Notching R&D Investment with Corporate Income Tax Cuts in China

By Zhao Chen, Zhikuo Liu, Juan Carlos Suárez Serrato and Daniel Yi Xu*

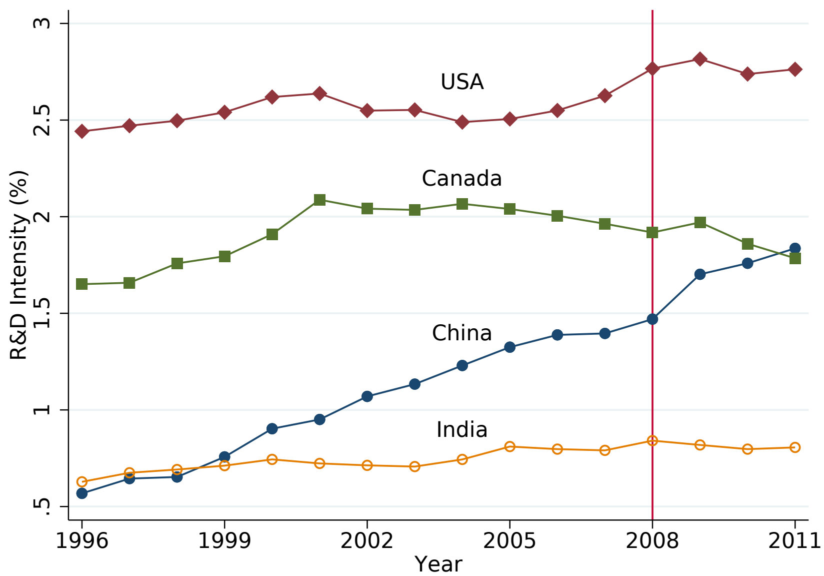
We study a Chinese policy that awards substantial tax cuts to firms with R&D investment over a threshold or "notch." Quasi-experimental variation and administrative tax data show a significant increase in reported R&D that is partly driven by firms relabeling expenses as R&D. Structural estimates show relabeling accounts for 24.2% of reported R&D and that doubling R&D would increase productivity by 9%. Policy simulations show that firm selection and relabeling determine the cost-effectiveness of stimulating R&D, that notch-based policies are more effective than tax credits when relabeling is prevalent, and that modest spillovers justify the program from a welfare perspective.
JEL: D24, O30, H25, H26
Keywords: R&D, China, relabeling, corporate taxes

The belief that innovation is crucial for economic growth inspires governments around the world to encourage R&D investment through tax incentives. While these incentives are meant to stimulate real R&D expenditures, firms can also respond by relabeling other expenses as R&D. Relabeling raises important questions about how tax incentives affect productivity growth. To what extent is reported R&D real or relabeled? How does relabeling affect estimates of the productivity effects of R&D? How should governments incentivize R&D while taking relabeling behavior into account?

We answer these questions using a novel administrative dataset of corporate tax returns of Chinese firms covering a period of sharp and changing tax incentives.

* Chen: China Center for Economic Studies, Fudan University, 600 Guoquan Rd, Shanghai, China, zhaochen@fudan.edu.cn. Liu: China Center for Economic Studies, Fudan University, 600 Guoquan Rd, Shanghai, China, lzhikuo@163.com. Suárez Serrato: Duke University & NBER, 213 Social Sciences, Durham, NC, 27705, jc@jcsuarez.com. Xu: Duke University & NBER, 213 Social Sciences, Durham, NC, 27705, daniel.xu@duke.edu. We are very grateful for guidance from Liran Einav (co-editor) and four anonymous referees, as well as for comments from Manuel Adelino, Ashish Arora, Pierre Bachas, Michael Best, Lysle Boller, Wesley Cohen, Dhammika Dharmapala, Michael Devereux, Rebecca Diamond, Bronwyn Hall, Jim Hines, Damon Jones, Henrik Kleven, Ben Lockwood, Jill Popadak, Jim Poterba, Nirupama Rao, Mark Roberts, Joel Slemrod, Stefanie Stantcheva, John Van Reenen, Mazhar Waseem, Shang-Jin Wei, Daniel Wilson, and Eric Zwick as well as seminar participants at ASSA, CEPR, Chicago (Booth & Becker Friedman), Cornell, Drexel, Duke (Fuqua & Econ), FRB Minneapolis, FRB Philadelphia, FRB San Francisco, Hitsotsubashi University, Hong Kong University, IIPF, LSE, MIT, National School of Development (PKU), NBER (Dev, Ent, China, CCER), NTA, Oxford, Penn State, Stanford (SIEPR & Econ), Toronto, UCLA, UCSD (Econ & Bunching Conference), University of Alberta, University of Maryland, University of Melbourne, Warwick University, University of Wisconsin, and ZEW MaTax. Dan Garrett, Yuxuan He, Xian Jiang, and Matt Panhans provided outstanding research assistance. This project is funded by NSF grant #17300024, and Suárez Serrato is also grateful for funding from the Kauffman Foundation. All errors remain our own.

Figure 1. : Cross-Country Comparison: R&D as Share of GDP



Note: This figure plots the aggregate R&D Intensity, i.e., R&D expenditure as a share of GDP, in the private sector for China, Canada, India, and the US. Chinese R&D intensity started in 1996 at 0.5%, a similar level to India. It increased dramatically, by more than threefold, to above 1.5% in 2011, on par with Canada. The R&D intensity of the US remained stable at 2.5% during the same period. The red line marks the year of the tax reform. Source: World Bank (1995-2011).

China is the perfect laboratory to study fiscal incentives for R&D. Figure 1 shows that China has experienced explosive growth in R&D investment even relative to its rapid GDP expansion. In addition, the government is focused on fostering technology-intensive industries as a source of future economic growth.

The tax incentive that we study—China's InnoCom program—provides substantial corporate income tax cuts to firms that report R&D investment over a given threshold or "notch." Before 2008, firms with an R&D intensity (R&D investment over revenue) above 5% could qualify for a special high-tech-firm status that was accompanied by a lower average tax rate of 15%—a large reduction from the statutory rate of 33%. After 2008, the government established three thresholds at 3%, 4%, and 6% for firms in different size categories. By changing average tax rates, as opposed to marginal incentives, the program generates very large incentives for firms to increase reported R&D. Section I describes this fiscal incentive and discusses the potential for relabeling of R&D.

We begin our analysis in Section II by showing graphically that tax notches have significant effects on the distribution of reported R&D intensity. We show that a large number of firms choose to locate at tax notches and that introducing the tax cut led to a large increase in R&D investment. Using a group of firms that were unaffected prior to 2008, we show that the bunching patterns are driven by the tax incentive and are not a spurious feature of the data. We quantify the percentage increase in R&D investment that is due to the tax notch using a bunching estimator (Kleven and Waseem, 2013). We find large increases in R&D investment of 25% for large firms, 17% for medium firms, and 10% for small firms in 2011.

We then analyze relabeling responses by exploiting the fact that, under Chinese Accounting Standards, R&D is reported as a subcategory of administrative expenses. Using our detailed tax data to separate R&D from other administrative expenses, we provide graphical evidence that firms may relabel non-R&D expenses as R&D to qualify for the tax cut. Specifically, we document that non-R&D expenses drop significantly at the R&D notches, which suggests that the increase in reported R&D is partly driven by relabeling of non-R&D expenses. We also study other forms of manipulation, including relabeling of other expenses as well as retiming of sales, and we do not find evidence of manipulation along these margins.

We develop a model of R&D investment and relabeling in Section III. Firms' decisions to invest or relabel depend on tax incentives, the effect of R&D on productivity, and the costs of relabeling, as well as on heterogeneous productivity and adjustment costs. The model shows that the InnoCom program incentivizes firms that would otherwise be at the low end of the R&D intensity distribution to bunch at the notch. Firms in the model can bunch either by increasing real R&D investments or by relabeling non-R&D expenses. The optimal real R&D investment decision and relabeling strategy depends on the relative strength of the cost of relabeling and the productivity elasticity of R&D. Our model allows for rich patterns of firm heterogeneity. First, firms face heterogeneous adjustment costs of investing in R&D, which rationalizes a highly dispersed R&D intensity distribution. Second, the model allows for random certification costs that account for non-R&D requirements of the InnoCom program and that explain why firms close to the notch may not participate in the program. Overall, the model captures competing mechanisms for bunching—real R&D vs. relabeling—and is rich enough to fit the main features of the data.

We estimate the model using a simulated method of moments approach in Section IV. The main parameters of the model—the productivity elasticity of R&D and the cost of relabeling—are informed by the bunching response in reported R&D, the relabeling response at the notch, and the joint distribution of R&D and productivity. By specifying the distributions of fixed and adjustment costs, the model also characterizes how firms select into the program, which allows us to study the effects of alternative policies. We estimate that, on average, 24.2% of the reported R&D investment is due to relabeling and that a 100% increase in real R&D would increase TFP by 9%. Our estimated model fits the data moments very well. The structural estimates are also consistent with reduced-form bunching estimates, which provide a valuable cross-validation of the model. Our results are also robust to a number of checks that ensure that our main conclusions are not artificially driven by the parameterization of the model.

In Section V, we use the estimated model to study how governments can best incentivize R&D in the presence of relabeling. We first study the effects of changing the size of the tax cut and the location of the notch. Policies with a larger tax cut and those with a notch at a lower R&D intensity select firms with lower productivity, higher adjustment costs, and greater motives to relabel. Firm selection into the program plays a crucial role in determining the economic effects of the program and the fiscal cost of incentivizing real R&D.

As a second use of the model, we compare the fiscal effectiveness of the InnoCom program to that of a linear tax credit. In a setting where firms have low incentives to relabel, a linear tax credit is more effective at stimulating R&D. However, a notch may be more effective than a linear tax credit when firms can relabel. The key intuition is that, under a linear tax credit, the government's monitoring efforts are spread across many firms, which lowers firms' relabeling costs. By focusing monitoring efforts on fewer firms, an InnoCom-style program can raise the cost of relabeling and incentivize real R&D at a lower fiscal cost. Governments may thus prefer to deviate from standard incentives in the presence of relabeling (e.g., Best et al., 2015).

Fiscal incentives for R&D are often motivated by the possibility that firms may under-invest in R&D in the presence of knowledge spillovers. As a final use of our model, we study the welfare effects of the InnoCom program by extending our empirical model into an equilibrium setting with potential R&D spillovers. In the absence of externalities, the InnoCom program distorts firm behavior and reduces tax revenue, leading to an overall reduction in welfare. We then calculate the magnitude of R&D spillovers that could justify the InnoCom program. The program is welfare neutral when spillovers are such that firm productivity increases by 6.9% in response to a doubling of average R&D investment in the economy. Since the empirical literature often estimates larger spillover effects (e.g., Lucking, Bloom and Van Reenen, 2019), InnoCom-style programs can possibly improve welfare and help alleviate the under-investment in R&D.

Overall, this paper shows that relabeling is an important concern for both understanding empirical facts surrounding R&D and designing policies aimed at encouraging innovation. Relabeling affects the measurement of actual R&D expenses, the contribution of R&D to TFP growth, and how tax incentives link fiscal costs to economic growth. Policies that may otherwise be suboptimal—such as notches—may be more effective at alleviating under-investment in R&D than standard tax credits, especially when such policies target firms with better prospects for technological improvement and limit the potential for relabeling.

This paper is related to a large literature analyzing the effects of tax incentives for R&D investment. Hall and Van Reenen (2000) and Becker (2015) survey this literature. The empirical evidence is concentrated in OECD countries, where micro-level data on firm innovation and tax records have become increasingly available. While earlier work relied on matching and panel data methods, there is an emerging literature that explores the effects of quasi-experimental variation in tax incentives for R&D.[1] This is the first paper to analyze R&D tax incentives in a large emerging economy such as China. It is also one of the first studies to use administrative tax data to study the link between fiscal incentives, R&D investment, and firm-level productivity.

Previous research has long highlighted relabeling as an important challenge to identifying the effects of tax incentives for R&D (Eisner, Albert and Sullivan, 1984; Mansfield and Switzer, 1985). This is a salient issue for policymakers in developed countries (GAO, 2009; Bloom, Van Reenen and Williams, 2019) and is likely a more severe problem in developing economies. We exploit unique firm-level data to jointly model and estimate firms' R&D bunching and relabeling decisions. Our policy simulations also improve our understanding of the effectiveness of different policies when firms may engage in relabeling.

Researchers and policymakers are concerned with the extent of misallocation of innovation resources in China. Wei, Xie and Zhang (2017) show that state-owned firms produce significantly fewer patents per yuan of investment than foreign or private domestic firms. König et al. (2018) argue that R&D investments in mainland China have smaller effects on productivity growth than those in Taiwan. Our results show that the seemingly low return to reported R&D is an artifact of relabeling and that tax incentives for R&D may be more costly in emerging economies where the corporate tax is imperfectly enforced (Cai, Chen and Wang, 2018).

Finally, our paper is related to a recent literature that uses bunching methods to estimate behavioral responses to taxation by analyzing the effects of sharp economic incentives.[2] While most of the literature studies kinks or notches in taxable income, the notch in the InnoCom program targets a particular action: R&D investment. We develop a simulated method of moments estimation approach that is consistent with results from reduced-form bunching estimators. The model clarifies the interpretation of reduced-form estimates, as suggested by Einav, Finkelstein and Schrimpf (2017).[3] Our model quantifies the extent of misreporting, measures the returns to real R&D, and simulates the effects of alternative policies. The model also clarifies how selection and relabeling determine the fiscal effectiveness and the welfare implications of a notch-based policy.[4]

[1] Recent examples include Agrawal, Rosell and Simcoe (2019), Dechezlepretre et al. (2016), Einiö (2014), Guceri and Liu (2019), Akcigit et al. (2018), and Rao (2016).

[2] Kleven (2016) provides a recent survey. While these methods have been used to study a wide range of behaviors, this paper is most related to a smaller literature analyzing firm-level responses (Devereux, Liu and Loretz, 2014; Patel, Seegert and Smith, 2016; Liu et al., 2019; Almunia and Lopez-Rodriguez, 2018; Bachas and Soto, 2019).

[3] Lockwood (2018) also notes that reduced-form effects from bunching on notches are not sufficient to analyze the effects of changes in policy. This result motivates the use of a structural model for policy analysis.

[4] Blinder and Rosen (1985) discuss selection patterns under which notches can be desirable, and Slemrod (2013) discusses administrative costs as a motivation for notches.

I. Fiscal R&D Incentives and the Chinese Corporate Income Tax

China had a relatively stable Enterprise Income Tax (EIT) system from 2000 to 2007. During this period, the EIT ran on a dual-track scheme with a base tax rate of 33% for all domestic-owned enterprises (DOEs) and a preferential rate for foreign-owned enterprises (FOEs) ranging from 15% to 24%. The government implemented a major corporate tax reform in 2008 that eliminated the dual-track system based on domestic/foreign ownership and established a common rate of 25%.[5]

This paper analyzes the InnoCom program, which targets qualifying high-tech enterprises (HTEs) and awards them a flat 15% income tax rate. Since a firm's average tax rate can fall from 33% to 15%, this tax incentive is economically very important and may lead firms to invest in projects with substantial fixed costs. This program is most important for DOEs, including both state- and privately-owned enterprises, as they are not eligible for many other tax breaks.

Table 1 outlines the requirements of the program and how they changed as part of the 2008 reform. A crucial requirement of the program is that firms must have an R&D intensity above a given threshold. The reform changed the threshold from a common R&D intensity of 5%, to a size-dependent threshold with a lower hurdle for medium and large firms, 4% and 3%, respectively, and a larger hurdle of 6% for small firms. This requirement provides a large fiscal incentive to invest above these thresholds, and the reform generates quasi-experimental variation across firms of different size and ownership categories. Notably, because the reform eliminated preferential tax rates for foreign firms, the incentive of FOEs to qualify for the InnoCom program grew after the reform.

In addition to increasing R&D intensity, the InnoCom program requires firms to employ college-educated workers and to sell "high-tech" products. Unlike the R&D intensity requirement, these guidelines—such as which products are classified as high tech—are easily influenced. It is also hard for tax authorities to verify the employment composition of a given firm. While these requirements are not sharp incentives, they increase the cost of participating in the program. Importantly, these costs may even prevent some firms from bunching at the notch despite having an R&D intensity immediately below the notch. To capture this cost of participating in the program, our model in Section III assumes that firms differ by an unobserved fixed cost of certification.

As a final program requirement, firms have to actively apply for the program and undergo a special audit. The reform improved enforcement of the program by changing the certifying agency from the Local Ministry of Science and Technology to a joint effort between the National Ministry of Science and Technology, the Ministry of Finance, and the National Tax Bureau. By focusing enforcement efforts on fewer firms, the InnoCom program increased the cost of relabeling

[5]We discuss details of other preferential tax policies in Appendix A.

Table 1—: **Requirements of the InnoCom Program**

Requirement	**Before 2008**	**After 2008**
R&D Intensity	5%	6% if sales $< 50M$ 4% if $50M <$ sales $< 200M$ 3% if sales $> 200M$
Sales of High Tech Products	60% of total sales	
Workers with College Degree	30% of workforce	
R&D Workers	10% of workforce	
Certifying Agency	Local Ministry of Science & Technology	Ministries of Science & Technology, Finance and National Tax Bureau

Note: Size thresholds in millions of RMB, where 50 M RMB $\approx$ 7.75 M USD and 200 M RMB $\approx$ 30 M USD.

R&D relative to a more standard setting where all firms are able to claim an R&D tax credit.[6]

Potential for Evasion and Relabeling

One concern is that firms' reported R&D investment is contaminated by evasion or relabeling. Relabeling of other expenses as R&D is a significant concern for policymakers (GAO, 2009) and for academics studying the effects of R&D investment (Eisner, Albert and Sullivan, 1984; Mansfield and Switzer, 1985). In our setting, the institutional environment limits some forms of evasion and suggests that the most likely form of relabeling is the miscategorization of administrative expenses as R&D.

The hypothesis that the entirety of the response is due to evasion is likely ruled out by the requirements of the InnoCom certification.[7] A second hypothesis is that firms manipulate their reported R&D intensity by reporting "phantom expenses" or by manipulating sales. China relies on a value-added tax (VAT) system with third-party reporting, and China's State Administration of Tax (SAT) keeps records of transaction invoices between a given firm and its third-party business partners. As in other settings (e.g., Kleven et al., 2011), this type of third-party reporting limits the degree to which firms can completely make up "phantom"

[6]The original government regulations also require that firms operate in a number of selected state-encouraged industries. Due to the breadth and vagueness of these industry definitions, this requirement does not constitute a substantial hurdle. In addition, after the reform, the state authorities further require that firms meet all these criteria in the previous three accounting years or from whenever the firm is registered, in case the firm is less than three years old.

[7]Part of this certification includes an audit of the firm's tax and financial standings. In addition, the Chinese State Administration of Tax, together with the Ministry of Science and Technology, conducts regular auditing of the InnoCom HTE firms.

R&D expenses.

From conversations with the State Administration of Tax as well as with corporate executives, we recognize that the most likely source of manipulation is the miscategorization of expenses. This is a natural channel for relabeling since, in the Chinese Accounting Standard, R&D is categorized under "Administrative Expenses," which includes various other expenses related to general management.[8] Thus, firms may relabel non-R&D administrative expenditures as R&D to over-report their R&D intensity. These types of expenses are easily shifted, and it may be hard to identify relabeling in any given audit. Relabeling may also be a way for firms to reach the R&D intensity threshold when it is hard for them to perfectly forecast their sales. A firm with unexpectedly high sales, for instance, might choose to characterize administrative expenses as R&D to meet the InnoCom requirement for a given year.[9] Our empirical strategy to detect relabeling leverages these institutional features and exploits the detailed cost reporting in our administrative tax data, which contain information on the breakdown of operating expenses and R&D expenses.

II. Descriptive Evidence of Firms' Responses to Tax Notches

We now describe our data and provide evidence that the R&D investment of Chinese manufacturing firms responds to the InnoCom program. We then show that part of this response may be due to relabeling. Specifically, we document stark bunching patterns precisely above the tax notches, and we show that the ratio of administrative expenses to sales drops sharply at the notch. These data patterns motivate our model in Section III and inform the structural estimation in Section IV.

A. *Data and Summary Statistics*

Our main data come from the Chinese State Administration of Tax (SAT, 2008-2011). The SAT is the counterpart of the IRS in China and is in charge of tax collection and auditing. Our data are comprised of administrative enterprise income tax records for the years 2008–2011 (Appendix B discusses our data sources). These panel data include information on firms' total production, sales, inputs, and R&D investment. The detailed cost breakdowns allow us to measure different subcategories of administrative expenses. We use these data to construct residualized measures of firm productivity.[10] The SAT's firm-level records of tax

[8]Examples include administrative worker salaries, business travel expenses, office equipment, etc. While we interpret changes in administrative expenses as relabeling, they may also be consistent with reallocating resources from other expenses toward R&D or more precise accounting of previously under-counted R&D expenses. In Section IV, we explore how this interpretation affects our estimates.

[9]We do not find systematic evidence that firms relabel R&D intensity through other means. In Section II, we show that sales are not manipulated around the R&D thresholds. Similarly, we do not find evidence of manipulation of other expenses.

[10]See Appendix C for details, where we also show that we obtain similar productivity estimates using the method of Ackerberg, Caves and Frazer (2015).

Table 2—: Descriptive Statistics

A. State Administration of Tax Data 2008–2011

	Mean	Std	p25	p50	p75	N
Sales (mil RMB)	118.263	1394.828	2.579	10.608	42.056	1202257
Fixed Asset (mil RMB)	32.912	390.406	0.402	2.089	10.743	1139038
# of Workers	175.402	852.494	17.000	48.000	136.000	1213497
R&D or not	0.081	0.273	0.000	0.000	0.000	1219630
R&D/Sales (%, if>0)	3.560	7.019	0.337	1.544	4.296	98258
Administrative Expense/Sales (%)	9.417	11.886	2.809	5.814	11.103	1171365
TFP	2.058	0.522	1.638	2.007	2.434	1100845

B. Annual Survey of Manufacturing 2006–2007

	Mean	Std	p25	p50	p75	N
Sales (mil RMB)	110.801	1066.080	10.760	23.750	59.513	638668
Fixed Asset (mil RMB)	42.517	701.282	1.630	4.492	13.370	638668
# of Workers	238.379	1170.327	50.000	95.000	200.000	638668
R&D or not	0.102	0.303	0.000	0.000	0.000	638668
R&D/Sales (%, if>0)	1.631	3.184	0.118	0.461	1.736	65267

Notes: Various sources; see Section II.A for details.

payments contain information on tax credits, such as the InnoCom program, as well as other major tax breaks. This allows us to precisely characterize the effective tax rate for individual manufacturing firms. We supplement these data with the Chinese Annual Survey of Manufacturing (ASM) (NBS, 2006-2007), which extends our sample to the years 2006–2007.

Table 2 reports descriptive statistics of the firms in our analysis sample. In panel A, we report summary statistics of our tax data for all surveyed manufacturing firms from 2008 to 2011. Our data are comprised of around 1.2 million observations, with about $300,000$ firms in each year. A total of 8% of the sample reports positive R&D. Among firms with positive R&D, the ratio of R&D to sales, i.e., R&D intensity, is highly dispersed. The 25th, 50th, and 75th percentiles are 0.3%, 1.5%, and 4.3%, respectively. The administrative expense-to-sales ratio, which is a potential margin for relabeling, is close to 5.8% at the median. While our measure of residualized TFP is normalized by construction, the distribution of productivity has a reasonable dispersion with an interquartile range of 0.8 log points. As one might expect, firms with higher R&D intensities also have higher values of TFP. For instance, large firms with R&D intensity below 3% have a (normalized) TFP of -1.5%, while firms with R&D intensity greater than 3% have an average TFP of 2.7%.

Panel B of Table 2 reports summary statistics of Chinese manufacturing firms with R&D activity in the ASM for the years 2006–2007. We have a similar sample size of around 300,000 firms per year. Firms in the ASM sample are noticeably larger than those in the SAT sample, and the difference is more pronounced when we look at lower quartiles (i.e., the 25th percentile) of the distribution of sales, fixed assets, and the number of workers. This is consistent with the fact that the ASM is weighted toward medium and large firms. The fraction of firms with positive R&D is slightly larger than 10%, and R&D intensity ranges from 0.1% to 1.7% at the 25th and 75th percentiles of this sample.

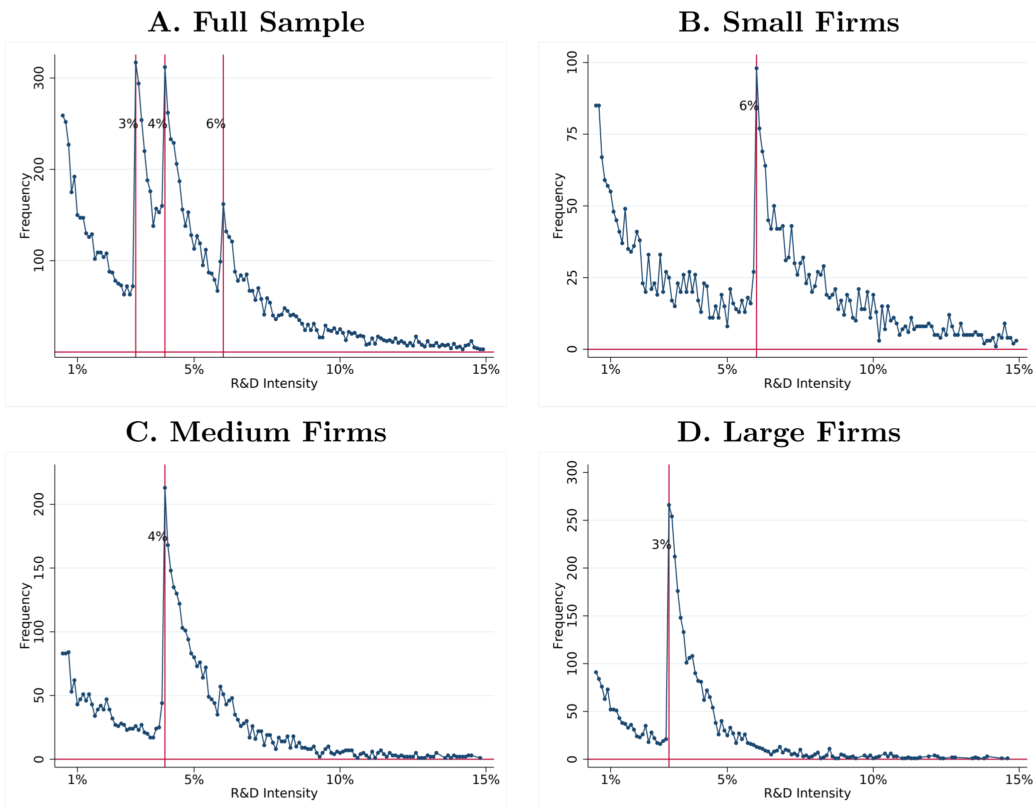
B. Bunching Response

We first analyze data from the post-2008 period since the multiple tax notches based on firm size generate rich variation in R&D bunching patterns. Figure 2 plots the empirical distribution of the R&D intensity of Chinese firms in 2011. We limit our sample to firms with R&D intensity between 0.5% and 15% to focus on firms with non-trivial innovation activities. The first panel in Figure 2 shows the histogram of overall R&D intensity distribution. There are clear bunching patterns at 3%, 4%, and 6% of R&D intensity, corresponding to the three program thresholds. This first panel provides strong prima facie evidence that fiscal incentives provided by the InnoCom program play an important role in firms' R&D investment choices.

To further validate that these R&D bunching patterns are motivated by this specific policy, we plot the histograms of R&D intensity for the three different size categories in the remaining panels of Figure 2. For firms with annual sales below 50 million RMB, we find clear bunching at 6%, and we find no evidence of bunching at other points. Similarly, for firms with annual sales between 50 million and 200 million RMB, we find bunching only at 4%, while for firms with more than 200 million RMB in annual sales, we observe bunching only at 3%. These patterns are consistent with the size-dependent tax incentive in the InnoCom program.[11]

[11]In comparison, Figure A.1 plots the empirical distribution of R&D intensity in the ASM for 2006–2007. The InnoCom tax incentive was not size-dependent before 2008 and kicked in uniformly at a 5% R&D intensity. It is reassuring that we observe the R&D intensity bunching solely at 5% and no significant spikes at 3%, 4%, and 6%.

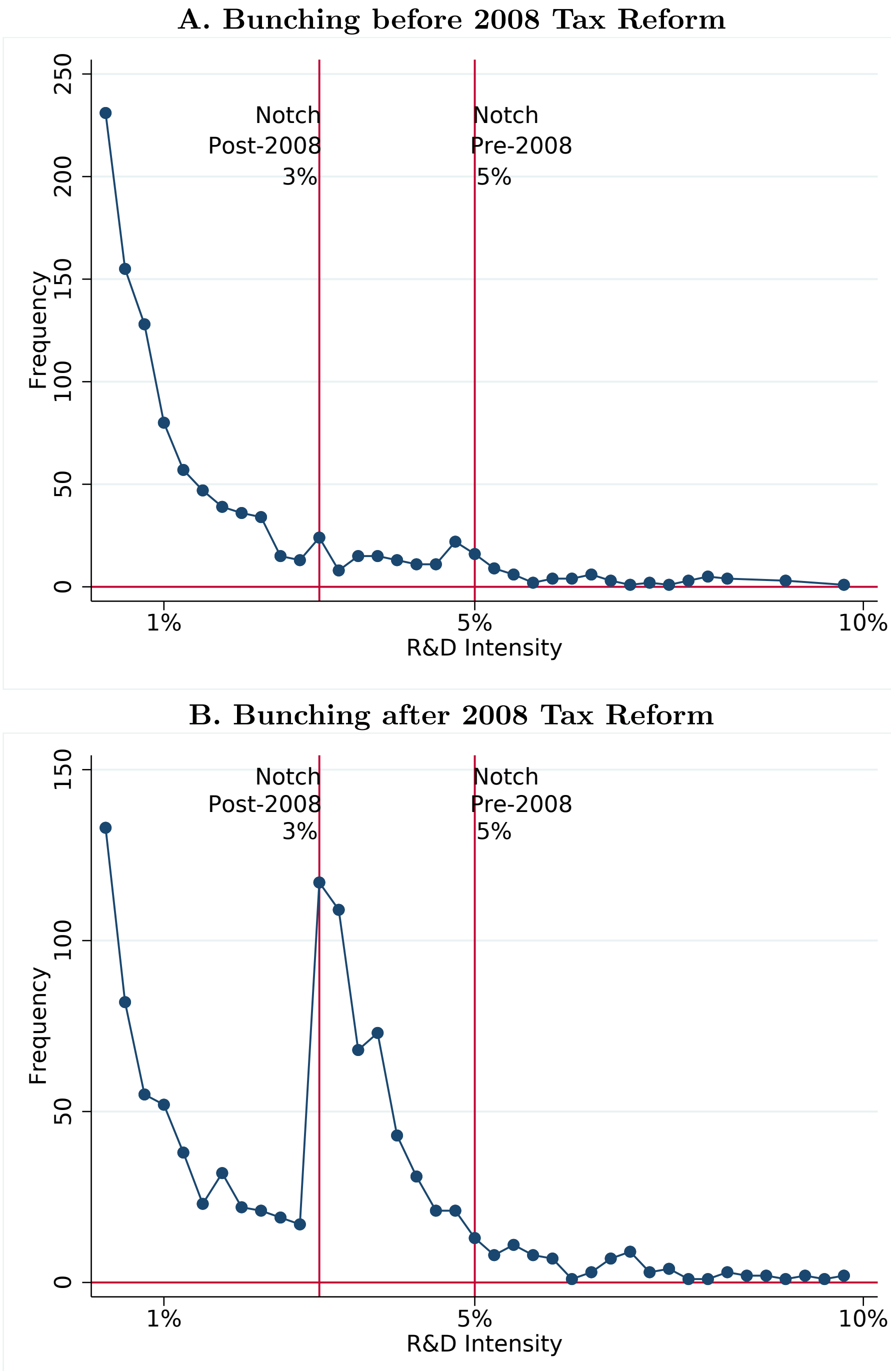
Figure 2. : Bunching at Different Thresholds of R&D Intensity (2011)



Note: This figure plots the empirical distribution of R&D intensity for all manufacturing firms with R&D intensity between 0.5% and 15% in the Administrative Tax Return Database. Panel A reports the pooled data distribution with all sizes of firms. Panels B, C, and D report the R&D intensity distribution of small, medium, and large firms, respectively. Note that large fractions of the firms bunch at the thresholds (6% for large, 4% for medium, and 3% for large) at which they qualify to apply for the InnoCom certification. Source: Administrative Tax Return Database. See Section II.A for details.

We now compare bunching patterns before and after the 2008 tax reform. Figure 3 compares the R&D intensity distribution for large FOEs before and after 2008. Large FOEs have no clear pattern of bunching before 2008. This is consistent with the fact that FOEs had a very favorable EIT treatment before the reform, which severely reduced the appeal of the InnoCom program. In contrast, FOEs start behaving like DOEs after 2008, when the InnoCom program became one of the most important tax breaks for FOEs. Their R&D intensity distribution shows a clear bunching pattern at 3% after the reform, which is the exact threshold required for these firms to qualify as HTEs. The figure demonstrates that the change in the EIT system had a large impact on firm behavior.

Figure 3. : Effects of the 2008 Tax Reform on the Bunching of Foreign-Owned Large Companies



Note: This figure compares the R&D intensity distribution for large foreign-owned firms before and after the 2008 tax reform. To make the two samples comparable, the figure plots only firms that we observe in both the SAT and ASM data. The tax reform eliminated the preferential corporate income tax for foreign-owned firms and increased their incentives to qualify for the InnoCom program. Compared with panel A, panel B shows that these firms increased their bunching behavior substantially after 2008. Source: Administrative Tax Return Database and Annual Survey of Manufacturers. See Section II.A for details.

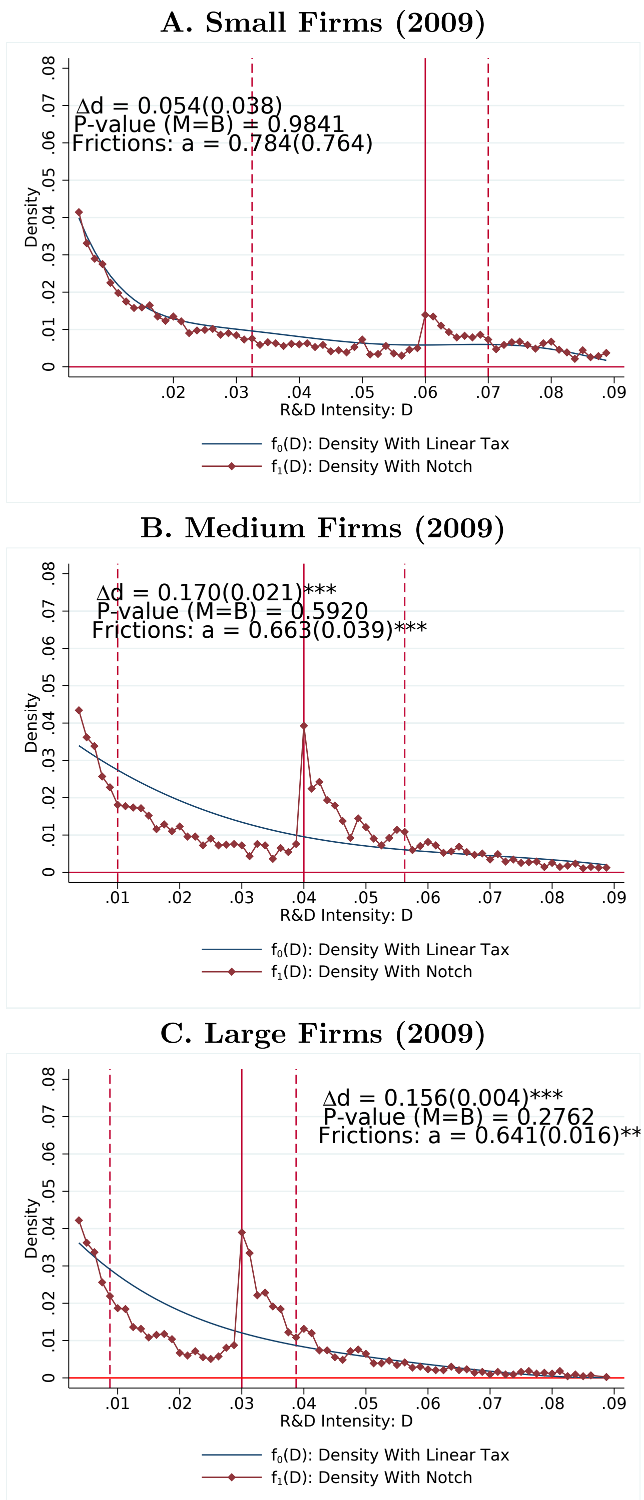
While Figures 2–3 show that the InnoCom program led to pronounced bunching patterns in the distribution of R&D intensity, these graphs alone do not allow us to quantify the overall increase in R&D. One approach to quantifying the increase in R&D is to use the observed density of R&D intensity, $f_1(\cdot)$, to infer the density in the counterfactual world without the InnoCom program, $f_0(\cdot)$. This approach relies on the assumption that only firms with R&D intensity in a given region $[d^{*-}, d^{*+}]$ respond to the program. This assumption allows us to use firms drawn from $f_1(\cdot)$ that are outside this region to estimate $f_0(\cdot)$. Following the literature (e.g., Kleven, 2016), we first group the data into bins of R&D intensity, d, and then estimate the following flexible polynomial:

$$c_j = \sum_{k=0}^{p} \beta_k \cdot (d_j)^k + \gamma_j \cdot \mathbf{1}\left[d^{*-} \leq d_j \leq d^{*+}\right] + \nu_j,$$

where c_j is the count of firms in the bin corresponding to R&D intensity d_j and p is the order of the polynomial regression. $\hat{c}_j = \sum_{k=0}^{p} \hat{\beta}_k \cdot (d)^k$ is then an estimate for $f_0(d)$. Intuitively, when only firms in the exclusion region $[d^{*-}, d^{*+}]$ respond to the program, $\hat{f}_0(d)$ will be equal to $f_1(\cdot)$ outside this region. To ensure that the estimation is not contaminated by firm responses to the program, d^{*-} and d^{*+} are determined by a data-driven procedure that ensures that $\hat{f}_0(\cdot)$ has the same mass over the excluded region as $f_1(\cdot)$.[12]

[12]Specifically, we use K-fold cross-validation to select p, d^{*-}, and d^{*+} assuming that $f_0(\cdot)$ is downward-sloping. We obtain standard errors by bootstrapping residuals. See Appendix D for details. Appendix E shows that our results are robust to excluding firms with extensive-margin responses. Appendix I shows that the assumption that only firms in $[d^{*-}, d^{*+}]$ respond to the program is consistent with our model in Section III.

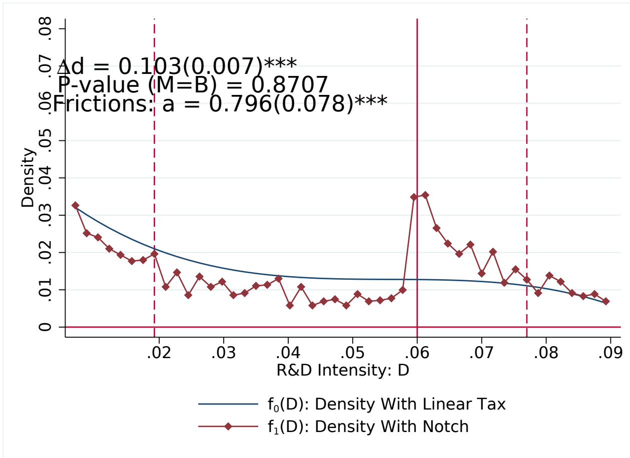
Figure 4. : Estimated Counterfactual Densities of R&D Intensity



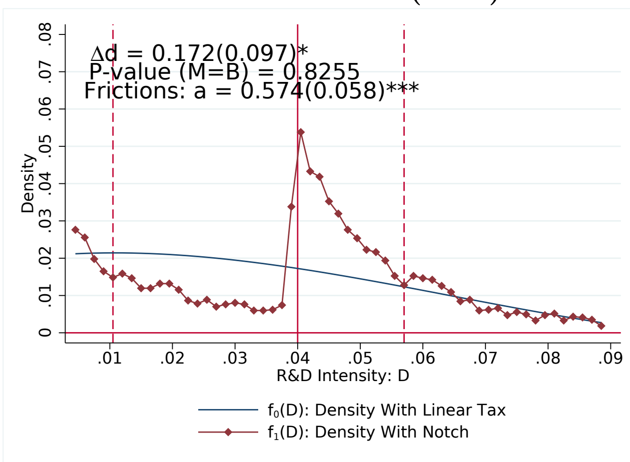
Note: This figure reports the results of our bunching estimator for small, medium, and large firms in 2009 and 2011. In each panel, we plot the empirical density of R&D intensity in red and the estimated counterfactual R&D intensity in blue. The lower bound d^{*-} and upper bound d^{*+} for the excluded region are indicated by vertical dashed lines. Δd is the percentage increase in R&D in the excluded region, and a^{*} is the fraction of firms that are constrained from participating in the program. We report the p-value of the test that the missing mass equals the excess mass. See Section II.B for details. Source: Administrative Tax Return Database.

Figure 4. : (Cont.) Estimated Counterfactual Densities of R&D Intensity

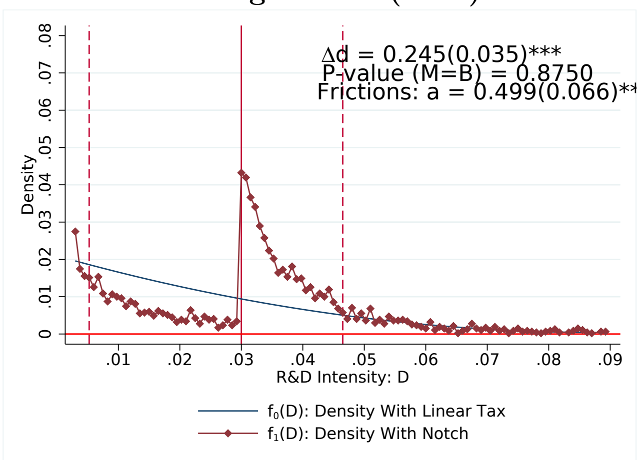
D. Small Firms (2011)



E. Medium Firms (2011)



F. Large Firms (2011)



Note: This figure reports the results of our bunching estimator for small, medium, and large firms in 2009 and 2011. In each panel, we plot the empirical density of R&D intensity in red and the estimated counterfactual R&D intensity in blue. The lower bound d^{*-} and upper bound d^{*+} for the excluded region are indicated by vertical dashed lines. Δd is the percentage increase in R&D in the excluded region, and a^* is the fraction of firms that are constrained from participating in the program. We report the p-value of the test that the missing mass equals the excess mass. See Section II.B for details. Source: Administrative Tax Return Database.

Figure 4 displays the results of this estimation. In each panel, the red line with diamond markers displays the observed distribution of R&D intensity $f_1(\cdot)$, the vertical dashed lines display the omitted region, and the blue line displays the estimated counterfactual density $\hat{f}_0(\cdot)$. To characterize the impact of bunching on average R&D intensity, we compute Δd as the percentage increase in average R&D intensity for firms in the exclusion region.[13] Panels A and D report results for small firms in 2009 and 2011, including the percentage increases in R&D over the excluded region of $\Delta d = 5.4\%$–10.3%. The size of these effects is constrained by the fact that many firms are not able to respond to the program. The fraction of firms that do not respond to the program in 2011 is $a^* = 79.6\%$.[14,15] Panels B and E show larger responses for medium firms, with Δd of 17%. These average increases are driven by heterogeneous firm-level responses. Firms immediately below the notch only require a marginal increase in R&D, while firms at d^{*-} see much larger R&D increases. Panels C and F report the results for large firms, where we estimate $\Delta d = 15.6\%$ for 2009 and $\Delta d = 24.5\%$ for 2011. These graphs also show that even large firms may be unable to satisfy some of the requirements of the program, since 50%–64% of firms that could have participated in the program opt not to do so. These results show that bunching patterns are persistent over time.[16] Appendix E shows that these bunching estimates are robust to a battery of specification tests.[17] Figure 4 contributes to our understanding of the effects of the InnoCom program by quantifying the average increase in R&D, by clarifying the significant heterogeneity in firm-level responses, and by showing that firms face idiosyncratic barriers to fulfilling the non-R&D requirements of the program.

[13]We use $f_1(\cdot)$ and $\hat{f}_0(\cdot)$ to directly calculate $\mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})]$ and $\mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]$, respectively. Δd is then the increase in R&D relative to the average R&D intensity in the exclusion region.

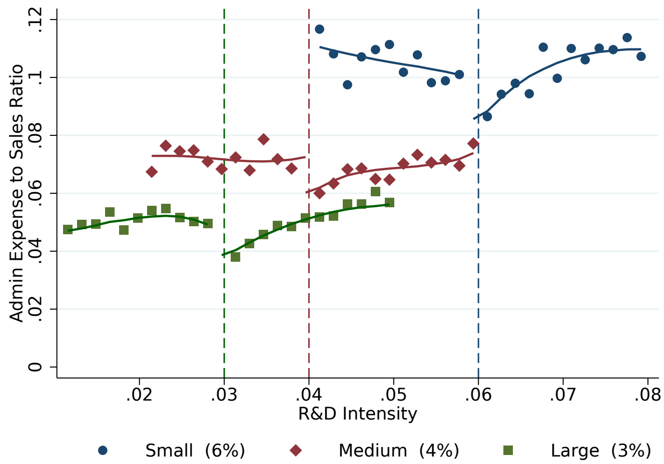
[14]Because the total mass of firms that could have responded is given by $\int_{d^*\_}^{\alpha} \hat{f}_0(v)dv$, for a given notch α, the fraction of firms that do not respond is $a^* = \int_{d^*\_}^{\alpha} f_1(v)dv / \int_{d^*\_}^{\alpha} \hat{f}_0(v)dv$. Note that small firms may be constrained in their ability to increase investment to a significant degree or to develop a new product. In addition, a higher failure rate among small firms implies that a long process of certification may never pay off in lower taxes.

[15]These graphs also report that we cannot reject the specification test that $\hat{f}_0(\cdot)$ has the same mass as $f_1(\cdot)$ over the excluded region for all types of firms.

[16]Consistent with the intent of the program, firms' bunching patterns are persistent over time: 76% of firms that report an R&D intensity greater than the notch in 2011 also bunched in 2010. For this reason, our model considers the choice of R&D as a medium-term investment plan.

[17]Specifically, we show that our estimator is able to recover a null effect in the absence of a notch and that our results are robust to excluding firms with extensive-margin responses and to excluding state-owned enterprises, low-tech firms, or low-profitability firms from the estimation. We also find similar estimates when we vary the choices of (p, d^{*-}, d^{*+}), and we even obtain similar estimates when we rely only on data above d^{*+} to estimate the counterfactual density. Our results are also robust to using data from large foreign firms before 2008 that were not subject to the incentives of the InnoCom program to inform the shape of the density in the excluded region. This check uses the insight of Blomquist and Newey (2017) that variation in non-linear incentives can help in identifying responses when bunching approaches are used.

Figure 5. : Empirical Evidence of Relabeling



Note: This figure plots the non-R&D administrative expense-to-sales ratio at each level of R&D intensity. The green dots/line are for the large firms, the red dots/line are for the medium firms, and the blue dots/line are for the small firms. The threshold of R&D intensity for firms to qualify for InnoCom certification differs by firm size: 6% for small firms, 4% for medium firms, and 3% for large firms. For each size category, there is a pronounced drop in the administrative expense-to-sales ratio when the R&D intensity approaches the required threshold. Source: Administrative Tax Return Database. See Section II for details.

C. Detecting Relabeling of R&D Investment

We now explore the degree to which the bunching response may be due to expense misreporting. Figure 5 explores how the ratio of non-R&D administrative expenses to sales is related to R&D intensity. For each size group, this figure groups firms into bins of R&D intensity and plots the mean non-R&D administrative expense-to-sales ratio for each bin. We report the data along with an estimated cubic regression of the expense ratio on R&D intensity with heterogeneous coefficients above and below the notches. The green squares are for large firms, red diamonds for medium firms, and blue dots for small firms. There is an obvious discontinuous jump downward at the notch for each size category. This drop suggests that some firms that report R&D intensity at the notch may partly relabel non-R&D expenses as R&D to qualify for the policy. When firms are farther away from the bunching threshold, there is no systemic difference in the administrative expense-to-sales ratio. This pattern is consistent with the hypothesis that firms miscategorize non-R&D expenses as R&D when they approach the bunching thresholds.[18]

[18]The existence of different thresholds across size groups also allows us to rule out other explanations for these discontinuities. In particular, there is no observable discontinuity when we impose the "wrong"

The structural breaks in Figure 5 are statistically significant for all three groups. Large firms see a drop of 0.8% of sales, which corresponds to 26% of the R&D intensity required to participate in the program. Small and medium firms see drops of 1.4% and 1.3%, respectively (see Table A.2). Because firms select into the program based on idiosyncratic factors (e.g., productivity, adjustment and certification costs), these estimates do not have a causal interpretation.[19] Nonetheless, these estimates present strong descriptive evidence that firms may respond to the InnoCom program by relabeling non-R&D expenses.

Lack of Sales Manipulation

The stark bunching patterns in Figures 2–4 raise the concern that firms may also manipulate their sales. There are two ways in which firms may do this. First, since the incentives of the InnoCom program are stated in terms of R&D intensity (R&D/Sales), firms could increase their R&D intensity by under-reporting sales. Panel A in Figure 6 plots firms' log sales relative to their R&D intensity. For each group of firms, we report average log sales for small bins of R&D intensity as well as an estimated cubic regression that is allowed to vary below and above each threshold. If firms under-report sales to achieve the target, we might expect a sudden drop in sales to the right of each threshold. In contrast, this figure shows that both the data and the estimated polynomial regressions are remarkably stable at each notch.[20] One reason for this result is that, in addition to the limits placed by third-party reporting in the VAT system, firm managers may not want to misreport sales, as these are seen as a measure of their job performance.

Second, if a firm wants to be categorized as a larger firm to qualify for a lower R&D intensity threshold, it may over-report sales. Panels B and C in Figure 6 show the histogram of firms around the size thresholds. Since larger firms face lower R&D intensity thresholds, we might expect firms to bunch on the right of the size threshold. These figures show that firms do not respond to the incentives by manipulating their size.[21] Overall, it does not appear that firms misreport sales to quality for the InnoCom program. One reason for this result is that, in addition to the limits placed by third-party reporting in the VAT system, firm managers may not want to misreport sales, as these are seen as a measure of their job performance.

thresholds of the other size groups. In Appendix G, we explore whether firms adjust other costs that are not in the administrative cost category, and we show that firms do not respond to the program by manipulating other expenses. We also conduct a similar set of analysis focusing on the ratio of R&D to total administrative expenses. In this case, expense miscategorization would result in discontinuous increases in this ratio at the notch. This is confirmed in Table A.4 and in Figure A.2.

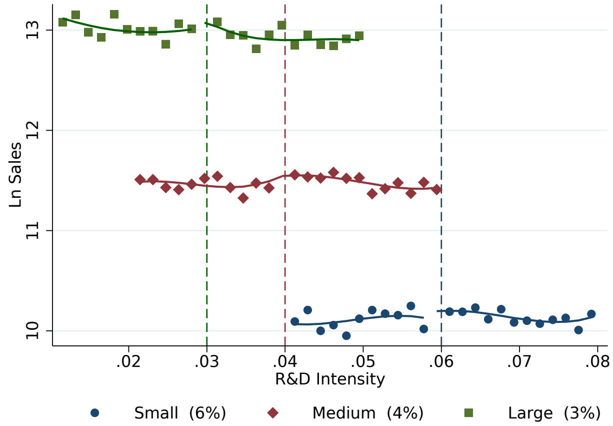
[19] Appendix F uses the methods of Diamond and Persson (2016) to estimate causal effects of the notch. Consistent with Figure 5, we estimate that the program led to a significant decrease in the average administrative cost ratio for firms in the excluded region, and we also find a statistically significant increase in TFP.

[20] Table A.3 reports statistically insignificant estimates of the structural breaks at these notches.

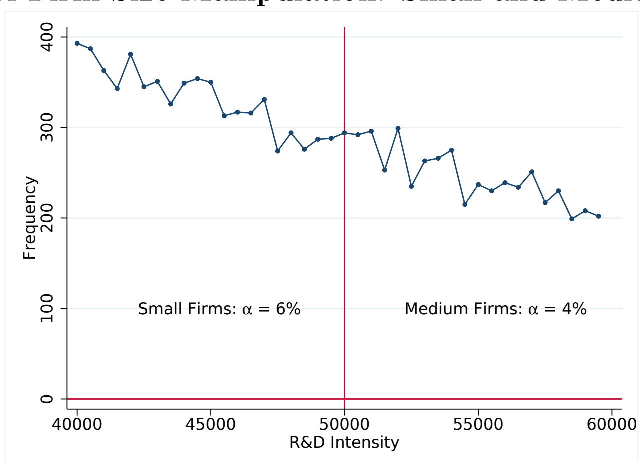
[21] In our estimations, we further restrict our sample to exclude firms that are close to the size threshold, and this does not affect our estimates.

Figure 6. : Lack of Sales Manipulation

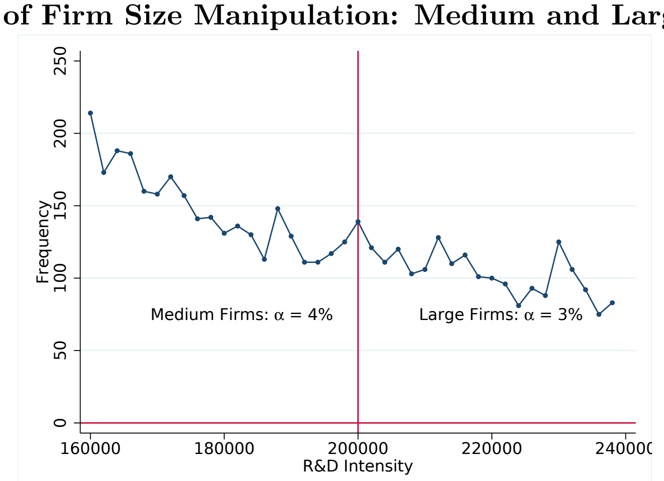
A. Lack of Sales Manipulation around R&D Intensity Thresholds



B. Lack of Firm Size Manipulation: Small and Medium Firms



C. Lack of Firm Size Manipulation: Medium and Large Firms



Note: This figure examines the potential manipulation of sales data. Panel A shows that firms do not manipulate sales by under-reporting their sales to reach their respective notch. Panels B and C show that firms do not attempt to over-report their sales to move into the next size category and thus reduce the threshold of R&D intensity needed to qualify for the InnoCom program. Overall, there is no evidence of sales manipulation. Source: Administrative Tax Return Database and Annual Survey of Manufacturers. See Section II for details.

The data patterns discussed in this section reveal a number of facts that motivate our model. First, the dispersed density of R&D intensity suggests firms face heterogeneous costs of adjusting R&D expenditures. Second, the InnoCom program led to significant increases in reported R&D investment for firms close to the notch. Third, the overall increase in R&D is driven by heterogeneous responses that depend on firms' pre-existing innovation activities. Fourth, differences in TFP between firms with low and high levels of R&D intensity suggest both that R&D investment may increase productivity and that firms may select into the InnoCom program partly based on heterogeneous adjustment costs of R&D investment. Fifth, the fact that many firms with R&D intensity close to the notch do not participate in the InnoCom program suggests firms face different obstacles that prevent them from obtaining the InnoCom certification. Finally, sharp drops in other administrative expenses at program notches suggest that firms inflate reported R&D expenditures by relabeling administrative expenses as R&D. A model of firm behavior that is consistent with these facts must therefore account for firm differences in underlying productivity as well as idiosyncratic costs of both adjusting R&D and obtaining the InnoCom certification. In addition, it is important to consider that firms may respond to the program by investing in R&D (to increase future productivity) or by relabeling other expenses (to obtain a preferential tax rate).

III. A Model of R&D Investment and Corporate Tax Notches

This section develops a model of R&D investment where firms can respond to notches in the corporate income tax schedule by investing in R&D and by relabeling non-R&D expenses. The model is motivated by the empirical facts in the previous section and shows that these data patterns inform structural parameters that are key for studying the effectiveness of alternative tax incentives.

A. Model Setup

Consider a firm i with a unit cost function $c(\phi_{it}, w_t) = w_t \exp\{-\phi_{it}\}$, where w_t is the price of inputs.[22] ϕ_{it} is log TFP and has the following law of motion:

$$\phi_{i,t} = \rho\phi_{i,t-1} + \varepsilon \ln(D_{i,t-1}) + u_{it}, \tag{1}$$

where $D_{i,t-1}$ is R&D investment and $u_{i,t} \sim$ i.i.d. $N(0, \sigma^2)$. Because our empirical analysis focuses on firms with non-trivial R&D, this law of motion applies to firms with $D_{i,t-1} > 0$.[23] This setup is consistent with the R&D literature where knowledge capital depreciates over time (captured by ρ) and is influenced by R&D expenditures (captured by ε).

[22] We provide additional model details in Appendix H. Note that any homothetic production function with Hicks-neutral productivity admits this representation.

[23] If firms do not engage in R&D, we assume that their productivity process is $\phi_{it} = \rho\phi_{i,t-1} + u_{it}$. In Appendix L, we further generalize our setup to allow knowledge spillovers across firms.

We assume that the firm faces a demand function with a constant elasticity: $\theta > 1$. This setup implies that firm sales are given by $\theta\pi_{it}$ and that we can write expected profits as follows:

$$\mathbb{E}[\pi_{it}] = \tilde{\pi}_{it} D_{i,t-1}^{(\theta-1)\varepsilon},$$

where $\tilde{\pi}_{it} \propto \mathbb{E}[\exp\{(\theta-1)\phi_{it}\}|\phi_{i,t-1}]$ measures the non-R&D expected profitability of the firm.

In our empirical setting, firms are only eligible to apply to the InnoCom program after demonstrating high levels of R&D over a three-year period (see Section I). Since firms commit to maintaining sustained levels of R&D to obtain the tax cut, the relevant investment margin is a medium-term decision. We therefore model the firm's investment decision as a two-period problem.

R&D Choice under a Linear Tax

We first model how R&D investment decisions would respond to a linear income tax:

$$\max_{D_{i1}} (1-t_1)\left(\pi_{i1} - D_{i1} - g(D_{i1}, \theta\pi_{i1})\right) + \beta(1-t_2)\tilde{\pi}_{i2} D_{i1}^{(\theta-1)\varepsilon}.$$

In addition to the direct R&D investment cost D_{i1}, firms pay a cost $g(D_{i1}, \theta\pi_{i1})$ to adjust their R&D. Following the investment literature, we adopt a quadratic formulation for $g(D_{i1}, \theta\pi_{i1}) = b \times \frac{\theta\pi_{1i}}{2}\left[\frac{D_{i1}}{\theta\pi_{1i}}\right]^2$. Absent adjustment costs, our model would predict a deterministic relationship between log R&D and log TFP. In reality, however, the distribution of R&D investment in China varies significantly across firms, even conditional on firm TFP. This variability reflects the fact that firms have different opportunities to improve their technology and face different costs of implementing R&D projects. Our model incorporates these real-world features by assuming that firms face heterogeneous adjustment frictions b of conducting R&D.

The optimal choice of D_{i1}^* is given by:[24]

$$FOC: \; -(1-t_1)\left(1 + b\left[\frac{D_{i1}}{\theta\pi_{i1}}\right]\right) + \beta(1-t_2)\varepsilon(\theta-1)D_{i1}^{(\theta-1)\varepsilon-1}\tilde{\pi}_{i2} = 0.$$

The marginal benefit of R&D depends on the potentially unobserved, firm-specific productivity ϕ_{i1}, as it determines non-R&D profitability, $\tilde{\pi}_{i2}$. The marginal cost, on the other hand, is linear in R&D and depends on the heterogeneous adjustment cost b. Intuitively, the law of motion for TFP (Equation 1) implies that increasing R&D has a proportional increase in the TFP of all units of production within a firm. As a result, firm's R&D expenditure is increasing in ϕ_{i1}. Since adjustment costs are proportional to firm size, they limit the scale effect of R&D investment

[24] As we discuss in Appendix H, we assume $(\theta-1)\varepsilon < 1$ to ensure a well-behaved second-order condition.

and play an important role connecting the distribution of TFP to the distribution of R&D intensity.[25]

R&D intensity, defined as the R&D-to-sales ratio, has an ambiguous relationship with ϕ_{i1}. To see this, we express the firm's FOC in terms of the choice of R&D intensity, $d_{i1} = \frac{D_{i1}}{\theta\pi_{i1}}$, such that

$$\underbrace{-(1-t_1)\left(1+bd_{i1}^{*}\right)}_{\text{Increase in Investment Cost}} \underbrace{+\beta(1-t_2)\varepsilon(\theta-1){d_{i1}^{*}}^{(\theta-1)\varepsilon-1}\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}}_{\text{Productivity Gain from R\&D}} = 0. \tag{2}$$

This equation shows that the relation between d_{i1}^{*} and ϕ_{i1} depends on whether the term $\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}$ is increasing or decreasing in TFP. Because ϕ_{i1} affects both expected profitability ($\tilde{\pi}_{i2}$) and current sales (π_{i1}), ε plays an important role in shaping the joint distribution of R&D intensity and TFP, a fact that we use in the estimation of our model.

A Notch in the Corporate Income Tax

Assume now that the tax in the second period has the following structure, modeled after the incentives in the InnoCom program:

$$t_2 = \begin{cases} t_2^{LT} & \text{if } d_{i1} < \alpha \\ t_2^{HT} & \text{if } d_{i1} \geq \alpha \end{cases},$$

where $t_2^{LT} > t_2^{HT}$ and where LT/HT stands for low-tech/high-tech. In practice, firms with high R&D intensity may not participate in the program if other constraints prevent them from hiring a sufficient number of technical employees, if they do not obtain a significant fraction of their sales from high-tech products, or if the compliance and registration costs are too high. We model these constraints by assuming that firms pay a fixed cost of certification: $c \times \theta\pi_{1i}$, where c varies across firms.

A firm decides whether to bunch by comparing the value of the firm from bunching, by setting $d_1^{*} = \alpha$, to the value of the firm at its optimal R&D intensity below the notch, i.e., d_{i1}^{*} from Equation 2. The value-to-sales ratio of the firm conditional on bunching, $\frac{\Pi(\alpha|t_2^{HT})}{\theta\pi_{i1}}$, is given by:

$$\frac{\Pi(\alpha|t_2^{HT})}{\theta\pi_{i1}} \equiv (1-t_1)\frac{1}{\theta}+\beta(1-t_2^{HT})\alpha^{(\theta-1)\varepsilon}\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}-(1-t_1)\left[\alpha\left(1+\frac{b\alpha}{2}\right)+c\right].$$

[25] Appendix K.3 shows that the results of our empirical model are robust to allowing for more flexible adjustment costs.

Similarly, the value-to-sales ratio at the interior optimal d_{i1}^*, $\frac{\Pi(d_{i1}^*|t_2^{LT})}{\theta\pi_{i1}}$, is:

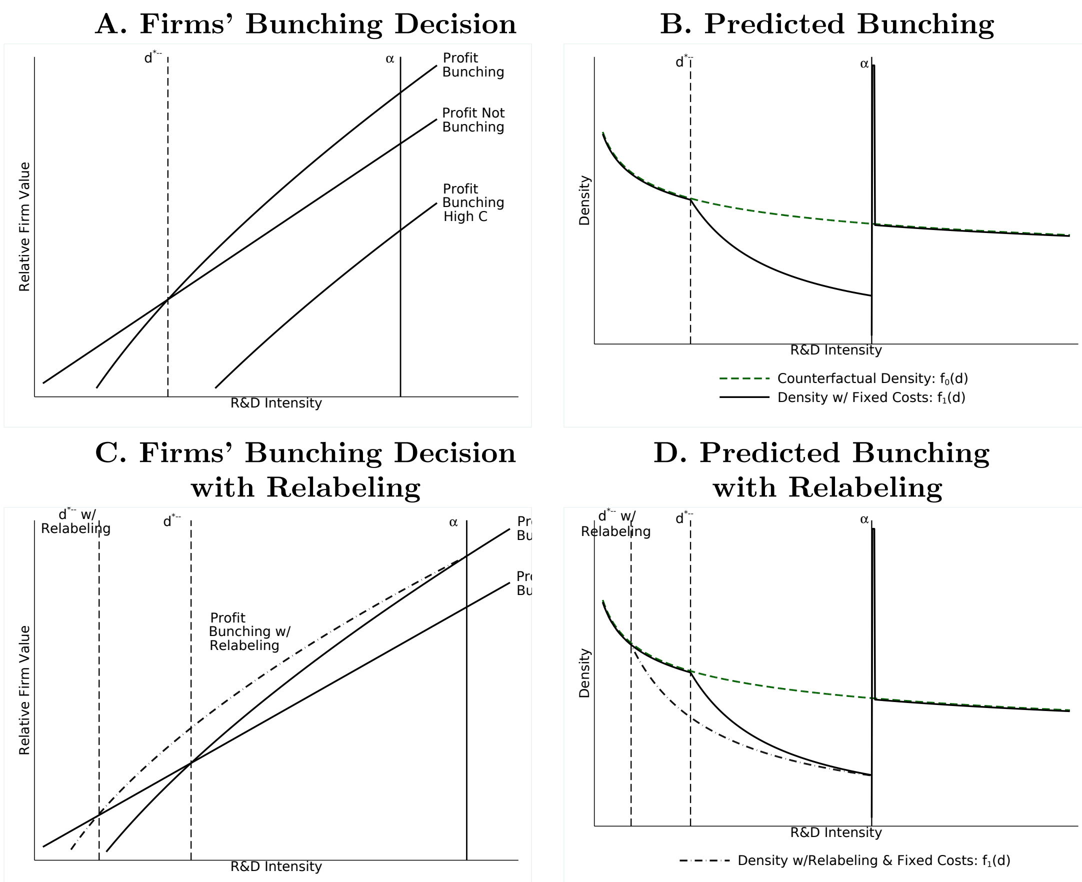
$$\frac{\Pi(d_{i1}^*|t_2^{LT})}{\theta\pi_{i1}} \equiv (1-t_1)\frac{1}{\theta}+\beta(1-t_2^{LT}){d_{i1}^*}^{(\theta-1)\varepsilon}\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}-(1-t_1)d_{i1}^*\left(1+\frac{bd_{i1}^*}{2}\right).$$

A firm that previously chose $d_{i1}^* < \alpha$ will bunch at the notch if $\frac{\Pi(\alpha|t_2^{HT})}{\theta\pi_{i1}} \geq \frac{\Pi(d_{i1}^*|t_2^{LT})}{\theta\pi_{i1}}$.

There are strong theoretical predictions regarding the effect of the tax notch on the cross-sectional distribution of R&D intensity. To build intuition, we refer to the simple case where the adjustment cost b is equal to zero. Substituting the term $\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}$ using Equation 2, we can express the decision to bunch or not as:

$$\underbrace{\left(\frac{d_{i1}^*}{\alpha}\right)^{1-(\theta-1)\varepsilon}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\frac{1}{(\theta-1)\varepsilon}-1-c}_{\text{Relative Profit from Bunching}} \geq \underbrace{\frac{d_{i1}^*}{\alpha}\left(\frac{1}{(\theta-1)\varepsilon}-1\right)}_{\text{Relative Profit from Not Bunching}}. \tag{3}$$

Figure 7. : Theoretical Bunching Predictions



Note: This figure provides intuition for when a firm decides to bunch and describes empirical implications of our model for R&D investment and bunching. Panel A visualizes Equation 3 by plotting the relative value from bunching, $\left(\frac{d^*_{i1}}{\alpha}\right)^{1-(\theta-1)\varepsilon}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\frac{1}{(\theta-1)\varepsilon}-1-c$, and the relative profit from not bunching, $\frac{d^*_{i1}}{\alpha}\left(\frac{1}{(\theta-1)\varepsilon}-1\right)$, as functions of the optimal R&D intensity level in the absence of the notch, d^*_{i1}. Absent fixed costs ($c=0$), the value from bunching exceeds the value of not bunching when $d^*_{i1}\approx\alpha$. All firms with $c=0$ and with $d^*_{i1}\in[d^{*-},\alpha]$ decide to bunch. When ε is small, the profit from bunching is steeper, which shifts the value of d^{*-} to the right and reduces the likelihood that firms will bunch. The firm value from bunching shifts down for $c>0$ so that firms with d^*_{i1} farther from α are less likely to bunch. When c is large enough, firms with $d^*_{i1}\approx\alpha$ may not participate in the program. Panel B shows how the incentives of the InnoCom program impact the density of R&D intensity, $f_1(d)$, relative to a counterfactual density without the program, $f_0(d)$. Panel C plots the relative firm value from relabeling (from Equation 5) and shows that, by flattening the slope of this line, relabeling decreases the R&D intensity of the marginal buncher. Panel D shows that the possibility of relabeling shifts d^{*-} to the left and increases the likelihood that firms will bunch. See Section III for details.

Panel A of Figure 7 visualizes this inequality by plotting the relative profits from bunching and not bunching as a function of R&D intensity. For firms that were already close to the notch $\left(\frac{d^*_{i1}}{\alpha} \approx 1\right)$, bunching has small costs and productivity benefits, but the tax cut $\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) > 1$ incentivizes firms to bunch. This figure shows that, when $c = 0$, bunching is optimal for firms with d^*_{i1} close to α. For firms farther from the notch (as d^*_{i1} decreases from α), the additional investment costs increase faster than the productivity benefits, which reduces firms' incentive to bunch. Let d^{*-} be the marginal firm such that Equation 3 holds with equality. Firms with $d^*_{i1} \in (d^{*-}, \alpha)$ would decide to bunch at the notch, since the difference between the left- and right-hand sides of Equation 3 is increasing in d^*_{i1}. It can also be shown that d^{*-} is decreasing in both $(\theta - 1)\varepsilon$ and $\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)$, so that we would observe more bunching if firms have a higher valuation of R&D or if the tax incentive is larger.

To visualize the role of fixed costs, Panel A of Figure 7 shows that the relative profit from bunching shifts down as c increases. This implies that firms with d^*_{i1} farther from α are less likely to bunch at the notch. When c is large enough, however, firms with $d^*_{i1} \approx \alpha$ may not be able to participate in the program. Panel B of Figure 7 depicts this prediction for the cross-sectional R&D intensity distribution. The green dashed line plots $f_0(d)$: the distribution of optimal R&D intensity under a linear tax. The black line plots $f_1(d)$: the density of R&D intensity with a notch.[26] In addition, the presence of adjustment costs implies that each firm's bunching decision depends on its idiosyncratic value of b. Firms with similar productivity will therefore differ in how they respond to the InnoCom program.[27]

B. *Real and Relabeled R&D Investment under a Tax Notch*

This section extends the model by allowing firms to inflate reported R&D expenditures by relabeling non-R&D costs as R&D. Denote a firm's reported level of R&D spending by $\tilde{D}_{i1}$. Firms qualify for the lower tax whenever $\tilde{D}_1 \geq \alpha\theta\pi_1$. We assume that firms face an expected cost of misreporting that is given by $h(D_{i1}, \tilde{D}_{i1})$, which represents the likelihood of being caught and the punishment from the tax authority. We further assume that the cost of misreporting is proportional to the reported R&D and depends on the percentage of misreported R&D, $\delta_{i1} = \frac{\tilde{D}_{i1} - D_{i1}}{\tilde{D}_{i1}}$, so that:

$$h(D_{i1}, \tilde{D}_{i1}) = \tilde{D}_{i1}\tilde{h}\left(\delta_{i1}\right),$$

[26] Note that, in the special case of no fixed costs, the range (d^{*-}, α) would be dominated by the notch α and there would be an empty region below the notch. This prediction is not consistent with the data patterns that we documented in Section II.

[27] Equation H.6 generalizes Equation 3 by including both adjustment and fixed costs.

where $\tilde{h}$ satisfies $\tilde{h}(0) = 0$ and $\tilde{h}'(\cdot) \geq 0$.[28]

Notice first that if a firm decides not to bunch at the level $\alpha\theta\pi_1$, it does not have an incentive to misreport R&D spending, as doing so would not affect total profits or the tax rate. However, a firm might find it optimal to report $\tilde{D}_1 = \alpha\theta\pi_1$ even if it actually invested in a lower level of R&D. Conditional on bunching, the firm's optimal relabeling strategy solves the following problem:

$$\max_{D_{i1}^K} \quad (1-t_1)\left(\pi_{i1} - D_{i1}^K - \theta\pi_{i1}c - \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}^K}{\theta\pi_{i1}}\right]^2\right) - \alpha\theta\pi_1\tilde{h}\left(\frac{\alpha\theta\pi_1 - D_{i1}^K}{\alpha\theta\pi_1}\right)$$
$$+\beta(1-t_2^{HT})\tilde{\pi}_{i2}(D_{i1}^K)^{(\theta-1)\varepsilon}$$

The first-order condition for relabeling in terms of the real R&D intensity $d_1^K = \frac{D_1^K}{\theta\pi_1}$ is then:

$$(4) \qquad 0 = \underbrace{-(1-t_1)\left(1+bd_{i1}^{K*}\right) + \tilde{h}'\left(1 - \frac{d_{i1}^{K*}}{\alpha}\right)}_{\text{Increase in Investment Cost and Reduction in Relabeling Cost}}$$
$$+ \underbrace{\beta(1-t_2^{HT})\varepsilon(\theta-1)d_{i1}^{K*(\theta-1)\varepsilon-1}\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}}_{\text{Productivity Gain from Real R\&D}}.$$

Comparing Equation 5 with the first-order condition Equation 2 for d_{i1}^* in the case without relabeling, we find that—despite the presence of relabeling—firms generally increase their real R&D intensity when they bunch, i.e., $d_{i1}^{K*} > d_{i1}^*$. The marginal incentive of investing in real R&D is higher for two reasons. First, since certified firms face a lower tax rate, $t_2^{HT} < t_2^{LT}$, the after-tax benefits of productivity improvements are larger. Second, real R&D investment also makes it less likely that a firm will be caught and punished for its relabeling behavior. This feature is known as the avoidance-facilitating effect, whereby real R&D lowers the marginal cost of relabeling (Slemrod and Gillitzer, 2013). Based on d_{i1}^{K*}, we define the fraction of relabeled R&D $\delta_{i1}^* = 1 - d_{i1}^{K*}/\alpha$ and the resulting firm value $\Pi(d_{i1}^{K*}, \alpha | t_2^{HT})$ from reporting R&D intensity α and conducting real R&D intensity d_{i1}^{K*}.

When firms can relabel, they decide whether to bunch by comparing the firm value from the optimal relabeling strategy, $\Pi(d_{i1}^{K*}, \alpha | t_2^{HT})$, with the firm value at the optimal interior solution, $\Pi(d_{i1}^*, d_{i1}^* | t_2^{LT})$. To gain further intuition, consider the simple case where $b = c = 0$. Using Equation 2 to simplify $\Pi(d_{i1}^{K*}, \alpha | t_2^{HT})$, it

[28]Our formulation of $\tilde{h}(\cdot)$ is consistent with general features of evasion cost functions in the literature (Slemrod, 2001). We assume that the misreporting cost depends on δ (the percentage of misreported R&D) because the InnoCom program is based on R&D intensity rather than total R&D expenditures. Appendix K.4 shows that the results of our empirical model are robust to an alternative relabeling cost function that can accommodate separable relabeling costs.

follows that firms decide to bunch when the following inequality holds:

$$(5)\underbrace{\left(\frac{d_{i1}^{K*}}{\alpha}\right)^{(\theta-1)\varepsilon}\left(\frac{d_{i1}^{*}}{\alpha}\right)^{1-(\theta-1)\varepsilon}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\frac{1}{(\theta-1)\varepsilon}-\frac{d_{i1}^{K*}}{\alpha}}_{\text{Relative Profit from Bunching}}-\underbrace{\frac{\tilde{h}(\delta_{i1}^{*})}{\alpha(1-t_1)}}_{\text{Relabeling Cost}}$$
$$\geq \underbrace{\frac{d_{i1}^{*}}{\alpha}\left(\frac{1}{(\theta-1)\varepsilon}-1\right)}_{\text{Relative Profit from Not Bunching}}.$$

Equations 3 and 5 are very similar and are identical in the case when $c = 0$ and $d_{i1}^{K*} = \alpha$ —i.e., when there is no relabeling $\delta_{i1}^{*} = 0$.

Panel C of Figure 7 visualizes Equation 5 to show how the possibility of relabeling impacts a firm's decision to bunch. Intuitively, since firms can elect to report truthfully ($\delta = 0$), firms' profits from bunching in the case with relabeling are greater than in the case without relabeling. Matching this intuition, the figure shows that the value of firms from bunching and relabeling is greater than in the case without relabeling. The figure also shows that, when relabeling is possible, the marginal firm (such that Equation 5 holds with equality) will have a lower threshold d^{*-}. Panel D of Figure 7 shows that we should see more bunching when firms can misreport R&D, such that the observed bunching patterns likely combine real increases in R&D with increases in relabeling. Therefore, while Equation 3 provides a tight connection between the extent of bunching and ε, Equation 5 shows that it is crucial to account for relabeling when bunching patterns are used to infer the returns to R&D.

IV. Structural Estimation

The previous section described a model motivated by the data patterns in Section II. The model links the observed bunching patterns to the distributions of productivity, adjustment costs, and certification costs, and allows firms to respond to tax incentives through productivity-enhancing investments in real R&D as well as through misreporting. This section proposes a method of simulated moments (MSM) framework to estimate the structural parameters of the model and uses these estimates to quantify the extent of relabeling and the increase in real R&D.

A. Estimation Framework

We first discuss how we parameterize the model. We begin by calibrating θ, which we set at $\theta = 5$ based on the survey by Head and Mayer (2014).[29] We use the fact that the evolution of productivity in Equation 1 is an AR(1) process with

[29]This value implies a gross markup of $\frac{\theta}{\theta-1} = 1.25$. We calibrate θ since, without data on physical production quantities, we are not able to separately identify this parameter from the productivity distribution.

persistence ρ and a normally distributed shock with variance σ^2. Given a value of θ, the persistence and volatility of log value-added of non-R&D performing firms map directly into ρ and σ^2, which yields the following calibrated values of $\rho = 0.725$ and $\sigma = 0.385$. This process implies a stationary normal distribution for the underlying productivity ϕ_1. Finally, we set $\beta = 0.925$.

We now parameterize the distributions of b and c, which we assume are *i.i.d.* across firms. We assume b is log-normally distributed, $b \sim \mathcal{LN}(\mu_b, \sigma_b^2)$, and that c has an exponential distribution, $c \sim \mathcal{EXP}(\mu_c)$. We adopt the following functional form for the costs of relabeling: $\frac{\exp\{\eta\delta\}-1}{\eta}$, where δ is the fraction of reported R&D corresponding to relabeling. While it is necessary to specify a functional form, this specification is quite flexible, as the function can be linear, convex, or concave depending on the value of η (e.g., Notowidigdo, 2019).

We use the method of simulated moments to estimate the parameters $\Omega = \{\varepsilon, \eta, \mu_b, \sigma_b, \mu_c\}$. For a given value of these parameters, we simulate productivity and adjustment and fixed costs for 30,000 firms. We determine whether each firm finds it optimal to bunch depending on the firm's optimal R&D investment conditional on not bunching (Equation 2) and the optimal relabeling strategy conditional on bunching (Equation 5). Based on these firm-level decisions, we compute data moments that are analogous to those discussed in Section II. We obtain the simulated moments by repeating this process 10 times and averaging over these instances. Our estimate of Ω minimizes the difference between data moments and moments generated by the distribution of simulated firms as measured by the criterion function:

$$Q(\Omega) = \begin{bmatrix} m^D(\Omega) \\ m^B(\Omega) \end{bmatrix}' W \begin{bmatrix} m^D(\Omega) \\ m^B(\Omega) \end{bmatrix},$$

where W is a bootstrapped covariance weighting matrix. $m^D(\Omega)$ and $m^B(\Omega)$ are moment conditions based on the descriptive statistics and on the bunching estimator, respectively. Because large firms account for more than 80% of all R&D investment (see Figure A.4), we use data for this group of firms to estimate the structural model.

$m^D(\Omega)$ includes four types of moments based on the data patterns in Section II. The first set of moments uses information from the histogram of R&D intensity. We include the fraction of firms falling in three equally spaced intervals below the 3% notch (i.e., [0.003, 0.012], [0.012, 0.021], and [0.021, 0.03]).[30] We summarize the top of the R&D intensity distribution by including moments that measure the fraction of firms falling in three equally spaced intervals between 5% and 9% (i.e., [0.05, 0.063], [0.063, 0.076], and [0.076, 0.09]). Second, we include the average R&D intensity for firms that potentially respond to the InnoCom program (i.e., over the interval [0.03,0.05]). Third, we include the average TFP for firms below and above the notch. As we discuss below, these moments play an important

[30] As in Figures 2–4, we exclude observations that are very close to conducting no R&D.

role in identifying key model parameters. Finally, we include the drop in the administrative cost ratio from Figure 5. This last moment plays an important role in disciplining the costs of relabeling.

Our initial model relies solely on the moments in $m^D(\Omega)$ to estimate the model. For robustness, we show that we obtain similar structural estimates when we also consider additional moments based on the bunching estimator $m^B(\Omega)$. These moments include the following: (1) the lower threshold of the excluded region d^{*-}; (2) the fraction of firms in the excluded region that do not bunch a^*; and (3) the percentage increase in R&D intensity over the excluded region Δd. In this case, our model parameters are additionally disciplined by the results from Figure 4.

Identification

While each of the simulated moments depends on multiple parameters, we give a heuristic description of the data patterns that identify each parameter.

Consider first the model that only relies on moments based on descriptive data patterns $m^D(\Omega)$. We start by discussing the identification of the distribution of fixed and adjustment costs. First, the parameters of the distribution of adjustment costs, μ_b and σ_b, are identified by the distribution of R&D intensity below the notch and in the top of the R&D intensity distribution. Next, given that the R&D intensity distribution is smooth, intuitively, there are three determinants of the excess mass of firms above the notch (over the interval [3,5]). Firms are more likely to bunch when the average certification cost μ_c is lower, when R&D has a larger effect on productivity ε, or when it is easier to relabel (lower η). The drop in the administrative cost ratio at the notch disciplines the relabeling cost η. The sorting of more productive firms into higher R&D intensity bins helps determine ε. Given η and ε, the magnitude of the certification cost μ_c is determined by the average R&D intensity right above the notch as well as the density of firm R&D right below the notch. This heuristic argument shows that our model is over-identified since our descriptive data patterns include the full empirical distribution of R&D intensity.

One benefit of using the additional moments in $m^B(\Omega)$ is that these moments compare the observed density of R&D to a flexibly estimated counterfactual density without the program. This density extracts additional information including the minimum bunching point d^{*-}, the average increase in reported R&D Δd, and the fraction of firms not bunching a^*. Similar to the excess mass of firms above the notch, these moments jointly inform the three parameters that determine bunching: ε, η, and μ_c, providing additional over-identifying restrictions.

Table 3—: Structural Estimates

A. Point Estimates

	TFP Elasticity of R&D	Relabeling Cost	Distribution of Adjustment Costs		Distribution of Fixed Costs
	ε	η	μ_b	σ_b	μ_c
Model 1: Excluding Bunching Moments					
Estimate	0.089	5.900	7.989	2.047	0.687
Standard Error	(0.002)	(0.493)	(0.086)	(0.076)	(0.062)
Model 2: All Moments					
Estimate	0.091	6.755	8.011	2.014	0.532
Standard Error	(0.002)	(0.449)	(0.075)	(0.073)	(0.012)

Note: Estimates based on calibrated values of $\theta = 5$, $\rho = 0.725$, and $\sigma = 0.385$. Model 1 estimates the structural parameters using all moments except the bunching estimates. Model 2 uses all the available moments to estimate the structural parameters. See Section IV for estimation details.

B. Simulated vs. Data Moments

	Data	Simulated	
		Model 1: Excluding Bunching	Model 2: All Moments
R&D Dist. Moments: $m^D(\Omega)$			
Below the notch (%)			
$[0.3, 1.2]$	0.373	0.382	0.379
$[1.2, 2.1]$	0.113	0.157	0.146
$[2.1, 3]$	0.067	0.080	0.069
Above manipulated region (%)			
$[5, 6.3]$	0.056	0.055	0.057
$[6.3, 7.6]$	0.026	0.037	0.038
$[7.6, 9]$	0.012	0.026	0.027
Mean R&D intensity $[3\%, 5\%]$	0.037	0.035	0.035
Average TFP below notch	-0.015	-0.017	-0.020
Average TFP above notch	0.027	0.023	0.025
Admin cost ratio break at notch	0.9%	0.8%	0.7%
Bunching Moments: $m^B(\Omega)$			
Bunching Point d^{-*}	0.009	(0.009)	0.010
Increase in Reported R&D: Δd	0.157	(0.124)	0.150
Fraction of firms not bunching	0.641	(0.738)	0.665

Note: This table compares the moments generated by our simulations with those from the data. The simulation is based on 30,000 firms. The moments that are not targeted by model 1 are in parentheses. The table shows our model does a remarkable job of matching 10 (13) moments from the data using a relatively parsimonious model based on 5 parameters.

B. Estimates of Structural Parameters

Table 3 reports estimates of our structural parameters: $(\varepsilon, \eta, \mu_b, \sigma_b, \mu_c)$. Panel A reports parameter estimates and standard errors for our two models. All the estimates are statistically significant in both models. We estimate remarkably similar parameters when we rely on the descriptive moments $m^D(\Omega)$ or when we also include the bunching moments $m^B(\Omega)$ in the estimation. Thus, while the bunching moments provide independent information, our model's quantification of the forces that generate the R&D bunching patterns are also consistent with those moments.

Consider the estimate of the returns to R&D, ε. The estimate from the full model in Table 3 panel A implies that doubling R&D increases measured TFP by 9%. Hall, Mairesse and Mohnen (2010) survey the extensive literature on this R&D elasticity in similar production function setups. Our estimate lies within the broad range of previous results, that is, between 2% and 17%. Since most previous studies use micro-data from developed countries, it is interesting to see that the returns to R&D of Chinese firms are comparable in magnitude.

Consider now the relabeling cost parameter, η. The estimates from both models are around 6. These values indicate that, at the margin, the cost of relabeling is highly convex in terms of δ. That is, it is easy for firms to overstate their R&D by a small amount, but the cost rises quickly for firms that are farther away from the required threshold α. To understand this result, note that the marginal benefit of relabeling includes reductions in investment costs and in adjustment costs, which include technological opportunity constraints. For this reason, firms that face a higher shadow cost of R&D (i.e., a higher b) will be more willing to engage in relabeling. On average, we calculate that firms' realized relabeling cost is 9.8% of the implicit R&D savings. Finally, the estimated certification cost is quite modest: for the firms that decide to bunch and certify as high-tech firms, the fixed certification cost is on average 4.4% of their expected profit.

Panel B of Table 3 compares the simulated moments with the data moments and shows that our models do a very good job of matching the data. The first model—based only on descriptive moments—replicates the distribution of firm-level R&D intensity, the bunching pattern, and the break in the administrative cost ratio very well. This model also matches the positive correlation between R&D intensity and measured productivity. Studying the predicted values of the (untargeted) bunching moments, we find that they match the data moments quite closely. The second column of Panel B reports the simulated moments for the full model. As would be expected, this model trades off a slightly better fit of the bunching moments for slight deviations from the baseline descriptive moments. However, these trade-offs are very minor: both models do a remarkable job of fitting the data.

Because the model is consistent with both sets of moments, one of the benefits of adding $m^B(\Omega)$ in the full model is an increase in the precision of the estimated parameters. While the full model features smaller standard errors for all the parameters, the biggest difference is in the standard error of μ_c, which drops from 0.06 to 0.01. This increased precision follows from the rationale that the bunching moments extract information from the counterfactual R&D intensity that we estimated in Section II.B, including the fraction of firms that are below the notch and that do not bunch. These additional restrictions reduced the uncertainty of the estimate for certification cost μ_c.

Benchmark Model Implications

Given our model estimates, we can simulate our full model to gain a deeper understanding of how heterogeneous firms respond to the existing policy.

First, we find that firms that comply with the policy are positively selected on several margins. Complier firms are, on average, 13.5% more productive than firms in the excluded region that do not comply with the policy. They also have idiosyncratic adjustment costs that are 24.3% lower than non-compliers, which indicates much better technological opportunities from R&D investment. Finally, they also have substantially smaller certification costs.

Second, our model shows that 24.2% of the reported R&D investment is due to relabeling, on average. This fraction is dispersed across firms, with the 10th percentile firm relabeling 4.3% and the 90th percentile relabeling 42.3%. This dispersion is driven mostly by dispersion in the adjustment costs, b. Conditional on firm productivity, firms with higher adjustment costs relabel a higher fraction of their R&D. Intuitively, firms with limited technological opportunities are willing to risk punishment for relabeling to reach the program threshold.

Lastly, we also find heterogeneous increases in real R&D for complying firms. Our model suggests that the distribution of real R&D investment is such that the 10th percentile firm sees an increase of 10.4%, the 90th percentile firm an increase of 29.0%, and the median firm an increase of 16.4%. This dispersion in investment then results in a dispersed distribution of gains in TFP.

C. Robustness and Sensitivity

We now show that our structural estimates are robust to relaxing many of the assumptions of our structural model. We discuss each of these cases in more detail in Appendix K.

We first investigate the parametric assumption that total factor productivity $\exp(\phi_1)$ follows a log-normal distribution. We find that the distribution of measured empirical TFP closely matches that of a log-normal distribution, which implies that this assumption is consistent with our data (see Appendix K.1).

In Appendix K.2, we discuss estimates from alternative models that allow heterogeneous εs and a constant b. While these models result in similar average values of ε and b, the models do not match the data as well as our benchmark model. Specifically, these models cannot match the joint distribution of TFP and R&D intensity.

One potential concern is that firms' adjustment costs may depend on the scale of a given firm. In Appendix K.3, we estimate an extended adjustment cost function that allows these costs to vary by firm size. Our results show that adjustment costs do not exhibit a firm-size bias and that we obtain very similar estimates of our main parameters with a more flexible adjustment cost function.

An additional concern is that our structural estimates may be influenced by the functional form of the relabeling costs. Appendix K.4 reports results from an alternative formulation that can accommodate relabeling costs that are separable from real choices. This model results in similar estimates of the productivity effects of R&D and implies a similar fraction of relabeled R&D as our baseline model.

As we mention in Section I, it is possible that the drop in administrative costs that we observe in Figure 5 may be partly driven by a real reallocation of resources. For instance, firms may reduce administrative costs if the tax incentive causes them to pay closer attention to their accounting of R&D expenses or if firms substitute inputs in response to the policy. In Appendix K.5, we explore this issue by assuming that 25% of the drop in administrative costs in Figure 5 is due to real responses and 75% is due to relabeling. As we show in Table A.7, while this assumption implies slightly larger costs of relabeling, it does not impact the rest of our structural estimates.

An important force in the model is the selection of firms into the InnoCom program. This selection is driven by differences in firm productivity and fixed costs, which we assume to be independently distributed. Appendix K.6 shows that our results are robust to allowing fixed costs to be correlated with firm productivity. Specifically, we show that an expanded model that allows an arbitrary correlation between c and ϕ yields a negligible correlation between these parameters and results in very similar estimates of our structural parameters.

As we discuss above, the productivity elasticity of R&D, ε, is partly identified by the productivity difference between firms above and below the notch. To ensure our estimates are robust to our measurement of productivity, in Appendix K.7, we report results where we replace these moments with alternative measures of firm productivity based on the methods of Ackerberg, Caves and Frazer (2015). Our results are robust to using these alternative productivity moments.

An additional way to validate our structural model is to test out-of-sample predictions. In Appendix F.1, we use the methods of Diamond and Persson (2016) to estimate treatment effects of the InnoCom program. As we show in Appendix K.8, the estimated model implies increases in firm-level TFP and relabeling that are consistent with reduced-form estimates of the effects of the InnoCom program on the administrative cost ratio and on TFP growth.

Finally, we evaluate the sensitivity of our point estimates to each individual moment. We calculate the local derivative of our estimated parameters in the full model with respect to each moment using the methods of Andrews, Gentzkow and Shapiro (2017). In general, the sensitivity matrix conforms with our heuristic discussion above. The joint distribution of TFP and R&D intensity are important determinants of ε. The extent of bunching, measured by the mean R&D intensity between $[3\%, 5\%]$, is also informative of the gains from innovation. The structural break in the administrative cost-to-sales ratio is by far the most important determinant of evasion cost η. We report the complete set of sensitivity results for ε and η in Figure A.10.

Overall, the structural model exploits the estimates from our descriptive and bunching analysis for identification and is able to replicate these data patterns quite well. While the structural model combines information from multiple moments and leverages functional form assumptions to increase the precision of the estimates, the benefit of the bunching approach is that it places no restrictions on the parameters of the model. By estimating a model that is consistent with both approaches, we reduce the risk that functional form assumptions are constraining the estimated parameters in ways that would bias the effects of the InnoCom program. For these reasons, the model provides a robust micro-foundation for simulating the effects of counterfactual policies.

V. Simulating Counterfactual Policies

We now use our model estimates to simulate the effects of alternative R&D tax incentives, and we quantify their implications for reported R&D investment, real R&D investment, tax revenue, productivity growth, and welfare. We first simulate alternative versions of the InnoCom program that vary the tax advantage and the location of the notch. We then compare our results with a counterfactual policy that follows a more standard investment tax credit. Finally, we consider whether knowledge spillovers can justify the InnoCom program from a welfare perspective.

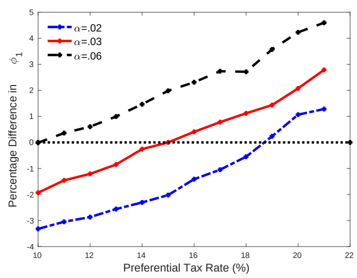
A. Alternative Notches and Tax Cuts

We analyze alternative versions of the InnoCom program that vary the tax advantage and the location of the notch for two reasons. First, even though standard policy recommendations avoid prescribing discontinuous incentives, notches are present in many settings (Slemrod, 2013) and may be justified in cases where governments can use them as a way to limit relabeling (Best et al., 2015). Second, given the explosive growth in R&D in China and the fact that the government has chosen to use this policy, it is important to understand the economic and fiscal consequences of this type of policy.

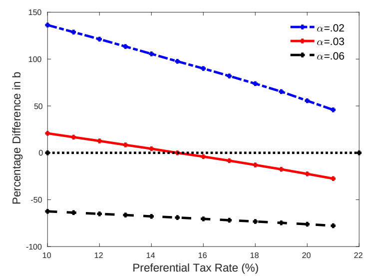
Figures 8-9 study the effects of changing the preferential tax rate for three values of the notch: 2%, 3%, and 6%. Each line shows the change in a given outcome from moving the preferential tax rate to between 10% and 22% for a given notch, relative to the current benchmark where $\alpha = 0.03$ and $t_2^{HT} = 15\%$.

Figure 8. : Simulated Counterfactual Policies: Selection and Relabeling

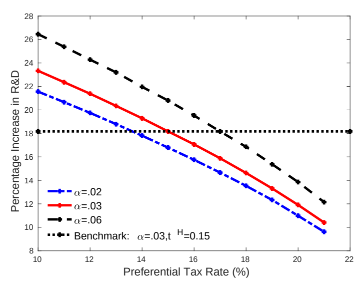
A. Mean ϕ_1 for Compliers Relative to Benchmark



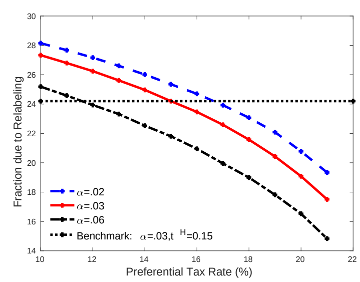
B. Mean b for Compliers Relative to Benchmark



C. Real R&D Increase for Compliers



D. Fraction due to Relabeling for Compliers

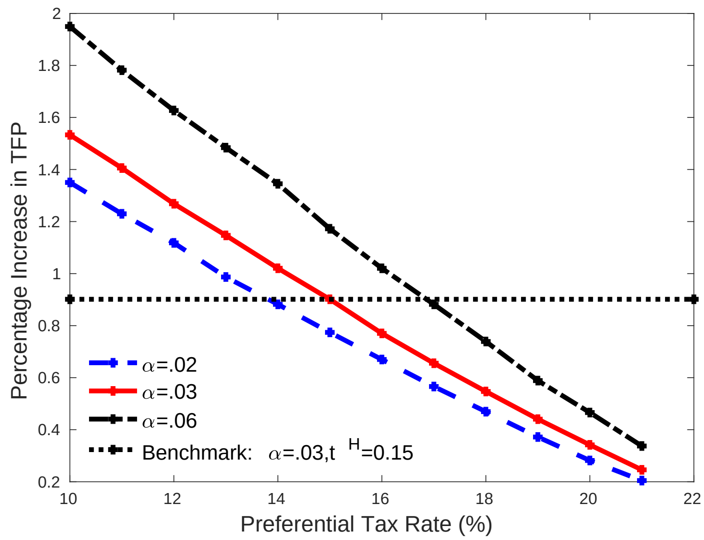


Note: These figures report the effects of different policy parameters on the selection of firms into the InnoCom program and on aggregate outcomes of interest. Panels A and B show that lower preferential tax rates select firms with higher adjustment costs and lower productivity. Panels C and D show how real and relabeled R&D respond to changes in parameters of the policy. See Section V for details on the structural model and the simulation.

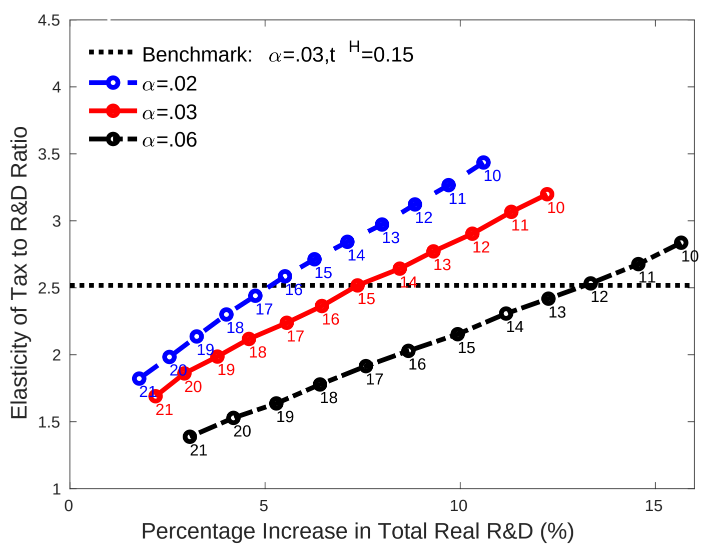
Panels A and B of Figure 8 analyze how changes in the policy parameters affect the characteristics of the compliers. We find that higher values of the notch lead to a selection of more productive firms and of firms with lower adjustment costs, on average. This graph also shows that, as we increase the tax break for high-tech firms (lower the preferential tax rate), the program selects firms with lower productivity and higher adjustment costs. The selection effect is more pronounced for adjustment costs than for productivity. For instance, when we change the threshold from 3% to 2%, the average adjustment cost for the compliers almost doubles, while the productivity is only around 2% lower. These results show that there are decreasing returns from expanding the InnoCom program by increasing the tax advantage and that a larger tax break might exacerbate misallocation of R&D by incentivizing R&D investment in firms with lower productivity and

Figure 9. : Simulated Counterfactual Policies: Productivity and Fiscal Cost of Stimulus

A. Average TFP Increase (Excluded Region)



B. Tax Revenue Cost of Stimulating Real R&D



Note: These figures report the effects of different policy parameters on aggregate outcomes of interest. Panel A shows how different reforms affect TFP. Panel B plots the elasticity of the tax cost to the government to the real R&D increase. This figure represents the fiscal cost curve of incentivizing R&D investment for the government and shows that notches that target larger firms have lower fiscal costs. See Section V for details on the structural model and the simulation.

higher adjustment costs.

Panels C and D of Figure 8 show how real R&D investment and relabeling respond to changes in the InnoCom program. Panel C shows that there is more real investment when firms face a lower preferential tax rate. However, the fraction of R&D due to relabeling also increases in the size of the tax cut. As panel D illustrates, when we set the notch threshold at 6%, moving the preferential tax rate from 21% to 10% increases the fraction of reported R&D attributable to relabeling by almost 10 percentage points.

Panel A of Figure 9 plots the average growth in productivity induced by the InnoCom program for firms in the excluded region. This effect is driven by two forces. First, as in panel C of Figure 8, complier firms invest more when the preferential tax rate is lower. Second, the fraction of firms that participate in the program also increases with a lower preferential tax rate. When $\alpha = 3\%$ and the preferential tax is reduced to 10%, the average firm sees a TFP increase of 1.4%. This is a larger increase than in the benchmark case, where firms see a 0.8% increase in TFP.

Finally, we use our simulations to answer the question: What is the lowest-cost policy for a government that wants to increase R&D by a given amount? To answer this question, we first estimate the elasticity of the tax revenue cost to the real increase in R&D investment for different values of α and t^{HT}. We then plot these ratios in panel B of Figure 9 according to the total increase in real R&D. This graph thus represents the cost frontiers for a government that wants to increase real R&D by a given amount. The current policy of $\alpha = 3\%$ and $t^{HT} = 15\%$ corresponds to a cost ratio of about 2.5. The black line shows that a policy defined by $\alpha = 6\%$ and $t^{HT} = 17\%$ would result in a similar increase in real R&D investment, but at a lower average cost. Alternatively, a policy defined by $\alpha = 6\%$ and a larger tax advantage $t^{HT} = 12\%$ would result in twice as large of an increase in R&D investment for a similar tax-to-R&D ratio. This result is driven by the fact that policies with larger α positively select more productive firms as well as firms with better technological opportunities. Nonetheless, as shown in panel D of Figure 8, policies with lower preferential tax rates invite relabeling.

These simulations show that the effectiveness of notch-based programs depends strongly on firm selection. Stronger incentives for R&D may misallocate R&D to firms with worse technological opportunities. Moreover, incentives that encourage R&D investment at the lowest cost to taxpayers may lead firms to engage in relabeling activities, which are likely socially undesirable.

B. R&D Tax Credit

A more common R&D subsidy policy is the R&D tax credit, which is prevalent in a large number of European and North American countries. We now use our estimated model to evaluate the effects of drastically changing the Chinese InnoCom program to an R&D tax credit system comparable to that of the US. While the US system has numerous accounting details, we define it by its two most fundamental features: the base amount $\bar{D}_i$ and the tax credit rate τ. The US government provides a credit of $\tau = 20\%$ for qualified R&D expenditures that exceed the base amount $\bar{D}_i$.[31]

If firms find it optimal to not misreport ($\delta^* = 0$), then the R&D tax credit effectively reduces the marginal cost of real R&D, D^K, by $(1 - t_1)\tau$. When there is no relabeling, an R&D tax credit is a relatively cheap way to induce incremental R&D investment. Indeed, the tax-to-R&D elasticity equals $(1-t_1)\tau \approx 0.15$, which is significantly more effective than the 2.5 elasticity of the benchmark InnoCom program. If we impose the estimated cost of relabeling of $\eta = 6.76$, as in our benchmark case, firms find it very costly to misreport and set $\delta^* = 0$. In this case, the R&D tax credit system is a superior policy.

In practice, however, the cost of relabeling η is likely to depend on the government's enforcement capability and on the number of firms that need to be monitored. Specifically, η is likely to be lower under an R&D tax credit system since the tax authority will need to audit *all* firms. This implies that individual firms will face lower costs of relabeling. With positive misreporting, the cost-effectiveness of the R&D credit quickly worsens. To see this, note that the R&D tax credit is calculated as:

$$(1 - t_1)\tau \left[\frac{D^{K*}}{1 - \delta^*} - D_1^*\right] \equiv (1 - t_1)\tau \left[(D^{K*} - D_1^*) + \frac{\delta^*}{1 - \delta^*} D^{K*}\right].$$

If firms relabel $\delta^* > 0$ of reported R&D, then the effective tax cost of inducing the marginal dollar of real R&D becomes $(1 - t_1)\tau[1 + \frac{\delta^*}{1-\delta^*}\frac{D^{K*}}{D^{K*}-D_1^*}]$. When the incremental real R&D, $D^K - D_1^*$, is small, the misreported R&D dominates the tax-to-real R&D elasticity. When we rescale the relabeling cost to match our benchmark relabeling of $\delta^* = 0.24$, our simulated model implies a tax-to-R&D elasticity of 4.13. This higher fiscal cost is largely driven by relabeling. Intuitively, firms were already at their interior optimum. The tax credit therefore induces mostly a relabeling response, with a very small increase in real R&D. In this case, the large relabeling response yields the surprising result that an InnoCom-style program is more effective at stimulating real R&D than a linear tax credit.

This analysis reveals that the choice of subsidy critically depends on the costs of relabeling. Using our model's estimates of firm-level R&D adjustment costs

[31]Since $\bar{D}_i$ typically depends on an average of R&D intensity in previous years, it is natural to assume that $\bar{D}_i = D_{i1}^*$, the interior optimum. We can thus set up the firm's optimal R&D decision problem as:

$$\max_{D^K,\delta}(1 - t_1)\left[\pi_1 - g(D^K, \theta\pi_1)\right] - D^K + t_1\left(\frac{D^K}{1 - \delta}\right) + (1 - t_1)\tau\left(\frac{D^K}{1 - \delta} - D_{i1}^*\right)$$
$$- \frac{D^K}{1 - \delta}h(\delta) + \beta(1 - t_2)E[\pi_2|D^K].$$

Note that the misreporting decision, δ, is separable from the real R&D choice, D^K. Thus, the optimal proportional evasion δ^* is determined by the evasion cost, η; the R&D tax credit, τ; and the corporate tax rate, t_1. Given the optimal evasion decision δ^*, firms choose real R&D amount D^K.

and returns to R&D, we search for the relabeling cost parameter that equalizes the fiscal cost of an R&D tax credit regime with the InnoCom program. We find that when we increase the evasion cost level such that it implies a lower fraction of relabeling of 13.85% (in contrast to 24.2% in our benchmark), the R&D tax credit policy achieves the same fiscal elasticity of 2.5. Therefore, a tax credit is a more cost-effective policy if the government can significantly increase the cost of relabeling. However, this may come at the cost of devoting additional government resources to detecting relabeling.

C. Welfare Implications

Governments often justify the use of fiscal incentives for R&D with the argument that innovative activities have positive spillover effects on the rest of the economy. When individual firms neglect these positive externalities, aggregate R&D investment may be lower than is socially optimal (see, e.g, Bloom, Van Reenen and Williams, 2019). We now study whether the InnoCom program can be justified as a tool to alleviate this market failure.

To consider this question, we extend our single-agent framework to an equilibrium setting and consider the aggregate implications of this policy.[32] As in Section III, individual firms i engage in monopolistic competition. Let C_t denote the CES composite good that is assembled from the output of all firms. Firm optimization implies that the price of the composite good in period t is given by $P_t = \frac{\theta}{\theta-1}\Phi_t^{-1}$, where $\Phi_t^{\theta-1} = \sum_i \exp\{(\theta-1)\phi_{i,t}\}$ is an aggregate measure of firms' log-productivity, $\phi_{i,t}$, and where θ denotes the constant elasticity of demand.

As in our empirical setting, we assume that a subset of firms, $N^{\text{R\&D}}$, engages in R&D.[33] We consider the role of spillovers by assuming that $\phi_{i,t}$ follows an expanded version of Equation 1:[34]

$$\phi_{i,t} = \rho\phi_{i,t-1} + \varepsilon \ln(D_{i,t-1}) + \zeta S_{t-1} + u_{it}, \quad \text{where} \quad S_{t-1} = \frac{1}{N^{\text{R\&D}}} \sum_{i=1}^{N^{\text{R\&D}}} \ln(D_{i,t-1}).$$

Past investments in R&D influence aggregate productivity Φ_t by directly increasing own-firm productivity as well as through potential spillovers effects when $\zeta > 0$.

[32]See Appendix L for detailed derivations. While previous analyses relied solely on individual firm decisions, the results in this section further assume that firms correctly anticipate the future prices implied by aggregate R&D.

[33]Table 2 shows that 8–10% of firms in our data engage in R&D. Since these firms are on average more productive, the sales share of the R&D sector is close to 35%.

[34]The evolution of log productivity for non-R&D-performing firms is similar but excludes the term $\varepsilon \ln(D_{i,t-1})$. In Appendix L, we show that, since the choice of $D_{i,t-1}$ is invariant to S_{t-1}, our previous analyses are not affected by the presence of spillover effects. For simplicity, we assume that S_{t-1} is a simple average of all R&D-performing firms; see Bloom, Van Reenen and Williams (2019) for a discussion of different weighted averages used in the empirical literature and Benhabib, Perla and Tonetti (2017); König et al. (2018) for a discussion of models of imitation and technology diffusion.

We consider a representative household that derives utility $C_t^{1-\gamma}G_t^{\gamma}$ from private consumption, C_t, and a public good, G_t. The household uses a per-period endowment L and after-tax firm profits to purchase C_t at price P_t. The government produces the public good G_t with a linear transformation of C_t, which is financed by taxing corporate profits.[35]

We now show how the InnoCom program impacts social welfare. To do so, we denote aggregate R&D expenditures gross of adjustment and fixed costs by:

$$D_1 = \sum_{i=1}^{N^{\text{R\&D}}} \left(D_{i,1} + g_i(D_{i,1}, \theta\pi_{i,1}) + \mathbb{I}(\text{InnoCom}_i)c_i\right),$$

where $\mathbb{I}(\text{InnoCom}_i)$ is an indicator for the event that firm i is in the InnoCom program. Similarly, $H_1 = \sum_i \mathbb{I}(\text{InnoCom}_i)h(D_{i,1}, \tilde{D}_{i,1})$ denotes aggregate relabeling costs and

$$\tau = \frac{(t^{LT} - t^{HT})\sum_i \mathbb{I}(\text{InnoCom}_i)\pi_{i,2}}{\sum_i \pi_{i,2}}$$

is the fiscal cost of the InnoCom program relative to aggregate profits. Social welfare is then:

$$\text{(6)} \quad \Phi_1(L - D_1 - H_1)\left(1 - \frac{t}{\theta}\right)^{1-\gamma}\left(\frac{t}{\theta}\right)^{\gamma} + \beta\Phi_2 L\left(1 - \frac{t}{\theta} + \frac{\tau}{\theta}\right)^{1-\gamma}\left(\frac{t}{\theta} - \frac{\tau}{\theta}\right)^{\gamma}.$$

Welfare in each period combines three factors. Welfare increases with Φ_t since higher productivity lowers the price of the composite good. Welfare is also increasing in the resources expended in a given period. Finally, welfare depends on the allocation of resources between private and public consumption.[36]

Equation 6 presents a welfare accounting of the costs and benefits of the InnoCom program. First, the InnoCom program lowers first-period spending by $D_1 + H_1$. Second, the fiscal cost of the InnoCom program raises the share of private consumption by $\frac{\tau}{\theta}$ at the expense of the public good. Finally, by increasing R&D investment, the InnoCom program raises Φ_2. This last effect is more pronounced when R&D has spillover effects on the productivity of other firms, i.e., $\zeta > 0$.

We calibrate three additional parameters to implement Equation 6. First, we use the fact that, in the absence of the InnoCom program, the tax rate $t = \gamma\theta$ maximizes social welfare. We thus set $\gamma = \frac{t}{\theta} = \frac{25\%}{5} = 5\%$, which is the value

[35] This setup builds on Samuelson (1954); Atkinson and Stern (1974) by incorporating a productive use of government funds that justifies the existing corporate tax rate. Corporate tax cuts would be trivially beneficial if tax revenue is not used for productive purposes, i.e., when $\gamma = 0$. Our results are robust to assuming that government production wastes a constant fraction of its budget, so that $G_t = (1 - \text{waste})C_t$.

[36] In the first period, the private expenditure share is $1 - \frac{t}{\theta}$, and the public goods share is $\frac{t}{\theta}$. In the case without the InnoCom program, where the optimal tax is given by $t = \theta\gamma$, the consumption mix of Equation 6 takes the familiar Cobb-Douglas form $(1-\gamma)^{1-\gamma}(\gamma)^{\gamma}$.

of γ that rationalizes the observed tax rate.[37] Second, we normalize L to equal payments to labor implied by our model. Finally, we calibrate the importance of the R&D sector such that the aggregate sales share of R&D-performing firms matches the share observed in our data.

We start by using Equation 6 to evaluate the welfare loss in the case where $\zeta = 0$. In the absence of spillover effects, the InnoCom program leads to (1) firm costs related to certification, compliance, and relabeling, (2) over-investment in R&D—a form of inter-temporal distortion—and (3) under-provision of public goods, which distorts the consumption mix. Our model estimates imply that welfare decreases by 0.14% in this case. While all three channels contribute to the welfare loss, the first channel has the largest effect. With the InnoCom program, aggregate efficiency, Φ_2/Φ_1, improves by 0.12% more than without the program. We find that, because the consumption loss from the additional R&D spending (including adjustment costs) increases by slightly more than 0.13%, the program has a very small inter-temporal distortion. The welfare loss of 0.14% is almost completely accounted for by the increase in certification and relabeling costs. The intuition is that, while channels (2) and (3) transfer resources across time or types of consumption, certification and relabeling costs are unproductive uses of resources.[38]

We now use Equation 6 to find the value $\hat{\zeta}$ such that the InnoCom program for large firms (i.e., $\alpha = 0.03, t^{HT} = 0.15$) yields the same welfare as the case without the InnoCom program. Using our estimated model, we find that $\hat{\zeta} = 0.069$. This value of $\hat{\zeta}$ implies that a firm's log-productivity would increase by 6.9% if all R&D firms doubled their R&D investment. When spillovers are small, i.e., $\zeta < 0.069$, the distortions discussed above exceed the gains from incentivizing R&D.

The InnoCom program can be justified from a welfare perspective as long as spillover effects are larger than $\hat{\zeta}$.[39] Compared to empirical estimates, $\hat{\zeta}$ is relatively small. For instance, Bloom, Mark and John (2013) and Lucking, Bloom and Van Reenen (2019) estimate significantly larger values of $\zeta \approx 0.20$. When we set $\zeta = 0.20$, we find that welfare increases by 0.27% and aggregate productivity increases by 0.53%. These results suggests that the InnoCom program may be a valuable policy tool to alleviate the under-investment in R&D.

The results of our policy simulations highlight the promises and limitations of an InnoCom-style program. Section V.B shows that such a program may be more

[37]Estimates of γ in the US range from 0.11 to 0.26 (Suárez Serrato and Wingender, 2014; Fajgelbaum et al., 2018).

[38]Slemrod (2006) discusses the compliance costs of business taxes and argues that compliance costs should be incorporated in welfare analyses of tax systems.

[39]Our framework makes three implicit assumptions that imply that our estimate of $\hat{\zeta}$ is a conservatively high value. First, our static model implies that firms expect an instant and sizable equilibrium price response to R&D tax policy, which may depress R&D investment. Second, by holding L constant, our model assumes that Φ_t is the only source of gains. Finally, in contrast to empirical approaches to estimating ζ that condition spillover pools on geographic or technological distance, we assume a broad spillover pool that includes all firms. Models with a more rigid equilibrium price response, where income can increase in response to productivity growth, or with narrower spillover pools would all imply a lower value of $\hat{\zeta}$.

effective than a linear tax credit at stimulating R&D investment when relabeling is a significant concern. While Section V.C shows that an InnoCom-style program can be justified from a welfare perspective under moderate spillover effects, the simulations in Section V.A also reveal the limits of this approach. Specifically, Figure 8 shows that the potential to scale-up InnoCom-style programs is limited by the fact that more generous tax credits or more accessible notches draw in firms that are less productive and that have higher adjustment costs, which exacerbates the prevalence of relabeling.

VI. Conclusions

Governments around the world devote considerable tax resources to incentivizing R&D investment. However, there is widespread concern that firms respond by relabeling other expenses as R&D expenditures. This paper takes advantage of a large fiscal incentive and detailed administrative tax data to analyze these margins in the important case of China. We provide striking graphical evidence consistent with both large reported responses and significant scope for relabeling. These results suggest misreporting of R&D may contaminate estimates of the effectiveness of R&D investment and may lead to misallocation of R&D toward firms with less innovative projects.

Optimal subsidies for R&D depend on the fiscal cost for the government and the potential positive externalities of R&D investment on other firms' productivity. We provide a useful metric that traces the government's trade-off between own-firm productivity growth and tax revenues. We also provide a bound on the size of the externality that would justify this government intervention.

Finally, while we find evidence consistent with relabeling, the unusual structure of the InnoCom program, characterized by pre-registration and auditing, may limit the scope of relabeling and evasion. In contrast, R&D investment tax credits may be more susceptible to relabeling in developing and even developed countries. As this paper demonstrates, accounting for relabeling has important implications for the design of R&D subsidies.

REFERENCES

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83(6): 2411–2451.

Agrawal, Ajay, Carlos Rosell, and Timothy S Simcoe. 2019. "Tax Credits and Small Firm R&D Spending." *American Economic Journal: Economic Policy, forthcoming.*

Akcigit, Ufuk, John Grigsby, Tom Nicholas, and Stefanie Stantcheva. 2018. "Taxation and Innovation in the 20th Century." National Bureau of Economic Research Working Paper 24982.

Almunia, Miguel, and David Lopez-Rodriguez. 2018. "Under the Radar: The Effects of Monitoring Firms on Tax Compliance." *American Economic Journal: Economic Policy*, 10(1): 1–38.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2017. "Measuring the Sensitivity of Parameter Estimates to Estimation Moments*." *The Quarterly Journal of Economics*, 132(4): 1553–1592.

Atkinson, A. B., and N. H. Stern. 1974. "Pigou, Taxation and Public Goods12." *The Review of Economic Studies*, 41(1): 119–128.

Bachas, Pierre, and Mauricio Soto. 2019. "Not(ch) Your Average Tax System: Corporate Taxation Under Weak Enforcement." World Bank Policy Research Working Paper 8524.

Becker, Bettina. 2015. "Public R&D Policies and Private R&D Investment: A Survey of the Empirical Evidence." *Journal of Economic Surveys*, 29(5): 917–942.

Benhabib, Jess, Jesse Perla, and Christopher Tonetti. 2017. "Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier." National Bureau of Economic Research Working Paper 23095.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem. 2015. "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan." *Journal of Political Economy*, 123(6): 1311 – 1355.

Blinder, Alan S., and Harvey S. Rosen. 1985. "Notches." *The American Economic Review*, 75(4): 736–747.

Blomquist, Sören, and Whitney Newey. 2017. "The Bunching Estimator Cannot Identify the Taxable Income Elasticity." National Bureau of Economic Research Working Paper 24136.

Bloom, Nicholas, John Van Reenen, and Heidi Williams. 2019. "A Toolkit of Policies to Promote Innovation." *Journal of Economic Perspectives*, 33(3): 163–84.

Bloom, Nicholas, Schankerman Mark, and Van Reenen John. 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica*, 81(4): 1347–1393.

Cai, Jing, Yuyu Chen, and Xuan Wang. 2018. "The Impact of Corporate Taxes on Firm Innovation: Evidence from the Corporate Tax Collection Reform in China." National Bureau of Economic Research Working Paper 25146.

Dechezlepretre, A., E. Einio, R. Martin, K.-T. Nguyen, and J. Van Reenen. 2016. "Do Fiscal Incentives increase innovation? An RDD for R&D." CEP Discussion Paper 1413.

Devereux, Michael P., Li Liu, and Simon Loretz. 2014. "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records." *American Economic Journal: Economic Policy*, 6(2): 19–53.

Diamond, Rebecca, and Petra Persson. 2016. "The Long-term Consequences of Teacher Discretion in Grading of High-Stakes Tests." Stanford University.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2017. "Bunching at the kink: Implications for spending responses to health insurance contracts." *Journal of Public Economics*, 146: 27 – 40.

Einiö, Elias. 2014. "R&D subsidies and company performance: Evidence from geographic variation in government funding based on the ERDF population-density rule." *Review of Economics and Statistics*, 96(4): 710–728.

Eisner, Robert, Steven H Albert, and Martin A Sullivan. 1984. "The new incremental tax credit for R&D: incentive or disincentive?" *National Tax Journal*, 171–183.

Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar. 2018. "State Taxes and Spatial Misallocation." *The Review of Economic Studies*, 86(1): 333–376.

GAO, United States. 2009. "The Research Tax Credit's Design and Administration Can Be Improved."

Guceri, Irem, and Li Liu. 2019. "Effectiveness of Fiscal Incentives for R&D: Quasi-experimental Evidence." *American Economic Journal: Economic Policy*, 11(1): 266–91.

Hall, Bronwyn, and John Van Reenen. 2000. "How effective are fiscal incentives for R&D? A review of the evidence." *Research Policy*, 29: 449–469.

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

Head, K., and T. Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." *Handbook of International Economics, Vol. 4.*

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics.*

Kleven, Henrik Jacobsen, and Mazhar Waseem. 2013. "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan." *The Quarterly Journal of Economics.*

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica*, 79(3): 651–692.

König, Michael, Kjetil Storesletten, Zheng Song, and Fabrizio Zilibotti. 2018. "From Imitation to Innovation: Where is all that Chinese R&D Going?"

Liu, Li, Ben Lockwood, Miguel Almunia, and Eddy Tam. 2019. "VAT Notches, Voluntary Registration, and Bunching." *Review of Economics and Statistics, forthcoming.*

Lockwood, Ben. 2018. "Malas Notches." *working paper.*

Lucking, Brian, Nicholas Bloom, and John Van Reenen. 2019. "Have R&D Spillovers Declined in the 21st Century?" *Fiscal Studies*, 40(4): 561–590.

Mansfield, Edwin, and Lorne Switzer. 1985. "The effects of R&D tax credits and allowances in Canada." *Research Policy*, 14(2): 97–107.

NBS. 2006-2007. "Annual Survey on Industrial Firms." National Bureau of Statistics.

Notowidigdo, Matthew. 2019. "The Incidence of Local Labor Demand Shocks." *Journal of Labor Economics, forthcoming.*

Patel, Elena, Nathan Seegert, and Matthew Grady Smith. 2016. "At a Loss: The Real and Reporting Elasticity of Corporate Taxable Income." SSRN: 2608166.

Rao, Nirupama. 2016. "Do Tax Credits Stimulate R&D Spending? The Effect of the R&D Tax Credit in its First Decade." *Journal of Public Economics*, 140.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *The Review of Economics and Statistics*, 36(4): 387–389.

SAT. 2008-2011. "Chinese State Administration of Tax." Enterpreise Tax Survey Dataset.

Slemrod, Joel. 2001. "A general model of the behavioral response to taxation." *International Tax andPublic Finance*, 8(2): 119–128.

Slemrod, Joel. 2006. "The (Compliance) Cost of Taxing Business." University of Michigan Working Paper.

Slemrod, Joel. 2013. "Buenas Notches: Lines and Notches in Tax System Design." *eJournal of Tax Research*, 11(3): 259–283.

Slemrod, Joel, and Christian Gillitzer. 2013. *Tax Systems.* MIT Press.

Suárez Serrato, Juan Carlos, and Philippe Wingender. 2014. "Estimating the Incidence of Government Spending." Mimeo.

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*, 49–70.

World Bank. 1995-2011. "World Bank Open Data." Research and development expenditure (% of GDP), Series GB.XPD.RSDV.GD.ZS.