FROM IMITATION TO INNOVATION: WHERE IS ALL THAT CHINESE R&D GOING?

By

Michael König, Kjetil Storesletten, Zheng Song, and Fabrizio Zilibotti

June 2020

COWLES FOUNDATION DISCUSSION PAPER NO. 2237



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

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

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

June 12, 2020

Abstract

We construct a model of firm dynamics with heterogenous productivity and distortions. The productivity distribution evolves endogenously as the result of the decisions of firms seeking to upgrade their productivity over time. Firms can adopt two strategies toward that end: imitation and innovation. The theory bears predictions about the evolution of the productivity distribution. 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 2007–2012. The estimated model fits the Chinese data well. We compare the estimates with those obtained using data for Taiwan and find the results to be robust. We perform counterfactuals to study the effect of alternative policies. We find large effects of R&D misallocation on long-run growth.

JEL Codes: O31, O33, O47

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

*We thank Joe Altonji, Ufuk Akcigit, Manuel Amador, Marco Bassetto, Chang-Tai Hsieh, Chad Jones, Patrick Kehoe, Juan Pablo Nicolini, Jesse Perla, Torsten Persson, Chris Tonetti, Daniel Xu, Gianluca Violante, and seminar participants at the ASSA Annual Meeting 2019, the Bank of Canada-IMF-University of Toronto Meeting on the Chinese Economy, Cambridge University, CIFAR, the Federal Reserve of Minneapolis, the National University of Singapore, New York University, Princeton University, Stanford University, Tsinghua University, the University of California at Berkeley, the University of Chicago, University College London, 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 Oslo, Department of Economics, kjetil.storesletten@econ.uio.no.

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

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

1 Introduction

This paper analyzes the determinants of productivity growth from both theoretical and empirical standpoints. We construct a dynamic equilibrium model of endogenous technical change where firms are heterogenous in productivity and distortions. The evolution of the productivity distribution hinges on the decisions of profit-maximizing firms seeking to upgrade their technology. To achieve this goal, firms face a binary choice: they can either attempt to learn better technologies through random interactions with other firms (*imitation*) or try to break new ground and independently discover new technologies (*innovation*). Focusing on innovation requires an investment and possibly entails some opportunity cost of learning through random interactions.[1] The state of firms' productivity determines the comparative advantage of the two alternative strategies. The further the firm is from the technology frontier, the more likely it is to succeed in learning through random interactions. Conversely, for firms close to the technology frontier, the scope for imitating other firms is limited, and they must break new ground in order to improve their technology. The investment decision is also affected by other firm characteristics, most notably, labor and capital market distortions (wedges).

The main contribution of the paper is a novel empirical implementation of the theory: we propose a structural estimation method exploiting the stationary equilibrium of the dynamic model. We estimate the model using the Simulated Method of Moments (SMM) by targeting moments of the empirical distribution of R&D and productivity growth that are salient in the theory. In our main application, we use data from manufacturing firms in mainland 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)).[2] However, there is only limited evidence about the return to these investments and their effect on aggregate growth.

We proxy the strategy of firms by their R&D investment behavior on the extensive margin and regard firms making R&D investments as innovators and firms not making R&D investments as imitators. We first show that the predictions of the theory conform with a set of cross-sectional data moments about the behavior and performance of Chinese firms. Then, we estimate the model. A parsimonious model fits the data reasonably well. Next, we extend the model to allow for heterogeneity across firms in R&D costs. We interpret these firm-specific "R&D wedges" as R&D subsidies and allow them to be correlated with firm-level TFP. The estimated pattern is suggestive of an active industrial policy, which is arguably a salient feature of China. The richer models with R&D wedge heterogeneity fit the target moments very accurately. Finally, motivated by the findings of Chen et al. (2018), we also 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 generous public subsidies. This model predicts that R&D investment is overestimated in the data.

We study the success of the theory in predicting a set of nontargeted moments. We first show that the estimated model generates conditional correlations that closely match the coefficients on key

[1] In our theory, the investment could either enhance or dampen firms' ability to absorb ideas from other firms. Our estimation suggests a trade-off: focusing on innovation reduces a firm's ability to absorb knowledge from other firms.

[2] The transition toward innovation-based growth is a central theme in the discourse of the Chinese political leadership. The 13th Five-Year Plan (2016–2020) emphasizes the importance of promoting research in strategic and frontier fields. The National Innovation-Driven Development Strategy Outline issued in June 2016 lays out an ambitious vision by which China should transform into an innovation-oriented economy by 2020 and into a technological innovation powerhouse by 2050. Consistent with this new emphasis, there has a been a boom in Chinese R&D investment, a standard measure of innovative investments. While China invested barely 1% of its GDP in the 1990s, R&D investments increased to 2.12% of GDP by 2017.

covariates in nonstructural linear regressions whose dependent variables are (i) the determinants of the probability of firms to engage in R&D; and (ii) the determinants of TFP growth rates. Second, the theory bears predictions about the (stationary) TFP distribution and the endogenous TFP growth rate. The estimated model replicates well the TFP distribution in the data, with both left and right Pareto tails that are quantitatively close to those observed in the data. Moreover, the theory predicts an annual TFP growth rate for the median firm of 3.8%, which is the same as in the data for China from 2007–12.

In an extension, we re-estimate the model using plant-level data from Taiwan, for which census data about R&D investments are also available. Taiwan is a natural comparison for mainland China, not only because of its geographic and cultural proximity, but also because of the structural similarities between the two economies. In both cases, economic development has been led by the manufacturing sector, first through the growth of low value-added industries and later through a transition to a higher level of technological sophistication (e.g., electronics and semiconductors). Both economies still have today a relatively large manufacturing sector accounting for about 30% of the respective GDPs. The results for Taiwan are qualitatively similar to those obtained for China, although both the productivity of R&D and the rate of technology diffusion appear to be higher for Taiwanese firms.

Finally, we perform counterfactual policy experiments. Among them, we consider a reduction in misallocation, as measured by the variance of output wedges. We measure these wedges using the empirical methodology proposed by Hsieh and Klenow (2009). That paper finds that misallocation is an important drag on TFP for China. In our paper, static output wedges affect the decision to invest in R&D and the dynamics of TFP distribution. Firms' R&D investments are motivated by the desire to improve productivity, which in turn increases future profits. This decision is subject to a classic appropriability problem: a positive output wedge reduces the gains associated with a future TFP increase. A negative wedge has the opposite effect. In an undistorted economy, the level of TFP is the main driver of the decision to invest in R&D: firms with high TFP invest in R&D while firms with low TFP find imitation to be a more attractive strategy because the probability of improving technology through random interaction is higher. Thus, R&D is highly correlated with TFP. The presence of heterogenous output wedges lowers this correlation because the decision to invest in R&D is driven more by the wedge and less by TFP–or, identically, a firms' size matters more than a firm's TFP. We find that halving the variance of output wedges has a large permanent effect on TFP growth. In the benchmark model, annual TFP growth increases from 3.8% to 5.1%.

This novel result points to a general insight that goes beyond China: a reduction in static misallocation brings about an increase in productivity growth by increasing the relative productivity of high-TFP firms. Thus, reducing the variance of output wedges increases the productivity dispersion across firms, both along the transition and in the stationary (long–run) distribution.

1.1 Literature review

Our study is broadly related to various streams of the growth and development literature. First, it contributes to literature on the determinants of success and failure in technological convergence (e.g., Hall and Jones (1999), Klenow and Rodriguez-Clare (1997), Acemoglu and Zilibotti (2001), Hsieh and Klenow (2010)). The importance of technology diffusion stretches back to the seminal work of Griliches (1957). The importance of R&D investment and spillovers for growth and technology diffusion is a core element of the neo-Schumpeterian theory à la Aghion and Howitt (1992); see also Griliches (1998). While many of these papers emphasize a process of creative destruction whereby new firms are carriers of innovation, recent research by Garcia Macia, Hsieh, and Klenow (2019) finds that a larger share of

the productivity improvements originate from incumbent firms. This evidence is consistent with the tenets of our theory.

The dichotomy between innovation and imitation in the process of development is emphasized by Acemoglu, Aghion, and Zilibotti (2006). The important role of misallocation as a determinant of aggregate productivity differences is related to the influential work of Hsieh and Klenow (2009). Our study builds on their methodology, although it attempts to endogenize the distribution of productivity across firms, which is instead exogenous in their work. The importance of misallocation in China is also emphasized, among others, by Song, Storesletten, and Zilibotti (2011), Hsieh and Song (2016), Cheremukhin et al. (2017), Brandt, Kambourov, and Storesletten (2017), and Tombe and Zhu (2019).

Our paper is also closely related to the recent theories describing the endogenous evolution of the distribution of firm size and productivity. These studies include Jovanovic and Rob (1989), Luttmer (2007, 2012), Ghiglino (2012), Perla and Tonetti (2014), Acemoglu and Cao (2015), Lucas and Moll (2015), König, Lorenz, and Zilibotti (2016), Benhabib, Perla, and Tonetti (2017), Akcigit et al. (2018), among others. While some of these papers emphasize the process of learning through random interactions, the only paper that focuses on the trade-off between innovation and imitation is the theoretical paper by König, Lorenz, and Zilibotti (2016) upon which the model of this paper builds.

Finally, our paper is related to the literature studying the effect of R&D on growth and the effect of policy on R&D investments. These studies include, among others, Klette and Kortum (2004), Lentz and Mortensen (2008), Acemoglu and Cao (2015), Akcigit and Kerr (2017), and Acemoglu et al. (2018).

1.2 Literature on China's policy to stimulate innovation

Our paper is also related to the empirical literature studying R&D policy in China. Ding and Li (2015) provide a comprehensive overview of the instruments adopted by the Chinese government intervention to foster R&D. The systematic policy intervention to stimulate innovation had its 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. Next, an 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, Hsieh, and Song 2016).

Some empirical studies attempt to evaluate the effects 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. In spite of 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 much 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 Motoyashi (2013) 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 2007–2013 to assess the effects of tax incentives on firm R&D expenditures and analyze how institutional conditions shape these effects. They show that tax incentives have significant effects on the R&D expenditure reported by firms. A 10% reduction in R&D user costs leads firms to increase R&D expenditures by 4% in the short run. They also document considerable effect heterogeneity: tax incentives significantly stimulate R&D in private firms but have less influence on state-owned enterprises' R&D expenditures.

Chen et al. (2018) 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 non-R&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 for Taiwan and mainland China. Section 4 presents the methodology of the structural estimation and Sections 5-7 lay out the results and the counterfactual experiments. Section 8 concludes. An appendix contains technical results and additional tables and figures. Further technical details are deferred to a webpage 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 homogenous final good by a Dixit-Stiglitz aggregator with a constant elasticity of substitution $\eta > 1$ between goods, implying $Y = \left(\int_0^1 Y_i^{(\eta-1)/\eta}\right)^{\eta/(\eta-1)}$.

Firms are owned by overlapping generations of two-period–lived manager-entrepreneurs as in Song, Storesletten, and Zilibotti (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 makes the R&D investment decision affecting next-period productivity with 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.

2.1 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. More formally, firms maximize profits taking factor prices as given, but their decisions are distorted by a set of firm-specific output wedges τ_i. We interpret τ_i as a tax levied by a government that redistributes the proceeds as a lump sum to consumers. Note that $\tau_i < 0$ indicates a negative wedge, or an implicit output subsidy.[3]

Since 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–i.e., 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.

2.2 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. Let $\hat{a}_i \equiv \log(A_i)$. 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: $\hat{a}_{i,t+1} = \hat{a}_{i,t} + \tilde{a}$, where $\tilde{a} > 0$ is a constant (thus, $\hat{a} \in \{\tilde{a}, 2\tilde{a}, ...\}$). We define $a \equiv \hat{a}/\tilde{a}$ and denote the ranking in the productivity ladder by $a \in \mathcal{N}^+$. Moreover, P denotes the productivity distribution, $P_1, P_2, ...$ denotes the proportion of firms at each rung of the ladder, and $F_a = \sum_{j=1}^{a} P_j$ 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.[4] We abstract from entry and exit.

[3] 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: $1-\tau_i \equiv (1+\tau_{Ki})^{-\alpha}(1+\tau_{Li})^{-(1-\alpha)}$.

[4] 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 possiblity.

Firms can increase their productivity through a binary choice between *innovation* and *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 can involve investments, the crux of the choice is the cost difference. Therefore, we normalize the cost of imitation to zero, and let the cost of innovation be nonnegative.[5]

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.

Innovation: An innovating firm can improve its productivity via two channels. First, it can discover something 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 $\overline{p} < 1$. 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 $p_i + (1 - p_i) \delta q (1 - F_a)$.[6]

The Trade-Off: 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

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

where P 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 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. Note that the ex-post productivity growth gap between innovating and imitating firms is increasing in the TFP level.[7]

The Law of Motion of Productivity: We can now write the law of motion of the distribution

[5] If we assumed both costs to be positive, some firms could decide to stay inactive, that is, neither innovate nor imitate. We ignore this possibility.

[6] We impose no restriction on 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.

[7] To see this, observe that for firms with current productivity a, the expected productivity gap in the next period between innovating and imitating firms is $\int_{Q(a,\tau;P)}^{\overline{p}} \{p_i + (1 - p_i) \delta q (1 - F_a)\} / (1 - G(Q(a, \tau; P))) \, dG(p) - q(1 - F_a)$. A larger a has two opposite effects on the expected productivity gap. On the one hand, a larger a lowers the probability to meet a firm with higher productivity, so F_a increases. This lowers the expected growth for imitating firms and contributes to a larger gap. On the other hand, a larger a lowers $Q(a, \tau; P)$, inducing firms with low p to start innovating. This negative selection (in terms of p) for innovating firms lowers their expected productivity gap. However, this is a second-order effect because the marginal firms are indifferent between innovation and imitation. A formal proof is available in the webpage appendix.

of log-productivity, $P_a(t)$. Define the indicator function

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

In plain words, χ^{im} is unity when the firm finds it optimal to imitate, while $\chi^{\text{in}}(a,p,\tau;P)$ 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:

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

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

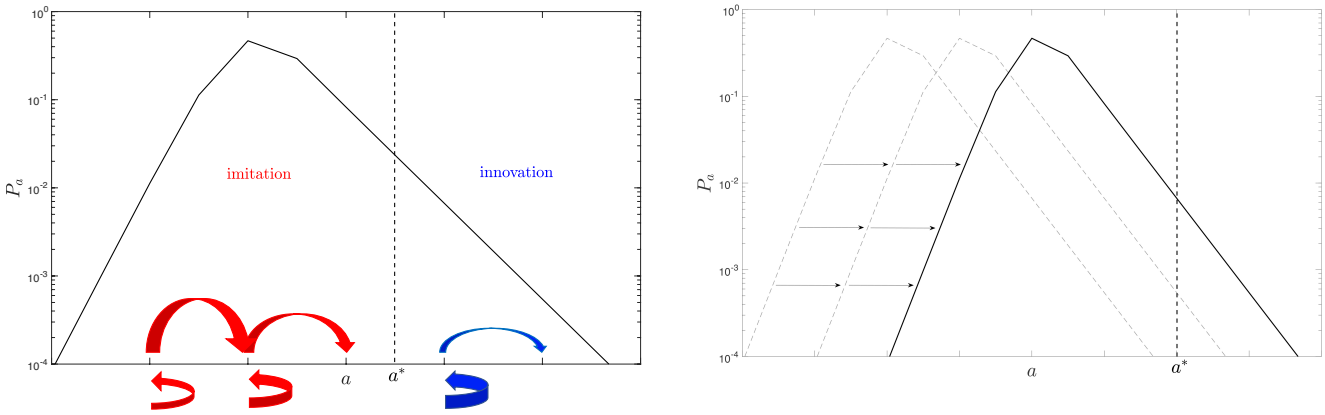
Proposition 1 *Consider the model of innovation-imitation described in the text whose equilibrium law of motion satisfies Equation (6), where each firm draws p from a distribution $G:[0,\overline{p}]\to[0,1]$. Let $\Delta t \to 0$ and assume that $q > \hat{p}$, where $\hat{p} \equiv \int_0^{\overline{p}} p\;dG(p)$. Assume the cost of both imitation and innovation is equal to zero. Then, there exists a traveling wave solution of the form $P_a(t) = f(a-\nu t)$ with velocity $\nu = \nu(q,\delta,g(p)) > 0$, with left and right Pareto tails. For a given t, P_a is characterized as follows: (i) for a sufficiently large, $P_a(t) = O\left(e^{-\rho(a-\nu t)}\right)$, where the exponent ρ is the solution to the transcendental equation $\rho\nu = \hat{p}\,(e^{\rho}-1)$; (ii) for a sufficiently small, $P_a(t) = O\left(e^{\lambda(a-\nu t)}\right)$, where the exponent λ is the solution to the transcendental equation $\lambda\nu = q(1-e^{-\lambda})$.*

Intuitively, a traveling wave is a productivity distribution that is stationary after removing the (endogenous) constant growth trend. More formally, $a^*(p,t+\nu t) = a^*(p,t)+\nu t$. 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, 2009). Although our model features no reflecting barrier strictly speaking, for each p and at a given t there is a threshold $a^*(p,t)$ such that all firms below $a^*(p,t)$ imitate. Among the imitators, less productive firms are more likely to meet a better firm and to make progress than are more productive firms. [8] Thus, the subdistribution of imitating firms catches up, which prevents the upper end of the distribution from diverging.

Interestingly, and different from earlier studies, our model also 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. Panel A displays the threshold and the force implying convergence.[9] Panel B illustrates the traveling wave.

Figure 1: Equilibrium Dynamics with a Stationary Distribution



Panel (a) Panel (b)

Note: Panel (a) displays the threshold a^* and the stationary distribution. Panel (b) plots a traveling wage.

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.[10]

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

[9] The solutions for ρ and λ are in an implicit form and involve transcendental equations. Standard methods allow one to show that the equations $\rho\nu = q(e^{\rho} - 1)$ 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.

[10] An analytical characterization of the growth rate is available for the case in which there is no heterogeneity in p. See the Webpage Appendix.

2.3 Productivity 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

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

where

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

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 R&D 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 revenue that we observe in the data. In the theory section, we set $\theta = 0$ for simplicity. In this case, the function Q is modified as follows:

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

where $E(\tau'|\tau)$ 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.

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.

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

$$\begin{aligned} &\frac{P_a(t+\Delta t) - P_a(t)}{\Delta t} \\ &= \sum_{j\in\{l,h\}} \mu_{\tau_j}(t) \times \int_0^{\overline{p}} \left[\begin{array}{c} \chi^{\text{in}}(a-1,p,\tau_j;P) \times (p+(1-p)\,\delta q\,(1-F_{a-1}(t)))\,P_{a-1}(t) + \\ +\chi^{\text{im}}(a-1,p,\tau_j;P) \times q\,(1-F_{a-1}(t))\,P_{a-1}(t) \\ -\chi^{\text{in}}(a,p,\tau_j;P) \times (p+(1-p)\,\delta q\,(1-F_a(t)))\,P_a(t) \\ -\chi^{\text{im}}(a,p,\tau_j;P) \times q(1-F_a(t))P_a(t) \end{array}\right] dG(p), \end{aligned} \tag{9}$$

where $\mu_{\tau_l}, \mu_{\tau_h}$ denote the proportion of low- and high-wedge firms, respectively.

The model is closed by the law of motion for μ_{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

$$\frac{\mu_{\tau_l}(t+\Delta t) - \mu_{\tau_l}(t)}{\Delta t} = (1-\rho_h)\,(1-\mu_{\tau_l}(t)) - (1-\rho_l)\,\mu_{\tau_l}(t),$$

where μ_{τ_l} converges in the long run to $\bar{\mu}_{\tau_l} \equiv (1-\rho_h)/((1-\rho_h)+(1-\rho_l))$.

The next proposition extends Proposition 1 to the case of costly R&D investments.[11]

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 $P_a(t) = f(a - \nu t)$ with velocity $\nu = \nu(q, \delta, G(p), \bar{c}, \tau_h, \tau_l, \bar{\mu}_{\tau_l}) > 0$, with left and right Pareto tails. Conditional on ν, the characterization of the tails is the same as in Proposition 1.*

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 non-R&D firms;
4. the gap in average TFP growth between R&D firms and non-R&D firms increases in TFP.

3 Data and descriptive evidence

We consider firm-level data for mainland China (henceforth, 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 in revenues in the industrial sector.[12] To estimate firm-level productivity growth, we focus on a balanced panel for all manufacturing firms in 2007–2012, focusing on 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.[13] 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.[14] 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. The difference is largely 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.

[11] The proof, which is an extension of the proof of Proposition 1, is available from the Webpage Appendix.

[12] The selection criteria changed to sales above 20 million yuan since 2010.

[13] 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.

Data on R&D are available for the periods 2001–2003 and 2005–2007. In an extension we also study the earlier sample of the balanced panel for 2001–2007 (using R&D data for 2001).

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

Table 1: Summary Statistics

	Number of Firms	Number of R&D Firms	Median Revenue (million USD)	Median Revenue (million USD)	Median R&D Intensity (%)	Median R&D Intensity (%)
	Balanced Panel of Chinese Firms					
2007	123368	18140	5.25	21.83	0.49	1.58
2012	123368	NaN	11.87	42.27	NaN	NaN
	Private Chinese Firms in the Balanced Panel					
2007	117983	15828	5.10	17.49	0.46	1.52
2012	117771	NaN	11.60	35.13	NaN	NaN
	Balanced Panel of Taiwanese Firms					
1999	11229	1487	1.66	7.82	1.14	2.80
2004	11229	1144	2.01	11.59	1.07	2.22

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

Our empirical analysis focuses on the extensive margin of the R&D decision, for three reasons. First, it is consistent with the 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. Third, the intensive margin is subject to a more severe measurement error, especially in China.[15]

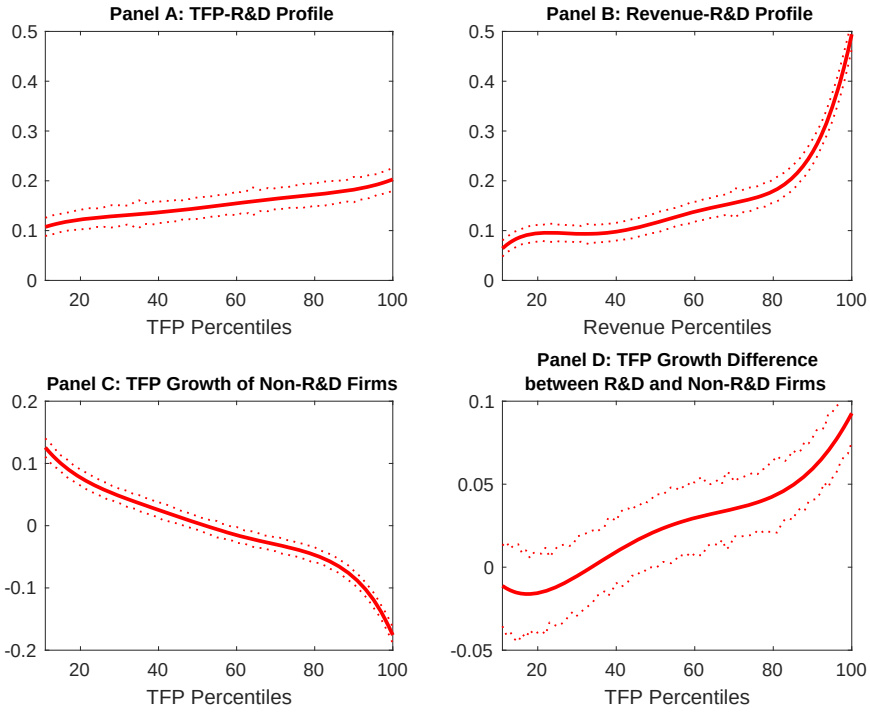
Figure 2 shows the distribution of R&D and TFP growth conditional on TFP in the initial year and conditional on firm size (where size is measured by value added). We estimate TFP following the methodology proposed by Hsieh and Klenow (2009), which is discussed in more detail in Section **??** To control for observable sources of firm 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. Firm-level value added is normalized by the median value in the industry to which each firm belongs.[16]

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 R&D is more attractive for more-productive firms. 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 may be a driver of both panels. However, the steeper profile in Panel B indicates that factors other than TFP must matter. According to our theory, firms with negative wedges–which we interpret as subsidies–are larger than is predicted by their productivity. The theory predicts that these subsidized firms are more likely to invest in R&D. Therefore, Panels A and B are consistent with the qualitative predictions of

[15]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. There are also policy incentives for firms to misreport the level of R&D expenditure (Chen et al., (2018)).

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

Figure 2: Chinese Firms in the Balanced Panel 2007–2012



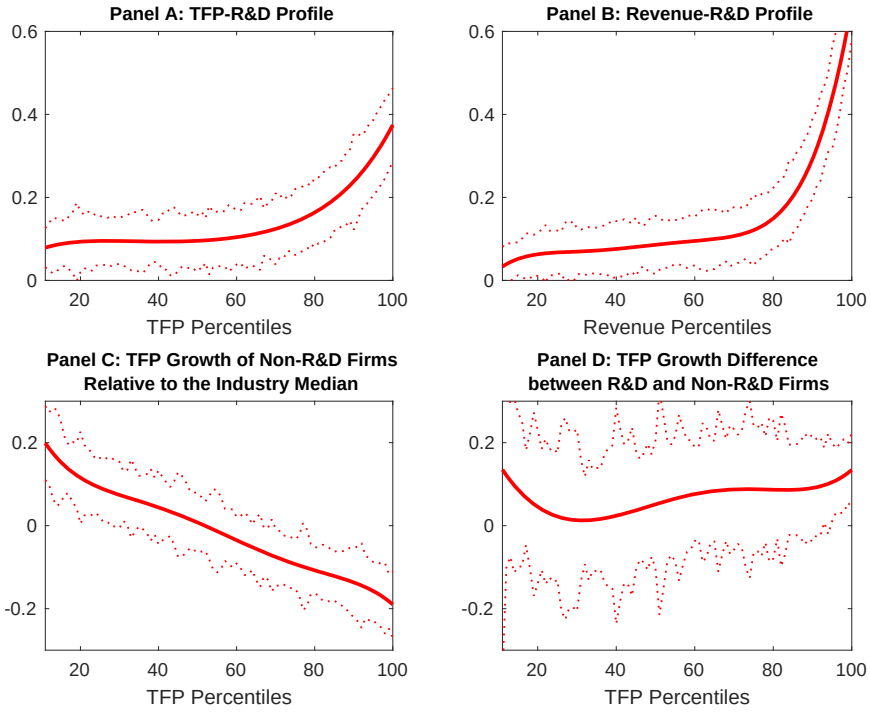
Note: The X-axis in Panels A, C, and D is the 2007 TFP percentile. The X-axis in Panel B is the 2007 value added percentile. The solid lines in Panel A and B plot the 2007 fraction of R&D firms in each TFP and value added percentile, respectively. The solid line in Panel C plots the median annualized 2007–2012 TFP growth among non-R&D firms in each TFP percentile. The solid line in Panel D plots the difference of the median 2007–2012 TFP growth between R&D and non-R&D firms. 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.

the theory.

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 non-R&D firms. In other words, there is strong convergence in productivity across non-R&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 especially 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 at $t+1$. Reversion to the mean would then exaggerate the convergence pattern. In the estimation section below, we model measurement error explicitly and allow it to be a codeterminant of the shape of the relationship in Panel C.

Finally, Panel D compares the TFP growth for R&D firms and non-R&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 non-R&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

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



Note: See Figure 2.

subsidies, as we discuss below.

The same patterns are confirmed by 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.

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

PANEL A: Correlations between firm characteristics and R&D decision in 2007

	(1) R&D$_d$	(2) R&D$_d$	(3) R&D$_d$	(4) R&D$_d$
log(TFP)	0.0623***	0.368***	0.343***	0.305***
	(0.00744)	(0.0284)	(0.0259)	(0.0232)
wedge		-0.410***	-0.378***	-0.332***
		(0.0357)	(0.0323)	(0.0296)
export$_d$			0.0528***	0.0544***
			(0.0134)	(0.0132)
SOE				0.205***
				(0.0232)
Industry effects	+	+	+	+
Age effects	+	+	+	+
Province effects	+	+	+	+
Observations	109,799	109,799	109,799	109,799
R-squared	0.143	0.208	0.211	0.224

PANEL B: Correlations between firm initial characteristics and TFP growth

	(1) TFP growth	(2) TFP growth	(3) TFP growth	(4) TFP growth	(5) TFP growth
log(TFP)	-0.0618***	-0.0617***	-0.0620***	-0.0615***	-0.0618***
	(0.00354)	(0.00355)	(0.00351)	(0.00361)	(0.00356)
R&D$_d$	0.0358***	0.0368***	0.0336***		
	(0.00421)	(0.00404)	(0.00348)		
export$_d$		-0.00578	-0.00647*	-0.00590	-0.00657*
		(0.00377)	(0.00354)	(0.00371)	(0.00350)
SOE			0.0288**		0.0286**
			(0.0115)		(0.0113)
R&D intensity$_h$				0.0441***	0.0412***
				(0.00599)	(0.00580)
R&D intensity$_m$				0.0419***	0.0382***
				(0.00690)	(0.00561)
R&D intensity$_l$				0.0254***	0.0225***
				(0.00351)	(0.00331)
Industry effects	+	+	+	+	+
Age effects	+	+	+	+	+
Province effects	+	+	+	+	+
Observations	109,799	109,799	109,799	109,799	109,799
R-squared	0.122	0.122	0.123	0.122	0.123

Note: In Panel A, 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. R&D intensity$_h$ is a dummy variable for high R&D intensity, which equals one if firm R&D expenditure over sales is in the 67th percentile and above (among all R&D firms). R&D intensity$_m$ is a dummy variable for medium R&D intensity, which equals one if firm R&D expenditure over sales is between the 33rd and 66th percentiles. R&D intensity$_l$ is a dummy variable for low R&D intensity, which equals one if firm R&D expenditure over sales is positive and below the 33rd percentile. The dependent variable in Panel B is tfp_g, the annualized firm TFP growth over the period of the balanced panel. We use the initial-period value for all the right-hand-side variables. Standard errors are reported in parenthesis. Observations are weighted by employment and standard errors are clustered by industry. We drop firms with TFP in the bottom 10 percentiles.

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 firm 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 wedge among the regressors.[17] 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 the regression analysis because we have information on R&D investments only for the year 2007. However, we have performed the same analysis on an earlier sample (2001–2007) for which R&D information is available in both the initial and final year. The results for the 2001–2007 panel are qualitatively similar to those reported in Table 2 and are robust to the inclusion of firm fixed effects: as firms become more productive over time they become more likely to perform R&D.

Panel B reports the coefficient of regressions whose left-hand side variable is the average TFP growth over the sample period 2007–2012. In these regressions, the number of observations is smaller, because we report the annualized five-year growth rate in order to mitigate the impact of noise and measurement error. 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) break the R&D dummy into three separate dummies, one per each tercile of R&D expenditure along the intensive margin. All three dummies are both statistically and economically significant. As expected, a higher (lower) investment in R&D is associated with a higher (lower) future TFP growth. The difference in growth rates between the upper and lower terciles is statistically significant. We have chosen to retain our focus on the extensive margin because we believe this mitigates the role of measurement error since it is arguably easier to measure R&D on the extensive margin than on the intensive margin. We have also checked the robustness of the results in Table 2 to different classifications of R&D firms, such as considering R&D firms as only those firms whose R&D spending amounts at least 0.5% of their revenue. The results are robust.

The descriptive results of this section are robust. A potential concern is that they are 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 also hold true when partitioning the sample into subgroups: exporting versus nonexporting firms, SOEs versus non-SOEs, and sorting firm by regions. The results are reported in the appendix.

We find the same patterns in the Taiwanese data, see Figure 3 and Appendix Table A1, with two noteworthy quantitative differences. First, the R&D-TFP profile in Panel A is steeper in Taiwan than in China. The percentage of Taiwanese R&D firms increases from 10% to over 35% as we move from the 60th to the 99th percentile of TFP. Second, Panel D has very large standard errors. However, the regression analysis in Table A1 establishes that there is a robust and highly significant positive

[17] The measurement of output wedges follows Hsieh and Klenow (2009) and 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.

correlation between R&D and future productivity growth similar to China.

4 Estimation

We calibrate three parameters–α, η, and θ–and estimate the rest from the structural model. The parameters α and η are standard production function parameters. We allow α to vary across industries and set α_j in (two-digit) industry j equal to the measured labor income share in the respective industries. Following Hsieh and Song (2015), we set $\eta = 5$. Finally, θ is identified from the variation of R&D expenditure across R&D firms with different TFP and revenue in the data. We return to this below.

4.1 Estimation procedure

4.1.1 Targeted moments and measurement error

We estimate the model by targeting 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. In particular, we consider four intervals of the distribution of TFP and size in each of the four panels: the 11th to 50th percentile, the 51st to 80th percentile, the 81st to 95th percentile, and the 96th percentile and above.[18] This choice yields sixteen target moments from the data. Figure A1 in the appendix plots the confidence intervals around these moments.

Measurement Error: The second set of targeted moments focuses on measurement error (m.e.). This is a common concern in models of misallocation *à la* Hsieh and Klenow (2009), where m.e. potentially affects both the measured moments of TFP and the imputed wedges. In our model, m.e. affects the target moments in Figure A1. 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. In the rest of the section, we propose a model of m.e. and discuss its estimation.

We assume that the observables–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 ν_y and ν_I. The notation with hats denotes observed variables, while no hat denotes the true variable.

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 2007–2012 period. Under this assumption, the time series within-period variation 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., $(1-\tau_i)\left(P_{it}Y_{it}\right) \propto (r_t)^{\alpha}\left(w_t\right)^{1-\alpha}\left(K_{it}\right)^{\alpha}\left(L_{it}\right)^{1-\alpha}$. Then, a constant τ_i implies that

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

[18] We focus 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.

where the notation Δ denotes $\Delta \ln X_t \equiv \ln X_t - \ln X_{t-1}$. The variance of true revenue growth can then be identified as follows:

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

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. The cross-sectional covariance is therefore unaffected by the aggregate input price growth. The variance of m.e. in revenue and inputs can then be written as

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

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

The empirical covariances in Equations (11) and (12) can be inferred using data on growth rates in revenue and inputs between 2007 and 2012.

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,

$$\hat{v}_{\mu a} = \left(\frac{\eta}{\eta-1}\right)^2 \hat{v}_{\mu y} + \hat{v}_{\mu I}.$$

We set $\eta = 5$ and use $\hat{\nu}_{\mu a}$ as a target moment in the joint estimation of the parameters of the model. Ideally, after estimating the variances $\hat{\nu}_{\mu a}$, $\hat{\nu}_{\mu y}$, and $\hat{\nu}_{\mu I}$, we could use two of them (the third being a combination of the other two) as target moments. However, handling two sources of m.e. is computationally infeasible within our structural estimation procedure.[19] Because the m.e. in inputs $\hat{\nu}_{\mu I}$ implied by Equation (12) is quantitatively small (see Appendix Table A2), the natural choice is to abstract from m.e. in inputs and use either $\hat{\nu}_{\mu a}$ or $\hat{\nu}_{\mu y}$ as our target. We focus on results based on targeting $\hat{\nu}_{\mu a}$ both because TFP is the main focus of our analysis and because this is the more conservative choice.[20]

The 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. More formally, we assume the following relationship:

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

[19] In the appendix, we show that with a single source of m.e., it is possible to analytically characterize how m.e. influences the moments of the model. Such an analytical mapping speeds up computations significantly. When there are multiple sources of m.e., the mapping can still be established by way of simulations. However, this process is time-consuming and leads to a computational curse at the structural estimation stage.

[20] The results are not sensitive to targeting $\hat{\nu}_{\mu a}$ or $\hat{\nu}_{\mu y}$. As a robustness exercise we estimate the benchmark model targeting $\hat{\nu}_{\mu a}$. The results are quantitatively very similar, although the theoretical fit of the targeted moments is slightly better when targeting $\hat{\nu}_{\mu y}$.

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,

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

Given the unbiased estimate of b, 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, $var\left(\hat{a}_{it}\right) = var(a_{it}) + \nu_{\mu a}$. The resulting unbiased estimates are $b = 0.779$ and $var\left(\varepsilon_{it}^{\tau}\right) = 0.042$ (compared to biased estimates of 0.802 and 0.047, respectively).[21]

Calibration of θ: Consider Equation (7). Recall that the parameter θ captures the elasticity of a firm's R&D cost to its TFP. We calibrate this parameter by targeting the ratio of R&D cost to revenue for firms in the top versus the bottom of the TFP distribution. Formally, we target the ratio $E\left(\psi_j|a_j\right)/E\left(\psi_i|a_i\right)$, where ψ denotes the ratio of R&D costs to revenue and a_i and a_j denote TFP in the i'th and j'th percentile. With the R&D cost in Equation (7), this ratio can be expressed analytically as

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

We use data on actual R&D costs and consider the R&D-to-sales ratio for firms in the top four percent of the TFP distribution versus firms in the 11th to 50th percentile. The parameter b is adjusted for m.e. in line with Equation (14). Equation (15) then implies an elasticity of $\theta \approx 0.25$.

Probability Distribution $G(p)$: 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 as discussed below.

4.1.2 Simulated method of moments

The model is estimated using SMM. Our estimation procedure searches for the parameters that minimize the distance between the targeted moments and the stationary distribution of the model. The fact that our model is analytically tractable is key for the feasibility of our estimation procedure. In particular, our SMM approach requires simulating the model under some parameter configuration, adding m.e. to the moments, and calculating the distance from the targeted empirical moments. One could in principle simulate the distribution of a large number of firms for every trial of a parameter configuration. However, this would be computationally very demanding. Our system of difference equations allows us to attain the stationary distribution very quickly. The appendix describes our analytical methodology for adjusting the stationary distribution for m.e. This approach is also very efficient in terms of computation time.

The numerical simulations have always converged to a unique stationary distribution irrespective of initial conditions, provided that the learning parameter q is sufficiently large. When this parameter is sufficiently low, there is no ergodic distribution.

The sample is randomly generated by bootstrapping for K times, where K is set to 500. Denote by $g_{m,k}$ the mth moment in the kth sample and by $g_m(\phi)$ the vector of the corresponding moments in

[21] Note that the distribution of τ_i is by construction consistent with Equation (13). 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 also positively autocorrelated.

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, e.g., Altonji and Segal (1996)).

5 Results

We first estimate the benchmark model and then consider three extensions.

5.1 Parsimonious model

Our benchmark model–which we label the *Parsimonious* model–estimates five parameters; q, $\bar{c}$, $\bar{p}$, δ, and $v_{\mu a}$. In this model, all firms have the same R&D cost parameter $\bar{c}$, implying that the R&D cost in terms of goods for firm i is given by $c_i(t) = \bar{c} \cdot \left(A_i(t)^{\theta} \bar{A}(t)^{1-\theta}\right)^{\eta-1}$, see Equation (7). Then, we consider more flexible specifications where we allow heterogeneity in the parameter $\bar{c}$. We focus on mainland China and study Taiwan as an extension.

The estimation results are displayed in column (1) of Table 3. The estimated coefficient $q = 0.164$ implies a significant convergence in TFP across imitating firms. We also find that δ is significantly smaller than unity, suggesting that firms focusing on innovation forgo potential learning through imitation. Investing in R&D reduces the probability of learning from a better firm by 83%. Firms trade off the R&D cost and the opportunity cost of learning through random interaction with the probability p of success through innovation. Recall here that although p is stochastic, a firm observes its realization before deciding whether to imitate or innovate. The estimated average probability of p is about $\bar{p}/2 = 0.071/2 \approx 3.6\%$. Given the costs and benefits of the two strategies, 11% of the firms invest in R&D in the estimated model which compares with 15.0% in the data. The estimated standard deviation of m.e. in TFP is $\sigma_{\mu a} = 0.486$. This implies that m.e. accounts for 23% of the variance of log TFP and 73% of the variance of TFP growth.[22]

Figure 4 shows that the Parsimonious model fits the data reasonably well. The model matches accurately the convergence pattern in Panel C and the differential growth between R&D and non-R&D in Panel D.[23] However, the model predicts steeper profiles for R&D-TFP (Panel A) and revenue-TFP (Panel B) than that of their empirical counterparts– a feature to which we return below. Since the number of moments exceeds the number of estimated parameters, the model can be tested for overidentification. The J-test value is 1.632, which establishes that the model is comfortably within the critical range at standard significance levels.[24]

[22] Note that the estimated variance of m.e. is somewhat lower than if we had calibrated it to match equations (11)-(12), which would have yielded a variance of 0.32.

[23] Measurement error has a strong impact on the estimates. Ignoring m.e. would increase the estimates for q and δ to $q = 0.614$ and $\delta = 0.622$, implying a substantially faster convergence. The reason is that in the absence of m.e., Panel C–TFP growth conditional on TFP for non-R&D firms–dictates a fast catching up rate for low-TFP firms, as discussed above.

[24] With five parameters and 17 moments (16 model moments plus one m.e. moment), we have 11 degrees of freedom. This yields a chi-squared critical value of 17.28 at a 10% level. Thus, the model cannot be statistically rejected.

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

	(1)	(2)	(3)	(4)
	Parsimonious	Flexible	Industrial Policy	Fake R&D
q	0.164	0.195	0.240	0.367
	(0.076)	(0.040)	(0.035)	(0.042)
δ	0.165	0.016	0.050	0.078
	(0.109)	(0.021)	(0.046)	(0.058)
$\bar{p}$	0.071	0.077	0.083	0.152
	(0.006)	(0.009)	(0.007)	(0.010)
$\bar{c}$	3.065	4.916	5.085	14.475
	(0.263)	(0.674)	(0.437)	(1.168)
$\sigma_{\mu a}$	0.486	0.421	0.397	0.373
	(0.037)	(0.013)	(0.014)	(0.030)
σ_c		1.508	1.622	1.164
		(0.093)	(0.081)	(0.291)
c_a			1.251	-3.743
			(0.159)	(1.539)
Fake R&D share μ				0.097
				(0.009)
J-Test	1.632	0.833	0.602	0.587

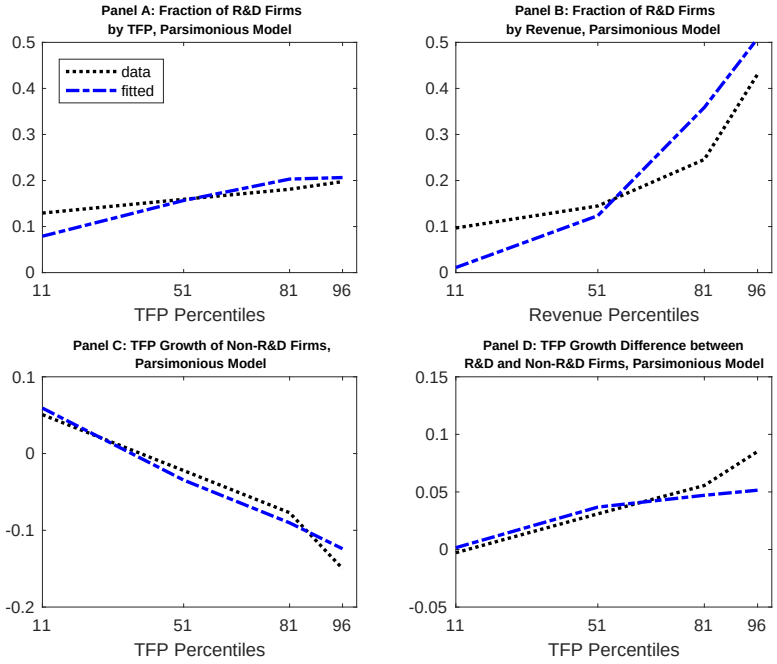
Note: Estimates of the four models. Bootstrapped standard errors in parentheses.

5.2 Heterogenous R&D costs

In this section, we consider more flexible versions of the model where firms face heterogenous R&D costs. We interpret this heterogeneity as either purely technological or arising from distortions. Similar to output wedges, firm-specific R&D wedges capture distortions and frictions that distort R&D decisions. This broad notion of wedges goes beyond formal R&D subsidies. For instance, it can be affected by government investments in infrastructure, science parks, etc. However, we will refer to the wedges as subsidies.

Given the results above, it is clear that heterogenous R&D can improve the fit of the model, especially, Panels A and B in Figure 4. Intuitively, heterogeneous R&D costs flatten the predicted profiles by making TFP and size less important for the firm's decision, which now also depends on the realization of the R&D cost.

Figure 4: China: Parsimonious Model



Note: Targeted moments for the Parsimonious model plotted against the corresponding empirical moments for the China 2007–2012 balanced panel. X-axis in Panel A, C and D is the first-period TFP percentiles defined on four intervals: the 11th to 50th, the 51st to 80th, the 81st to 95th and the 96th to 100th. 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. The solid line in Panel C plots the median annualized TFP growth among non-R&D firms in each TFP interval. The solid line in Panel D plots the difference of the median TFP growth between R&D and non-R&D firms. 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 fifth-order polynomial. The dotted lines plot the 95-percent confidence intervals by bootstrapping.

We consider three extensions to the benchmark model. In the *Flexible* model, we assume that subsidies are randomly distributed across firms. Then, we explore a model in which the subsidies are correlated with observable firm characteristics, capturing the idea that local and central governments may target their support to firms with certain characteristics. We refer to this specification as the *Industrial Policy* model. We also study a version of the Industrial Policy model where some firms can strategically misreport R&D expenditures in order to attract subsidies without actually distorting their investment decisions. We label this the *Fake R&D* model.

Flexible model: In this model, R&D subsidies are random and independent of firm characteristics. More formally, the effective R&D cost is given by

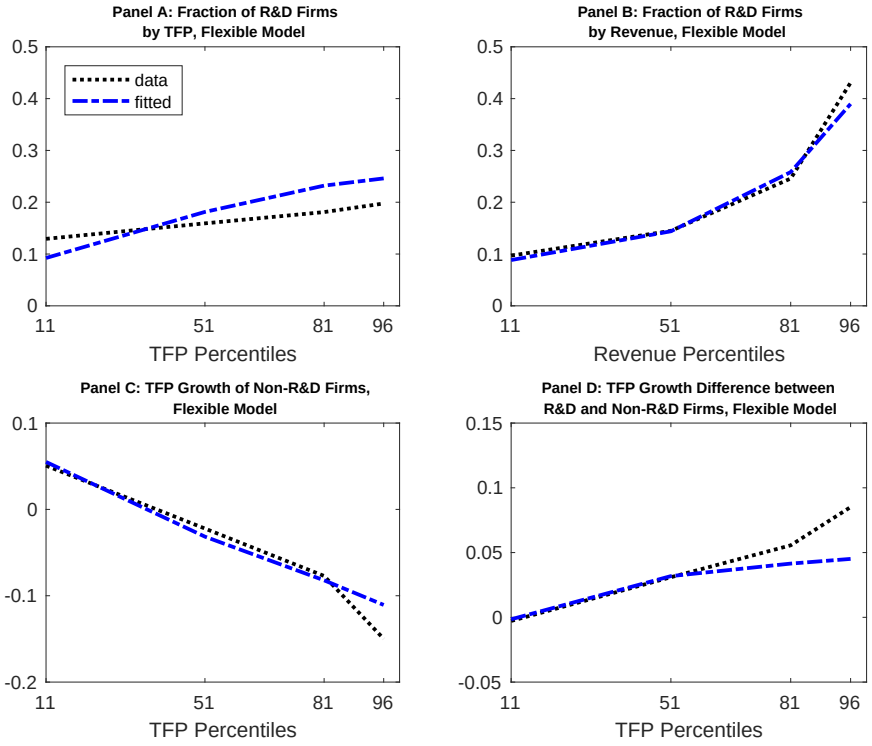
$$c_i(t) = \left[\bar{c} - \exp\left(\xi_i(t) - \frac{\sigma_c}{2}\right) + 1\right] \cdot \left(A_i(t)^{\theta} \bar{A}(t)^{1-\theta}\right)^{\eta-1},$$

where $\xi_i \sim \mathcal{N}\left(0, \sigma_c^2\right)$.[25] Note that $E\left[\bar{c} - \exp\left(\xi_i(t) - \sigma_c/2\right) + 1\right] = \bar{c}$, so σ_c is a mean-preserving spread.

[25]The derivation of Equation (15) is for the Parsimonious model without wedges. Since R&D wedges are often of a non-pecuniary nature, we assume that they do not show up in the R&D expenditure measured in the data. Under this assumption, the identification of θ remains valid in all our specifications of R&D costs and wedges.

Figure 5 shows that the Flexible model fits the data better than the Parsimonious model.[26] Note that the estimates of the structural parameter are quantitatively similar to those of the Parsimonious model, with two exceptions. First, the average (potential) R&D cost $\bar{c}$ is estimated to be larger than the cost $\bar{c}$ in the Parsimonious model. The reason is selection: firms with large subsidies have so low (potentially negative) costs that they will engage in R&D even if the probability of success p is low. In order to match the average share of R&D firms, it is then necessary for the average cost to be higher than in the Parsimonious model. Second, the parameter δ, capturing the "second chance" for R&D firms to imitate, now has a lower estimate because in the Flexible model more low-TFP firms engage in R&D. If δ were larger, these firms would grow too fast relative to the data, that is, the model would fail to match the pattern in Panel D.

Figure 5: China: Flexible Model



Note: See Figure 4.

Industrial Policy model: Next, we allow the R&D subsidies to be correlated with firm-specific TFP. The rationale for this model is to capture a notion of industrial policy, namely that the government might want to induce firms with particular characteristics in particular locations to engage in R&D. We do not impose any sign on the correlation in the estimation. We assume the effective R&D cost c_i to be of the form

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

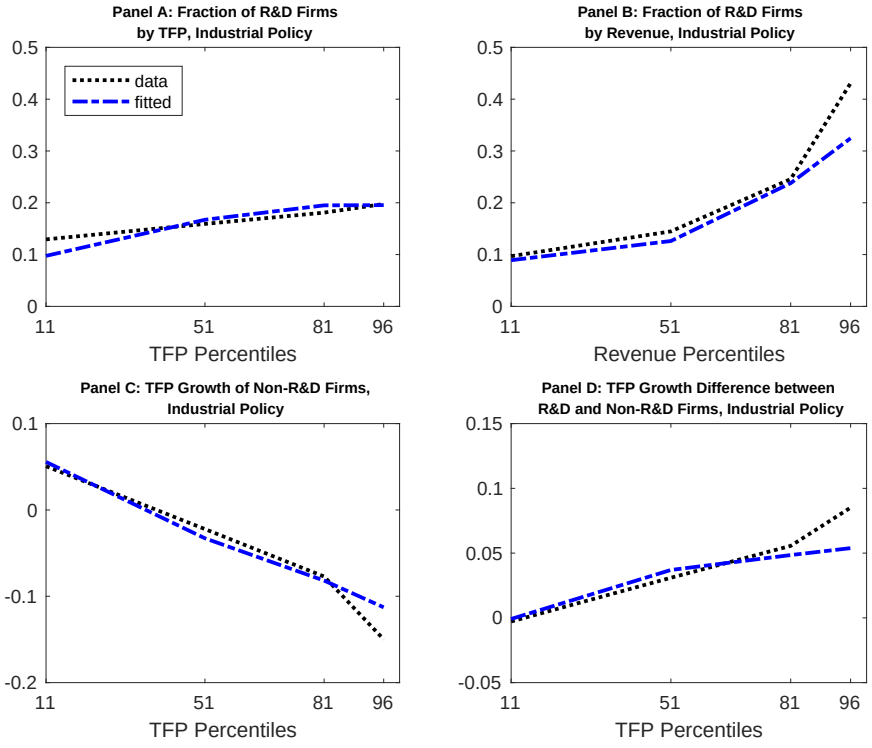
[26] Relative to the Parsimonious model, the Flexible model lowers the J-statistic by 50%, from 1.632 to 0.833.

where G is the cumulative density of a_i. A positive c_a means the industrial policy favors low-TFP firms, while a negative c_a means the policy favors high-TFP firms. In particular, the expected value (across realizations of ξ) of the cost term in brackets is equal to $\bar{c}$ for firms with the lowest a (for which $G(a) \to 0$) and $\bar{c} + c_a$ for firms with the highest a (for which $G(a) \to 1$). The Industrial Policy model nests the Flexible model as a particular case when $c_a = 0$.

Column (3) in Table 3 reports the estimation result for the Industrial Policy model. The fit of the model improves slightly relative to the Parsimonious and Flexible models of columns (1) and (2)–the J-statistic is now just 37% of the counterpart in the Parsimonious model. Figure 6 shows the fit of the targeted moments for the Industrial Policy model. Visually, the Industrial Policy model now fits more accurately both the TFP-R&D relationship (Panel A) and the value added-R&D relationship (Panel B).

The new parameter c_a is estimated to be positive. This indicates a negative correlation between TFP and R&D wedges: firms with higher TFP tend to recieve less subsidies. The reason for this result can be appreciated by comparing the results in Panel A and D of Figures 4 and 5. In the Flexible model, the share of firms engaging in R&D increases too steeply with TFP. Allowing R&D subsidies to be targeted to firms with low TFP reduces the predicted correlation between R&D investment and TFP and flattens the profile in Panel A. Moreover, the R&D decision for firms with high TFP becomes more dependent on a high relaization of p, which increases the TFP growth for R&D firms with high TFP, thereby improving the fit in Panel D.

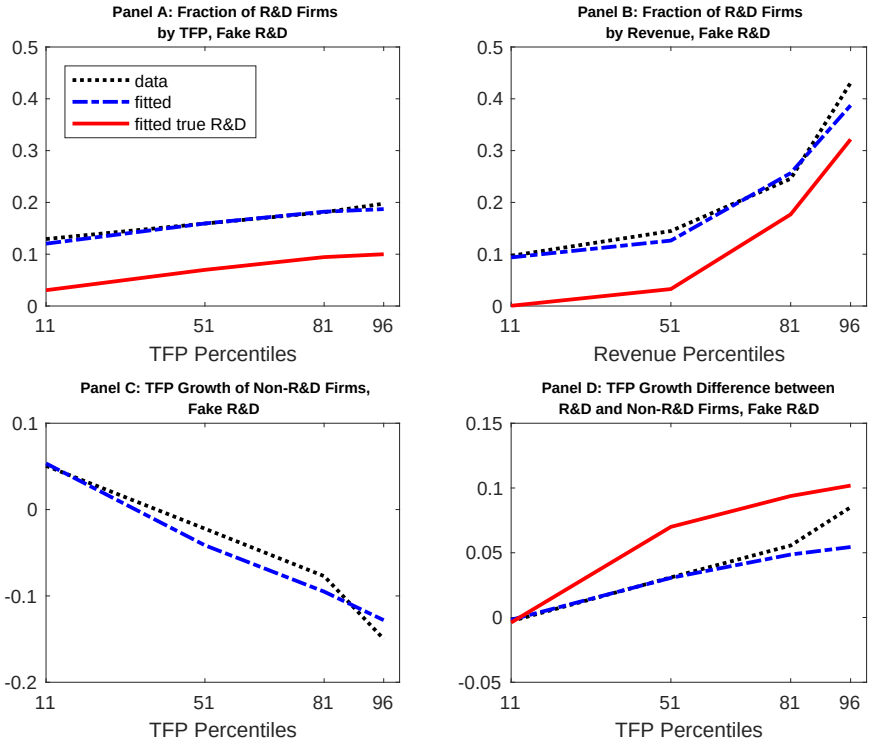
Figure 6: China: Industrial Policy Model



Note: See Figure 4.

Fake R&D model: Recent literature (Chen et al. 2018) suggests that many Chinese firms respond to subsidies by reclassifying operational expenditure as R&D expenditure. To explore this hypothesis,

Figure 7: China: Fake R&D Model



Note: See Figure 4.

we augment our theory with a simple model of moral hazard. We extend the Industrial Policy model by assuming 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 will be classified as R&D firms in the data while actually being non-R&D firms. More precisely, we assume that a share μ of firms (i.i.d. over time) is in the privileged position of being able to 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 effect of R&D toward zero, because fudging firms are classified as R&D firms in the data. The share μ of privileged firms is the only additional parameter in the structural estimation of the model.

Column (4) in Table 3 reports the results. According to our estimate, a share $\mu = 10\%$ of firms has the opportunity to fake R&D. The fit of the model improves only marginally relative to the Industrial Policy model, as illustrated in Figure 7 and from the fact that the J-statistic is almost unaffected. Each panel 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. The proportion of genuine R&D firms at each percentile can be inferred from the model given the estimate of μ. Panel D is especially revealing. For high-TFP firms, the productivity growth difference between R&D and non-R&D firms is as high as 10.2%. This is consistent with the casual observation that China has a small number of firms that are highly innovative and internationally successful (e.g., Huawei, Tencent). The model now almost perfectly fits the data, including the nonlinearities in Panel B.

The parameters $\bar{p}$ and c_a change markedly relative to the Industrial Policy model. Naturally, the productivity of R&D parametrized by $\bar{p}$ is now significantly larger because it excludes firms that fudge R&D: relative to the Industrial Policy model, the average probability of success more than doubles, from 0.042 to 0.076. Therefore, in this model, high-TFP firms have a stronger comparative advantage in pursuing innovation. Conversely, many low-TFP firms that report spending on R&D are fudgers.[27]

Comparing models: The estimated models are nested in the sense that the Parsimonious model is a special case of the Flexible model (when $\sigma_c = 0$), the Flexible model is a special case of the Industrial Policy model (when $c_a = 0$), and the Fake R&D model is a special case of Industrial Policy model (when $\mu = 0$). This allows us to provide a formal statistical F-test of whether the Parsimonious model can be rejected in favor of the extensions involving R&D wedges.[28]

Given the values for the J-statistics, the F-statistic for the (restricted) Parsimonious model against the heterogeneous wedges models are 10.6, 8.6, and 5.3, for the Flexible model, the Industrial Policy model, and the Fake R&D model, respectively. The corresponding critical values (at a 1% level) are 9.5, 7.6, and 7.0, respectively. However, if we test the Industrial Policy model and the Fake R&D model against the Flexible model, the Flexible model cannot be rejected in favor of the more general models.

In conclusion, the Parsimonious model–which is derived directly from the stylized theory–appears to be already quite successful in fitting the data, although some of its partial failures can be ameliorated by more flexible specifications. The Parsimonious model can statistically be rejected in favor of two of the Flexible model and the Industrial Policy model although it cannot be rejected in favor of the Fake R&D model. We show below that the models with heterogeneous cost improve upon the ability of the Parsimonious model to match nontargeted moments.

5.3 Nontargeted moments

In this section, we discuss the quantitative predictions of the model for nontargeted moments of the data.

5.3.1 Indirect inference

The moments used to estimate the model involve the TFP-R&D profile, the size-R&D profile, and the TFP-growth profile (Panels A, B, and C of Figure 2). The estimation targets pairwise correlations. However, we do not target explicitly the conditional relationships, i.e., the joint correlation structure between these variables. We now use indirect inference methods to investigate whether the model is consistent with these conditional correlations.

We first run multiple linear regressions whose dependent variable is the discrete choice of engaging in R&D and the right hand-side variables are TFP levels and the wedges . Panel A of Table 4 shows the results. Column (1) restates the empirical relationship in Panel A of Table 2 that the probability the firm does R&D is increasing both in TFP and in output subsidies (i.e., a negative coefficient on the output wedge τ). When simulating the model and running the same regression on the simulated data (including m.e.), we find the same qualitative conditional relationships for all models. In the Parsimonious model, the R&D decision has elasticities to TFP and wedges that are somewhat too

[27] The coefficient c_a becomes negative in the Fake R&D model, implying that TFP and subsidies are positively correlated because in the Fake R&D model the observed R&D profile would be essentially flat if $c_a = 0$. The observed R&D-TFP correlation would then be close to zero. To be consistent with the upward-sloping empirical profile in Panel A, the coefficient c_a must then be negative.

[28] The F-test is asymptotically valid because the J-statistics are asymptotically Chi distributed.

large in absolute value, while the heterogeneous cost models have elasticities that are much closer to the elasticities in the data and spot on for the Fake R&D model.[29]

In Panel B of Table 4 we run linear regressions with the firm-specific TFP growth as the dependent variable and both the initial TFP level and R&D decision (in the initial year) as explanatory variables. Again, 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. When simulating the model and running the same regression on the simulated data, we find that all models have the same qualitative conditional relationships as in the data. However, all models imply a slightly too large (in absolute value) elasticity of initial TFP. In terms of the elasticity of TFP growth to the R&D decision, the Parsimonious and Industrial Policy models have elasticities that are very close to the empirical values.

Based on the evidence from indirect inference documented here, we conclude that our model fits well the joint correlation structure between TFP growth, the R&D decision, initial TFP level, and size (wedge).

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

PANEL A: : Correlations between firm characteristics and R&D decision in 2007

	(1) Empirical $R\&D_d$	(2) Parsimonious Model	(3) Flexible Model	(4) Industrial Policy	(5) Fake R&D
log(TFP)	0.368***	0.7392***	0.3189***	0.2777***	0.3634***
	(0.0284)	(0.0148)	(0.0178)	(0.0175)	(0.0169)
wedge	-0.410***	-0.8647***	-0.3073***	-0.2842***	-0.4159***
	(0.0357)	(0.0180)	(0.0217)	(0.0212)	(0.0203)
Industry effects	+				
Age effects	+				
Province effects	+				

PANEL B: : Correlations between firm initial characteristics and TFP Growth

	(1) Empirical TFP growth	(2) Parsimonious Model	(3) Flexible Model	(4) Industrial Policy	(5) Fake R&D
log(TFP)	-0.0618***	-0.0827***	-0.0778***	-0.0839***	-0.0997***
	(0.00354)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
$R\&D_d$	0.0358***	0.0339***	0.0290***	0.0322***	0.0514***
	(0.0042)	(0.0005)	(0.0005)	(0.0005)	(0.0007)
Industry effects	+				
Age effects	+				
Province effects	+				

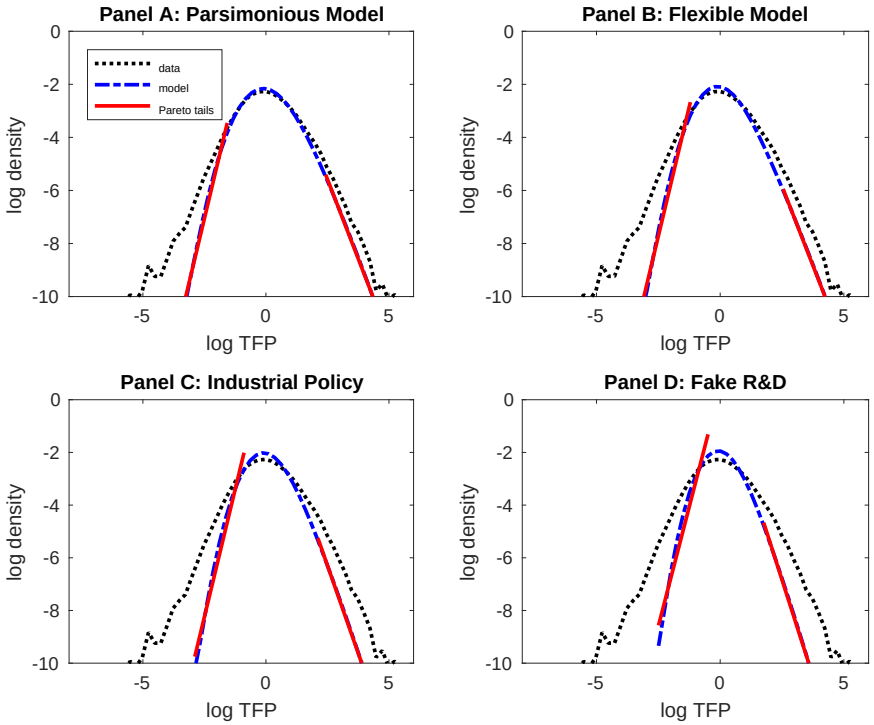
Note: The first column reports the regression results from the data. The second and third columns report the results from the simulated data. See Table 2 for variable definitions.

[29]The standard errors on the parameters estimated from the models 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.

5.3.2 Productivity distribution

The structural model features a *traveling wave*, namely, a stationary productivity distribution that shifts upward at an endogenous rate. The distribution is tent-shaped with two Pareto tails (see Proposition 2). We can simulate the model under the estimated parameters and compare the stationary TFP distribution and the growth rate predicted by the theory with the empirical distribution for China. Since we have not targeted any aspect of the TFP distribution in our estimation, this comparison provides evidence of the models' ability to match an important and untargeted empirical moment (recall that all moments we target are in terms of quantiles and abstract from the absolute dispersion).

Figure 8: TFP Distribution



Note: The dotted lines represent the log density of log TFP in the data. The dashed line is the steady-state log density of log TFP in the Parsimonious model (Panel A), Flexible model (Panel B), Industrial Policy model (Panel C) and Fake R&D model (Panel D), respectively. The solid lines are the estimated Pareto tails of the steady-state TFP distribution implied by each of the models.

Figure 8 displays the entire TFP distributions for each version of the model. The fit is overall good. Visually, the theoretical and empirical distributions are close one to another. Overall, the model better matches the upper tail than the lower tail of the empirical TFP distribution. The model cannot replicate the thick lower tail in the empirical distribution.[30] Consequently, the models generate a lower TFP dispersion than in the data. For example, the Parsimonious model accounts for 77% of the empirical variance of log TFP. The estimated left and right Pareto tail parameters in the Parsimonious model are 3.9 and -2.4. Fitting Pareto distributions in the data yields 1.3 and -1.7.[31]

[30] It is useful to recall that most existing theories in the literature on the endogenous evolution of productivity distribution feature no lower tail whatsoever.

[31] The corresponding numbers for the Industrial Policy model are 58%, 3.9, and -2.7.

The models have predictions about the speed of growth of the traveling wave. The Parsimonious model implies a steady-state annualized productivity growth rate of 3.75% (versus 4.1% in the Industrial Policy model). The empirical counterpart for this measure is the TFP growth rate for the median firm in our balanced panel, which is 3.75% for 2007–2012.[32]

5.3.3 Patents

Our theory predicts that firms that are larger and more productive size 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 2007–2012 balanced panel of matched firms, there are 14,492 firms (out of ca. 123,000) that were granted one or more invention patents for which they applied for during 2007–2012. The total number of invention patents applied for by these firms in that period is 146,896. This implies that NBS firms with patents in the balanced panel have on average 10.1 patents.[33] Note that, since the average time for granting a patent is about three years, a lot of patents sought during 2007–2012 have been granted after 2016. The magnitudes above are therefore a lower bound to the actual number of patents.

Consider Figure 9. Panel A 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 non-R&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 non-R&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.

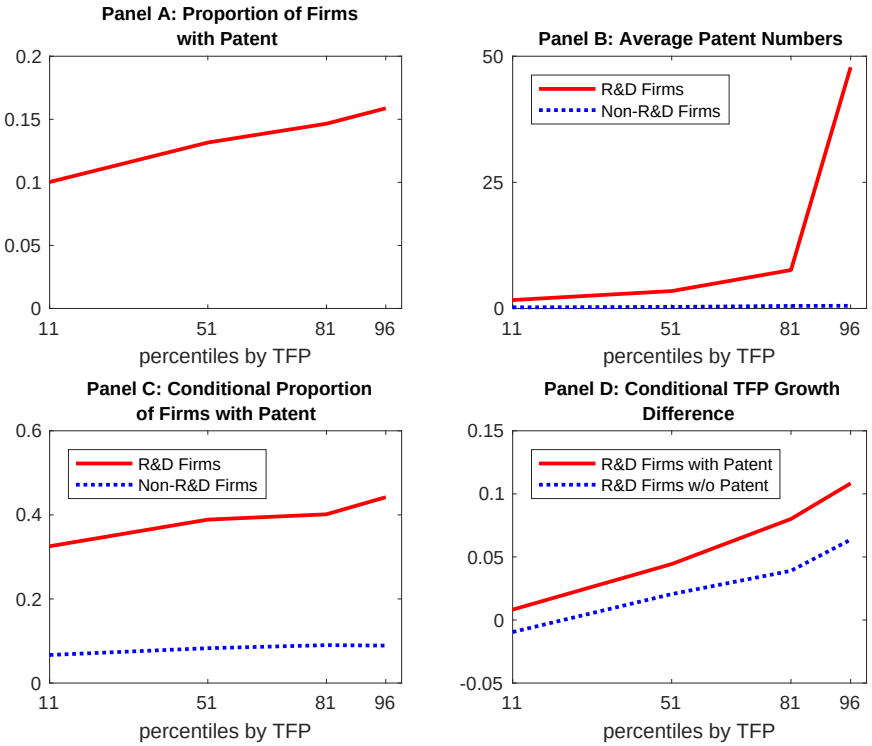
Panel D displays the most interesting finding. Firms with a positive number of patents experience larger TFP growth than R&D firms without a patent. The gap increases with TFP, being largest for the top two deciles of the TFP distribution. This is consistent with our model, where R&D firms that are successful 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, the 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 81st to the 96th TFP percentile (Panel B), precisely consistent with the prediction of the Fake R&D model that it is mostly low-TFP firms that have an incentive to opportunistically report R&D expenditure without actually engaging in it.

[32] The aggregate empirical TFP growth could stem from three different sources: growth in average TFP of surviving firms, changes in higher-order moments of the TFP distribution during transitional dynamics, and entry and exit. In steady state, our model incorporates only the first of these sources. This is why it is appropriate to compare the steady-state growth in the model to the TFP growth of the median firm in the empirical balanced panel of firms.

[33] In the unbalanced panel, 28,081 NBS firms (out of a total of 275 thousand) have one or more invention patents that were applied for during 2007–2012. The total number of invention patents applied for by these NBS firms during that period is 228,634. That is to say, NBS firms with patents applied for 8.1 patents on average in that period.

Figure 9: Patents of Chinese Firms, Balanced Panel 2007–2012



Note: All patents in the figure refer to invention patents applied during 2007–2012. We group all firms in the 2007–2012 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 non-R&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 non-R&D firms (solid line). The solid line in Panel D plots the TFP growth difference between R&D firms with patents and non-R&D firms. The dotted line plots the TFP growth difference between R&D firms without patents and non-R&D firms.

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 (2001–2007). The goal of this analysis is twofold. First, we want to assess the extent to which the estimates are robust across different samples. 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 that has commonalities with mainland China.

6.1 China 2001–2007

We first consider data for mainland China in the earlier period of 2001–2007. Table 5 shows the results. Similar to the results for the 2007–2012 period, our model also matches the earlier period very well: the J-statistic is roughly the same for both periods, although slightly better for the 2001–2007 period. Appendix Figure A2 displays the fit of the targeted moments for the Parsimonious model. Overall, the fit of the model is very good.

The estimated parameters are qualitatively similar to those for 2007–2012, although there are some quantitative differences. The estimates of m.e. are essentially identical for the two periods and across

all models. However, the estimated productivity of R&D, $\bar{p}$, is less than half the estimates for 2007–2012. The productivity of imitation (i.e., technological convergence) is also substantially lower than in the benchmark sample: the parameter q is around 40% of the estimate in Table 3. The estimates of the probability of a "second chance" to imitate (δ), however, is substantially larger for the earlier period for all models but the Flexible model. This indicates a low opportunity cost of R&D.[34]

Table 5: Estimation, Chinese Firm Balanced Panel 2001–2007.

	(1)	(2)	(3)	(4)
	Parsimonious	Flexible	Industrial Policy	Fake R&D
q	0.066	0.054	0.086	0.172
	(0.031)	(0.047)	(0.035)	(0.044)
δ	0.432	0.132	0.365	0.395
	(0.105)	(0.017)	(0.104)	(0.057)
$\bar{p}$	0.028	0.033	0.032	0.068
	(0.001)	(0.001)	(0.002)	(0.002)
$\bar{c}$	1.178	2.284	1.942	5.120
	(0.032)	(0.051)	(0.296)	(0.088)
$\sigma_{\mu a}$	0.476	0.472	0.446	0.408
	(0.026)	(0.016)	(0.017)	(0.009)
σ_c		0.993	0.969	0.126
		(0.076)	(0.134)	(0.073)
c_a			0.386	0.025
			(0.296)	(0.039)
Fake share μ				0.131
				(0.015)
J-Test	1.375	0.727	0.701	0.637

Note: Estimates of the four models for the Chinese 2001–2007 sample. Bootstrapped standard errors in parentheses.

In summary, the results for China in the period 2001–2007 indicate a less important role for R&D investments as a driver of productivity growth compared to the 2007–2012 period. The results are consistent with a technological interpretation: the productivity of R&D was intrinsically lower in

[34]Part of this may be due to the large number of R&D firms that make positive but very small investments in R&D. In regressions analogous to columns (4) and (5) in 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-sales. However, the effect is relatively small even for large firms.

the earlier period, and the difference between R&D firms and non-R&D firms was smaller. This interpretation is in line with the argument of Acemoglu, Aghion, and Zilibotti (2006), who argue that the importance of R&D investment increases as an economy approaches the world technology frontier.

6.2 Taiwan

Consider now the data for Taiwan. The results are summarized in Table 6.[35] Appendix Figure A3 displays the fit of the targeted moments for the Parsimonious model. The visual fit of the Flexible and the Industrial Policy models are almost identical. Moreover, the formal measure of fit – the J-statistic – is better than in the mainland China samples.

In terms of parameter estimates, the results are qualitatively identical to the results for China, and also highly robust across models. Quantitatively, however, there are some differences. First we estimate larger m.e. than in the Chinese data. Second, the productivity of R&D is larger in Taiwan: the estimated average probability of success in innovation ($\bar{p}/2$) is estimated to be around 7%, that is, more than twice as large as for the benchmark China sample. Third, 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 0.35 and 0.43. Finally, note that since the Parsimonious model already fits the data for Taiwan quite well, there is relatively little improvement to be gained from allowing dispersion in R&D wedges.

In conclusion, the model fits the Taiwanese data even better than for mainland China. The results for the two economies are broadly consistent. However, technology diffusion through imitation is somewhat faster for Taiwanese firms whose R&D investments are also more productive.

7 Counterfactuals

In this section, we run some counterfactual policy experiments based on the estimated model. We base our results on the Parsimonious model because the economic mechanism is most transparent in this model. The results are similar in the Industrial Policy model (details are available upon request).

Our primary counterfactual experiment is an exogenous reduction in misallocation. Hsieh and Klenow (2009) document large static efficiency gains from reducing misallocation in China and India. Our model also features dynamic effects of misallocation through the R&D investment channel and the estimated model allows us to evaluate these effects. Conditional on TFP, firms facing negative wedges (i.e., subsidies) are more likely to invest in R&D, while firms facing positive wedges (i.e., taxes) are less likely to do so. Thus, the dispersion in firm-specific wedges flattens the TFP-R&D profile, inducing more low-TFP firms and fewer high-TFP firms to invest in R&D. In other words, static misallocation makes R&D decisions depend more on firm size and less on TFP, in line with the steeper R&D profile in Panel B than in Panel A of Figure 4.

We study the dynamic effects of an unanticipated 50% reduction in the variance of the logarithm of output wedge, $\log(1-\tau_i)$.[36] Figure 10 displays the transition of the economy from the initial steady state to the new steady state with lower output wedge dispersion. The two upper panels show the evolution of the growth rates in the cross-sectional average TFP and in aggregate TFP (defined as in

[35] We have omitted the results for the Fake R&D model for Taiwan because the estimated share of R&D fudgers, μ, is close to zero. The Fake R&D model is therefore practically identical to the Industrial Policy model for Taiwan.

[36] Recall that wedges and TFP are related according to $\log(1-\tau_i) = b * log(A_i) + \varepsilon_i^{\tau}$, see eq. (13). Thus, $var(\log(1-\tau)) = b^2 \cdot var(\log(A)) + var(\varepsilon_i^{\tau})$. In the counterfactuals, we reduce both b^2 and $var(\varepsilon_i^{\tau})$ by 50% so that $var(\log(1-\tau))$ would be reduced by 50% if the TFP distribution were held constant.

Table 6: Estimation, Taiwanese Firm Balanced Panel 1999–2004

	(1)	(2)	(3)
	Parsimonious	Flexible	Industrial Policy
q	0.428	0.357	0.349
	(0.116)	(0.165)	(0.179)
δ	0.403	0.062	0.070
	(0.206)	(0.189)	(0.162)
$\bar{p}$	0.138	0.155	0.154
	(0.031)	(0.036)	(0.039)
$\bar{c}$	6.352	8.869	9.793
	(0.996)	(1.581)	(2.072)
$\sigma_{\mu a}$	0.615	0.591	0.595
	(0.052)	(0.094)	(0.127)
σ_c		1.808	1.842
		(0.195)	(0.219)
c_a			-0.907
			(1.042)
J-Test	1.081	0.67	0.659

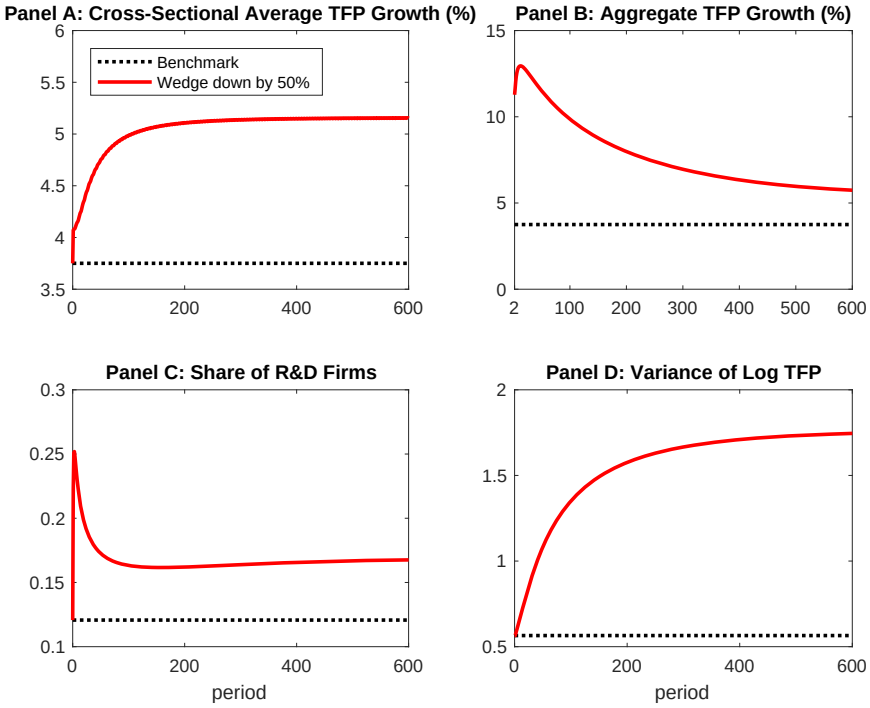
Note: Estimates of the four model for Taiwan. Bootstrapped standard errors in parentheses.

Hsieh and Klenow (2009)), respectively.[37] In balanced growth, the two statistics would be identical because the higher-order moments of the TFP distribution would be constant over time. Therefore, the only source of TFP growth in balanced growth would be that the cross-sectional average TFP increases. However, along the transition, the higher-order moments change and this can give rise to transitional growth. In our experiment of reduced output wedges, the cross-sectional variance of log TFP increases sharply (Panel D), due to the upper tail of the TFP distribution becoming heavier.

The contribution to aggregate TFP growth from an increase in the growth rate cross-sectional average TFP is small on impact–from 3.8% in the initial steady state to 4%–and increases gradually to 5.2% in the long run (Panel A). This positive contribution of the increase in the cross-sectional average TFP is dwarfed by the contribution of the higher-order moments. For example, for the first 110 periods after the reduction in wedges, aggregate TFP growth hovers above 5% per year. The fact that aggregate TFP growth exceeds the cross-sectional average TFP growth is driven by the increase in the cross-sectional dispersion in TFP (Panel D). Over time the growth rate in aggregate TFP falls.

[37] To ease the visualization, Panel B does not display the initial jump in TFP associated with an improvement in the static allocation. This static effect happens on impact and does not contribute to growth after the initial period.

Figure 10: TFP Distribution



Note: The graph displays the transition for the benchmark Parsimonious model from the initial steady state after a 50% reduction in the variance of wedges. Panel A shows the growth rate of cross-sectional average TFP. Panel B displays the growth rate in aggregate TFP. Panel C plots the share of firms engaging in R&D, and Panel D documents the cross-sectional variance of log TFP.

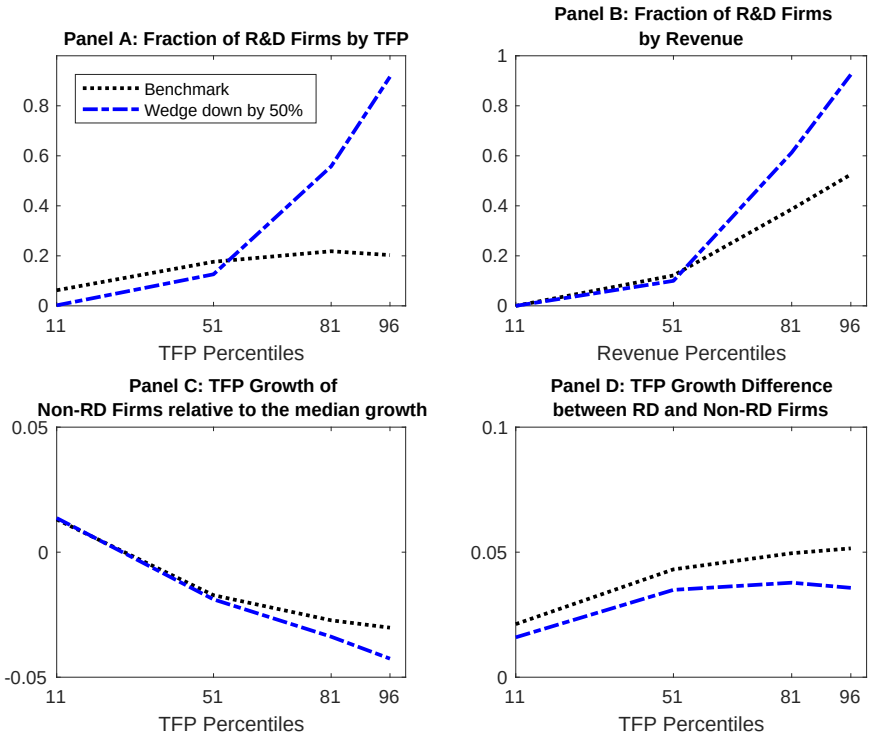
After about 500 periods the increase in the variance of $\log(TFP)$ stagnates, and the aggregate growth rate converges to the growth rate of the (cross-sectional) average TFP.

Panel C shows the evolution of the share of R&D firms. This share doubles upon impact because output wedges deter many high-TFP firms from investing in R&D despite their comparative advantage. This is only partially offset by some subsidized low-TFP firms investing in R&D. Note that firms with very low TFP do not invest in R&D even if they receive a larger output subsidy. The reason is that these firms have a higher expected productive growth when they imitate than if they try to innovate. High-TFP firms grow on average faster in the counterfactual than in the benchmark equilibrium because investing in R&D accords with their comparative advantage. This pulls these firms ahead in the distribution resulting in an increasing spread of the TFP distribution (Panel D). In turn, this growing spread makes imitation more attractive over time, inducing a declining share of R&D firms over time. In the new balanced growth, the share of R&D firms is higher (17%), and the productivity distribution has a fatter tail.

Figure 11 displays a steady state comparison for the set of target moments in the estimation.[38] Panel A shows that the TFP-R&D profile is much steeper in the counterfactual than in the benchmark economy, in line with the argument above that lower distortions should make the R&D more correlated with TFP. In this counterfactual economy, 92% of the firms in the top 4% of the TFP distribution (96th–

[38]Since this figure shows no data, we abstract from m.e. For this reason, the figure is different from Figure 2 where we added m.e. to the simulations in order to make them comparable with the data.

Figure 11: Steady-State Moments for Counterfactual Model



Note: The graph displays the moments in steady state for the counterfactual experiment where the variance of wedges are reduced by 50% relative to the benchmark Parsimonious model. See also the caption to Figure 4.

99th percentiles) and more than half the firms in the 81st–95th percentiles of the TFP distribution invest in R&D. In contrast, almost no firms below the median invests in R&D. Moreover, the Revenue-R&D profile (panel B) is now more similar to the TFP-R&D profile – recall that, in the limit case with no output wedges, the two panels would be identical because, in that case, the firm size would be perfectly correlated with TFP. Panel D shows that the TFP growth difference between R&D and non-R&D firms is smaller in the counterfactual, especially for high TFP levels. This is due to a selection effect: wedges deter firms from investing R&D, except for those that draw a very high p. This implies positive selection on p. Fewer distortions lower the positive selection, especially for high TFP levels. (Recall that most of those firms engage in R&D when distortions are lower.) Therefore, the gap in TFP growth is larger in the highly distorted benchmark economy than in the less distorted counterfactual.

The main conclusion of this experiment is that static misallocation has large quantitative effects on future productivity growth. Moreover, it compresses the dispersion in the stationary TFP distribution. In our model, we assume a competitive labor market, so inequality in TFP and firm size has no effect on wage inequality. However, in models with labor market frictions (e.g., search and ex-post rent sharing between workers and firms) and in models with assortative matching between workers and firms, reducing wedges and misallocation would lead to higher wage inequality.

Finally, we consider some additional counterfactuals. Table 7 summarizes the comparison across balanced growth equilibria. In columns (1) and (2), we change the structural parameters $\bar{p}$ and q, respectively, so that R&D technology in China has the same productivity as in Taiwan. Other parameters, including the distribution of output wedges, remain as estimated for China. In each case, we

observe a significant increase (90% and 63%, respectively) in the steady-state TFP growth rate. Interestingly, the long-run growth rate responds significantly not only to an increase in the productivity of R&D but also to an increase in the probability of a chance to imitate. In column (3), we force all firms to invest in R&D. The long-run growth rate is now much lower in the counterfactual economy–about one-third of the growth rate in the estimated benchmark. This result is driven by the empirical finding that focusing on innovation (R&D) entails a large opportunity cost in terms of forgone ability to learn from random interactions. Therefore, forcing all firms, including low-TFP ones, to focus on innovation yields a significant efficiency loss and lower growth. In columns (4) and (5), we change the baseline cost of innovation $\bar{c}$ so that, respectively, 5% and 20% of the Chinese firms invest in R&D. The growth rate responds as expected: a small increase in $\bar{c}$ reduces the TFP growth rate, while a small reduction in $\bar{c}$ increases it. Interestingly, increasing innovation through a R&D subsidy across the board yields the standard positive effect on growth as predicted by models of endogenous technical change (e.g., Romer (1990), Aghion and Howitt (1992)). Note however, that the effect of R&D subsidies is nonlinear and becomes negative for sufficiently large subsidies. This is illustrated by the (extreme) case of column (3), where all firms invest in R&D.

Table 7: Counterfactuals, Parsimonious Model

	(1)	(2)	(3)	(4)	(5)	(6)
	Output wedge down by 50%	Taiwan's $\bar{p}$	Taiwan's q	All firms do R&D	Increasing $\bar{c}$ so that share of R&D firms = 5%	Decreasing $\bar{c}$ so that share of R&D firms = 20%
Fraction of R&D Firms (%)	16.8 (12.1)	26.9 (12.1)	9.5 (12.1)	100 (12.1)	5.0 (12.1)	20.0 (12.1)
Steady State TFP Growth (%)	5.15 (3.75)	7.11 (3.75)	6.11 (3.75)	1.23 (3.75)	2.61 (3.75)	4.69 (3.75)

Note: Counterfactual experiments for the Parsimonious model. The statistics do not include measurement error.

8 Conclusion

In this paper, we have constructed a theory of firm dynamics and productivity growth where firms are heterogenous in productivity and are subject to a variety of distortions. Each firm can improve its productivity either through learning from random interactions or through investing in R&D to try to discover new technologies. Both strategies yield a stochastic payoff. However, both the productivity 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 using data for China and Taiwan. The estimated model provides a good quantitative fit of salient targeted and nontargeted moments of the data. The return to R&D investment is significantly higher in Taiwan than in mainland China. Part of the gap could be explained by Chinese firms fudging and overreporting R&D.

From a methodological perspective, our paper represents one the first attempts to estimate models

of random interactions and endogenous productivity distribution using firm-level data. We find it encouraging that the model well fits the data for two independent data sets and performs well in predicting nontargeted moments. Our structural approach lends itself well to perform policy-relevant counterfactuals.

References

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

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr (2018). "Innovation, Reallocation, and Growth," *American Economic Review* 108(11), 255-294.

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

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

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

Akcigit, Ufuk, and William R. Kerr (2018). "Growth through Heterogeneous Innovations," *Journal of Political Economy* 126(4), 1374-1443.

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

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

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

Asl, Farshid Maghami, and Galip Ulsoy (2003). "Analysis of a system of linear delay differential equations," *Journal of Dynamic Systems, Measurement, and Control* 125(2), 215–223.

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

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

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

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

Benhabib, Jess, Jesse Perla, and Christopher Tonetti (2017). "Reconciling models of diffusion and innovation: a theory of the productivity distribution and technology frontier," NBER Working Paper 23095.

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

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

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

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

Chen, Zhao, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi Xu (2018)."Notching R&D Investment with Corporate Income Tax Cuts in China," NBER Working Paper 24749.

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

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

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

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

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

Garcia-Macia, Daniel, Chang-Tai Hsieh, and Peter J. Klenow (2019). "How destructive is innovation? *Econometrica* 87(5) 1501-1541.

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

Griliches, Zvi (1957). "Hybrid corn: An exploration in the economics of technological change," *Econometrica* 25(4), 501-522.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Lichtenberg, Frank R. (1992). "R&D investment and international productivity differences," NBER Working Paper 4161.

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

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

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

McFadden, Daniel (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica* 57(5), 995-1026.

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

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

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

Tombe, Trevor, and Xiaodong Zhu (2019). "Trade, migration and productivity: A quantitative analysis of China," *American Economic Review* 109 (5), 1843-72.

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

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

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

Technical Appendix

Proof of Proposition 1.

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

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

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

Let $\Delta t \to 0$. The difference equation governing the evolution of the log-productivity distribution can then be broken down as follows:

$$\begin{aligned} \frac{\partial P_a(t)}{\partial t} &= \lim_{\Delta t \to 0} \frac{P_a(t+\Delta t) - P_a(t)}{\Delta t} \\ &= \begin{cases} q\left[(1-F_{a-1}(t))P_{a-1}(t) - (1-F_a(t))P_a(t)\right] & \text{if} \quad a < a^*(t), \\ \begin{pmatrix} q(1-F_{a-1}(t))P_{a-1}(t) \\ -G(Q(a))\left[q(1-F_a(t))P_a(t)\right] \\ -\int_{Q(a)}^{\overline{p}}\left[(p+(1-p)\delta q(1-F_a(t)))P_a(t)\right]dG(p) \end{pmatrix} & \text{if} \quad a = a^*(t)+1, \\ \begin{pmatrix} G(Q(a-1)) \times q(1-F_{a-1}(t))P_{a-1}(t) \\ +\int_{Q(a-1)}^{\overline{p}}\left[(p+(1-p)\delta q(1-F_{a-1}(t)))P_{a-1}(t)\right]dG(p) \\ -G(Q(a)) \times q(1-F_a(t))P_a(t) \\ -\int_{Q(a)}^{\overline{p}}\left[(p+(1-p)\delta q(1-F_a(t)))P_a(t)\right]dG(p) \end{pmatrix} & \text{if} \quad a > a^*(t)+1. \end{cases} \end{aligned} \tag{17}$$

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:

$$\begin{aligned} \frac{\partial F_a(t)}{\partial t} &= \sum_{b=1}^{a} \frac{\partial P_b(t)}{\partial t} \\ &= \begin{cases} -q(1-F_a(t))(F_a(t)-F_{a-1}(t)) & \text{if} \quad a \leq a^*(t) \\ \begin{pmatrix} -G(Q(a))\,q(1-F_a(t))(F_a(t)-F_{a-1}(t)) \\ -\int_{Q(a)}^{\overline{p}}\begin{bmatrix} (p+(1-p)\delta q(1-F_a(t))) \times \\ (F_a(t)-F_{a-1}(t)) \end{bmatrix} dG(p) \end{pmatrix} & \text{if} \quad a > a^*(t) \end{cases} \end{aligned} \tag{18}$$

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

be rewritten as:

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

Note that $\partial F_a(t)/\partial t \leq 0$ (and conversely, $\partial H_a(t)/\partial t \geq 0$). 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, Lorenz, and Zilibotti (2016). Differentiating the traveling wave solution yields $\partial F_a(t)/\partial t = -\nu\tilde{f}'$. Therefore, having defined $x = a - \nu t$, we can rewrite (18) as:

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

or, identically,

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

Consider, first, the range $x \leq x^*$. Using the upper part of (20) yields the following Delay Differential Equation (DDE):[39]

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

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:

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

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

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

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

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

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

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. ■

Adding Measurement Error to Model Moments

This appendix describes the approach we use to add m.e. in TFP – implied by the estimated m.e. in revenue – to the trues distribution of TFP. For computational reasons it is necessary to be able to go very quickly from the true TFP Distribution to the observed TFP distribution.

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

Suppose we have a distribution of a, denoted by $P(a)$, simulated from our model. The first task is to convert $P(a)$ to $P(\hat{a})$ – 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 $P(\hat{a}|a)$. For $j \in \{2, \cdots, N-1\}$, we have

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

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

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

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

We now derive the **c**onditional TFP growth. Let us start with observed TFP growth of imitating firms.

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

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

The posterior distribution of a follows

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

To obtain the posterior distribution of μ, first notice that

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

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

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

Then, the posterior distribution of μ follows

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

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

Additional Tables

Table A1: 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)	0.0866***	0.614***	0.573***
	(0.00782)	(0.0184)	(0.0189)
wedge		-0.720***	-0.673***
		(0.0234)	(0.0236)
$export_d$			0.0879***
			(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)	-0.0639***	-0.0657***	-0.0664***
	(0.00458)	(0.00473)	(0.00553)
$R\&D_d$	0.103***	0.0928***	
	(0.0162)	(0.0168)	
$export_d$		0.0386***	0.0386***
		(0.00630)	(0.00626)
R&D $intensity_h$			0.0719**
			(0.0262)
R&D $intensity_m$			0.106***
			(0.0278)
R&D $intensity_l$			0.0966***
			(0.0150)
Industry effects	+	+	+
Age effects	+	+	+
Observations	9,996	9,996	9,996
R-squared	0.081	0.083	0.084

Note: See Table 2 for the same variable definitions and regression specifications with two exceptions. We run pooled regressions with year fixed effects in Panel A. There is no province dummy for Taiwan.

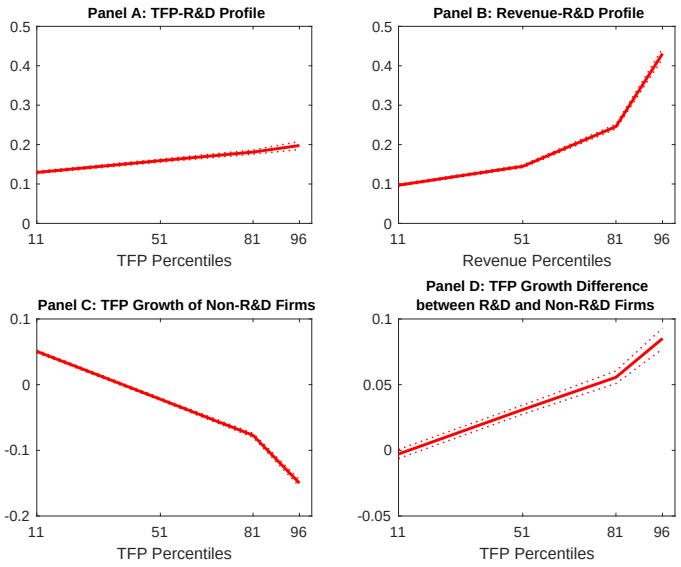
Table A2: Measurement Error Moments

	Empirical averages				Implied estimates		
	$var(\Delta Y)$	$var(\Delta I)$	$cov(\Delta Y, \Delta I)$	$var(TFP)$	ν_Y^2	ν_I^2	ν^2
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: Y, I and TFP represent $\log(P_{it}Y_{it})$, $\log(K_{it}^{\alpha}L_{it}^{(1-\alpha)})$, and $\log(A_{it})$, respectively. 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.

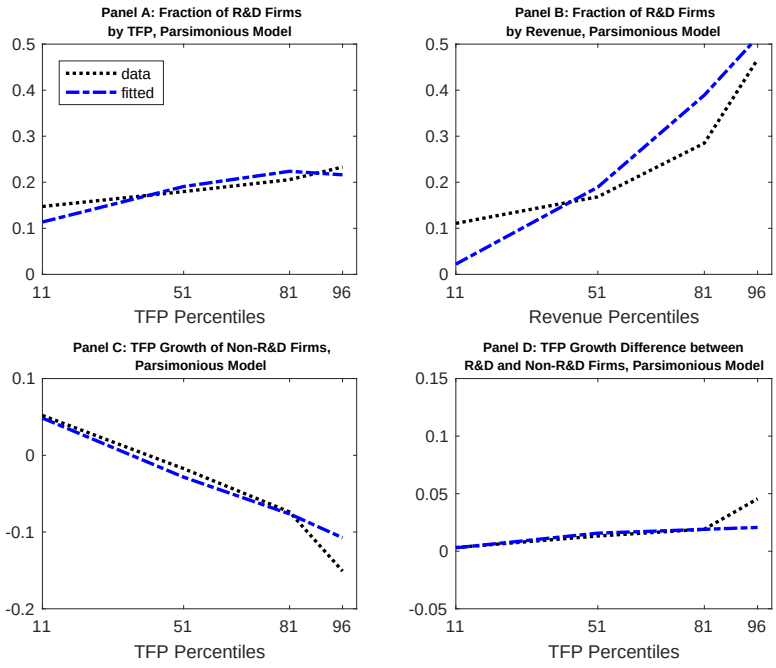
Additional Figures

Figure A1: Chinese Firms in the Balanced Panel 2007–2012



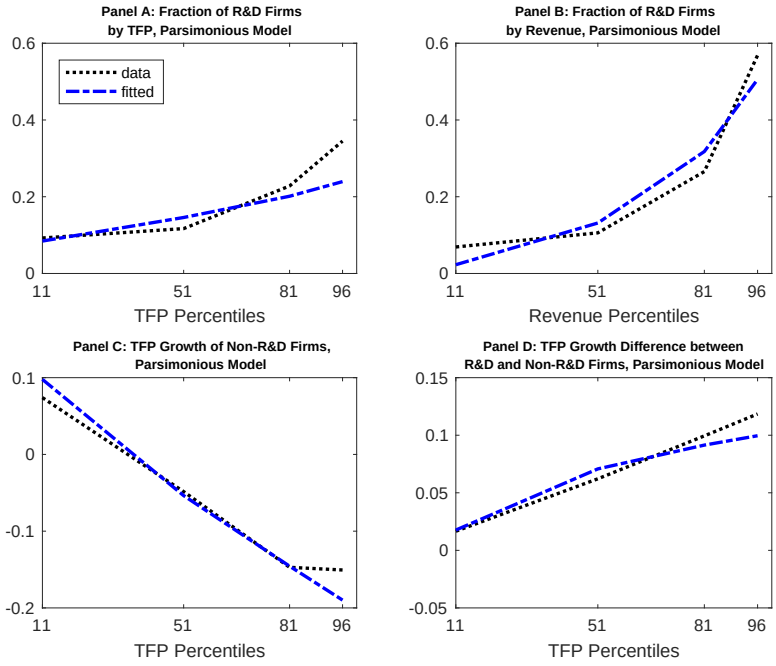
Note: empirical moments for the balanced panel for China 2007–2012. See also Figure 4.

Figure A2: China 2001–07: Parsimonious Model



Note: See Figure 4.

Figure A3: Taiwan 1999–2004: Parsimonious Model



Note: See Figure 4.