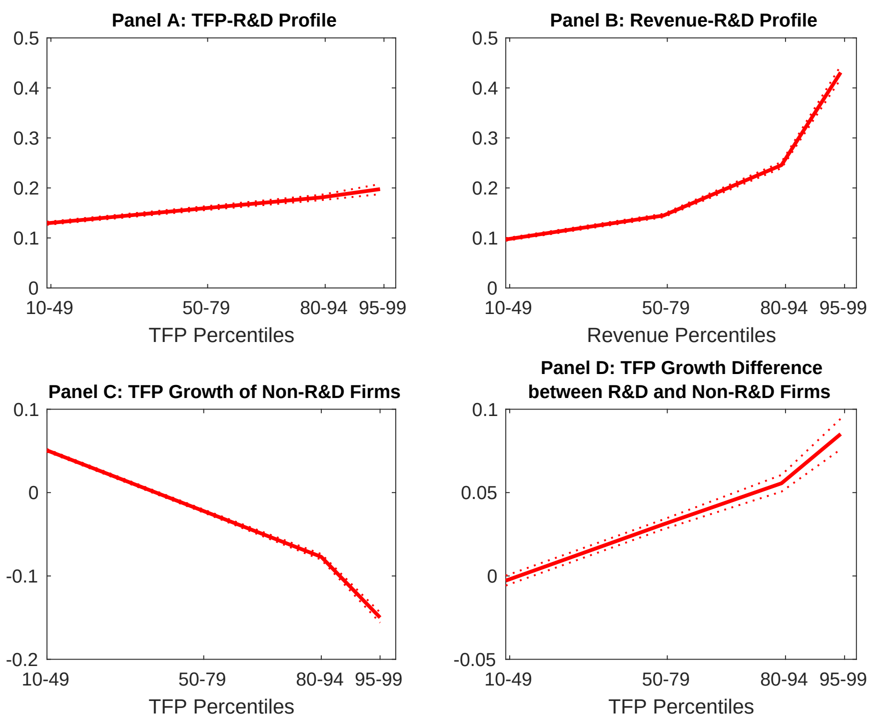
Regression results for Taiwan

In Table A2 we report the regression results discussed in the text for Taiwan. By comparing Table A2 with Table 2, it is clear that all qualitative results are the same for Taiwan as for our 2007-12 China data. However, R&D in Taiwan are more highly correlated with TFP and more negatively correlated with wedges, and TFP growth is more strongly related to R&D. In particular, the coefficients on TFP and wedges in Panel A (explaining the R&D decision) of Table A2 are about twice as large in magnitude as in Table 2. Morever, the coefficient on R&D in Panel B (explaining TFP growth) of Table A2 are about three times as large in magnitude as in Table 2.

Target moments of the empirical distribution

Figure A3 displays the empirical moments that we target in the estimation. Each observation represents a quantile of the distribution based on Figure 2.

Figure A3: Chinese Firms in the Balanced Panel 2007–12



Note: The figure shows the empirical moments of the balanced panel for China 2007–12. See also Figure 4. The dotted lines represents standard errors.

Table A2: Balanced Panel of Taiwanese Firms, 1999–2004.

PANEL A: Correlations between firm characteristics and R&D decision.

	(1) R&D$_d$	(2) R&D$_d$	(3) R&D$_d$
log(TFP)	0.087***	0.614***	0.573***
	(0.0078)	(0.0184)	(0.0189)
wedge		-0.720***	-0.673***
		(0.0234)	(0.0236)
export$_d$			0.088***
			(0.0227)
Industry effects	+	+	+
Age effects	+	+	+
Year effects	+	+	+
Observations	44,326	44,326	44,326
R-squared	0.219	0.396	0.404

PANEL B: Correlations between firm initial characteristics and TFP growth.

	(1) TFP growth	(2) TFP growth	(3) TFP growth
log(TFP)	-0.064***	-0.066***	-0.066***
	(0.0046)	(0.0047)	(0.0055)
R&D$_d$	0.103***	0.093***	
	(0.0162)	(0.0168)	
export$_d$		0.039***	0.039***
		(0.0063)	(0.0063)
R&D intensity$_h$			0.072**
			(0.0262)
R&D intensity$_m$			0.106***
			(0.0278)
R&D intensity$_l$			0.097***
			(0.0150)
Industry effects	+	+	+
Age effects	+	+	+
Observations	9,996	9,996	9,996
R-squared	0.081	0.083	0.084

Note: The table reports the analogues of the regression results in Table 2 for the Taiwanese firms. There are two differences relative to Table 2: (1) Panel A reports pooled regressions with year fixed effects (for Taiwan, data for R&D expenditure is available for multiple years); and (2) there is no province dummy for Taiwan.

C Estimation

Measurement error: Mapping

This appendix describes how we incorporate measurement error when calculating the theoretical moments. We provide an analytical mapping from the theoretical distribution of a variable x to the observed distribution of true x plus m.e. This analytical mapping is critical to speed up the structural estimation avoiding a computational curse that would arise if we had to rely on simulations.

The presentation focuses on TFP. The approach for adding m.e. to the theoretical (true) distribution of value added is equivalent, replacing a and μ with y and μ_y below.

Denote by $\hat{a}$ and a the observed and true log TFP: $\hat{a} = a + \mu$, where μ is m.e. Consider the following discrete state space: $a \in \{\delta, \cdots, N\delta\}$, $\hat{a} \in \{\delta, \cdots, N\delta\}$, and $\mu \in \{-N^{\mu}\delta, \cdots, -\delta, 0, \delta, \cdots, N^{\mu}\delta\}$. We set $N^{\mu} = 4$.

Let the theoretical distribution of a be denoted by $\Pi(a)$. The first task is to convert $\Pi(a)$ to $\Pi(\hat{a})$ – i.e., the distribution of observed TFP with m.e., which can be compared with the data. To this end, we first derive the transition matrix $\Pi(\hat{a}|a)$. For $j \in \{2, \cdots, N-1\}$, we have

$$\Pi(\hat{a} = a_j | a = a_i) = \Pi(\mu = (j-i)\delta). \tag{26}$$

For $j = 1$ or N, we have $\Pi(\hat{a} = a_1 | a = a_i) = \sum_{k \geq i-1} \Pi(\mu = -k\delta)$ and $\Pi(\hat{a} = a_N | a = a_i) = \sum_{k \geq N-i} \Pi(\mu = k\delta)$. So, the unconditional probability of $\hat{a}$ is

$$\Pi(\hat{a} = a_j) = \sum_i \Pi(\hat{a} = a_j | a = a_i) \Pi(a = a_i). \tag{27}$$

Note that when $\Pi(\hat{a})$ is observable while $\Pi(a)$ is unknown, one can use $\Pi(\hat{a} = a_j | a = a_i)$ in (26) to back out $\Pi(a)$ by solving the system of equations in (27).

We now derive the conditional TFP growth. Let us start with observed TFP growth of imitating firms.

$$\begin{aligned} \mathbb{E}^{IM}[\Delta\hat{a}|\hat{a}] &= \mathbb{E}^{IM}[\Delta a + \Delta\mu|\hat{a}] \\ &= \mathbb{E}[q(1-F(a))|\hat{a}] - \mathbb{E}[\mu|\hat{a}] \\ &= \sum_i q(1-F(a))\Pi(a = a_i|\hat{a} = a_j) - \sum_k k\delta\Pi(\mu = k\delta|\hat{a} = a_j). \end{aligned} \tag{28}$$

To go from the theoretical (conditional) distribution of true TFP growth conditional on true a to TFP growth with m.e. conditional on $\hat{a}$, we need conditional probabilities of $\Pi(a = a_i|\hat{a} = a_j)$ and $\Pi(\mu = k\delta|\hat{a} = a_j)$.

The posterior distribution of a follows

$$\Pi(a = a_i|\hat{a} = a_j) = \frac{\Pi(\hat{a} = a_j|a = a_i)\Pi(a = a_i)}{\Pi(\hat{a} = a_j)}, \tag{29}$$

To obtain the posterior distribution of μ, first notice that

$$\begin{aligned} \Pi(\hat{a} = a_j \cap \mu = k\delta) &= \Pi(\hat{a} = a_j|\mu = k\delta)\Pi(\mu = k\delta) \\ &= \Pi(a = a_{j-k})\Pi(\mu = k\delta). \end{aligned}$$

for $j \in \{2, \cdots, N-1\}$. Note that for $j = 1$ or N, we have the following boundary cases:

$$\begin{aligned}
\Pi(\hat{a} = a_1 | \mu = -i\delta)\,\Pi(\mu = -i\delta) &= \sum_{k \le i+1} \Pi(a = a_k)\,\Pi(\mu = -i\delta), \\
\Pi(\hat{a} = a_N | \mu = i\delta)\,\Pi(\mu = i\delta) &= \sum_{k \ge N-i} \Pi(a = a_k)\,\Pi(\mu = i\delta).
\end{aligned}$$

Then, the posterior distribution of μ follows

$$\begin{aligned}
\Pi(\mu = k\delta | \hat{a} = a_j) &= \frac{\Pi(\hat{a} = a_j | \mu = k\delta)\,\Pi(\mu = k\delta)}{\Pi(\hat{a} = a_j)} \\
&= \frac{\Pi(\hat{a} = a_j \cap \mu = k\delta)}{\Pi(\hat{a} = a_j)}.
\end{aligned} \tag{30}$$

We can thus use (28), together with (29) and (30), to generate TFP growth of imitating firms with measurement errors.

Measurement error: Moments

Table A3 reports the empirical moments we use to derive the moments involving measurement error.

Table A3: Measurement Error Moments

	Empirical variances				Implied m.e. variance		
	$var(\Delta Y)$	$var(\Delta I)$	$cov(\Delta Y, \Delta I)$	$var(TFP)$	$\hat{v}_{\mu y}$	$\hat{v}_{\mu I}$	$\hat{v}_{\mu a}$
2007-2012 China	0.456	0.328	0.156	1.059	0.150	0.086	0.320
2001-2007 China	0.470	0.098	0.043	1.269	0.214	0.027	0.361
1999-2004 Taiwan	1.128	0.124	-0.005	2.363	0.567	0.065	0.950

Note: The first four columns refer to variances of growth in revenue, inputs, and TFP, respectively. Y, I and TFP represent $\log(P_{it}Y_{it})$, $\log(K_{it}^{\alpha}L_{it}^{(1-\alpha)})$, and $\log(A_{it})$, respectively. The fourth column is cross-sectional dispersion in TFP. We use the full sample (i.e., keeping the firms with initial TFP in the bottom ten percentiles). The results in the trimmed sample are similar. The implied variance of measurement error is derived from equations (11)-(12) and the expression for m.e. in TFP.

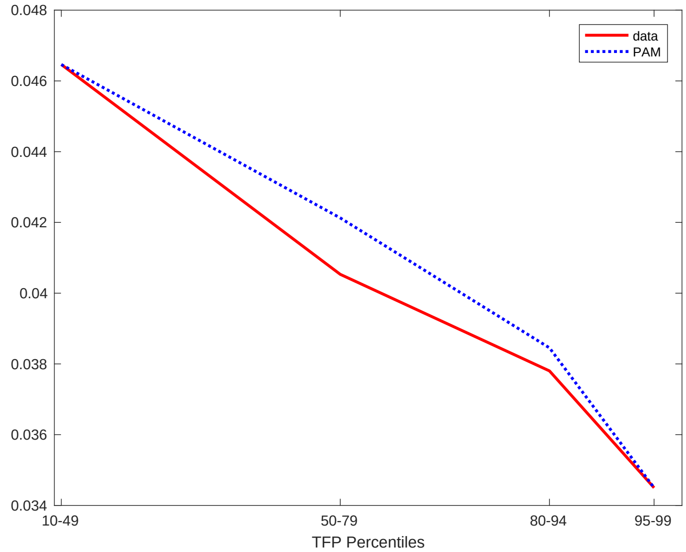
Calibration of θ

Figure A4 displays how the ratio of R&D intensity for R&D firms changes with TFP in the data (solid line) and in the benchmark PAM model. The figure shows R&D intensity for each quantile relative to the intensity for firms in the lowest quantile, normalized to unity. The technological parameter θ is calibrated to match the slope between the first and last quantile in Figure A4.

D Results

In this section we report robustness estimation results referred to in the text.

Figure A4: Ratio of R&D to Value Added, PAM vs. Data



Note: The figure shows the average ratio of R&D to value added for R&D firms in the data (solid line) and the Parsimonious model (dotted line) for four quantiles of the TFP distribution.

Estimation of the Fake R&D model

Figure A5 shows the fit of the Fake R&D model (FRM). The blue line in each panel represents moments from simulated data based on firms claiming to do (or not to do) R&D. The corresponding moments for a true classification of R&D investments are represented by red lines in the figure.

Intensive Margin

In this section, we document the fit of the two models estimated in Section 5.3.

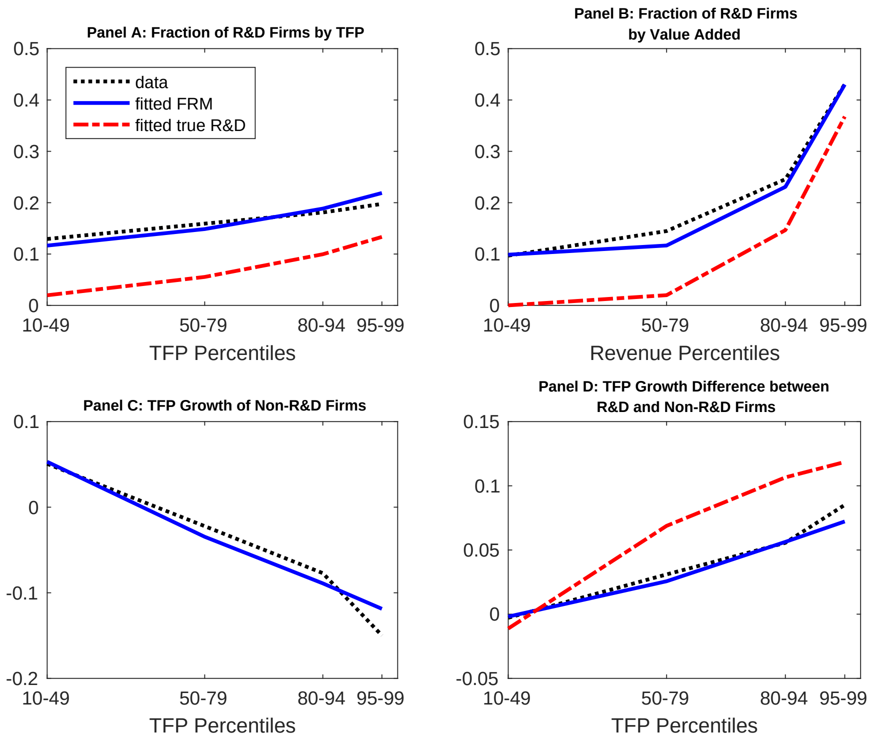
Figure A6 is the equivalent of Figure A3 and shows how the 16 target moments are affected by applying the more stringent classification of innovation, based on Figure A2. Figure A6 also shows the fit of the PAM when reestimated based on these adjusted moments. The estimated coefficients for this version of the PAM model are reported in columns (5)–(6) of Table 3.

Figure A7 shows the fit of the PAM with two R&D technologies (small and large R&D projects). The estimated coefficients for the PAM model with two R&D technologies are reported in columns (7)–(8) of Table 3. The empirical moments for the high- and low-intensive R&D firms are illustrated by black dotted lines in Figure A7.

E Estimating the Model on Different Samples

Figures A8 and A9 show the fit of the PAM and IPM for China in the earlier sample (2001–07) and for Taiwan.

Figure A5: China: Fake R&D Model



Note: Each panel of Figure A5 displays three schedules: (i) the dotted line shows the moments in the data, (ii) the dashed line shows the fit of the model (which refers to measured R&D), and (iii) the solid line shows results restricted to the firm which, according to the model predictions, truly perform R&D. See also Figure 4.

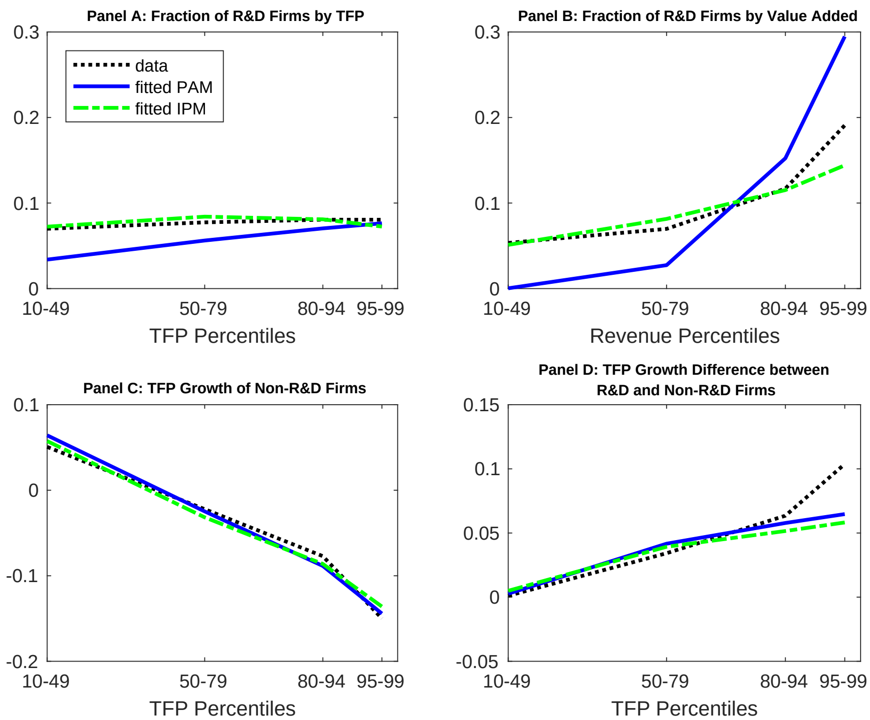
F Counterfactuals

This section provides robustness analysis for the counterfactuals.

Models with heterogeneity in innovation costs

Figures A10–A11 are the analogues of Figures 8–9 in the manuscript for the IPM that we estimated on the benchmark balanced sample for China, 2007–12. Table A4 is the analogue of Table 6 in the manuscript for the IPM.

Figure A6: Models with Higher R&D Cutoff



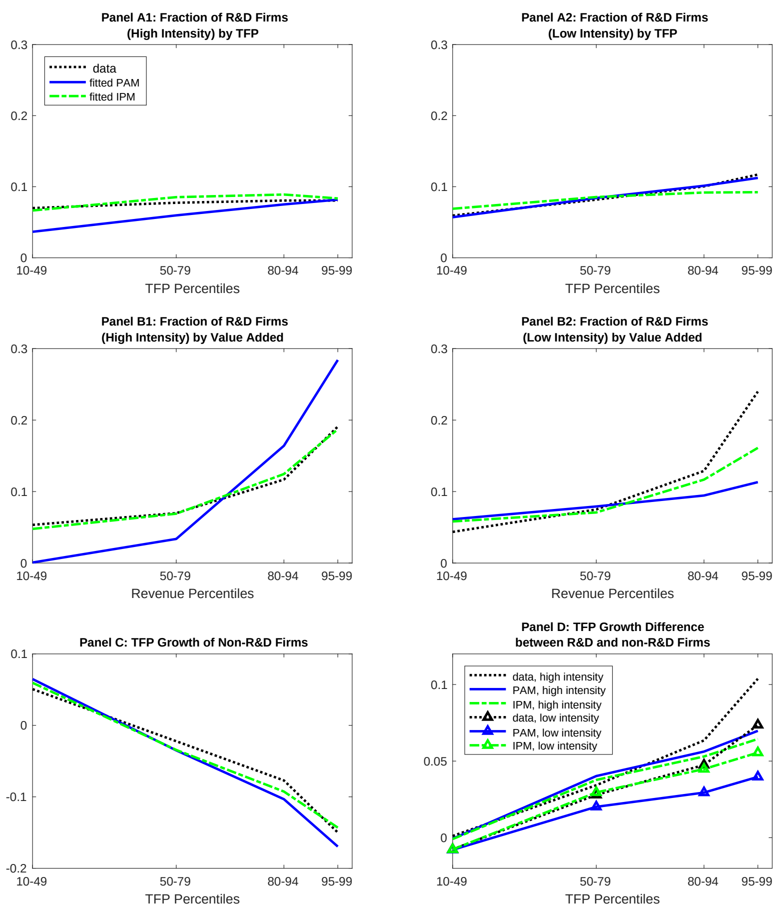
Note: The models are estimated to match empirical moments where only firms with R&D intensity exceeding (1.73% of value added are classified as innovative firms. See Figure 4 for additional information.

Table A4: Counterfactuals, Industrial Policy Model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IPM estim. model	50% lower output wedges	Taiwan's q	Taiwan's $\bar{p}$ and $\bar{c}$	Taiwan's $\bar{p}$, $\bar{c}$, and q	Increase $\bar{c}$ so share R&D firms = 5%	Decrease $\bar{c}$ so share R&D firms = 20%	All firms do R&D
Fraction of R&D Firms (%)	14.9	16.0	14.8	8.24	8.17	5	20	100
Steady State TFP Growth (%)	5.05	6.20	5.11	5.71	5.78	3.64	5.39	4.41

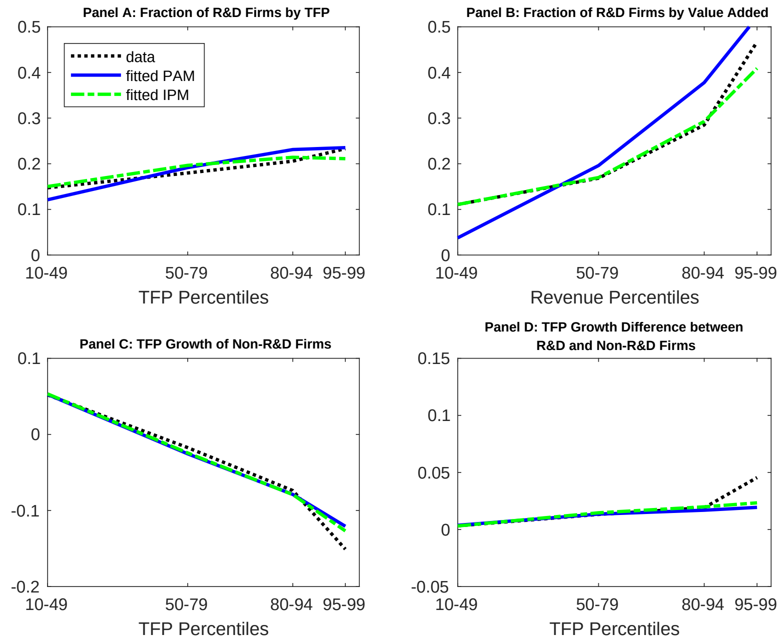
Note: The table reports statistics for the counterfactual experiments for the IPM discussed in the text. Column (1) reports the predicted moments of the estimated IPM for comparison.

Figure A7: High- and low-R&D firms



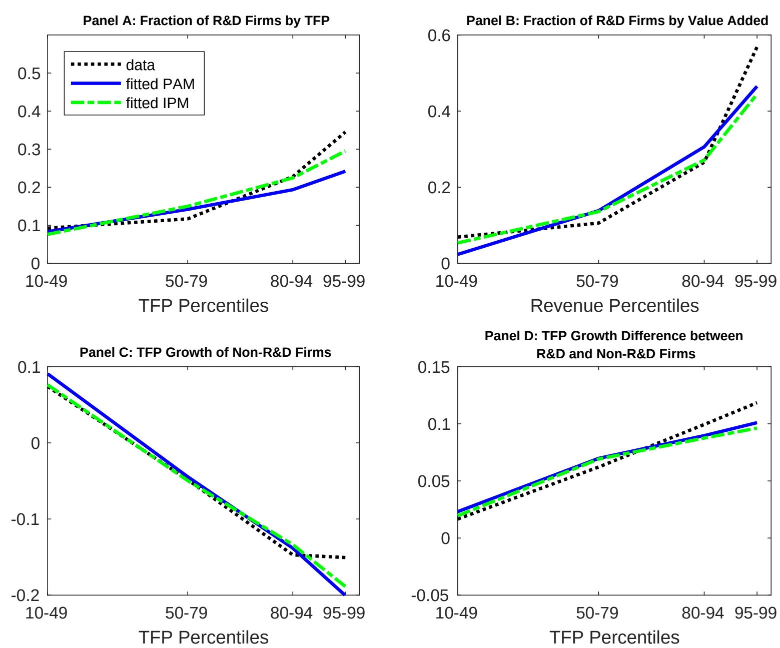
Note: The figure shows empirical and theoretical moments for the extension to two R&D technologies (Section 5.3). Panels A1-A2 and B1-B2 correspond to Panels A and B in Figure 4, reported separately for firms with high- versus low-cost R&D technology. Panels C and D correspond to their counterparts in Figure 4,

Figure A8: China 2001–07: PAM and IPM



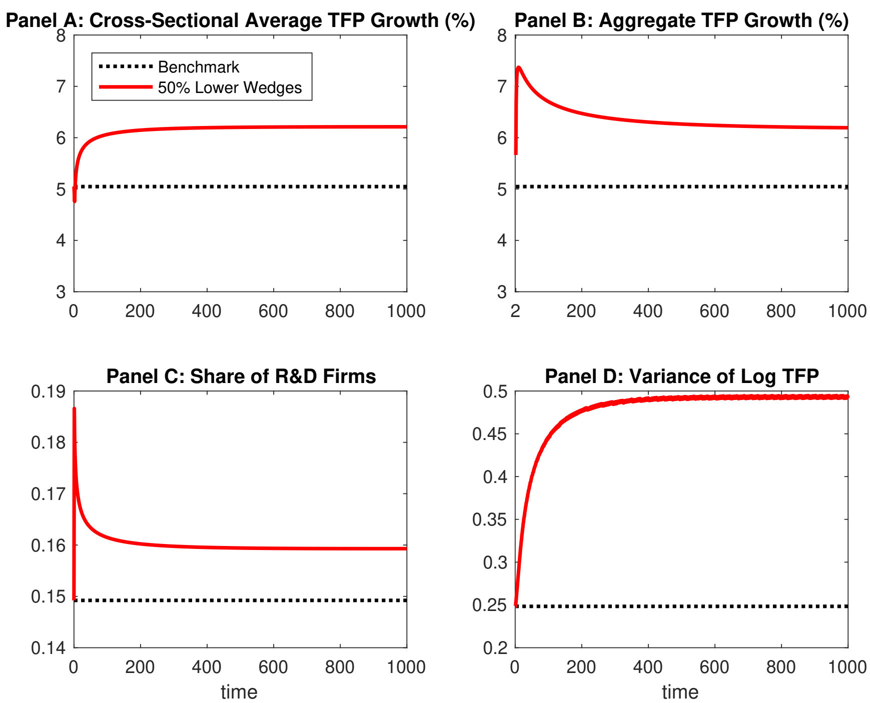
Note: See Figure 4.

Figure A9: Taiwan 1999–2004: PAM and IPM



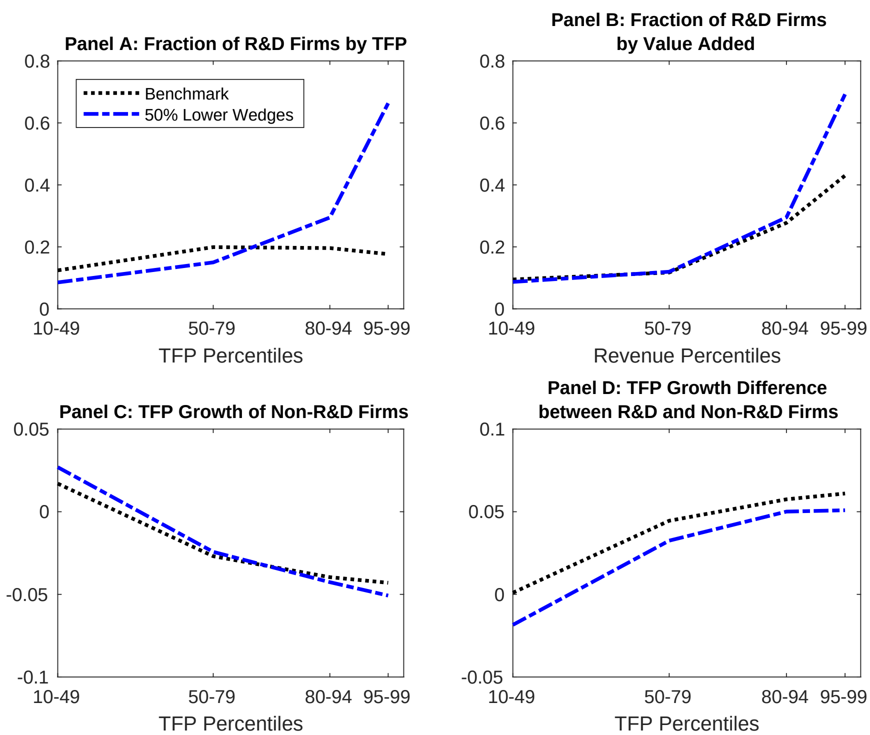
Note: See Figure 4.

Figure A10: Transition: Lower Wedges in IPM



Note: This figure is the IPM analogue of Figure 8.

Figure A11: Steady-State: Lower Wedges in IPM



Note: This figure is the IPM analogue of Figure 9.