Rising Skill Supply, Technological Change, and Innovation: A Quantitative Exploration of China*

Shijun Gu† Chengcheng Jia‡

March 21, 2025

Abstract

Can labor market reforms promote firm innovation and lead to productivity gains in the production sector? In this paper, we examine how college expansion in China contributes to the rapid growth of firm R&D investment and productivity. In our model, heterogeneous firms make endogenous R&D decisions, requiring them to allocate skilled workers between the production and R&D sectors. We structurally estimate the model using firm-level data on the level and distribution of R&D, as well as macro-level data on skill prices and its sectoral allocation. Quantitative analysis reveals that the combined effects of R&D-sector-biased technological shock, skill-biased technological change, and a positive skilled labor supply shock lead to a 12 percent increase in Total Factor Productivity (TFP) between 2004 and 2018. In particular, the skilled labor supply shock accounts for approximately one-fifth of the TFP improvement implied by the model. Counterfactual analysis shows that continuing to expand college enrollment would have an increasing but diminishing effect on firm productivity.

Keywords: R&D, TFP, Skilled Labor, Chinese Economy
JEL classification: O31, O32, J24

*Preliminary and incomplete - please do not cite without authors' approval. We are grateful to Dirk Drueger, Ayse Imrohoroglu, Helu Jiang, Vincenzo Quadrini, Kjetil Storesletten, and Le Xu, as well as seminar and conference participants at the Federal Reserve Bank of Cleveland, 2022 Asian Meeting of the Econometric Society, and 3rd Forum of Chinese Macroeconomic Scholars for helpful comments and conversations. The views stated herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System.

†School of Economics, Central University of Finance and Economics, China. Email: shijungu@cufe.edu.cn
‡Federal Reserve Bank of Cleveland, US. Email: Chengcheng.Jia@clev.frb.org

1 Introduction

During the past two decades, China has been striving to transition from a capital investment-driven economy to an innovation-driven economy to secure sustainable growth. This shift is evident in the rapid growth of R&D intensity (the ratio of aggregate R&D expenditures to sales) in the industrial sector, which increased from 0.6 percent in 2004 to 1.2 percent in 2018. As emphasized by Romer (2000) and Bloom, Van Reenen, and Williams (2019), increasing the quantity of innovative activity requires increasing the supply of workers with the human capital needed to carry out research, and this increase in supply would boost R&D by reducing the equilibrium price of R&D workers. In this paper, using China's higher education expansion policy as an example, we explore to what extent rising skills supply can foster firm innovation and promote total factor productivity (TFP) growth.

In our model, heterogeneous firms make endogenous R&D choices to improve their productivity. The novel aspect of the theory is that firms decide how to allocate skilled workers between the production and R&D sectors. Skill prices are determined by the equilibrium conditions of the labor market. We calibrate two aggregate shocks—a productivity shock biased toward the R&D sector and a skill-biased technological shock—that drive the increasing demand for skilled workers. Meanwhile, the college expansion program in China increased the supply of skilled workers. Together, these factors led to a 12 percent increase in TFP between 2004 and 2018, with the skill supply shock contributing to one-fifth of the TFP growth. Counterfactual experiments also show that continuing to expand college capacity would increase TFP by an additional 3 percent as the skill price is lowered further.

We begin by documenting two simultaneous empirical trends in China since 2004: the rising share of college-educated workers in the labor force and the increasing expenditures on innovation. These patterns lead us to conjecture that the growing supply of college-educated labor may contribute to the rapid growth in firms' R&D investment in China. Furthermore, the data also indicate a growing importance of skilled labor in the process of firm innovation, as firms allocate more skilled labor to perform R&D tasks and R&D activities become more labor-intensive. As a result, in our model, we allow firms to endogenously choose the allocation of skilled labor between the R&D and production sectors. Finally, we also find that the research productivity of R&D workers has significantly improved. This motivates us to consider technology shocks biased toward the R&D sector and skilled labor as drivers of the rising demand for R&D workers.

In the empirical section, we also provide suggestive evidence of the determinants of firms' R&D decisions and the effects of R&D on productivity growth using firm-level data, which contain information on firms' production variables, balance sheets, and R&D expenditures.

We find that the fraction of firms engaging in R&D activities (referred to as "R&D firms" in the rest of the paper) increases with productivity levels and asset levels. These empirical findings imply that financial frictions play a significant role in firms' R&D decisions, highlighting the need to incorporate financial frictions into our model.

We then develop a model of heterogeneous firms with endogenous R&D decisions under financial frictions. In the economy, workers are heterogeneous in their labor productivity and education levels. We classify workers with a bachelor's degree as skilled labor. There is a continuum of firms that are heterogeneous in their productivity and net worth. All firms produce homogeneous manufacturing goods using a decreasing-returns-to-scale production technology that combines capital, skilled labor, and unskilled labor, where capital demand is subject to a collateral constraint.

In each period, firms make a decision on whether to perform R&D and how much *ideas* (the output of R&D expenditures) to produce. The cost of R&D consists of a fixed cost incurred every period and a variable cost, which includes wage expenses for skilled workers and expenditures on intermediate inputs. If a firm chooses not to perform R&D, its productivity evolves according to an AR(1) process. If a firm invests in R&D, its productivity in the next period improves on top of the AR(1) process, with the magnitude of the improvement depending on the amount of R&D *ideas* it produces. The endogenous R&D decisions create a trade-off for firms between current-period profits and future productivity gains.

We calibrate model parameters by matching the stationary equilibrium of our model with the Chinese economy in the early 2000s. We employ a standard two-step calibration strategy. First, a subset of parameters is selected outside the model. Second, the remaining parameters, reflecting unique characteristics of the Chinese economy, are internally calibrated to match the aggregate and distributional patterns in China's Annual Survey of Industry (ASI) and Economic Census. Under our calibration, the R&D production function exhibits a higher degree of decreasing returns to scale compared to the manufacturing good production function, and the two input factors are gross substitutable. Doubling the amount of R&D ideas leads to a 6 percent increase in productivity in the following period. Our model's predictions closely align with the empirical data for the moments that we do not explicitly target in the calibration.

We feed three aggregate shocks into the estimated model to capture all relevant factors driving the rapid growth in firms' R&D investment between 2004 and 2018. Specifically, the skilled labor supply shock is calibrated to match the change in the share of college graduates in the urban labor force, the skill-biased technological shock is constructed to match the change in the college wage premium, and the R&D-sector-biased technological shock is calibrated to match the change in the R&D expenditures-to-output ratio in the

industrial sector.

The combined effect of the three shocks leads to a 128 percent increase in R&D ideas per R&D worker. In the data, patent per R&D worker increased by 144 percent, aligning with our model's prediction. The experiment generates a 12 percent rise in TFP, with the skilled wage rate increasing by 38 percent. Although the skilled wage rate is higher, firms allocate more skilled labor to perform R&D tasks, with the share of R&D workers in the skilled labor force increasing from 19 percent to 25 percent. This is because skilled labor becomes significantly more productive due to the skill-biased technological shock.

To isolate the effect of rising skilled labor supply, we simulate a counterfactual economy in which the share of skilled labor is fixed at its 2004 level. In this case, the skilled wage rate is 158 percent higher than in the benchmark model, implying a significantly higher cost of employing skilled labor. As a consequence, firms allocate significantly less skilled workers to perform R&D tasks (15 percent). Since firms can substitute out of the use of skilled labor when performing R&D, total R&D expenditures drop only slightly. However, over one-fifth of TFP growth is lost, as the positive effect of skill-biased technological change is offset by the substantially increasing cost of skilled labor.

Lastly, since the share of college-educated workers is expected to grow gradually in the near future, we use our model to predict the long-run impact of the college expansion program on firm productivity. We find that continuing to expand college capacity would increase TFP by an additional 3 percent, as the skilled labor wage is further reduced. However, the effect on TFP would diminish due to the increase in the equilibrium wage of unskilled workers. More specifically, although producing R&D ideas becomes cheaper, firms' profits eventually decrease because unskilled labor is a crucial input in the production function of manufacturing goods. This would hinder firms' ability to invest in innovation. These results shed light on the optimal design of both education and innovation policies.

Related Literature

Our paper connects three strands of literature. First, we contribute to the growing literature that studies the recent surge of R&D expenditure and TFP in China. Our paper is most closely related to König et al. (2020) and Buera and Fattal Jaef (2018). Both of these papers model endogenous firms' R&D decisions in the presence of misallocation. Relative to König et al. (2020), we also add the context of China's college expansion program and study how the connection between skilled labor supply and firms' R&D decisions are affected by financial frictions. Previous research has studied other drivers of firms' R&D investments in China, including institutional factors (Hu and Jefferson (2009)), the effects of tax incentives, (Li

(2012), Dang and Motohashi (2015), Jia and Ma (2017), and Dai and Wang (2019)), and the possibility of relabeling Chen et al. (2021) other expenses as R&D.

Second, our paper is related to the literature on the connection between human capital and innovation. Previous literature has provided empirical evidence (e.g. Aghion et al. (2009), Toivanen and Väänänen (2016), Aghion et al. (2017)) and theoretical explanations (e.g., Akcigit, Pearce, and Prato (2020) and Bloom, Van Reenen, and Williams (2019)) for the relationship between human capital, firm R&D and productivity growth. In the context of China, Che and Zhang (2018) show empirical evidence that the policy induced increase in college graduates in China leads to more productivity growth in human-capital intensive industries. Ma (2024) is the first one that evaluates the effects of the increase in the share of skilled labor on firms' innovation choices in a quantitative model. Ma (2024) emphasizes the linkage between innovation and export decisions, whereas our novelty is to incorporate financial frictions in studying the effect of rising share of skilled labor on firm R&D activities in China.

Third, our paper extends the existing quantitative macro literature on the effects of financial frictions on on aggregate productivity. The majority of previous papers on this topic focus on how financial frictions reduce TFP by distorting the allocation of inputs across firms. We contributes to this literature by assuming each firm's productivity is endogenous and each firm makes optimal R&D decisions. Previous studies have provide both empirical evidence (see i.e., Brown, Martinsson, and Petersen (2012), Peia and Romelli (2022), and Mezzanotti and Simcoe (2023)) and theoretical explanations (see i.e., Ottonello and Winberry (2024)) of how financial frictions alter firms' R&D decisions. In the context of the Chinese economy, Wei, Xie, and Zhang (2017) show evidence that the returns of private R&D investments are significantly heterogeneous due to resource misallocations. Resource misallocation has been identified as a key feature in the Chinese economy (e.g. Hsieh and Klenow (2009), Song, Storesletten, and Zilibotti (2011), and Brandt, Tombe, and Zhu (2013)).

2 Motivating Facts and Descriptive Evidence

In this section, we first briefly describe the institutional background of China's college expansion program and show its impact on the supply of skilled labor. (Section 2.1) We then summarize the changes in both the quantity and research productivity of R&D workers. (Section 2.2) Finally, we present suggestive evidence on the determinants and effects of R&D decisions at the firm level.

2.1 Rising Share of Skilled Labor

In China, university education is highly controlled by the state. The Ministry of Education manages the admission process and assigns quotas of admitted students in each year. In 1999, the central government initiated the "college expansion program". As a result, the number of college graduates increased from slightly more than 1 million to almost 4 million between 2004 and 2018.[1] (Figure 1, Panel A) The proportion of college graduates population aged 22 in the respective years shown on the x-axis of the graph increased from approximately 5 percent to 19 percent.

In addition to enrolling more college students, China has implemented other higher education reforms, such as promoting world-class research universities, increasing research funding, encouraging international collaboration, and adopting a more flexible college entrance exam system to improve educational quality and accessibility. Appendix A.2 provides a detailed breakdown of educational attainments by age groups from 2000 to 2020.

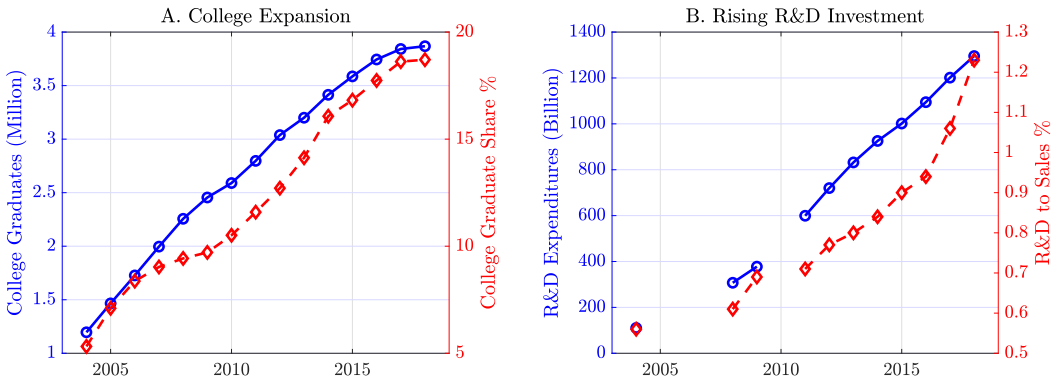


Figure 1: College Expansion and Rising R&D Investment in China

Note: College graduates refers to the number of four-year college degree recipients in the respective years, measured in millions. College graduate share refers to the ratio of college graduates to the population aged 22. R&D expenditures refers to the total R&D spending of all industrial enterprises above a designated size, measured in billions of RMB. R&D to sales refers to the ratio of R&D expenditures to sales revenue for all industrial enterprises above a designated size. Data sources include the China Statistical Yearbook and the China Economic Census Yearbook.

2.2 Cost, Quantity, and Productivity of R&D Workers

As shown in Panel B in Table 1, the influx of workers with college degrees into the labor market coincided with a rapid rise in aggregate R&D expenditure. Total R&D expenditure increased approximately eight times between 2004 and 2018, while the R&D-to-sales ratio rose from around 0.6 percent to 1.2 percent.

[1] We only account for 4-year college and do not account for 2- or 3-year vocational college programs.

In this paper, we conjecture that the increasing supply of college-educated labor can promote firm innovation in China, as skilled workers are essential inputs for R&D tasks. Data on the use of R&D workers and research output in Chinese industrial enterprises from 2004 to 2018 provide further evidence for our hypothesis.

The first two rows of Table 1 show that the number of workers engaged in R&D tasks in Chinese industrial enterprises increased 5.5-fold over 14 years, and the share of R&D personnel in total employment rose from 0.8 percent to 3.6 percent.[2] Meanwhile, as shown in the third and fourth rows, the share of wage expenditures for R&D workers in total R&D spending increased by nearly 12 percentage points,[3] and an increasingly larger proportion of college-educated workers were assigned to R&D tasks. These data not only reflect a significant rise in the demand for skilled labor capable of conducting R&D tasks but also highlight the growing importance of skilled labor in the process of firm innovation.

Next, we discuss China's R&D output, which is measured by patent production, and R&D productivity, which is measured by patents per R&D worker. From 2004 to 2018, as shown in the fifth and sixth rows of Table 1, the number of patent applications filed by Chinese industrial enterprises increased more than 18.5-fold, and the number of patent applications per R&D worker increased by 255 percent. This reflects a significant improvement in the research productivity of R&D workers in China over the past two decades. To control the potential influence of patent quality on research productivity, we made two adjustments to the patent application data. First, the reported patent application data include only invention patents, excluding utility model patents and design patents. Second, we also calculated patents granted per R&D worker, which accounts only for successfully granted patents. After implementing the second adjustment, we found that the research productivity of R&D workers increased by 144 percent.[4]

[2]Here, we refer R&D workers as the number of employees engaged in R&D tasks in industrial enterprises, multiplied by their average hours spent on R&D tasks. This adjustment is made primarily because some R&D workers are involved in R&D activities on a part-time basis.

[3]A widely known issue in Chinese plant-level data is that the National Bureau of Statistics (NBS) understates actual labor cost. For example, the aggregate share of wage cost is roughly 30 percent, which is significantly lower than the aggregate labor share in the manufacturing sector in the Chinese national account. This inconsistency is also documented in Bai, Hsieh, and Qian (2006) and Hsieh and Klenow (2009) A possible explanation for this inconsistency is that the NBS only reports wage income but does not provide information on non-wage benefits, including bonus, pension, insurance, and housing allowance. As a result, when calculating the cost share of labor in R&D activities, we assume non-wage benefit is a fixed portion of the total labor compensation, and adjust the raw data (22.5 percent in the 2004 update of China Economic Census Yearbook) by the same factor (1.6), following Bai, Hsieh, and Qian (2006) and Hsieh and Klenow (2009).

[4]One might argue that the significant rise in patents per R&D worker in China stems from the government's stricter enforcement of patent laws rather than an improvement in the research productivity. This concern might indeed have been valid in the early 2000s, as China strengthened its crackdown on patent infringement to meet the requirements for joining the WTO. However, we find that the largest increase in

Table 1: R&D Workers: Cost, Quantity, and Productivity

	2004	2008	2013	2018
Skill share and allocation				
R&D workers (×1,000)	542	1,230	2,494	2,981
R&D workers cost share %	36.0	32.5	43.7	47.9
Share of R&D worker in labor force %	0.82	1.39	2.55	3.57
Share of R&D worker in skilled labor force %	21.2	20.3	28.7	31.2
Patent production				
Patent applications (×1,000)	20	59	205	371
Patent applications per R&D worker	0.04	0.05	0.08	0.13
Patent granted per R&D worker	0.014	0.016	0.021	0.034
Patent application approval rate %	37.9	32.3	25.2	28.0

Note: R&D workers refers to the number of employees engaged in R&D tasks in industrial enterprises in the respective years, multiplied by their average hours spent on R&D tasks. Share of R&D workers in labor force refers to the number of R&D workers divided by the total number of employees in industrial enterprises. R&D workers cost share refers to the ratio of compensation for R&D workers to total R&D expenditures. Share of R&D workers in skilled labor force refers to the ratio of the number of R&D workers to the number of college-educated workers. Patent applications refers to the total number of invention patent applications filed by industrial enterprises in the respective years. Patent applications per R&D worker refers to the ratio of the total number of invention patent applications to the number of R&D workers. Patents granted per R&D worker refers to the ratio of the total number of invention patents granted to the number of R&D workers. Patent application approval rate refers to the ratio of the total number of invention patents granted to the total number of invention patent applications. Data sources include the China Statistical Yearbook and the China Economic Census.

2.3 Firm-Level Evidence on R&D

In this section, we use firm level data to investigate the determinants and the effects of R&D expenditure. We obtain firm level data from the Annual Survey of Industries (ASI) conducted by China's National Bureau of Statistics, which is a census of all state-owned firms and the private firms in the manufacturing industry with annual revenue above 5 million RMB.[5]

To investigate how firms' R&D decisions are affected by other factors, we do the following regression analysis using a linear probability model. The dependent variable is a dummy variable whose value is 1 for R&D firms, and 0 for non-R&D firms. The main independent variable is the firm's TFP in the current year,[6] Columns (1) - (3) in Table 2 show that the

patents per R&D worker occurred between 2013 and 2018, a period when China's patent protection system had already been well-established. This suggests that stricter enforcement of patent laws is not the primary driver of the increase in patents per R&D worker.

[5]We do not use ASI after 2008 because because value-added firm output is no longer reported after 2008. Information on R&D is only available in the year 2001 and 2005 - 2007.

[6]We estimate firm TFP using the Solow residual method, which requires a specification of the production function. We assume a decreasing return to scale production function, taking the form of $Y_t = Z_t K_t^{\alpha} L_t^{\theta}$,

fraction of R&D-firms robustly increases with firm TFP, after controlling for the amount of equity in the firm and whether the firm's ownership type.[7] In addition, firms with higher net worth are more likely to invest in R&D. State-owned enterprises (SOEs) make more R&D investments compared to private firms.

Table 2: Characteristics of Firm R&D Decisions

	(1) $R\&D_d$	(2) $R\&D_d$	(3) $R\&D_d$	(4) $\log(TFP_{t+1})$	(5) $\log(TFP_{t+1})$
$\log(TFP_t)$	0.033	0.021	0.028	0.681	0.698
	(0.004)	(0.003)	(0.003)	(0.018)	(0.011)
log(equity)		0.071	0.066		
		(0.005)	(0.005)		
SOE			0.121		
			(0.014)		
$R\&D_d$				0.050	
				(0.014)	
log(R&D)					0.032
					(0.004)
R-squared	0.149	0.250	0.259	0.452	0.530

Notes: The dependent variable, $R\&D_d$ is a dummy variable with the value 1 being R&D firm, and 0 for non-R&D firm. $\log(TFP)$ is the logarithm of firm TFP. Equity is calculated as the difference between total asset and total debt. SOE_d is a dummy variable with the value 1 being a state-owned enterprises. All regressions include industry, start year, and province fixed effect. Regressions are weighted by the size of total wage bills. Standard errors are clustered at industry level and are reported in parenthesis. We use a balanced panel between 2001 and 2002.

To further quantify the effects of R&D expenditure on productivity growth, we estimate the following TFP process:

$$\log(TFP)_{i,t+1} = \rho \log(TFP)_{i,t} + \gamma\ X_{i,t} + \epsilon_{i,t+1} \tag{1}$$

where $X_{i,t}$ measures R&D activities. We first estimate whether making positive R&D expenditure increases productivity in the next period. To do so, we let $X_{i,t}$ be the dummy variable taking the value 1 for R&D firms, and 0 for non-R&D firms. Then, for all R&D

where Y_t, K_t, and L_t are value-added output, capital, and labor of a firm, correspondingly. To account for the heterogeneity in labor inputs, we use total wage bills to proximate L_t. We chose $\alpha = 0.40$ and $\theta = 0.45$ consistent with Bai, Hsieh, and Qian (2006). We also include control variables of the firm's equity value and ownership type. We also regress TFP (in logs) on start year, province, and 2-digit industry to control for variation caused by these factors.

[7]The dummy variable SOE takes the value 1 if a firm is a state-owned enterprises according to its registration type, or the state has an absolute or relative controlling share of the firm. The two definitions of state ownership (registration type and controlling share) do not always overlap since China started the reform on SOEs in 1998, and since then, a growing fraction of registered SOEs were corporatized. See Hsieh and Song (2015) and Gu and Jia (2022) for more information.

firms, we test whether making greater R&D investments leads higher growth rate in the next period. To this end, we restrict the sample to be all R&D firms, and $X_{i,t} = \log(R\&D)_{i,t}$ is the logarithm of the amount of R&D expenditure.

The last two columns in Table 2 show that the auto-correlation coefficient of the productivity process is roughly 0.7 at an annual rate. Conditional on the current TFP level, making positive R&D investments increases TFP by 5 percent on average in the next period. In addition, conditional on making positive R&D investments, doubling R&D increases TFP in the next period by 3 percent. This number is at the lower end in the range of estimates of the private returns of R&D. Most empirical studies on this topic are based on micro data in developed countries. Hall, Mairesse, and Mohnen (2010) review this literature and find this R&D elasticity parameter to have a broad range between 2 percent and 25 percent. Chen et al. (2021) use a different source of data from the Chinese State Administration of Tax and find the R&D elasticity parameter to be 9 percent from 2006 to 2011.

Summary. In this section, we provide the institutional background of China's college expansion program and present suggestive evidence on the connection between the increasing supply of skilled labor and the growing expenditure on R&D. In addition, we show that both the quantity and productivity of R&D workers increased significantly from 2004 to 2018, suggesting that the rise in research productivity in the R&D sector may be another crucial driver. Furthermore, using firm-level data, we show that the probability of making positive R&D investments increases with a firm's productivity and net worth, and that R&D investments significantly improve productivity in the following period.

The data presented in this section lead us to model skilled labor as an input in R&D, with firms optimally making R&D decisions and allocating skilled workers between production and R&D. In addition, the data suggest that two types of aggregate shocks drive the growth in R&D expenditure: a labor supply shock that increases the share of skilled workers and a sector-biased technology shock that improves the research productivity of R&D workers. The goal of our subsequent analysis is to quantify the effects of these shocks on R&D expenditure and total factor productivity (TFP) in the Chinese economy.

3 Model

We develop a heterogeneous-firms model with endogenous R&D decisions. All firms produce a homogeneous manufacturing good for sale, and may choose to invest in R&D to improve their firm-specific productivity in the next period. Firms' capital demand may be constrained due to financial frictions, which also affect their decisions on R&D expenditure. We outline

the model assumptions in Section 3.1, formulate the recursive problem in Section 3.2, and define the stationary equilibrium in Section 3.3.

3.1 Environment

Time is discrete and the horizon is infinite with $t \in \{1, 2, ..., \infty\}$.[8] There is no aggregate uncertainty.

Demographics. There is a continuum of firms of measure one. The total labor supply is also normalized to one. N_t^s denotes the share of skilled labor, corresponding to the share of workers with a bachelor-or-above degree in the data.[9] The remaining workers, N_t^u, make up the unskilled labor force.

Labor productivity and education policy. We assume workers are heterogeneous in their labor productivity h, which follows a Gamma distribution $\Gamma(\zeta_1, \zeta_2)$ where the two shape parameters, ζ_1 and ζ_2, are time invariant and will be calibrated externally (see Section 4.1).

Workers do not make education choices. Rather, the education policy decides a productivity cutoff, χ_t, above which all individuals are selected in receiving college education and becoming skilled workers. The cutoff is time variant, reflecting the expansion of the higher education program in China.

Manufacturing good production. All firms produce a homogeneous manufacturing good and operate under a decreasing-returns- to-scale technology. Firms are heterogeneous in productivity, and demand skilled labor, unskilled, labor, and capital in production.

The production function of the manufacturing good is given by

$$y = F(k, n_{s,y}, n_u) = Azk^{\alpha} \left[\gamma (S_t n_{s,y})^{\sigma} + (1-\gamma) n_u^{\sigma}\right]^{\frac{\theta}{\sigma}}, \tag{2}$$

where A is the aggregate productivity level, and z is the idiosyncratic productivity level. $n_{s,y}$ and n_u denote the efficient units of skilled and unskilled labor used in production, respectively. S_t is a skilled biased technology shock that will be calibrated in Section 5. α is capital share, θ is labor share, γ is skilled labor share in producing the manufacturing good, and σ controls the elasticity of substitution between skilled and unskilled labor.

[8] The time subscript indicates that the time variation in the corresponding variable can be treated as an aggregate shock later in Section 5. Otherwise, we omit the time index.

[9] Those who have completed vocational or part-time college degrees are not treated as skilled workers in our model. We set a relatively high bar for skilled workers because in our model, all skilled workers are capable of doing R&D, and R&D requires more advanced skills than part-time or vocational college programs typically provide.

Financial market. In our model, we assume that firms borrow intra-period loans to finance their capital demands. The cost of capital is the sum of the capital rental rate r and the depreciation rate of capital δ. We assume the financial market is imperfect, and firms are subject to a collateral constraint. A firm's choice of capital may be constrained, and the borrowing limit depends on the firm's net worth, a. Specifically, the maximum loan that a firm can have is given by

$$k \leq \lambda a, \tag{3}$$

where λ is a parameter reflecting the tightness of the financial constraint.

R&D. In each period, firms may pay a fixed cost to perform R&D activities (become R&D-active firms). The magnitude of the fixed cost depends on the firm's current productivity, given by fz^{η}, where f and η govern the scale and shape of the fixed cost.

Conditional on being "R&D active", a firm then chooses the amount of R&D activities it performs. We call the output of R&D activities as "ideas" (x). Ideas are firm-specific, meaning they can only improve firm-specific productivity in the next period and cannot be traded. Firms combine two input factors in producing R&D ideas: skilled labor, $n_{s,x}$, and manufacturing good as intermediate inputs, i.

The production function of R&D is given by

$$x = J(n_{s,x}, i) = B_t \left[\psi \left(S_t n_{s,x}\right)^{\kappa} + (1-\psi) i^{\kappa}\right]^{\frac{\mu}{\kappa}} \tag{4}$$

where B_t is a time-varying technology shock for R&D activities. (There is no idiosyncratic technology variation in ideas production.) ψ is the share of skilled labor in the R&D sector, κ controls the elasticity of substitution between skilled labor and intermediate input, and $\mu < 1$ implies the production of the R&D ideas exhibits decreasing returns to scale. The skilled biased technology shock, S_t, also affects the labor productivity of the skilled labor.

Productivity process. If a firm chooses not to perform R&D (an R&D-inactive firm), its idiosyncratic productivity z follows the standard AR(1) process in logs

$$\log(z') = \rho \log(z) + \epsilon', \ \epsilon' \sim \mathrm{N}(0, \sigma_{\epsilon}), \tag{5}$$

where ρ governs the persistence of productivity process and σ_{ϵ} is the standard deviation of a *i.i.d.* shocks, ϵ.

On top of the stochastic AR(1) process, the next-period productivity of an R&D-active firm have an endogenous increase, and the size of the increase depends on the amount of

R&D ideas it makes in the current period. Specifically, the productivity process on an R&D active firm follows:

$$\log(z') = \rho\log(z) + \phi\log(1+x) + \epsilon',\ \epsilon' \sim \mathrm{N}(0, \sigma_\epsilon), \tag{6}$$

where ϕ controls the elasticity of productivity gains with respect to the amount of R&D ideas.

Government policies. The government taxes all profit-making firms at a uniform rate τ_c, and provides proportional subsidies to all R&D-active firms for their R&D expenses at the rate of τ_x.[10]

3.2 A Dynamic Problem of R&D Choices

Firms are heterogeneous in productivity and net worth. At the beginning of each period, after observing their net worth and productivity shocks, firms decide whether to become an R&D-active or R&D-inactive firm. If a firm decides to be an R&D-inactive firm, it faces a standard intra-period profit-maximization problem and a inter-period consumption-saving decision. If a firm decides to be an R&D active firm, it pays a fixed cost, and in addition to the profit-maximization and consumption-saving decisions, it chooses the amount of ideas to produce in R&D activity. For a given amount of R&D ideas to produce, the firm optimally allocates intermediate goods input and skilled labors to minimize the cost in R&D activity.

Profit maximization. Conditional on the current period's productivity and net worth, all firms maximizes their operation profit[11]

$$\pi(a,z) = \max_{k,n_{s,y},n_u} \{F(k,n_{s,y},n_u) - (r+\delta)k - w_s n_{s,y} - w_u n_u\} \tag{7}$$

subject to

$$0 \le k \le \lambda a, n_{s,y} \ge 0, n_u \ge 0 \tag{8}$$

where w_s is the wage rate for the skilled labor, w_u is the wage rate for the unskilled labor.

R&D cost minimization. Let x denote the amount of ideas that an R&D-active firm decides to produce. It then chooses the optimal intermediate goods input, i, and skilled

[10] In reality, China has more complex tax incentives for R&D expenses, as extensively explored by Chen et al. (2021)). We abstract from this complexity to focus on the other drivers of firms' R&D decisions.

[11] The operation profit defined here does not include the fixed and variable expenditures of R&D.

labor, $n_{s,x}$, to minimize the cost of R&D production. The R&D cost-minimization problem can be expressed as:

$$\Omega(x) = \min_{i,n_{s,x}} \{i + w_s n_{s,x}\} \tag{9}$$

subject to

$$x \leq J(n_{s,x}, i), i \geq 0, n_{s,x} \geq 0 \tag{10}$$

where $\Omega(x)$ is the minimum cost of producing x units of R&D ideas.

Dynamic problem. A firm chooses whether or not to be an R&D-active firm by comparing $v^E(a,z)$ and $v^R(a,z)$, where $v^E(a,z)$ and $v^R(a,z)$ denote the value of an R&D-inactive and an R&D-active firm, respectively. The value of a firm conditional on their net worth and productivity is given by

$$v(a,z) = \max\left\{v^E(a,z), v^R(a,z)\right\}. \tag{11}$$

An R&D-inactive firm maximizes current-period's operating profit and then decides the amount of net worth to carry over to the next period. Its dynamic problem can be expressed as:

$$v^E(a,z) = \max_{c,a'}\left\{u(c) + \beta \mathbb{E}_{z'}\left[v(a',z')\right]\right\} \tag{12}$$

subject to

$$c + a' + \tau_c \pi(a,z) = (1+r)a + \pi(a,z) \tag{13}$$

$$c \geq 0, a' \geq 0 \tag{14}$$

The evolution of the productivity z is according to Equation 5.

R&D-active firms face a trade-off where an increase in the amount of ideas they produce leads to a reduction in their current-period's profit but an increase in their next-period's productivity. The dynamic problem of an R&D-active firm can be expressed as:

$$v^R(a,z) = \max_{c,a',x}\left\{u(c) + \beta \mathbb{E}_{z'}\left[v(a',z')\right]\right\} \tag{15}$$

subject to

$$T(a,z,x) = \pi(a,z) - f z^{\eta} - (1-\tau_x)\Omega(x) \tag{16}$$

$$c + a' + \tau_c \max\{T(a,z,x), 0\} = (1+r)a + T(a,z,x) \tag{17}$$

$$c \geq 0, a' \geq 0, x \geq 0 \tag{18}$$

where productivity z evolves according to Equation 6 and $T(a,z,x)$ is the taxable corporate income, which equals the operation profit less the fixed and variable R&D expenses.

3.3 Definition of Equilibrium

We study a competitive equilibrium for a small open economy, in which the world's interest rate is r. Firms are indexed by individual states $\mathbf{s} = \{a, z\}$. The recursive competitive equilibrium is defined by a set of value function $\{v(\mathbf{s}), v^E(\mathbf{s}), v^R(\mathbf{s})\}$, the allocation of R&D-active firms $X^R = \{c(\mathbf{s}), a'(\mathbf{s}), k(\mathbf{s}), n_u(\mathbf{s}), n_{s,y}(\mathbf{s}), i(\mathbf{s}), n_{s,x}(\mathbf{s})\}$, the allocation of the R&D inactive firms $X^E = \{c(\mathbf{s}), a'(\mathbf{s}), k(\mathbf{s}), n_u(\mathbf{s}), n_{s,y}(\mathbf{s})\}$, government policies $\{\tau_c, \tau_x\}$, prices $\{r, w_s, w_u\}$, and the joint cumulative distribution of firms $\Psi = \{\Psi^E(\mathbf{s}), \Psi^R(\mathbf{s}))\}$ such that

1. The value and policy functions solve the maximization problem described in Section 3.2, given factor prices and government policies.

2. The skilled labor market clears

$$\int_{\chi}^{\infty} \Gamma(h; \zeta_1, \zeta_2) dh = \int n_{s,y}(\mathbf{s}) d\Psi^R(\mathbf{s}) + \int n_{s,y}(\mathbf{s}) d\Psi^E(\mathbf{s}) + \int n_{s,x}(\mathbf{s}) d\Psi^R(\mathbf{s}) \tag{19}$$

3. The unskilled labor market clears

$$\int_{0}^{\chi} \Gamma(h; \zeta_1, \zeta_2) dh = \int n_u(\mathbf{s}) d\Psi^R(\mathbf{s}) + \int n_u(\mathbf{s}) d\Psi^E(\mathbf{s}) \tag{20}$$

4. The distribution Ψ is a fixed point where its transition is consistent with the policy functions and the law of motion for Ψ:

$$\Psi = \Phi(\Psi) \tag{21}$$

   where Φ is a one-period ahead transition operator such that $\Psi' = \Phi(\Psi)$.

4 Calibration

The initial steady state of the model is calibrated to the Chinese economy in 2004. We first explain the parameters that are calibrated externally. (Section 4.1) The rest of the parameters, which represent the unique features in the Chinese economy, are estimated internally within the model. (Section 4.2) Section 4.3 evaluates the performance of our calibration by comparing a set of non-targeted moments generated by our model with their empirical counterparts.

4.1 Externally Calibrated Parameters

The model is calibrated at an annual frequency. Table 3 list all the parameters that we calibrated externally. The values we assign to these parameters are either standard in the literature or can be directly measured from the data.

The factor-neutral aggregate TFP level, A, is normalized to unity. The degree of decreasing returns to scale is set to be $\alpha+\theta=0.85$, which is standard in the literature. We set the depreciation rate of capital $\delta=0.10$, and the capital share parameter $\alpha=0.40$, consistent with the estimates of the Chinese economy in the early 2000s reported by Bai, Hsieh, and Qian (2006). The elasticity of substitution between skilled and unskilled labor σ is set to 0.30 following Li (2010). The capital rental rate is set to 6 percent.

Table 3: Externally Calibrated Parameters

Parameter		Value	Source
Production			
Total factor productivity	A	1.00	Normalization
Span-of-control (production)	$\alpha+\theta$	0.85	Standard
Share of capital	α	0.40	Bai, Hsieh, and Qian (2006)
Capital depreciation rate	δ	0.10	Bai, Hsieh, and Qian (2006)
Elas. of substitution btw. skills	σ	0.30	Li (2010)
Capital rental rate	r	0.06	Standard
Skill endowment			
Measure of workers	N	1.00	Normalization
Share of skills	N_t^s	0.04	China Population Census
Distribution (scale parameter)	ζ_1	9.15	CFPS 2010-18
Distribution (rate parameter)	ζ_2	0.11	CFPS 2010-18
Government policy			
Corporate income tax rate	τ_c	0.33	State Taxation Administration
R&D subsidy rate	τ_x	0.06	2000 R&D Census

The total supply of workers is normalized to one. The share of skilled labor is set to 0.04, matching the share of the working age urban population that have college-or-above degrees in the 2000 Population Census in China.[12] Since the college expansion program was initiated in 1999 and the implementation took a gradual process, the share of skilled labor was stable in the early 2000s. In addition, we use the math tests for the adult in the labor force from the China Family Panel Studies (CFPS) to proxy the Γ distribution of labor

[12]Ideally, we should use data on the educational attainment of workers in the industrial sector, since our firm-level data are taken from the Annual Survey of Industries (ASI). However, the ASI did not release this information except in 2004, so we use data on urban workers from the Population Census as a proxy. By comparing the 2000 Population Census and the 2004 ASI, we have confirmed that the educational attainment of urban workers in the Population Census is similar to that of workers employed in the industrial sector in the ASI.

productivity. In the initial steady state, the labor productivity at the cutoff for skilled labor is 63 percent higher than the average labor productivity. (Figure 6 in Appendix B plots the labor productivity distribution and the cutoff for college admission.)

The corporate income tax rate is 33 percent, mirroring the tax rate for domestic firms in the early 2000s. The R&D subsidy rate is 6 percent, which is calculated as the aggregate tax deduction for R&D investments, divided by aggregate R&D expenditure.[13]

4.2 Internally Calibrated Parameters

The remaining 12 parameters are internally estimated within the model, which are summarized in Table 4. The targeted moments and their empirical and model-predicted values are displayed in Table 5. These parameters are calibrated to replicate the unique aggregate and distributional features of Chinese manufacturing firms, especially with regard to patterns of R&D investments. The majority of targeted moments are calculated using the 2001-02 update of the ASI and 2004 update of China Economic Census Yearbook (CECY), with the exception of the skill wage premium, which is derived from the 2002 update of Chinese Household Income Project (CHIP).

Table 4: Internally Calibrated Parameters

Parameter		Value
Preference		
Discount factor	β	0.855
Manufacturing good production		
Borrowing constraint	λ	2.839
Skills share in production	γ	0.200
R&D good production		
R&D sector-biased technology	B	0.927
Span-of-control (R&D)	μ	0.720
Skills share in R&D	ψ	0.573
Elas. of substitution	κ	0.741
R&D fixed cost		
Fixed cost, level ($\times 100$)	f	0.044
Fixed cost, curvature	η	4.523
Productivity process		
Persistence of productivity	ρ_z	0.699
SD of productivity shocks	σ_ϵ	0.451
Scale effect of R&D	ϕ	0.063

We choose the 12 parameters to match 12 data moments using Simulated Method of

[13]Source: China Economic Census in 2004.

Moments (SMM).[14] Even though every targeted moment is determined simultaneously by all parameters, in what follows, we discuss each of the moments in relation to the parameter for which the moment yields the most identification power.

Preferences. We choose the discount factor $\beta = 0.855$ to match the aggregate asset-to-output ratio. A higher β means firms are more patient and thus maintain a higher level of asset.

Manufacturing good production. We choose the borrowing constraint $\lambda = 2.839$ to target the aggregate firm debt to capital ratio (leverage ratio) of 0.563, following the standard practice in the literature. (Karabarbounis and Macnamara (2021) and Ottonello and Winberry (2024)) γ is calibrated to match the wage premium of skilled workers in the 2002 update of the CHIP.[15]

R&D fixed cost. The two parameters that govern the fixed cost of R&D are calibrated to target two moments in the firm-level data: the share of R&D-active firms (6.2 percent in both model and data[16]), and how the share of R&D-active firms depends on their current productivity levels. The later is obtained by regressing the dummy variable of being an R&D-active firm on the firm's productivity level in logs.

R&D ideas production. The effect of R&D investments on productivity is governed by four parameters in the R&D function (the productivity level, B, the span-of-control parameter, μ, the cost share of skilled labor, ψ, and the substitution parameter κ) and the parameter ϕ in the productivity process.

We estimate $\psi = 0.573$ to target the 0.82 percent share of R&D workers in the labor force.[17] The parameter κ has no steady-state implications. Instead, it directly determines

[14] Specifically, we minimize the weighted sum of squared percentage differences between data and model moments. The vector of parameters, Θ, is chosen to minimize the minimum-distance-estimator criterion function

$$f(\Theta) = [\mathbf{m}_{data} - \mathbf{m}_{model}(\Theta)]' \mathbf{W} [\mathbf{m}_{data} - \mathbf{m}_{model}(\Theta)],$$

where $\mathbf{m}_{data}$ and $\mathbf{m}_{model}(\Theta)$ are the vectors of moments in the data and the model, and $\mathbf{W}_{ii} = \text{diag}(\omega_i)$ is a diagonal weighting matrix, where i indexes the i_{th} moment. We place additional weight, ω_i, on the data we view as more important to match and normalize $\sum_{t=1}^{12} \omega_i = 1$.

[15] To calculate the wage premium of skilled workers, we run a Mincer regression using the 2002 CHIP data. Specifically, the wage premium is estimated from the following regression: $log(Y_i) = \beta_0 + \beta_1 COL_i + \beta EXP_i + \beta_e EXP_i^2 + \beta_4 X_i + \epsilon_i$ where Y_i is the labor income, COL_i is a dummy variable of having a college-or-above degree, EXP_i is the years of working experience, and X_i controls for other characteristics.

[16] CECY2004 reports that among 276,474 industrial firms in China, 17,075 firms report positive R&D investments

[17] TO ADD

the flexibility of R&D-active firms to substitute between two input factors when the price or factor-biased productivity of one factor changes. As discussed in Section 2.2, the share of R&D workers in the labor force increased dramatically from 2004 to 2018. We set $\kappa = 0.74$ to match the observed magnitude of this increase (from 0.82% to 3.57%) in the data when all aggregate shocks are fed into the model.[18] The other three parameters are calibrated jointly.

A larger B means a firm produces more R&D ideas for a given amount of inputs. A larger ϕ means a firm's next-period's productivity increases by more for a given amount of R&D ideas. A firm is willing to invest more on R&D if either B or ϕ is larger. Therefore, various combination of the two parameters could yield the same aggregate R&D expenditure to output ratio. We calibrate these two parameters together by targeting the 1.9 percent aggregate R&D intensity in the data and the effect of R&D investments on next-period's productivity, which is estimated to be 0.032 in the ASI. ϕ is estimated to be 0.063, indicating that doubling the amount of R&D ideas would result in an average of 6.3 percent increase in firm productivity in the next period.

Both a greater B and a greater μ yield a higher R&D intensity. However, at given R&D intensity, a larger μ shifts the distribution of R&D expenditure to the most productive firms, which is equivalent to a lower median R&D expenditure-to-output ratio. This is because a larger μ means that the production of R&D displays less decreasing returns to scale. Therefore, the combination of the aggregate and the median R&D expenditure-to-output ratio help us identify both B and μ.[19]

Productivity process. ρ_z and σ_ϵ govern the productivity process of R&D-inactive firms and the exogenous component of the productivity process of R&D firms. We estimate ρ_z by running a regression of a firm's next-period productivity on its current productivity using the 2001-02 ASI data. We estimate σ_ϵ by measuring the standard deviation of productivity growth in the 2001-02 ASI data.

4.3 Model Performance

To evaluate our model performance, we compare the non-targeted moments between our model predictions and the data counterparts, and report them in the lower panel of Table 5.

When calibrating the parameters that govern the fixed cost R&D production, we target

[18] We will describe how we construct these aggregate shocks in Section Sec5:aggshocks

[19] Cao et al. (2024) estimates the skill intensity is 0.22 for incremental innovation and 0.39 for radical innovation in China. In our estimation, the parameter μ and ψ together imply that the elasticity of R&D ideas to the skilled labor input is 0.29, which is in line with the literature.

Table 5: Targeted and Non-targeted Moments: Model vs. Data

Moment	Model	Data
Targeted		
Aggregate asset to output ratio	1.913	1.923
Aggregate leverage ratio	0.560	0.563
College wage premium	1.640	1.640
Aggregate R&D to output ratio %	1.914	1.914
Median R&D to output ratio %	2.084	2.062
Share of R&D workers in labor force %	0.818	0.818
Δ Share of R&D workers (18 vs. 04) %	2.758	2.749
Share of R&D firms	0.062	0.062
Regress R&D dummy on $\log(z)$	0.034	0.034
Regress $\log(z')$ on $\log(z)$	0.700	0.698
SD of productivity growth	0.500	0.499
Regress $\log(z')$ on $\log(i + w_s n_{s,x})$	0.032	0.032
Non-targeted		
Reg. R&D decision on log asset	0.073	0.071
R&D to output 95th percentile	0.252	0.265
R&D to output 90th percentile	0.155	0.163
R&D concentration Top 1%	0.574	0.566
R&D concentration Top 5%	0.820	0.776
R&D concentration Top 10%	0.908	0.863
R&D workers cost share	0.391	0.360

how the decision of becoming R&D active relates to a firm's productivity. To evaluate our model performance, we check if our calibrated model can reproduce the relationship between the decision of becoming R&D active and a firm's net worth. To this end, using model-generated statistics, we regress the dummy variable of being an R&D-active firm on the firm's productivity level (in log) and asset level (in log). The estimated regression coefficient is in our model (0.073) is closely aligned with the one in the data (0.071). This implies that our calibrated model can capture how financial frictions shape firms' R&D decisions in China, as we haven shown in Section 2.3.

In Section 4.2, we intentionally target the share of R&D workers in the labor force. We now check to see whether the cost of R&D workers as a share of total R&D expenditures is similar between the model and the data.

The last row of Table 5 shows that, in the data, 36.0 percent of R&D expenditures is allocated to R&D worker compensation, while the model counterpart is 39.1 percent. The ability of our model to simultaneously match the labor cost share of R&D expenditure and the population share of R&D workers stems from the assumption of heterogeneous labor productivity. Skilled workers have higher average productivity and, consequently, receive higher labor income.

When calibrating the parameters of the R&D production function, we use the aggregate and the median R&D expenditure-to-output ratio. We now check if our calibrated model can reproduce the the distribution of R&D expenditure in the data. We will show later show how the distribution of R&D expenditure can be shaped by aggregate shocks, and how the distribution of R&D expenditure can decide the effect of aggregate shocks on TFP.

To evaluate our model predictions on the distribution of R&D investments, we first compare the most R&D intensive firms in our model and in the data. Specifically, we calculate the R&D-to-output ratio for firms at the top 5 percent and 10 percent of the R&D expenditure-to-output distribution. Table 5 show that our model predictions are consistent with the data.

Next, we examine whether the concentration of R&D expenditure aligns with the data, which is not a target in the calibration. Our calibrated model reproduces these empirical features well. For example, among all R&D-active firms, the top 1 percent firms that have the largest R&D investments account for 56.6 percent of aggregate R&D investments in the data and 57.4 percent in the model.

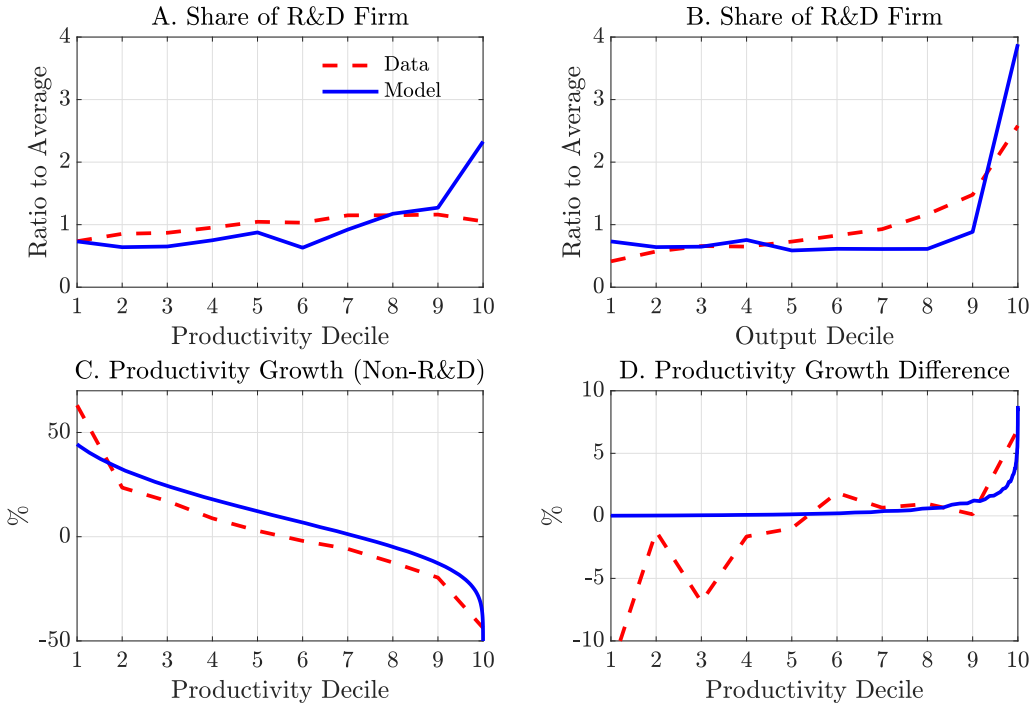


Figure 2: Model Performance

When calibrating the fixed cost of R&D, we target the share of R&D in all firms and the estimated linear relation ship between becoming R&D active and the productivity level. To examine our model performance, we check the share of R&D firms across all productivity

and output levels. Figure 2 show that while the data show more productive firms are more likely to invest in R&D, this positive relationship is very moderate (Panel A). In contrast, the positive correlation between the share of R&D-active firms and output levels are more pronounced, especially at the top end of the output distribution (Panel B). Our model can reproduce both of these features. The reason is that in our model, both productivity levels and asset levels affect firms' R&D decisions. A firm at the top of the productivity distribution may choose not to invest in R&D if it does not have enough net worth. The reason is that with a low asset level, the firm's future capital demand is constrained, and thereby the profitability of productivity enhancement is limited. On the other hand, the output size is determined by both productivity and net worth. Therefore, our model reproduces the empirical pattern of a more significant relationship between the share of R&D firms and output sizes, and less significant relationship between the share of R&D firms and productivity levels.

Lastly, we check if our calibration of the productivity process can reproduce the productivity growth patterns in the data. First, we plot the empirical and the model-predicted productivity growth for non-R&D firms across the productivity distribution. (Figure 2 Panel C) Our model predictions closely match the data pattern that higher productivity firms have a lower productivity growth, implying a large discount factor (ρ) in the productivity process. Second, we plot the growth difference between R&D firms and non-R&D firms across the productivity distribution. Our model predicts that at the middle and lower end of the productivity distribution, holding current productivity level fixed, R&D and non-R&D firms have similar productivity growth rate. In contrast, at the upper end of the productivity distribution, R&D firms have a much faster productivity growth than non-R&D firms. This is because under the calibrated productivity process, more productive firms have a faster depreciation of productivity, and therefore they invest more in R&D to improve their future productivity.

5 Quantitative Analysis

We now turn to quantify the impact of the rise in R&D expenditure on TFP as well as the role of the college expansion program. We make a steady-state comparison between the Chinese economy in 2004 and 2018. We first construct the aggregate shocks that drive the increase in R&D intensity during this period. (Section 5.1) Next, we analyze the aggregate and the distribution effects of the combination of these aggregate shocks. (Section 5.2) In particular, we isolate the impact of the skilled labor supply shock by simulating a counterfactual economy without the college expansion policy. Lastly, we predict the long-run effect of the college expansion program by simulating counterfactual economies with continued

increase in the share of skilled labor. (Section 5.3)

5.1 Constructing Aggregate Shocks

We model three aggregate shocks that drive the increase in R&D intensity from 2004 to 2018. The first aggregate shock is the exogenous increase in the share of skilled labor, which is calibrated externally using Chinese Population Census data. As reported by the 2020 Chinese Population Census, the share of skilled labor—measured as the fraction of urban workers aged 23 to 55 holding a bachelor's degree or higher—has increased from 4 percent in 2004 to 14 percent in 2018. As we model a Gamma distribution of labor productivity, the expansion of college admission indicates that the cutoff labor productivity is shift to the left. (See Figure 6 in Appendix B.) As individuals with lower labor productivity enroll in colleges, the productivity cutoff of workers holding a college degree becomes 35 percent higher than the average.

As the supply of skilled labor increases and the supply of unskilled labor decreases, ceteris paribus, the wage premium for skilled labor should decrease. To explain the relatively stable wage premium in the data,[20] we feed in a skill-biased technology shock, S_t, which changes from 1 (normalization) to 2.25.

Table 6: Calibrating the Aggregate Shocks

Parameter		2004	2018	Target	Model	Data
Share of skilled workers	N_t^s	0.04	0.14	Share of college-educated labor	0.14	0.14
Skill-biased tech. change	S_t	1.00	2.25	College wage premium	1.64	1.64
Sector-biased tech. change	B_t	0.93	3.10	Aggregate R&D to output ratio %	4.22	4.22

Non-Target	Model	Data
Δ patent per R&D worker %	128	144

We construct the third shock to capture all the residual factors that explain the expansion of the R&D sector relative to the production sector, and call it the R&D-sector specific technology shock, B_t. We calibrate this shock to target the change in the aggregate R&D-to-output ratio, which increases from 1.91 percent to 4.22 percent. This shock can be regarded as an increase in the productivity of R&D, reflecting Chinese firms have become more innovative in advancing technology frontiers. It can also be regarded as improvement in the administrative and legal environment that protects intellectual property, making firms more capable of secure the profits of their R&D efforts.

We validate our construction of the aggregate shocks by comparing our model prediction on the *ideas* per researcher with the data counterpart of granted patent per R&D worker.

[20] Add a footnote explain the wage premium is fixed at what level.

Our calibrated shocks generate a 128 percent increase in *ideas* produced per R&D worker, and the data show a 144 percent increase.[21]

5.2 Quantifying the Combined Effect of Aggregate Shocks

In this section, we examine the aggregate and distributional effects of the aggregate shocks. In our analysis, we adjust the equilibrium wages of skilled and unskilled workers to ensure that both labor markets clear. Our main goal is to quantify how much the combination of these aggregate shocks contribute to the surge in R&D expenditure and how much the surge in R&D expenditure contribute to the increase in aggregate TFP.

Table 7: Quantifying the Impact of Aggregate Shocks

	Benchmark		Tech. Change	
	w/o ΔN_t^s	w/ ΔN_t^s	w/o ΔN_t^s	w/ ΔN_t^s
	(1)	(2)	(3)	(4)
A. R&D input				
Aggregate R&D to output ratio	1.9	2.4	4.0	4.2
Median R&D expenses to output	2.1	2.4	3.7	4.0
Fraction of R&D firms	6.2	8.2	16.4	17.7
R&D concentration Top 1%	57.4	59.6	68.9	70.8
R&D concentration Top 5%	81.9	82.9	87.1	88.0
R&D concentration Top 10%	90.8	91.3	93.2	93.7
Share of R&D workers in skilled labor force %	19.0	31.1	14.8	25.0
B. Productivity (%Δ from bench.)				
Total factor productivity	-	1.6	9.6	12.0
Average productivity	-	0.1	1.2	1.5
TFP Losses	-	-0.4	-2.3	-2.9
C. Financial metrics (%Δ from bench.)				
Manufacturing output	-	-1.1	24.3	27.4
After-tax profit	-	-2.2	10.4	9.8
Net worth	-	-1.4	14.2	14.1
D. Skill price (%Δ from bench.)				
College wage	-	-46.5	157.7	37.9
Non-college wage	-	13.2	18.9	37.9

Note: In Panel A, we report the absolute values in each case. All statistics shown in Panel B-D display the changes from the benchmark steady state. Case (1) display the moments from the benchmark economy. In case (2), we feed in the skilled labor supply shock but fix B_t at the benchmark economy. In case (4), we feed in both aggregate shocks described in Section 5.1. In case (3), we keep the skilled labor supply fixed at the benchmark economy but only consider the impact of B_t shock. In all experiments, we solve for the equilibrium skill prices.

Table 7 presents the effects of the aggregate shocks on R&D input, productivity, financial

[21] Cao et al. (2024) reports that the number of eventually granted patents per researcher in China increased from 0.14 in 2004 to 0.27 in 2014, using a different data source. This finding is consistent with our calculations, although data beyond 2014 is unavailable.

metrics, and wage for skilled labor. The difference between columns (1) and (4) shows the combined effect of all three aggregate shocks.

R&D input. The aggregate R&D to output ratio is our target when calibrating the sector-biased technology shock, B_t. The model predicted median R&D expenditure to output ratio increased by a similar size. The fraction of R&D firms increased from 6.2 percent in the initial steady state to 17.7 percent in the final steady state, whereas in the data, the fraction of R&D firms increased to 28 percent.

The reason that our model underpredict the increase in the share of R&D firms is that we do not model any fiscal or industrial policies implemented by the Chinese government to stimulate firms' R&D investment, such as tax deductions and exemptions. Studies have shown that firms have benefited from these policies by over-reporting their R&D expenditure. (Chen et al. (2021)) Therefore, both the officially released R&D intensity and the share of R&D firms might be overstated, and particularly the latter. This is because R&D investment is heavily concentrated in large firms and large private enterprises, which are subject to more stringent tax regulations. In contrast, small and medium-sized enterprises (SMEs) are more likely to inflate their R&D expenditures in order to qualify as "high-tech enterprises" and gain tax benefits.

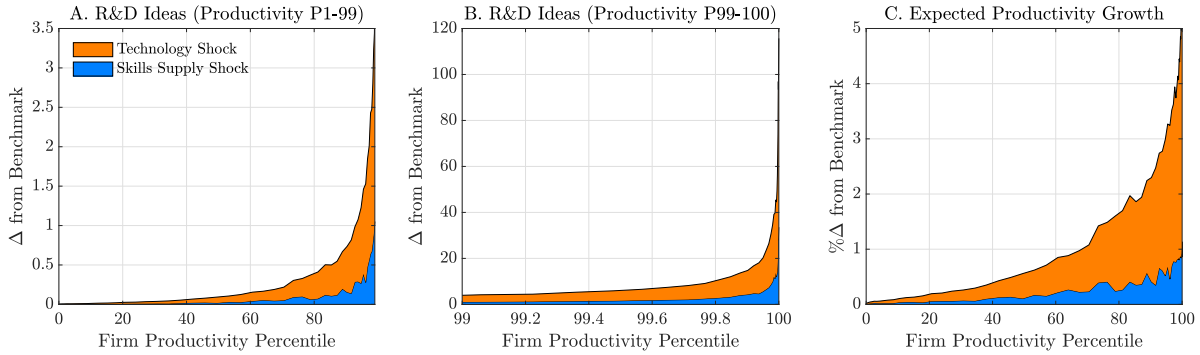


Figure 3: Decomposing the Impacts of Aggregate Shocks

The combined effect of the aggregate shocks make R&D expenditure more concentrated at the top, especially at the top 1 percent. The share of R&D investment from the top 1 percent of the largest innovators increases by 13 percentage points (from 57 percent to 70 percent). (Table 7) In addition, we plot the change in R&D output from the initial steady state across the productivity distribution. As shown in Figures 3.A and 3.B, the aggregate shocks makes more productive firms, particularly at the top 1 percent, increase R&D expenditure by a larger amount than the rest of the firms.

Productivity growth. As firms invest more in R&D following the aggregate shocks, productivity improves. The combination of the three aggregate shocks results in a 12.0 percent increase in total factor productivity (TFP), much greater than the increase in average productivity, which is only 1.5 percent. The reason is that after the aggregate shocks, firms that are already at the top of the productivity distribution invest more on R&D than the rest, and thereby become more productive. As illustrated in Figure 3.C, the expected productivity growth of the most productive firms (top 1%) increases by approximately 5 percentage points, while firms up to the third quartile increase by less than 1 percentage point. As firms at the top of the productivity distribution use a larger amount of production factors (labor and capital) than the rest of the firms do, the productivity growth of highly productive firms has a disproportionately large impact on economy-wide TFP. Therefore, the increase in TFP greatly exceeds the increase in average productivity.[22]

Allocative efficiency Allocative efficiency improves due to the aggregate shocks. The TFP losses[23] due to the presence of financial frictions have decreased by 3.4 percent. This seems counter-intuitive at first glance, because as firms' productivity increases, their optimal capital demand should increase, and this would imply a tighter constraint and thereby greater capital misallocation, which is equivalent to a greater TFP loss all else being equal.

The primary reason for the improvement in allocative efficiency is the increase in aggregate firm net worth, which rose by 14.1 percent. This increase in firms' net worth is driven by two factors. First, we find that firms' after-tax profits increased by 9.8 percent, allowing them to allocate more resources towards accumulating net worth, thereby helping them to overcome the borrowing constraint. Second, firms also expect to them to invest more in R&D and hence a higher productivity in the future, due to the combined effect of all aggregate shocks. Therefore, firm has the incentive to accumulate net worth to self finance a higher capital demand in the future.

Wages The labor supply shock alone increases the equilibrium wage for skilled workers and decreases the equilibrium wage for non-skilled workers. However, both the skilled biased and the sector biased technology changes increase the demand for skilled labor. The combined

[22] The average productivity is calculated as $\int z d\mu$, where μ denotes the distribution of all firms. The (measured) total factor productivity is calculated as $\text{TFP} = Y / \left(AK^{\alpha} [\gamma (N_{s,y})^{\sigma} + (1-\gamma) N_u^{\sigma}]^{\frac{\theta}{\sigma}} \right)$, where the upper-case letters denote the aggregate variables.

[23] To quantify TFP losses, we apply the TFP accounting framework developed in Karabarbounis and Macnamara (2021). We first calculate an efficiency level of TFP, which can be computed as $\text{TFP}_e = (\int z^{\frac{1}{1-\alpha-\theta}} d\mu)^{1-\alpha-\theta}$. The TFP loss is defined as the percentage difference between the efficient and measured level of TFP: $\text{TFP losses} = (\text{TFP}_e / \text{TFP} - 1) \times 100$.

effect leads to a larger wage increase for skilled worker, and a slightly widened wage gap between skilled and unskilled workers.

5.3 The Effect of the Rising Share of Skilled Labor

In this section, we first isolate the actual effect of the labor supply shock from 2004 to 2018 that was caused by the college expansion program. Next, we analyze a hypothetical scenario in which the share of skilled labor continues to increase in the future.

5.3.1 Isolating the effect of labor supply shock

The effect of the rising share of skilled labor without other shocks can be seen by comparing columns (1) and (2) in Table 7. The effect of the rising share of skilled labor conditional on the two technological shocks can be seen by comparing column (3) and (4) of Table 7.

The most direct effect of the labor supply shock shows up on the wage premium of skilled workers. Without an increase in skill supply, skilled labor wages would rise significantly as both of the technological shocks drive up the demand for skilled workers, and thereby increasing the cost of innovation. The blue (lower) areas depicted in Figures 3.A and 3.B illustrate the effects of rising skilled labor supply on the production of R&D ideas across firms with varying productivity levels. We observe that the lower costs of innovation, resulting from reduced skilled wages, disproportionately enhance the R&D ideas produced by high-productivity firms. Consequently, when we eliminate the skill supply shock, the degree of the concentration of R&D ideas declines. As shown in Figure 3.C, the expected productivity growth for highly productive firms would decrease by approximately 1 percentage point, which is significantly greater than the decline for less productive firms.

We found that, in absence of rising skilled labor supply, the growth in TFP would decline by one-fourth (3.1 percentage points), the growth in manufacturing output would drop by one-fifth (3.4 percentage points), the increase in firm net worth would decline by one-sixth (1.8 percentage points), and the rise in after-tax profits would also fall by one-sixth (1.6 percentage points). Moreover, by comparing columns (1) and (2) of Table 7, we can quantify the impact of skill supply shocks when there are no research productivity shocks. When feeding the same skill supply shocks into the model, we find that their effects on the aforementioned variables are significantly reduced.

In summary, we show that the increase in skill supply is complementary to the rise in research productivity, and both of which promotes R&D investment on the intensive and extensive margins. Lower prices for skilled labor particularly incentivize high-productivity firms to invest more in innovation, leading to substantial enhancements in aggregate produc-

tivity. Furthermore, accounting for firm heterogeneity is essential for estimating the impacts of research productivity and skill supply shocks, as we observe that the R&D choices of the most productive firms (top 1%) have a disproportionally large effect on TFP.

5.3.2 Predicting the long run impact

What will happen if the college expansion program continues and more unskilled workers turn into skilled labor in the future? To study this question, we perform the following exercise: we hold the skill biased technology shock and the sector biased technology shock at their 2018 level, and vary the share of skilled workers from 4 percent to 50 percent. We adjust the equilibrium wages of skilled and unskilled workers to clear both labor markets as we change the share of skilled workers. Figure 4 shows the results.

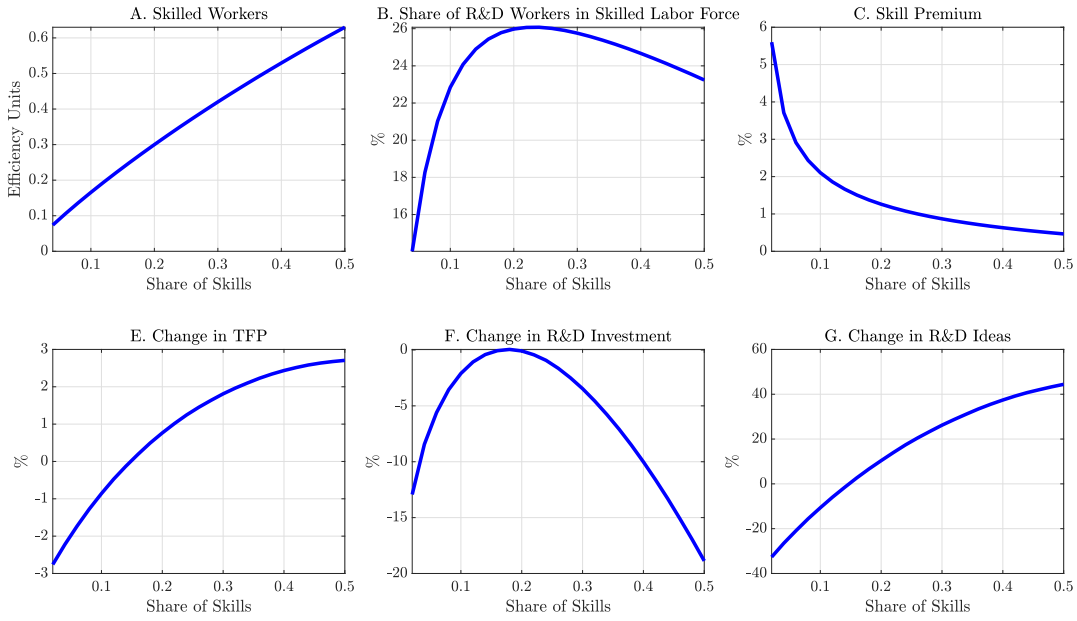


Figure 4: R&D Activities under Counterfactual Share of Skills

Panel (A) of Figure 4 shows the associated increase in the efficiency unit of skilled labor. Due to the influx of skilled workers, their wage premium decreases. (Panel C) Aggregate TFP increases at first, but the rate of increase slows down. (Panel E). The reason is that as unskilled labor becomes more expensive, firms' after-tax profit starts to decrease. Therefore, R&D becomes less affordable. In addition, companies have less incentive to invest in R&D, when expecting a decreasing return of R&D in the future.

The total number of R&D workers increases at a diminishing rate, which also reflects in the change in R&D output (Panel G). Aggregate R&D input peaks when the share of skilled

labor is around 18 percent (Panel F). The reason that R&D investment has a humped shape is that as the share of skilled workers increases, the unit price of an R&D worker becomes cheaper but at the same time firms hires more R&D workers. The former dominates the latter when the share of skilled workers exceeds 18 percent.

6 Conclusion

This paper examines the extent to which the increasing number of college graduates contributes to the rapid growth of firms' R&D investments and productivity in China from 2004 to 2018. We first provide motivating evidence on the relationship between the rising share of skilled labor and the rapid growth of R&D investment. We then construct a heterogeneous firms model in which firms make endogenous R&D decisions under financial frictions and allocate skilled labor between the production and R&D sectors. We structurally estimate the model using firm-level data on the level and distribution of R&D, as well as macro-level data on skill prices and their sectoral allocation.

To generate the observed growth in R&D investments, we construct three aggregate shocks: an R&D-sector-biased technological shock, a skill-biased technological shock, and a skilled labor supply shock, all of which are disciplined by the data. Our model suggests that the combination of these shocks generates a 12 percent increase in TFP, with the skilled labor supply shock accounting for approximately one-fifth of the model-predicted TFP improvement. Our model also implies that continuing to expand college capacity would have a large but diminishing effect on firm productivity, shedding light on the optimal design of education and innovation policies.

References

Aghion, Philippe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen. 2017. "The social origins of inventors." Tech. rep., National Bureau of Economic Research.

Aghion, Philippe, Leah Boustan, Caroline Hoxby, and Jerome Vandenbussche. 2009. "The causal impact of education on economic growth: evidence from US." *Brookings papers on economic activity* 1 (1):1–73.

Akcigit, Ufuk, Jeremy G Pearce, and Marta Prato. 2020. "Tapping into talent: Coupling education and innovation policies for economic growth." Tech. rep., National Bureau of Economic Research.

Bai, Chong-En, Chang-Tai Hsieh, and Yingyi Qian. 2006. "The return to capital in China." Tech. rep., National Bureau of Economic Research.

Bloom, Nicholas, John Van Reenen, and Heidi Williams. 2019. "A toolkit of policies to promote innovation." *Journal of economic perspectives* 33 (3):163–184.

Brandt, Loren, Trevor Tombe, and Xiaodong Zhu. 2013. "Factor market distortions across time, space and sectors in China." *Review of Economic Dynamics* 16 (1):39–58. URL https://doi.org/10.1016/j.red.2012.10.002.

Brown, James R, Gustav Martinsson, and Bruce C Petersen. 2012. "Do financing constraints matter for R&D?" *European Economic Review* 56 (8):1512–1529.

Buera, Francisco J and Roberto N Fattal Jaef. 2018. "The dynamics of development: innovation and reallocation." *World Bank Policy Research Working Paper* (8505).

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

Che, Yi and Lei Zhang. 2018. "Human capital, technology adoption and firm performance: Impacts of China's higher education expansion in the late 1990s." *The Economic Journal* 128 (614):2282–2320.

Chen, Zhao, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi Xu. 2021. "Notching R&D investment with corporate income tax cuts in China." *American Economic Review* 111 (7):2065–2100.

Dai, Xiaoyong and Fang Wang. 2019. "Does the high-and new-technology enterprise program promote innovative performance? Evidence from Chinese firms." *China Economic Review* 57:101330.

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.

Gu, Shijun and Chengcheng Jia. 2022. "Firm dynamics and SOE transformation during China's Economic Reform." *Journal of International Money and Finance* 127:102672.

Hall, Bronwyn H, Jacques Mairesse, and Pierre Mohnen. 2010. "Measuring the Returns to R&D." In *Handbook of the Economics of Innovation*, vol. 2. Elsevier, 1033–1082.

Hsieh, Chang-Tai and Peter J. Klenow. 2009. "Misallocation and manufacturing TFP in China and India." *The Quarterly Journal of Economics* 124 (4):1403–1448. URL https://doi.org/10.1162/qjec.2009.124.4.1403.

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* (1):295–366. URL https://doi.org/10.1353/eca.2016.0005.

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.

Karabarbounis, Marios and Patrick Macnamara. 2021. "Misallocation and financial frictions: The role of long-term financing." *Review of Economic Dynamics* 40:44–63.

König, Michael, Zheng Michael Song, Kjetil Storesletten, and Fabrizio Zilibotti. 2020. "From imitation to innovation: Where is all that Chinese R&D going?" Tech. rep., National Bureau of Economic Research.

Li, Ben. 2010. "Multinational production and choice of technologies: New evidence on skill-biased technological change from China." *Economics Letters* 108 (2):181–183.

Li, Xibao. 2012. "Behind the recent surge of Chinese patenting: An institutional view." *Research policy* 41 (1):236–249.

Ma, Xiao. 2024. "College expansion, trade, and innovation: Evidence from china." *International Economic Review* 65 (1):315–351.

Mezzanotti, Filippo and Timothy Simcoe. 2023. "Research and/or Development? Financial Frictions and Innovation Investment." Tech. rep., National Bureau of Economic Research.

Ottonello, Pablo and Thomas Winberry. 2024. "Capital, Ideas, and the Costs of Financial Frictions." Tech. rep., National Bureau of Economic Research.

Peia, Oana and Davide Romelli. 2022. "Did financial frictions stifle R&D investment in Europe during the great recession?" *Journal of International Money and Finance* 120:102263.

Romer, Paul M. 2000. "Should the government subsidize supply or demand in the market for scientists and engineers?" *Innovation policy and the economy* 1:221–252.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing like China." *American Economic Review* 101 (1):196–233. URL https://doi.org/10.1257/aer.101.1.196.

Toivanen, Otto and Lotta Väänänen. 2016. "Education and invention." *Review of Economics and Statistics* 98 (2):382–396.

Wei, Shang-Jin, Zhuan Xie, and Xiaobo Zhang. 2017. "From" Made in China" to" Innovated in China": Necessity, prospect, and challenges." *Journal of Economic Perspectives* 31 (1):49–70.

Appendices

A Data

A.1 Education Attainments across Age Groups

Table A1 summarizes the share of population with college and post-graduate degrees for different age groups in 2000, 2010, and 2020. The most significant change is shown in the age group of 25 to 29 years old, as regular higher education degrees are mostly obtained in one's 20s. From 2000 to 2020, the share of population with college or higher degrees increases from around 4 percent to around 25 percent. The changes for elder group of the population from 2000 to 2020 are mostly due to the completion of part-time degree programs, which is arguably of less quality. We address this issue later in the quantitative analysis.

Table A1: Education Attainments across Age Groups

	College and above %			Post-graduate %		
Age	2000	2010	2020	2000	2010	2020
25-29	4.3	15.6	25.6	0.4	2.1	3.6
30-34	4.3	10.7	20.8	0.4	1.4	2.6
35-39	4.1	7.3	18.9	0.4	0.8	2.5
40-44	2.3	5.9	12.4	0.2	0.5	1.6
45-49	2.0	5.0	7.9	0.1	0.5	0.8
50-54	2.1	2.8	5.6	0.1	0.2	0.5
55-59	3.8	2.2	4.5	0.1	0.1	0.5
All	3.5	7.5	13.9	0.3	0.9	1.8

Source: Authors' tabulation from China's Population Census.

A.2 R&D Statistics in Chinese Statistical Yearbook

Figure 5 plots aggregate statistics from the Chinese Statistical Yearbook 2000 - 2018, which provides annual data on aggregate R&D statistics for all above-the-scale companies.[24] Panel (A) shows the substantial increase in R&D to GDP ratio in China. To gain insights into the drivers of the increase in aggregate R&D intensity, we calculate the changes at the extensive and at the intensive margin. The extensive margin are measured as the fraction of firms making positive R&D expenditure (panel (B)). The fraction of firms performing R&D was less than 10 percent in 2004, and increased to over 25 percent by 2019. Interestingly, most of

[24]The scale was 5 million RMB in annual revenue before 2011, and it was increased to 20 million RMB after 2011.

the increase occurred after 2011. Changes at the intensive margin is shown as the increase in average R&D investment (panel (C)), which is computed as the ratio of total firm R&D expense over the number of R&D firms. The average R&D expenditure increased from 6 million to 12 million between 2004 and 2011, but then decreased to about 8 percent by 2019. Panel (B) and panel (C) show that the increase in R&D expenditure happened both at the extensive margin and at the intensive margin. In addition, panel (D) shows the fraction of R&D workers in all payroll employment increased from about 0.5 percent in 2000 to more than 3.5 percent in 2019.[25] The size of increase in R&D workers is similar to that in the aggregate R&D intensity.

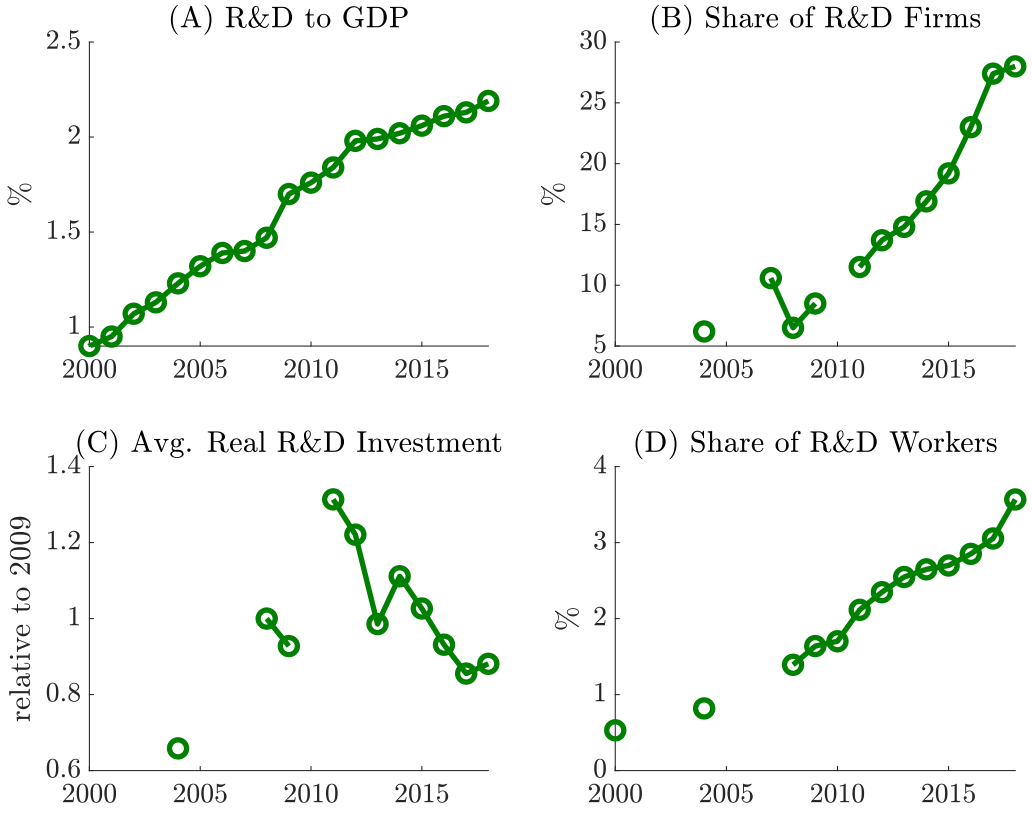


Figure 5: Aggregate R&D Statistics in China

Source: Authors' tabulation from Chinese Statistical Year Book.

We use the 2019 Yearbook as an example to explain how to find and calculate the aggregate statistics of firm R&D reported above. Most information is reported in Chapter 20-4 (Summary of R&D Activities for Above-the-Scale Companies).

The total amount of R&D workers is measured by full-time equivalents (*R&D quan shi*

[25] See Appendix A.2 for the definition of R&D workers.

dang liang), which accounts for both full-time and part-time employees involved in R&D. To get the fraction of R&D workers in the total workforce, we divide *R&D quan shi dang liang* by the aggregate payroll employment, where the aggregate payroll employment can be found in Chapter 13-3 (Major Indicators for Above-the-Scale Companies).

The share of R&D firms for all above-the-scale companies can be directly found in Chapter 20-4. To calculate the average of R&D investments, we first calculate the average R&D investments in nominal terms as the ratio of total R&D expenses over the number of R&D firms, both of which can be found in Chapter 20-4. Next, we adjust this nominal value by CPI using 2004 as the base year.

B Labor Productivity

This figure plots the calibrated Γ distribution of labor productivity, and the productivity cutoff for college admission in 2004 and in 2018.

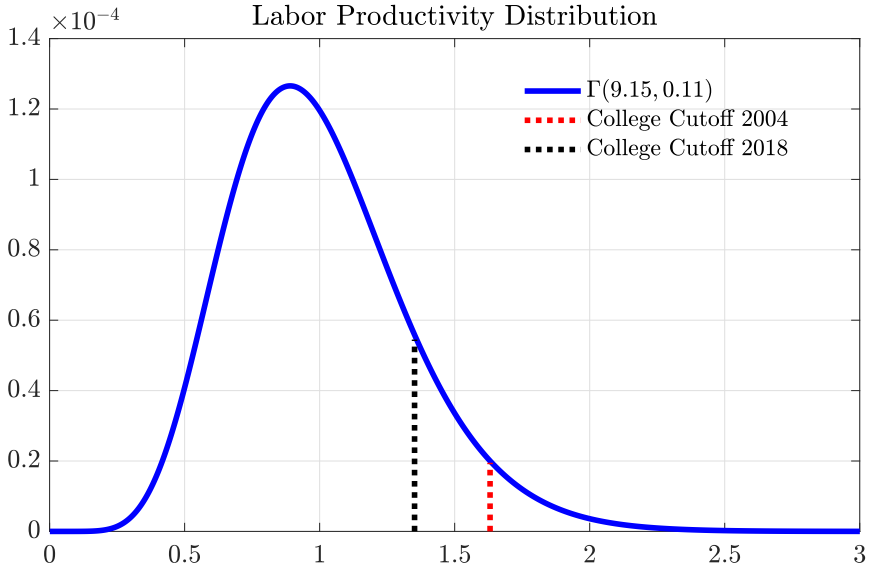


Figure 6: Labor Productivity Distribution and Cutoff for Skilled Labor