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

Zhao Chen
Fudan University

Zhikuo Liu
Fudan University

Juan Carlos Suárez Serrato
Duke University
& NBER

Daniel Yi Xu
Duke University
& NBER

February 2021

Abstract

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 Codes: D24, O30, H25, H26.
Keywords: R&D, China, relabeling, corporate taxes

*We are very grateful for 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.

1 Introduction

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. 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 2 describes this fiscal incentive and discusses the potential for relabeling of R&D.

We begin our analysis in Section 3 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 4. 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 5. 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 6, 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

[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).

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]

2 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

[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.

[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 sales $> 50M$ & 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

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

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

[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.

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" 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.

[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.

[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 undercounted R&D expenses. In Section 5, 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 3, we show that sales are not manipulated around the R&D thresholds. Similarly, we do not find evidence of manipulation of other expenses.

3 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 4 and inform the structural estimation in Section 5.

3.1 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 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,

[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).

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.

3.2 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]

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.

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

[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%.

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. $\hat{c}_j = \sum\limits_{k=0}^{p} \hat{\beta}_k \cdot (d)^k$ 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]

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 B 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 C and D 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 E 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

[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 4.

[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.

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.

3.3 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]

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

[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.

[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" 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.

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.

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

[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.

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).

4 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.

4.1 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).

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 2). Since firms commit to

[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.

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 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.

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

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

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:

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

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:

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

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]

4.2 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:

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

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:

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

[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 3.

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

[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.

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:

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

Comparing Equation 4 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 follows that firms decide to bunch when the following inequality holds:

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

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.

5 Structural Estimation

The previous section described a model motivated by the data patterns in Section 3. 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.

5.1 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 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 4). Based on these firm-level decisions, we compute data moments that are analogous to those discussed in Section 3. 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:

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

[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.

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 3. 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 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

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

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.

5.2 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 3.2, 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.

5.3 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 2, 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.

6 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.

6.1 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 $t_2^{HT} = 15\%$.

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 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.

6.2 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

[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.

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 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.

6.3 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 4, 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.

[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.

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 $\zeta > 0$.

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:

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

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]

[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.

[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 $(1-\gamma)^{1-\gamma} (\gamma)^{\gamma}$.

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 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,

[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

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 6.2 shows that such a program may be more effective than a linear tax credit at stimulating R&D investment when relabeling is a significant concern. While Section 6.3 shows that an InnoCom-style program can be justified from a welfare perspective under moderate spillover effects, the simulations in Section 6.1 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.

7 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.

spillover pools would all imply a lower value of $\hat{\zeta}$.

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.

Auerbach, Alan J. 1984. "Taxes, firm financial policy and the cost of capital: An empirical analysis." *Journal of Public Economics*, 23(1-2): 27–57.

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. 2009. "The Impact of Uncertainty Shocks." *Econometrica*, 77(3): 623–685.

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.

Chen, Zhao, Xian Jiang, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Xu. 2019. "Tax Policy and Lumpy Investment Behavior: Evidence from China's VAT Reform." National Bureau of Economic Research Working Paper 26336.

Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman. 2009. "Power-law distributions in empirical data." *SIAM review*, 51(4): 661–703.

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.

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

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

Head, Keith, Thierry Mayer, and Mathias Thoenig. 2014. "Welfare and Trade without Pareto." *American Economic Review*, 104(5): 310–16.

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?"

Kratz, Marie, and Sidney I. Resnick. 1996. "The qq-estimator and heavy tails." *Communications in Statistics. Stochastic Models*, 12(4): 699–724.

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.

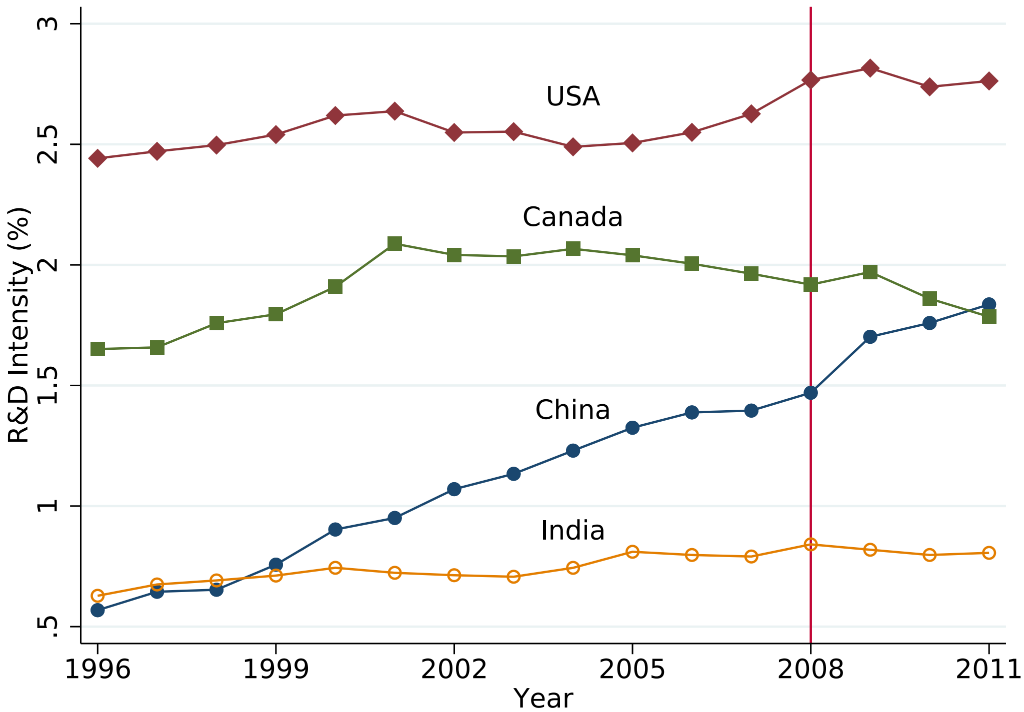
Suárez Serrato, Juan Carlos, and Philippe Wingender. 2014. "Estimating the Incidence of Government Spending." Mimeo.

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.

Wilson, Daniel. 2009. "Beggar Thy Neighbor? The In-State, Out-of-State, and Aggregate Effects of R&D Tax Credits." *The Review of Economics and Statistics*, 91(2): 431–436.

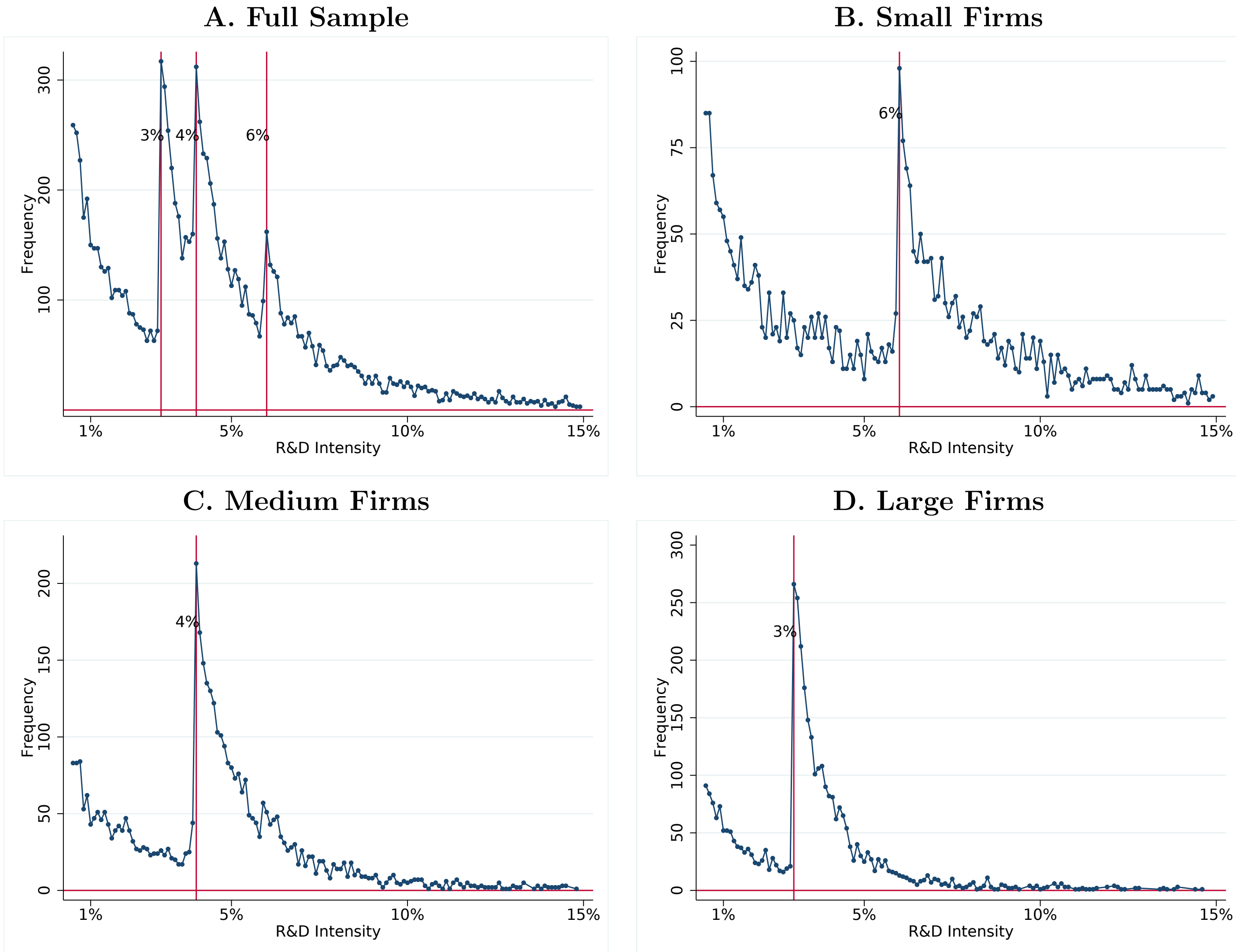
World Bank. 1995-2011. "World Bank Open Data." Research and development expenditure (% of GDP), Series GB.XPD.RSDV.GD.ZS.

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



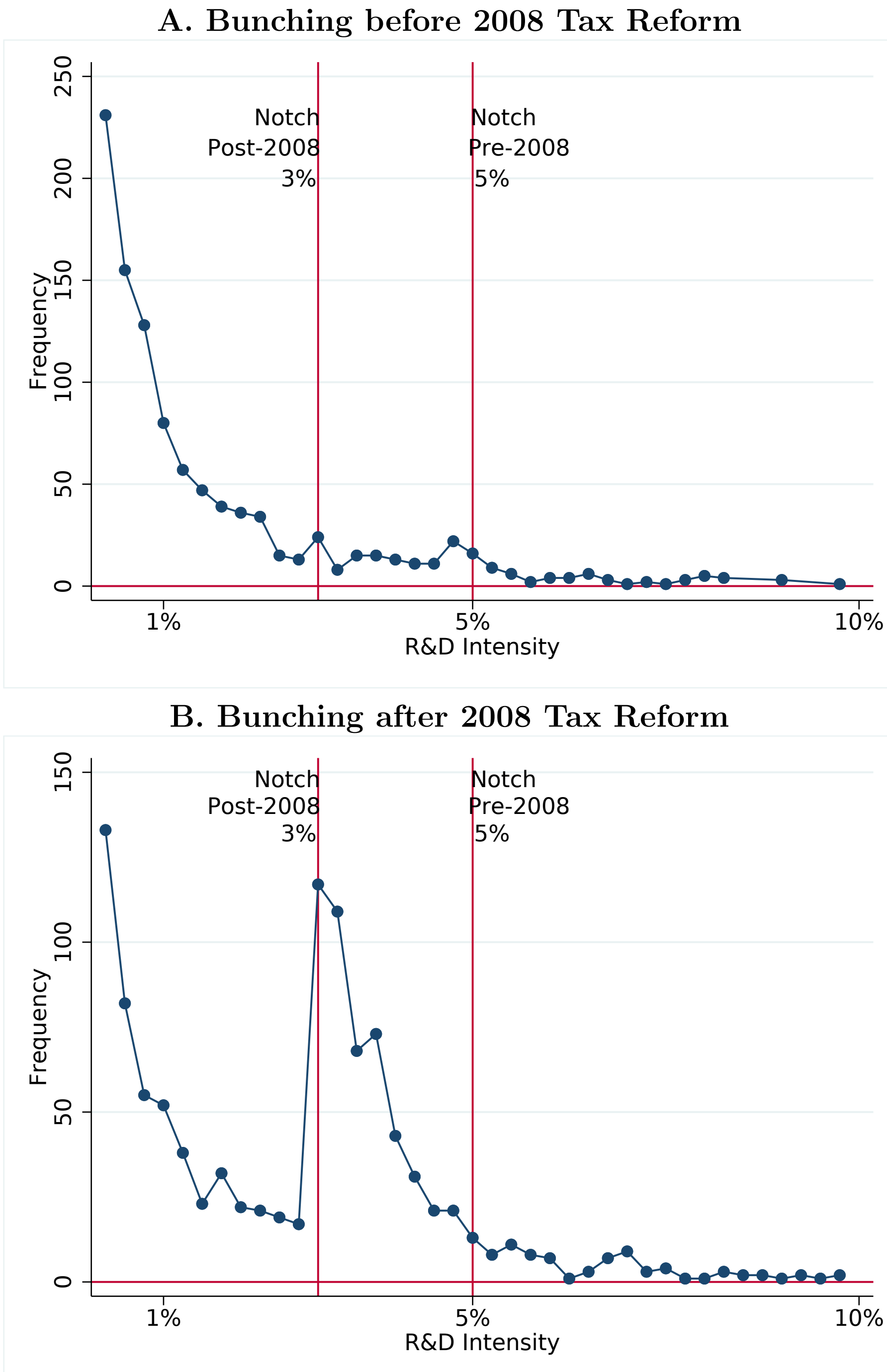
Notes: This figure plots the aggregate R&D Intensity, i.e., R&D expenditure as a share of GDP, in the private sector for China, Canada, India, and the US. Chinese R&D intensity started in 1996 at 0.5%, a similar level to India. It increased dramatically, by more than threefold, to above 1.5% in 2011, 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).

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



Notes: 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 3.1 for details.

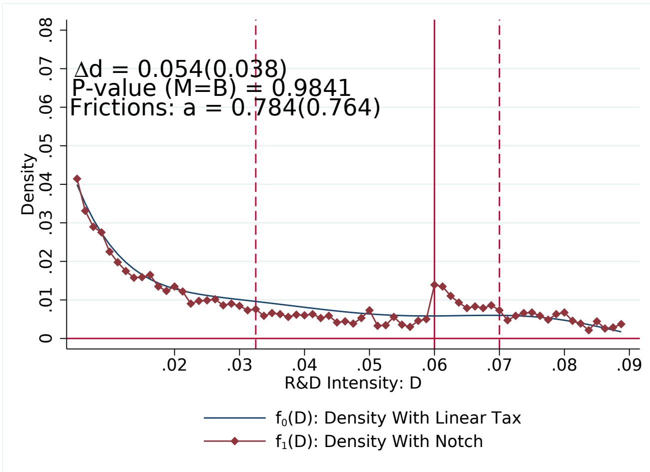
Figure 3: Effects of the 2008 Tax Reform on the Bunching of Foreign-Owned Large Companies



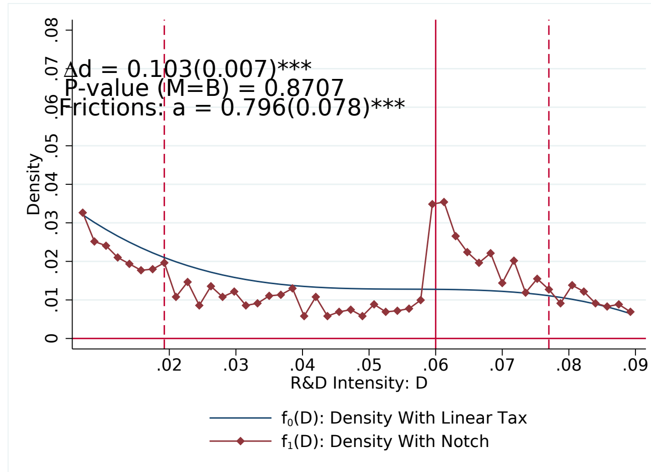
Notes: This figure compares the R&D intensity distribution for large foreign-owned firms before and after the 2008 tax reform. To make the two samples comparable, the figure plots only firms that we observe in both the SAT and ASM data. The tax reform eliminated the preferential corporate income tax 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 3.1 for details.

Figure 4: Estimated Counterfactual Densities of R&D Intensity

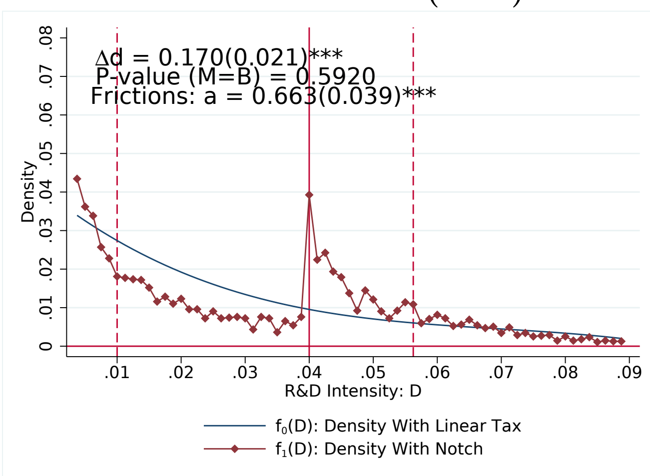
A. Small Firms (2009)



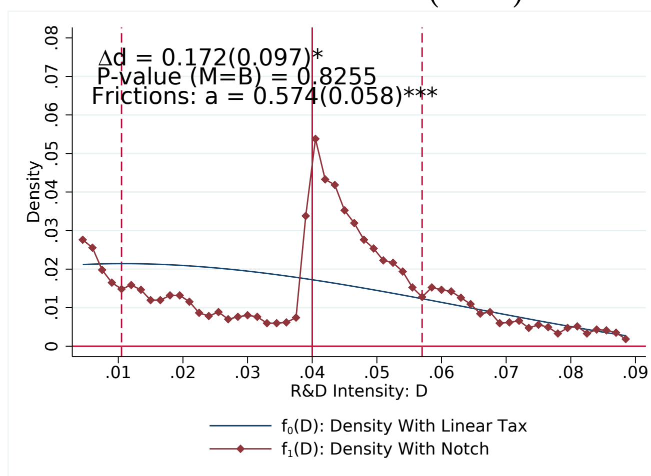
B. Small Firms (2011)



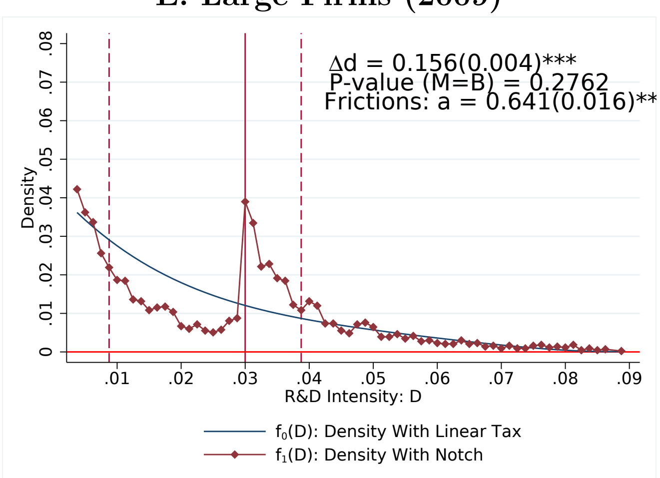
C. Medium Firms (2009)



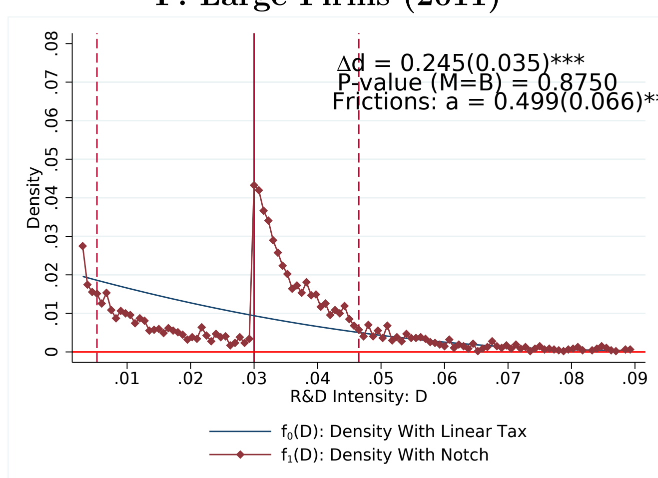
D. Medium Firms (2011)



E. Large Firms (2009)

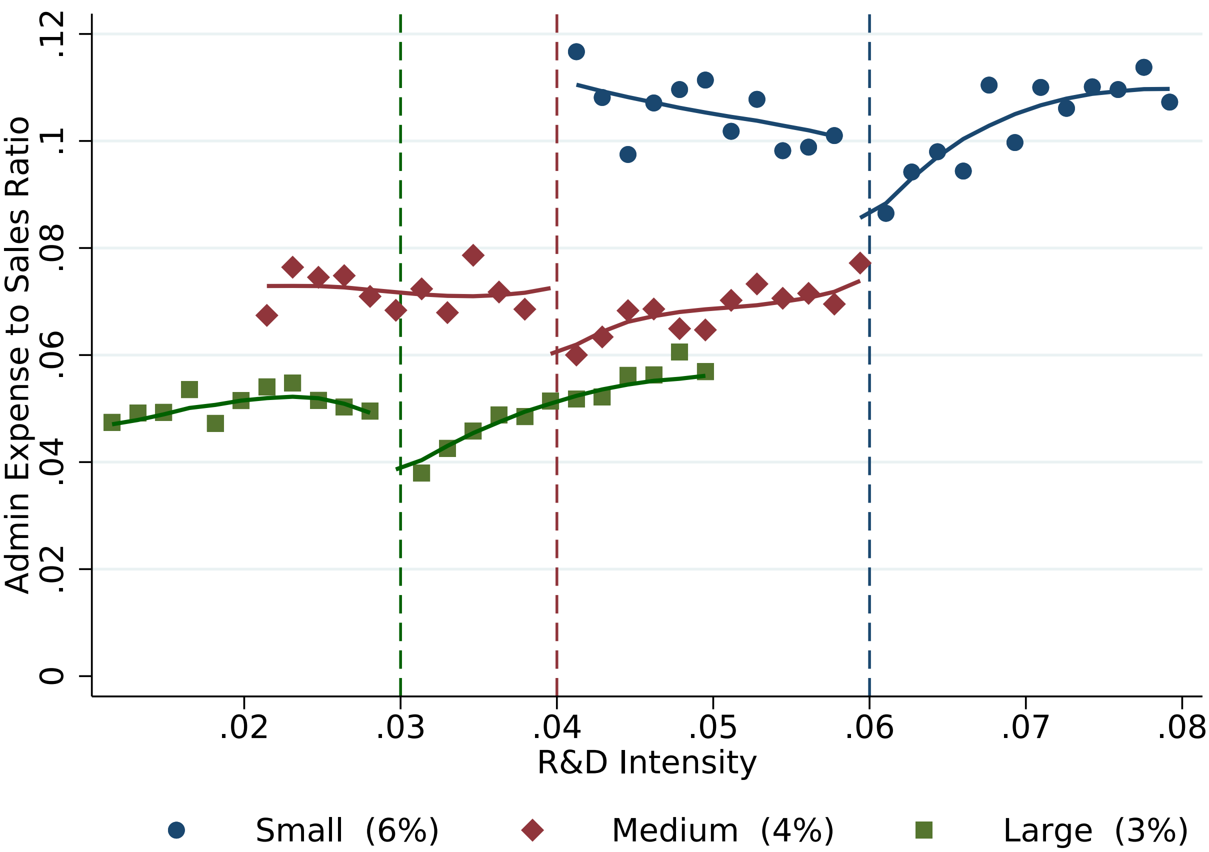


F. Large Firms (2011)



Notes: 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 3.2 for details. Source: Administrative Tax Return Database.

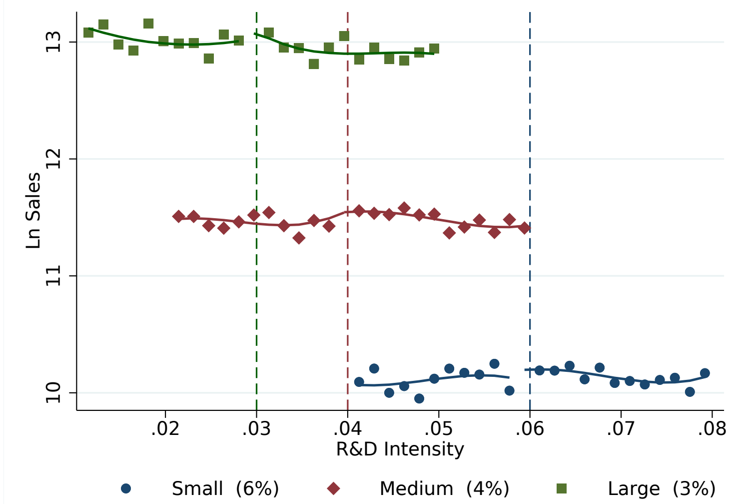
Figure 5: Empirical Evidence of Relabeling



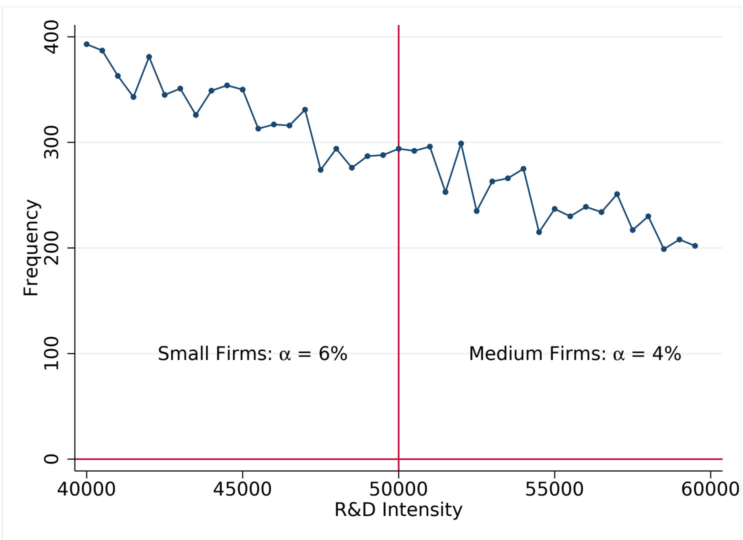
Notes: This figure plots the non-R&D administrative expense-to-sales ratio at each level of R&D intensity. The green dots/line are for the large firms, the red dots/line are for the medium firms, and the blue dots/line are for the small firms. The threshold of R&D intensity for firms to qualify for InnoCom 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 3 for details.

Figure 6: Lack of Sales Manipulation

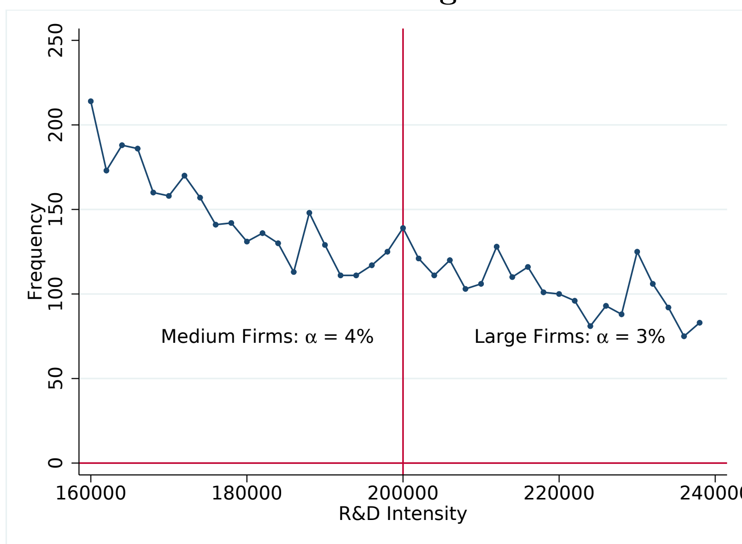
A. Lack of Sales Manipulation around R&D Intensity Thresholds



B. Lack of Firm Size Manipulation
Small and Medium Firms

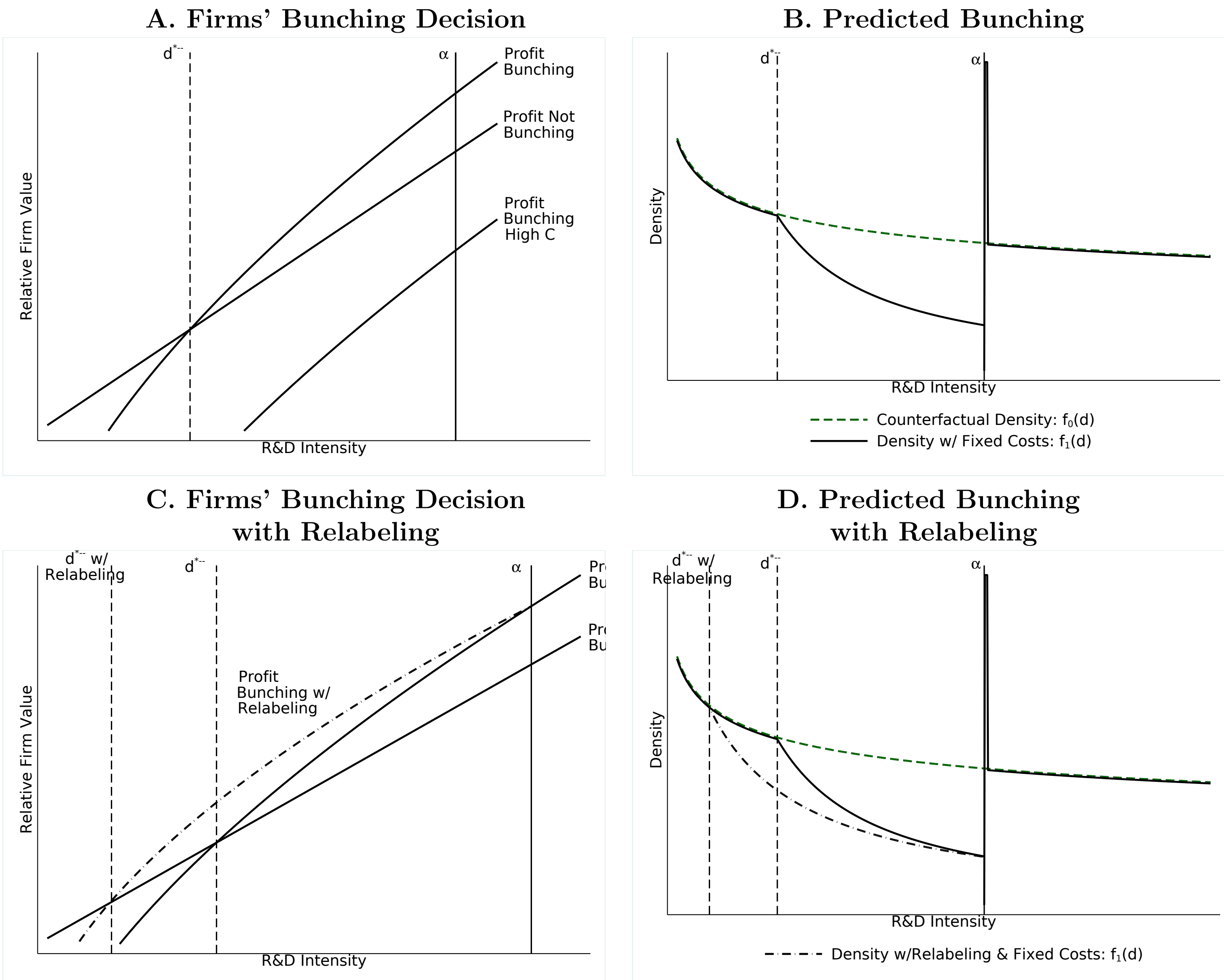


C. Lack of Firm Size Manipulation
Medium and Large Firms



Notes: 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 3 for details.

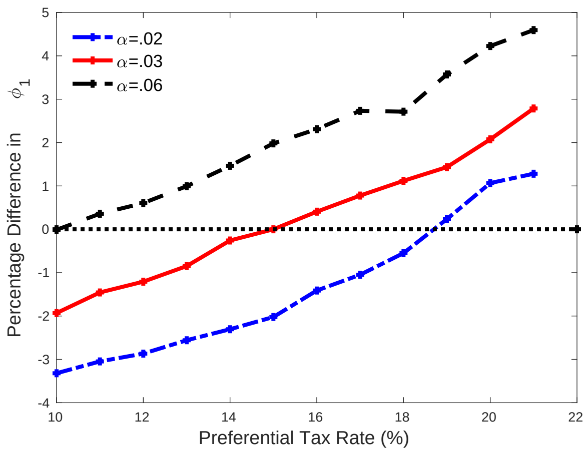
Figure 7: Theoretical Bunching Predictions



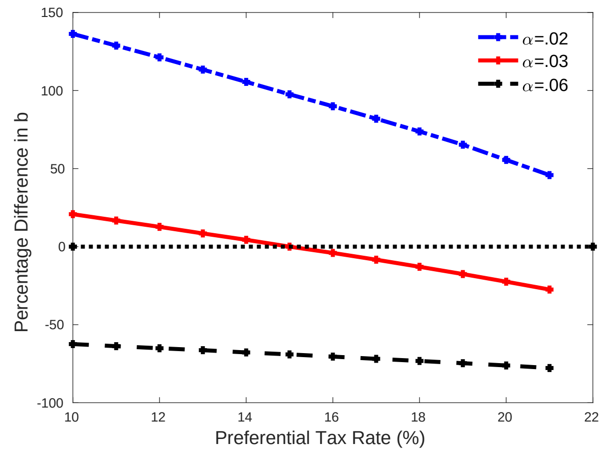
Notes: This figure provides intuition for when a firm decides to bunch and describes empirical implications of our model for R&D investment and bunching. Panel A visualizes Equation 3 by plotting the relative value from bunching, $\left(\frac{d_{i1}^*}{\alpha}\right)^{1-(\theta-1)\varepsilon}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\frac{1}{(\theta-1)\varepsilon}-1-c$, 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 4 for details.

Figure 8: Simulated Counterfactual Policies: Selection and Relabeling

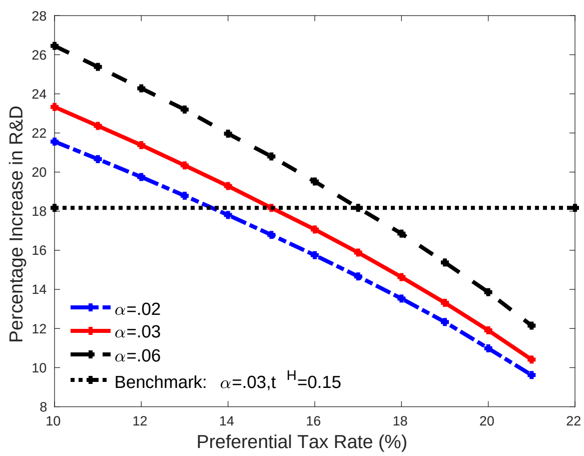
A. Mean ϕ_1 for Compliers Relative to Benchmark



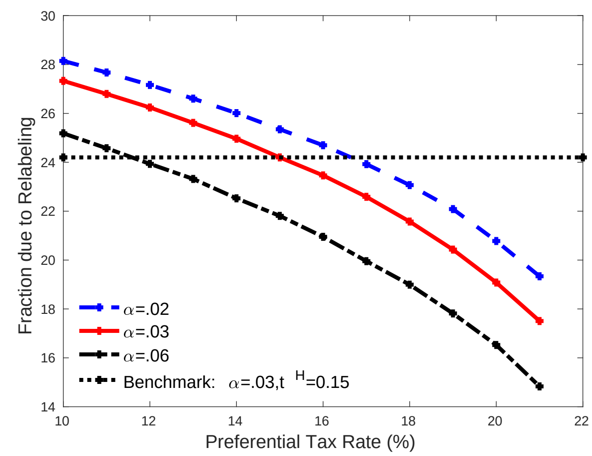
B. Mean b for Compliers Relative to Benchmark



C. Real R&D Increase for Compliers



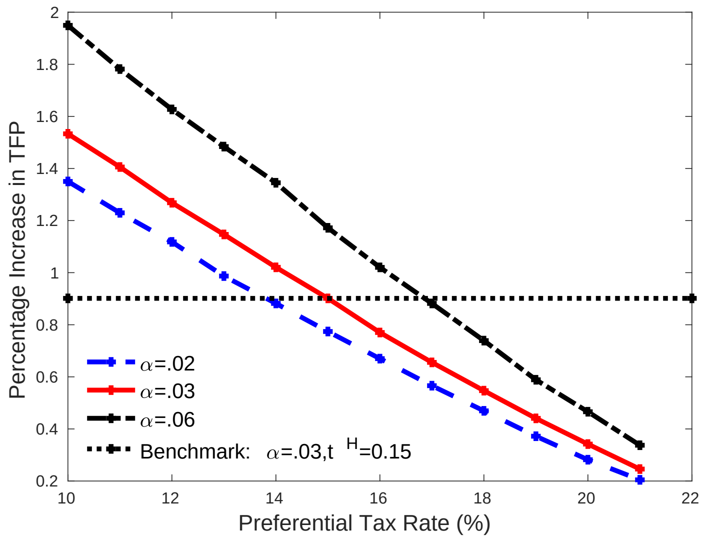
D. Fraction due to Relabeling for Compliers



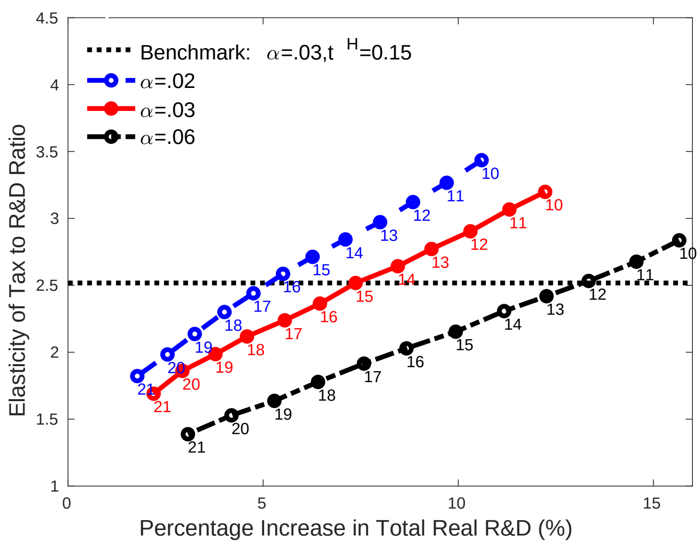
Notes: These figures report the effects of different policy parameters on the selection of firms into the InnoCom program and on aggregate outcomes of interest. Panels A and B show that lower preferential tax rates select firms with higher adjustment costs and lower productivity. Panels C and D show how real and relabeled R&D respond to changes in parameters of the policy. See Section 6 for details on the structural model and the simulation.

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



Notes: 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 6 for details on the structural model and the simulation.

Table 2: Descriptive Statistics

A. State Administration of Tax Data 2008–2011

	Mean	Std	p25	p50	p75	Observations
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	Observations
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 3.1 for details.

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)

Notes: 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 5 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

Notes: 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.

Online Appendix: Not For Publication

This appendix contains multiple additional analyses. Appendix A includes additional details on the Chinese corporate income tax system. Appendix B describes in more detail the data that we use in our analysis. Appendix C discusses the estimation of our measure of log TFP. Appendix D discusses details of the implementation of the bunching estimator. Appendix E discusses additional robustness checks of our bunching estimates. Appendix F describes details of the implementation of the estimator of Diamond and Persson (2016). Appendix G shows that firms do not respond to the InnoCom program by manipulating sales expenses. Appendix H provides a detailed derivation of the model. Appendix I links the model to more traditional bunching estimates used in the public finance literature. Appendix J provides additional details behind the structural estimation. Appendix K explores the robustness of our structural estimation. Finally, Appendix L provides details on the welfare implications of the program.

A Additional Details on the Chinese Corporate Income Tax System

China had a relatively stable Enterprise Income Tax (EIT) system in the early part of our sample, from 2000 to 2007. During that period, the EIT ran on a dual-track tax scheme with the base tax rate for all domestic-owned enterprises (DOE) at 33% and that for foreign-owned enterprises (FOE) ranging from 15% to 24%. The preferential treatment of FOEs has a long history dating to the early 1990s, when the Chinese government started to attract foreign direct investment in the manufacturing sector. The government offered all new FOEs located in the Special Economic Zones (SEZ) and Economic and Technology Development Zones (ETDZ) a reduced EIT of 15%. It also offered a reduced EIT of 24% for all FOEs located in urban centers of cities in the SEZs and ETDZs. The definition of foreign-owned is quite broad: it includes enterprises owned by Hong Kong, Macau, and Taiwan investors. It also includes all joint-venture firms with a foreign share of equity larger than 25%. The effective tax rates of FOEs are even lower since most had tax holidays that typically left them untaxed for the first 2 years and then halved their EIT rate for the subsequent 3 years.

In addition to the special tax treatments of FOEs, the Chinese government started the first round of the West Development program in 2001. Both DOEs and FOEs that are located in West China and are part of state-encouraged industries enjoy a preferential tax rate of 15%. West China is defined as the provinces of Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Ningxia, Qinghai, Xinjiang, Inner Mongolia and Guangxi. Finally, there is also a small and medium enterprise tax break, which is common in other countries. However, the revenue threshold is as low as $\$50,000$, making this tax break effectively irrelevant for our sample.

The Chinese government implemented a major corporate tax reform in 2008 to eliminate the dual-track system based on domestic/foreign ownership and established a common rate of 25%. Some of the existing tax breaks for FOEs were gradually phased out. For instance, FOEs that had previously paid an EIT of 15% paid a tax rate of 18% in 2008, 20% in 2009, 22% in 2010, and 24% in 2011. In contrast, the West Development program will remain in effect through 2020.[40]

B Data Sources

We connect two large firm-level databases of Chinese manufacturing firms. The first is the relatively well-studied Chinese Annual Survey of Manufacturing (ASM), an extensive yearly survey of Chinese manufacturing firms. The ASM is weighted towards medium and large firms and includes all Chinese manufacturing firms with total annual sales of more than 5 million RMB (approximately $800,000) as well as additional state-owned firms with lower sales. This survey provides detailed information on the ownership, location, production, and balance sheet of manufacturing firms. This dataset allows us to measure total firm production, sales, inputs, and, for a few years, detailed skill composition of the labor force. We supplement these data with a separate Chinese National Bureau of Statistics survey that includes firms' reported R&D. We use these data for the years 2006–2007.

The second dataset that we use is the administrative enterprise income tax records from the Chinese State Administration of Tax (SAT). The SAT is the counterpart of the IRS in China and is in charge of tax collection and auditing. In addition, the SAT supervises various tax assistance programs such as the InnoCom program. The SAT keeps its own firm-level records of tax payments as well other financial statement information used in tax-related calculations. We acquire these administrative enterprise income tax records for 2008–2011, which allows us to construct detailed tax rate information for individual manufacturing firms. Our main sample of analysis includes firms with a positive tax liability. We also use these data to construct residualized measures of firm productivity.[41] The scope of the SAT data is slightly different from that of the ASM data, but there is a substantial amount of overlap for the firms that conduct R&D. For instance, the share of total R&D that can be matched with ASM records is close to 85% for 2008.

[40] After the phase-out of the FOE preferential tax treatment, InnoCom became the largest preferential tax program based on the EIT in China. From a financial accounting perspective, firms in China can potentially choose to expense or amortize R&D expenses. This is consistent with generally accepted accounting principles (GAAP). However, while Chinese accounting principles allow firms in research-intensive industries to book a restricted fraction of "development R&D" as an intangible asset and to amortize it over time, R&D expenditures are predominantly expensed. More importantly, from a tax accounting perspective, the Chinese State Administration of Tax (SAT) defines the taxable profit based on *total R&D expenditure*, regardless of whether it was amortized or expensed in financial accounts. In our model, we follow the SAT in computing taxable profits by immediately deducting all R&D expenditures from operating profit.

[41] We discuss the details of this procedure in Appendix C.

C Estimation of Residual Productivity

This appendix describes how we construct the empirical measure of firm-level productivity, $\hat{\phi}_{it}$. First, we use the structure in our model of constant elasticity demand to write firm revenue (value-added) as:

$$\ln r_{it} = \left(\frac{\theta - 1}{\theta}\right) [\kappa \ln k_{it} + (1 - \kappa) \ln l_{it} + \phi_{it}],$$

where l_{it} is the labor input, which we assume may be chosen in each period. Second, we obtain the following relation from the first-order condition of cost minimization for the variable input l_{it}:

$$\ln s^l_{it} \equiv \ln\left(\frac{wl_{it}}{r_{it}}\right) = \ln\left[(1 - \kappa)\left(\frac{\theta - 1}{\theta}\right)\right] + v_{it},$$

where $v_{it} \sim iid$, and $E[v_{it}] = 0$ is measurement error or a transitive shock in factor prices. Third, we obtain a consistent estimate of $(1 - \kappa)(\frac{\theta-1}{\theta})$ for each 3-digit manufacturing sector. Finally, given our benchmark value of $\theta = 5$, we construct a residual measure of log TFP as follows:

$$\hat{\phi}_{it} = \frac{\theta}{\theta - 1} \ln r_{it} - \hat{\kappa} \ln k_{it} - (1 - \hat{\kappa}) \ln l_{it}.$$

Robustness of TFP Estimates

We also follow the empirical literature to directly estimate sector-specific production functions using the method of Ackerberg, Caves and Frazer (2015), estimated by Chen et al. (2019). We use only data from firms that do not perform R&D and thus are not affected by the notch in the InnoCom program. We then construct measured productivity based on these production function estimates. We find estimates similar to those from the cost-share based "index number" approach. For instance, the correlation of the labor coefficient in the production function across the two methods is 0.7 across 30 3-digit manufacturing industries. More importantly, the two estimates of measured TFP have a correlation of 0.88.

D Cross-Validation of p and (d^{*-}, d^{*+}) in Bunching Analysis

We follow Diamond and Persson (2016) in using a data-based approach to selecting the excluded region (i.e., (d^{*-}, d^{*+})), and the degree of the polynomial, p. In particular, we use K-fold cross-validation to evaluate the fit of a range of values for these three parameters.

Our cross-validation procedure searches over values of p and possible discrete values of $d^{*-} < \alpha$ and $d^{*+} > \alpha$ that determine the excluded region. Given the monotonically decreasing shape of the R&D intensity distribution, we restrict the estimated β_k's to result in a decreasing density.

For each triple (p, d^{*-}, d^{*+}), the procedure estimates the model in $K = 5$ training subsamples of the data and computes two measures of model fit on corresponding testing subsamples of the data. First, we test the hypothesis that $f_0(\cdot)$ and $f_1(\cdot)$ have equal mass over the exclusion restriction. Second, we compute the sum of squared errors across the test subsamples. We select the combination of parameters that minimizes the sum of squared errors, among the set of parameters that do not reject the test of equality of the first test at the 10% level.

Note that a common practical problem in the literature is the higher frequency in the reporting of round numbers. As Figure 2 in Section 3 demonstrates, our data does not display the round-number problems that are often present in other applications.

Finally, we obtain standard errors by bootstrapping the residuals from the polynomial regression, generating replications of the data, and re-estimating the parameters.

E Robustness of Bunching Estimates

This section explores the robustness of our bunching estimates. First, we show in panel A of Figure A.5 that our estimator is able to recover a null effect in the absence of the policy. This panel estimates the effect of a non-existent notch on the pre-2008 distribution of R&D intensity of large foreign firms, which were not subject to the incentives of the InnoCom program, and finds a small and negative estimate of Δd.

Second, we explore the potential for firms' extensive-margin responses to bias our estimates. If the bunching that we observe is driven by firms that previously did not perform any R&D, the missing mass would not equal the excess mass. This would lead us to underestimate Δd. In panel B of Figure A.5, we use data for large firms in 2011, and we restrict the sample to firms that had positive R&D in 2009 and 2010. This panel shows that we obtain a very similar estimate of Δd when we rule out extensive-margin responses.

Third, we show that our results are robust to using data from before 2008 for large foreign firms that were not subject to the incentives of the InnoCom program to inform the shape of the density in the excluded region. Panel C of Figure A.5 shows that using these data results in very similar estimates of both the counterfactual density and Δd.[42]

Fourth, Figure A.6 estimates the counterfactual density of R&D intensity when we exclude certain groups of firms from the data. Panel A analyzes data on large firms from 2011 and shows that excluding state-owned enterprises from our data does not have a meaningful effect on our estimate of Δd. Similarly, panels B and C show that excluding firms with low profitability and firms that are not in designated high-tech industries, respectively, results in very similar estimates

[42] As discussed in Blomquist and Newey (2017), variation in non-linear incentives can help in identifying responses when bunching approaches are used. We combine this non-manipulated density with the density in 2011, $f_1(d)$, by ensuring that the combined density is continuous at the boundaries of the excluded region, d^{*-} and d^{*+}.

of the effects of the notch on R&D investment.

Fifth, Figure A.7 shows that our estimates of counterfactual densities are robust to the choice of (p, d^{*-}, d^{*+}). This figure shows that restricting (p, d^{*-}, d^{*+}) to the second-best estimate either with $p = 3$ (panel A) or $p = 4$ (panel B) results in very similar estimates. Panel C of this graph further restricts the estimation to have $p = 2$ and to only rely on data such that $d > d^{*+}$ to recover the counterfactual density. This panel shows that even relying only on data beyond the bunching region results in very similar estimates.

Overall, estimates from the bunching analysis consistently show that firms respond to the InnoCom program by increasing their reported R&D intensity.

F ITT Estimates on Productivity, Relabeling, and Tax Revenue

Our structural estimates in Section 5 quantify the cost of relabeling and the productivity effects of R&D. This appendix discusses an alternative and complementary approach to quantifying the effects of the InnoCom program on relabeling and productivity. Because firms select into the program by manipulating R&D, comparing firms that participate in the program to those that do not can result in biased estimates of the effects of the program. To obtain unbiased estimates, we follow a treatment effects approach that compares the (observed) average outcome of firms that could have participated in the program to a counterfactual average without the InnoCom program.

Diamond and Persson (2016) develop an estimator that formalizes this comparison and quantifies the average effect of the program on a given outcome Y:[43]

$$ITT^Y = \mathbb{E}[Y|\text{Notch}, d \in (d^{*-}, d^{*+})] - \mathbb{E}[Y|\text{No Notch}, d \in (d^{*-}, d^{*+})], \tag{F.1}$$

where we define the manipulated region (d^{*-}, d^{*+}) to include all firms that could have responded to the program. While $d^{*+} = \alpha$ in theory, in practice, firms bunch in a neighborhood above α, as can be seen in Figure 2. Equation F.1 compares the average potential outcome of firms in the region (d^{*-}, d^{*+}), which includes firms that do not respond to the program, as well as firms whose R&D intensity would be above the notch without the program. For this reason, we interpret this quantity as an intent to treat (ITT).

[43]Bachas and Soto (2019) implement a similar approach to analyze the effects of notches on other outcomes.

F.1 Implementing the Estimator of Diamond and Persson (2016)

For a given outcome Y_{it}, such as TFP, R&D or administrative costs, the estimate is given by:

$$\begin{aligned}\widehat{ITT^{Y_t}} &= \mathbb{E}[Y_t|\text{Notch}, d_{t_1} \in (d^{*-}_{t_1}, d^{*+}_{t_1})] - \mathbb{E}[Y_t|\text{No }\widehat{\text{Notch}, d_{t_1}} \in (d^{*-}_{t_1}, d^{*+}_{t_1})] \\ &= \left[\frac{1}{N^{Exc.}} \sum_{d_{i,t_1} \in \left(d^{*-}_{t_1}, d^{*+}_{t_1}\right)} Y_{it}\right] - \left[\int_{d^{*-}_{t_1}}^{d^{*+}_{t_1}} \hat{f}_0(r) E[Y_{it}|d_{t_1} \widehat{= r, \text{No}} \text{ Notch}]dr\right]. \end{aligned} \tag{F.2}$$

When Y_t is R&D or administrative costs, we estimate contemporaneous effects, so that $t = t_1$. For the case when Y_t is TFP, we study the effect of the program in time t_1 on future TFP ($t > t_1$). We interpret this estimate as an intent to-treat (ITT).[44] For example, the ITT on $Y = \ln d$ measures Δd, the percentage increase in R&D intensity over the excluded region.

The first quantity in Equation F.2 is the observed average value of a given outcome Y_{it} over the excluded region. The second quantity is a counterfactual average value of Y_{it}. We construct this counterfactual by combining the counterfactual density of R&D intensity that we estimated as part of the bunching analysis ($\hat{f}_0(\cdot)$) with an estimated average value of the outcome conditional on a given value of R&D. We estimate $E[Y_{it}|\widehat{d_{t_1}, \text{No}} \text{ Notch}]$ using a flexible polynomial regression of Y_{it} on R&D intensity over the same excluded region used to estimate $\hat{f}_0(\cdot)$:[45]

$$Y_{it} = \underbrace{\sum_{k=0}^{p} \beta_k \cdot (d_{it_1})^k}_{E[Y_t|d_{t_1}=d, \text{No Notch}]} + \gamma \cdot \mathbf{1}\left[d^{-*} \leq d_{it_1} \leq d^{+*}\right] + \delta Y_{it_1} + \phi_s + \nu_{it},$$

where we exclude observations in the manipulated region and control for industry fixed effects ϕ_s and lagged outcomes Y_{it_1} when $t > t_1$. Armed with an estimate of $E[Y_{it}|d_{t_1}, \text{No Notch}]$, we then compute the counterfactual average value for firms in the excluded region by integrating $E[Y_{it}|d_{t_1}, \text{No Notch}]$ relative to the counterfactual density $f_0(d)$.

The interpretation of Equation F.2 as a treatment effect relies on two assumptions: first, that we can consistently estimate the counterfactual density $f_0(d)$ and, second, that the InnoCom program does not change the relationship between R&D intensity and a given outcome outside the excluded region. This assumption allows us to uncover the relationship between an outcome and the running variable. We can then use this relationship to approximate $E[Y_{it}|d_{t_1}, \text{No Notch}]$ inside

[44] As detailed in our model, firms self-select into the treatment depending on whether they face fixed or adjustment costs that prevent them from obtaining the high-tech certification. This selection implies that we cannot use data just beneath the threshold as a control group for firms above the threshold. Our procedure does not rely on such comparisons across firms but instead relies on the assumption that $E[Y_{it}|d_{t_1}, \text{No Notch}]$ is smooth around the notch and that it may be approximated with data outside the excluded region that, by definition, is not subject to a selection problem.

[45] Note that this regression is not causal. Its role is purely to predict the outcome over the excluded region. We obtain standard errors for ITT estimates in Equation F.2 by bootstrapping this equation as well as the estimates of the counterfactual density.

the excluded region. Given a consistent estimate of $f_0(d)$, this assumption holds trivially for the case where $Y = \ln d$. When we estimate the effect on TFP growth, this assumption implies that the only effect of the program on TFP growth is through real R&D investment. This assumption is consistent with the model in the previous section. A similar argument applies to the case of relabeling through administrative costs. Finally, note that this approach has the advantage that it places no restrictions on the distributions of fixed costs, adjustments costs, and productivity and does not rely on functional form assumptions for relabeling costs and the effects of R&D on firm-level productivity.

Estimation of $\mathbb{E}[Y|d]$ for the ITT Analysis

We now discuss estimates of the functions $\mathbb{E}[Y|d, \text{No Notch}]$. We focus on large firms since, as shown in Figure A.4, they account for the vast majority of R&D in the economy. In addition, all analyses report the effects of the notch in 2009 on outcomes in 2009 and 2011.

We estimate $\mathbb{E}[Y|d, \text{No Notch}]$ using the following regression:

$$Y_{it} = \underbrace{\sum_{k=0}^{p} \beta_k \cdot (d_{it_1})^k}_{E[Y_t|d_{t_1}=d, \text{No Notch}]} + \gamma \cdot \mathbf{1}\left[d^{-*} \le d_{it_1} \le d^{+*}\right] + \delta Y_{it_1} + \phi_s + \nu_{it},$$

where we use the same exclusion region as in panel E of Figure 4 (see Appendix D for details), and we use either quadratic or cubic polynomials for each outcome.[46]

Figure A.8 shows the data for a given outcome as a function of R&D intensity in 2009 (blue circles) along with the fitted values from these regressions (red lines). The size of the circles indicates the weights based on the number of observations in each bin. Panel A considers the case of log R&D intensity. Since this is a mechanical function of R&D intensity, we know what $\mathbb{E}[Y|d, \text{No Notch}]$ should look like. This figure shows that, even though the polynomials are driven by data outside of the exclusion region, we are able to fit non-linear functions very well. Other panels show that the red lines provide a good fit for data outside of the exclusion region. As firms self-select into the InnoCom program, we cannot evaluate the fit inside the exclusion region, since these patterns may be due to selection. Finally, note that we allow the user cost to have a discontinuous jump in panel C since, in contrast to other outcomes, we would expect participation in the program to have a mechanical effect on the user cost of R&D.[47]

[46] To comply with data availability policies, we first collapse the cleaned data into the bins of R&D intensity displayed in panel E of Figure 4. For each bin, we make available the count of firms and the median value of a given variable. We estimate this regression on the binned data, where we weight each bin by the number of firms in each bin.

[47] Diamond and Persson (2016) allow discontinuities in their estimates of $\mathbb{E}[Y|d, \text{No Notch}]$ since, in their application, manipulation above the notch may have a direct effect on outcomes. In our case, we would not expect a direct effect of the program on firm-level outcomes apart from the effects related to tax incentives, which would mechanically affect the user cost of R&D.

F.2 ITT Estimates

Panel A of Table A.1 presents estimates of ITT effects of the InnoCom program on several outcomes. We find that R&D investment for firms in the excluded region increased by 14.6% in 2009, which is very close to the bunching estimate of Δd of 15.6%. We also find a decrease in the administrative cost ratio of 9.6%. We find that administrative costs decreased, relative to the average value of this ratio, by 0.33% of firm sales. Finally, we study how the decision to invest in R&D in 2009 affected productivity in 2011. We find that between 2009 and 2011, the policy led to an increase in TFP of 1.2%. These results show that, while the policy induces relabeling, it also leads to real R&D investment and productivity gains.

To relate our estimates to the existing literature, we obtain estimates of the elasticity of R&D investment to the user cost of capital (UCC). Panel A of Table A.1 shows that the policy lowered the UCC in 2009 by 7.1%.[48] The second panel of Table A.1 presents estimates of UCC elasticities obtained by taking the ratio of the ITT on R&D to the ITT on the UCC, along with bootstrapped confidence intervals. The first row shows that reported R&D increased by 2% for every 1% decrease in the user cost. When we use the approximation above to obtain an estimate of the real increase in R&D, we obtain a user cost elasticity closer to 1.3. Notice that the empirical literature focused on OECD countries (see Hall and Van Reenen, 2000; Becker, 2015) has typically found an elasticity ranging from 0.4 to 1.8 based on direct R&D tax credit programs. Thus, our estimates indicate that, once we correct for the relabeling behavior of Chinese manufacturing firms, their user cost elasticity is comparable to those in more developed economies.

As an alternative metric, we consider how much it costs the government to increase R&D investment in terms of foregone revenue. Panel A of Table A.1 shows that the policy reduced corporate tax revenues by 12.8%. Thus, we find that, for every 1% increase in R&D, there was a 0.88% decrease in tax revenue. This statistic is a useful ingredient for deciding whether the InnoCom policy is too expensive or whether externalities from R&D investment merit further subsidies. However, this statistic does not line up perfectly with the government's objective, since part of the response may be due to relabeling and since this estimator relies on the average percentage increase, which may differ from the percentage increase in total R&D. Our structural model section bridges this gap by computing the fiscal cost of raising real R&D and by showing how the fiscal cost depends on the design of the InnoCom program.

[48] We compute the user cost of R&D by generating an equivalent-sized tax credit. This credit is the ratio of tax savings to R&D investment. We then use the standard Hall and Jorgenson (1967) formula as in Wilson (2009).

G Lack of Manipulation of Other Expenses

In Figure 5, we show a significant downward break in the administrative expense-to-sales ratio at the notches for each firm size category. Given the fact that administrative expenses and R&D are categorized together under the Chinese Accounting standard, we think that is the natural place to find suggestive evidence of relabeling behavior. In this section, we address the question of whether other types of expenses might also illustrate similar empirical patterns. We plot a similar graph to Figure 5 in Figure A.3 for the sales expense-to-sales ratio for all three size categories. We find that there are no detectable discontinuities at the notches for all firms. Note that, while there is a drop for small firms at the 6% notch, Table A.5 shows that this drop is not statistically significant. This analysis suggests that the drops that we observe in administrative costs are likely not due to substitution of inputs and are likely due to relabeling.

H Detailed Model Derivation

This appendix provides additional details behind the derivation of the model.

H.1 Model Setup

Consider a firm i with a constant-returns-to-scale production function given by:

$$q_{it} = \exp\{\phi_{it}\}V_{it}$$

where V_{it} is static input bundle with price w_t and where ϕ_{it} is log TFP, which follows the law of motion given by:

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

where $D_{i,t-1} > 0$ is R&D investment and $u_{i,t} \sim$ i.i.d. $N(0, \sigma^2)$. This setup is consistent with the R&D literature where knowledge capital depreciates over time (captured by ρ) and is influenced by R&D expenditure (captured by ε). In a stationary environment, it implies that the elasticity of TFP with respect to a permanent increase in R&D is $\frac{\varepsilon}{1-\rho}$.

The unit cost function for this familiar problem is simply given by:

$$c(\phi_{it}, w_t) = \frac{w_t}{\exp\{\phi_{it}\}}.$$

The firm faces a constant elasticity demand function given by:

$$q_{it} = p_{it}^{-\theta} B_t,$$

where $\theta > 1$ is the demand elasticity and B_t is the aggregate demand shifter. In a given period, the firm chooses p_{it} to:

$$\max_{p_{it}} \quad p_{it}^{1-\theta} B_t - p_{it}^{-\theta} B_t c(\phi_{it}, w_t).$$

The profit-maximizing p_{it} gives the familiar constant markup pricing:

$$p_{it}^{*} = \frac{\theta}{\theta - 1} c(\phi_{it}, w_t),$$

where $\frac{\theta}{\theta-1}$ is the gross markup. Revenues then equal production costs multiplied by the gross-markup:

$$\text{Revenue}_{it} = \left(\frac{\theta}{\theta - 1} \frac{w_t}{\exp(\phi_{it})} \right)^{1-\theta} B_t.$$

Head and Mayer (2014) survey estimates of θ from the trade literature. While there is a broad range of estimates, the central estimate is close to a value of 5, which implies a gross markup of 1.25. We also normalize the input cost $w_t \equiv 1$ for the rest of our analysis. We can then write per-period profits as:

$$\pi_{it} = \frac{1}{\theta} \text{Revenue}_{it} = \frac{(\theta - 1)^{\theta-1}}{\theta^{\theta}} \left[\exp(\phi_{it})\right]^{\theta-1} B_t.$$

Uncertainty and R&D investment enter per-period profits through the realization of log TFP ϕ_{it}. We can write expected profits as follows:

$$\begin{aligned} \mathbb{E}[\pi_{it}] &= \frac{(\theta - 1)^{\theta-1}}{\theta^{\theta}} B_t \left[\exp(\rho(\theta - 1)\phi_{i,t-1} + (\theta - 1)^2 \sigma^2/2)\right] D_{i,t-1}^{(\theta-1)\varepsilon} \\ &= \tilde{\pi}_{it} D_{i,t-1}^{(\theta-1)\varepsilon}, \end{aligned}$$

where $\tilde{\pi}_{it}$ denotes the expected profit without any R&D investment.

We follow the investment literature and model the adjustment cost of R&D investment with a quadratic form that is proportional to revenue $\theta\pi_{i1}$ and that depends on the parameter b:

$$g(D_{it}, \theta\pi_{it}) = \frac{b\theta\pi_{it}}{2} \left[\frac{D_{it}}{\theta\pi_{it}}\right]^2.$$

H.2 R&D Choice under Linear Tax

Before considering how the InnoCom program affects a firm's R&D investment choice, we first consider a simpler setup without such a program. In a two-period context with a linear tax, the firm's inter-temporal problem is given by:

$$\max_{D_1} \, (1 - t_1) \left(\pi_{i1} - D_{i1} - g(D_{i1}, \theta\pi_{i1})\right) + \beta(1 - t_2)\tilde{\pi}_{i2} D_{i1}^{(\theta-1)\varepsilon},$$

where the firm faces an adjustment cost of R&D investment given by $g(D_{i1}, \theta\pi_{i1})$. This problem has the following first-order condition:

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

Notice first that if the tax rate is constant across periods, the corporate income tax does not affect the choice of R&D investment.[49] In the special case of no adjustment costs (i.e., $b = 0$), the optimal choice of D_{i1} is given by:

$$D_{i1}^{*} = \left[\frac{\beta(1-t_2)(\theta-1)\varepsilon}{1-t_1}\tilde{\pi}_{i2}\right]^{\frac{1}{1-(\theta-1)\varepsilon}}. \tag{H.2}$$

We observe that the choice of R&D depends on the potentially unobserved, firm-specific factor ϕ_{i1} that influences $\tilde{\pi}_{i2}$. If $(\theta-1)\varepsilon < 1$, then R&D investment is increasing in firm's current productivity ϕ_{i1}.

Since the InnoCom program focuses on R&D intensity (i.e., the R&D-to-sales ratio), we also rewrite the FOC in terms the optimal R&D intensity $d_{i1}^{*} \equiv \frac{D_{i1}^{*}}{\theta\pi_{i1}}$:

$$\underbrace{(1-t_1)\left(1+bd_{i1}^{*}\right)}_{\text{Increase in Investment Cost}} = \underbrace{\beta(1-t_2)\varepsilon(\theta-1)(d_{i1}^{*})^{(\theta-1)\varepsilon-1}\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}}_{\text{Productivity Gain from R\&D}}. \tag{H.3}$$

When the profit return of R&D $\varepsilon(\theta-1)$ is larger (smaller) than the depreciation rate of knowledge $(1-\rho)$, firms' R&D intensity d_{i1} is increasing (decreasing) in firm's current TFP ϕ_{i1} and size. In our data, R&D intensity is weakly positively correlated with firm TFP and size. We use this pattern to discipline our key parameter ε in our model estimation.

We now write the optimal firm value-to-sales ratio as:

$$\frac{\Pi(d_{i1}^{*}|t_2)}{\theta\pi_{i1}} = (1-t_1)\left[\frac{1}{\theta} + d_{i1}^{*}\left(\frac{1}{(\theta-1)\varepsilon}-1\right) + (d_{i1}^{*})^{2}\left(\frac{b}{(\theta-1)\varepsilon}-\frac{b}{2}\right)\right], \tag{H.4}$$

where we use Equation H.3 to substitute $\frac{\tilde{\pi}_{i2}(d_{i1}^{*})^{(\theta-1)\varepsilon}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}$ with $\frac{(1-t_1)\left(1+bd_{i1}^{*}\right)}{\beta(1-t_2)\varepsilon(\theta-1)(d_{i1}^{*})^{-1}}$.

Second-Order Condition To ensure that our model results in a well-defined solution, we confirm that the second-order condition holds at the estimated values. The SOC is given by:

$$SOC: \ -(1-t_1)\left(b\left[\frac{1}{\theta\pi_{i1}}\right]\right) + \beta(1-t_2)\varepsilon(\theta-1)((\theta-1)\varepsilon-1)(D_{i1}^{*})^{(\theta-1)\varepsilon-2}\tilde{\pi}_{i2} < 0.$$

It is sufficient to have $(\theta-1)\varepsilon < 1$ for the second-order condition to hold. We can also use the implicit function theorem to show that the R&D decision D_{i1}^{*} is increasing in ϕ_{i1} if $(\theta-1)\varepsilon < 1$, which is consistent with numerous empirical studies.

H.3 A Notch in the Corporate Income Tax

Assume now that the tax in the second period has the following structure that mirrors the incentives in the InnoCom program:

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

[49] This simple model eschews issues related to source of funds, as in Auerbach (1984).

$t_2^{LT} > t_2^{HT}$ and where α is the R&D intensity required to obtain the high-tech certification and LT/HT stands for low-tech/high-tech. In addition, we introduced a fixed cost of certification c such that firms need to pay $c \times \theta\pi_{i1}$ to obtain the tax benefit when they pass the R&D intensity threshold.

We first calculate the optimal profit of the firm conditioning on bunching at the notch, $\Pi(\alpha\theta\pi_1|t_2^{HT})$:

$$\Pi(\alpha\theta\pi_1|t_2^{HT}) = (1-t_1)\left(\pi_{i1} - \theta\pi_{i1}(\alpha+c) - \frac{b\theta\pi_{i1}}{2}\left[\frac{\alpha\theta\pi_{i1}}{\theta\pi_{i1}}\right]^2\right) + \beta(1-t_2^{HT})(\alpha\theta\pi_{i1})^{(\theta-1)\varepsilon}\tilde{\pi}_{i2}.$$

Let $\frac{\Pi(\alpha|t_2^{HT})}{\theta\pi_{i1}}$ be the value-to-sales ratio of the firm conditional on bunching at the notch. We can write it as:

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

We can similarly substitute $\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}$ with $\frac{(1-t_1)\left(1+bd_{i1}^*\right)}{\beta(1-t_2^{LT})\varepsilon(\theta-1)(d_{i1}^*)^{\varepsilon(\theta-1)-1}}$ and express the value-to-sales ratio as

$$(1-t_1)\left[\frac{1}{\theta} + \alpha\left(\left(\frac{d_{i1}^*}{\alpha}\right)^{1-(\theta-1)\varepsilon}(1+bd_{i1}^*)\frac{(1-t_2^{HT})}{(1-t_2^{LT})}\frac{1}{\varepsilon(\theta-1)} - \left(1+\frac{c}{\alpha}\right) - \frac{\alpha b}{2}\right)\right]. \tag{H.5}$$

A firm will bunch at the notch if $\frac{\Pi(\alpha|t_2^{HT})}{\theta\pi_{i1}} \geq \frac{\Pi(d_{i1}^*|t_2)}{\theta\pi_{i1}}$, which occurs when:

$$\begin{aligned}\left(\frac{d_{i1}^*}{\alpha}\right)^{1-(\theta-1)\varepsilon}&\left(1+\alpha b\left(\frac{d_{i1}^*}{\alpha}\right)\right)\frac{(1-t_2^{HT})}{(1-t_2^{LT})}\frac{1}{\varepsilon(\theta-1)} - \left(1+\frac{c}{\alpha}\right) - \frac{\alpha b}{2}\\ &\geq \frac{d_{i1}^*}{\alpha}\left(\frac{1}{(\theta-1)\varepsilon} - 1\right) + \alpha\left(\frac{d_{i1}^*}{\alpha}\right)^2\left(\frac{b}{(\theta-1)\varepsilon} - \frac{b}{2}\right).\end{aligned} \tag{H.6}$$

For each specific realization of adjustment and fixed costs (b, c), we define the marginal firm with interior optimal R&D intensity $d_{b,c}^{*-}$ such that Equation H.6 holds with equality.

H.4 R&D Choice under Tax Notch with Relabeling

Assume now that firms may misreport their costs and shift non-R&D costs to the R&D category. Following conversations with CFOs of large Chinese companies, we model relabeling as a choice to misreport expenses across R&D and non-R&D categories. Misreporting expenses or revenues overall is likely not feasible, as firms are subject to third-party reporting (see, e.g., Kleven et al. (2011)).

Denote a firm's reported level of R&D spending by $\tilde{D}_{i1}$. The expected cost of misreporting to the firm is given by $h(D_{i1}, \tilde{D}_{i1})$. We assume that the cost of misreporting is proportional to the reported R&D, $\tilde{D}_{i1}$, and depends on the percentage of misreported R&D, $\delta_{i1} = \frac{\tilde{D}_{i1} - D_{i1}}{\tilde{D}_{i1}}$, so that:

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

We also assume that $\tilde{h}$ satisfies $\tilde{h}(0) = 0$ and $\tilde{h}'(\cdot) \geq 0$.

The effects of the InnoCom program are now as follows:

$$t_2 = \begin{cases} t_2^{LT} & \text{if } \tilde{D}_1 < \alpha\theta\pi_1 \\ t_2^{HT} & \text{if } \tilde{D}_1 \geq \alpha\theta\pi_1 \end{cases},$$

Notice first that, if a firm decides not to bunch at the level $\alpha\theta\pi_1$, there is no incentive to misreport R&D spending, as it does not affect total profits and does not affect the tax rate. However, a firm might find it optimal to report $\tilde{D}_1 = \alpha\theta\pi_1$ even if the actual level of R&D is lower. We start by characterizing the firm's optimal relabeling strategy δ_{i1}^* conditional on bunching and its resulting payoff function $\Pi(\alpha\theta\pi_1, D_1^{*K}|t_2^{HT})$.

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

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:

$$\underbrace{-(1-t_1)\left(1 + bd_{i1}^{K*}\right) + \tilde{h}'\left(1 - \frac{d_{i1}^{K*}}{\alpha}\right)}_{\text{Increase in Investment Cost and Reduction in Relabeling Cost}} \quad + \underbrace{\beta(1-t_2^{HT})\varepsilon(\theta-1)d_{i1}^{K*(\theta-1)\varepsilon-1}\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}}_{\text{Productivity Gain from Real R\&D}} = 0. \tag{H.7}$$

We again substitute for the expected productivity components of the firm decision, i.e., $\tilde{\pi}_{i2}$ with the interior optimal R&D d_{i1}^* using Equation H.3. The FOC for d_1^K effectively characterizes the optimal relabeling strategy $\delta_{i1}^* \equiv 1 - \frac{d_{i1}^{K*}}{\alpha}$.

The firm decides to bunch if the profits from the optimal relabeling strategy $\Pi(\alpha, d_{i1}^{K*}|t_2^{HT})$ are greater than when the firm is at the optimal interior solution (and truthful reporting) $\Pi(d_{i1}^*, d_{i1}^*|t_2^{LT})$. We write this in terms of a value-to-revenue ratio comparison and obtain:

$$\underbrace{\left(\frac{d_{i1}^*}{d_{i1}^{K*}}\right)^{1-(\theta-1)\varepsilon}(1+bd_{i1}^*) \times \frac{(d_{i1}^{K*}/\alpha)}{(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) - \frac{c}{\alpha} - \frac{d_{i1}^{K*}}{\alpha} - \frac{b}{2\alpha}(d_{i1}^{K*})^2}_{\text{Relative Profit from Bunching}}$$

$$-\underbrace{\frac{\tilde{h}(1-d_{i1}^{K*}/\alpha)}{(1-t_1)}}_{\text{Relabeling Cost}} \geq \underbrace{\frac{d_{i1}^*}{\alpha}\left(\frac{1}{(\theta-1)\varepsilon} - 1\right) + \alpha\left(\frac{d_{i1}^*}{\alpha}\right)^2\left(\frac{b}{(\theta-1)\varepsilon} - \frac{b}{2}\right)}_{\text{Relative Profit without Bunching}}. \tag{H.8}$$

The marginal firm $d_{b,c}^{*-}$ in this case is determined by Equation H.7 and Equation H.8 when it holds with strict equality.

I Connecting the Model with Bunching Estimates

Section 5 estimates the model in Section 4 by matching the descriptive statistics from Section 3. This section shows that our model is also directly connected to bunching estimates, which are commonly used in the public finance literature.

I.1 Empirical Implications for Bunching on R&D

Figure 7 provides intuition linking the model to bunching estimates. Panel B plots $f_0(d)$: the counterfactual distribution of R&D intensity under a linear tax. In a world where firms face no adjustment or fixed costs, all of the firms in the range (d^{*-}, α) would bunch at the notch. Denote B as the missing mass relative to the counterfactual distribution over this range. To see how the model relates to the extent of bunching, note that a larger value of ε or a lower relabeling cost will result in a larger missing mass B and a lower value of d^{*-}—i.e., the marginal bunching firm has a lower R&D intensity.

Panel B also plots $f_1(d)$, which shows that, in the presence of fixed and adjustment costs, some firms do not respond to the incentives in the InnoCom program. For given values of (b, c), a firm will be constrained from responding if $d < d^{*-}_{b,c}$, an event that we denote by $\mathbb{I}[d < d^{*-}_{b,c}]$. The fraction of constrained firms at a given value of r in the range (d^{*-}, α) is given by:

$$\mathbb{P}r(\text{Constrained}|r) = \int\limits_{b,c} \mathbb{I}[r < d^{-}_{b,c}] f_0(r, b, c) d(b, c) = f_1(r),$$

where $f_0(r, b, c)$ is the joint density of R&D intensity, fixed costs, and adjustment costs and where the second equality notes that we observe this fraction of firms in the data. It follows from this expression that measures of $\mathbb{P}r(\text{Constrained}|r)$ are informative of the distributions of b and c.

The area B can be computed as follows:

$$\begin{aligned} B &= \int\limits_{d^{*-}}^{\alpha} \int\limits_{b,c} \mathbb{I}[r \geq d^{-}_{b,c}] f_0(r, b, c) d(b, c) dr = \int\limits_{d^{*-}}^{\alpha} \int\limits_{b,c} (1 - \mathbb{I}[r < d^{-}_{b,c}]) f_0(r, b, c) d(b, c) dr \\ &= \int\limits_{d^{*-}}^{\alpha} (f_0(r) - \mathbb{P}r(\text{Constrained}|r)) dr = \int\limits_{d^{*-}}^{\alpha} (f_0(r) - f_1(r)) dr. \end{aligned} \tag{I.1}$$

The first line shows that B depends on the distribution of fixed and adjustment costs. The second line shows that frictions result in a smaller bunching mass B by subtracting the fraction of constrained firms. The observed degree of bunching B is therefore a function of ε, relabeling costs, adjustment costs, and certification costs.

This discussion highlights how bunching estimates can inform the parameters of the model. Specifically, the model predicts that higher values of ε, lower costs of relabeling, and lower fixed and adjustment costs will result in larger values of B, lower values of d^{*-}, and lower values of $\mathbb{P}r(\text{Constrained}|d)$.

I.2 Percentage Increase in the R&D Intensity of the Marginal Firm

We now connect the values of B, d^{*-}, and $\mathbb{P}r(\text{Constrained}|d)$ to estimates of the effects of the InnoCom program on the increase in R&D using approximations that are common in this literature (e.g., Kleven and Waseem, 2013). Recall that:

$$B = \int_{d^{*-}}^{\alpha} (1 - \mathbb{P}r(\text{Constrained}|r)) f_0(r) dr.$$

Using the assumption that $\mathbb{P}r(\text{Constrained}|d)$ does not depend on d, we obtain:

$$\begin{aligned} B &= (1 - \mathbb{P}r(\text{Constrained})) \int_{d^{*-}}^{\alpha} f_0(r) dr \\ &\approx (1 - \mathbb{P}r(\text{Constrained})) f_0(\alpha) \alpha \underbrace{\frac{\alpha - d^{*-}}{\alpha}}_{\Delta D^*}, \end{aligned} \tag{I.2}$$

where the second line approximates the integral under the assumption that $f_0(d)$ is flat in the interval (d^{*-}, α). While the assumptions behind this approximation may be strong, they provide a useful approximation for ΔD^* based on B and $\mathbb{P}r(\text{Constrained})$:

$$\Delta D^* \approx \frac{B}{f_0(\alpha)\alpha(1 - \mathbb{P}r(\text{Constrained}))}.$$

This equation can be implemented using the following counterfactual estimates for $f_0(\alpha)$ and B:

$$\widehat{f_0(\alpha)} = \sum_{k=0}^{p} \hat{\beta}_k \cdot (\alpha)^k \quad \text{and} \quad \hat{B} = \sum_{d_j = d^{*-}}^{\alpha} \left(\sum_{k=0}^{p} \hat{\beta}_k \cdot (d_j)^k - c_j \right).$$

Following Kleven and Waseem (2013), it is possible to estimate the fraction of constrained firms at an R&D intensity α^- such that firms would be willing to jump to the notch even if R&D had no effects on productivity:[50]

$$a^*(\alpha^-) = \frac{\mathbb{P}r(\widehat{\text{Constrained}}|\alpha^-)}{\widehat{f_0(\alpha^-)}} = \frac{c_{\alpha^-}}{\sum_{k=0}^{p} \hat{\beta}_k \cdot (\alpha^-)^k}.$$

Note that this expression differs from our expression for $a^* = \int_{d^{*-}}^{\alpha} f_1(v)dv / \int_{d^{*-}}^{\alpha} \hat{f}_0(v)dv$ because the latter considers the average fraction of firms that respond to the program over the interval (d^{*-}, α).

I.3 Connecting the Model with ITT Estimates

Our model generates intuitive predictions for the ITT effects on R&D, relabeling through administrative costs (ADM), and TFP. If some of the reported R&D intensity is real activity, our model

[50] The "money-burning" point is easy to compute. Note that the tax benefit is given by Profits $\times (t^{HT} - t^{LT})$ and the cost of jumping to the notch is Sales $\times (\alpha - \alpha^-)$, which implies that $\alpha^- = \alpha - (t^{HT} - t^{LT}) \times \frac{\text{Profits}}{\text{Sales}}$. The average net profitability ratio in our data of 7% implies that firms in the range $(\alpha - .07 \times (t^{HT} - t^{LT}), \alpha)$ are not able to respond to the incentives of the InnoCom program. For the case of large firms, we have $(\alpha^-, \alpha) = (2.3\%, 3\%)$.

would predict that $ITT^{TFP} \geq 0$. According to our model for the evolution of TFP in Equation 1, we would find larger values of ITT^{TFP} for larger values of the parameter ε. We expect to find $ITT^{ADM} < 0$ if a fraction of the reported R&D is due to relabeling of administrative costs. Intuitively, if firms over-report R&D by under-reporting administrative costs, the average ADM over the excluded region would be artificially low. Our model predicts small values of ITT^{ADM} if firms face large costs of relabeling. Finally, consider the case where the outcome of interest is reported R&D intensity. In this case, ITT^{d} only depends on the counterfactual density $f_0(d)$. Our model predicts a larger fraction of compliers if ε is large or if relabeling costs are low.

This discussion shows that this treatment effects approach can also quantify the extent of relabeling and the effects of the InnoCom program on productivity. In addition, the estimates of Equation F.1 complement the model-based approach by providing additional moments that can inform the parameters of the model.

J Details of Structural Estimation

This appendix provides details on the structural estimation.

We first discuss how we compute the moments used in our structural estimation. To comply with data availability policies, we first collapse the cleaned data into the bins of R&D intensity displayed in panel E of Figure 4. For each bin, we make available the count of firms and the average value of a given variable. We use these collapsed data to compute the following moments:

$m^D(\Omega)$: **R&D Distribution Moments:**

1. Density of R&D intensity for three intervals below the notch (at 3%): [0.3, 1.2], [1.2, 2.1], and [2.1, 3]. We compute these moments by aggregating bin counts into these three categories. The variance of these moments is computed using the closed-form expression for the variance of a multivariate Bernoulli distribution.

2. Density of R&D intensity for three intervals above the manipulated region: [5, 6.3], [6.3, 7.6], and [7.6, 9]. We compute these moments by aggregating bin counts into these three categories. The variance of these moments is computed using the closed-form expression for the variance of a multivariate Bernoulli distribution.

3. Average R&D intensity over the interval [3,5]. We compute this moment by averaging the bin averages based on the number of firms in each bin. We compute the variance of this moment by bootstrapping over the number of firms for each bin.

4. Average TFP below the notch (at 3%). We compute this moment by averaging the bin averages based on the number of firms in each bin. We compute the variance of this moment by bootstrapping over the number of firms for each bin.

5. Average TFP above the notch (at 3%). We compute this moment by averaging the bin averages based on the number of firms in each bin. We compute the variance of this moment by bootstrapping over the number of firms for each bin.

6. Administrative cost ratio break at notch. We implement a version of Figure 5 using the binned data. That is, we estimate a regression with third-order polynomials above and below the notch that is weighted by the number of observations in each bin. The moment used in the estimation is the difference between the residuals of this regression for the bins above and below the notch, which matches our estimate of the structural break reported in Table A.2. We obtain the variance of this moment by bootstrapping this difference over the number of firms for each bin.

$m^B(\Omega)$: **Bunching Moments:**

7. Bunching point (R&D intensity of marginal buncher). We obtain this estimate based on the procedure described in Appendix D. To estimate its variance, we solve Equation I.2 for d^{*-} and compute the variance of the resulting expression.

8. Increase in reported R&D (in the manipulated region). We use the observed density $f_1(\cdot)$ and the estimated counterfactual density $\hat{f}_0(\cdot)$ (see Section 3 and Appendix D) to compute $\mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})]$ and $\mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]$, respectively. We then compute the increase in R&D intensity relative to the observed average over the exclusion region. We compute the variance of this moment by bootstrapping the counterfactual average over the excluded region.

9. Fraction of firms not bunching. We use the observed density $f_1(\cdot)$ and the estimated counterfactual density $\hat{f}_0(\cdot)$ (see Section 3 and Appendix D) to compute $a^* = \int_{d^*-}^{\alpha} f_1(v)dv / \int_{d^*-}^{\alpha} \hat{f}_0(v)dv$. We compute the variance of this moment by bootstrapping the counterfactual average over the excluded region.

We now discuss the empirical implementation of the model. We simulate 30,000 firms that differ in their initial productivity ϕ_{i1}. For a given guess of the structural parameters of the model, Ω, we first determine the R&D intensity conditional on not bunching by solving Equation H.3 for each firm. We then determine the optimal relabeling strategy conditional on bunching using Equation H.7. Equation H.8 then determines whether a firm bunches at the notch. Using the optimal R&D investment and relabeling strategies of these 30,000 simulated firms, we then compute distributional and bunching moments. We repeat this process 10 times and average the moments over these instances to compute $m^D(\Omega)$ and $m^B(\Omega)$. The estimate of the structural parameters, $\hat{\Omega}$, is determined by minimizing the criterion function as discussed in Section 5.

K Robustness of Structural Model Assumptions

In this section, we conduct a few additional robustness checks of the parametric and modeling assumptions that we make in our structural estimation.

K.1 Parametric Distribution of Firm Productivity

In our benchmark model, we micro-found the cross-sectional log TFP distribution from a normal AR(1) process. We use the persistence and volatility of log value-added for non-R&D firms to calibrate the persistence parameter $\rho = 0.725$ and variance parameter $\sigma = 0.385$.[51] The assumption of this process requires the cross-sectional distribution of firm TFP $\exp(\phi_1)$ to be log-normal. Since we construct firm-level TFP in our data, we can check this parametric assumption directly with the TFP data.

We use ideas proposed by Kratz and Resnick (1996) and Head, Mayer and Thoenig (2014) in this robustness check. The basic idea is to compare the distribution of measured productivity with the distribution implied by a given functional form. We first construct the *empirical* CDF of firms' measured TFP as $\hat{F}_i, i = 1, 2, ...N$, with i ranked based on firm TFP.

According to the log-normal parametric assumption, the *theoretical* CDF is $\Phi_N(\frac{\ln TFP - \mu_{tfp}}{\sigma_{tfp}})$, with Φ_N as the standard normal CDF. For a number $F_i \in [0, 1]$, we can then write $\ln TFP_i$ as:

$$\ln TFP_i^{LN} = \mu_{tfp} + \Phi_N^{-1}(F_i)\sigma_{tfp}.$$

A commonly used alternative parametric assumption is that firm TFP follows a Pareto distribution. Building on the idea above, the implied $\ln TFP_i$ for a value $F_i \in [0, 1]$ is given by:

$$\ln TFP_i^P = \ln \underline{\alpha_{tfp}} - \frac{1}{\alpha_{tfp}} \ln(1 - F_i)$$

where $\underline{\alpha_{tfp}}$ and α_{tfp} are the Pareto scale and shape parameters.

With our frequency estimate $\hat{F}_i$, we can then impose the log-normal formula to predict $\ln \hat{TFP^{LN}}_i$ and the Pareto formula, which yields $\ln \hat{TFP^P}_i$. This procedure allows us to evaluate how reasonable the log-normal or Pareto parametric assumptions are by comparing these predictions with the observed distribution of log TFP, $\ln \hat{TFP}_i$.

Figure A.9 shows the predicted TFP from imposing the log-normal CDF on the left panel and the Pareto CDF on the right panel. The predicted TFP from imposing the log-normal CDF tracks the 45 degree line, implying that this assumption is broadly consistent with the data. By contrast,

[51] An alternative approach would be to estimate these parameters in a panel regression of TFP on lagged TFP and lagged R&D (as in our Equation 1). Estimating this equation directly from the data is challenging for two reasons: (1) endogeneity concerns and (2) the fact that, as we show, reported R&D is manipulated, which is a form of measurement error. Given the assumption of constant markup in the model, the firm size distribution maps directly into firms' underlying heterogeneity, which allows us to recover these parameters from the firm size distribution.

the predicted TFP based on the Pareto CDF does not fit the data well. This graph provides strong evidence that log-normal is a reasonable parametric assumption for the TFP distribution.[52]

K.2 Heterogeneity in the TFP Elasticity: ε

Our benchmark model assumes firms have heterogeneous technological opportunities for R&D investment, driven by the heterogeneity in adjustment costs, b. An alternative way of modeling the heterogeneity in firms' technological opportunities is to allow heterogeneity in ε.[53] As we show in this appendix, our average estimates of ε and b do not depend on which can be heterogeneous. However, models where ε is allowed to be heterogeneous produce worse fits of the data. We therefore believe that our benchmark model is superior to models with heterogeneous values of ε.

To investigate how this alternative setup affects our results, we estimated models where ε follows a Beta distribution $B(\alpha_\varepsilon, \beta_\varepsilon)$ between 0 and an upper bound of $\bar{\varepsilon}$. We chose the Beta distribution since its probability density function is highly flexible in the interval $[0, \bar{\varepsilon}]$. We estimated two versions of the heterogeneous-ε model. In Model A, we assume a symmetric Beta distribution, i.e., $\alpha_\varepsilon = \beta_\varepsilon$, and jointly estimate α_ε and $\bar{\varepsilon}$. In Model B, we impose $\bar{\varepsilon} = 1/(\theta - 1) = 0.25$, a value that guarantees the second-order condition of firm's R&D choice problem. We then estimate α_ε and β_ε. The results are reported in Table A.6.

Several findings are worth highlighting. First, several of the parameters are close to our baseline estimates. We estimate average values of ε of 0.0725 and 0.102 in Models A and B, respectively.[54] These values are close to our benchmark estimate of 0.09. Similarly, the average adjustment cost parameter is 7.666 and 8.271 for the two cases, which bracket our benchmark estimate. This is not surprising since adjustment costs are primarily identified by the distribution of R&D intensity away from the notch.

Second, these models imply small variations in the forces that lead firms to bunch at the notch. Model A combines a slightly lower mean value of ε with lower values of the fixed cost parameter and the cost of relabeling. Because this model implies a lower productivity impact of R&D, the model attempts to match observed bunching patterns via lower certification and relabeling costs. In contrast, Model B combines a slightly larger mean estimate of ε of 0.102 with larger values of the parameters that govern the fixed costs and the cost of relabeling. While a larger average value of ε would lead to more bunching in this model, this force is limited by fixed and relabeling costs.

An important question is whether these alternative models yield better or worse fits of the data moments. Panel B of Table A.6 shows that both of these models result in significantly lower values

[52] Head, Mayer and Thoenig (2014) draws a similar conclusion for export sales using micro-level data of French and Chinese exporters. We also follow Clauset, Shalizi and Newman (2009) and formally conduct a goodness-of-fit test based on a Kolmogorov-Smirnov (KS) statistic. The Pareto distribution is strongly rejected with our measured TFP data.

[53] Note that our data variation cannot separately identify heterogeneity in both ε and b.

[54] Recall that $\varepsilon \sim \bar{\varepsilon} \times \text{Beta}(\alpha, \beta)$ so that the mean value of ε is $\frac{\bar{\varepsilon}\alpha}{\alpha+\beta}$.

for average TFP both below and above the notch. Model A, in particular, results in a smaller gap between TFP above and below the notch, which may explain the smaller average estimate of ε. In addition, these alternative models also have a harder time matching the extent of bunching, reflected in lower values of Δd^*. These findings indicate that despite obtaining similar estimates of key model parameters, our benchmark model of heterogeneous adjustment cost is a preferable model for our data.

K.3 Robustness of the Adjustment Cost Function

We now explore the robustness of our structural model to variations in the adjustment cost function. Our baseline model assumes:

$$g(D_{it}, \theta\pi_{it}) = \frac{b\theta\pi_{it}}{2}\left[\frac{D_{it}}{\theta\pi_{it}}\right]^2.$$

While this function follows previous research in normalizing adjustment costs relative to firm scale (e.g., Bloom, 2009), a potential concern is that this size-dependence may constrain our estimates. To explore this possibility, we now consider a generalized function that can vary the dependence on firm size:

$$g(D_{it}, \theta\pi_{it}) = b\frac{(\theta\pi_{it})^{1-\nu}}{2}\left[\frac{D_{it}}{\theta\pi_{it}}\right]^2.$$

With this function, an R&D intensity of d has the overall cost of:

$$d\left(1 + b\frac{d}{(\theta\pi_{it})^{\nu}}\right) \times \theta\pi_{it}.$$

While our baseline specification assumes $\nu = 0$, this formulation allows us to test whether adjustment costs for R&D intensity are increasing or decreasing as a function of firm size. For instance, a positive value of ν would imply that larger firms are subject to smaller adjustment costs.

Column 3 in Table A.7 reports estimates of this extended model. We estimate a small and positive value of ν that is not statistically different from zero. We do not find a significant improvement in the fit of our moments, despite having one more parameter. Overall, we find similar estimates of the effect of R&D on TFP ($\varepsilon = 0.091$) and on the extent of relabeling (fraction of relabeled R&D at 24.2%). These results suggest that our main estimates are not biased by a potential size-dependent property of adjustment costs.

K.4 Robustness of the Relabeling Cost Function

We now consider the robustness of our model to the formulation of the cost of relabeling. Our baseline model assumes that the cost of misreporting is proportional to the reported R&D, $\tilde{D}_{i1}$, and depends on the percentage of misreported R&D, $\delta_{i1} = \frac{\tilde{D}_{i1} - D_{i1}}{\tilde{D}_{i1}}$, so that:

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

While this assumption is consistent with the literature on evasion (e.g., Slemrod and Gillitzer, 2013), one potential concern is that our results may depend on the degree to which real behavior interacts with avoidance behavior—what Slemrod and Gillitzer (2013) call the "avoidance facilitating effect of real activity."

To explore the robustness of our results, we consider an alternative cost of evasion. We modify the formulation above by normalizing relative to firm sales $\theta\pi_{i1}$, as opposed to reported R&D investment, so that:

$$h(D_{i1}, \tilde{D}_{i1}) = \theta\pi_{i1}\tilde{h}\left(\frac{\tilde{D}_{i1} - D_{i1}}{\theta\pi_{i1}}\right).$$

Normalizing by firm sales as a proxy for firm size is intuitive for two reasons. First, costs that are not proportional to firm size will result in large-scale evasion by large firms and no evasion by small firms, which is not consistent with our empirical results. Second, the policy itself is targeted toward R&D intensity (R&D over sales), making it likely that the government monitors R&D intensity directly. Importantly, this formulation only depends on the difference $(\tilde{D}_{i1} - D_{i1})$ and is separable from R&D. Using our specification for $\tilde{h}(\cdot)$ implies:

$$h(D_{i1}, \tilde{D}_{i1}) = \theta\pi_{i1}\frac{\exp\left\{\eta\frac{\tilde{D}_{i1} - D_{i1}}{\theta\pi_{i1}}\right\} - 1}{\eta}.$$

From this expression, it follows that for small values of η (i.e., as $\eta \to 0$), we have:

$$h(D_{i1}, \tilde{D}_{i1}) = \eta(\tilde{D}_{i1} - D_{i1}).$$

That is, this specification accommodates a linear, separable cost of evasion that is also independent of firm size.

Column 4 in Table A.7 reports estimates from this alternative model. Importantly, we find similar estimates of the R&D effects on TFP of 10%. Because the coefficient of the cost of evasion function, η, now has a different magnitude, it is not directly comparable to our benchmark estimate. However, this new specification implies that 23.2% of the increase in R&D is due to relabeling, which is similar to our baseline of 24.2%.

However, Panel B of Table A.7 shows that the fit of several of the moments is worse under this specification of the cost of relabeling. For instance, this model predicts less bunching—both in lower values of the average R&D intensity between 3% and 5% and in Δd^*—and results in a worse fit of the moments measuring average TFP below and above the notch.

Overall, this robustness check shows that our model is not constraining the cost of evasion in a way that biases our main estimates or that prevents us from improving the fit of the data.

K.5 Potential Real Responses to Administrative Costs

As we note in Section 6 (footnote 9), it is possible that the structural break in administrative costs that we document in Figure 5 is partly driven by "reallocating resources from other expenses toward R&D or more precise accounting of previously undercounted R&D expenses." To study the sensitivity of our results to this possibility, we estimate an alternative version of the model where we multiply the coefficient from the structural break regression by $(1-\xi)$, where $\xi = 0.25$.

Column 5 of Table A.7 reports the results from this model. Remarkably, we obtain very similar estimates for most of our structural parameters. The only exception is the cost of evasion parameter η. Under the assumption that 75% of the drop is due to real responses, our estimate of η increases from our baseline of 6.755 to 7.655. This increase makes sense, since a larger cost of relabeling would result in a lower extent of relabeling.

This model is informative of the sensitivity of our results to how we interpret the estimates from Figure 5. First, the productivity effects of R&D are not significantly affected by this alternative formulation. Second, while the estimate of η increases in a predictable way, this increase is relatively small and does not affect our main results. In this case, the fraction of the relabeled R&D is 21.7%, which is lower than our baseline finding but still quite substantial.

K.6 Robustness to Allowing a Correlation between Fixed Costs and Productivity

An important force in our model is the firm-level decision of whether to bunch. Our baseline model assumes that firm-level TFP ϕ is distributed independently of firms' adjustment and fixed costs. One potential concern is that firm-level TFP may be correlated with adjustment costs and that this correlation may bias our main estimates.

We explore this possibility by considering a non-zero correlation between the certification cost c and productivity ϕ. We specify the dependence of the certification cost on productivity as follows:

$$\hat{c} = c \times \exp\{\kappa\phi\}.$$

The exponential form allows the costs of certification to be smaller or larger depending on ϕ. We then estimate the parameter κ along with the other parameters. This model nests our baseline model in the case that $\kappa = 0$.

One way that this variation can improve the fit of the model is through the selection channel. That is, if our baseline model is not properly capturing the selection of high-productivity firms into the program, by allowing for a dependence of $\hat{c}$ on ϕ, this extended model can help improve the fit of the ITT of TFP. A different pattern of selection would also have different implications for the policy simulations.

To estimate this parameter, we use an additional moment in our structural estimation: the ITT effect on TFP. Based on the discussion above, this moment can help us discern the degree to which higher-productivity firms have smaller certification costs, which would imply a positive value of κ.

Column 6 of Table A.7 reports the results of this estimation. We find a very small estimate of κ that is not statistically different from zero. We also do not see a significant change in the fit of our model, suggesting that assuming that $\kappa = 0$ in our baseline model does not bias our main estimates.

K.7 Robustness to Measures of Firm Productivity

As we discuss in Appendix C, our main measure of firm productivity is based on industry-level cost shares. For robustness, we also obtain measures of productivity following Ackerberg, Caves and Frazer (2015, ACF). Column 7 of Table A.7 reports the results of a structural estimation that replaces two of our baseline data moments (average TFP below and above the notch) with their counterparts based on this alternative measure of productivity. Panel A shows that we obtain very similar estimates of our structural parameters using this alternative measure of firm productivity.

Panel B reports simulated moments and shows that this model continues to fit all of the other moments quite well. Because this model does not target the TFP moments reported in column 1, we now evaluate the fit of these moments. The ACF measure of productivity yields an average TFP below the notch of -2.9% (c.f., -1.5% in our baseline) and of 2.3% (c.f., 2.7% in our baseline) above the notch. The simulated moments for these moments are -2.3% and 2.4%, which is quite a good fit given that the model also matches 11 other moments.

K.8 Validating Model Estimates Using ITT Estimates as Out-of-Sample Predictions

As an out-of-sample validation of our model, we consider additional moments based on the treatment effects of the program. Appendix F discusses the estimation of these moments, $m^{ITT}(\Omega)$, which include the treatment effects on the administrative expense ratio and on TFP growth. These moments are useful since the ITT estimates do not depend on a particular structure.

Let $\omega = \{\phi_1, b, c\}$ denote a firm with random draws of its fundamentals—i.e., productivity, adjustment cost, and fixed cost. We construct moments that match the empirical and simulated counterparts of the ITT estimates:[55]

$$m^{ITT}(\Omega) = \int_{d^{\text{No Notch}}(\omega)\in(d^{*-},d^{*+})} E[Y(\omega;\text{Notch}) - Y(\omega;\text{No Notch})]dF_\omega(\Omega) - \widehat{ITT^Y},$$

[55]Note that the simulated ITT restricts the support of $\omega = \{\phi_1, b, c\}$ to firms in the excluded region.

where $\widehat{ITT^Y}$ is an estimate from Section F.

The ITT estimate on measured TFP growth is related to ε. Note, however, that this estimate combines three mechanisms: the returns to R&D, selection into the treatment, and the potential for relabeling. In practice, we find that the relabeling margin plays an important role in influencing these ITT moments. For this reason, the ITT estimate on the administrative expense ratio is also related to both η and ε.

Table A.8 shows that our model implies ITT effects on productivity and administrative costs that are similar to those measured in the data.

L Welfare Analysis

This appendix provides details on the analysis in Section 6.3.

L.1 Setup: Firm Optimization

Firms produce a composite good with a CES technology $Q_t = \left[\sum_i q_{i,t}^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}}$. The residual demand for firm i's variety implies that this firm sets $p_{i,t}$ to solve

$$\max_{p_{i,t}} \ (1-t)\left[p_{i,t}^{1-\theta} - \frac{w_t}{\exp\{\phi_{i,t}\}} p_{i,t}^{-\theta}\right] B_t,$$

where $w_t = 1$, $B_t = I_t P_t^{\theta-1}$, which implies $p_{i,t} = \frac{\theta}{\theta-1}\left(\frac{1}{\exp\{\phi_{i,t}\}}\right)$ and taxable profits of $\pi_{i,t} = \frac{(\theta-1)^{\theta-1}}{\theta^\theta}\left(\frac{1}{\exp\{\phi_{i,t}\}}\right)^{1-\theta} B_t$.

Industry prices are given by:

$$P_t^{1-\theta} = \sum_i p_{i,t}^{1-\theta} = (\frac{\theta}{\theta-1})^{1-\theta} \sum_i \left(\frac{1}{\exp\{\phi_{i,t}\}}\right)^{1-\theta} = \left(\frac{\theta}{\theta-1}\right)^{1-\theta} \frac{1}{\Phi_t^{1-\theta}}$$

where $\Phi_t^{\theta-1} = \sum_i \exp\{\phi_{i,t}\}^{\theta-1}$.

Overall profits are given by:

$$\begin{aligned} \Pi_t &= \sum_i \pi_{i,t} = \sum_i \frac{(\theta-1)^{\theta-1}}{\theta^\theta}\left(\frac{1}{\exp\{\phi_{i,t}\}}\right)^{1-\theta} B_t \\ &= \frac{(\theta-1)^{\theta-1}}{\theta^\theta} \Phi_t^{-(1-\theta)} B_t \\ &= \frac{1}{\theta} I_t. \end{aligned}$$

That is, overall industry profits are proportional to I_t. Note, however, that even though industry profits are constant, each individual firm still has an incentive to invest in R&D and to participate in the InnoCom program, since:

$$\pi(\phi_{it}) = \frac{I_t}{\theta}\left(\frac{\exp\{\phi_{i,t}\}}{\Phi_t}\right)^{\theta-1}. \tag{L.1}$$

We can further write the expected profit (conditional on R&D decision D_{i1}) as

$$\frac{I}{\theta\Phi_2^{\theta-1}}\exp\{\phi_{i1}\}^{\rho(\theta-1)}D_{i1}^{(\theta-1)\varepsilon}\exp\left(\frac{(\theta-1)\sigma^2}{2}\right).$$

Given the fact that Φ_2 is an equilibrium object, we take it into account when solving the R&D choice of individual firms. In other words, the marginal benefit of R&D is decreasing in Φ_2. For this reason, we find the equilibrium that is consistent with firms' beliefs on Φ_2 and their optimal decisions.

L.2 Case 1: No R&D

We consider a representative household that has labor endowment L and owns all the production units. There is an outside sector that pins down worker wages such that $w_t = 1$. Firms operate in a monopolistic competitive environment such that household budget $I_t \equiv L + \Pi_t = \theta\Pi_t$. The total operating profit of firms is $\frac{1}{\theta-1}L$ and $I_t = \frac{\theta}{\theta-1}L$.

The consumer values two goods, a composite good C and a public good G with utility $U_t = C_t^{1-\gamma}G_t^{\gamma}$. For simplicity, we assume that the government/public sector simply purchases the composite good and transforms it into public good with a linear technology. In other words, the total demand for the composite good is $Q_t = C_t + G_t$.

The government finances G_t with a tax t on profits Π_t. Based on the firm optimization problem above, the price of the composite good is $P_t = \frac{\theta}{\theta-1}\Phi_t^{-1}$, and the aggregate profit is $\Pi_t = \frac{I_t}{\theta}$. Therefore:

$$G_t = t\frac{\Pi_t}{P_t} = t\frac{L}{\theta}\Phi_t$$

Consumption is then:

$$C_t = \frac{L + (1-t)\Pi_t}{P_t} = \frac{L + (1-t)\frac{I_t}{\theta}}{\frac{\theta}{\theta-1}\Phi_t^{-1}} = \frac{L}{\theta}(\theta - t)\Phi_t$$

The government sets t to maximize:

$$\frac{L\Phi_t}{\theta}[\theta - t]^{1-\gamma}[t]^{\gamma}$$

Taking the FOC:

$$\frac{L\Phi_t}{\theta}[\theta - t]^{1-\gamma}[t]^{\gamma} \times \left[-\frac{1-\gamma}{\theta - t} + \frac{\gamma}{t}\right] = 0$$

implies:

$$\frac{\gamma}{1-\gamma} = \frac{t}{\theta - t} \Longrightarrow \gamma = \frac{t}{\theta}.$$

Assuming $t = 0.25, \theta = 5$ implies $\gamma = 0.05$.

L.3 Case 2: R&D Investment without InnoCom Program

As in Section 4, we consider a two-period model. Since the firm only invests in R&D in the first period, the consumption bundles in period 2 correspond to those in case 1. The slight difference for the consumption bundles in the first period is that the total net firm profit of period 1 is $\Pi_1 - D_1$, where $D_1 = \sum_i (D_{i,1} + g_i(D_{i,1}))$. From the representative household perspective, it is important to note that its budget becomes $I_1 \equiv L + (\Pi_1 - D_1) = \theta \Pi_1$. As a result, $\Pi_1 = \frac{L-D_1}{\theta-1}$, $I_1 = \frac{\theta}{\theta-1}(L - D_1)$. The consumption bundles in the first period follow by substituting L with $L - D_1$ in case 1.

Utility is then:

$$\frac{(L-D_1)\Phi_1}{\theta}[\theta - t]^{1-\gamma}[t]^{\gamma} + \beta\frac{L\Phi_2}{\theta}[\theta - t]^{1-\gamma}[t]^{\gamma} = \left[\frac{(L-D_1)\Phi_1}{\theta} + \beta\frac{L\Phi_2}{\theta}\right] \times [\theta - t]^{1-\gamma}[t]^{\gamma}.$$

Since t does not affect the choice of D_1, the optimal tax rate in the two-period case is the same as in the one-period case.

L.4 Case 3: R&D Investment with the InnoCom Program

Let $\mathbb{I}(\text{InnoCom}_i)$ denote the event that firm i is part of the InnoCom program and redefine:

$$D_1 = \sum_i \left(D_{i,1} + g_i(D_{i,1}) + \mathbb{I}(\text{InnoCom}_i)c_i\right).$$

Note that, in contrast to the linear tax case, $D_{i,1} + g_i(D_{i,1})$ are affected for those firms that participate in the program. Aggregate relabeling costs are given by:

$$H_1 = \sum_i \mathbb{I}(\text{InnoCom}_i)h(\tilde{D}_{i,1}).$$

The value of the tax credit for the InnoCom program is given by:

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

where t^{LT} is the standard rate and t^{HT} is the preferential rate for high-tech firms in the InnoCom program.

Income in the first period decreases by the relabeling penalty H_1, so that $I_1 = \frac{\theta}{\theta-1}(L - D_1 - H_1)$. This also implies that $G_1 = t\frac{L-D_1-H_1}{\theta}\Phi_1$ and that $C_1 = \frac{L-D_1-H_1}{\theta}(\theta - t)\Phi_1$. Utility in the first period is then:

$$\frac{(L-D_1-H_1)\Phi_1}{\theta}[\theta - t]^{1-\gamma}[t]^{\gamma}.$$

In the second period, G decreases by the tax credit, which also increases after-tax profits. $I_2 = L + \Pi_2 = \frac{\theta}{\theta-1}L$ so $\Pi_2 = \frac{L}{\theta-1}$. This implies that:

$$G_2 = \frac{t\Pi_2 - TC}{P_2} = \frac{t\frac{L}{\theta-1} - TC}{\frac{\theta}{\theta-1}\Phi_2^{-1}} = \left(t\frac{L}{\theta} - TC\frac{\theta-1}{\theta}\right)\Phi_2.$$

Consumption is then:

$$C_2 = \frac{L + (1-t)\Pi_2 + TC}{P_2} = \frac{L + (1-t)\frac{L}{\theta-1} + TC}{\frac{\theta}{\theta-1}\Phi_2^{-1}} = \left(\frac{L}{\theta}(\theta - t) + TC\frac{\theta-1}{\theta}\right)\Phi_2.$$

Utility in the second period is:

$$\Phi_2\left(\frac{L}{\theta}(\theta - t) + TC\frac{\theta-1}{\theta}\right)^{1-\gamma}\left(t\frac{L}{\theta} - TC\frac{\theta-1}{\theta}\right)^{\gamma}$$

and overall utility is now:

$$\frac{\Phi_1}{\theta}(L - D_1 - H_1)(\theta - t)^{1-\gamma}t^{\gamma} + \frac{\beta\Phi_2}{\theta}(L(\theta - t) + TC(\theta-1))^{1-\gamma}(tL - TC(\theta-1))^{\gamma}.$$

To further simplify this expression, let τ denote the InnoCom tax credits as a fraction of overall profits in the second period, and note that $\tau = \frac{TC}{\Pi_2} = \frac{TC(\theta-1)}{L}$. We then have:

$$\frac{\Phi_1}{\theta}(L - D_1 - H_1)(\theta - t)^{1-\gamma}t^{\gamma} + \frac{\beta\Phi_2}{\theta}L(\theta - t + \tau)^{1-\gamma}(t - \tau)^{\gamma},$$

which is Equation 6 in the text.

L.5 Implementation and Spillovers

We now discuss how we implement this framework to account for knowledge spillovers. To do so, we take into account the empirical fact that not all firms engage in R&D. We therefore assume that $N^{\text{R\&D}}$ firms engage in R&D. For R&D-performing firms, we expand the baseline model by assuming that:

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

where $S_{t-1} = \frac{1}{N^{\text{R\&D}}}\sum_{i\in N^{\text{R\&D}}}\ln(D_{i,t-1})$ is the R&D spillover pool. Log productivity for non-R&D firms evolves as follows:

$$\phi_{i,t} = \rho\phi_{i,t-1} + \zeta S_{t-1} + u_{it}.$$

Note that, because firms do not internalize their individual R&D impact on the spillover pool, they take S_{t-1} as given.

We now show that S_{t-1} does not impact R&D investment decisions. Let $\tilde{\phi}_{i,t} = \phi_{i,t} - \zeta S_{t-1}$ and $\tilde{\Phi}_t = \Phi_t \times \exp\{-\zeta S_{t-1}\}$ denote firm and aggregate productivity measures net of spillover effects. Because $\frac{\exp\{\phi_{i,t}\}}{\Phi_t} = \frac{\exp\{\tilde{\phi}_{i,t}\}}{\tilde{\Phi}_t}$, it follows that firm profits in Equation L.1 are maximized by the same R&D investment decision regardless of the value of S_{t-1}.

While spillovers do not affect the R&D investment decisions of individual firms, positive spillovers raise aggregate productivity in the second period through $\Phi_2 = \tilde{\Phi}_2 \times \exp\{\zeta S_1\}$. Therefore, we can use Equation 6 to evaluate welfare in both the case with and the case without spillover effects.

To implement Equation 6, we need to compute Φ_2 as an equilibrium object. This is because the overall level of R&D in the economy lowers Φ_2 as well as expected profits. Our implementation of Equation 6 requires that the resulting equilibrium be consistent with firms' belief on Φ_2 and therefore with their optimal investment decisions.

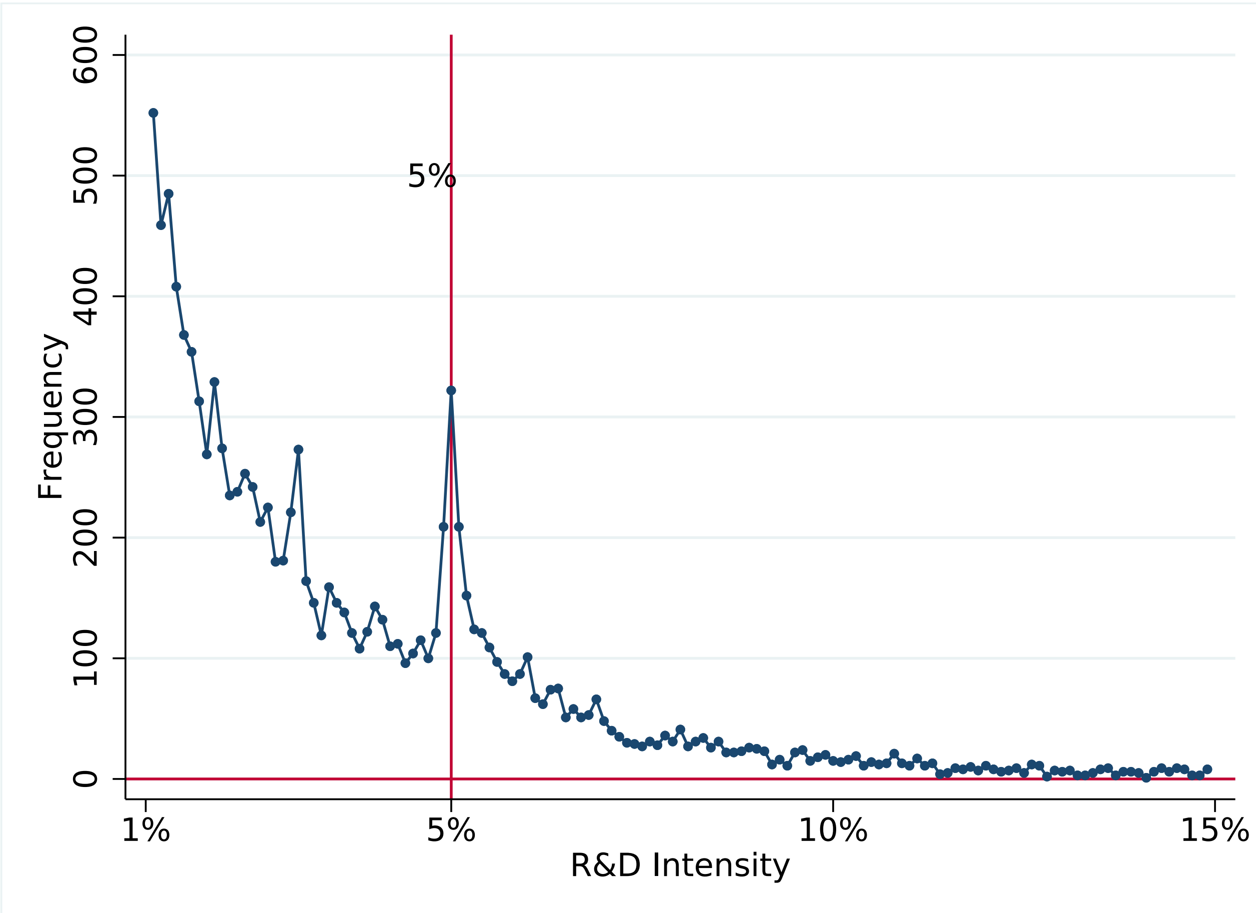
To compute such an equilibrium, it helps to write:

$$\Phi_t^{\theta-1} = (\Phi_t^{\text{R\&D}})^{\theta-1} + (\Phi_t^{\text{Non-R\&D}})^{\theta-1},$$

where $(\Phi_t^{\text{R\&D}})^{\theta-1} = \sum_{i \in N^{\text{R\&D}}} \exp\{\phi_{i,t}\}^{\theta-1}$ and similarly for $\Phi_t^{\text{Non-R\&D}}$. Further, the sales share of R&D-performing firms is equal to $\frac{(\Phi_t^{\text{R\&D}})^{\theta-1}}{\Phi_t^{\theta-1}}$. In a given simulation, we can compute $\Phi_2^{\text{R\&D}}$ from our model output. We then pick $\Phi_2^{\text{Non-R\&D}}$ so that the sales share of R&D firms in the second period equals 35%. Note that, in the case with spillovers, we adjust $\Phi_2^{\text{Non-R\&D}}$ to account for the effect of spillovers.

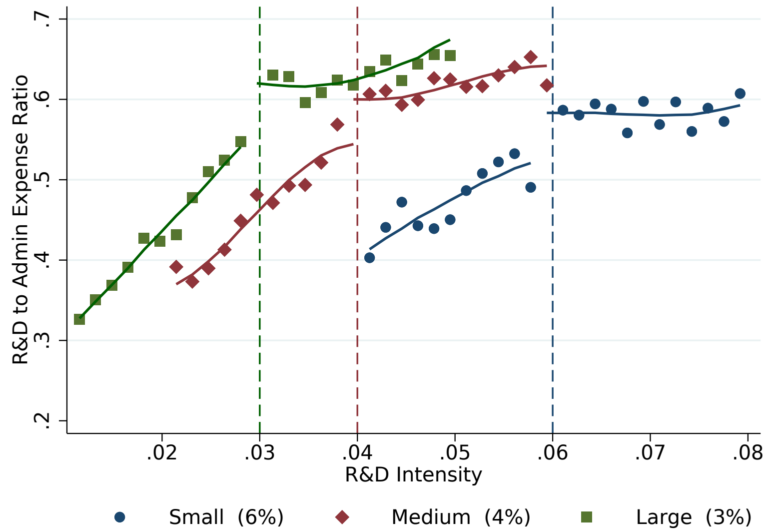
Appendix Graphs

Figure A.1: Bunching at 5% R&D Intensity (2005–2007)



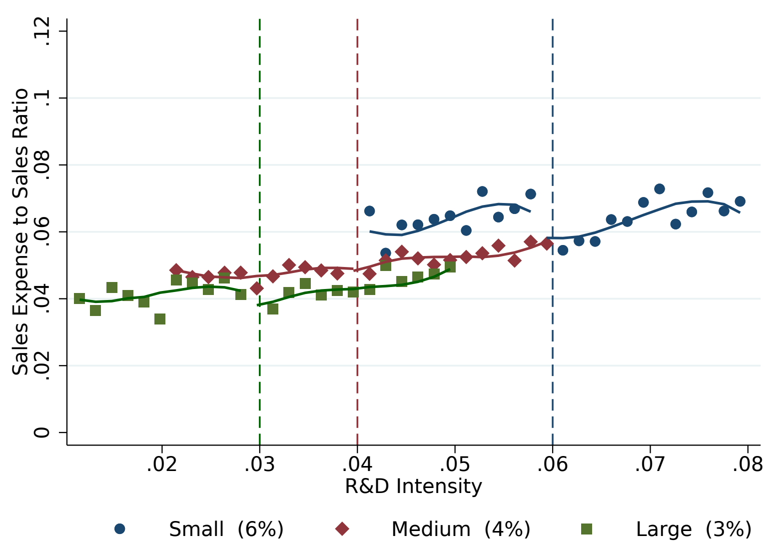
Notes: This figure plots the R&D intensity distribution of manufacturing firms conducting R&D during the period of 2005 to 2007. We include the firms with R&D intensity between 1% and 15%. There is a significant bunching of firms at the 5% threshold. Source: Annual Survey of Manufacturers. See Section 3.1 for details.

Figure A.2: Alternative Empirical Evidence of Relabeling



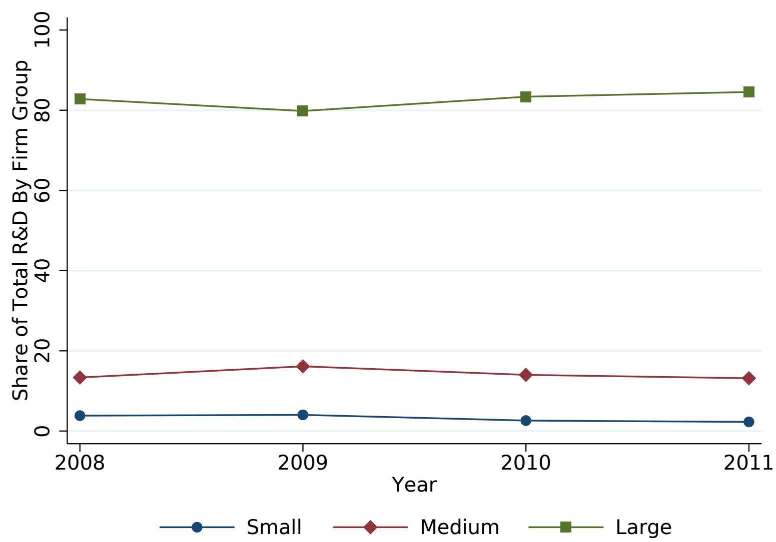
Notes: This figure summarizes the ratio of R&D to administrative expenses for small, medium, and large firms in our sample. The figure shows that this ratio jumps discontinuously across the thresholds of R&D intensity prescribed by the InnoCom program. This suggests that firms manipulate their reported R&D intensity by relabeling non-R&D administrative expenses as R&D. See Table A.4 for estimates of the structural break.

Figure A.3: Lack of Manipulation of Sales Expenses



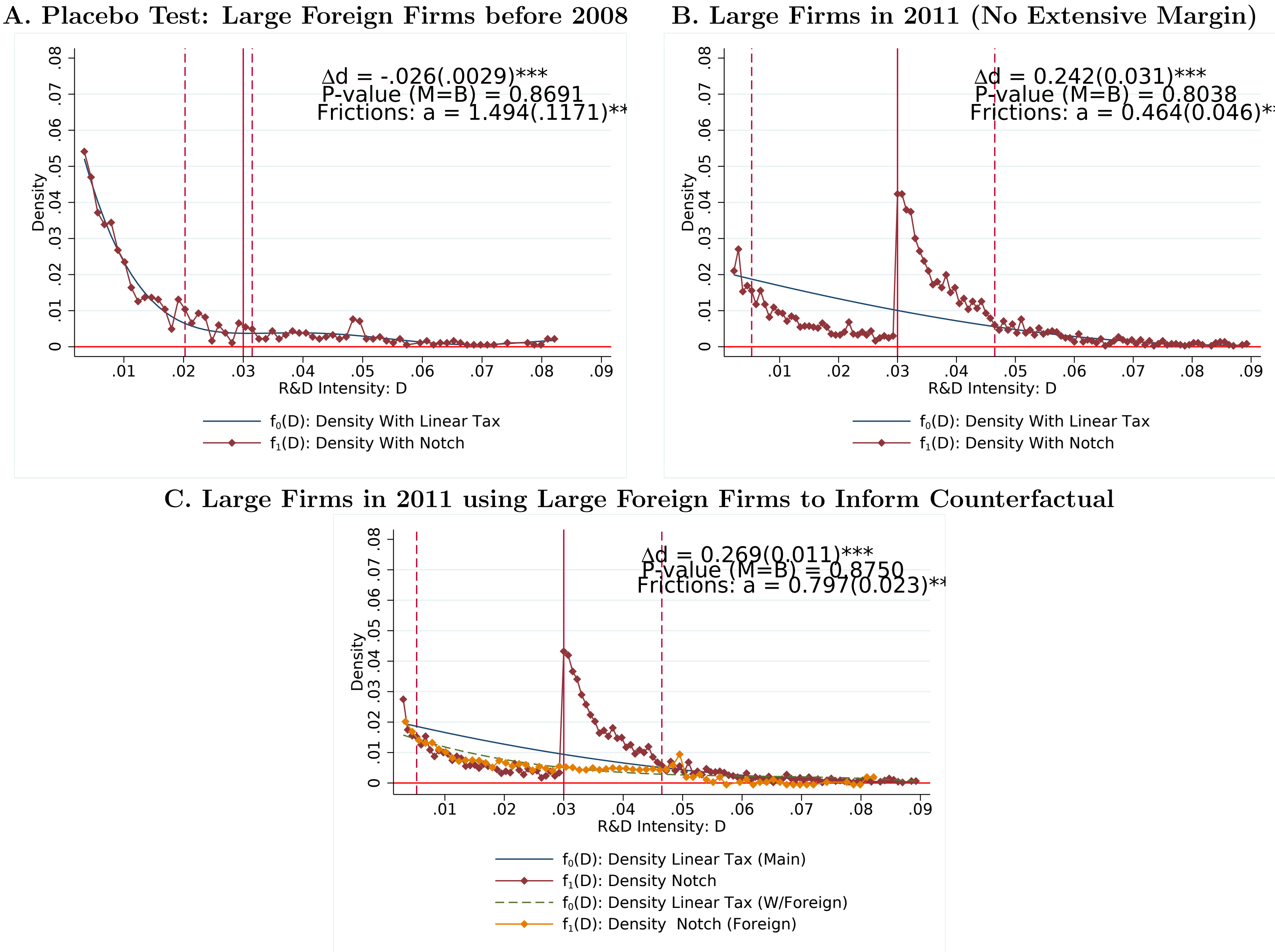
Notes: This figure shows the binned plot of the sales expense-to-sales ratio for each firm size category. Table A.5 shows that we do not find a detectable drop in this ratio at the notches.

Figure A.4: Aggregate Implications



Notes: This figure summarizes the share of total R&D accounted for by the small, medium, and large firms in our sample. As the figure illustrates, the large firms account for more than 80% of total R&D and thus are the most important group for the aggregate implications of the policy.

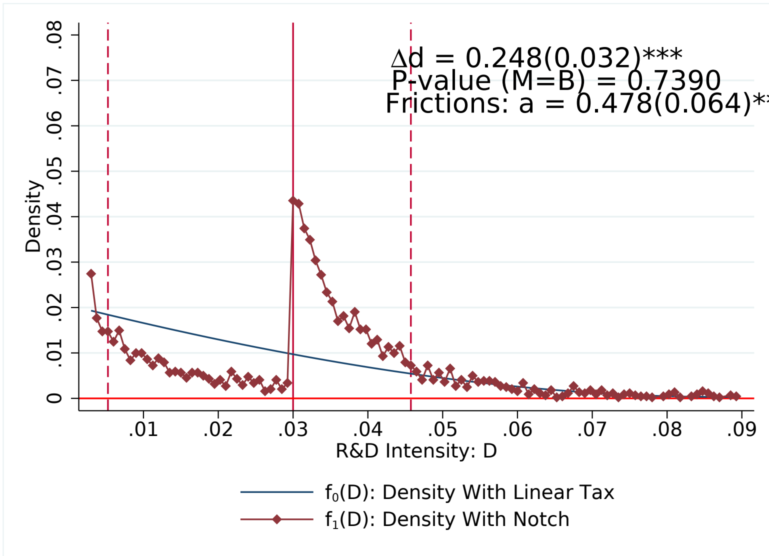
Figure A.5: Robustness of Bunching Estimates



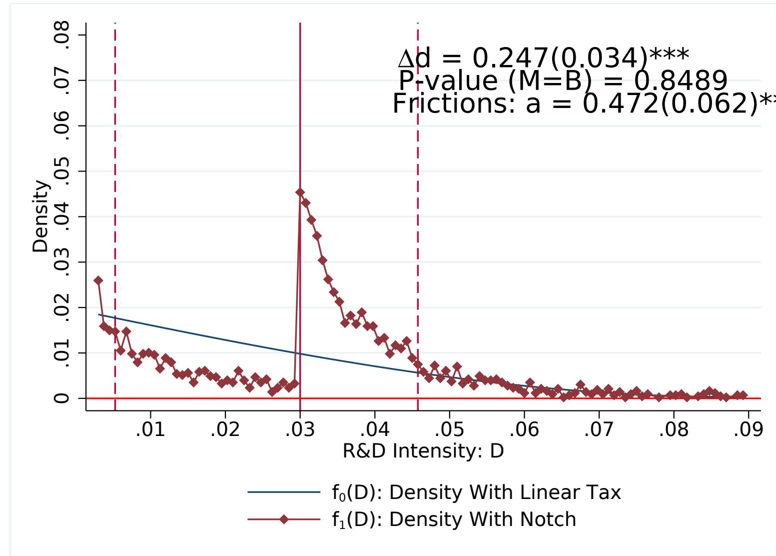
Notes: This figure reports robustness checks of our bunching estimator in panel C of Figure 4. Panel A reports a placebo test where we use the data from large foreign firms before 2008. Panel B implements our bunching estimator for large firms that had engaged in R&D in previous years. Panel C uses large foreign firm's R&D intensity before 2008 to inform the counterfactual distribution. See Appendix E for details.

Figure A.6: Robustness of Bunching Estimates to Dropping Groups of Firms

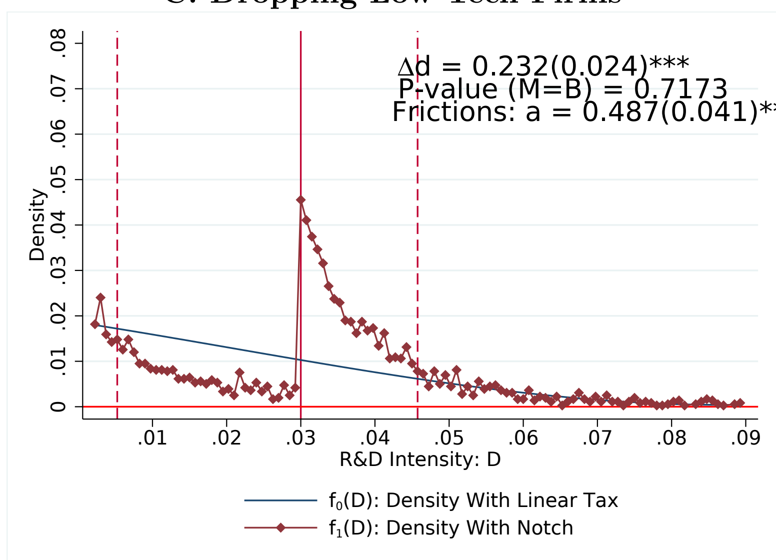
A. Dropping SOEs



B. Dropping Low-Profitability Firms



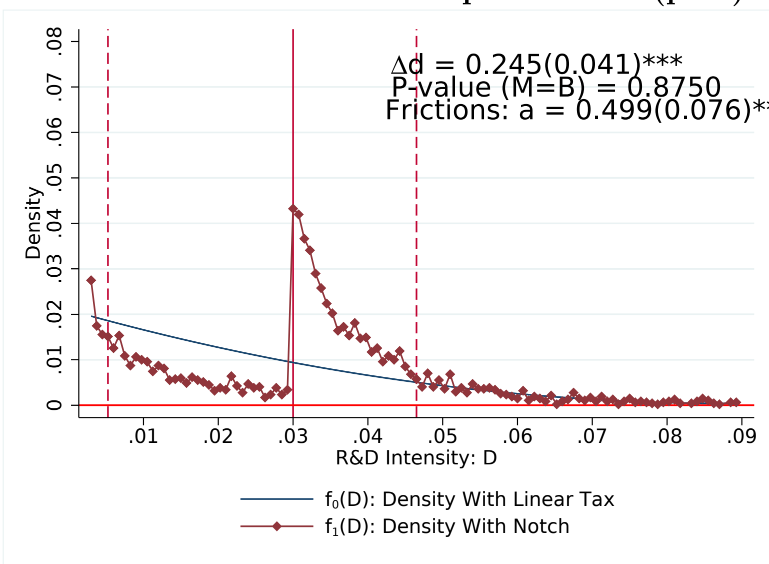
C. Dropping Low-Tech Firms



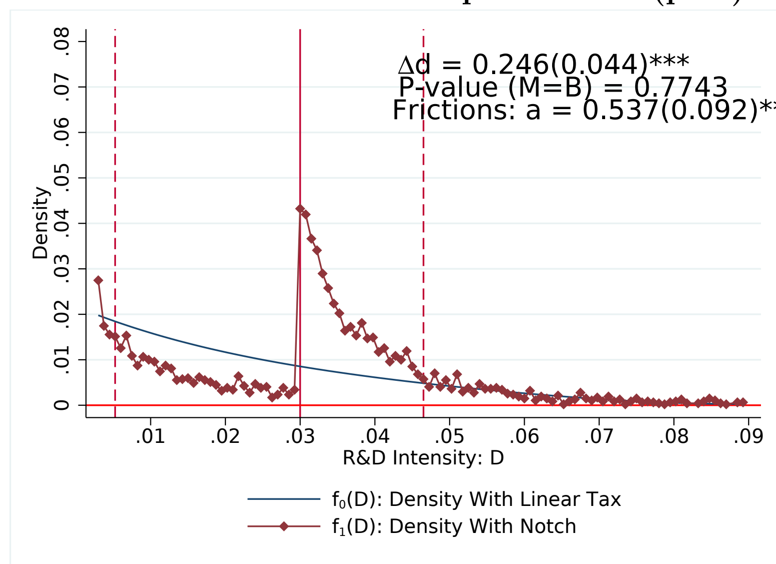
Notes: This figure presents robustness checks of the benchmark bunching analysis for large firms in 2011. In panel A, we drop state-owned enterprises. In panel B, we drop the bottom 20% of firms in terms of profitability. In panel C, we drop all firms not classified in the high-tech industries defined by the Chinese government. These graphs show that our benchmark results are robust across these subsamples. See Appendix E for details.

Figure A.7: Robustness of Bunching Estimates to Specification of Counterfactual Density

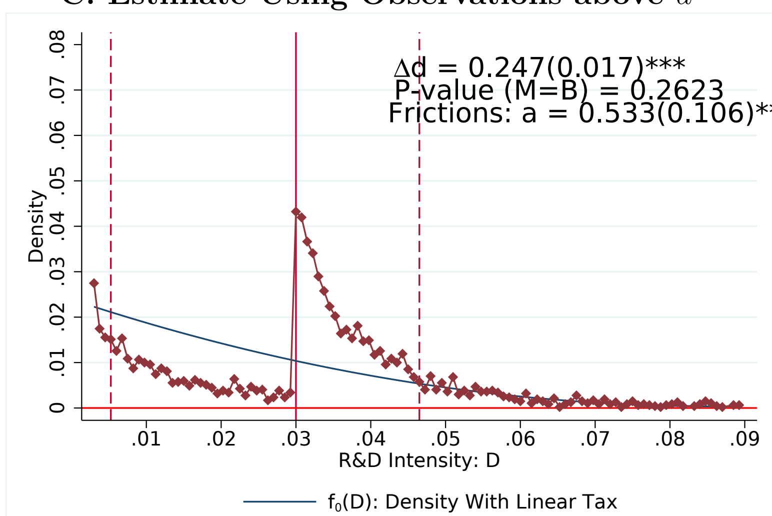
A. Second-Best Choice of Specification (p=3)



B. Second-Best Choice of Specification (p=4)

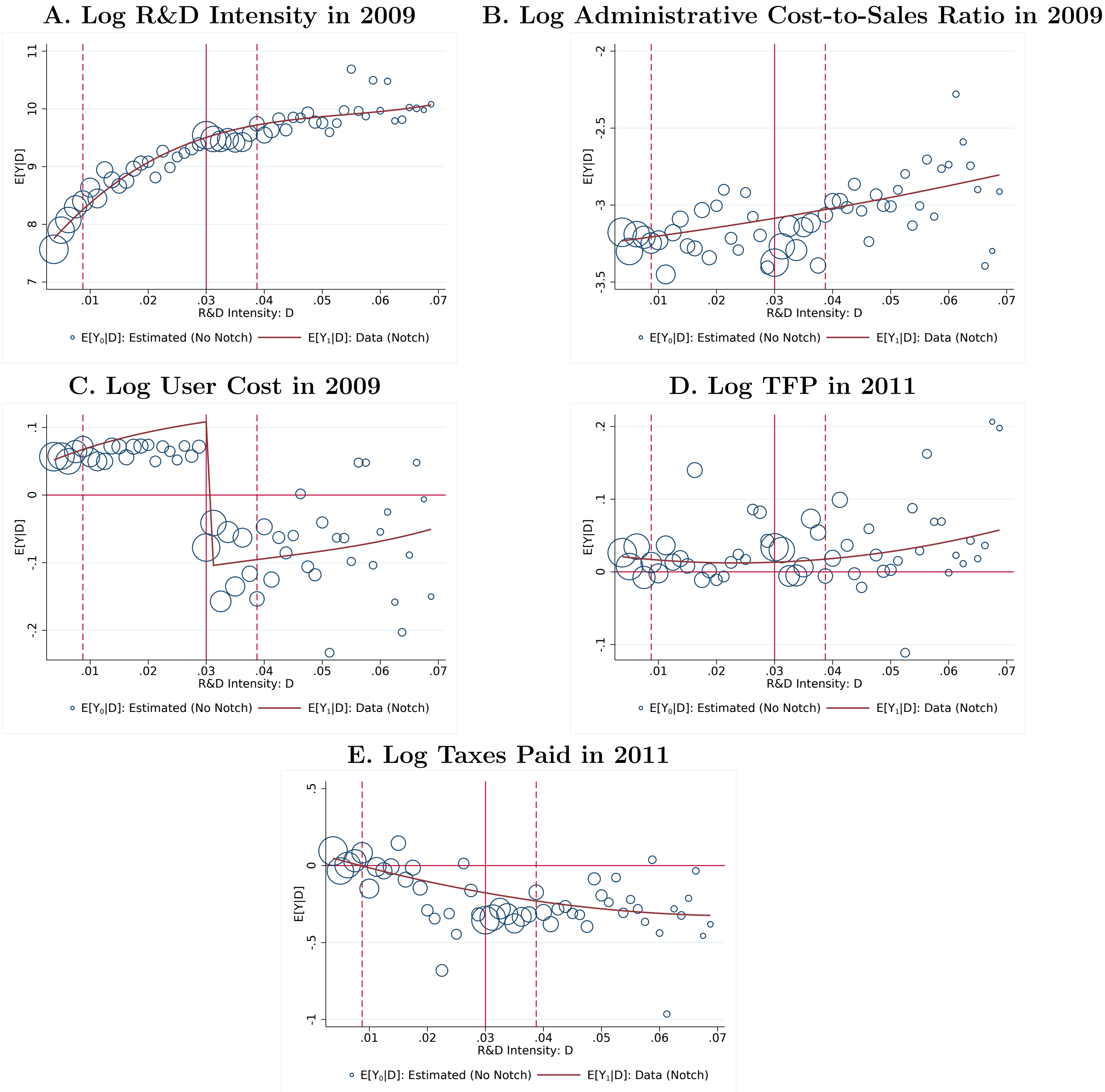


C. Estimate Using Observations above d^{*+}



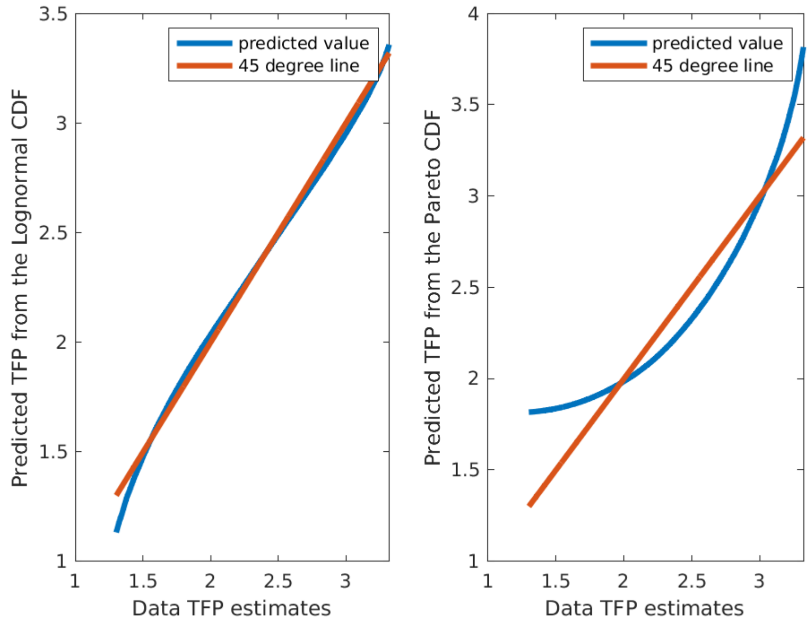
Notes: This figure conducts robustness checks of the benchmark bunching analysis for large firms in 2011. As discussed in Appendix D, we select (p, d^{*-}, d^{*+}) via cross-validation. In panel A, we use the second-best choice for the specification of (p, d^{*-}, d^{*+}). As in our benchmark case, $p = 3$. In panel B, we further restrict $p = 4$, and we select (d^{*-}, d^{*+}) via cross-validation. In panel C, we use the same value of d^{*+} as in our benchmark case, and we only use data above this value when estimating the counterfactual density. These graphs show that our benchmark results are robust to how we specify (p, d^{*-}, d^{*+}).

Figure A.8: Estimated Values of $E[Y|d]$ for the ITT Analysis



Notes: This figure reports the polynomial regression of binned outcome variables on R&D intensity. The size of each circle indicates the weights based on the number of observations accounted for by each bin. We leave out all the observations in the manipulated region. Overall, these graphs show a good fit of the data outside of the exclusion region. The fit in the exclusion region cannot be evaluated since the data patterns may be due to selection. See Appendix F.1 for more details.

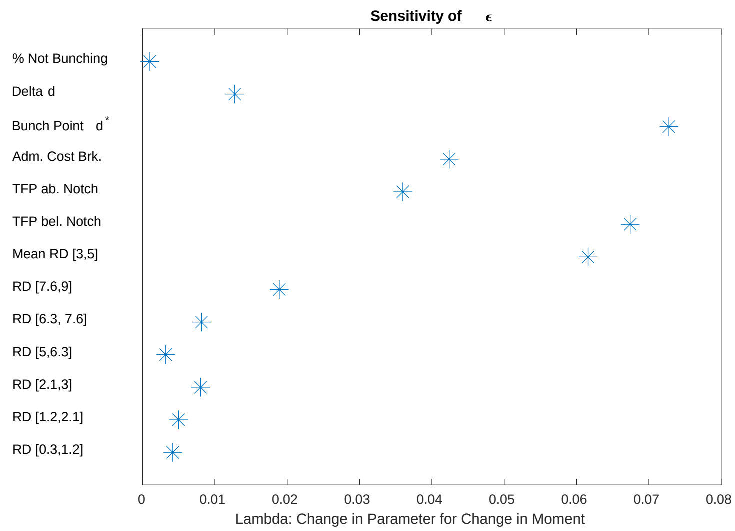
Figure A.9: Observed TFP and Predicted TFP under a Log-Normal vs. Pareto Distribution



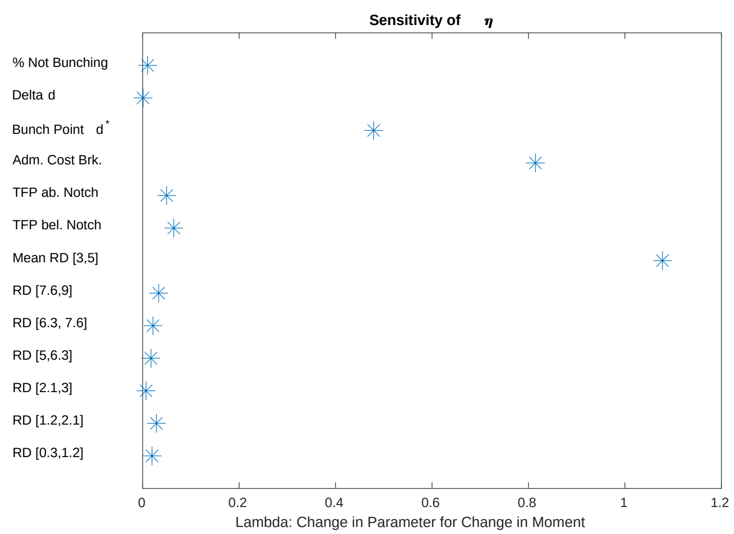
Notes: This figure reports the predicted TFP from imposing a log-normal CDF, a Pareto CDF, and the 45 degree linear line. For both cases, we trim the observed TFP data at 1% and 99%. The figure shows that the predicted TFP from imposing the log-normal CDF tracks observed TFP quite well. It thus provides strong evidence that log-normal is a reasonable parametric assumption for the TFP distribution.

Figure A.10: Sensitivity Analysis

A. Sensitivity Analysis for ε



B. Sensitivity Analysis for η



Notes: This figure reports the results of a sensitivity analysis based on Andrews, Gentzkow and Shapiro (2017). We report estimates of the sensitivity matrix Λ, which captures how a local change in each moment affects the parameter estimates.

Table A.1: Estimates of Treatment Effects

A. Estimates of Intent-to-Treat (ITT) Effects

	ITT	SE	T-Stat	Bootstrap 5th Perc.	Bootstrap 95th Perc.
2009					
Admin Costs	-0.096	0.025	-3.887	-0.134	-0.053
Admin Costs (levels)	-0.003	0.001	-3.742	-0.005	-0.002
R&D	0.146	0.064	2.272	0.036	0.247
R&D (real)	0.090	0.044	2.078	0.022	0.165
User Cost of Capital	-0.071	0.037	-1.930	-0.134	-0.010
2011					
Tax	-0.128	0.018	-7.186	-0.160	-0.101
TFP	0.012	0.006	1.999	0.002	0.022

B. Estimates of User-Cost-of-Capital Elasticities

	Estimate	Bootstrap 5th Perc.	Bootstrap 95th Perc.
Reported R&D to UCC	-2.052	-8.345	-0.093
Real R&D to UCC	-1.272	-5.441	-0.055
Tax to Reported R&D	-0.879	-2.860	-0.462

Notes: This table reports estimates of ITT effects of the notch on various outcomes. Panel B reports ratios of the estimates in panel A. Standard errors computed via bootstrap. See Section 3.1 for details on data sources and Appendix F for details on the estimation. Source: Administrative Tax Return Database.

$$ITT = \frac{1}{N^{Excluded}} \sum_{i \in (D^{*-}, D^{*+})} Y_i - \int_{D^{*-}}^{D^{*+}} \hat{f}_0(r) E[Y | r\widehat{d, \text{No}}\,\text{Notch}] dr$$

Appendix Tables

Table A.2: Manipulation of the Administrative Expense-to-Sales Ratio

	(1) Small	(2) Medium	(3) Large
Structural Break	-0.014**	-0.013***	-0.008***
	(0.007)	(0.004)	(0.003)
Observations	5,016	8,336	8,794

Notes: This table reports estimates of the structural break at the notches in Figure 5. The table shows that the ratio of administrative expenses to sales drops across the notches of the InnoCom program, which suggests that firms qualify for the InnoCom program by relabeling non-R&D expenses as R&D. See Section 3.1 for details on data sources and Section 3.3 for details on the estimation. Standard errors in parentheses. Source: Administrative Tax Return Database.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table A.3: Lack of Sales Manipulation at R&D Intensity Thresholds

	(1) Small	(2) Medium	(3) Large
Structural Break	0.108	-0.021	0.055
	(0.103)	(0.067)	(0.114)
Observations	1,096	1,952	1,665

Notes: This table reports estimates of the structural break at the notches of panel A in Figure 6. The table shows that firms do not manipulate their sales to qualify for the InnoCom program. See Section 3.1 for details on data sources and Section 3.3 for details on the estimation. Standard errors in parentheses. Source: Administrative Tax Return Database.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table A.4: Alternative Estimates of Manipulation of Administrative Expenses

	(1) Small	(2) Medium	(3) Large
Structural Break	0.053**	0.056***	0.054**
	(0.026)	(0.020)	(0.022)
Observations	3,544	5,710	5,597

Notes: This table reports estimates of the structural break at the notches in Figure A.2. The table shows that the ratio of administrative expenses to R&D jump across the notches of the InnoCom program, which suggests that firms qualify for the InnoCom program by relabeling non-R&D expenses as R&D. See Section 3.1 for details on data sources and Section 3.3 for details on the estimation. Standard errors in parentheses. Source: Administrative Tax Return Database.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Lack of Manipulation of Sales Expenses at R&D Intensity Thresholds

	(1) Small	(2) Medium	(3) Large
Structural Break	-0.002	-0.000	-0.001
	(0.006)	(0.004)	(0.004)
Observations	4,774	8,064	8,600

Notes: This table reports estimates of the structural break at the notches in Figure A.3. The table shows that firms do not manipulate sales expenses to qualify for the InnoCom program. See Section 3.1 for details on data sources and Appendix G for details on the estimation. Standard errors in parentheses. Source: Administrative Tax Return Database.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table A.6: Structural Estimates with Heterogeneous ε

A. Point Estimates

	Model A ($\alpha_\varepsilon = \beta_\varepsilon$)				
	Distribution of TFP Elasticity of R&D		Relabeling Cost	Adjustment Cost	Distribution of Fixed Costs
	α_ε	$\bar{\varepsilon}$	η	μ_b	μ_c
Estimate	1.001	0.145	7.711	7.666	0.482
SE	0.058	0.003	0.809	0.065	0.024

	Model B ($\bar{\varepsilon} = 1/(\theta - 1)$)				
	Distribution of TFP Elasticity of R&D		Relabeling Cost	Adjustment Cost	Distribution of Fixed Costs
	α_ε	β_ε	η	μ_b	μ_c
Estimate	2.616	3.766	7.147	8.271	0.741
SE	0.101	0.177	0.366	0.061	0.039

Notes: This table reports estimates of structural parameters of the model in Appendix K. Estimates based on calibrated values of $\theta = 5$, $\rho = 0.725$, and $\sigma = 0.385$.

B. Simulated vs. Data Moments

		Data	Model A	Model B
R&D Distribution Moments: $m^D(\Omega)$				
Below the notch (%):	$[0.3, 1.2]$	0.373	0.396	0.371
	$[1.2, 2.1]$	0.113	0.163	0.193
	$[2.1, 3]$	0.067	0.056	0.057
Above manipulated region (%):	$[5, 6.3]$	0.056	0.053	0.046
	$[6.3, 7.6]$	0.026	0.024	0.030
	$[7.6, 9]$	0.012	0.011	0.020
Mean R&D intensity $[3\%, 5\%]$		0.037	0.035	0.034
Average TFP below notch		-0.015	-0.028	-0.034
Average TFP above notch		0.027	0.009	0.004
Admin cost ratio break at notch		-0.009	-0.005	-0.007
Bunching Moments: $m^B(\Omega)$				
Bunching Point d^{-*}		0.009	0.012	0.011
Increase in Reported R&D: Δd		0.157	0.138	0.142
Fraction of firms not bunching		0.641	0.611	0.642

Notes: This table compares the moments generated by our simulations with those from the data. The simulation is based on 30,000 firms. Appendix K discusses details of each of these robustness checks.

Table A.7: Robustness of Structural Estimates

A. Parameter Estimates	Data	Baseline	Alternative Adjustment Cost	Alternative Relabeling Cost	Real Admin Cost Response	TFP-Fixed Cost Correlation	Alternative TFP Measure
TFP Elasticity of R&D ε		0.091	0.091	0.104	0.090	0.091	0.092
		0.002	0.002	0.003	0.002	0.002	0.002
Adjustment Cost (Mean) μ_b		8.011	8.018	8.545	8.064	8.014	7.914
		0.075	0.096	0.114	0.084	0.097	0.076
Adjustment Cost (Dispersion) σ_b		2.014	2.011	2.366	2.012	2.016	2.035
		0.073	0.077	0.113	0.072	0.075	0.078
Fixed Costs μ_c		0.532	0.532	0.652	0.513	0.532	0.532
		0.012	0.027	0.032	0.020	0.024	0.013
Relabeling Cost η		6.755	6.754	2.713	7.655	6.759	6.114
		0.449	0.491	0.160	0.446	0.452	0.348
Adjustment Cost (Scale) ν			0.001				
			0.004				
TFP-Fixed Cost Correlation κ						-0.001	
						0.029	
B. Simulated vs. Data Moments							
R&D Distribution Moments: $m^D(\Omega)$							
Below the notch (%): $[0.3, 1.2]$	0.373	0.379	0.379	0.414	0.395	0.380	0.359
$[1.2, 2.1]$	0.113	0.146	0.146	0.133	0.150	0.146	0.141
$[2.1, 3]$	0.067	0.069	0.069	0.059	0.067	0.069	0.069
Above manipulated region (%): $[5, 6.3]$	0.056	0.057	0.057	0.057	0.055	0.057	0.060
$[6.3, 7.6]$	0.026	0.038	0.038	0.041	0.036	0.038	0.042
$[7.6, 9]$	0.012	0.027	0.026	0.030	0.025	0.026	0.029
Mean R&D intensity [3%, 5%]	0.037	0.035	0.035	0.035	0.035	0.035	0.035
Average TFP below notch	-0.015	-0.020	-0.020	-0.025	-0.018	-0.020	-0.023
Average TFP above notch	0.027	0.025	0.025	0.033	0.024	0.025	0.024
Admin cost ratio break at notch	-0.009	-0.007	-0.007	-0.007	-0.007	-0.007	-0.008
Bunching Point d^{-*}	0.009	0.010	0.010	0.010	0.010	0.010	0.009
Increase in Reported R&D: Δd	0.157	0.150	0.150	0.149	0.145	0.150	0.157
Fraction of firms not bunching	0.641	0.665	0.665	0.655	0.658	0.665	0.670
ITT TFP	0.012					0.009	

Notes: This table compares the moments generated by our simulations with those from the data. The simulation is based on 30,000 firms. Appendix K discusses the details of each of these robustness checks.

Table A.8: Structural Estimates and ITT Moments as Over-Identifying Moments

Simulated vs. Data Moments

	Data	Simulated	
		Excl. Bunching	All
ITT Moments: $m^{ITT}(\Omega)$			
ITT TFP	0.012	(0.007)	(0.009)
ITT administrative cost ratio	-0.33%	(-0.20%)	(-0.22%)

Notes: This table compares the moments generated by our simulations with those from the data. The simulation is based on $30,000$ firms. The table shows that our model matches these moments that are not targeted in the estimation.