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

Zhao Chen
Fudan University

Zhikuo Liu
Shanghai University of Finance and Economics

Juan Carlos Suárez Serrato
Duke University
& NBER

Daniel Yi Xu
Duke University
& NBER

March 2019

Abstract

We analyze the effects of a Chinese policy that awards substantial corporate tax cuts to firms that increase R&D investment over a given threshold, or notch. We exploit this quasi-experimental variation with administrative tax data in order to shed light on longstanding questions on the effects of fiscal incentives for R&D. We find large responses of reported R&D using a cross-sectional "bunching" estimator that is new to the R&D literature. We also find significant increases in firm-level productivity, even though about 30% of the increase in R&D is due to relabeling of administrative expenses. Anchored by these reduced-form effects, we estimate a structural model of R&D investment and relabeling that recovers a 9.8% return to R&D. We simulate alternative policies and show that firm selection into the program and the relabeling of R&D determine the cost-effectiveness of the policy, and the effects on productivity growth.
JEL Codes: D24, O30, H25, H26. Keywords: R&D, China, relabeling, corporate taxes

*We are very grateful for comments from Manuel Adelino, Ashish Arora, Pierre Bachas, Michael Best, 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 University, Penn State, Stanford (SIEPR & Econ), Toronto, UCLA, UCSD (Econ & Bunching Conference), University of Alberta, University of Maryland, University of Melbourne, Warwick University, and ZEW MaTax. Dan Garrett, Yuxuan He, 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

It is widely believed that economic growth is highly dependent on innovation and, in particular, on R&D investment. For this reason, governments around the world encourage R&D investment through tax incentives. As China's development through industrialization reaches a mature stage, the country's leaders have focused their efforts on fostering technology-intensive industries as a source of future growth. Figure 1 compares the explosive growth of R&D in China to the experience of other countries and shows that China has now equaled or surpassed developed-country levels of R&D intensity. This paper analyzes the effects of China's InnoCom program: a large fiscal incentive for R&D investment in the form of a corporate income tax cut. We exploit a novel administrative dataset of corporate tax returns of Chinese firms as well as sharp and changing tax incentives to provide new estimates of the effects of fiscal incentives on R&D investment and productivity growth.

This paper analyzes quasi-experimental variation in the InnoCom program to answer two questions that are of both policy and economic interest. First, is R&D investment responsive to fiscal incentives and, if so, do firms engage in evasion or manipulation of reported R&D in response to the tax incentives? Quantifying these effects is crucial for governments to determine the fiscal cost of the marginal increase of real R&D investment. Second, what is the effect of fiscal incentives on productivity growth, and how much do firms value R&D investment in terms of future profits? These questions are central to the design of government programs to encourage R&D investment.

Answers to these questions are often confounded by the lack of large and plausibly exogenous variation in tax incentives. Small fiscal incentives are unlikely to have measurable effects on R&D investment since R&D is often constrained by existing technological opportunities and usually requires fixed and adjustment costs. A second concern is that comparisons of investment and productivity across different firms will result in upwardly-biased returns of R&D since firms with better prospects for innovation are likely to invest more heavily in R&D. Additionally, it is often hard to determine whether firm responses to tax incentives for R&D investment correspond to real activity or to relabeling of expenses. If measured R&D is contaminated by relabeling, this might result in an upwardly-biased estimate of the user-cost elasticity of R&D investment and a downwardly-biased estimate of the R&D elasticity of TFP.

We overcome these concerns by leveraging an unusual and large fiscal incentive for R&D investment in China. Before 2008, firms with an R&D intensity (R&D investment over revenue) above 5% could qualify for a special status as high-tech firms that was accompanied by a lower average tax rate of 15%—a large reduction from the standard rate of 33%. After 2008, the government established three thresholds of 3%, 4%, and 6% for firms of different size categories. The use of average, as opposed to marginal incentives, creates a notch in the corporate income tax that generates very large incentives for firms to invest in R&D. Our tax data precisely measure a firm's R&D investment and exposure to fiscal incentives. In addition, we leverage the detail in our administrative data to analyze firm-level outcomes of interest, such as productivity, and to explore whether firms respond to the tax incentive by relabeling non-R&D expenses.

Our main result is that firms are highly responsive to the tax incentives in the InnoCom program, but that a significant fraction of the response is due to relabeling of non-R&D expenses. Despite the relabeling response, we find the program led to a significant increase in productivity and that accounting for relabeling results in larger estimates of the effects of R&D on productivity in our structural analysis. We use these insights to simulate the effects of alternative policies and show that the cost-effectiveness of the tax incentive for R&D is driven by firm selection into the program, which influences the effects of the policy on investment, relabeling, and productivity growth.

Our analysis proceeds in four steps. We first provide descriptive evidence that the R&D notches have significant effects on firms' reported R&D intensity and that part of this response may be due to relabeling of non-R&D expenses. We show that a large number of firms choose to locate at the notches and that introducing the tax cut led to a large increase in R&D investment. We use a group of firms unaffected by the incentive prior to 2008 to show that the bunching patterns are driven by the tax incentive and are not a spurious feature of the data. 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 in order to qualify for the tax cut.

Second, we develop a rich model of firm behavior where R&D investment and relabeling decisions depend on tax incentives, the effect of R&D on productivity, the costs of relabeling, as well as on heterogeneity in firm productivity and adjustment costs. The model shows that, as long as firm productivity is smoothly distributed across the population, the InnoCom program leads to excess bunching at the R&D notch relative to a tax system without a notch. We derive a bunching estimator that relates the bunching patterns to the percentage increase in R&D following methods similar to those in Saez (2010) and Kleven and Waseem (2013). Our model also predicts increases in relabeling and productivity that depend on the returns to R&D. We then show that these predictions can be quantified empirically by linking our model to new methods developed by Diamond and Persson (2016).

In our third step, we measure the effects of the InnoCom program on reported R&D investment, relabeling, and productivity, as well as on other outcomes of policy interest, such as tax revenues. Using the bunching estimator, we quantify the percentage increase in R&D investment that is due to the tax notch. We find large increases in R&D investment of 31% for large firms, 21% for medium firms, and 11% for small firms in 2011. Our bunching estimates are supported by a number of robustness checks. First, we use a set of firms not affected by the program to show that the bunching pattern is not due to some spurious feature of the data but is indeed caused by the program. Second, we use these unaffected firms to inform the bunching estimate of the counterfactual density and we find very similar results.[1] Finally, these estimates are robust to excluding firms with extensive margin responses, state-owned enterprises, and to different specifications of the bunching estimator or the exclusion region, in addition to other robustness checks.[2]

We then estimate the effects of the InnoCom program on relabeling, productivity, and tax revenues. Even though a significant fraction of the response is consistent with relabeling, we find persistent and statistically significant effects of the InnoCom program on future productivity. In particular, between 2009 and 2011, the program led to an increase of 1.2% in productivity for the firms exposed to the fiscal incentive. We then calculate the elasticity of R&D investment to the change in the user cost that is induced by the InnoCom program, and we find an elasticity of 2 for reported R&D, and, once we account for relabeled administrative costs, an elasticity of 1.3 for real R&D investment.

Finally, we propose a simulated method of moments approach to estimate the structural parameters of our model, including costs of relabeling, the effect of real R&D on TFP, and the distributions of fixed and adjustment costs. Our estimates imply that, on average, 30% of the reported R&D investment is due to relabeling, and that a 100% increase in real R&D would increase TFP by 9.8%. We then use these estimates to simulate the effects of counterfactual policies that change the size of the tax cut and the location of the notch. We find that firm selection into the program plays a crucial role in determining the economic effects of the program. In particular, policies that lead to greater increases in aggregate R&D also tend to select firms with lower productivity, higher adjustment costs, and that have greater motives for relabeling. This lowers the effectiveness of the policy and increases the cost to the government of incentivizing real R&D. Finally, we compare the effects of the InnoCom program to those of a simulated linear tax credit. In a setting with high costs of relabeling, the linear tax

[1] As discussed by Blomquist and Newey (2017) and Bertanha et al. (2018), budget set variation is useful in identifying bunching estimators from notches. In practice, we obtain very similar estimates of the counterfactual density when using the unaffected firms in the estimation.

[2] Specifically, we obtain very similar results when we exclude SOEs, firms that had extensive margin responses during our sample period, low profitability firms, or low tech firms. We also obtain similar estimates of the counterfactual distribution when we use a set of firms that were not affected by the InnoCom program, when using different parametric choices for the density or the exclusion region, or when we estimate the counterfactual density using only data from the right tail of the distribution.

credit is more effective at stimulating R&D investment. However, when relabeling costs are low, an InnoCom-style program with a notch can incentivize real R&D at a lower fiscal cost than a linear tax credit.

This paper contributes to several literatures. First, 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. Hall and Van Reenen (2000) find a dollar-for-dollar effect of tax credits on R&D investment. 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. Examples include Agrawal et al. (2014), Dechezlepretre et al. (2016), Einiö (2014), Guceri and Liu (2015), Akcigit et al. (2018), and Rao (2015). To our knowledge, 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.[3]

Second, a previous literature has long documented "relabeling" as an important challenge to identifying the real impact of tax incentives for R&D (Eisner et al. (1984), Mansfield and Switzer (1985)). This is a salient issue for policymakers in developed countries (GAO, 2009) and is likely a more severe problem in developing economies (Bachas and Soto (2015), Best et al. (2015)). Our paper exploits unique data on firm expenditures 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 evasion, as in Best et al. (2015).

Third, although there has been a dramatic increase in innovation activities in China, researchers and policymakers are concerned that innovation resources could be misallocated in China. Wei et al. (2017) show that state-owned firms produce significantly fewer patents-per-yuan of investment than foreign or private domestic firms. In a closely related paper, König et al. (2018) compare the effects of R&D on productivity growth in Taiwan and mainland China, and find that R&D investments are significantly less effective in mainland China. They conjecture that misreported R&D in China may explain this discrepancy. Our paper validates this conjecture by using detailed micro-level data to examine an important policy that can lead firms to misreport R&D investment.

Finally, our paper is related to a recent literature that uses "bunching" methods to recover estimates

[3] As noted in the literature, optimal policies for R&D investment rely on estimates of the social returns to R&D investment (e.g., Bloom et al. (2013)), in addition to the firm-level effects of R&D. Our results characterize the costs to the government of increasing R&D through fiscal incentives, which can be used to evaluate policies given an estimate of spillovers from R&D.

of behavioral responses to taxation by analyzing the effects of sharp economic incentives, such as kinks or notches in tax schedules.[4] As detailed below, the R&D tax incentive creates a jump, or notch, in the after-tax profit function, generating similar incentives to those in Kleven and Waseem (2013) and Best et al. (2015). However, in contrast to this literature, the incentive generated by the notch targets a particular action: increasing R&D investment. We exploit this feature of our setting to estimate treatment effects of the program on R&D investment, relabeling, tax revenues, and growth in productivity using an estimator recently developed by Diamond and Persson (2016). Finally, we develop a simulated method of moments estimation approach that uses reduced-form estimates from the bunching estimators to recover structural parameters. We use the model to quantify the extent of misreporting, measure the returns to real R&D, and to simulate the effects of alternative policies.[5]

The rest of the paper is organized as follows. Section 2 provides a description of the fiscal incentive for R&D investment and discusses the potential for relabeling of R&D expenses in China. Section 3 discusses the data and provides descriptive evidence of the effects of the tax incentive on R&D investment and relabeling. Section 4 develops a model of R&D investment that links traditional estimates of productivity with bunching estimators. Section 5 describes our results on the real and relabeling responses to the InnoCom program. Section 6 culminates with the estimation of the structural parameters of the model and the simulation of counterfactual policies; Section 7 concludes.

2 Fiscal R&D Incentives and the Chinese Corporate Income Tax

China had a relatively stable Enterprise Income Tax (EIT) system from 2000-2007. During this period, the EIT ran on a dual-track scheme with a base tax rate of 33% for all domestic-owned enterprises (DOE) and a preferential rate for foreign-owned enterprises (FOE) ranging from 15% to 24%. The Chinese government implemented a major corporate tax reform in 2008 in order to eliminate the dual-track system based on domestic/foreign ownership and established a common rate of 25%.[6]

This paper analyzes the InnoCom program, which targets qualifying high tech enterprises (HTE) and awards them a flat 15% income tax rate. Since the average tax rate of the firm can fall from 33%

[4]These methods, pioneered by Saez (2010), have been used by researchers analyzing a wide range of behaviors. Kleven (2015) provides a recent survey. Our project is most related to a smaller literature analyzing firm-level responses (Devereux et al. (2014), Patel et al. (2016), Liu and Lockwood (2015), Almunia and Lopez-Rodriguez (2015), Bachas and Soto (2015)) as well as to papers analyzing the effect of constraints to optimizing behavior (Kleven and Waseem (2013), Best and Kleven (2015), Gelber et al. (2014)).

[5]The model allows us to clarify the interpretation of cross-sectional estimates by addressing issues discussed in Einav et al. (2015). Lockwood (2018) also notes that reduced-form effects from bunching in notches are not sufficient to analyze the effects of changes in policy. He shows changes to a notch generate first order effects on welfare, which are captured by our structural model. Similarly, Blomquist and Newey (2017) and Bertanha et al. (2018) note that cross-sectional estimators may not identify structural parameters without variation in non-linear incentives. As we discuss in Section 5, we use data from an unaffected set of firms to overcome this concern.

[6]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 and Technology	Ministries of Science and 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.

to 15%, the tax incentive of this program is economically very important and may lead firms to invest in projects with substantial fixed costs. This program is most important for DOEs, including both state- and privately-owned enterprises, as they are not eligible for many other tax breaks.

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

In addition to increasing R&D intensity, the InnoCom program requires that at least 30% of the firm's employees must have a college degree, and at least 10% of the firm's total employment should be devoted to R&D. Finally, in order for firms to qualify for the program, they have to actively apply and submit to a special audit.[7] The reform improved compliance with 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.[8]

[7] Our data does not detail whether firms comply with the non-R&D requirements of the program. To the extent that these additional requirements are binding, they will limit firms from responding by bunching at the R&D notch. Our model accounts for these additional certification requirements and application costs by assuming that firms differ by an unobserved fixed cost of certification (see Section 4.3).

[8] The original government regulations also require that firms operate in a number of selected state-encouraged industries. However, 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. In particular, 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 et al. (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 mis-categorization of administrative expenses as research expenses.

The hypothesis that the entirety of the response is due to evasion is likely ruled out by the requirements of the InnoCom certification in order to obtain the preferential tax rate.[9] A second unlikely form of evasion is the reporting of "phantom expenses" or the manipulation of 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.[10]

From conversations with the State Administration of Tax as well as corporate executives, we recognize that the most important source of manipulation is mis-categorization of expenses. Specifically, in the Chinese Accounting Standard, R&D is categorized under "Administrative Expenses," which includes various other expenses that are related to corporate governance.[11] This raises the possibility that firms relabel non-R&D administrative expenditures as R&D in order to over-report their R&D intensity. These type of expenses are easily shifted, and it may be hard to identify relabeling in any given audit. In particular, since the threshold of R&D depends on sales, it might be hard for firms to perfectly forecast their expenses. A firm with unexpectedly high sales, for instance, might choose to characterize administrative expenses as R&D in order to meet the InnoCom requirement in any given year.[12] Our empirical strategy to detect relabeling leverages these institutional features and exploits the detailed cost reporting in our administrative tax data, which contains information on the breakdown of operating expenses and R&D expenses.

[9]First, the certification process requires firms to maintain the required R&D intensity for a period of three years and firms often use specialized consulting firms to ensure that they satisfy the standards set by the Ministry of Science and Technology. Second, 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.

[10]As in other settings (e.g., Kleven et al. (2011)), it is hard for companies to report expenses that are not reported by third-party vendors. For these reasons, it is very hard for firms to completely make up "phantom" R&D expenses.

[11]Examples include administrative worker salary, 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 towards R&D, or with more precise accounting of previously-undercounted R&D expenses. In Section 6 we explore how this interpretation affects our estimates.

[12]In Section 3 we show sales are not manipulated around the R&D thresholds.

3 Descriptive Evidence of Firms' Responses to Tax Notches

We now describe our data, provide descriptive evidence suggesting that R&D investment by Chinese manufacturing firms is responsive to the InnoCom program, and we 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.

3.1 Data and Summary Statistics

Our main data come from the Chinese State Administration of Tax (SAT). The SAT is the counterpart to the IRS in China and is in charge of tax collection and auditing. Our data are comprised of administrative enterprise income tax records for years 2008–2011.[13] These panel data include information on firms' total production, sales, inputs, and R&D investment. In particular, the detailed cost breakdowns allow us to measure different subcategories of administrative expenses. We also use these data to construct residualized measures of firm productivity.[14] 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 relatively well-studied Chinese Annual Survey of Manufacturing (ASM), which extends our sample to 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. 8% of the sample reports positive R&D. Among firms with positive R&D, the ratio of R&D to sales ratio, 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.

In panel B, we report summary statistics of Chinese manufacturing firms with R&D activity in the ASM for years 2006–2007. We have a similar sample size of around $300,000$ firms each year. Firms in the ASM sample are noticeably larger than those in the SAT sample, and the difference is more pronounced when we look at the lower quartiles (i.e. 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 towards medium and large firms. The fraction of firms with positive R&D is slightly larger than 10%, and

[13] We discuss our data sources in detail in Appendix B.

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

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, which correspond to the three thresholds where the corporate income tax cut kicks-in. 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, the remaining panels of Figure 2 plot the histograms of R&D intensity for the three different size categories specified by the InnoCom program. 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 only find bunching at 4%, while for firms with more than 200 million RMB in annual sales, we only observe bunching at 3%. These patterns are consistent with the size-dependent tax incentive in the InnoCom program.[15]

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 becomes one of the most important tax breaks for FOEs. Their R&D intensity distribution shows a very distinguishable bunching at 3% after the reform, which is the exact threshold required for these firms to qualify as HTEs. The figure illustrates clearly that the change in the EIT system had a large impact on firm behavior.[16]

[15]In comparison, Figure A.1 plots the empirical distribution of R&D intensity in the ASM for years 2006–2007. The tax incentive of the InnoCom 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%.

[16]Similarly, Figure A.2 shows the effect on small firms (sales below 50 million RMB) DOEs who saw an increase in the R&D intensity threshold from 5% to 6%. While there is a stable bunching pattern at 5% for years 2006 and 2007, it almost completely disappears in 2008 and shifts to 6%.

3.3 Detecting Relabeling of R&D Investment

We now explore the degree to which the bunching response may be due to expense mis-reporting. Figure 4 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 suggests that some firms that report R&D intensity at the notch may not change their real R&D investment, but may instead mis-categorize non-R&D expenses to comply with the policy. Once the firms get 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 mis-categorize non-R&D expenses into R&D when they get close to the bunching thresholds.[17]

The structural breaks in Figure 4 are statistically significant for all three groups (see Table A.1). As we discuss in Section 5.2, however, these estimates do not have a causal interpretation. Nonetheless, they present strong descriptive evidence that firms may respond to the InnoCom program by relabeling non-R&D expenses.[18]

Lack of Sales Manipulation

The stark bunching patterns in these figures raise the concern that firms may also manipulate their sales. There are two ways 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 5 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-reported sales in order 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.[19]

[17] The existence of different thresholds across size groups also allows us to rule out other explanations for these discontinuities. In particular, we find that when we impose the "wrong" thresholds of the other size groups, there is no observable discontinuity. In Appendix D, we explore whether firms adjust other costs that are not in the administrative cost category, and we show firms do not respond to the program by manipulating other expenses.

[18] We also conduct a similar set of analysis focusing on the ratio of R&D to total administrative expenses. In this case, expense mis-categorization would result in discontinuous increases in this ratio at the notch. This is confirmed in Table A.3 and in Figure A.3.

[19] Table A.2 reports estimates of the structural breaks at these notches, which are statistically insignificant.

Second, if a firm wants to be categorized as a larger firm, it may over-report sales in order to qualify for a lower R&D intensity threshold. Panels B and C in Figure 5 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 are not responding to the incentives by manipulating their size.[20] Overall, it does not appear firms mis-report sales in order to comply with 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 mis-report sales as this is seen as a measure of their job performance.

Overall, Figures 2-4 provide strong qualitative evidence that firms actively respond to the incentives in the InnoCom program by increasing reported R&D investment and by relabeling administrative costs as R&D. Our quantitative analysis focuses on measuring the size of the change in R&D investment, analyzing the degree to which the response is due to relabeling, and studying how relabeling may influence the effect of R&D on productivity.

4 A Model of R&D Investment and Corporate Tax Notches

This section develops a model of R&D investment where firms may respond to notches in the corporate income tax schedule in China by investing in R&D and by relabeling non-R&D expenses. The objectives of the model are three-fold. First, the model shows that a standard model of firm investment and relabeling may produce the patterns described in Section 3. Second, the model motivates a bunching estimator for the increase in R&D investment, as in Saez (2010) and Kleven and Waseem (2013), as well as an estimator of treatment effects on relabeling and productivity, as in Diamond and Persson (2016). We present estimates of these effects in Section 5. Finally, the model relates the extent of bunching and the treatment effects on relabeling and productivity to structural parameters of the model, which we estimate in Section 6.

4.1 Model Setup

We start with a simple model and develop extensions to allow for relabeling, and for fixed costs of certification and adjustments costs of R&D investment. Full details of the model are presented in Appendix E.

Consider a firm i with a unit cost function $c(\phi_1, w_t) = c(w_t)\exp\{-\phi_{it}\}$, where w_t is the price of

[20] 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.

inputs.[21] ϕ_{it} is log-TFP and has the following law of motion:

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

where $D_{i,t-1} \geq 0$ 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 expenditures (captured by ε).

We assume the firm faces a demand function with a constant elasticity: $\theta > 1$. This implies that we can write expected profits as follows:

$$\mathbb{E}[\pi_{it}] \quad = \quad \mathbb{E}[\pi_{it}|D_{i,t-1} = 0]D_{i,t-1}^{(\theta-1)\varepsilon} = \tilde{\pi}_{it}D_{i,t-1}^{(\theta-1)\varepsilon},$$

where $\tilde{\pi}_{it} \equiv \mathbb{E}[\pi_{it}|D_{i,t-1} = 0] \propto \mathbb{E}[\exp\{(\theta-1)\phi_{it}\}|\phi_{i,t-1}]$ measures the expected profitability of the firm.

R&D Choice Under A Linear Tax

Consider first how firms' R&D investment decisions would respond to a linear income tax. We analyze the firm's inter-temporal problem as a two-period investment decision:[22]

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

The optimal choice of D_{i1}^* is given by:[23]

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

The choice of R&D depends on potentially-unobserved, firm-specific factors ϕ_{i1}, as they influence expected profits, $\tilde{\pi}_{i2}$. We can recover these factors by inverting the first order condition and writing $\tilde{\pi}_{i2}$ as a function of D_{i1}^*:

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

We now write the value of the firm, $\Pi(D_{i1}^*|t_2)$, as a fraction of firm sales, $\theta\pi_{i1}$, by substituting Equation 2 into the objective function:

$$\frac{\Pi(d_{i1}^*|t_2)}{\theta\pi_{i1}} = (1-t_1)\left[\frac{1}{\theta} + d_{i1}^*\left(\frac{1}{(\theta-1)\varepsilon} - 1\right)\right]. \tag{3}$$

[21] Note that any homothetic production function with Hicks-neutral technical change admits this representation.

[22] Firms commit to a medium-term set of R&D investments in order to participate in the InnoCom program (see Section 2). For this reason, we view the relevant margin for firms as a medium-term decision that we characterize in a two-period context.

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

Equation 3 expresses the firm's problem in terms of the choice of R&D intensity, $d_{i1}^* = \frac{D_{i1}^*}{\theta \pi_{i1}}$, as in the InnoCom program.[24]

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. This tax structure induces a notch in the profit function at $d_{i1} = \alpha$, where α is the R&D intensity required to obtain the high-tech certification. Figure 6 presents two possible scenarios following this incentive. Panel A shows the example of a firm that finds it optimal to choose a level of R&D intensity below the threshold. At this choice, the first order condition with a linear tax holds and the optimal value of the firm is given by Equation 3. From this panel, we can observe that a range of R&D intensity levels below the threshold are dominated by choosing an R&D intensity that matches the threshold level α. Panel B shows another firm that is indifferent between the internal solution and "bunching" at the notch. This firm is characterized both by Equation 2 and by having equal value from R&D intensities of d_{i1}^* and α.

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. Using Equation 2, we can write this equation as:

$$\frac{\Pi(\alpha|t_2^{HT})}{\theta \pi_{i1}} = (1-t_1)\left[\frac{1}{\theta} + \alpha\left(\left(\frac{d_{i1}^*}{\alpha}\right)^{1-(\theta-1)\varepsilon}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\frac{1}{(\theta-1)\varepsilon} - 1\right)\right].$$

Compared to Equation 3, this equation shows a larger R&D intensity (since $d_{i1}^* < \alpha$), which increases the cost of investment. The additional investment results in higher profits because of the productivity effect from the additional investment in R&D, $\left(\frac{d_{1i}^*}{\alpha}\right)^{-(\theta-1)\varepsilon} > 1$, and because of the tax benefit, $\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) > 1$.

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

$$\underbrace{\left(\frac{d_{i1}^*}{\alpha}\right)^{1-(\theta-1)\varepsilon}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\frac{1}{(\theta-1)\varepsilon} - 1}_{\text{Relative Profit from Bunching}} \geq \underbrace{\frac{d_{i1}^*}{\alpha}\left(\frac{1}{(\theta-1)\varepsilon} - 1\right)}_{\text{Relative Profit from Not Bunching}}. \tag{4}$$

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. For firms farther from the notch (as

[24] Firm value is given by $\Pi(D_{i1}^*|t_2) = (1-t_1)\left[\pi_{i1} + D_{i1}^*\left(\frac{1}{(\theta-1)\varepsilon} - 1\right)\right]$, where we substitute Equation 2 into the firm's objective function.

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 4 holds with equality, as in panel B of Figure 6. In this simple model, 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 4 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.

4.2 Real and Relabeled R&D Investment Under Tax Notch

This section extends the model by allowing for firms to misreport their costs and shift non-R&D costs to the R&D category. We show that, while the bunching predictions from the previous sections remain unaffected, the interpretation of the reported bunching response is now a combination of real and relabeled activity.

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 $h(D_{i1}, \tilde{D}_{i1})$, which represents the likelihood of being caught, and the punishment from the tax authority. We assume that the cost of mis-reporting 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:[25]

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

We also assume that $\tilde{h}$ satisfies $\tilde{h}(0) = 0$ and $\tilde{h}'(\cdot) \geq 0$. Finally, define $\Pi(D_{i1}, \tilde{D}_{i1}|t)$ as the value function of a firm's inter-temporal maximization problem when the firm invests D_{i1} on R&D, declares investment of $\tilde{D}_{i1}$, and faces tax t in period 2.

Firms qualify for the lower tax whenever $\tilde{D}_1 \geq \alpha\theta\pi_1$. 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 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 a lower level of R&D.

Consider now the optimal relabeling strategy of a firm conditional on bunching. The first order condition for relabeling implies the following condition:[26]

$$\underbrace{\left(\frac{d_{i1}^*}{\alpha(1-\delta_{i1}^*)}\right)^{1-(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)}_{\text{Productivity Loss from Relabeling}} = \underbrace{\frac{\left((1-t_1) - \tilde{h}'(\delta_{i1}^*)\right)}{\alpha(1-t_1)}}_{\text{Reduction in Investment Cost and Increase in Relabeling Cost}}, \tag{5}$$

[25]We assume that the mis-reporting cost depends on δ, the percentage of mis-reported R&D rather than the level of mis-reported R&D based on our specific institutional setting: the InnoCom program is based on R&D intensity rather than total R&D expenditure.

[26]We provide a detailed derivation in Appendix E.

where we use Equation 2 to express the first order condition in terms of the interior optimum R&D intensity, d^*_{i1}. When deciding how much to relabel, the firm trades-off lower productivity gains and increased costs of relabeling with the decrease in investment costs.

The firm decides to bunch if the profits from the optimal relabeling strategy are greater than when the firm is at the optimal interior solution, which occurs when:

$$\underbrace{\left(\frac{d^*_{i1}}{\alpha(1-\delta^*_{i1})}\right)^{1-(\theta-1)\varepsilon} \times \frac{(1-\delta^*_{i1})}{(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) - (1-\delta^*_{i1})}_{\text{Relative Profit from Bunching}} - \underbrace{\frac{\tilde{h}(\delta^*_{i1})}{\alpha(1-t_1)}}_{\text{Relabeling Cost}} \geq \underbrace{\frac{d^*_{i1}}{\alpha}\left(\frac{1}{(\theta-1)\varepsilon}-1\right)}_{\text{Relative Profit from Not Bunching}}. \tag{6}$$

Equations 4 and 6 are very similar and are identical in the case when $\delta^*_{i1} = 0$, such that there is no relabeling. When $\delta^*_{i1} > 0$, the cost of investment and the productivity gains are smaller, but the firm also incurs a cost of relabeling.

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. However, since the relative profit from not bunching has not changed, this implies that misreporting allows more firms to bunch than in the case without relabeling. Panel C of Figure 6 shows this intuition graphically. It depicts a firm that would not bunch absent the ability to relabel R&D. With relabeling, the firm reports an R&D intensity of α, while real R&D intensity is $(1-\delta^*_{i1})\alpha \geq d^*_{i1}$. Thus, when relabeling is possible, the marginal firm, such that Equation 6 holds with equality, will imply a lower threshold d^{*-}. This implies that we should see more bunching when firms are able to misreport R&D, and that the observed bunching patterns are a combination of real increases in R&D as well as increases in reported R&D that are due to relabeling of other expenses.

4.3 Adjustment Costs of Investment and Fixed Certification Cost

We now enrich the model to allow for firms to have random and heterogeneous adjustment and fixed costs, since they are a salient feature of the environment and help fit the data patterns described in Section 3.

First, as is common in studies of R&D investment, the distribution of R&D investment in China has large variability even conditional on firm TFP. In a world without the InnoCom program, our model would predict a deterministic relationship between R&D and TFP. In reality, firms face heterogeneous adjustment frictions of conducting R&D. We follow the investment literature and adopt a quadratic formulation for adjustment costs that is governed by: $b \times \frac{\theta\pi_{1i}}{2}\left[\frac{D_i}{\theta\pi_{1i}}\right]^2$. This term represents both fiscal costs of installing new equipment as well as limits to technological opportunity. Intuitively, the law of motion for TFP allows for strong returns to scale, as it implies that increasing R&D will have a proportional increase in the TFP of all units of production within a firm. Since the adjustment costs are proportional to firm size, they limit the returns to scale in R&D investment.

Second, while our model predicts that all firms with $d \in [d^{*-}, \alpha]$ would bunch at the notch, we find some firms do not obtain the InnoCom certification despite being very close to the notch. This is consistent with the guidelines of the program discussed in Section 2 that show that a greater-than-notch R&D intensity is not a sufficient condition for participating in the program. Indeed, firms with high R&D intensity may not participate in the program due to constraints that prevent them from hiring the sufficient number technical employees, if they do not obtain a significant fraction of their sales from high-tech products, or due to compliance and registration costs. We model these constraints by assuming that firms pay a fixed cost of certification: $c \times \alpha\theta\pi_{1i}$. We see this term as representing the cost to the firm of complying with the additional requirements of the program, such as hiring additional high-tech workers.

Appendix E shows that a firm's choice of whether to bunch is determined by a similar condition to Equation 6. In this case, however, the identity of the marginal firm depends on a given set of values b and c, which we denote $d^-_{b,c}$.[27] As expected, we find that $d^-_{b,c}$ is increasing (smaller response) with both adjustment, b, and fixed, c, costs. As before, for a given set of values b and c, $d^-_{b,c}$ is decreasing (larger response) in the profitability elasticity of R&D ($(\theta-1)\varepsilon$), and increasing in the relabeling cost. We now redefine $d^{*-} = \min\limits_{b,c} d^-_{b,c}$ as the smallest R&D intensity for which there is a marginal firm.

4.4 Empirical Implications for Bunching on R&D

We now describe how we use the model to quantify the distributional patterns described in Section 3. Figure 7 provides the intuition for this procedure. Panel A provides a counterfactual distribution of R&D intensity, d^*_{i1}, under a linear tax. Denote this counterfactual density by $h_0(d)$. Panel A demonstrates the effect of the notch on the distribution of R&D intensity in a world of unconstrained firms. In this case, there is a range of R&D intensity that is dominated by the threshold α, as shown by the density of R&D intensity with a notch, $h_1(d)$. Firms with an internal solution in this range will opt to bunch at the notch. Define the missing mass in the range $[d^{*-}, \alpha]$, relative to the counterfactual distribution, as B.

The prediction in panel A of Figure 7 is quite stark in that no firms are expected to locate in the dominated interval. The presence of fixed and adjustment costs may constrain firms from responding to the incentives in the InnoCom program. For given values of (b, c), a firm will be constrained from

[27] Another potential modeling choice is to allow for heterogeneous returns to R&D ϵ across the firms. We discuss about this possibility in Appendix J.

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 d in the range (d^{*-}, α) is given by

$$\mathbb{P}r(\text{Constrained}|d) = \int\limits_{b,c} \mathbb{I}[d < d^-_{b,c}] h_0(d,b,c) d(b,c) = h_1(d),$$

where $h_0(d,b,c)$ is the joint density of R&D intensity, and fixed and adjustment costs, and where the second equality notes that we observe this fraction of firms in the data.

Panel B of Figure 7 describes graphically how allowing for random adjustment and fixed costs affects the predicted bunching pattern. In particular, the area B can now be computed as follows:

$$\begin{aligned} B &= \int\limits_{d^{*-}}^{\alpha} \int\limits_{b,c} \mathbb{I}[d \geq d^-_{b,c}] h_0(d,b,c) d(b,c) dd = \int\limits_{d^{*-}}^{\alpha} \int\limits_{b,c} (1 - \mathbb{I}[d < d^-_{b,c}]) h_0(d,b,c) d(b,c) dd \\ &= \int\limits_{d^{*-}}^{\alpha} (h_0(d) - \mathbb{P}r(\text{Constrained}|d)) dd = \int\limits_{d^{*-}}^{\alpha} (h_0(d) - h_1(d)) dd. \end{aligned} \quad (7)$$

This equation shows that the extent of bunching measured by the area B is determined by the threshold d^{*-}, and by the joint distribution of counterfactual R&D intensity. Our model predicts a larger area B when firms have larger valuations for R&D, $(\theta - 1)\varepsilon$, lower relabeling costs, or smaller fixed and adjustment costs.

As in Kleven and Waseem (2013), we can relate the bunching patterns to the behavior of the marginal firm. Defining $\Delta D^* = \frac{\alpha - d^{*-}}{\alpha}$ as the percentage increase in R&D intensity relative to the notch, we have:[28]

$$\Delta D^* \approx \frac{B}{\alpha h_0(\alpha)(1 - \mathbb{P}r(\text{Constrained}))}. \quad (8)$$

4.5 Model Implications for Relabeling and Productivity

In addition to the bunching predictions, our model predicts that firms that bunch may engage in relabeling, and that their future TFP will increase to the extent that the reported R&D investment constitutes real activity. However, since firms select into the program by manipulating R&D, comparing firms that participate in the program to those that do not will result in a biased estimate of the effects of the program. Instead, we compare the observed outcomes of the firms that could have participated in the program to a counterfactual value of the same outcomes. For this purpose, we define the manipulated region (d^{*-}, d^{*+}) to include all firms that could have responded to the program.[29] This is similar to the previous section that compares the observed density of R&D under InnoCom program with a counterfactual density without the program.

[28] Appendix F contains details of this approximation.

[29] In practice, firms bunch in a neighborhood above α, as can be seen in Figure 2.

Diamond and Persson (2016) develop an estimator that formalizes this comparison by quantifying the average effect of the program on a given outcome Y:[30]

$$ITT^Y = \mathbb{E}[Y|\text{Notch}, d \in (d^{*-}, d^{*+})] - \mathbb{E}[Y|\text{No Notch}, d \in (d^{*-}, d^{*+})]. \tag{9}$$

Note that $\mathbb{E}[Y|\text{Notch}, d \in (d^{*-}, d^{*+})]$ is directly observed in the data. We discuss the econometric approach to estimating $\mathbb{E}[Y|\text{No Notch}, d \in (d^{*-}, d^{*+})]$ in Section 5.2. Equation 9 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).[31]

Our model has intuitive predictions for ITT^Y's on R&D, relabeling, and TFP. To see this, note that $ITT^Y \approx B(\bar{Y} - \underline{Y})$, where $\underline{Y}$ is the counterfactual average value of Y for compliers with $d_0 \in (d^{*-}, \alpha)$ and $\bar{Y}$ is the average value of Y for compliers with $d_1 \in (\alpha, d^{*+})$.[32] This expression simply states that the ITT is approximately equal to the average treatment effect among compliers multiplied by the excess mass from Equation 7. If some of the reported R&D intensity is real activity, our model would predict $ITT^{TFP} > 0$. 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, $\overline{ADM}$ 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 of R&D intensity and $ITT^d \approx B\frac{(d^{*+} - d^{*-})}{2}$. Our model would predict a larger fraction of compliers B if ε is large or if relabeling costs are low. Section 6 discusses how we link estimated treatment effects to structural parameters.

5 Effects on Investment, Relabeling, and Productivity

This section presents estimates of the causal effects of the InnoCom program on investment, relabeling, and productivity. Section 5.1 estimates the investment response from the bunching estimator. Section 5.2 presents estimates of treatment effects on relabeling, productivity, and tax revenues.

[30] Bachas and Soto (2015) implement a similar approach to analyze the effect of notches on other outcomes.

[31] Conceptually, we can partition the firms in the region (d^{*-}, d^{*+}) into compliers, never-takers, and always-takers. In our setting, never-taker firms are firms below the notch that are constrained from responding to the policy. Always-taker firms are firms that are already above the notch without the program. By assuming that there are no defier firms, we can show that Equation 9 has the interpretation of an intent-to-treat. See Appendix F for a detailed discussion.

[32] See Appendix F for details of this approximation.

5.1 Bunching Estimates of Investment Response

We follow the literature (see, e.g., Kleven (2015)) by estimating the counterfactual density of R&D intensity, $h_0(\cdot)$, using a flexible polynomial that ignores the effects of data around the threshold. Mechanically, we first group the data into bins of R&D intensity and then estimate the following regression:

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

where c_j is the count of firms in the bin corresponding to R&D intensity $d_j = \frac{D_{j1}}{\theta\pi_1}$, and where (d^{*-}, d^{*+}) is the region excluded in the estimation. An estimate for $h_0(d)$ is now given by $\hat{c}_j = \sum\limits_{k=0}^{p} \hat{\beta}_k \cdot (d)^k$. Similarly, we obtain counterfactual estimates for $h_0(\alpha)$ and B as follows:

$$\widehat{h_0(\alpha)} = \sum_{k=0}^{p} \hat{\beta}_k \cdot (\alpha)^k \quad \text{and} \quad \hat{B} = \sum_{d_j = d^{*-}}^{\alpha} \left(\sum_{k=0}^{p} \hat{\beta}_k \cdot (d_j)^k - c_j \right).$$

Intuitively, a larger missing mass B is related to a larger increase in R&D intensity. Indeed, these quantities provide a first cut of the effect of the program on R&D investment since:[33]

$$\Delta d \equiv \frac{\mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})] - \mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]}{\mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]} \approx \frac{B}{2\alpha h_0(\alpha)}. \tag{10}$$

Finally, we estimate the fraction of constrained firms relative to the counterfactual density at the R&D intensity α^- such that firms would be willing to jump to the notch even if R&D had no effects on productivity:

$$a^*(\alpha^-) = \frac{\mathbb{P}r(\widehat{\text{Constrained}}|\alpha^-)}{\widehat{h_0(\alpha^-)}} = \frac{c_{\alpha^-}}{\sum\limits_{k=0}^{p} \hat{\beta}_k \cdot (\alpha^-)^k},$$

which allows us to implement Equation 8 for the percentage increase of the marginal buncher, ΔD^*.[34]

Implementing the bunching estimator requires choosing the degree of the polynomial, p, and selecting the excluded region, (d^{*-}, d^{*+}). We use a data-based approach to selecting these parameters by cross-validating the choice of these values such that the missing mass below the notch equals the excess match above the notch.[35] Finally, we obtain standard errors by bootstrapping the residuals from the original regression, generating 5000 replicates of the data, and re-estimating the parameters.

[33] We provide a detailed derivation for Δd in Appendix F.

[34] 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 $\alpha^- = \alpha - (t^{HT} - t^{LT}) \times \frac{\text{Profits}}{\text{Sales}}$. Using the average net profitability ratio in our data of 7%, this 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 the large firms we have $(\alpha^-, \alpha) = (2.3\%, 3\%)$.

[35] This procedure ensures that we do not overfit the data with an overly flexible polynomial, and provides an objective approach to selecting 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. We describe this procedure in detail in Appendix G.

Figures 8-9 display the results of the bunching estimator for the three different notches for 2009 and 2011. The red line with diamond markers displays the observed distribution of R&D intensity $h_1(\cdot)$, the vertical dashed lines display the data-driven choices of the omitted region, and the blue line displays the estimated counterfactual density $h_0(\cdot)$. These graphs also report the fraction of firms that are constrained below the notch point, $a^*(\alpha^-)$, the overall percentage increase in R&D intensity in the excluded region, Δd, the increase in R&D intensity for the marginal firm in Equation 8, ΔD^*, as well as the p-value of the test that the missing mass equals the excess mass.

Panel A of Figure 8 shows a percentage increase in R&D over the excluded region of $\Delta d = 5.6\%$ for small firms in 2009. This small increase is due to the fact that many firms are not able to respond to the program, $a^*(\alpha^-) = 74\%$. As these are small firms, many firms may be constrained in their ability to increase investment to a significant degree, to develop a new product, or to increase the fraction of their workforce with college degrees. In addition, a higher failure rate among small firms implies that a long process of certification may never pay off in lower taxes. However, the marginal firm sees a significant increase, since $\Delta D^* = 38\%$. The specification test shows that using the missing mass or the excess mass results in statistically indistinguishable estimates.

Panels B and C show larger responses for medium and large firms in 2009. These counterfactual densities imply an increase in R&D intensity of 13.3% for medium firms, and of 14.9% for large firms. However, marginal firms see larger increases of 78.2% and 69.4% for medium and large firms, respectively. These graphs show that a significant fraction of firms are constrained from responding to the program (66% for medium and 57% for large firms). These patterns show that even large and medium firms may be unable to satisfy some of the requirements of the program. Using the missing mass and the excess mass results in statistically indistinguishable estimates of the increases in R&D for both types of firms.

Figure 9 shows similar qualitative patterns for 2011, although we find that the fraction of constrained firms is now smaller in all cases. We also find larger increases in R&D of 31% for large firms, 21% for medium firms, and 11% for small firms. These effects are estimated with a high degree of precision as standard errors are often an order of magnitude smaller than the estimates.

We now explore the robustness of our estimates. First, we show in panel A of Figure 10 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 distribution of R&D intensity of large foreign firms before 2008, which were not subject to the incentives of the InnoCom program, and finds a small and insignificant estimate of Δd. Second, we explore the potential for firms' extensive margin responses to bias our estimates. If the bunching we observe is driven by firms who previously did not perform any R&D,

the missing mass would not equal the excess mass. This would lead us to underestimate both the excess mass and Δd. In panel B of Figure 10 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. Finally, we show that our results are robust to using data from large foreign firms before 2008 who were not subject to the incentives of the InnoCom program in order to inform the shape of the density in the excluded region. Panel C of Figure 10 shows that using these data results in very similar estimates of both the counterfactual density and Δd.[36] Appendix H explores further robustness checks. First, as we show in Figure A.6, our results are not sensitive to excluding state-owned enterprises, low-tech firms, or low-profitability firms in the estimation. Additionally, as we show in Figure A.7, our results are not sensitive to the choice of (p, d^{*-}, d^{*+}) and we even obtain similar estimates when we only rely on data above d^{*+} to estimate the counterfactual density.

5.2 ITT Estimates on Productivity, Relabeling, and Tax Revenue

We now use an estimator of treatment effects developed by Diamond and Persson (2016) to estimate the effects of the InnoCom program on productivity, relabeling, and on fiscal costs. The intuition of the estimator is to compare the observed aggregate mean outcome for firms in the excluded region to a suitable counterfactual. For a given outcome Y_{it}, where $t \geq t_1$, the estimate is:

$$\begin{aligned} \widehat{ITT^{Y_t}} &= \mathbb{E}[Y_t|\text{Notch}, d_{t_1} \in (d^{*-}_{t_1}, d^{*+}_{t_1})] - \widehat{\mathbb{E}[Y_t|\text{No Notch}, d_{t_1}} \in (d^{*-}_{t_1}, d^{*+}_{t_1})] \\ &= \frac{1}{N^{Excluded}} \sum_{d_{i,t_1} \in \left(d^{*-}_{t_1}, d^{*+}_{t_1}\right)} Y_{it} - \int_{d^{*-}_{t_1}}^{d^{*+}_{t_1}} \hat{h}_0(d_{t_1}) E[Y_{it}|\widehat{d_{t_1}, \text{No}}\ \text{Notch}] dd_{t_1}, \end{aligned} \tag{11}$$

where the excluded region corresponds to year t_1, but where the outcome year t may correspond to t_1 or to a later period. As we discuss in Section 4.5, we interpret this estimate as an intent-to-treat (ITT).[37] For example, the ITT on $Y = \ln d$ measures the percentage increase in R&D intensity over the excluded region, Δd, without imposing the approximation of Equation 10. Finally, we obtain estimates of the elasticity of R&D investment to the user cost of capital (UCC) by taking the ratio of the ITT on R&D to the ITT on the UCC.

[36] As discussed in Blomquist and Newey (2017), variation in non-linear incentives can help in identifying responses when using bunching approaches. We combine this un-manipulated density with the density in 2011, $h_1(d)$, by ensuring that the combined density is continuous at the boundaries of the excluded region, d^{*-} and d^{*+}.

[37] 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.

The first quantity in Equation 11 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 density of R&D intensity, $\hat{h}_0(\cdot)$, estimated as part of the bunching analysis, with an estimated average value of the outcome conditional on a given value of R&D. We estimate $E[Y_{it}|\widehat{d_{t_1}, \text{No}} \text{ Notch}]$ as a flexible polynomial regression of Y_{it} on R&D intensity over the same excluded region used to estimate $\hat{h}_0(\cdot)$:[38]

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

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 $h_0(d)$.

Panel A of Table 3 presents estimates of ITT effects of the InnoCom program on several outcomes. We focus on large firms since they account for more than 90% of all R&D investment (see Figure A.5) and we study how the decision to invest in R&D in 2009 affects productivity and tax payments in 2011. 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 14.9%. We also find a decrease in the administrative cost ratio of 9.6%. When compared with the average value of this ratio, we find that administrative costs decreased by 0.33% of firm sales. We use this estimate to construct an approximation to the fraction of R&D investment that was relabeled. Compared to the implied increase in R&D intensity, this would imply that $\left(\frac{0.33\%}{0.89\%} \approx\right)$ 37% of the increase in R&D intensity was due to relabeling.[39] Note that this approximation is imperfect because it assumes that all firms engage in the same relabeling activity. As our model in Section 4.2 shows, the fraction of relabeling may vary across firms that are closer or farther away from the notch. The structural model in Section 6 relaxes this strong assumption. Nonetheless, this estimate would imply that the real increase in R&D investment was closer to 9%. The last 2009 outcome that we analyze is the effect of the policy on the user cost of R&D, where we find a decrease of 7.1%.[40]

[38] 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 11 by bootstrapping this equation as well as the estimates of the counterfactual density.

[39] We can approximate the increase in R&D intensity with $\alpha(1 - a^*(\alpha^-))\Delta D^* \approx 0.89\%$ for large firms in 2009.

[40] To compute the user cost of R&D, we first generate an equivalent-sized tax credit by dividing the tax savings form the policy by the R&D investment, and then use the standard Hall and Jorgenson (1967) formula as derived by Wilson (2009).

Panel A of Table 3 also reports the effects of the policy on outcomes in 2011. We find that between 2009 and 2011, the policy led to an increase in TFP of 1.2%. Finally, we observe an overall decrease in corporate tax revenues of 12.8%.[41]

The second panel of Table 3 presents estimates of user cost of capital elasticities 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) and 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 re-labeling behavior of the Chinese manufacturing firms, their user cost elasticity is comparable to those in more developed economies. Finally, as an alternative metric, we consider how much it costs the government to increase R&D investment in terms of foregone revenue. For every 1% increase in R&D, we find that 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. The structural model in the next section bridges this gap by computing the fiscal cost of raising total real R&D, and by showing how the fiscal cost depends on the design of the InnoCom program.

6 Structural Estimation and Simulation of Counterfactual Policies

The empirical estimates from the previous sections are crucial to evaluating the effects of the current program on reported R&D investment, suspected relabeling activities, and firm productivity. However, the previous analysis does not allow us to quantify the effect of real R&D on firms' productivity, or how the fraction of reported R&D investment that is due to relabeling varies across firms. Similarly, these estimates cannot be used to evaluate the effects of alternative policies, since changes to the policy will affect firm selection into the program, as well as investment and relabeling activities.

[41] We explore robustness of these estimates in Table A.5, where we show that the ITT estimates are robust to using an alternative, second-best parametrization of the counterfactual density of R&D intensity.

6.1 Structural Estimation

We now propose a method of simulated moments (MSM) framework to estimate the structural parameters of the model in Section 4 by matching the estimates from Section 5 to simulated counterparts.

We first discuss how we parametrize the model. We begin by calibrating θ, which we set at $\theta = 5$ based on the survey by Head and Mayer (2014).[42] 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 of variance σ^2. Given a value of θ, the persistence and volatility of log sales 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 also implies a stationary normal distribution of the underlying productivity ϕ_1 which we use as model fundamentals.[43]

We now parametrize the distributions of b and c, which we assume are distributed *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, $c \sim \mathcal{EXP}(\mu_c)$.[44] We adopt the following functional form for the costs of relabeling: $\frac{\exp\{\eta\delta\}-1}{\eta}$, where δ is the fraction due to relabeling. Note that this function may be linear, convex, or concave depending on the value of η (see, e.g., Notowidigdo (2013)). We use the method of simulated moments to estimate the set of parameters $\Omega = \{\varepsilon, \eta, \mu_b, \sigma_b, \mu_c\}$, where ε is the productivity effect of R&D.

To implement the MSM estimator, we form the criterion function:

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

where W is a bootstrapped weighting matrix. $h^B(\Omega)$ and $h^{ITT}(\Omega)$ are moment conditions based on our bunching and ITT estimators, respectively. $h^B(\Omega)$ is based on our estimates of d^{*-}, d^{*+}, and the distribution of R&D intensity based on these cutoffs. That is, we choose our model parameters so that our simulated data can rationalize the bunching patterns estimated in Section 5.1. In addition to this unconditional empirical density, we also require that the model match the joint distribution of firms' measured TFP and R&D intensity. As we discuss below, these moments play an important role in identifying key model parameters.

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

[43] Appendix J investigates the parametric assumption that total factor productivity $exp(\phi_1)$ follows a log normal distribution. We find that the distribution of measured TFP closely matches that of a log normal distribution, which implies that imposing this assumption is consistent with our data.

[44] In Appendix J we discuss estimates from an alternative model that allows for heterogeneous ε's and a constant b. While this model result in similar average values of ε and b, the model does not match the data as well as our benchmark model since it cannot match the joint distribution of TFP and R&D intensity.

We use the treatment effects on reported R&D, administrative expense ratio, and TFP from Section 5.2 to form the last set of moments, $h^{ITT}(\Omega)$. Let $\omega = \{\phi_1, b, c\}$ denote a firm with random draws of its fundamentals of productivity, adjustment cost, and fixed cost. We construct moments that match the empirical and simulated counterparts of the ITT estimates:[45]

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

where $\widehat{ITT^Y}$ is an estimate from Section 5.2.

Identification

While each of the simulated moments depends on multiple parameters, we give a heuristic description of the data patterns that identify each parameter. We start with the most central parameter: the returns to R&D, ε. One interesting observation is that, while the bunching patterns certainly inform this parameter, the bunching patterns alone are not able to separately identify ε and the unobserved heterogeneous adjustment and fixed costs. This is intuitive since both the benefit and cost of R&D enter the optimal choices of innovating firms. Two additional sets of moments help to separately identify these parameters. First, we rely on the model insight that firms' R&D decisions are not distorted below d^{*-} and above d^{*+}. Thus, the ranking of firms' measured productivity across these regions is determined by ε, and is not affected by the InnoCom program. For this reason, including the joint distribution of TFP and R&D intensity in $h^B(\Omega)$ helps to identify ε. Second, the ITT estimates on reported R&D and measured TFP also help to discipline ε. Note, however, that these estimates combine three distinctive forces: 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 too. For this reason, the ITT estimate on the administrative expense ratio is also crucial in order to pin down both η and ε.

Given ε and η, the identification of the distributions of adjustment and fixed costs is quite intuitive. First, the parameters of the distribution of adjustment costs, μ_b and σ_b, are identified by the counterfactual distribution of R&D intensity below d^{*-} and above d^{*+}. Next, the fraction of firms that bunch and the ITT on reported R&D inform the parameter of the distribution of fixed costs of certification: μ_c. Finally, the location of d^{*-} is jointly determined by all the parameters.

[45]Note that we restrict the support of firm fundamentals $\omega = \{\phi_1, b, c\}$ by requiring the counterfactual R&D to be in the excluded region.

Estimates of Structural Parameters

We estimate the model using a Laplace-type estimator that is based on Markov Chain Monte Carlo (MCMC), following Chernozhukov and Hong (2003). This procedure provides a numerically attractive way of obtaining point estimates and conducting inference. We construct the weighting matrix W based on the bootstrapped covariance matrix of our data moments.

Table 4 reports estimates of our structural parameters: $(\varepsilon, \eta, \mu_b, \sigma_b, \mu_c)$. Panel A reports the parameter estimates and the standard errors. All the estimates are statistically significant. Consider the estimate for ε. The estimate from panel A implies that doubling R&D increases measured TFP by 9.8%. Hall et al. (2010) surveys the extensive literature on R&D elasticity in similar production function setups. Our estimate lies within the broad range of previous result between 2% and 17%. Almost all of these previous studies use micro-data from developed countries, so it is interesting to see that the returns to R&D of Chinese firms are comparable in magnitude. The estimated relabeling cost parameter is 5.7, which indicates that, at the margin, the cost of relabeling is highly convex in terms of δ. In other words, 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. Note that firms benefit from relabeling by lowering investment and adjustment costs, which include technological opportunity constraints. Thus, firms that face a higher shadow cost of R&D (i.e. higher b) will be more willing to engage in relabeling. On average, we calculate that firms' realized relabeling cost is 4.7% of the implicit R&D savings. Finally, the estimated certification cost is quite modest: for the firms who decide to bunch and certify as high-tech firms, the fixed certification cost is on average 2% of their realized profit.

Panel B compares the simulated moments with the data moments and shows that our model does a very good job of matching the data. The model replicates the distribution of firm-level R&D intensity and the bunching pattern almost perfectly. It also captures the positive correlation between R&D intensity and measured productivity very well. The ITT estimates are the moments with the largest bootstrapped standard deviations. For this reason, they are matched less precisely based on our optimal weighting matrix. In particular, our model predicts a slightly smaller ITT on TFP.

Finally, we evaluate the sensitivity of our point estimates to each individual moment. We calculate the local derivative of our estimated parameters with respect to each moment using the methods of Andrews et al. (2017). The recovered sensitivity matrix is reasonable and conforms to the heuristic discussion above. We find that the joint distribution of TFP and R&D intensity are important determinants of ε. For instance, with a small change in the average TFP of firms above d^{*+}, ε would increase by around 10 percent from its estimated value. In contrast, we find that ε is not very sensitive to changes in the ITT of TFP. These methods also allow us to consider the potential that part of the reduction in administrative expenses is not due to relabeling.[46] If half of the decrease in administrative costs is not related to relabeling, our sensitivity analysis shows that ε would decrease by 0.002, which is a very modest amount. 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 reduced-from analysis for identification, is able to replicate these data patterns quite well, and provides a useful micro-foundation for simulating the effects of counterfactual policies.

Benchmark Model Implications

Given our model estimates, we can simulate our benchmark 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, 9.64% more productive than firms in the excluded region that do not comply with the policy. They also have idiosyncratic adjustment costs that are 34.5% 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 30.3% of the reported R&D investment is due to relabeling, on average. This fraction is dispersed across firms, with the 10th percentile firm relabeling 6.8%, and the 90th percentile relabeling 51.9%. 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 the punishment from relabeling in order to achieve the program threshold.

Third, we also find heterogeneous increases in real R&D for complying firms. Our models suggests that the distribution of real R&D investment is such that the 10th percentile firm sees an increase of 10.2%, the 90th percentile firm increases by 25.0%, and the median firm increases by 15.1%. This dispersion in investment then results in a dispersed distribution of gains in TFP.

6.2 Simulation of 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, and productivity growth. We first simulate alternative versions of the InnoCom program that vary the tax

[46] For instance, administrative costs may reduce if the tax incentive causes firms to pay closer attention to their accounting of R&D expenses, or if firms substitute inputs in response to the policy.

advantage and the location of the notch. We then compare our results with a counterfactual policy that follows a more standard investment tax credit.

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 may use them as a way to limit relabeling (Best et al., 2015). Second, given the explosive growth in R&D in China and that the government has chosen to use this policy, it is important to understanding the economic and fiscal consequences of this type of policy.

Figure 11 studies 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 between 10% to 22% for a given notch, relative to the current benchmark where $\alpha = 0.03$ and $t_2^{HT} = 15\%$.

Panels A and B 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 preferential tax rate), the program selects firms with lower productivity and higher adjustment costs. The selection effect is more pronounced based on adjustment costs than on 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 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 22% to 10% increases the fraction of reported R&D due to relabeling by almost 15 percentage points.

Panel E 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, complier firms invest more with a lower preferential tax rate. 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 increases 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 F according to the total increase in real R&D. This graph thus represents 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 2.8.[47] 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 a twice-as-large 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 will positively select more productive firms, and firms with better technological opportunities. Nonetheless, as shown in panel D, this policy would also be accompanied by more relabeling.

These simulations show that the effectiveness of these type of programs depends strongly on firm selection. As incentives for R&D increase, this may lead to misallocation of 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 that are likely socially undesirable.

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. In this section, we use our estimated model fundamentals to evaluate the possibility of drastically changing the Chinese InnoCom program into a R&D tax credit system comparable to that of the U.S. While the U.S. R&D tax credit system has numerous accounting details, we define it by the two most fundamental features: the base amount $\bar{D}_i$ and the tax credit rate τ. The U.S. government provides a credit of $\tau = 20\%$ to the tax payers' qualified R&D

[47]This ratio is greater than the value of 0.88 reported in Table 3 since this number accounts for relabeling of R&D, and, since the percentage increase in total real R&D is disproportionately determined by the high R&D intensity firms, which have smaller than average increase in R&D spending. When we calculate the same quantity as in Table 3, we obtain a comparable value of 0.92.

expenditure that exceed the base amount $\bar{D}_i$.[48]

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 is equals $(1 - t_1)\tau \approx 0.15$, which is significantly more effective than the 2.8 elasticity of the benchmark InnoCom program. If we impose the estimated cost of relabeling of $\eta = 5.663$, 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.

However, there are reasons to suspect that the tax enforcement will be more difficult under an R&D tax credit system since the tax authority will need to audit *all* the firms. This implies that individual firms will face lower costs of relabeling. With positive misreporting, the cost-effectiveness of the R&D credit quickly worsens. To see this, note that the R&D tax credit is calculated as

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

If firms relabel δ of reported R&D, then the effective tax cost of inducing per incremental 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. If we set $\delta^* = 0.3$, as in our benchmark case, then the tax-to-R&D elasticity is 7.97. In this case, the InnoCom program is a more cost-effective policy.

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 searched for the relabeling cost parameter that makes the policy maker indifferent between an R&D tax credit regime and our benchmark case. We find that when $\eta = 1.324$, which implies a fraction of relabeling of 9.9% (in contrast to 30.3% in our benchmark), the R&D tax credit policy achieves the same fiscal elasticity of 2.8. Therefore, a tax credit is a more cost-effective policy if the government can increase the cost of evasion such that $\eta > 1.324$. However, this may come at the cost of devoting additional government resources to detecting relabeling.

[48]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.

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. Despite the relabeling responses, we find significant effects on firm-level productivity that are consistent with sizable returns to R&D.

Optimal subsidies for R&D will depend on the fiscal cost for the government and whether R&D investment has external effects. This paper provides a useful metric that traces the government's tradeoff between own-firm productivity growth and tax revenues. If R&D is believed to have positive externalities on other firms' productivity, our estimates then provide a bound on the size of the externality that would justify government intervention.

Finally, while we find evidence consistent with relabeling, the unusual structure of the InnoCom program may limit the scope of relabeling and evasion through pre-registration and auditing. 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 may have large effects on the design of R&D subsidy policies, and future research should explore the potential for relabeling in other contexts.

References

Agrawal, Ajay, Carlos Rosell, and Timothy S Simcoe, "Do tax credits affect r&d expenditures by small firms? evidence from canada," Technical Report, National Bureau of Economic Research 2014.

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

Almunia, Miguel and David Lopez-Rodriguez, "Heterogeneous Responses to Effective Tax Enforcement: Evidence from Spanish Firms," Technical Report, Warwick University 2015.

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

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

Bachas, Pierre and Mauricio Soto, "Not(ch) Your Average Tax System: Corporate Taxation Under Weak Enforcement," Technical Report, UC Berkeley 2015.

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

Bertanha, Marinho, Andrew H. McCallum, and Nathan Seegert, "Better Bunching, Nicer Notching," *working paper*, 2018.

Best, Michael Carlos and Henrik Jacobsen Kleven, "Housing Market Responses to Transaction Taxes: Evidence From Notches and Stimulus in the UK," Technical Report, London School of Economics 2015.

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

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

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

Chernozhukov, Victor and Han Hong, "An MCMC approach to classical estimation," *Journal of Econometrics*, 2003, *115* (2), 293 – 346.

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

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

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

Einav, Liran, Amy Finkelstein, and Paul Schrimpf, "Bunching at the kink: implications for spending responses to health insurance contracts," Technical Report, Stanford University 2015.

Einiö, Elias, "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*, 2014, *96* (4), 710–728.

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

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

Gelber, Alexander M., Damon Jones, and Daniel W. Sacks, "Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test," Technical Report, UC Berkeley 2014.

Guceri, Irem and Li Liu, "Effectiveness of fiscal incentives for R&D: a quasi-experiment," Technical Report, Citeseer 2015.

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

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

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

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

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

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

Kleven, Henrik Jacobsen, "Bunching," Technical Report, London School of Economics 2015.

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

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

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

Liu, Li and Ben Lockwood, "VAT Notches," Technical Report, CESifo Working Paper Series No. 5371 2015.

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

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

Notowidigdo, Matthew, "The Incidence of Local Labor Demand Shocks," *working paper*, 2013.

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

Rao, Nirupama, "Do Tax Credits Stimulate R&D Spending? The Effect of the R&D Tax Credit in its First Decade," Technical Report, Wagner School, New York University 2015.

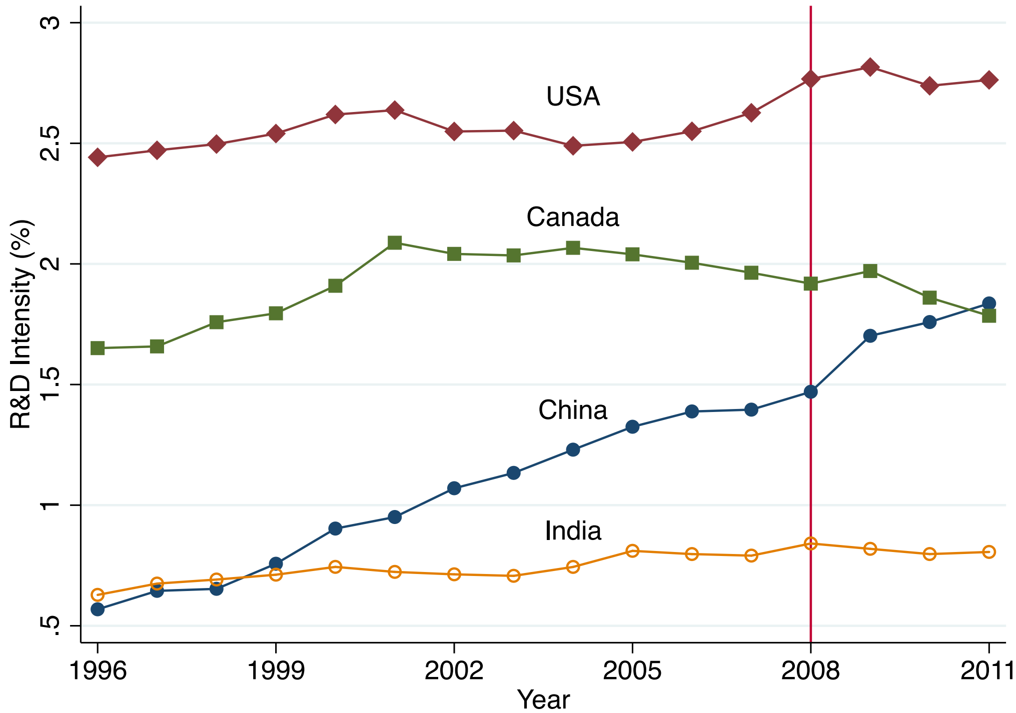
Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2010.

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

Wei, Shang-Jin, Zhuan Xie, and Xiaobo Zhang, "From "Made in China" to "Innovated in China": Necessity, Prospect, and Challenges," *Journal of Economic Perspectives*, 2017, pp. 49–70.

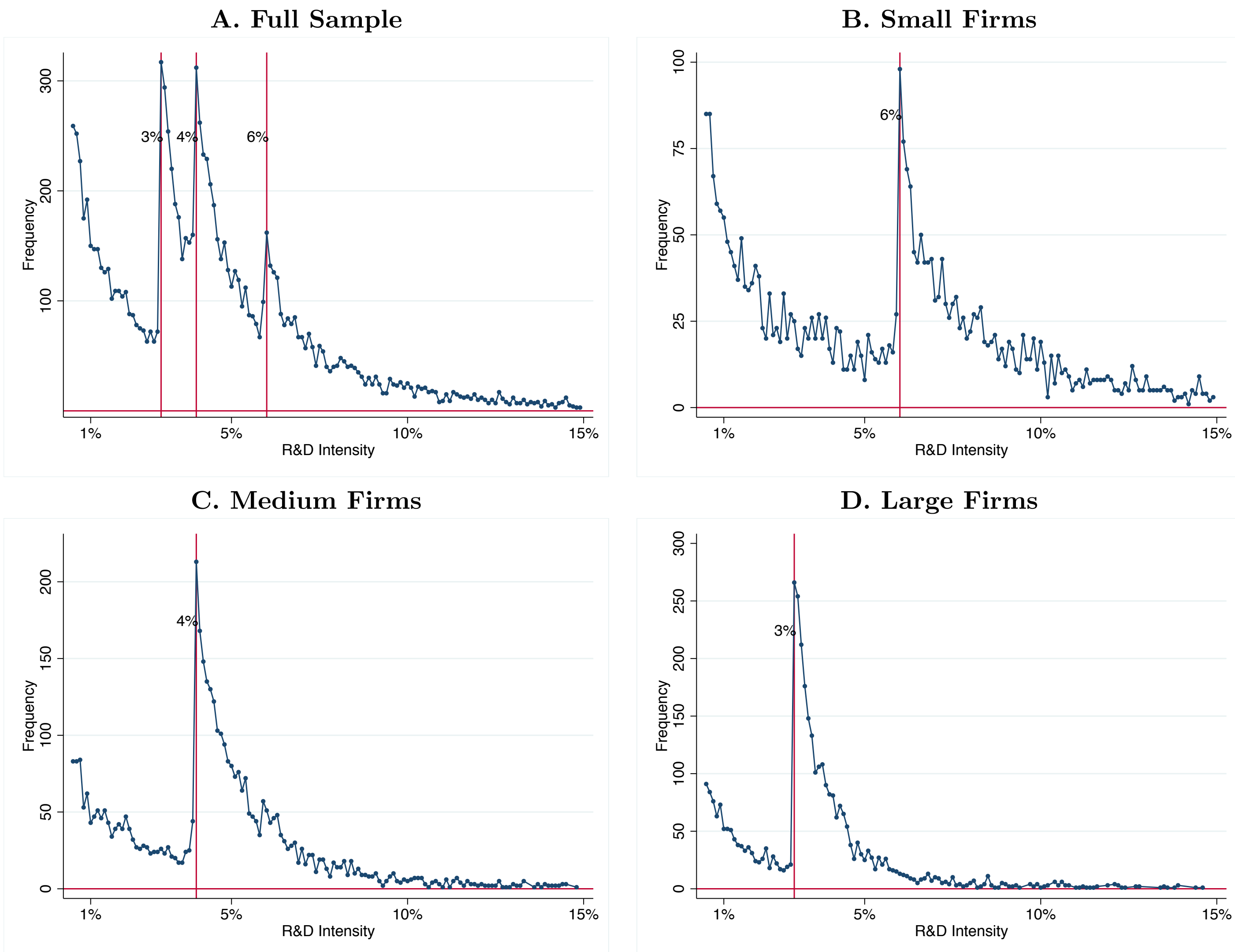
Wilson, Daniel, "Beggar Thy Neighbor? The In-State, Out-of-State, and Aggregate Effects of R&D Tax Credits," *The Review of Economics and Statistics*, 2009, *91* (2), 431–436.

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 share of GDP, in the private sector for China, Canada, India, and US. Chinese R&D intensity started at 0.5% in 1996, a similar level to India. It increased dramatically by more than three-fold to above 1.5% in 2011, on par with Canada. The R&D intensity of the U.S. remained stable at 2.5% during the same period. Chinese R&D intensity improved from 1/5 in 1996 to around 2/3 of the U.S level in 2011. The red line marks the year of the tax reform. Source: World Bank.

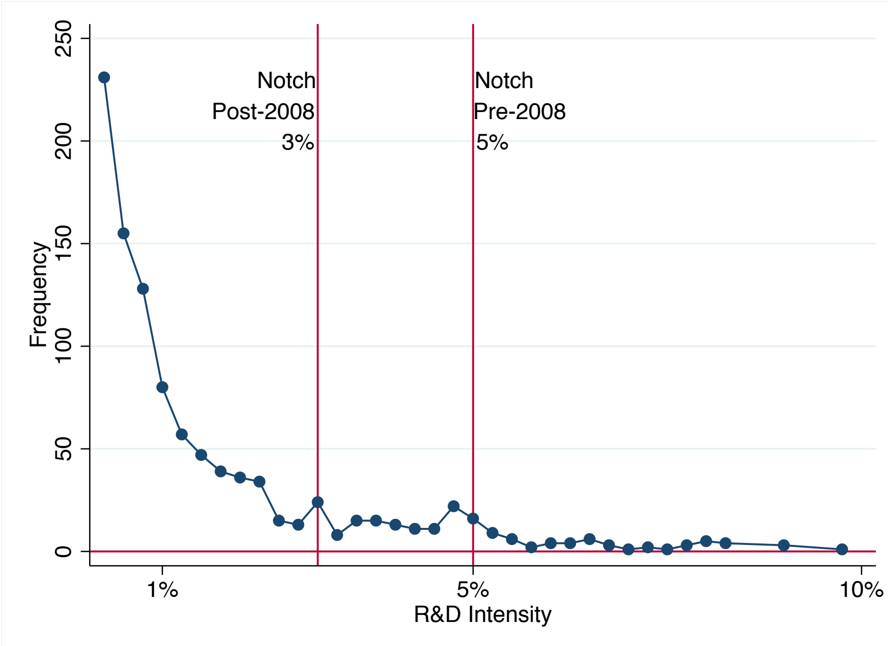
Figure 2: Bunching at Different Thresholds of R&D Intensity (2011)



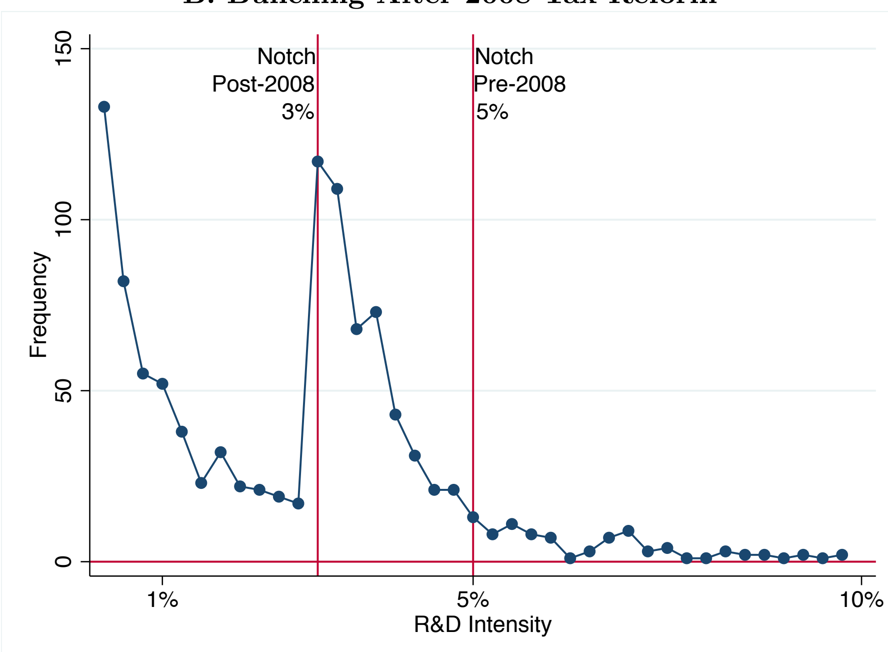
NOTES: This figure plots the empirical distribution of R&D intensity for all manufacturing firms that has R&D intensity between 0.5% and 15% in the Administrative Tax Return Database. Panel A reports the pooled data distribution with all sizes of firms. Panels B, C, and D report the R&D intensity distribution of the firms that have been classified as "Small", "Medium", and "Large" respectively. Note that large fractions of the firms "bunch" at the thresholds (6% for large, 4% for medium, and 3% for large) that qualify them 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

A. Bunching Before 2008 Tax Reform

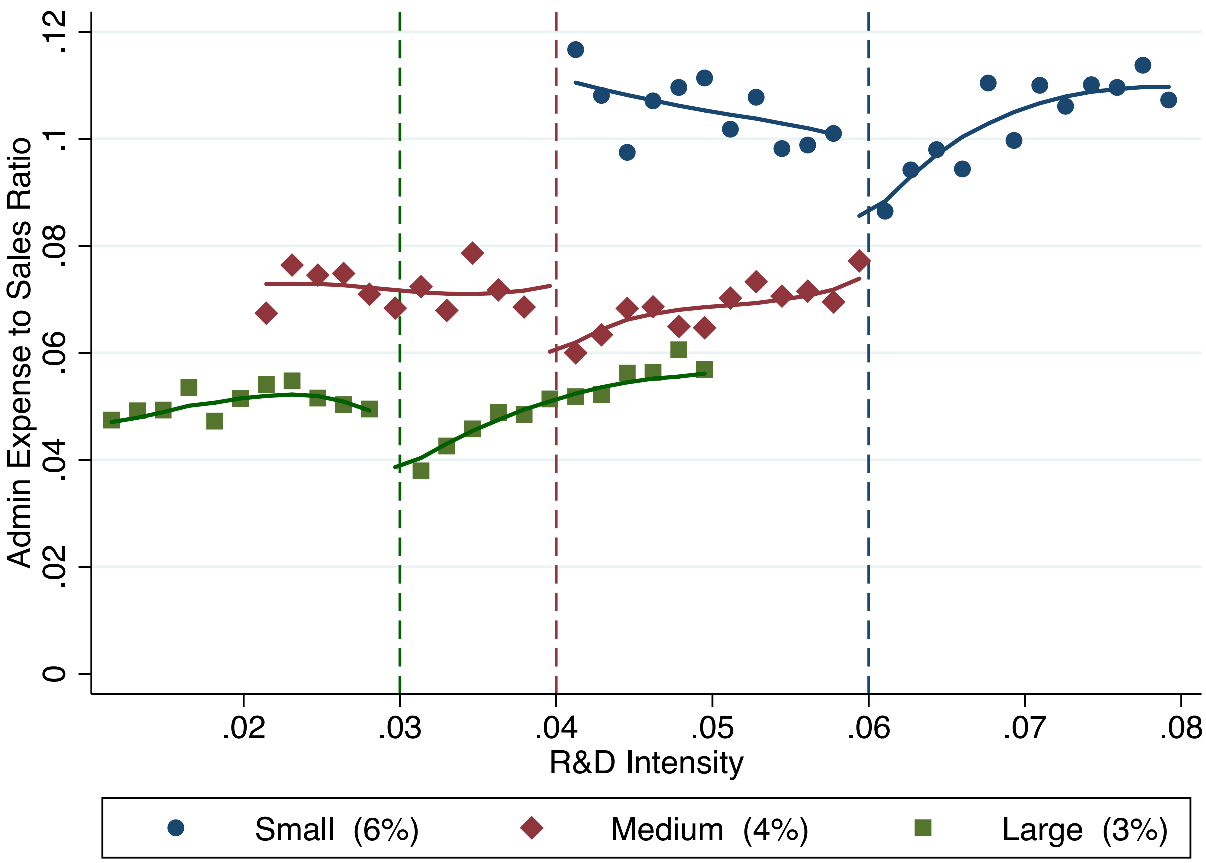


B. Bunching After 2008 Tax Reform



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 only plots 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 increase their bunching behavior substantially after 2008. The R&D intensity is concentrated around the 3% threshold. Source: Administrative Tax Return Database and Annual Survey of Manufacturers. See Section 3.1 for details.

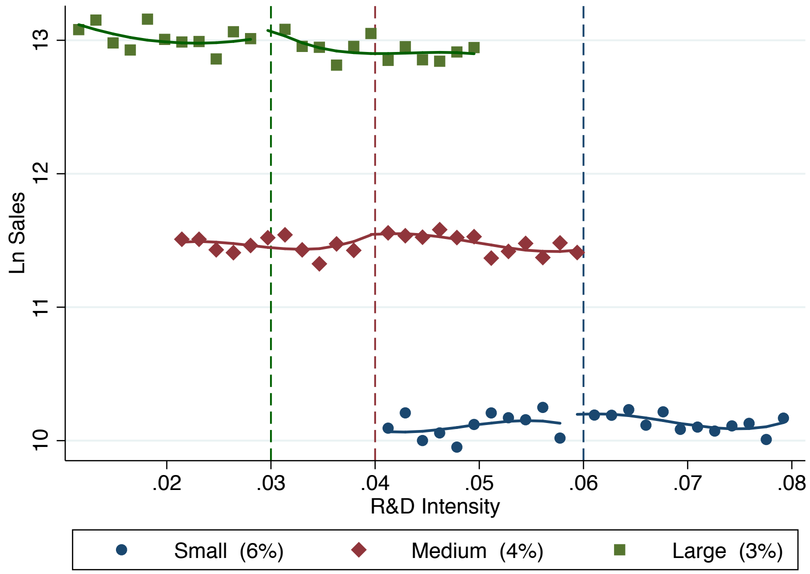
Figure 4: Empirical Evidence of Relabeling



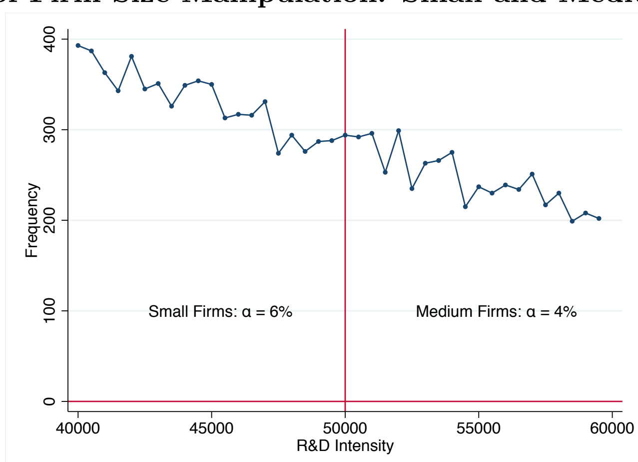
NOTES: This figure plots the non-R&D administrative expense to sales ratio at each level of R&D intensity. The green dots/line are for the large firms, the red dots/line are for the medium firms, and the blue dots/line are for the small firms. The threshold of R&D intensity for firms to qualify applying for InnoCom certification differs across firm size: 6% for small firms, 4% for medium firms, and 3% for large firms. For each size category, there is a pronounced drop of the administrative expense to sales ratio when the R&D intensity approaches the required threshold. Source: Administrative Tax Return Database. See Section 3.1 for details on data sources and Section 4 for details on the estimation.

Figure 5: Lack of Sales Manipulation

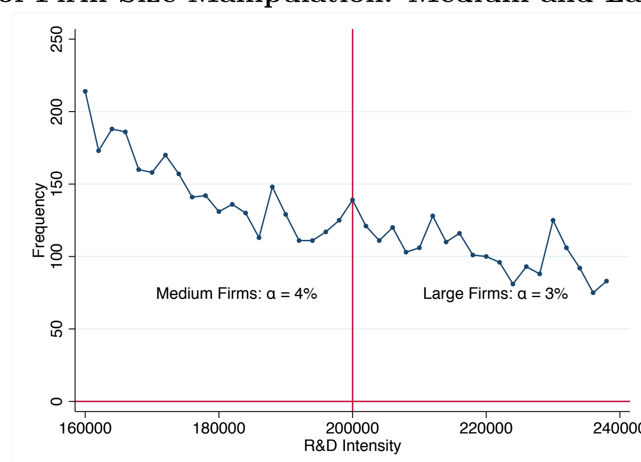
A. Lack of Sales Manipulation Around R&D Intensity Threshold



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

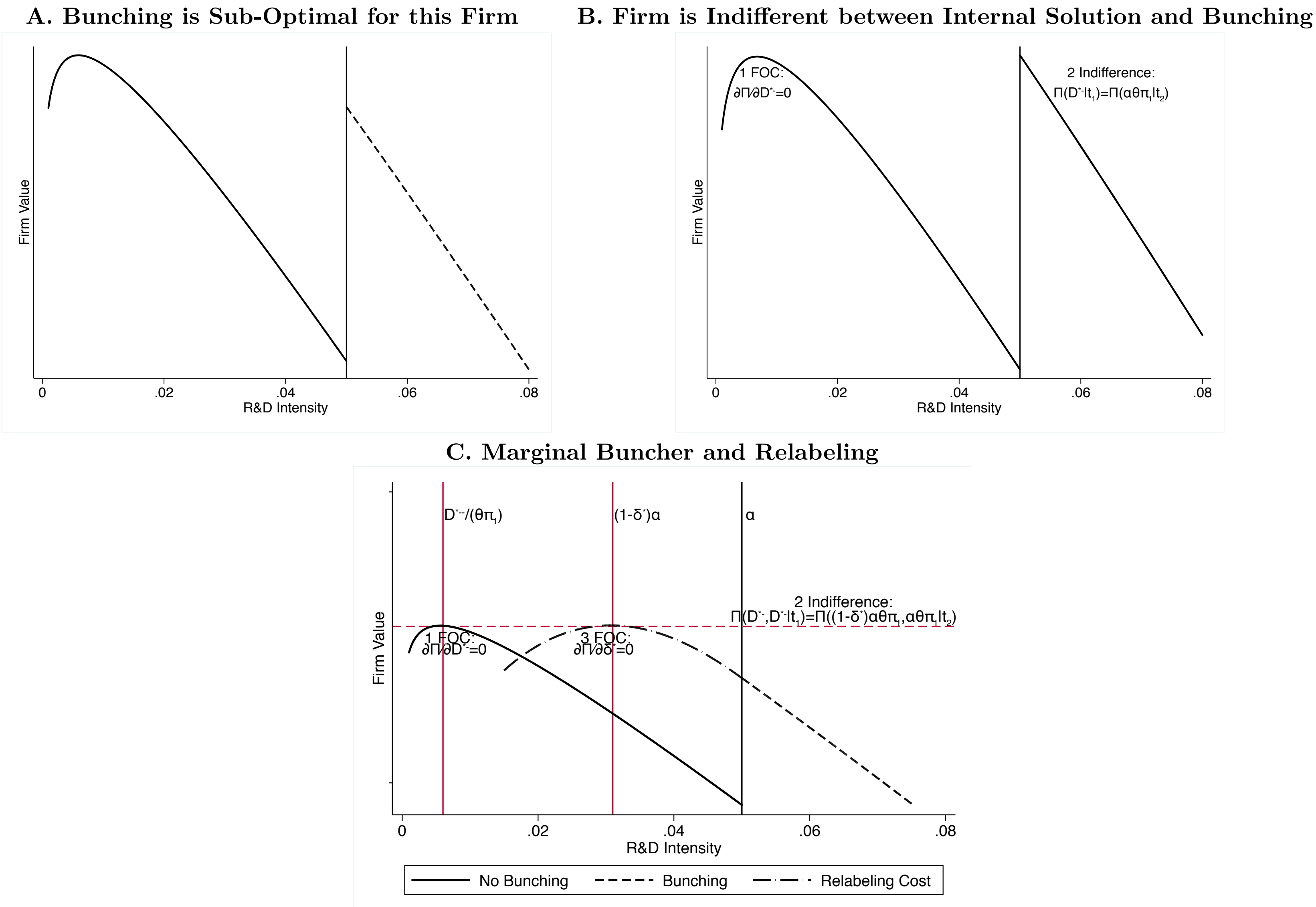


C. Lack of Firm Size Manipulation: Medium and Large Firms



NOTES: This figure examines the potential manipulation of sales data. Panel A shows firms do not manipulate sales by under-reporting their sales in order to reach their respective notch. Panels B and C show firms do not attempt to over-report their sales in order to move into the next size category and thus reduce the threshold of R&D intensity for qualifying the InnoCom program. Overall, there is little evidence for sales manipulation. Source: Administrative Tax Return Database and Annual Survey of Manufacturers. See Section 3.1 for details.

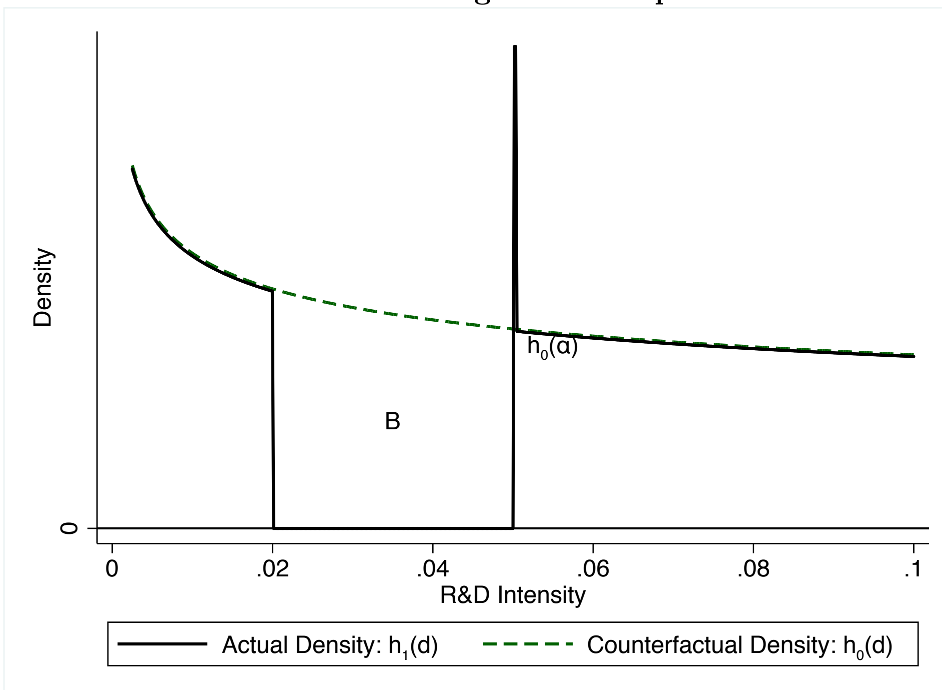
Figure 6: Induced Notch in Profit Functions



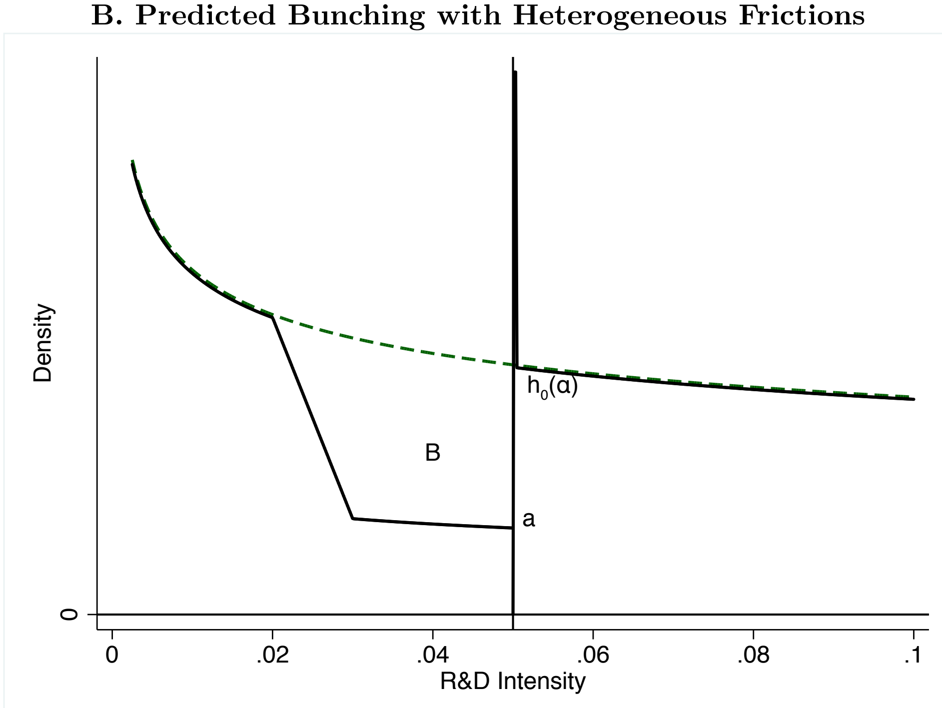
NOTES: This figure provides the intuition of when a firm decides to bunch or not. In panel A, we characterize a firm whose value of performing the interior optimal level of R&D is larger than bunching at the threshold. In panel B, we characterize another firm whose value of performing interior optimal level of R&D is exactly equal to bunching at the threshold. The fundamental determinant of this relationship is the unobserved firm heterogeneity in ϕ_1, which is reflected by the interior optimal R&D. Panel C shows that when firms can relabel non-R&D expenses as R&D expenses, the marginal firm which is indifferent between bunching or not will have a lower interior optimal R&D level level D^{*-}. See Section 4 for details.

Figure 7: Theoretical Predictions of Bunching

A. Predicted Bunching in the Simple Model

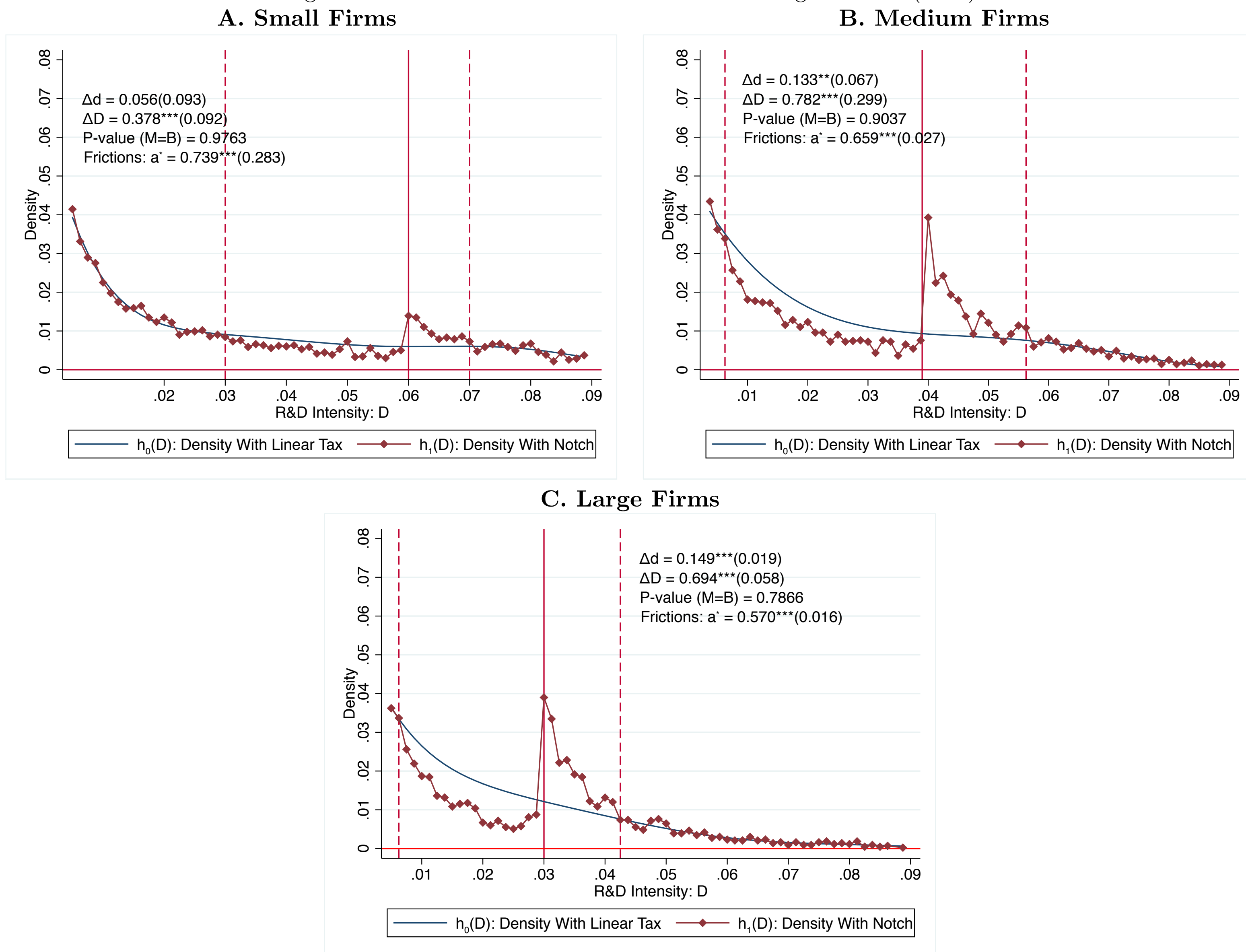


B. Predicted Bunching with Heterogeneous Frictions



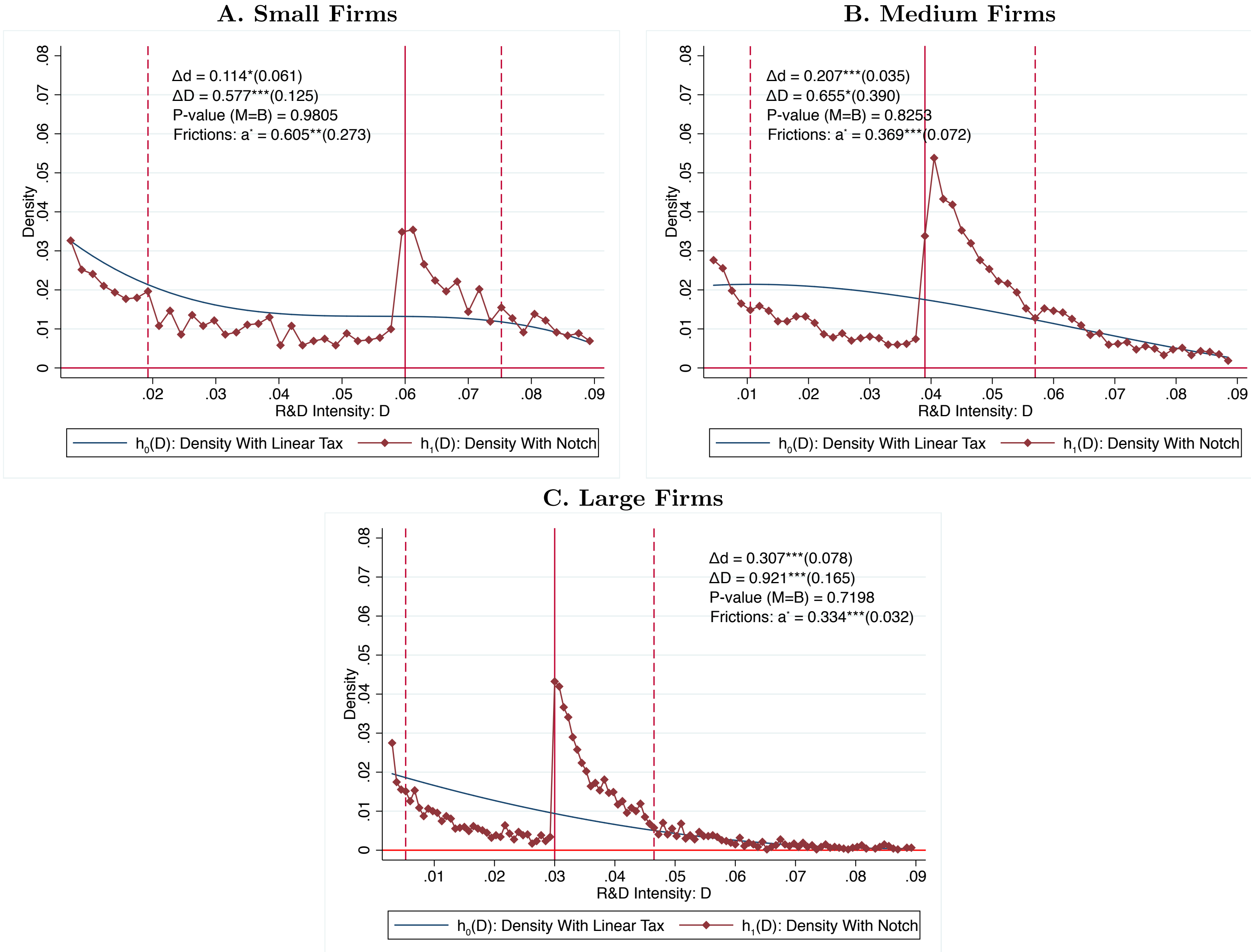
NOTES: This figure describes empirical implications of our model for R&D investment and bunching. Panel A plots the implied empirical R&D density distribution in our baseline model of R&D investment with productivity as the only source of heterogeneity at the firm-level. The model predicts that all the firms between the marginal firm and the notch will bunch, creating a dominated interval in the density. Panel B plots an enriched model where firms' R&D decision is subject to heterogeneous adjustment costs and a fixed cost of certification. These heterogeneities create frictions such that not all the firms in the dominated interval bunch on the notch. See Section 4 for details.

Figure 8: Estimates of Excess Mass from Bunching at Notch (2009)



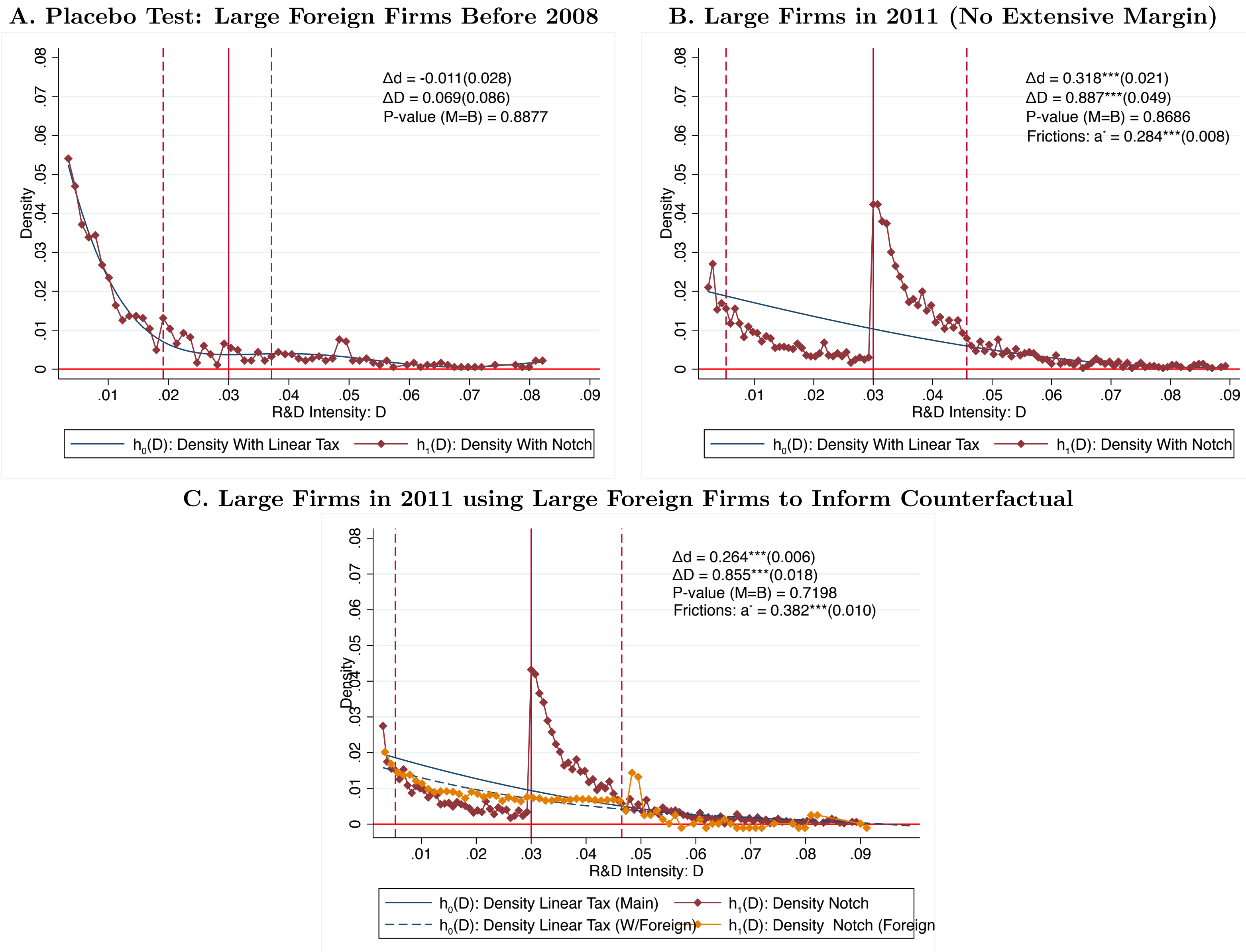
NOTES: This figure reports the results of our bunching estimator for small, medium, and large firms in 2009. In each panel, we plot the empirical density of R&D intensity in red and the estimated counterfactual R&D intensity in blue. The lower bound d^{*-} and upper bound d^{*+} for the excluded region are indicated by vertical dashed lines. Δd is the percentage increase in R&D in the excluded region, ΔD is the increase for the marginal firm, a^* is the fraction of firms that are constrained from participating in the program, and we report the p-value of the test that the missing mass equals the excess mass. See Section 5.1 for details. Source: Administrative Tax Return Database.

Figure 9: Estimates of Excess Mass from Bunching at Notch (2011)



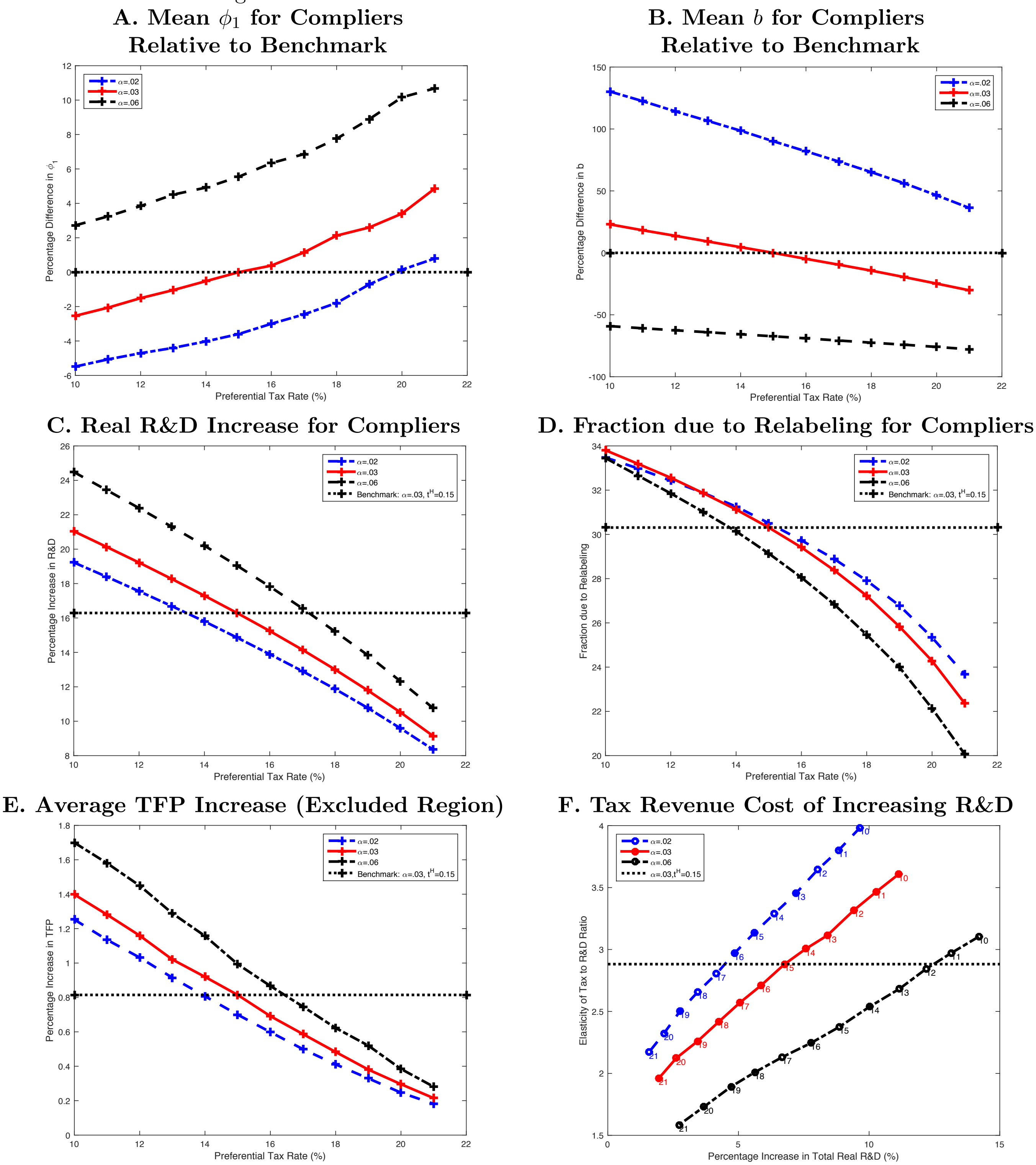
NOTES: This figure reports the results of our bunching estimator for small, medium, and large firms in 2011. In each panel, we plot the empirical density of R&D intensity in red and the estimated counterfactual R&D intensity in blue. The lower bound d^{*-} and upper bound d^{*+} for the excluded region are indicated by vertical dashed lines. Δd is the percentage increase in R&D in the excluded region, ΔD is the increase for the marginal firm, a^* is the fraction of firms that are constrained from participating in the program, and we report the p-value of the test that the missing mass equals the excess mass. See Section 5.1 for details. Source: Administrative Tax Return Database.

Figure 10: Robustness of Bunching Estimates



NOTES: This figure reports robustness checks of our bunching estimator in panel C of Figure 9. 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 which already performed R&D in previous years. Panel C uses large foreign firm's R&D intensity before 2008 to inform the counterfactual distribution. See Section 5.1 for details.

Figure 11: Simulated Effects of Counterfactual Policies



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, and panel E shows how this affects TFP. Finally, panel F 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 6.2 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: 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	-3.822	-0.136	-0.054
Admin Costs (level)	-0.003	0.001	-3.686	-0.005	-0.002
R&D	0.146	0.065	2.245	0.037	0.251
R&D (real)	0.090	0.044	2.074	0.022	0.165
User Cost	-0.071	0.037	-1.929	-0.130	-0.009
2011					
Tax	-0.128	0.018	-7.293	-0.159	-0.101
TFP	0.012	0.006	1.953	0.001	0.022

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

	Estimate	Bootstrap 5th Perc.	Bootstrap 95th Perc.
Reported R&D to User Cost (2009)	-2.052	-7.919	-0.016
Real R&D to User Cost (2009)	-1.272	-4.900	-0.010
Tax to Reported R&D (2011)	-0.879	-2.730	-0.458

NOTES: This table reports estimates of ITT effects of the notch on various outcomes. Panel B reports ratios of estimates in panel A. Standard errors computed via bootstrap. See Section 3.1 for details on data sources and Section 5 for details on the estimation. Source: Administrative Tax Return Database.

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

Table 4: 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
Estimate	0.098	5.663	8.581	1.648	0.629
SE	0.004	0.175	0.216	0.137	0.043

NOTES: This table reports estimates of structural parameters of the model in Section 4. Estimates based on calibrated values of $\theta = 5$, $\rho = 0.725$, and $\sigma = 0.385$. See Section 6 for estimation details.

B. Simulated vs. Data Moments

	Simulated	Data
Probability Mass for $d < d^{-*}$	0.284	0.280
Fraction not Bunching	0.676	0.675
Probability Mass for $d > d^{+*}$	0.198	0.189
Bunching Point d^{-*}	0.75%	0.88%
ITT reported R&D	0.162	0.146
ITT TFP	0.008	0.012
ITT administrative cost ratio	-0.27%	-0.33%
Average TFP for $d < d^{-*}$	-0.032	-0.032
Average TFP for d between d^{-*} and d^{+*}	0.003	0.000
Average TFP for $d > d^{+*}$	0.056	0.056

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 our model does a remarkable job of matching 10 moments from the data using a relatively parsimonious model based on 5 parameters. See Section 6 for estimation details and discussion of how these moments inform the structural parameters in our model.

Online Appendix: Not For Publication

This appendix contains multiple additional analyses. Appendix A includes additional details of the Chinese corporate income tax system. Appendix B describes in more detail the data we use in our analysis. Appendix C discusses the estimation of our measure of log-TFP. Appendix D shows that firms do not respond to the InnoCom program by manipulating sales expenses. Appendix E provides a detailed derivation of the model. Appendix F shows that the missing mass in the bunching analysis can be used to approximate the effects of the notch on R&D investment. Appendix G discusses details of the implementation of the bunching estimator. Appendix H discusses additional robustness checks of our bunching estimates. Appendix I describes details of the implementation of the ITT estimator. Finally, Appendix J explores the robustness of our structural estimation by showing that the actual distribution of TFP is very close to being log-normal and by discussing estimates of an alternative structural model with heterogeneous ε.

A Additional Details of 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 - 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 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 Zone (SEZ) and Economic and Technology Development Zone (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 $\$50,000$, and is effectively irrelevant for our sample.

The Chinese government implemented a major corporate tax reform in 2008 in order 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 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 until 2020.

B Data Sources

We connect three 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 additional state-owned firms with lower sales. This survey provides detailed information on ownership, location, production, and the 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 this data with a separate survey by the Chinese National Bureau of Statistics that includes firms' reported R&D. We use these data for years 2006–2007.

The second dataset we use is the administrative enterprise income tax records from Chinese State Administration of Tax (SAT). The SAT is the counterpart to 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 have acquired these administrative enterprise income tax records from 2008–2011, which allows us to construct detailed tax rate information for individual manufacturing firms. We also use these data to construct residualized measures of firm productivity.[49] The scope of the SAT data is slightly different from the ASM, but there is a substantial amount of overlap for the firms which conduct R&D. For instance, the share of total R&D that can be matched with ASM records is close to 85% in 2008.

The third dataset we use is the list of firms that are enrolled in the InnoCom program from 2008–2014. For each of these manufacturing firms, we have the exact Chinese name, and the year it was certified with high-tech status. This list is available from the Ministry of Science and Technology website, and we have digitized it in order to link it to the SAT and ASM data. We use these data to

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

cross-validate the high-tech status recorded in the SAT data.

C Estimation of Residual Productivity

This appendix describes how we construct an 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 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:

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

D Lack of Manipulation of Other Expenses

In Figure 4, 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 the 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 4 in Figure A.4 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.4 shows that this drop is not statistically significant. This analysis suggests that the drops we observe in administrative costs are likely not due to substitution of inputs, and are likely due to relabeling.

E Detailed Model Derivation

E.1 Model Setup

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

$$q_{it} = \exp\{\phi_{it}\} F(K_{it}, \cdots, V_{it}),$$

where $K_{it}, \cdots, V_{it}$ are static inputs with prices w_{it}, and where ϕ_{it} is log-TFP which follows the law of motion given by:

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

where $D_{i,t-1} \geq 0$ 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 cost function for this familiar problem is given by:

$$C(q; \phi_{it}, w_{it}) = qc(\phi_{it}, w_{it}) = q\frac{c(w_{it})}{\exp\{\phi_{it}\}},$$

where $c(\phi_{it}, w_{it}) = \frac{c(w_{it})}{\exp\{\phi_{it}\}}$ is the unit cost function.

The firm faces a constant elasticity demand function given by:

$$p_{it} = q_{it}^{-1/\theta},$$

where $\theta > 1$. Revenue for the firm is given by $q_{it}^{1-1/\theta}$. In a given period, the firm chooses q_{it} to

$$\max_{q_{it}} q_{it}^{1-1/\theta} - q_{it}c(\phi_{it}, w_{it}).$$

The profit-maximizing q_{it} is given by:

$$q_{it}^* = \left(\frac{\theta - 1}{\theta} \frac{1}{c(\phi_{it}, w_{it})}\right)^{\theta}.$$

Revenue is then given by:

$$\text{Revenue}_{it} = \left(\frac{\theta}{\theta - 1} \frac{1}{c(\phi_{it}, w_{it})}\right)^{\theta-1} = \frac{\theta}{\theta - 1} q_{it}^* c(\phi_{it}, w_{it})$$

That is, revenues equal production costs multiplied by a gross-markup $\frac{\theta}{\theta-1}$. 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 around 1.2. Per-period profits are then given by:

$$\pi_{it} = \frac{1}{\theta - 1} q_{it}^* c(\phi_{it}, w_{it}) = \frac{(\theta - 1)^{\theta-1}}{\theta^\theta} c(\phi_{it}, w_{it})^{1-\theta}.$$

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

$$\begin{aligned}\mathbb{E}[\pi_{it}] &= \frac{(\theta-1)^{\theta-1}}{\theta^{\theta}}c(\rho\phi_{i,t-1}+(\theta-1)\sigma^2/2, w_{it})^{1-\theta}D_{i,t-1}^{(\theta-1)\varepsilon} \\ &= \mathbb{E}[\pi_{it}|D_{i,t-1}=0]D_{i,t-1}^{(\theta-1)\varepsilon} = \tilde{\pi}_{it}D_{i,t-1}^{(\theta-1)\varepsilon},\end{aligned}$$

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 depends on the parameter b:

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

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

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

Notice first that if the tax rate is constant across periods, the corporate income tax does not affect the choice of R&D investment.[50] In the special case of no adjustment costs (i.e., $b=0$), the optimal choice of D_{i1} is given by:

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

Even in the general case (unrestricted b), we also observe that the choice of R&D depends on potentially-unobserved, firm-specific factor ϕ_{i1} that influences $\tilde{\pi}_{i2}$. A useful insight for the proceeding analysis is that we can recover these factors from D_{1i} as follows:

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

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

Substituting $\tilde{\pi}_{i2}$ into the objective function, we can write the value of the firm as

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

Rewriting this equation in terms of firm's optimal R&D intensity $d^*_{i1} = \frac{D^*_{i1}}{\theta\pi_{i1}}$, the value-to-sales ratio is

$$\frac{\Pi(d^*_{i1}|t_2)}{\theta\pi_{i1}} = (1-t_1)\left[\frac{1}{\theta} + d^*_{i1}\left(\frac{1}{(\theta-1)\varepsilon}-1\right)+(d^*_{i1})^2\left(\frac{b}{(\theta-1)\varepsilon}-\frac{b}{2}\right)\right]. \tag{E.3}$$

Second Order Condition

This problem may feature multiple solutions. To ensure our model results in sensible solutions, we confirm 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 $(\theta-1)\varepsilon < 1$ in order for the second order condition to hold. We can also use the implicit function theorem to show that R&D decision D^*_{i1} is increasing in ϕ_{i1} if $(\theta-1)\varepsilon < 1$, which is consistent with numerous empirical studies.

E.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:

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

$t_2^{LT} > t_2^{HT}$ 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 costs of certification c such that firms need to pay $c \times \alpha\theta\pi_{i1}$ to obtain the tax benefit when they pass the R&D intensity threshold. Intuitively, this tax structure induces a notch in the profit function at $d_1 = \alpha$. Figure 6 presents two possible scenarios following this incentive. Panel A shows the situation where the firm finds it optimal to choose a level of R&D intensity below the threshold. At this choice, the first order condition of the linear tax case holds and the optimal level of R&D is given by Equation E.1. From this panel, we can observe that a range of R&D intensity levels below the threshold are dominated by choosing an R&D intensity that matches the threshold level α. Panel B shows a situation where the firm that is indifferent between the internal solution of panel A and the "bunching" solution of panel B. The optimal choice of R&D for this firm is characterized both by Equation E.1 and by equating $d^*_1 = \alpha$.

Which of the two scenarios holds depends on determinants of the R&D investment decision that may vary at the firm level and are summarized by $\tilde{\pi}_{i2}$, adjustment and fixed costs b, c, as well as on the degree to which R&D investment is valued by firms in terms of future profits (i.e. $\varepsilon(\theta-1)$). However, as long as $\tilde{\pi}_{i2}$ and $(b,\,c)$ are smoothly distributed around the threshold α, this incentive will lead a mass of firms to find $d_1=\alpha$ optimal and thus "bunch" at this level.

We first calculate the optimal profit of the firm conditioning on bunching at the notch, $\Pi(\alpha\theta\pi_1|t_2^{HT})$, by substituting for the unobserved components of the firm-decision, i.e. $\tilde{\pi}_{i2}$, using Equation E.1 to obtain:

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

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 again in terms of the optimal interior R&D intensity d_{i1}^* as

$$\frac{\Pi(\alpha|t_2^{HT})}{\theta\pi_{i1}}=(1-t_1)\left[\frac{1}{\theta}+\alpha\left(\left(\frac{d_{i1}^*}{\alpha}\right)^{1-(\theta-1)\varepsilon}(1+bd_{i1}^*)\frac{(1-t_2^{HT})}{(1-t_2^{LT})}\frac{1}{\varepsilon(\theta-1)}-(1+c)-\frac{\alpha b}{2}\right)\right]. \quad \text{(E.4)}$$

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

$$\begin{aligned}
&\left(\frac{d_{i1}^*}{\alpha}\right)^{1-(\theta-1)\varepsilon}\left(1+\alpha b\frac{d_{i1}^*}{\alpha}\right)\frac{(1-t_2^{HT})}{(1-t_2^{LT})}\frac{1}{\varepsilon(\theta-1)}-(1+c)-\frac{\alpha b}{2} \\
&\qquad \geq\frac{d_{i1}^*}{\alpha}\left(\frac{1}{(\theta-1)\varepsilon}-1\right)+\alpha\left(\frac{d_{i1}^*}{\alpha}\right)^2\left(\frac{b}{(\theta-1)\varepsilon}-\frac{b}{2}\right) \qquad \text{(E.5)}
\end{aligned}$$

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 E.5 holds with equality.

E.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 mis-reporting is proportional to the reported R&D, $\tilde{D}_{i1}$, and depends on the percentage of mis-reported 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 $\Pi(\alpha\theta\pi_1, D_1^{*K}|t_2^{HT})$. We again substitute for the unobserved components of the firm-decision, i.e. $\tilde{\pi}_{i2}$ with the interior optimal R&D D_{i1}^* using Equation E.1:

$$\begin{aligned} \max_{D_{i1}^K} \quad & (1-t_1)\left(\pi_{i1} - D_{i1}^K - \alpha\theta\pi_{i1}c - \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}^K}{\theta\pi_{i1}}\right]^2\right) - \alpha\theta\pi_1\tilde{h}\left(\frac{\alpha\theta\pi_1 - D_{i1}^K}{\alpha\theta\pi_1}\right) \\ + \quad & \frac{(1-t_1)(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{D_{i1}^K}{D_{i1}^*}\right)^{(\theta-1)\varepsilon}\left(1 + b\left[\frac{D_{i1}^*}{\theta\pi_{i1}}\right]\right)D_{i1}^* \end{aligned}$$

The first order condition is:

$$\begin{aligned} \left(1 + b\left[\frac{D_{i1}^{*K}}{\alpha\theta\pi_{i1}}\right]\right) \quad = \quad & \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\left(\frac{D_{i1}^{*K}}{D_{i1}^*}\right)^{(\theta-1)\varepsilon-1}\left(1 + b\left[\frac{D_{i1}^*}{\theta\pi_{i1}}\right]\right) \\ + \quad & \tilde{h}'\left(\frac{\alpha\theta\pi_1 - D_{i1}^{*K}}{\alpha\theta\pi_1}\right)\frac{1}{1-t_1} \end{aligned}$$

This equation defines the optimal relabeling strategy δ_{i1}^* as an implicit function of the interior optimal R&D intensity d_{i1}^* as the following:

$$\left(\frac{d_{i1}^*}{\alpha(1-\delta_{i1}^*)}\right)^{1-(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)(1 + \alpha b\left(\frac{d_{i1}^*}{\alpha}\right)) = \frac{1}{\alpha}\left[1 + b\left(\frac{d_{i1}^*}{\alpha}\right) - \frac{\tilde{h}'(\delta_{i1}^*)}{(1-t_1)}\right] \tag{E.6}$$

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

$$\begin{aligned} & \underbrace{\left(\frac{d_{i1}^*}{\alpha(1-\delta_{i1}^*)}\right)^{1-(\theta-1)\varepsilon}(1 + bd_{i1}^*) \times \frac{(1-\delta_{i1}^*)}{(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) - c - (1-\delta_{i1}^*) - \frac{\alpha b}{2}(1-\delta_{i1}^*)^2}_{\text{Relative Profit from Bunching}} \\ & \qquad - \underbrace{\frac{\tilde{h}(\delta_{i1}^*)}{\alpha(1-t_1)}}_{\text{relabeling Cost}} \geq \underbrace{\frac{d_{i1}^*}{\alpha}\left(\frac{1}{(\theta-1)\varepsilon} - 1\right) + \alpha\left(\frac{d_{i1}^*}{\alpha}\right)^2\left(\frac{b}{(\theta-1)\varepsilon} - \frac{b}{2}\right)}_{\text{Relative Profit Not Bunching}}. \end{aligned} \tag{E.7}$$

The marginal firm $d_{b,c}^{*-}$ in this case is determined by Equation E.6 and Equation E.7 when it holds with strict equality.

F Bunching Approximations

This section provides detailed derivations of expressions that approximate changes in the R&D investment with the estimated density.

F.1 Percentage Increase in R&D Intensity of Marginal Firm

As in Kleven and Waseem (2013), we can approximate the behavior of the marginal firm with the quantities B and $h_0(\alpha)$. We first consider the special case without frictions, and note that

$$B = \int_{d^{*-}}^{\alpha} h_0(u)\,du \approx h_0(\alpha)\left(\alpha - d^{*-}\right) = h_0(\alpha)\alpha\underbrace{\frac{\alpha - d^{*-}}{\alpha}}_{\Delta D^*}. \tag{F.1}$$

The first part of Equation F.1 makes the point that the excess mass B will equal the fraction of the population of firms that would have located in the dominated region. This quantity is defined by the integral of the counterfactual distribution $h_0(\cdot)$ over the dominated interval, which is given by (d^{*-}, α). The second part of Equation F.1 approximates this integral by multiplying the length on this interval by the value of the density at α. Simplifying this expression and solving for ΔD^* we obtain:

$$\Delta D^* \approx \frac{B}{h_0(\alpha)\alpha}.$$

Thus, in order to estimate ΔD^*, it suffices to have an estimate of the counterfactual density $h_0(\cdot)$, and to use this to recover the quantities B and $h_0(\alpha)$. Note that while ΔD^* is the percentage increase relative to the notch, the percentage increase relative to the initial point of the marginal firm is given by: $\frac{\Delta D^*}{1-\Delta D^*} = \frac{\alpha - d^{*-}}{d^{*-}}$. Similarly, the increase in R&D intensity for the marginal firm is given by $\alpha\Delta D^* = \alpha - d^{*-}$.

In the case of heterogeneous frictions, we may obtain a similar approximation if we assume that the probability of being constrained from responding to the program does not depend on d. While this may be a strong assumption, it provides a useful approximation for B. To see this, note that

$$\begin{aligned} B &= \int_{d^{*-}}^{\alpha} \int_{b,c} \mathbb{I}[d \geq d^{-}_{b,c}] h_0(d,b,c) d(b,c) dd \\ &= \int_{d^{*-}}^{\alpha} \int_{b,c} \mathbb{I}[d \geq d^{-}_{b,c}] h_0(b,c|d) d(b,c) h_0(d) dd \\ &= \int_{d^{*-}}^{\alpha} (1 - \mathbb{P}r(\text{Constrained}|d)) h_0(d) dd, \end{aligned}$$

where the second line uses the definition of conditional probability, and the third line integrates over (b, c). Using the assumption that $\mathbb{P}r(\text{Constrained}|d)$ does not depend on d and using the same approximation as in Equation F.1, we obtain:

$$\begin{aligned} B &= (1 - \mathbb{P}r(\text{Constrained})) \int_{d^{*-}}^{\alpha} h_0(d)dd \\ &\approx (1 - \mathbb{P}r(\text{Constrained}))h_0(\alpha)\alpha \underbrace{\frac{\alpha - d^{*-}}{\alpha}}_{\Delta D^*}. \end{aligned}$$

The formula for ΔD^* now becomes:

$$\Delta D^* \approx \frac{B}{h_0(\alpha)\alpha(1 - \mathbb{P}r(\text{Constrained}))}.$$

F.2 Average Percentage Increase in R&D Intensity

We now derive an approximation for the average percentage increase in R&D due to the notch. We begin by writing the average R&D intensities in both situations as:

$$\mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})] = \int_{d^{*-}}^{d^{*+}} dh_0(d)dd \approx \underbrace{\frac{\alpha + d^{*-}}{2}}_{\underline{d}} \int_{d^{*-}}^{\alpha} h_0(d)dd + \underbrace{\frac{d^{*+} + \alpha}{2}}_{\bar{d}} \int_{\alpha}^{d^{*+}} h_0(d)dd$$

$$\mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})] = \int_{d^{*-}}^{d^{*+}} dh_1(d)dd \approx \underbrace{\frac{\alpha + d^{*-}}{2}}_{\underline{d}} \int_{d^{*-}}^{\alpha} h_1(d)dd + \underbrace{\frac{d^{*+} + \alpha}{2}}_{\bar{d}} \int_{\alpha}^{d^{*+}} h_1(d)dd$$

We can then write the change in R&D intensity as:

$$\begin{aligned} \mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})] - \mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})] &\approx \bar{d} \underbrace{\int_{\alpha}^{d^{*+}} (h_1(d) - h_0(d))dd}_{B} \\ &+ \underline{d} \underbrace{\int_{d^{*-}}^{\alpha} (h_1(d) - h_0(d))dd}_{-B} \\ &= B(\bar{d} - \underline{d}), \end{aligned} \tag{F.2}$$

where we use the fact that the excess mass above the notch is equal to the missing mass below the notch.

Taking the following approximation of $\mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]$:

$$\begin{aligned} \mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})] &= \int_{d^{*-}}^{d^{*+}} dh_0(d)dd \approx \int_{d^{*-}}^{d^{*+}} \alpha h_0(\alpha)dd \\ &= \alpha h_0(\alpha)(d^{*+} - d^{*-}) = 2\alpha h_0(\alpha)(\bar{d} - \underline{d}), \end{aligned}$$

we obtain:

$$\frac{\mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})] - \mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]}{\mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]} = \frac{B}{2\alpha h_0(\alpha)}. \tag{F.3}$$

Note that, while these derivations do not explicitly include the role of heterogeneous frictions, these expressions are not affected by the presence of heterogeneous frictions.

F.3 Identification of Intent-to-Treat Effect

The ITT estimates are identified by firms that "comply" with the tax incentive. To see this, note:

$$\begin{aligned}
\mathbb{E}[Y|\text{No Notch}, d \in (d^{*-}, d^{*+})] &= \underbrace{\int_{d^{*-}}^{\alpha} Y h_0(d) \times \mathbb{P}r(\text{Constrained}|d) dd}_{\text{Never Takers}} \\
&+ \underbrace{\int_{d^{*-}}^{\alpha} Y h_0(d) \times (1 - \mathbb{P}r(\text{Constrained}|d)) dd}_{\text{Compliers}} + \underbrace{\int_{\alpha}^{d^{*+}} Y h_0(d) dd}_{\text{Always Takers}}
\end{aligned}$$

Similarly, we can write

$$\begin{aligned}
\mathbb{E}[Y|\text{Notch}, d \in (d^{*-}, d^{*+})] &= \underbrace{\int_{d^{*-}}^{\alpha} Y h_1(d) dd}_{\text{Never Takers}} \\
&+ \underbrace{\int_{\alpha}^{d^{*+}} Y h_1(d) \times (1 - \mathbb{P}r(\text{Constrained}|d)) \times \mathbb{I}[d_0 \in (d^{*-}, \alpha)] dd}_{\text{Compliers}} \\
&+ \underbrace{\int_{\alpha}^{d^{*+}} Y h_1(d) \mathbb{I}[d_0 \in (\alpha, d^{*+})] dd}_{\text{Always Takers}},
\end{aligned}$$

where we assume that there are no defier firms that would be above the notch without the InnoCom program, but would be below the notch with the InnoCom program. Noting that $h_0(d) \times \mathbb{P}r(\text{Constrained}|d) = h_1(d)$, and that $h_1(d) \times \mathbb{I}[d_0 \in (\alpha, d^{*+})] = h_0(d)$, we can write the ITT^Y as:

$$ITT^Y = \int_{\alpha}^{d^{*+}} Y h_1(d)(1 - \mathbb{P}r(\text{Constrained}|d)) \mathbb{I}[d_0 \in (d^{*-}, \alpha)] dd - \int_{d^{*-}}^{\alpha} Y h_0(d)(1 - \mathbb{P}r(\text{Constrained}|d)) dd, \tag{F.4}$$

which is just the change in the average of firms in the excluded region that is driven by the compliers.

Approximation of Intent-to-Treat Effect

Finally, we can obtain more intuition behind the ITT estimates by noting that:

$$B = \int_{\alpha}^{d^{*+}} h_1(d)(1 - \mathbb{P}r(\text{Constrained}|d))\mathbb{I}[d_0 \in (d^{*-}, \alpha)]dd = \int_{d^{*-}}^{\alpha} h_0(d)(1 - \mathbb{P}r(\text{Constrained}|d))dd.$$

Using this fact, the following expression is an approximation of Equation F.4:

$$ITT^Y \approx B(\bar{Y} - \underline{Y}) \tag{F.5}$$

where $\underline{Y}$ is the counterfactual average value of Y for compliers with $d_0 \in (d^{*-}, \alpha)$ and $\bar{Y}$ is the average value of Y for compliers with $d_1 \in (\alpha, d^{*+})$. This equation gives a discrete treatment effect interpretation to the ITT by showing that the ITT is driven by the amount of switching of compliers between the "below notch" and "above notch" regions, given by B, and the change in the outcome associated from being in the "above notch" region. Note that this approximation implies a constant treatment effect. While we do not rely on this assumption in our analysis, we find it useful in order to build intuition for the interpretation of the ITT estimates.

G 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 $p < 7$, and all 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 the excess mass (above the notch) equals the missing mass (below the notch). 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 between the missing and excess mass at the 10% level.

Note that a common practical problem in the literature is the higher frequency in the reporting of "round numbers." As Figures 2 and A.1 in Section 3 demonstrate, our data does not display "round-number" problems that are often present in other applications.

Finally, we obtain standard errors by bootstrapping the residuals from the series regression, generating 5000 replicates of the data, and re-estimating the parameters.

H Robustness of Bunching Estimates

This section discusses additional robustness checks of our results in Section 5.1. 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 of the effects of the notch on R&D investment.

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 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.

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

Section 5.2 discusses the estimation of the ITT effects of the notch on our outcomes of interest. The ITT estimates depend on estimates of the counterfactual distribution, $h_0(d)$, as well as the predicted value of the outcome over the excluded region, $\mathbb{E}[Y|d, \text{No Notch}]$. In this section we discuss estimates of these functions. We focus on large firms since, as shown in Figure A.5, 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. The counterfactual density of interest is presented in panel C of Figure 8.

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

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

where we use the same exclusion region as in panel C of Figure 8 (see Appendix G for details), and we use a quadratic polynomial for each outcome. Figure A.8 shows the average value of our outcomes 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 indicate 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 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 for 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.[51]

J Robustness of Structural Model Assumptions

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

Parametric Distribution of Firm Productivity

In our benchmark model, we micro-found the cross-sectional TFP distribution from a Normal AR(1) process. We use the persistence and volatility of the sales for non-R&D firms to calibrate the persistence parameter $\rho = 0.725$ and variance parameter $\sigma = 0.385$. The assumption of this process restricts the cross-sectional distribution of firm TFP $exp(\phi_1)$ to be Lognormal. Since we have constructed firm-level TFP in our data, it allows us to check this parametric assumption directly with the TFP data.

We use ideas proposed by Kratz and Resnick (1996) and Head et al. (2014) in this robustness check. The basic idea is to construct the *empirical* CDF of our sample firms' measured TFP as $\hat{F}_i, i = 1, 2, ...N$, with i ranked based on firm TFP and the Nth firm of the highest measured TFP. With the Log-normal parametric assumption, we know the *theoretical* CDF is $F_{LN}(lnTFP) = \Phi(\frac{lnTFP - \mu_{tfp}}{\sigma_{tfp}})$, with Φ as the standard Normal CDF. Thus, we can write the $lnTFP$ of each quantile i as:

$$lnTFP_i = \mu_{tfp} + \Phi^{-1}(F_i)\sigma_{tfp}.$$

With our frequency estimate $\hat{F}_i$, we can then predict the "theoretical" $ln\hat{TFP}_i$ using the formula above. Notice that we have used the parametric Normal assumption in this calculation. This procedure allows

[51] Diamond and Persson (2016) allow for discontinuities in their estimates of $\mathbb{E}[Y|d, \text{No Notch}]$ since, in their application, being manipulated 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.

us to evaluate how reasonable the Lognormal parametric assumption is by comparing the empirical fit of $ln\hat{TFP}$ and $lnTFP$.

In Figure A.9, we show that the predicted TFP from imposing the Lognormal CDF tracks the 45 degree linear line, i.e., the data quite well. It thus provides strong evidence that Lognormal is a reasonable parametric assumption for the TFP distribution.

Heterogeneity in the TFP Elasticity: ε

Our benchmark model assumes firms have heterogeneous technological opportunities of R&D investment that is driven by heterogeneity in adjustment costs, b. An alternative way of modeling the heterogeneity in firms' technological opportunities is to allow for heterogeneity in ε.[52] 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 was allowed to be heterogeneous produced worse fits of the data. Specifically, these models predict that R&D intensity is not increasing in TFP, which is contrary to what we observe in the real world. For this reason, we believe our benchmark model is superior to models with heterogeneous values of ε.

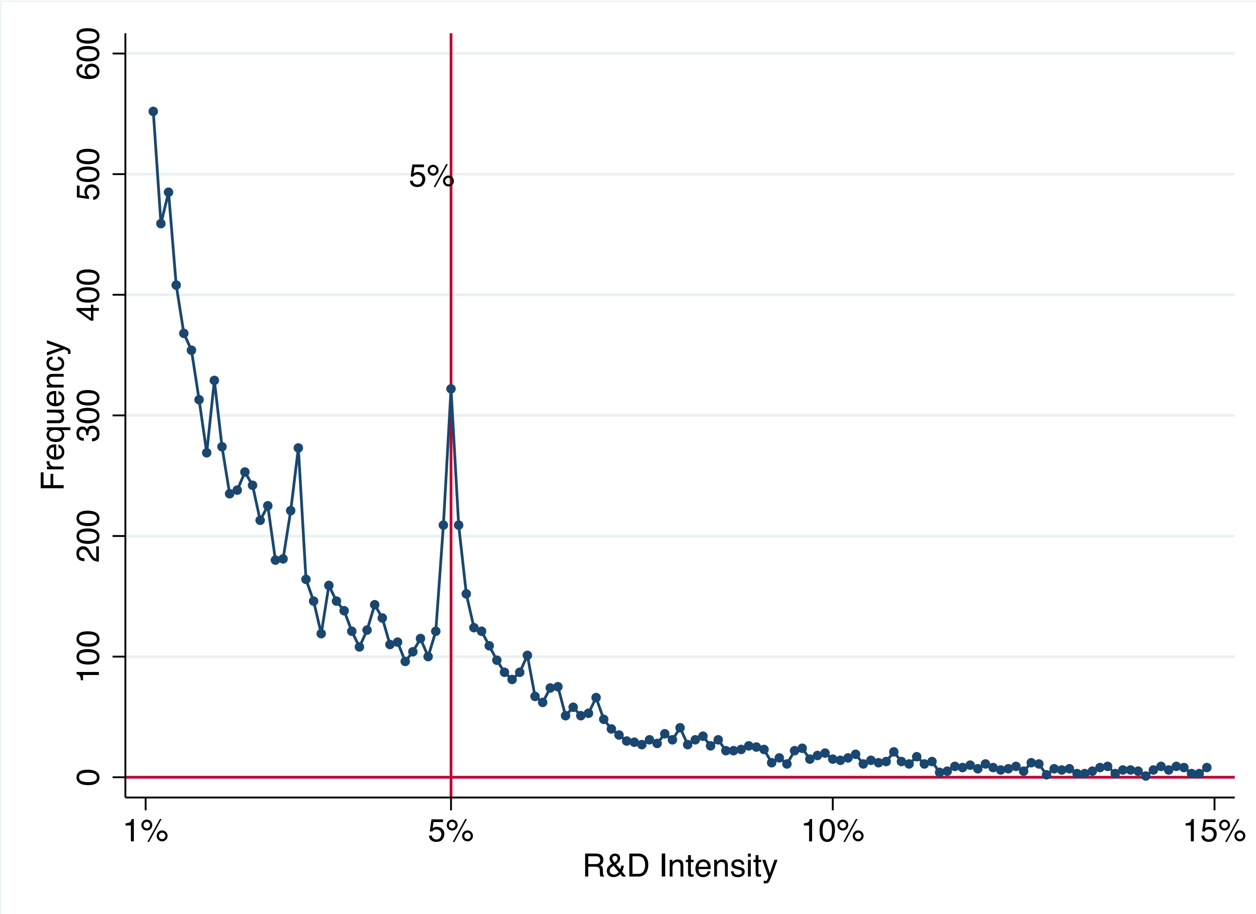
To investigate how this alternative setup affects our results, we estimated models where ε follows a Beta distribution $B(\alpha_\varepsilon, \beta_\varepsilon)$ between 0 and an upper bound of $\bar{\varepsilon}$. We chose the Beta distribution since its probability density function is highly flexible in the interval $[0, \bar{\varepsilon}]$. We estimated two versions of the heterogeneous-ε model. In Model A, we restrict the Beta distribution to be symmetric, 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, the implied mean ε are 0.113 and 0.114 in Model A and B, respectively. These values are comparable to our benchmark value of 0.098. Second, the average adjustment cost parameter is 8.659 and 8.677 for the two cases, again very similar to our benchmark estimate. However, the set of moments summarizing firm TFP at different R&D intensity regions had noticeably worse fit than the our benchmark. When ε is heterogeneous, our model predicts a non-monotonic relationship between TFP and R&D intensity, which is inconsistent with the positive correlation we observe in the data. This is because despite the fact that firm R&D itself is increasing in TFP, its R&D *Intensity* becomes decreasing in TFP when the value of ϵ is small. Combined, these findings indicate that despite obtaining similar estimates of key model parameters, our benchmark model of heterogeneous adjustment cost is a a preferable model for our data.

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

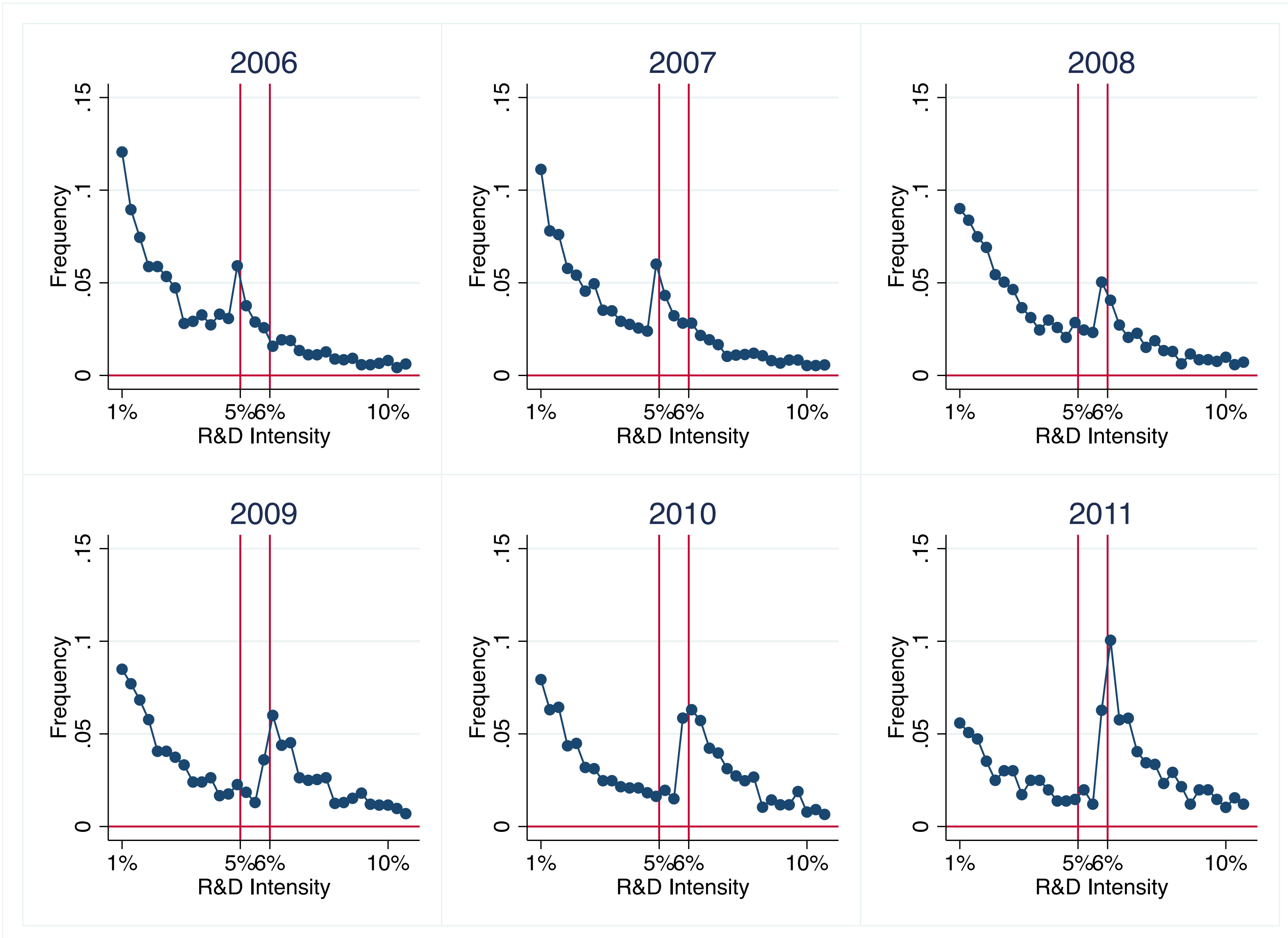
Appendix Graphs

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



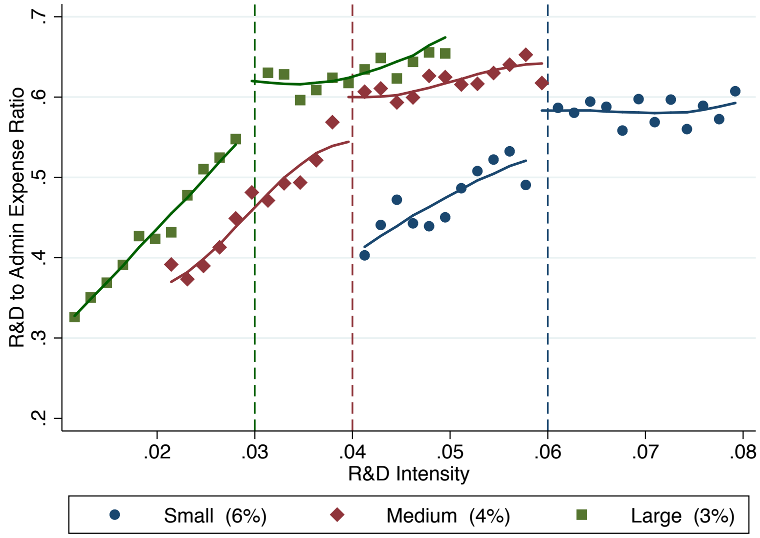
NOTE: This figure plots the R&D intensity distribution of manufacturing firms conducting R&D during the period of 2005 to 2007. We include the firms that had the R&D intensity between 1% and 15%. There is a significant bunching of firms at the 5% threshold. Source: Annual Survey of Manufacturers. See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

Figure A.2: Effects of the 2008 Tax Reform on the Bunching of Domestic-Owned, Small Companies



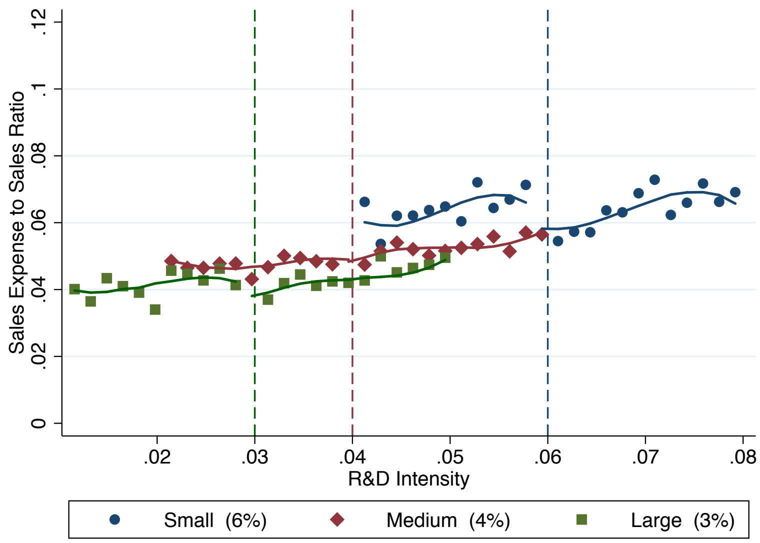
NOTE: This figure plots the R&D intensity distribution of a balanced panel of domestic-owned small companies. These firms' qualifying threshold changed from 5% to 6% due to the 2008 tax reform. These firms gradually adjusted towards the new threshold of 6% from 2008 to 2011. Source: Administrative Tax Return Database and Annual Survey of Manufacturers.
See Section 3.1 for details.

Figure A.3: Alternative Empirical Evidence of Relabeling



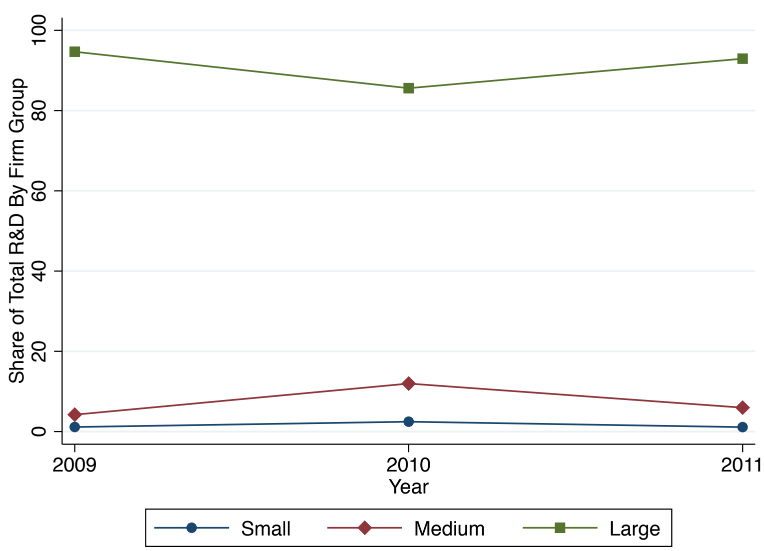
NOTE: This figure summarizes the ratio of R&D to administrative expense for small, medium, and large firms in our sample. This figure shows that this ratio jumps discontinuously across the thresholds of R&D intensity prescribed by the InnoCom program. This suggests firms manipulate their reported R&D intensity by relabeling non-R&D administrative expenses as R&D. See Table A.3 for estimates of the structural break.

Figure A.4: Lack of Manipulation of Sales Expenses



NOTE: This figure shows the binned plot of sales expense-to-sales ratio for each size categories of firms. Table A.4 shows that we do not find a detectable drop in this ratio at the notches.

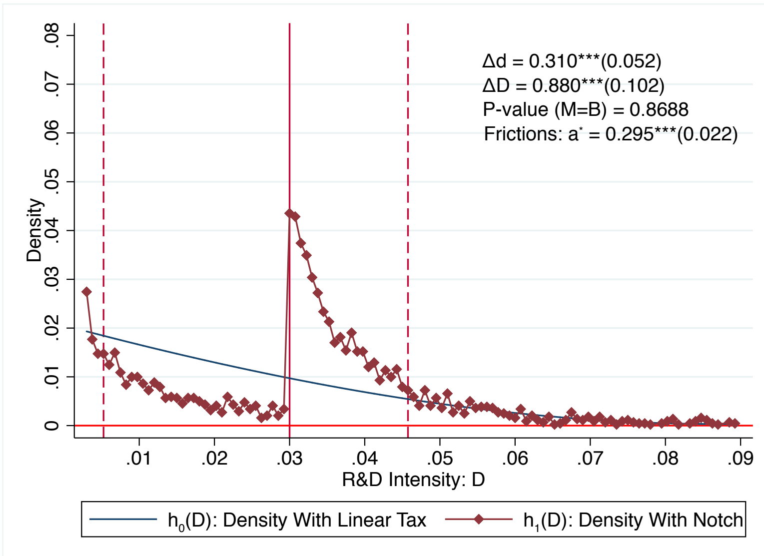
Figure A.5: Aggregate Implications



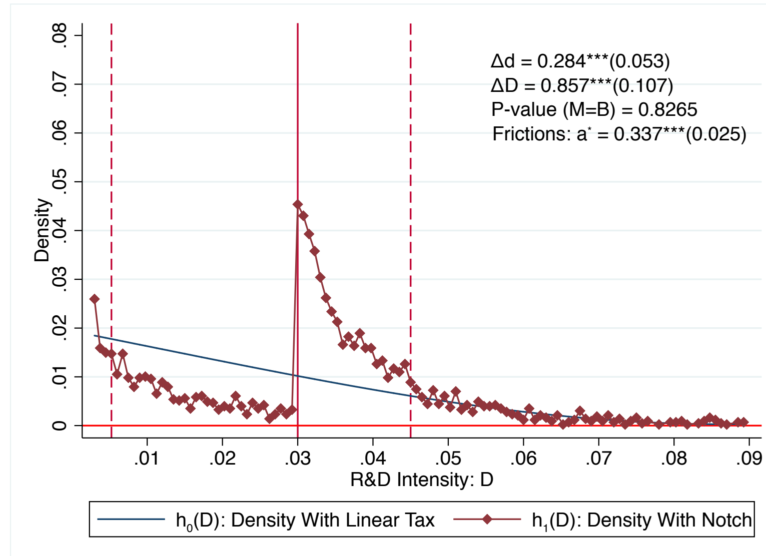
NOTE: This figure summarizes the share of total R&D accounted for by the small, medium, and large firms in our sample. As it illustrates, the large firms account for more 90% of the total R&D and thus is the most important group for aggregate implications of the policy.

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

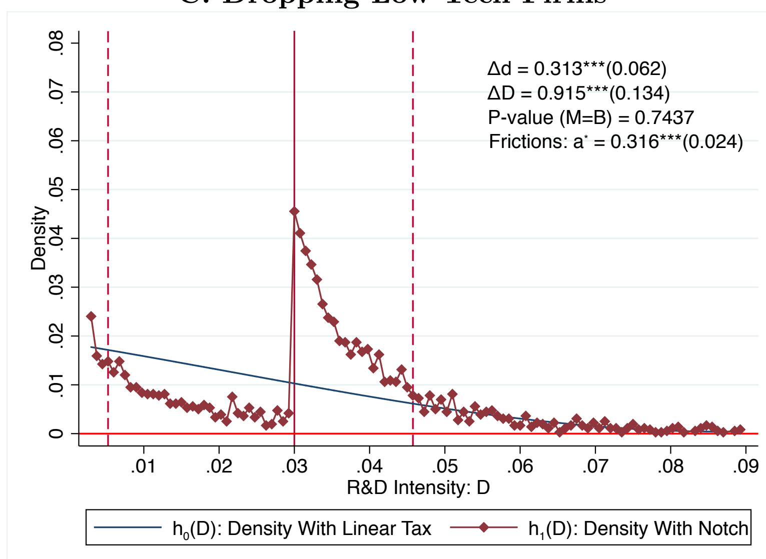
A. Dropping SOEs



B. Dropping Low Profitability Firms

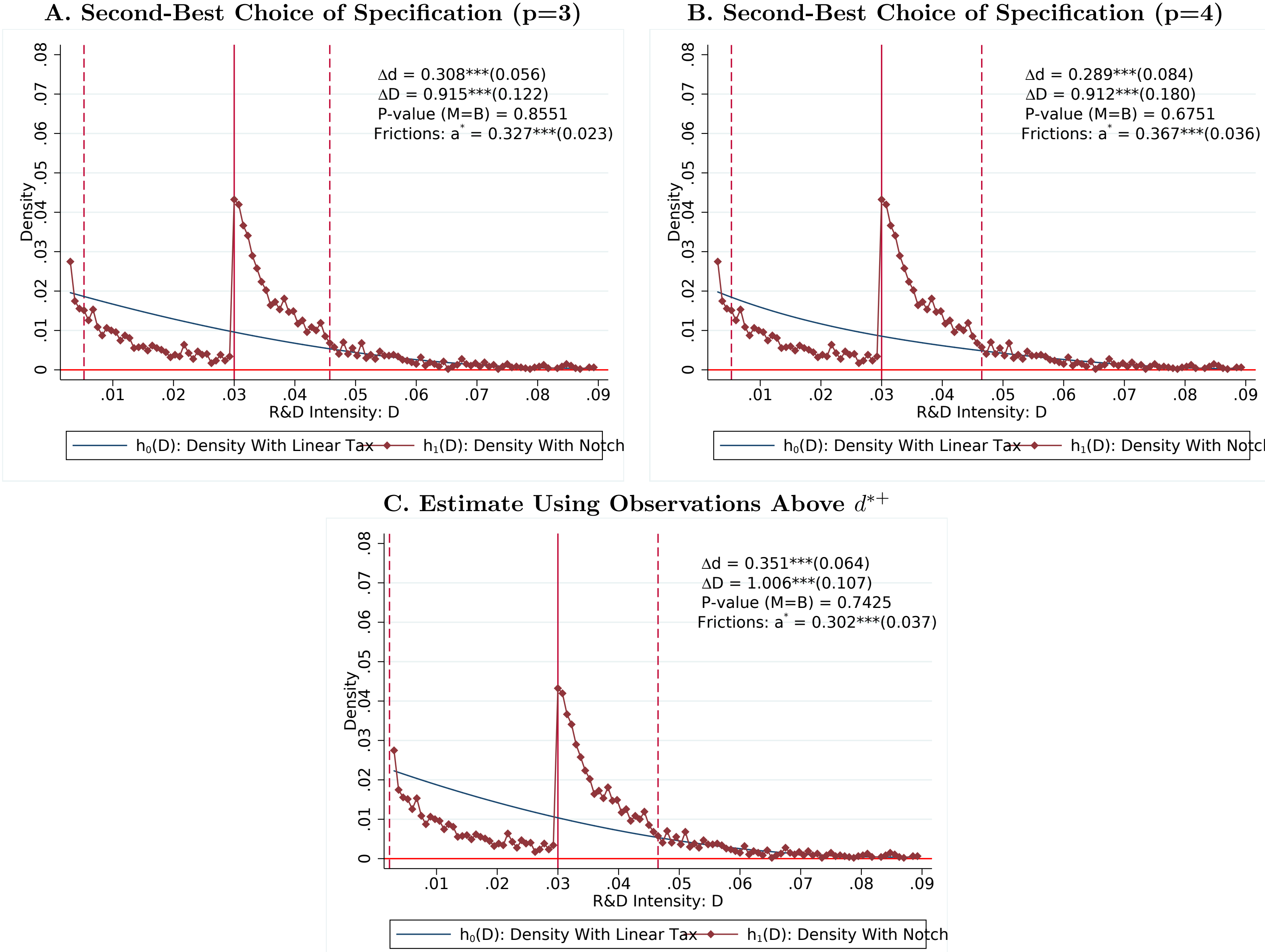


C. Dropping Low Tech Firms



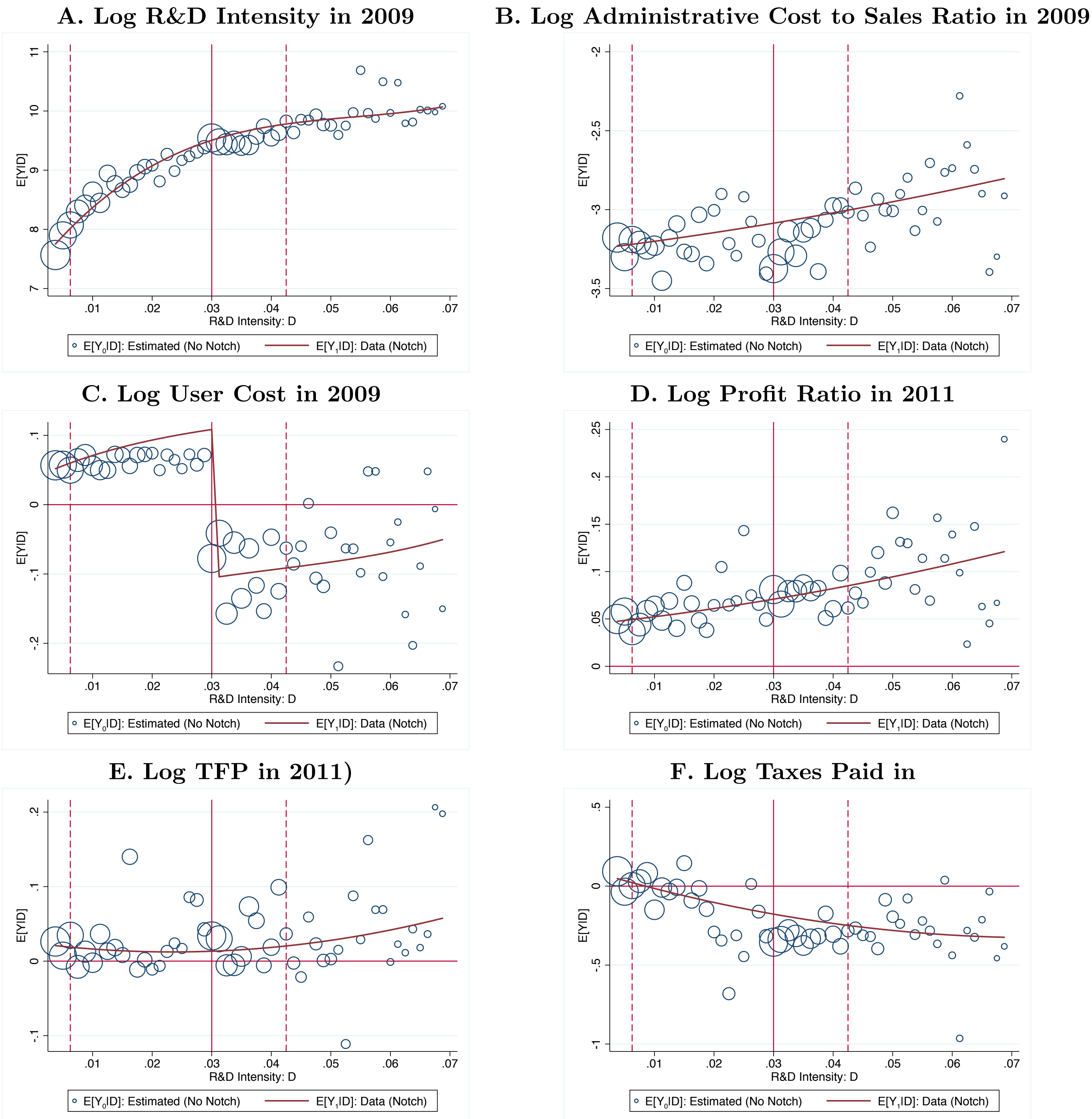
NOTE: This figure conducts robustness checks of the benchmark bunching analysis for large firms in 2011. In panel A, we drop the State-owned enterprises. In panel B, we drop the lowest 20% profitability firms. In panel C, we dropped all the firms that are not classified in the "High Tech" industries defined by the Chinese government. These graphs shows our benchmark results are robust across these subsamples.

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



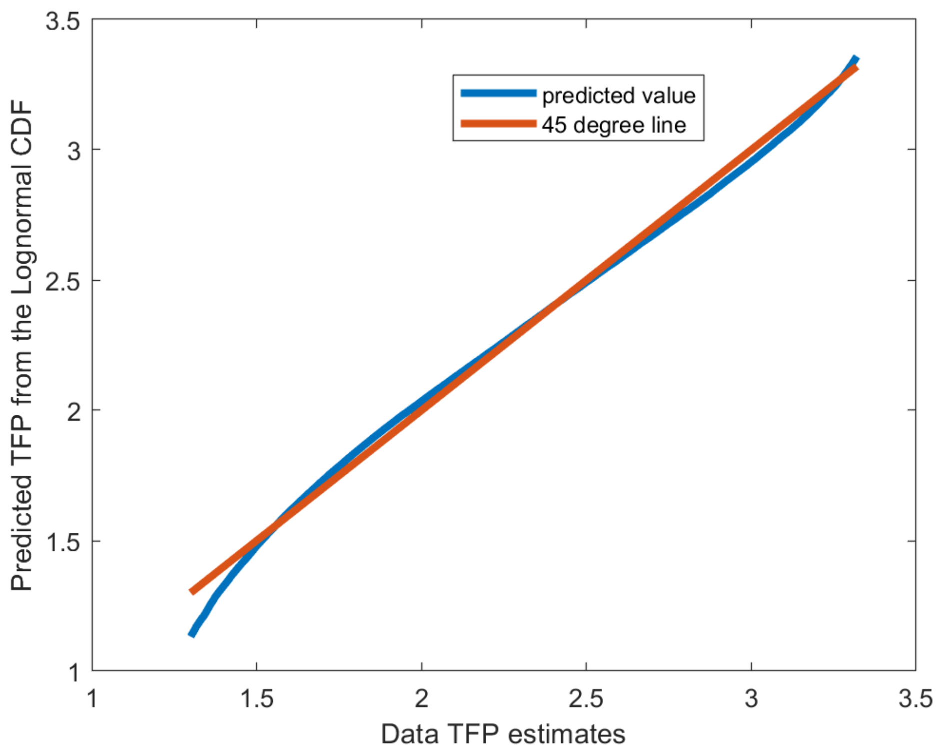
NOTE: This figure conducts robustness checks of the benchmark bunching analysis for large firms in 2011. As discussed in Appendix G, we select (p, d^{*-}, d^{*+}) via cross-validation. In panel A, we use the second-best choice for the specification of (p, d^{*-}, d^{*+}). As in our benchmark case, $p = 3$. 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 shows our benchmark results are robust to how we specify (p, d^{*-}, d^{*+}).

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



NOTE: 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 on 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 I for more details.

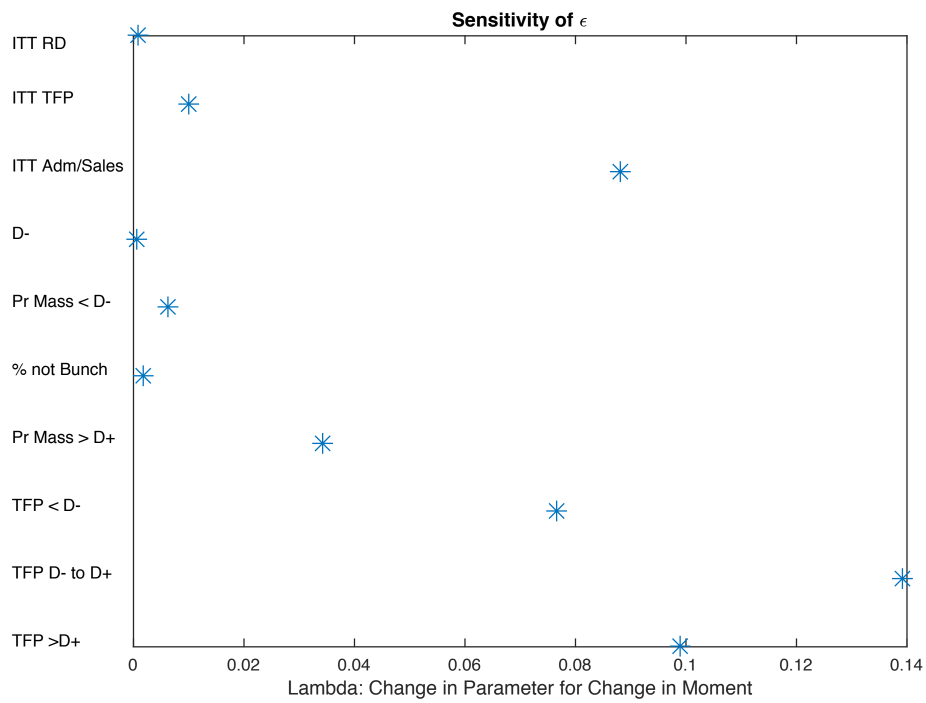
Figure A.9: Data TFP and Predicted TFP under Log Normal Distribution



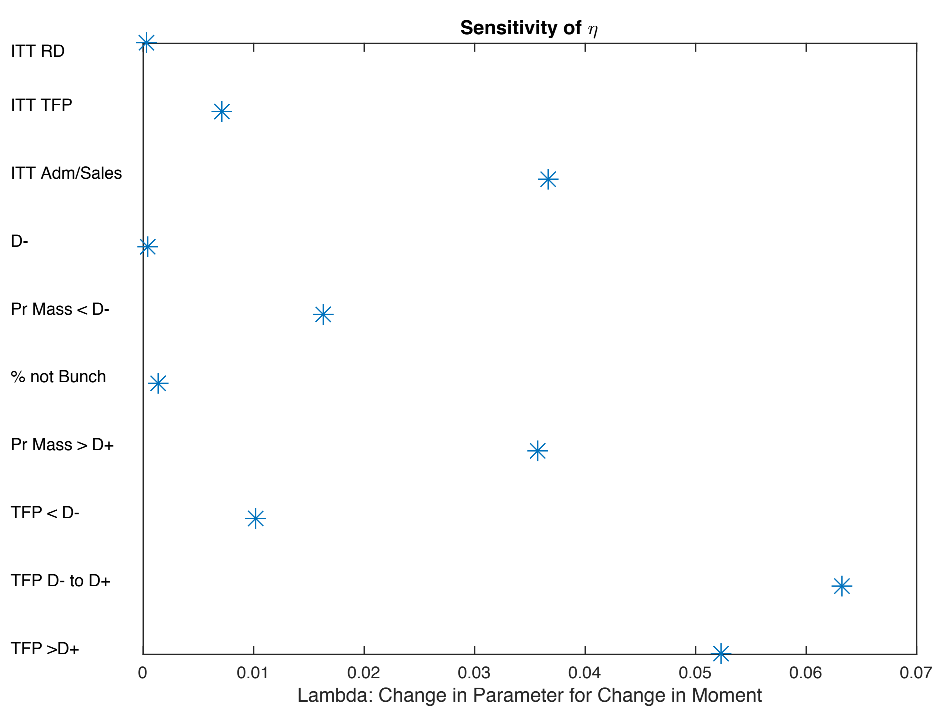
NOTE: This figure reports the predicted TFP from imposing Lognormal CDF and the 45 degree linear line. It shows that the predicted TFP tracks the data TFP quite well. It thus provides strong evidence that Lognormal is a reasonable parametric assumption for the TFP distribution.

Figure A.10: Sensitivity Analysis

A. Sensitivity Analysis for ε



B. Sensitivity Analysis for η



NOTE: This figure reports results of sensitivity analysis based on Andrews et al. (2017). We report the sensitivity matrix Λ, which captures how a local change in each moment affects the parameter estimates. To make it comparable across parameters, we scale the Λ to present the magnitude in terms of percent of each parameter.

Appendix Tables

Table A.1: 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 4. The table shows that the ratio of administrative expenses to sales drops across the notches of the InnoCom program, which suggests 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.2: 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 5. The table shows that firms do not manipulate their sales to comply with 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.3: 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.3. The table shows that the ratio of administrative expenses to R&D jump across the notches of the InnoCom program, which suggests 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.4: 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.4. The table shows that firms do not manipulate sales expenses to comply with the InnoCom program. See Section 3.1 for details on data sources and Section D for details on the estimation. Standard errors in parentheses. Source: Administrative Tax Return Database.

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

Table A.5: 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	-0.095	0.025	-3.809	-0.137	-0.054
Admin Costs (levels)	-0.004	0.001	-3.700	-0.005	-0.002
R&D	0.146	0.065	2.255	0.037	0.250
R&D (real)	0.087	0.042	2.051	0.021	0.158
User Cost of Capital	-0.071	0.037	-1.919	-0.131	-0.009
2011					
Tax	-0.130	0.018	-7.345	-0.158	-0.101
TFP	0.012	0.006	1.930	0.002	0.022

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

	Estimate	Bootstrap 5th Perc.	Bootstrap 95th Perc.
Reported R&D to UCC (2009)	-1.914	-7.845	-0.016
Real R&D to UCC (2009)	-1.030	-4.823	-0.012
Tax to Reported R&D (2011)	-1.153	-2.751	-0.459

NOTES: This table reports robustness of estimates of ITT effects of the notch on various outcomes. Relative to Table 3, this table uses an alternative, second-best estimate of the density of counterfactual R&D distribution. Panel B reports ratios of estimates in panel A. Standard errors computed via bootstrap. See Section 3.1 for details on data sources and Section 5 for details on the estimation. Source: Administrative Tax Return Database.

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

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	1.287	0.226	4.838	8.659	1.004
SE	0.192	0.009	0.504	0.044	0.055
	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	3.256	3.860	6.265	8.677	1.029
SE	0.091	0.160	0.103	0.063	0.049

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

B. Simulated vs. Data Moments

	Model A	Model B	Data
Probability Mass for $d < d^{-*}$	0.324	0.296	0.280
Fraction not Bunching	0.676	0.665	0.675
Probability Mass for $d > d^{+*}$	0.259	0.217	0.189
Bunching Point d^{-*}	0.78%	0.90%	0.88%
ITT reported R&D	0.145	0.141	0.146
ITT TFP	0.008	0.009	0.012
ITT administrative cost ratio	-0.25%	-0.24%	-0.33%
Average TFP for $d < d^{-*}$	-0.029	-0.029	-0.032
Average TFP for d between d^{-*} and d^{+*}	-0.050	-0.039	0.000
Average TFP for $d > d^{+*}$	0.075	0.135	0.056

NOTES: This table compares the moments generated by our simulations with those from the data. The simulation is based on 30,000 firms.