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 2018

Abstract

We analyze the effects of a large fiscal incentive for R&D investment in China that awards a lower *average* corporate income tax rate to qualifying firms. The sharp incentives of the program generate notches, or jumps, in firm values, and vary over time and across firm characteristics. We exploit a novel link between survey and administrative tax data of Chinese firms to estimate investment responses, the potential for evasion, as well as effects on productivity and tax payments. We find large responses of reported R&D using a cross-sectional "bunching" estimator that is new in the R&D literature. We also find evidence that firms relabel administrative expenses as R&D to qualify for the program, and that about a third of the increase in R&D may be due to relabeling. These effects imply user cost elasticities of 2 for the reported response, and 1.3 for the real response. Using the panel structure of the data, we estimate that the program increased firm productivity by 1.2% for targeted firms. Finally, we estimate a structural model of R&D investment and relabeling, and simulate the effects of counterfactual policies. We recover an elasticity of real R&D to TFP of 9.8%, and show that the cost-efficiency of the program depends on the selection of firms into the program. These results are crucial ingredients for designing policies that trade-off corporate tax revenue with productivity growth.
JEL Codes: D24, O30, H25, H26.

*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, Jill Popadak, Jim Poterba, Adriano Rampini, Nirupama Rao, Mark Roberts, Leslie Robinson, Stefanie Stantcheva, Daniel Wilson, and Eric Zwick as well as seminar participants at ASSA, CEPR, Chicago Booth, Duke (Fuqua and Econ), FRB Philadelphia, Hong Kong University, IIPF, LSE, National School of Development (PKU), MIT, NBER Development, NBER Entrepreneurship, NBER Chinese Economy, NBER-CCER Conference, NTA, Oxford University, Penn State, Stanford (SIEPR), UCSD, University of Melbourne, Warwick University, and ZEW MaTax. Dan Garrett, Yuxuan He, and Matt Panhans provided outstanding research assistance. Suárez Serrato is grateful for funding from the Kauffman Foundation. All errors remain our own.

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 for the country, which has led to an explosive growth in R&D investment. Figure 1 compares this growth to the experience of other countries and shows that China has now equalled or surpassed developed-country levels of R&D intensity. This paper analyzes the effects the InnoCom program, a large fiscal incentive for R&D investment in the form of a corporate income tax cut. We exploit a novel link between tax return data and survey data 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 sets of 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 cost of the marginal yuan of R&D investment in terms of foregone tax revenue. Second, what is the effect of fiscal incentives on firm-level and aggregate productivity growth, and how much do firms value R&D investment in terms of future profits? These questions are central to the decision of whether and to what degree governments should encourage R&D investment through tax subsidies.

Answers to these questions are often confounded by the lack of large and plausibly exogenous variation in tax incentives. Since R&D usually requires both fixed and adjustment costs, small fiscal incentives are unlikely to have large effects on R&D investment, especially at the individual firm level. A second concern is that, as firms with better prospects for innovation are likely to invest more heavily, comparisons of investment and profitability across different firms yield upward biases in the value of R&D investment to firms. In addition, an outstanding question is 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. Before 2008, firms with an R&D intensity (R&D investment over revenue) above 5% qualified 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. The combination of administrative tax data and survey data provides a new way to precisely measure a firm's R&D investment, exposure to the fiscal incentives, as well as firm-level outcomes of interest, such as productivity. In addition, we leverage the unusual detail in our administrative data to analyze whether firms respond to the tax incentive by relabeling non-R&D expenses.

Overall, we find that firms are highly responsive to the tax incentives in the InnoCom program, and that a significant fraction of the response is due to relabeling of non-R&D expenses. However, we find the program led to large increases in productivity, and that accounting for relabeling behavior results in larger estimates of the effects of R&D on productivity. We use these insights to simulate the effects of alternative policies, and show that firm selection into the program plays a crucial role in determining the effects of the policy on investment, relabeling, and aggregate 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 threshold, 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. Our detailed tax data allows us to separate R&D from other administrative expenses, which we use to show patterns consistent with a significant relabeling response.

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 evasion, as well as on heterogeneity in firm productivity and adjustment costs. Our analysis characterizes the profit function of the firm that is indifferent between the level of R&D implied by the notch and a level of investment below the notch. The model shows that as long 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 Kleven and Waseem (2013) and Saez (2010). Our model also predicts an increase in relabeling, and an increase in productivity that depends on the effect of R&D on productivity, as well as on the fraction of the reported response that corresponds to real activity. 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 provide causal estimates of 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. We first use the bunching estimator to quantify the percentage increase in R&D investment that is due to the tax incentive. Consistent with our descriptive evidence, we find large increases in R&D investment of 30% for large firms, of 20% for medium firms, and of 11% for small firms in 2011. These intent-to-treat estimates mask the behavior of complier and non-complier firms. On average, firms that comply with the program increase investment by 46% for large firms, of 33% for medium firms, and of 29% for small firms.

We then provide causal estimates of the InnoCom program on relabeling, productivity, and tax revenues. We find estimates of intent-to-treat effects that confirm an increase in reported R&D investment and a decrease in administrative costs. We 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. 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 and profitability. In particular, between 2009 and 2011, the program led to an increase of 5.9% in profitability, and 8.4% in productivity for every 100% increase in reported R&D. While the effects of the program on profitability lessen the fiscal cost of the government, we find that increasing reported R&D investment by 10% cost the government a 8.8% decrease in corporate tax revenues.

Finally, we propose a simulated method of moments approach to estimate the structural parameters of our model, including costs of evasion, the effect of R&D on TFP, and the distributions of fixed and adjustment costs. We then use these estimates to simulate the effects of counterfactual policies that change the current policy parameters. We find that firm selection into the program plays a crucial role in determining the economic effects of the program. In particular, if firms have heterogeneous adjustment costs, the firms that participate may not be the most productive. Selection into the program generates misallocation where low productivity firms with low adjustment costs may receive large tax benefits that do not accrue to high productivity firms with high adjustment costs. This lowers the efficiency of the policy and results in a lower ratio of productivity growth to tax expenditures.

The paper relates to several literatures. First, this paper is related to a large literature analyzing tax incentives for R&D investment. Becker (2015) and Hall and Van Reenen (2000) survey evidence of R&D tax incentives, and Hall and Van Reenen (2000) find a dollar-for-dollar effect of tax credits on R&D investment. The recent empirical evidence so far is concentrated in OECD countries, where micro-level data of firm innovation and/or tax records have become increasingly available.[1] While earlier work typically relied on matching and panel data methods, there is an emerging literature that explores the impact of tax incentives on R&D incentives in a quasi-experimental setup, in particular, by exploiting policy discontinuities. Examples include Agrawal et al. (2014), Bøler et al. (2015), Dechezlepretre et al. (2016), Einiö (2014), Guceri and Liu (2015), 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.[2] It is also one of the first studies that combine administrative tax data with industry survey data to study the link between fiscal incentives, R&D investment, and firm-level productivity.

Second, a previous literature has long documented "relabeling" as an important challenge to identifying the real impact of tax incentive on R&D (see Hall and Van Reenen (2000), Eisner et al. (1984), and Mansfield and Switzer (1985)). This issue is likely more severe in a developing economy setting (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 behaviors. Our policy simulations also inform our understanding of the efficiency of different policies when firms may engage in evasion, as in Best et al. (2015). In particular, size-based policies may be preferable to investment tax credits in developing countries if they substantially increase the cost of evasion.

Third, our paper is related to a recent literature that uses non-parametric methods to recover estimates of behavioral responses to taxation by analyzing the effects of sharp economic incentives, such as kinks or notches in tax schedules, on aggregate patterns of "bunching" in distributions of economic activity.[3] 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 combines the estimates of treatment effects on relabeling and productivity with the bunching estimator to recover structural parameters.[4]

[1] For instance, see Agrawal et al. (2014) and Czarnitzki and Licht (2006) for Canada, Einiö (2014) for Finland, Mulkay and Mairesse (2013) for France, Almus and Czarnitzki (2003) and Hussinger (2008) for Germany, Lach (2002) for Israel, Bøler et al. (2015) for Norway, González and Pazó (2008) for Spain, Griffth et al. (2001), Guceri and Liu (2015), and Dechezlepretre et al. (2016) for the UK, and Rao (2015) for the U.S.

[2] Ding and Li (2015) provide a recent review of the effects of Chinese innovation policy.

[3] 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)).

[4] This model allows us to clarify the interpretation of cross-sectional estimates by addressing issues discussed in Einav et al. (2015). Similarly, Blomquist and Newey (2017) note that cross-sectional estimators may not identify structural parameters without variation in the non-linear incentives. We use data from an unaffected set of firms to overcome this concern, and we also study the sensitivity of our structural parameters to changes in the reduced-form moments using the methods of Andrews et al. (2017).

The rest of the paper is organized as follows. Section 1 provides a description of the fiscal incentive for R&D investment, and discusses the potential for relabeling of R&D expenses in China. Section 2 discusses the data, and Section 3 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 evasion responses to the InnoCom program, and how accounting for evasion affects estimates of the effects of R&D on firm-level productivity. Section 6 culminates with the estimation of the structural parameters of the model, and the simulation of counterfactual policies; Section 7 concludes.

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

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%.[5]

Our project analyzes the "InnoCom" program, which targets qualifying "high tech" enterprises (HTE) and provides them a flat 15% income tax rate. This program is most important for DOEs, including both state-owned and domestically private-owned enterprises, as they are not eligible for many other tax breaks. Prior to 2008, the certification process was administered by the local Ministry of Science and Technology, which established a long list of prerequisites. The most important determinants for certification are the following:[6]

1. At least 30% of the firm's (technician) employees must have a college degree, and at least 10% of the firm's total employment should be devoted to R&D.

2. The firm's R&D intensity (ratio of R&D expenditure to total sales) must be greater than or equal to 5%. In addition, more than 60% percent of the R&D expenditure must be incurred within China.

3. The sales of "high tech" products must account for more than 60% of the firm's total sales.

The program thus generates a large fiscal incentive to invest more than 5% of sales on R&D, which we model in Section 4.

[5]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. It 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 which has foreign share of equity larger than 25%. The effective tax rates of FOEs are even lower since most had tax holidays, typically tax free for the first 2 years or when the firm becomes profitable, and then half the 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, but the revenue threshold is as low as $50,000 and is effectively irrelevant for our sample.

[6]The original government regulations also require that the 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.

Corporate Income Tax Reform of 2008

In addition to leveraging the cross-sectional implications of the InnoCom program, we also exploit changes in tax rates across time. 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%.[7] In concert with this reform, the Ministry of Science and Technology reformed the InnoCom program by streamlining the application process, teaming-up with the Ministry of Finance and the National Tax Bureau to improve compliance, and by changing the threshold requirement of R&D intensity as a function of firms' sales. The post-2008 requirements are as follows:

1. Firms with sales below 50 million RMB must maintain an R&D intensity at, or above 6%.
2. Firms with sales above 50 million RMB, but below 200 million RMB must maintain an R&D intensity at, or above 4%.
3. Firms with sales above 200 million RMB must maintain an R&D intensity at, or above 3%.
4. More than 60 percent of R&D expenditures must be incurred within China

The rest of the pre-2008 requirements remain in effect. In addition, 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.

The InnoCom program has several desirable characteristics that allow us to avoid common problems that arise when estimating the effects of fiscal incentives on R&D investment. First, researchers often lack plausibly exogenous variation in fiscal incentives. As firms with better prospects for innovation are likely to invest more in R&D, comparisons of investment and profitability across firms with different levels of R&D may result in upwardly biased estimates of the value of R&D investment to firms. The InnoCom program generates sharp counterfactual predictions for the distribution of R&D intensity by changing firms' *average tax rate*, which generates a notch in firms' after-tax value functions. This allows us to use cross-sectional estimation methods (e.g., Saez (2010), Kleven and Waseem (2013), and Diamond and Persson (2016)) to identify causal effects of the tax incentives on firm investment and productivity.

A second concern is that, since R&D usually requires large fixed costs, even randomly assigned incentives might not have the statistical power to detect meaningful responses. Since the average tax rate of the firm can fall from 33% to 15%, the incentives implied by this program are economically very important and may lead firms to invest in projects with substantial fixed costs.

Potential for Evasion and Relabeling

A final concern is that the reported R&D investment might not represent a real change in investment, but instead might be a form of tax evasion. This concern is important when interpreting the reported elasticity of R&D investment as real activity, and may loom large when measuring the effects of R&D investment on productivity. To our knowledge, the current literature is not able to circumvent this problem. We now discuss features of the institutional environment that limit some forms of evasion and suggest the that the most likely form of evasion is the mis-categorization of administrative expenses as research expenses.

[7] Some of the existing previous tax breaks for FOEs were also gradually phased-out. For instance, FOEs which 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.

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. 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 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. These factors likely eliminate the possibility for all-out evasion.

A second unlikely form of evasion is the reporting of "phantom expenses." China relies on a value-added tax (VAT) system with third-party reporting, and China's State Administration of Tax (SAT) keeps records of transaction invoices between a given firm and its third-party business partners. As in other settings (e.g., Kleven et al. (2011)), it is hard for companies to report expenses that are not reported by third-party vendors. For these reasons, it is very hard, if not impossible, for firms to completely make up "phantom" R&D expenses.

From conversations with the State Administration of Tax as well as corporate executives, we recognize that the most important source of evasion is expense mis-categorization. Specifically, in the Chinese Accounting Standard, R&D is categorized under "Administrative Expenses," which also includes various other expenses that are related to corporate governance.[8] This raises the possibility that firms reallocate the non-R&D administrative expenditure into 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. For these reasons, we choose to focus on this form of evasion since the institutional setting limits other types of evasion.

Our empirical strategy to detect relabeling leverages these institutional features and exploits the detailed cost reporting in our administrative tax data. In particular, our administrative tax data contains detailed information on the breakdown of operating expenses and R&D expenses. This allows us to test whether firms that respond to the InnoCom program change spending in categories that are more likely to be subject to manipulation, such as administrative or clerical services.

2 Data and Summary Statistics

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 data 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

[8]Examples include administrative worker salary, business travel expenses, office equipments, etc.

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.[9] 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, for the year of 2008, the share of total R&D that can be matched with ASM records is close to 85%.

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

Summary Statistics

Table 1 reports descriptive statistics of all the firms in our analysis sample. In panel A, we report the summary statistics of our main dataset from the SAT for all surveyed manufacturing firms from 2008 to 2011. As discussed in Section 1, the 2008 tax reform creates an interesting pre- and post-test for FOEs, as these firms did not have an incentives to obtain the high-tech certification prior to 2008. Similarly, the change in the R&D intensity threshold across size-groups allows us to trace the response of firms across time.

Our data are comprised of around 1.2 million observations and about $300,000$ firms in each sample year. On average, 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-percentile are 0.3%, 1.5%, and 4.3%, respectively. The administrative expense to sales ratio, which we use as a measure of misreporting to detect evasion, 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 interquantile range of 1.8%.

We also report input and output variables that we used to construct measures of firm performance. As in standard micro-level producer data, these variables are all quite dispersed and skewed, and their means are much larger than their medians. For instance, the mean sales is 118.2 million RMB, while the median firm's sales is 10.6 million RMB. Similarly, the average number of workers is 175, while the median is 48. The summary statistics are quite stable over the four years, which is why we only report pooled moments.

In panel B, we report the summary statistics of Chinese manufacturing firms with R&D activity in the Annual Survey of Manufacturing during the period 2006–2007. Since the National Statistical Office of China stops reporting firm R&D activity after 2007, we mostly use these firms in our descriptive evidence analysis. We have a similar sample size of around $300,000$ each year, although the firms in the ASM sample are noticeably larger than those in the SAT sample. The difference is more pronounced when we look at the lower quartile (i.e. 25%) 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. Interestingly, the firms in the ASM sample do not appear to invest more in R&D despite being larger.

[9]We discuss the details of this procedure in Appendix A.

The fraction of positive R&D firms is slightly higher than 10%, however, R&D intensity ranges from 0.1% to 1.7% at the 25th and 75th percentile in this sample.

3 Descriptive Evidence of Firms' Responses to Tax Notches

In this section, we provide descriptive evidence suggesting that R&D investment by Chinese manufacturing firms is responsive to the fiscal incentives of the InnoCom program, and that part of this response may be due to relabeling. In particular, we document stark bunching patterns precisely above tax notches, and we show that the ratio of administrative expenses to sales drops sharply at the notch.

3.1 Bunching Response

We first analyze data from the post-2008 period as the phasing out of the dual-track system provides for cleaner comparisons across firms. Moreover, 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 of R&D intensity between 1% 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 firm's 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 ranges specified by the InnoCom program. For firms with annual sales less than 50 million RMB in sales, 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 annual sales, we only observe bunching at 3%. These patterns are consistent with the size-dependent tax incentive programs laid out in the InnoCom program. Moreover, these plots allay concerns of potential "round number problems" that might occur if firms report rounded versions of true data and that are present in other bunching studies (e.g., Kleven and Waseem (2013)) as there are no other significant spikes in the data.

Next, we analyze the sample of data from the pre-2008 period, and we report in Figure 3 the empirical distribution of Chinese firms' R&D intensity during 2006–2007. Recall that the tax incentive of the InnoCom was not size-dependent before 2008, and kicked-in uniformly at a 5% R&D intensity level. In addition, our pre-2008 data has information of each firm's employee education based on the Census of Manufacturing conducted in 2004. This allows us to refine our sample to firms with more than 30% college educated workers, consistent with the requirement of InnoCom program. It is reassuring here that we observe the R&D intensity bunching solely at 5%, and no significant spikes at 3%, 4%, and 6%. The contrast of R&D intensity bunching patterns across different time periods provides further evidence that Chinese firms respond actively to the tax notches based on R&D intensity.

Bunching Response to the Tax Reform of 2008

The previous figures look at the cross-sectional distribution of R&D intensity and show a striking pattern of bunching for both pre and post-2008 periods. We now explore some of the variation over time in the

Chinese corporate income tax system described in Section 1.

Consider first the behavior of FOEs in the large category (sales above 200 million RMB) as the incentive to invest in R&D changes dramatically for these firms after 2008. Before 2008, most of the large FOEs benefited from the dual-tax system and faced an EIT rate between 15% to 24%. These firms were not likely to obtain the HTE certification as they saw little to no tax benefits from the InnoCom program. However, when the dual-tax system was phased-out in 2008, the InnoCom program becomes the most important tax incentive program for large FOEs.[10] In Figure 4, we compare the R&D intensity distribution for the large FOEs before and after 2008. To make the two samples comparable, we only use those firms that we were able to match between the SAT and ASM data. The figure illustrates clearly that the changing EIT system has a large impact on firm behavior. Large FOEs have no clear pattern of bunching before 2008, in contrast to DOEs that show a clear bunching at 5% of R&D intensity level. This is consistent with the fact that FOEs already faced very favorable EIT treatment during that period. In contrast, FOEs start behaving like DOEs after 2008. Their R&D intensity distribution starts to show a very distinguishable bunching at the 3% level, which is the exact threshold required for these firms to qualify as HTEs.

We now consider the behavior of "small" (sales below 50 million RMB) DOEs. This is an interesting group of firms since it is the only category that saw an increase in the required R&D intensity threshold from 5% to 6%. Figure 5 shows this adjustment process. Similar to the previous case, we restrict our analysis to those firms that we can match across samples over time. While there is a stable bunching pattern at 5% for years 2006 and 2007, it almost completely disappears in 2008. However, it takes a few additional years for this group of firms to gradually increase their R&D to generate a clear bunching at 6%. This pattern is indicative of adjustment cost or other constraints that a firm needs to overcome when they start to increase R&D investment.

Lack of Sales Manipulation

The stark bunching patterns in these figures raise the concern that firms may manipulate their sales. There are two ways firms may do this. First, if a firm wants to be categorized as a larger firm, they may over-report their sales in order to qualify for a lower R&D intensity threshold. Second, 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.

We analyze both types of misreporting but we note that it is unlikely that firms will manipulate sales. First, firm managers would not want to under-report sales as this is seen as a measure of their performance on the job. Second, China's VAT system with third-party reporting makes it hard for firms to manipulate their revenue since, in the case of over-reporting, they would have to provide evidence for these phantom sales. Nonetheless, there is the possibility that firms that are close to achieving either the size threshold or the intensity threshold may manipulate the timing of their sales by accelerating or slowing-down transactions that are close to the end of the year.

Consider first the case of manipulating sales for the purpose of reaching the R&D intensity threshold. Panel A in Figure 6 plots firms' log sales relative to their R&D intensity. For each group of firms, we report average log sales for small bins of R&D intensity as well as an estimated cubic regression that is allowed to vary below and above each threshold. If firms under-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

[10]Since most of these firms are located in coastal Special Economic Zones or in Economic and Technology Development Zones, the Western Development program usually does not apply.

figure shows that both the data and the estimated polynomial regressions are remarkably stable at each notch. Table A.1 reports estimates of the structural break at the notch and shows that we do not detect evidence of sales manipulation.

We now consider whether firms manipulate their sales to qualify for a lower threshold. Panels B and C in Figure 6 show the histogram of firms around the size thresholds. Since larger firms face lower R&D intensity thresholds, we might expect firms to bunch on the right of the size threshold. These figures show that firms are not responding to the incentives by manipulating their size.[11]

3.2 Detecting Relabeling of R&D Investment

We now explore the degree to which the bunching response may be due to expense mis-reporting. As mentioned above, under Chinese Accounting Standards, R&D is categorized under "Administrative Expenses." For this reason, we look for evidence of evasion by studying the ratio of non-R&D administrative expenses to sales. Figure 7 explores how this ratio is related to R&D intensity, and whether this ratio changes discontinuously at the relevant notches. For each size group, this figure groups firms into bins of R&D intensity and plots the mean non-R&D admin 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 dots are for large sales firms, red for medium sales firms, and blue for small firms. For each size category, there is an obvious discontinuous jump downward at each threshold. Once the firms get further away from the bunching threshold, there is no systemic difference of the admin expense-to-sales ratio for firms with either low or high R&D intensities. This pattern is very consistent with the hypothesis that firms mis-categorize non-R&D expenses into R&D when they get close to the bunching thresholds.[12]

In Table A.2, we report the estimated jump at the notch from the series regression to further quantify the size of the downward jump for each size group. The coefficient of structural break is highly significant for all three groups. The large, medium, and small sales firms reduce their admin expense-to-sales ratio by 1.4%, 1.3%, and 0.8%, respectively. Comparing the drop to the R&D intensity at the notch, we find that $\frac{-\beta^{Evasion}}{\alpha}$ is on average 23.3% for large sales firms, 32.9% for medium sales firms, and 26.9% for small sales firms. As we discuss in Section 5.2, these estimates do not have a causal interpretation; however, they present strong descriptive evidence that firms may respond to the InnoCom program by relabeling non-R&D expenses.

As a robustness check, we 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.1. We also explore the degree to which evasion is related to firm liquidity. In Table A.4, we analyze whether the jump in the non-R&D administrative expense-to-sales ratio is larger for firms with more current assets. This table shows that mis-reporting may be larger for firms with high current asset ratios.[13]

Combined, these figures 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 will focus on measuring the size of the change in R&D invest-

[11] 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.

[12] The existence of different thresholds across size groups also allows us to conduct a set of falsification tests. In particular, we find that when we impose the "wrong" thresholds of the other size groups, there is no observable discontinuity.

[13] Appendix B provides additional analyses suggesting that a fraction of the reported R&D activity may be relabeled by contrasting the effect of reported R&D on TFP above and below the notch.

ment, analyzing the degree to which the response is due to relabeling, and studying how evasion 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 objective of the model is three-fold. First, the model shows that a standard model of firm investment and evasion may produce the patterns described in Section 3.2. 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 causal treatment effects on relabeling and productivity, as in Diamond and Persson (2016). We present estimates of these causal 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.

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 C.

4.1 Model Setup

Consider a firm i with a unit cost function $c(\phi_1, p_t) = c(p_t)\exp\{-\phi_{it}\}$, where p_t is the cost of inputs.[14] ϕ_{it} is log-TFP and which follows the law of motion given by:

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

where $D_{i,t-1}$ is R&D investment, and $u_{i,t} \sim$ i.i.d. $N(0, \sigma^2)$. This setup is consistent with the R&D literature where knowledge capital is depreciated (captured by ρ) and influenced by continuous 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 $\frac{\varepsilon}{1-\rho}$.

We assume the firm faces a constant elasticity demand function: $p_{it} = q_{it}^{-1/\theta}$. This implies that we can write expected profits as follows:

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

R&D Choice Under A 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:

$$\max_{D_1} (1-t_1)(\pi_{i1} - D_{i1}) + \beta(1-t_2)\mathbb{E}[\pi_{i2}].$$

The optimal choice of D_{i1} given by:

$$D_{i1} \quad = \quad \left[\frac{1}{(\theta-1)\varepsilon}\frac{1-t_1}{\beta(1-t_2)}\frac{1}{\mathbb{E}[\pi_{i2}|D_{i1}=0]}\right]^{\frac{1}{(\theta-1)\varepsilon-1}}. \tag{2}$$

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

Notice first that if the tax rate is constant across periods, the corporate income tax does not affect the choice of R&D investment.[15]

This equation shows that the optimal R&D choice has a constant elasticity with respect to the net of tax rate, so that

$$\frac{d\ln D_{i1}}{d\ln(1-t_2)} = \frac{1}{1-(\theta-1)\varepsilon}.$$

In particular, this elasticity suggest that firms that have a higher valuation of R&D (greater value of $(\theta-1)\varepsilon$) will be more responsive to tax incentives.

The choice of R&D depends on potentially-unobserved, firm-specific factors, as they influence $\mathbb{E}[\pi_{i2}|D_{i,t-1}=0]$. An important insight from this analysis is that we can recover these factors from D_{1i} as follows:

$$\mathbb{E}[\pi_{i2}|D_{i1}=0] = \frac{1}{(\theta-1)\varepsilon}\frac{1-t_1}{\beta(1-t_2)}D_{i1}^{1-(\theta-1)\varepsilon}. \tag{3}$$

Substituting into the objective function, we can write the value of the firm as:

$$\Pi(D_{i1}|t_2) = (1-t_1)\left[\pi_1 + D_{i1}\left(\frac{1}{(\theta-1)\varepsilon}-1\right)\right]. \tag{4}$$

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:

$$t_2 = \begin{cases} t_2^{LT} & \text{if } D_1 < \alpha\theta\pi_1 \\ t_2^{HT} & \text{if } D_1 \geq \alpha\theta\pi_1 \end{cases},$$

where sales equal $\theta\pi_1$, $t_2^{LT} > t_2^{HT}$, and where LT/HT stands for low-tech/high-tech. Intuitively, this tax structure induces a notch in the profit function at $D_1 = \alpha\theta\pi_1$, where α is the R&D intensity required to attain the high-tech certification. Figure 8 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 2. 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 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 2 and by $D_1 = \alpha\theta\pi_1$.

Whether the firm finds it optimal to set R&D intensity equal to the notch threshold depends on firm-level conditions that are summarized by $\mathbb{E}[\pi_{i2}|D_{i,t-1}=0]$, 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 $\mathbb{E}[\pi_{i2}|D_{i,t-1}=0]$ is smoothly distributed around the threshold α, this incentive will lead a mass of firms to find $D_1 = \alpha\theta\pi_1$ optimal and thus "bunch" at this level. Our analysis proceeds by studying when firms decide to bunch and by characterizing the R&D intensity of the firm that is marginal between both solutions in terms of the R&D intensity.

Let $\Pi(\alpha\theta\pi_1|t_2^{HT})$ be the value of the firm conditional on bunching at the notch. Using the optimal choice for an internal solution in Equation 3, we manipulate $\Pi(\alpha\theta\pi_1|t_2^{HT})$ by substituting the unobserved

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

components of the firm-decision, i.e. $\mathbb{E}[\pi_{i2}|D_{i1}=0]$:

$$\Pi(\alpha\theta\pi_1|t^{HT}) = (1-t_1)(\pi_1 - \alpha\theta\pi_1) + \frac{(1-t_1)D_{i1}}{(\theta-1)\varepsilon}\left(\frac{\alpha\theta\pi_1}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right).$$

We then obtain the relative profit from bunching by subtracting profits in the first period $(1-t_1)\pi_1$ and by dividing by the after-tax cost of R&D investment from bunching at the notch $(1-t_1)\alpha\theta\pi_1$:

$$\tilde{\Pi}(\alpha\theta\pi_1|t_2^{HT}) = \left(\frac{d}{\alpha}\right)^{1-(\theta-1)\varepsilon} \times \frac{1}{(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) - 1, \tag{5}$$

where we define $d = \frac{D_{i1}}{\theta\pi_1}$ as the R&D intensity of the firm in the internal solution (without bunching). A similar manipulation of Equation 4 results in the relative profit from not bunching:

$$\tilde{\Pi}(D_{i1}|t_2^{LT}) = \frac{d}{\alpha} \times \left(\frac{1}{(\theta-1)\varepsilon} - 1\right). \tag{6}$$

A firm will decide to bunch if the relative profit from doing so (Equation 5) is greater than the relative profit from not bunching (Equation 6). Comparing Equations 5 and 6, we see that Equation 5 shows a larger cost of investment in the first period (since $d < \alpha$) and higher profits in the second period. Profits are higher because of the tax benefit, $\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) > 1$, and because of the productivity effect from the additional investment in R&D, $\left(\frac{\alpha}{d}\right)^{(\theta-1)\varepsilon} > 1$.

Note that when a firm is already close to the notch, *i.e.*, $\frac{d}{\alpha} \approx 1$, it will be very likely to bunch since the tax cut ensures $\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) > 1$ and therefore $\tilde{\Pi}(\alpha\theta\pi_1|t_2^{HT}) > \tilde{\Pi}(D_{i1}|t_2^{LT})$. As d decreases from α, the value of bunching decreases faster than the value of the firm when the firm does not bunch.[16] Let d^{*-} be the R&D intensity of the firm that is indifferent between bunching and not bunching (*i.e.*, $\tilde{\Pi}(\alpha\theta\pi_1|t_2^{HT}) = \tilde{\Pi}(D_{i1}^{*-}|t_2^{LT})$), and note that this firm satisfies:

$$\underbrace{\left(\frac{d^{*-}}{\alpha}\right)^{1-(\theta-1)\varepsilon} \times \frac{1}{(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) - 1}_{\text{Relative Profit Bunching}} = \underbrace{\frac{d^{*-}}{\alpha} \times \left(\frac{1}{(\theta-1)\varepsilon} - 1\right)}_{\text{Relative Profit Not Bunching}}. \tag{7}$$

In this simple model, firms with $d \in [d^{*-}, \alpha]$ would decide to bunch at the notch.

To gain intuition behind Equation 7, note that the decision to bunch is influenced by firm-level conditions that are summarized by $\mathbb{E}[\pi_{i2}|D_{i,t-1}=0]$, as well as on the degree to which R&D investment is valued by firms in terms of future profits (i.e., $\varepsilon(\theta-1)$). Our model uses Equation 3 as a sufficient statistic of firm-level determinants of R&D investments to provide a link between the increase in the investment, d^{*-} to α, the profitability elasticity of R&D to the firm, $(\theta-1)\varepsilon$, and the magnitude of the tax incentive, $\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)$. It can 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

As discussed above, one mechanism driving the large bunching responses we observe might be the manipulation of reported R&D investment. This section extends the model by allowing for firms to misreport

[16] Note that, while Equation 6 is linear in R&D intensity, Equation 5 is concave in d (as long as $(\theta-1)\varepsilon < 1$).

their costs and shift non-RD costs to the R&D category. We show that the bunching predictions from the previous sections remain unaffected. However, the interpretation of the reported bunching response is now a combination of real and relabeled activity. While relabeling obscures the link between bunching and the firms' valuation R&D, we show that we may uncover firms' valuation of R&D in addition to their costs of misreporting by analyzing the model's implications for productivity and relabeling.

Denote a firm's reported level of R&D spending by $\tilde{D}_1$. The expected cost of misreporting to the firm is given by $h(D_1, \tilde{D}_1)$. We assume that the cost of mis-reporting is proportional to the reported R&D, $\tilde{D}_1$, and depends on the percentage of misreported R&D, $\delta = \frac{\tilde{D}_1 - D_1}{\tilde{D}_1}$, so that:

$$h(D_1, \tilde{D}_1) = \tilde{D}_1 \tilde{h}(\delta).$$

We also assume that $\tilde{h}$ satisfies $\tilde{h}(0) = 0$ and $\tilde{h}'(\cdot) \geq 0$. Finally, define $\Pi(D_1, \tilde{D}_1|t)$ as the value function of a firm's inter-temporal maximization problem when the firm faces tax t in period 2, invests D_1 on R&D, and declares investment of $\tilde{D}_1$.

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:[17]

$$\underbrace{\left(\frac{d}{\alpha(1-\delta^*)}\right)^{1-(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)}_{\text{Productivity Benefit from Reducing Evasion}} = \underbrace{\frac{((1-t_1) - h'(\delta^*))}{\alpha(1-t_1)}}_{\text{Investment Cost and Reduction in Evasion Cost}}. \tag{8}$$

The optimal fraction of misreported R&D, δ^*, depends on the firm's R&D intensity if it did not bunch, d. Since the firm is bunching, the R&D reported by the firm is $\alpha\theta\pi_1$. The firm increases its real R&D investment to $D^{*K} = (1-\delta^*)\alpha\theta\pi_1$, which is such that $\alpha\theta\pi_1 \geq D^{*K} \geq D_{i1}$. As the firm increases the fraction of relabeled R&D, δ, the productivity benefits and the costs of investment decrease, but there is also an increase in the cost of evasion. Note that δ^* is increasing in the tax advantage $\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)$ and decreasing in both the valuation of R&D, ε, and the marginal cost of evasion.

The firm decides to bunch if the profits from the optimal relabeling strategy $\Pi(\alpha\theta\pi_1, D_1^{*K}|t_2^{HT})$ are greater than the profits from not bunching $\Pi(D_{i1}, D_{i1}|t_2^{LT})$. Notice that the relative profits from not bunching are still characterized by Equation 6. As in the case without evasion, we obtain an expression for the relative profits from bunching as follows:

$$\underbrace{\left(\frac{d}{\alpha(1-\delta^*)}\right)^{1-(\theta-1)\varepsilon} \times \frac{(1-\delta^*)}{(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) - (1-\delta^*)}_{\text{Relative Profit from Bunching}} - \underbrace{\frac{h(\delta^*)}{\alpha(1-t_1)}}_{\text{Evasion Cost}} \tag{9}$$

Equations 5 and 9 are very similar and are identical in the case when $\delta^* = 0$, such that there is no evasion. When $\delta^* > 0$, the cost of investment is smaller, the productivity gains are also smaller, and the firm also incurs a cost of evasion.

To understand the implications for bunching from this equation, consider the firm that is indifferent between bunching and potentially misreporting, and not bunching. Panel (c) of Figure 8 shows that if

[17] Appendix C shows the detailed derivation and shows that this results is robust to including fixed and adjustment costs.

the firm is willing to misreport R&D in order to reach the notch, it will have a lower internal solution than if evasion were not possible. The indifference condition of the marginal firm is now:

$$\underbrace{\left(\frac{d^{*-}}{\alpha(1-\delta^{*-})}\right)^{1-(\theta-1)\varepsilon} \times \frac{(1-\delta^{*-})}{(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) - (1-\delta^{*-})}_{\text{Relative Profit from Bunching}} - \underbrace{\frac{h(\delta^{*-})}{\alpha(1-t_1)}}_{\text{Evasion Cost}}$$

$$= \underbrace{\frac{d^{*-}}{\alpha} \times \left(\frac{1}{(\theta-1)\varepsilon} - 1\right)}_{\text{Relative Profit from Not Bunching}} . \tag{10}$$

Note that since the firm can always elect to report truthfully ($\delta = 0$), the relative profit from bunching in the case with evasion is greater than in the case without evasion. 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. That is, the marginal firm with relabeling will have a lower threshold d^{*-}, which implies that we should see more bunching when firms are able to misreport R&D. However, this also implies 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 Costs

Our model provides a link between firms' valuation of R&D and the patterns described in Section 3. However, the simple model in the previous section predicts bunching patterns that are counterfactual to what we observe in the data. First, as in 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. 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 1, 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 new products, or due to compliance and registration costs. Finally, the literature on R&D investment suggests that R&D is subject to adjustment costs. If this were the case and adjustment costs limited firms responses, using Equation 7 to link the bunching response to $(\theta-1)\varepsilon$ would result in a downwardly biased value of $(\theta-1)\varepsilon$.

We thus augment our model to allow for firms to face adjustment costs of investment and fixed costs of certification. 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.

We also assume that firms pay a fixed costs of certification that is given by: $c \times \alpha\theta\pi_{1i}$. 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, in addition to the costs of complying with the InnoCom program.

Appendix C shows that for given values of (b, c), we obtain a similar result to Equation 10, which links the R&D intensity of the marginal firm to the effect of R&D on profitability. In this case, however, the marginal firm depends on the values of (b, c), which we denote $d^{-}_{b,c}$. As expected, we find that $d^{-}_{b,c}$ is increasing (smaller response) with both adjustment, b, and fixed, c, costs. We also allow for firms of similar pre-existing productivity to have heterogeneous adjustment and fixed costs. For this reason, we now redefine $d^{*-} = \min_{b,c} d^{-}_{b,c}$ as the smallest R&D intensity for which there is a marginal firm.

Our augmented model results in a reasonable distribution of R&D intensity in the case without a notch, does not predict a "hole" in the distribution near the notch, and allows for firms with similar productivity levels to engage in different patterns of investment depending on their fixed and adjustment costs. As we show in the following sections, the model also allows us to link the bunching response to the increase in R&D and the parameters governing firms' valuation for R&D and costs of evasions in a manner that is robust to the presence of adjustment costs.

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 9 provides the intuition for this procedure. Panel (a) provides a counterfactual distribution of R&D intensity under a linear tax. Denote this counterfactual density by $h_0(\cdot)$. 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 levels 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, which generates the bunching patterns. Define the missing mass in the range $[d^{*-}, \alpha]$, relative to the counterfactual distribution, as B.

The prediction in Panel (a) of Figure 9 is quite stark in that no firms are expected to locate in the dominated interval. As discussed above, 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 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

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

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.[18]

Panel (b) of Figure 9 describes graphically how allowing for this degree of heterogeneity, in addition to frictions, affects the predicted bunching pattern. In particular, the area B can now be computed as follows:

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

To understand the empirical content underlying this bunching prediction, it can be shown that the percentage increase in R&D intensity for firms that may potentially respond to the incentive by

[18]We view this formulation as a micro-foundation for the constraints discussed in Kleven and Waseem (2013).

bunching can be expressed as a function of the missing mass B and the counterfactual density at the notch:[19]

$$\Delta d \equiv \frac{\mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})] - \mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]}{\mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]} \approx \frac{B}{2\alpha h_0(\alpha)}, \tag{11}$$

where $d^{*+} > \alpha$ is chosen to capture the extent of bunching.

As in Kleven and Waseem (2013), we can also 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:

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

4.5 Model Implications for Evasion 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. We formalize these predictions by linking our model to the estimator for treatment effects proposed by Diamond and Persson (2016). As in the case of the average increase in the R&D of Equation 11, we study the average effect on a given outcome Y over the region (d^{*-}, d^{*+}):

$$\mathbb{E}[Y|\text{Notch}, d \in (d^{*-}, d^{*+})] - \mathbb{E}[Y|\text{No Notch}, d \in (d^{*-}, d^{*+})] = \int_{d^{*-}}^{d^{*+}} Y h_1(d) dd - \int_{d^{*-}}^{d^{*+}} Y h_0(d) dd. \tag{12}$$

The first thing to notice about this quantity is that $\mathbb{E}[Y|\text{Notch}, d \in (d^{*-}, d^{*+})]$ is directly observed in the data. In Section 5.2 we discuss the econometric approach to estimating $\mathbb{E}[Y|\text{No Notch}, d \in (d^{*-}, d^{*+})]$.

To interpret this treatment effect note that the region (d^{*-}, d^{*+}) includes firms that do not respond to the program, as well as firms whose R&D intensity is already above the notch. Conceptually, we can partition the firms in the region (d^{*-}, d^{*+}) into compliers, never-takers, and always-takers. In our setting, the never-taker firms are firms below the notch that are constrained from responding to the policy. The always-taker firms are firms that are already above the notch. By assuming that there are no defier firms, we can show that Equation 12 has the interpretation of an intent-to-treat, and that this effect is identified by the behavior of complier firms that respond to the incentives of the program:[20]

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

To see that this equation represents the behavior of the compliers, note that the first integral evaluates the average value of Y for firms that were previously below the notch, denoted by $\mathbb{I}[d_0 \in (d^{*-}, \alpha)]$, and that were not constrained in their response. The second integral compares this value to the average value for the same complier firms under the counterfactual scenario where there is no notch.

Our model for the evolution of TFP predicts a tight connection between the ITT for productivity and the ITT for R&D. To see this, note that for a given firm we would expect to observe:

$$\phi_2^1 - \phi_2^0 = \rho(\phi_1^1 - \phi_1^0) + \varepsilon(\ln d_1^1 - \ln d_1^0) + (u_1^1 - u_1^0),$$

[19] Appendix D contains details of these approximations. Note that in practice we may compute the left-hand-side of this equation without an approximation by evaluating the expectations using the estimated counterparts of $h_0(d)$ and $h_1(d)$. Note that the approximation of ΔD^* relies on the assumption that $\mathbb{P}r(\text{Constrained})$ does not depend on d.

[20] In our setting, defier firms are those that would be above the notch without the InnoCom program and below the notch in the presence of the InnoCom program. Appendix D provides details of this derivation.

where the superscript 1 corresponds to the notch and 0 corresponds to the no-notch case, and where subscripts denote time periods. Averaging over the firms in the excluded region we find:[21]

$$ITT^{\phi_2} = \varepsilon ITT^{\ln d_1}.$$

If complier firms respond to the InnoCom program by relabeling administrative expenses as R&D expenses this relation is adjusted by replacing $ITT^{\ln d_1}$ with the ITT on real investment. In this case, our model also predicts a negative ITT on administrative expenses that is informative of firms' cost of evasion. Section 6 discusses how we link estimated treatment effects to structural parameters even in the case where these relations might not admit a closed-form expression.

5 Effects on Investment, Relabeling, 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.

5.1 Bunching Estimates of Investment Response

We now describe how we estimate $h_0(\cdot)$ to recover the empirical quantities B and $h_0(\alpha)$. We follow the literature (see, e.g., Kleven (2015)) by estimating a flexible polynomial on a subset of data that excludes the area around the threshold, and by using the fitted polynomial on the excluded region as an estimate of $h_0(\cdot)$. 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 level $d_j = \frac{D_{j1}}{\theta\pi_1}$, and where (d^{*-}, d^{*+}) is the region excluded in the estimation. Given the monotonically decreasing shape of the R&D intensity, we restrict the estimated β_k's to result in a decreasing density.

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 a counterfactual estimate for $h_0(\alpha)$ and B as follows:

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

Finally, an estimate of the fraction of constrained firms relative to the counterfactual density is given by:

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

where α^- is the value of R&D such that a firm would be willing to jump to the notch even if R&D had no effects on productivity.[22]

[21] Note that $\mathbb{E}[\phi_1^1 - \phi_1^0] = \mathbb{E}[u_1^1 - u_1^0] = 0$ by construction.

[22] This "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 $(\alpha^-, \alpha) = (2.3\%, 3\%)$.

Implementing the bunching estimator requires choosing the degree of the polynomial, and selecting the excluded region. 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. For each value, 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.[23] Finally, we obtain standard errors by bootstrapping the residuals from the series regression, generating 5000 replicates of the data, and re-estimating the parameters.

Figures 10-11 display the results of the bunching estimator for the three different notches for 2009 and 2011. The red line 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 $h_0(\cdot)$. Each of these graphs also reports the percentage increase in R&D intensity for complier firms, $\Delta d/(1-a^*)$, the fraction of firms that are constrained below the notch point, $a^*(\alpha^-)$, and the p-value of the test that the missing mass and the excess mass are of the same magnitude.

Panel (a) of Figure 10 shows an increase in R&D intensity of 19% for small firms in 2009. This estimate corresponds to the response of complier firms that are not otherwise constrained in their ability to respond to the incentives of the InnoCom program. The specification test shows that using the missing mass or the excess mass results in statistically indistinguishable estimates. We also find that 74% of the firms are not able to respond to the incentive. 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.

Panels (b) and (c) show the same set of results for medium and large firms in 2009. We find similar increases in R&D intensity of 49% and 35%, respectively. In both cases, using the missing mass and the excess mass results in statistically indistinguishable estimates of the increases in R&D. The estimated fraction of firms that face constraints to respond to the program is now 66% and 57%, respectively. When we analyze these firms, we find that most of these firms have low profitability, or are already benefitting from other tax credits. Both of these features would lower the incentive to be certified by the InnoCom program. Figure 11 shows similar qualitative patterns for 2011, where we find that the fraction of constrained firms is now smaller in all cases, and the average increase in R&D is greater. It is worth noting that these effects are estimated with a high degree of precision as standard errors are often an order of magnitude smaller than the estimates.

Table 2 provides further details behind these statistics. The first column of the table reports Δd, the percentage increase in R&D intensity for all of the firms in the excluded region. This statistic is always smaller than when we adjust for the fraction $a^*(\alpha^-)$ of firms that are constrained from responding to the policy. Note that this statistic measures the same quantity as the ITT estimator for log-R&D intensity ($\ln d$) in the framework of Section 4.5 but relies on the approximation in Equation 11. For large firms

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

in 2009, we see an increase in R&D intensity of 14.9% among all firms in the excluded region.

Column (4) reports the percentage increase in R&D intensity relative to the notch for the marginal buncher. This effect represents the largest possible response for complier firms. While understanding the behavior of firms of different sizes is interesting from an economic perspective, policy makers may be interested in the aggregate increase in R&D across the economy. Figure A.2 shows that the vast majority of R&D is conducted by firms in the large sales category, so it makes sense to focus on these firms when when mapping these estimates to the patterns in Figure 1. In Column (5) we report the level increase in R&D intensity by multiplying ΔD^* by $\alpha(1 - a^*(\alpha^-))$. This column shows that, for large firms in 2011, the marginal bunchers increased their R&D intensity by 1.9 percentage points, which would indeed contribute to the aggregate growth in R&D intensity in Figure 1.

We now explore the robustness of our estimates. First, we show in Panel (a) of Figure 12 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 firms in 2008, which were not subject to the incentives of the InnoCom program, and finds a small and insignificant estimate of $\Delta d/(1-a^*)$. Second, we explore the potential for firms' extensive margin responses to bias our estimate. 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 $\Delta d/(1 - a^*)$. In Panel (b) of Figure 12 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, if slightly smaller, estimate of $\Delta d/(1 - a^*)$ when we rule out extensive margin responses. Finally, one potential concern with estimating the counterfactual density using a polynomial regression is that a large excluded region may result in a biased estimate of the counterfactual density. We assuage this concern by using data from large firms in 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. 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^{*+}.[24] Panel (c) of Figure 12 shows that using these data to inform the shape of the counterfactual density in the excluded region results in very similar estimates of both the counterfactual density and $\Delta d/(1 - a^*)$.

5.2 ITT Estimates on Productivity, Relabeling, and Tax Collections

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_{i,t_2}, the estimator is:

$$\begin{aligned} \widehat{ITT}^{Y_{t_2}} &= \mathbb{E}[Y_{t_2}|\text{Notch}, d_{t_1} \in (d^{*-}_{t_1}, d^{*+}_{t_1})] - \mathbb{E}[Y_{t_2}|\text{No Notch}, d_{t_1} \in (d^{*-}_{t_1}, d^{*+}_{t_1})] \\ &= \frac{1}{N^{Excluded}} \sum_{d_{i,t_1} \in (d^{*-}, d^{*+})} Y_{i,t_2} - \int_{d^{*-}}^{d^{*+}} \hat{h}_0(d_{t_1}) E[Y_{i,t_2}|\widehat{d_{t_1}, \text{No}} \text{ Notch}] dd_{t_1}. \end{aligned} \tag{13}$$

The first quantity is the observed average value of a given outcome Y_{i,t_2} over the excluded region. The second quantity is a counterfactual average value of Y_{i,t_2}, which is constructed by combining the

[24]Let $h_0^{2008}(d)$ denote the un-manipulated density of large firms in 2008. We combine $h_1(d)$ with $\tilde{h}_0^{2008}(d) = \alpha + \beta h_0^{2008}(d)$, where $\beta = \frac{h_1(d^{*+}) - h_1(d^{*-})}{h_0^{2008}(d^{*+}) - h_0^{2008}(d^{*-})}$ and $\alpha = h_1(d^{*-}) - \beta h_0^{2008}(d^{*-})$. As discussed in Blomquist and Newey (2017), variation in non-linear incentives can help in identifying responses when using bunching approaches.

counterfactual 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.

Since the estimator compares averages over the excluded region, which includes compliers and non-compliers, we interpret it as an intent-to-treat (ITT). As an 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 11. One way to think of this counterfactual is from the point of view of the law of iterated expectations. As the quantity $E[Y|\widehat{d_{t_1}, \text{No}}\,\text{Notch}]$ recovers the average value of a given outcome had there been no notch, the integral simply averages this function of d_{t_1} over the excluded region with respect to the counterfactual density of R&D, $\hat{h}_0(d_{t_1})$. Taking ratios of these estimates produce Wald estimates of treatment effects.

In order to implement this estimator, we estimate $E[Y_{i,t_2}|\widehat{d_{t_1}, \text{No}}\,\text{Notch}]$ as a flexible polynomial regression of Y_{i,t_2} on R&D intensity over the same excluded region used to estimate $\hat{h}_0(\cdot)$:[25]

$$Y_{i,t_2} = \underbrace{\sum_{k=0}^{p} \beta_k \cdot (d_{i,t_1})^k}_{E[Y_{t_2}|d_{t_1}=d,\text{No Notch}]} + \gamma \cdot \mathbf{1}\left[d^{-*} \leq d_{i,t_1} \leq d^{+*}\right] + \delta Y_{i,t_1} + \phi_s + \nu_i.$$

Figure 13 presents a visual example for the case of administrative costs, where we estimate a cubic regression of the admin expense to sales ratio on R&D intensity in 2009, and where the excluded region corresponds to Panel (c) of Figure 10. As in Figure 7, we observe a significant drop in the ratio after the notch that is likely due to relabeling of expenses to qualify for the InnoCom program. 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 prevents the econometrician from using data just beneath the threshold as a control group for firms above the threshold.

In contrast, our procedure does not rely on such comparisons across firms, but instead relies on the assumption that $E[Y_{i,t_2}|d_{t_1}, \text{No Notch}]$ 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. As shown by Figure 13, this flexible polynomial fits the data outside of the region very well. Moreover, we observe from Figure 7 that small- and medium-sized firms have smooth and flat relations between administrative expenses and R&D intensity around the 3% level, which suggests that our estimate of $E[Y_{i,t_2}|d_{t_1}, \text{No Notch}]$ represents a valid counterfactual. Armed with an estimate of $E[Y_{i,t_2}|d_{t_1}, \text{No Notch}]$, we then compute an average value for firms in the excluded region by combining this estimate with an estimate of the counterfactual density, which in this case corresponds to Panel (c) of Figure 10. The resulting ITT estimate in Equation 13 thus compares the observed average outcome over the excluded region, to a counterfactual average over the same region.[26] We obtain standard errors for ITT estimates in Equation 13 by bootstrapping this procedure.

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 al R&D investment (see Figure A.2), 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. We also find a decrease in the administrative cost ratio of 9.6%. When compared with the average value of the ratio, we find that administrative costs decreased

[25]Note that this regression is not causal. Its role is purely to predict the outcome over the excluded region.

[26]In particular, this estimate does not rely on comparisons of firms that are close to the notch, as in the case of a regression discontinuity.

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 increase in R&D intensity of 0.89% (column (5) in Table 2), this would imply that $\left(\frac{0.33\%}{0.89\%} \approx\right)$ 37% of the increase in R&D intensity was due to relabeling. 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%.[27]

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%. We find a comparable effect on the profit ratio of 0.9%. Finally, we observe an overall decrease in corporate tax revenues of 12.8%.

The second panel of Table 3 presents ratios of these estimates along with bootsrapped 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. Comparing the R&D increase with the growth in the profit ratio, we find that doubling R&D would lead to an increase in the profit ratio of 5.9%.[28] We find slightly larger effects on TFP, with an increase of 8.4%. In order to relate this estimate to our parameter ε, note that this effect reflects the cumulative effect over two time periods, which should be discounted by $1+\rho \approx 1.725$. These reduced-form estimates are thus consistent with a value of $\varepsilon \approx 0.049 \left(\approx \frac{0.084}{1.725}\right)$. However, if we take into account that about 37% of the reported R&D response is due to relabeling, we would obtain an estimate of $\varepsilon \approx 0.077 \left(\approx \frac{0.084}{(1.75)(1-0.37)}\right)$. Given the strong assumptions we need in order to calculate the fraction of relabeled R&D, we rely on the structural model in Section 6 to generate our main estimate of ε. However, it is worth pointing out that estimates of the effects of R&D on TFP that do not account for relabeling are likely downwardly-biased. Finally, we consider how much it costs the government to increase R&D investment in terms of foregone revenue. We find that doubling R&D investment would cost the government about 88% of corporate tax revenue. This estimate is also downwardly-biased to the extent that the R&D response is due to relabeling. These estimates are crucial ingredients for deciding whether the InnoCom policy is too expensive, or whether externalities from R&D investment merit further subsidies. In the next section we refine these estimates by accounting for relabeling of R&D, and by analyzing how changes in the InnoCom program would affect the government's cost of incentivizing R&D investment.

6 Structural Estimation and Simulation of Counterfactual Policies

While the causal estimates discussed in the previous section describe the effects of the current policy, the evaluation of alternative policies requires a model of firm selection into the policy, as well as how investment and relabeling decisions affect productivity.

[27] 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).

[28] The interpretation of this ratio deserves caution as it represents the effects of increasing R&D as well as other effects of the InnoCom program, such as the tax cut. See Jones (2015) for a useful exposition of the economics of such restrictions.

6.1 Structural Estimation

This section proposes a method of simulated moments (MSM) framework to estimate the structural parameters of the model in Section 4 by matching the estimated 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).[29] We then calibrate the distribution of underlying productivity ϕ_1. We use the fact that the evolution of productivity in Equation 1 is an AR(1) process that implies a stationary normal distribution with persistence ρ and variance σ^2. Given a value of $\theta = 5$, 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 $\sigma = 0.385$. Finally, we use the following functional form for the cost of evasion $\frac{\exp\{\eta\delta\}-1}{\eta}$ (see, e.g., Notowidigdo (2013)). Note that this function may be linear, convex, or concave depending on the value of η.

We then use the method of simulated moments to estimate the structural model parameters including productivity effect of R&D, ε, the cost of evasion, η, and distributions of the adjustment costs, b, and fixed costs, c. We assume that b and c are distributed *i.i.d.* across firms, that b is log-normally distributed, so that $b \sim \mathcal{LN}(\mu_b, \sigma_b^2)$, and that c has an exponential distribution, so that $c \sim \mathcal{EXP}(\mu_c)$. In summary, our simulated sample will discipline the set of parameters $\Omega = \{\varepsilon, \eta, \mu_b, \sigma_b, \mu_c\}$.

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

$$Q(\Omega) = \left[\begin{array}{c} h^B(\Omega) \\ h^{ITT}(\Omega) \end{array} \right]' W \left[\begin{array}{c} h^B(\Omega) \\ h^{ITT}(\Omega) \end{array} \right],$$

where W is a weighting matrix. $h^B(\Omega)$ and $h^{ITT}(\Omega)$ are moment conditions that are related to 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. In other words, 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.

We use the treatment effects on reported R&D, admin 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 generate the simulated model counterpart of our ITT estimates to construct moments of the form:[30]

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

where $\widehat{ITT^Y}$ is an estimate from Section 5.2. As a simple example, consider the case where Y is next-period productivity. Our model predicts that $E[\phi_2(\omega;\text{Notch}) - \phi_2(\omega;\text{No Notch})] = \varepsilon[\ln d^{\omega,\text{Notch}} - \ln d^{\omega,\text{No Notch}}]$. This shows that the estimated effects on firm productivity will inform the values of ε. While there is no closed-form expression for the fraction of relabeled R&D, we can form a similar moment to match the estimated effect on the admin expense ratio, which will inform η, the cost of evasion.

[29] 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.

[30] 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.

6.1.1 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 and 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 (and the evasion cost parameter η) plays an important role in influencing these ITT moments. For this reason, the ITT estimate on the admin 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 bunches (in the excluded region) and the ITT on reported R&D inform parameter of the distribution of fixed costs of certification: μ_c. Finally, the location of d^{*-} is jointly determined by all the parameters.

6.1.2 Estimates

Table 4 reports estimates of our structural parameters: $(\varepsilon, \eta, \mu_b, \sigma_b, \mu_c)$. We follow Chernozhukov and Hong (2003) and use a Laplace-type estimator that is based on Markov Chain Monte Carlo (MCMC) to estimate our model. 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.

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%. Since the InnoCom program requires that firms commit to a three-year increase in R&D, the interpretation of this coefficient is closer to a medium-run effect. The estimated evasion cost parameter is 5.663, which indicates that, at the margin, the cost of evasion 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 the benefits of relabeling for a firm come both in the form of lower costs of investment, as well as lower adjustment costs, which include costly technological opportunity constraints. Thus, firms who face a higher shadow cost of R&D (i.e. higher b) will be more willing to relabel R&D. On average, we calculate that firms' realized evasion cost is 4.7% of the 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. However, given the fact that we attribute all of the reduction in administrative expense to relabeling, the model prediction can be interpreted as a lower bound on the productivity gains from real R&D effort.

Finally, we evaluate the sensitivity of our point estimates to each individual moment using the methods of Andrews et al. (2017). We calculate the local derivative of our estimated parameters with respect to each moment. 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, if we double the TFP of firms above d^{*+}, ε would increase substantially by close to 0.05. These methods also allow us to consider the potential that part of the reduction in administrative expenses is not due to evasion.[31] If half of the decrease in admin 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 in Figure A.3.

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.

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 focus on policies that are reasonably close to the form of the InnoCom program. We maintain the structure of an average corporate income tax cut when firm R&D intensity is above a certain threshold, and we vary the location of the threshold to explore differences in both firm-level and aggregate responses.

We take our structural estimates and the existing policy as a benchmark (where $\alpha = 3$ and $t_2^{HT} = 15\%$), and explore how different combinations of α and t_2^{HT} change the excluded region, how firms select into the policy by obtaining the InnoCom certification, and how firms adjust their real R&D investment, as well as their relabeling behavior. These individual-level responses help us further digest the effects on aggregate outcomes like tax revenue losses and productivity growth.

Figure 14 studies the effects of changing the preferential tax rate for three values of the notch: 2%, 3%, and 6%. The first two panels analyze how the characteristics of the compliers depend on the policy parameters. We find that higher values for 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. This indicates that the firms that comply with the new policy have a much worse set of technological opportunities. 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.

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

Panel (c) shows that, for every level of the notch, there is more real R&D investment when firms benefit from larger tax breaks. When we move from a small tax break of 22% towards a large one of 10%, the increase of the real R&D for compliers increase by around 10 percent for every level of the notch. This magnitude is broadly consistent with the user cost elasticity that we estimated in our reduced-form analysis. On the other hand, the fraction of the total response that is due to evasion is also increasing in the size of the tax break. As panel (d) illustrates, when we set the notch threshold at 0.06, moving the preferential tax rate from 22% to 10% increases the fraction of reported R&D due to relabeling by almost 15 percentage points. The increase is slightly less pronounced when the threshold is lower. As we discuss above, firms' evasion motives are largely driven by their heterogeneous technology opportunity b. When the notch threshold is harder to achieve, the selection of high adjustment cost firms into the program is increasingly driven by evasion. Panel (e) then plots the average growth in productivity induced by these R&D incentives for firms that may potentially respond to the policy (in the excluded region). This effect is a combination of the effect on real R&D that we documented in panel (c), as well as the effect on the fraction of compliers, which is larger when there is a lower preferential tax. We see that when 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 ($\alpha = 0.03$) where firms see a 0.8% increase in TFP, and is due both because there is an increase in the fraction of compliers, as well as an increase in real R&D investment.

Finally, panel (f) plots the ratio of the change in taxes to the change in total real R&D investment. This ratio represents the average cost to the government of increasing real R&D investment. We compute this ratio for different values of α and t^{HT} and plot these combinations according to the tax-to-R&D ratio and the total increases in real R&D. This graph thus represents cost frontiers for a government that wants to increase R&D by a given amount. The current policy of $\alpha = .03$ and $t^{HT} = 0.15$ corresponds to a cost-ratio of about 2.3. This ratio is much greater than that reported in Table 3, 0.88, since this accounts for heterogeneous relabeling of R&D. The black line shows that a policy defined by $\alpha = .06$ and $t^{HT} = 0.15$ would result in a similar increase in real R&D investment, but at a lower average cost. Alternatively, a policy defined by $\alpha = .06$ and a larger tax advantage $t^{HT} = 0.12$ would result in a larger increase in R&D investment for a similar tax-to-R&D ratio. However, as shown in Panel (d), this policy would also be accompanied by more evasion. These graphs show how firm selection into the program depends on different policy choices that result in non-trivial tradeoffs between encouraging R&D investment at the lowest cost to taxpayers, introducing misallocation across firms with different adjustment costs, and incentivizing relabeling activities.

7 Conclusions

Governments around the world devote considerable tax resources to incentivize 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. Despite the relabeling responses, we find significant effects on firm-level productivity and profitability 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 the 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-firm productivity, our estimates provide a bound on the size of the externality that would justify government intervention.

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

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

Almus, Matthias and Dirk Czarnitzki, "The effects of public R&D subsidies on firms' innovation activities: the case of Eastern Germany," *Journal of Business & Economic Statistics*, 2003, *21* (2), 226–236.

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.

Berger, Philip G, "Explicit and implicit tax effects of the R & D tax credit," *Journal of Accounting Research*, 1993, pp. 131–171.

Bérubé, Charles and Pierre Mohnen, "Are firms that receive R&D subsidies more innovative?," *Canadian Journal of Economics/Revue canadienne d'économique*, 2009, *42* (1), 206–225.

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, Mark Schankerman, and John Van Reenen, "Identifying technology spillovers and product market rivalry," *Econometrica*, 2013, *81* (4), 1347–1393.

_ , Mark Shankerman, and John Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, July 2013, *81* (4).

Bøler, Esther Ann, Andreas Moxnes, and Karen Helene Ulltveit-Moe, "R&d, international sourcing, and the joint impact on firm performance," *The American Economic Review*, 2015, *105* (12), 3704–3739.

Bronzini, Raffaello and Eleonora Iachini, "Are Incentives for R&D Effective? Evidence from a Regression Discontinuity Approach," *American Economic Journal: Economic Policy*, November 2014, *6* (4), 100–134.

_ **and** _ , "Are incentives for R&D effective? Evidence from a regression discontinuity approach," *American Economic Journal: Economic Policy*, 2014, *6* (4), 100–134.

Busom, Isabel, "An empirical evaluation of the effects of R&D subsidies," *Economics of innovation and new technology*, 2000, *9* (2), 111–148.

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

Clausen, Tommy H, "Do subsidies have positive impacts on R&D and innovation activities at the firm level?," *Structural Change and Economic Dynamics*, 2009, *20* (4), 239–253.

Czarnitzki, Dirk and Georg Licht, "Additionality of public R&D grants in a transition economy," *Economics of Transition*, 2006, *14* (1), 101–131.

_ **, Petr Hanel, and Julio Miguel Rosa**, "Evaluating the impact of R&D tax credits on innovation: A microeconometric study on Canadian firms," *Research Policy*, 2011, *40* (2), 217–229.

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.

Ding, Xuedong and Jun Li, *Incentives for Innovation in China : Building an Innovative Economy*, Taylor and Francis, 2015.

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.

_ **, Martin Antoine, Nguyen Ralf et al.**, "Do tax incentives for research increase firm innovation? An RD Design for R&D," Technical Report 2016.

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.

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

González, Xulia and Consuelo Pazó, "Do public subsidies stimulate private R&D spending?," *Research Policy*, 2008, *37* (3), 371–389.

_ , Jordi Jaumandreu, and Consuelo Pazó, "Barriers to innovation and subsidy effectiveness," *RAND Journal of Economics*, 2005, pp. 930–950.

Görg, Holger and Eric Strobl, "The effect of R&D subsidies on private R&D," *Economica*, 2007, *74* (294), 215–234.

Griffth, Rachel, Stephen Redding, and John Reenen, "Measuring the Cost-Effectiveness of an R&D Tax Credit for the UK," *Fiscal Studies*, 2001, *22* (3), 375–399.

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

_ et al., "Will the real R&D employees please stand up? Effects of tax breaks on firm level outcomes," Technical Report 2016.

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.

_ and Marta Wosinska, "Effectiveness of the California R&D Tax Credit," 1999.

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.

Hilber, Christian A. L. and Teemu Lyytikäinen, "Housing Transfer Taxes and Household Mobility: Distortion on the Housing or Labour Market?," Technical Report, Government Institute for Economic Research VATT Working Papers 47/2013 2013.

Hussinger, Katrin, "R&D and subsidies at the firm level: An application of parametric and semiparametric two-step selection models," *Journal of applied econometrics*, 2008, *23* (6), 729–747.

Jaffe, Adam B and Trinh Le, "The impact of R&D subsidy on innovation: a study of New Zealand firms," Technical Report, National Bureau of Economic Research 2015.

Jones, Damon, "The Economics of Exclusion Restrictions in IV Models," Working Paper 21391, National Bureau of Economic Research July 2015.

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.

Lach, Saul, "Do R&D subsidies stimulate or displace private R&D? Evidence from Israel," *The journal of industrial economics*, 2002, *50* (4), 369–390.

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

Lokshin, Boris and Pierre Mohnen, "How effective are level-based R&D tax credits? Evidence from the Netherlands," *Applied Economics*, 2012, *44* (12), 1527–1538.

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

Mulkay, Benoît and Jacques Mairesse, "The R&D tax credit in France: assessment and ex ante evaluation of the 2008 reform," *Oxford Economic Papers*, 2013, p. gpt019.

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.

Pomeranz, Dina, "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, August 2015, *105* (8), 2539–69.

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.

_ , "Do tax credits stimulate R&D spending? The effect of the R&D tax credit in its first decade," *Journal of Public Economics*, 2016.

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

Takalo, Tuomas, Tanja Tanayama, and Otto Toivanen, "Estimating the benefits of targeted R&D subsidies," *Review of Economics and Statistics*, 2013, *95* (1), 255–272.

Wallsten, Scott J, "The effects of government-industry R&D programs on private R&D: the case of the Small Business Innovation Research program," *The RAND Journal of Economics*, 2000, pp. 82–100.

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

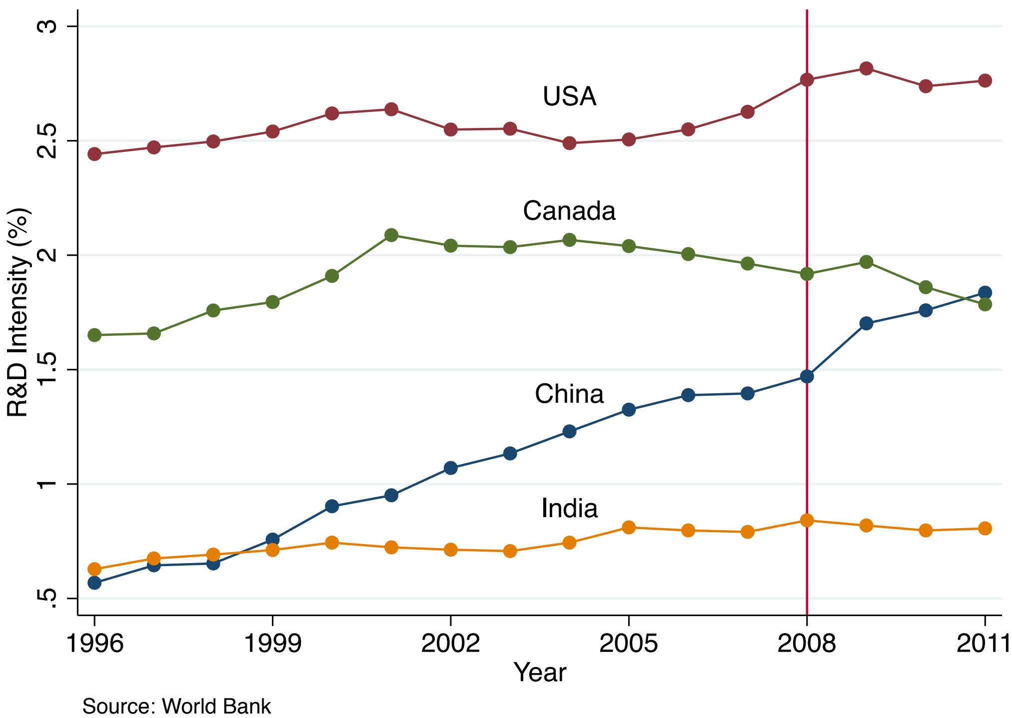
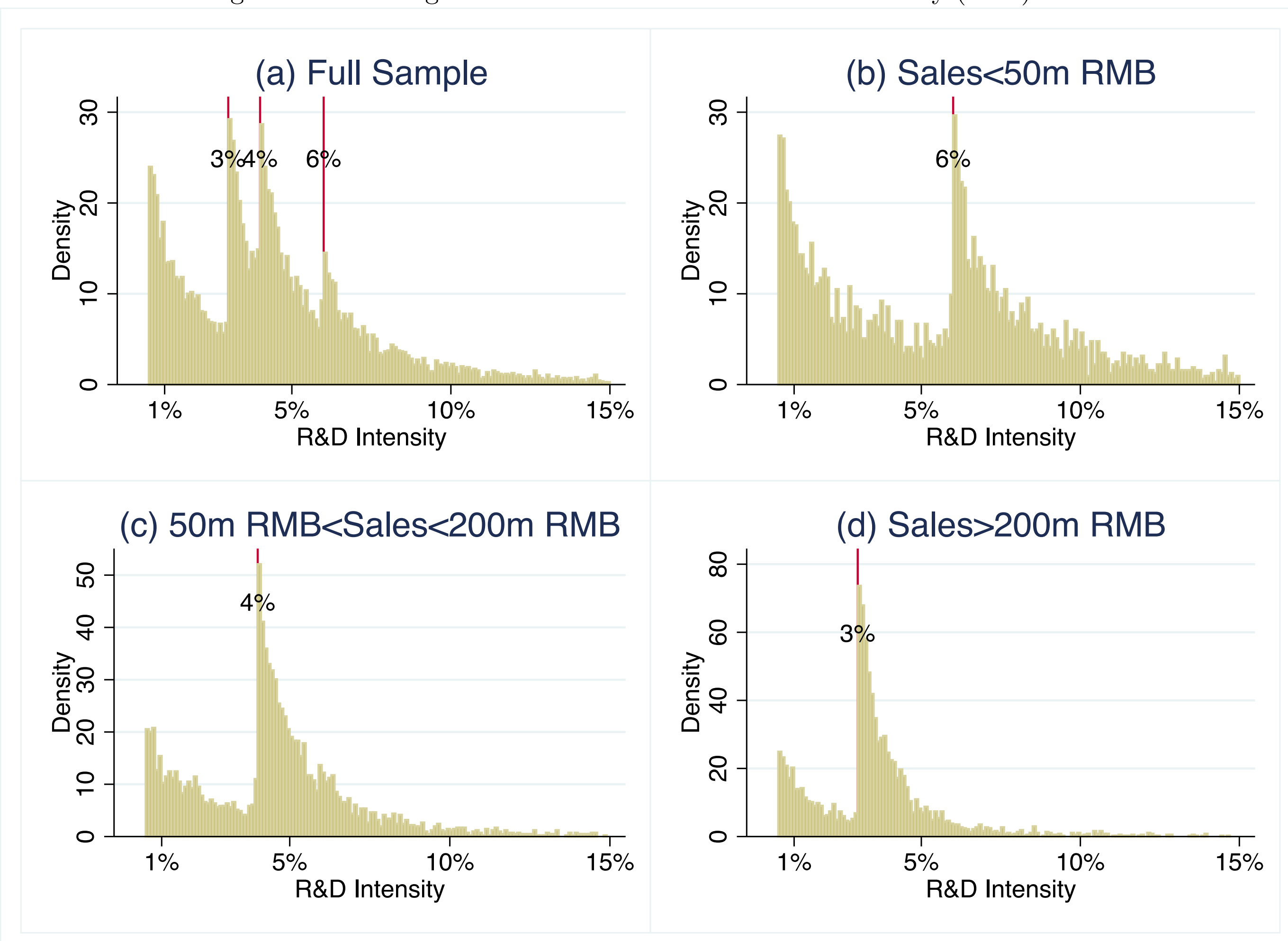
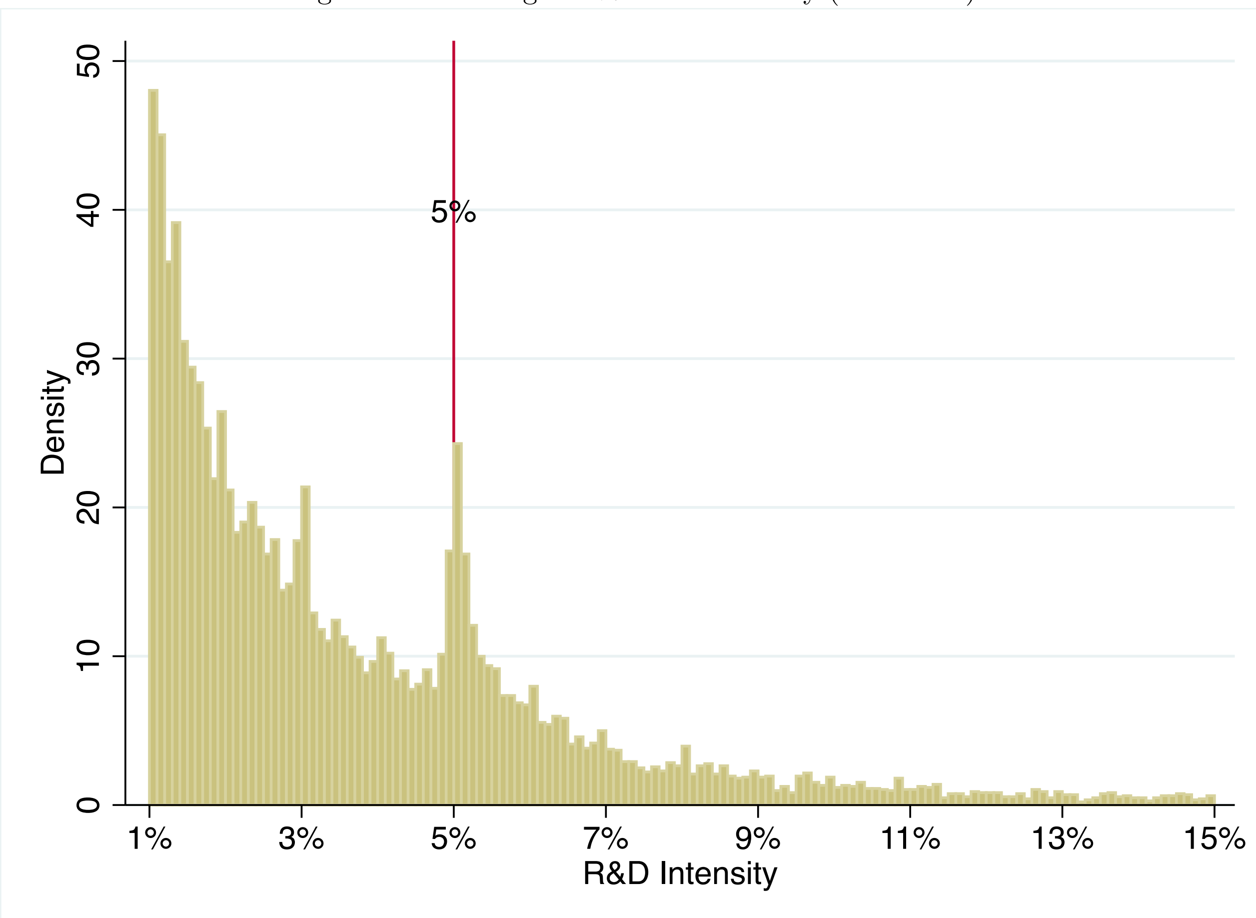


Figure 2: Bunching at Different Thresholds of R&D Intensity (2011)



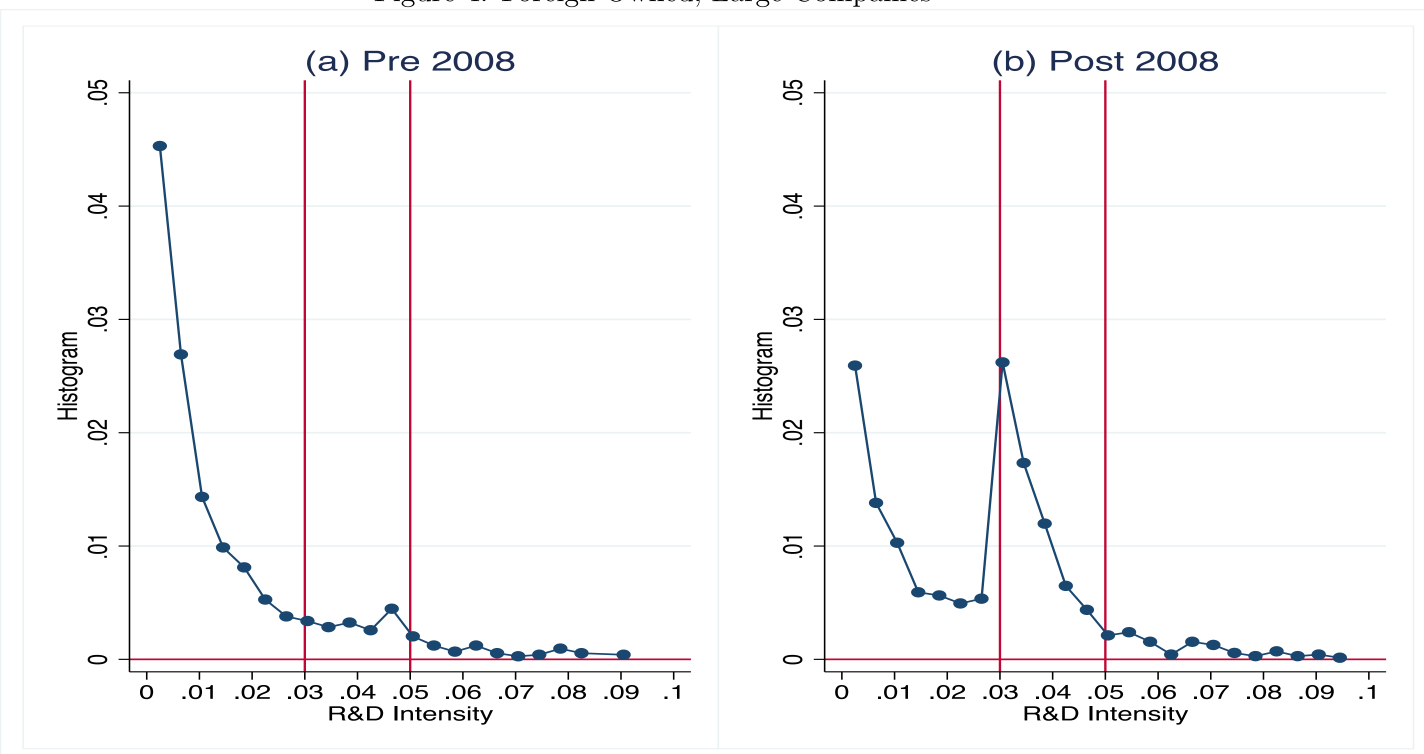
Source: Administrative Tax Return Database. See Section 2 for details.

Figure 3: Bunching at 5% R&D Intensity (2005-2007)



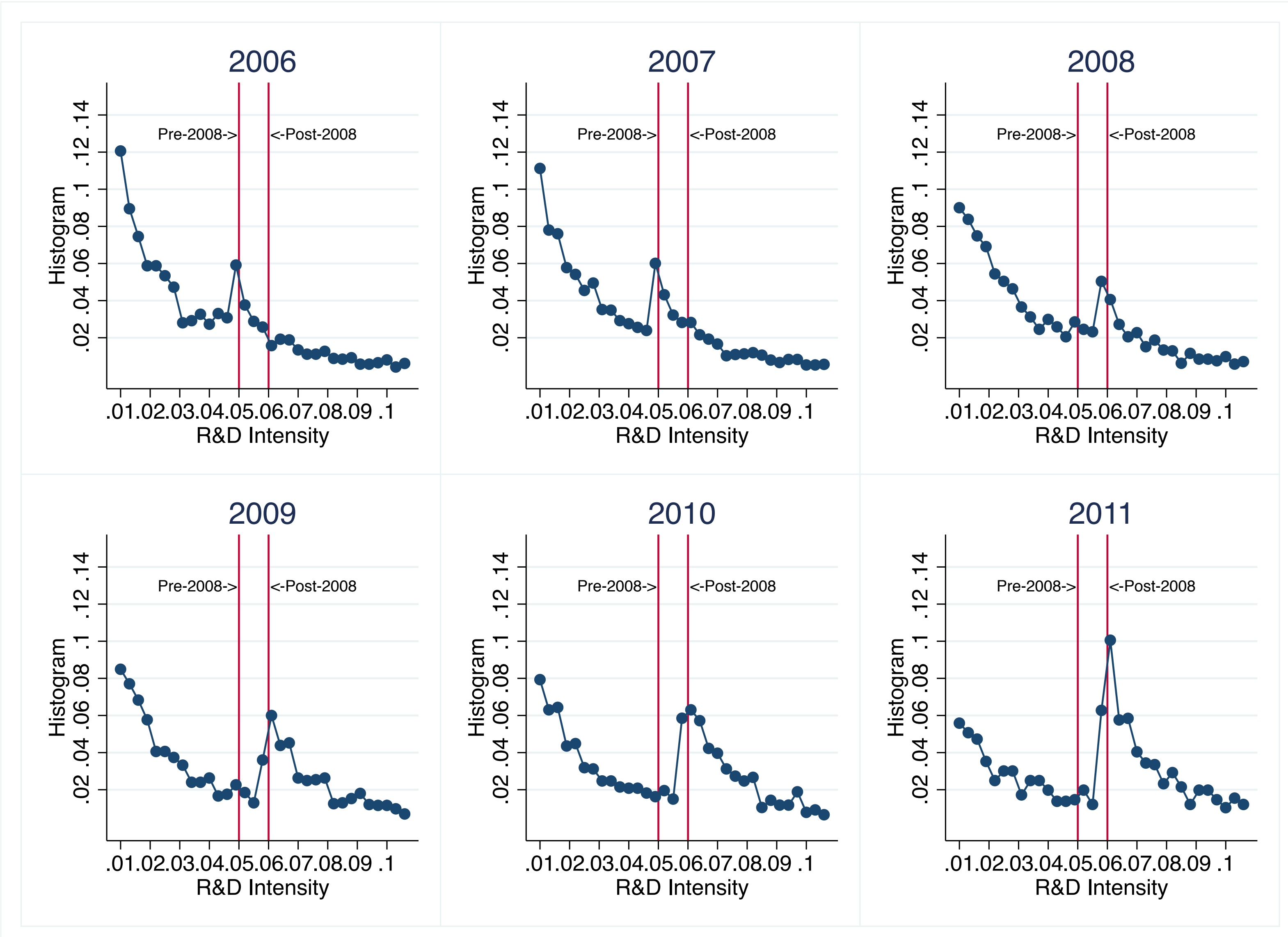
Source: Annual Survey of Manufacturers. See Section 2 for details.

Figure 4: Foreign-Owned, Large Companies



Source: Administrative Tax Return Database and Annual Survey of Manufacturers.
See Section 2 for details.

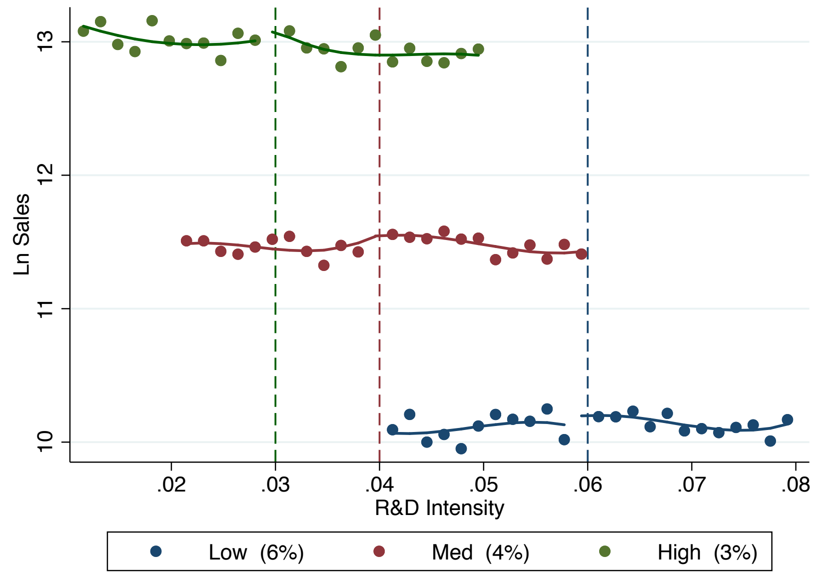
Figure 5: Domestic-Owned, Small Companies



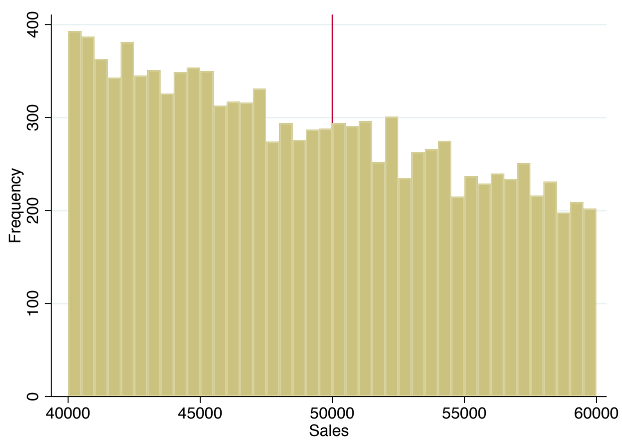
Source: Administrative Tax Return Database and Annual Survey of Manufacturers.
See Section 2 for details.

Figure 6: Lack of Sales Manipulation

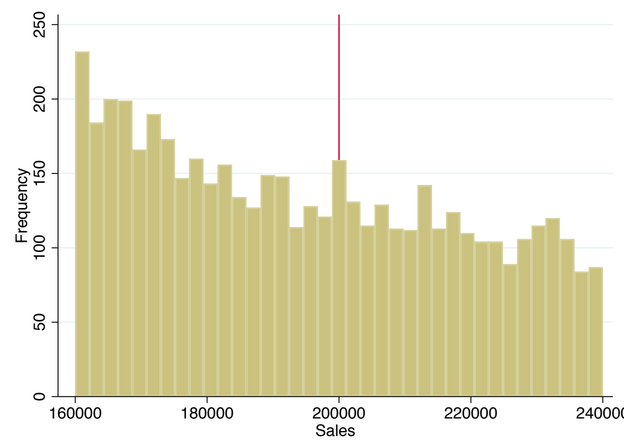
(a) Lack of Sales Manipulation Around R&D Intensity Threshold



(b) Lack of Firm Size Manipulation: Small and Medium Firms

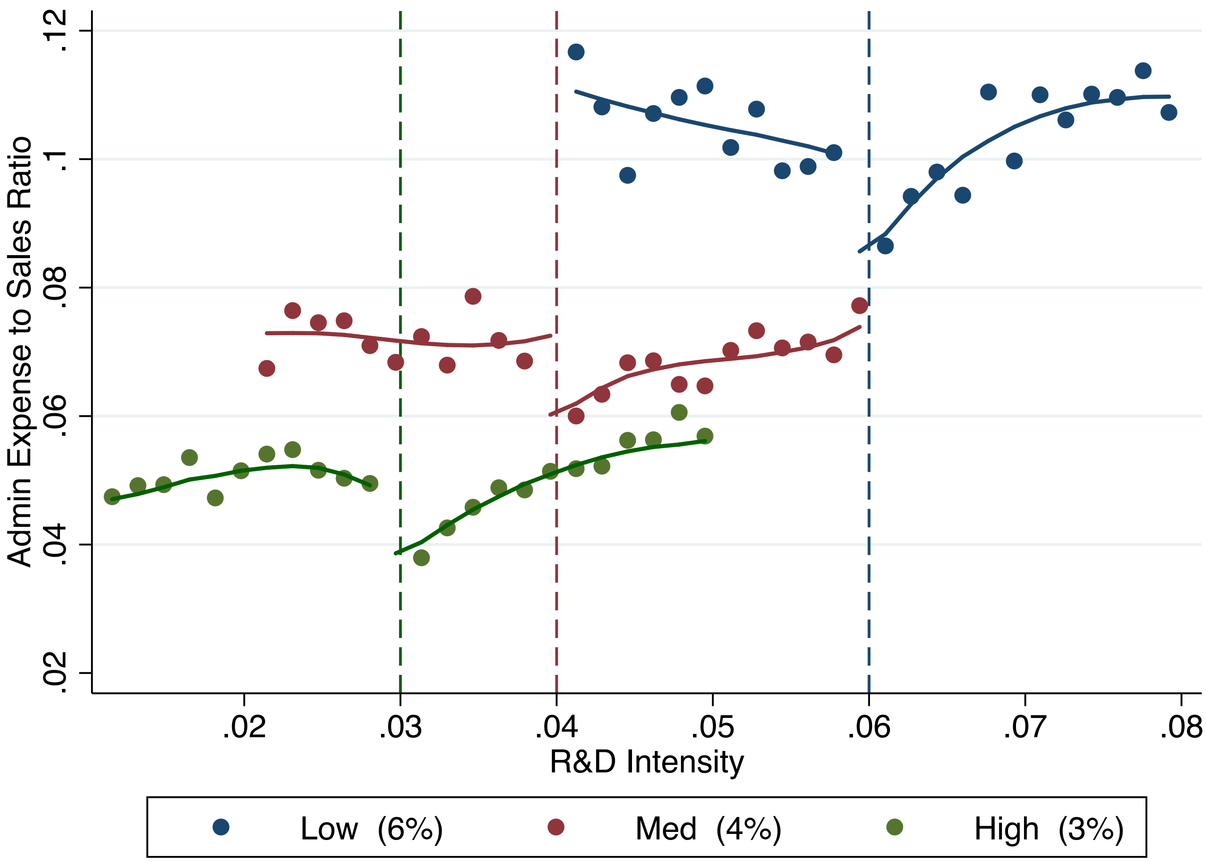


(c) Lack of Firm Size Manipulation: Medium and Large Firms



Source: Administrative Tax Return Database and Annual Survey of Manufacturers.
See Section 2 for details.

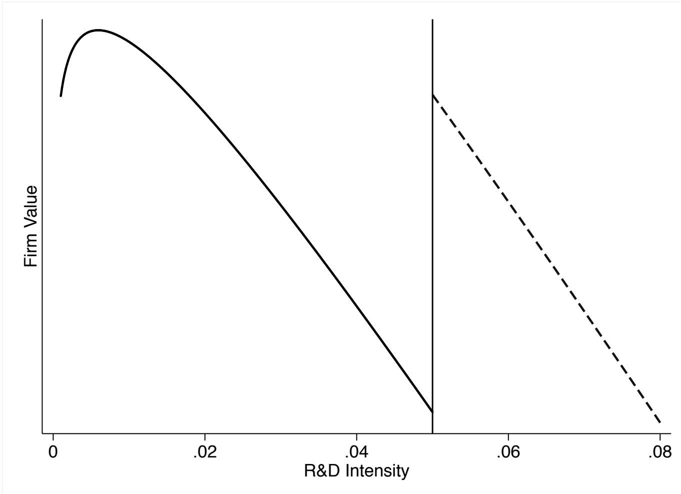
Figure 7: Empirical Evidence of Evasion



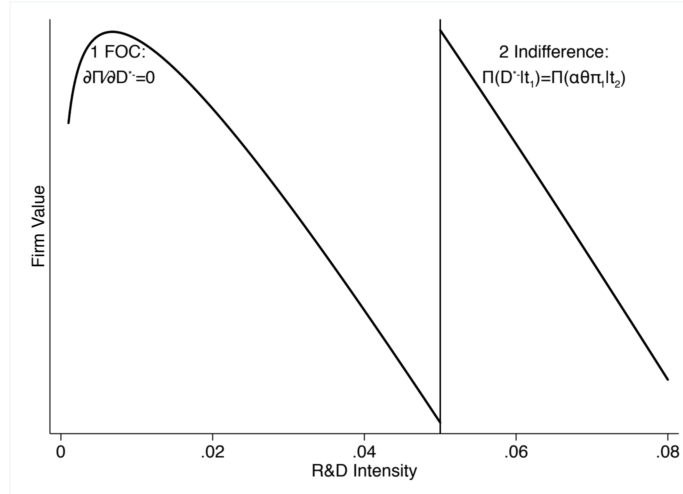
Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 4 for details on the estimation.

Figure 8: Induced Notch in Profit Functions

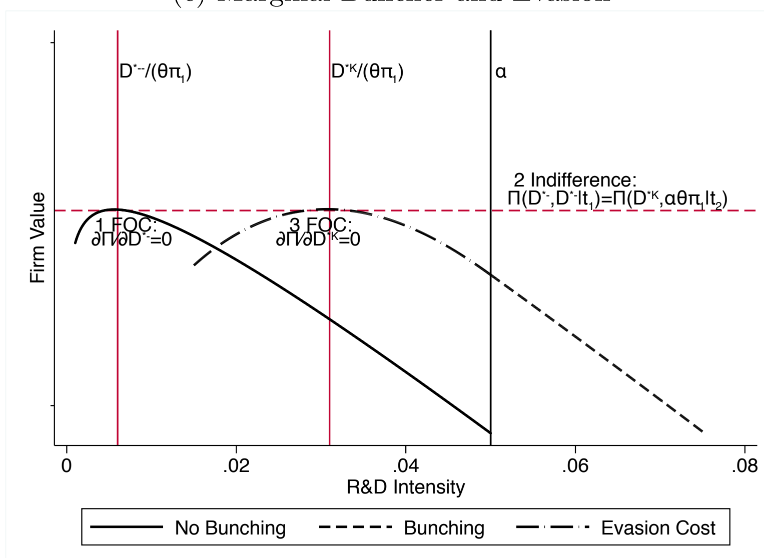
(a) Bunching is Sub-Optimal for Firm



(b) Firm is Indifferent between Internal Solution and Bunching



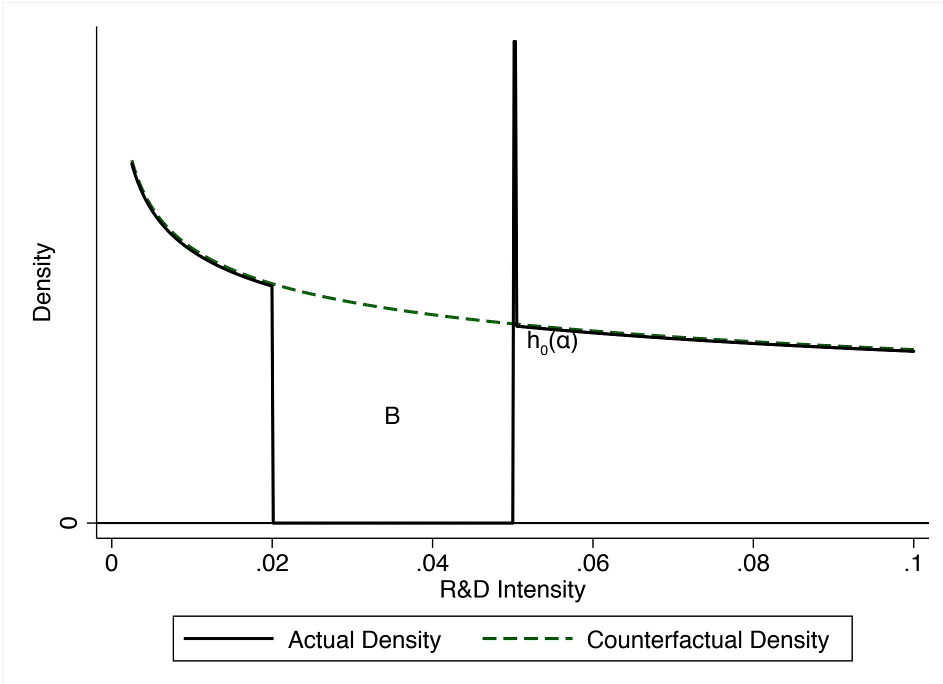
(c) Marginal Buncher and Evasion



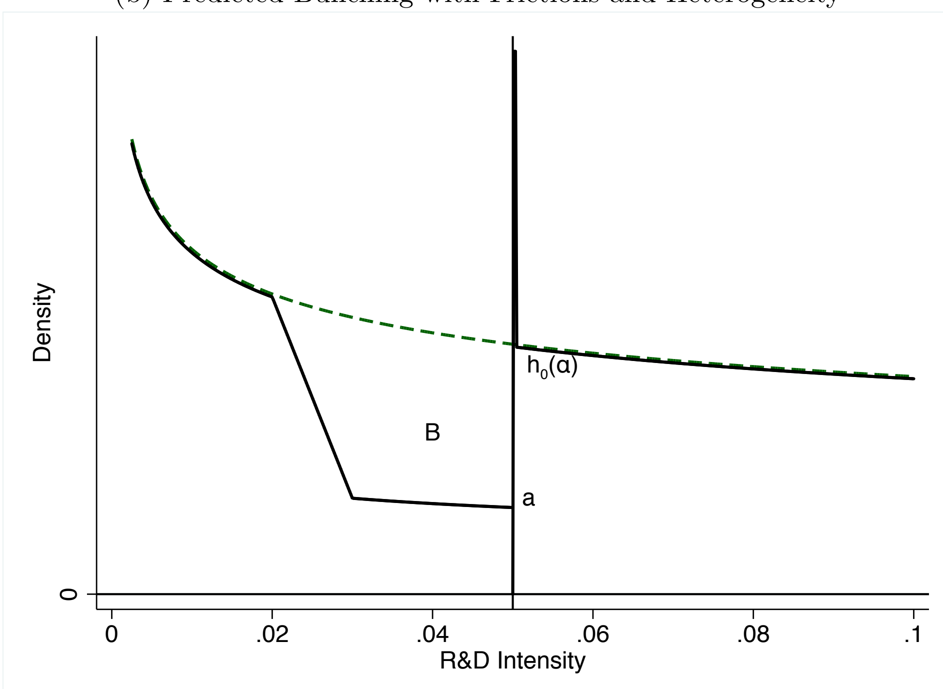
Notes: See Section 4 for details.

Figure 9: Theoretical Predictions of Bunching

(a) Predicted Bunching

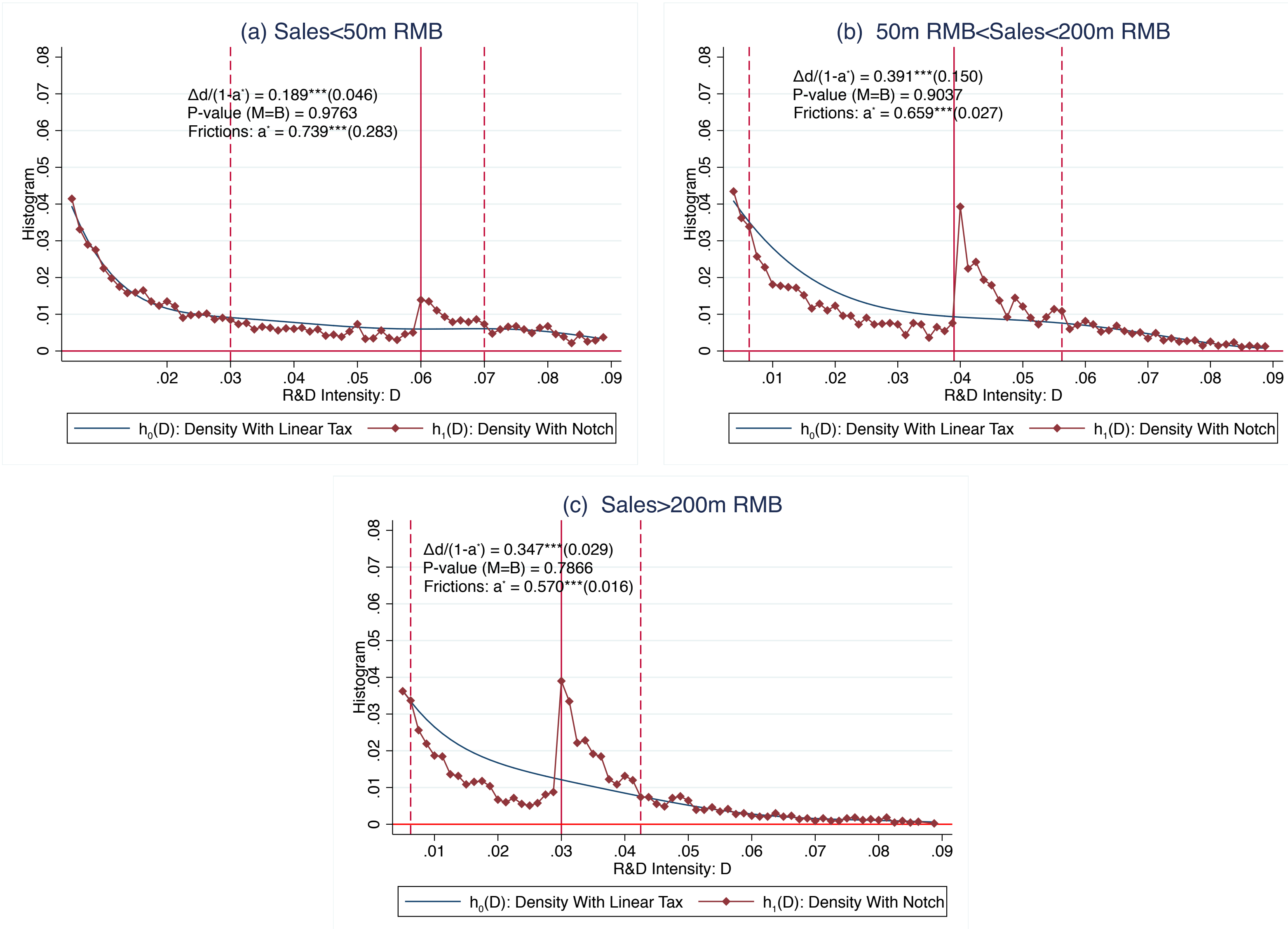


(b) Predicted Bunching with Frictions and Heterogeneity



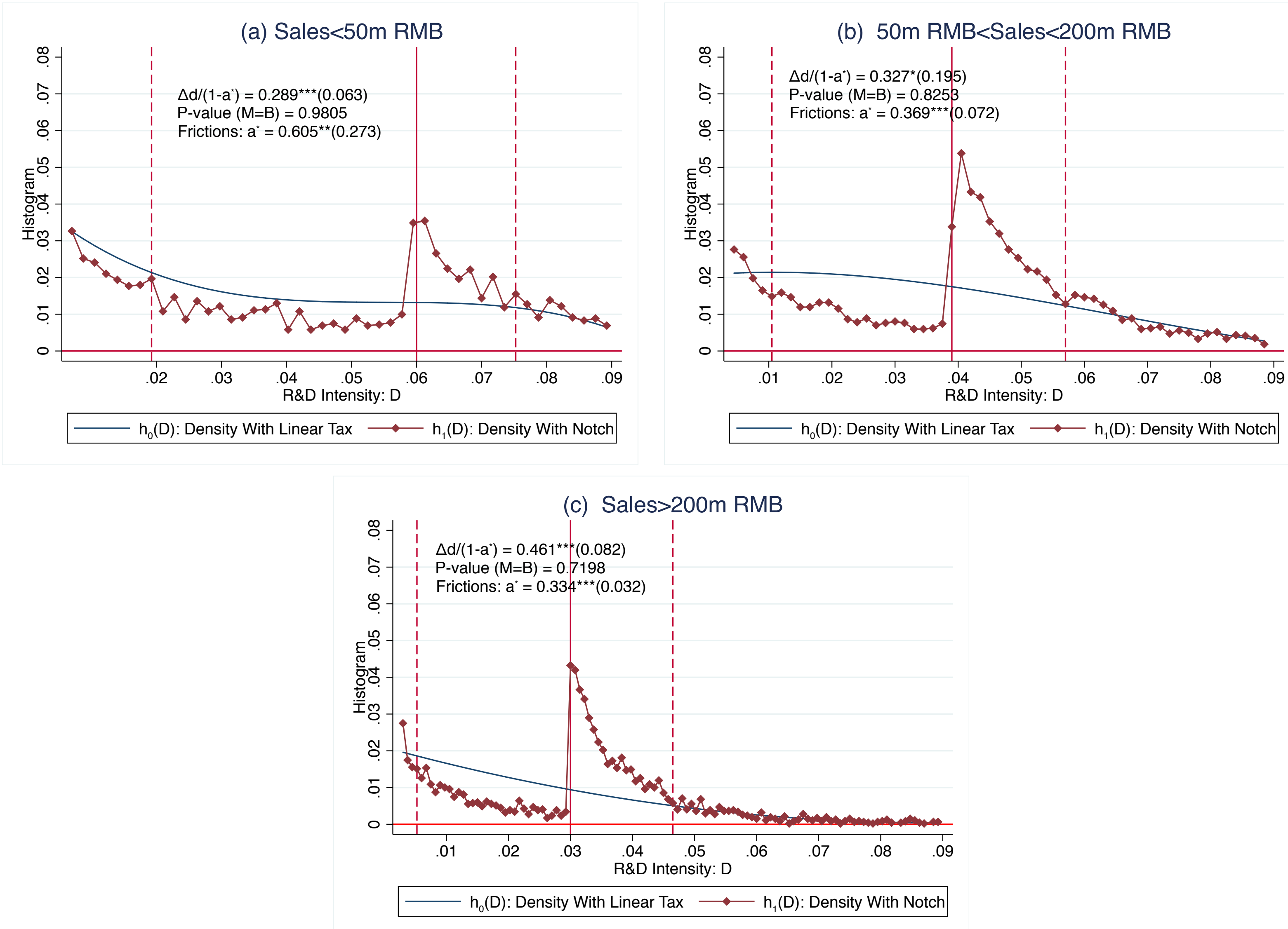
Notes: See Section 4 for details.

Figure 10: Estimates of Excess Mass from Bunching at Notch (2009)



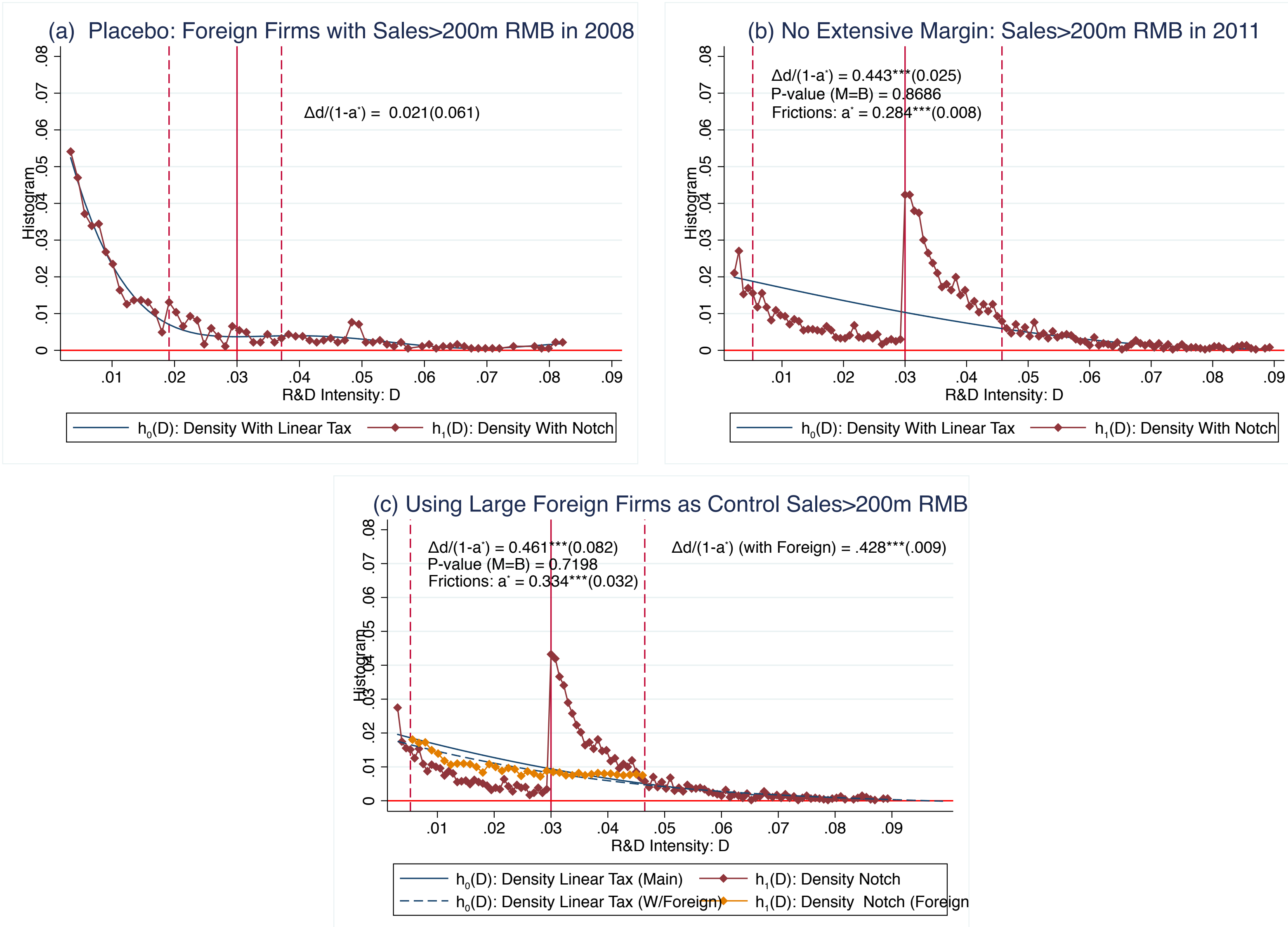
Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 4 for details on the bunching estimator.

Figure 11: Estimates of Excess Mass from Bunching at Notch (2011)



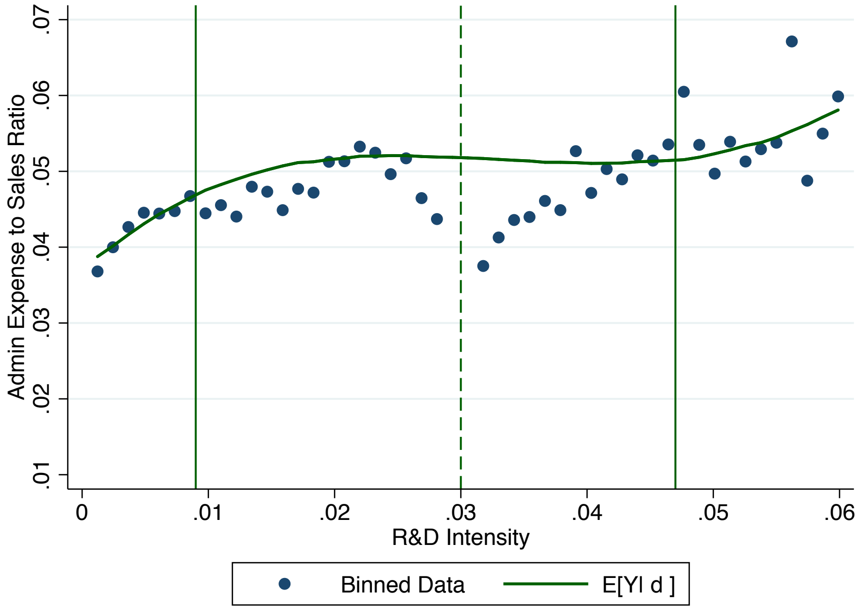
Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 4 for details on the bunching estimator.

Figure 12: Robustness of Bunching Estimates



Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 4 for details on the bunching estimator.

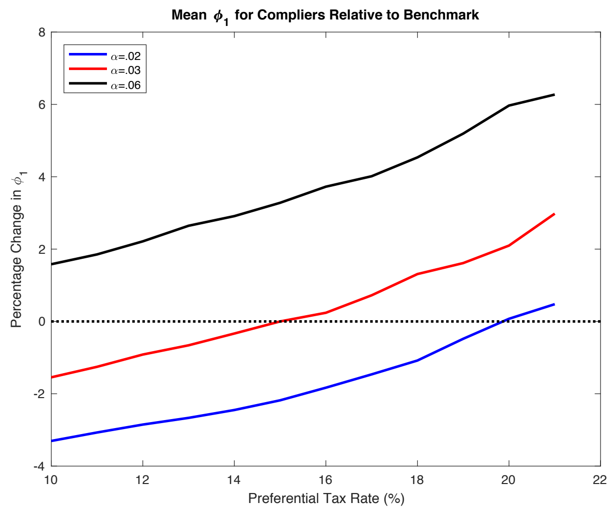
Figure 13: Estimates of Excess Mass from Bunching at Notch (2009) and ITT on Admin Cost Ratio



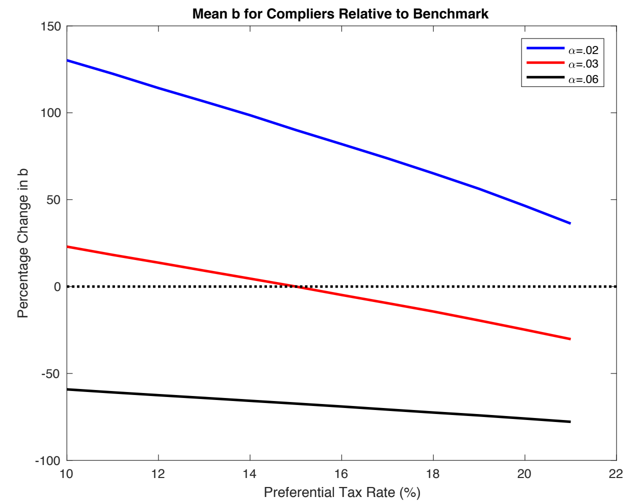
Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 4 for details on the bunching estimator.

Figure 14: Simulated Effects of Counterfactual Policies

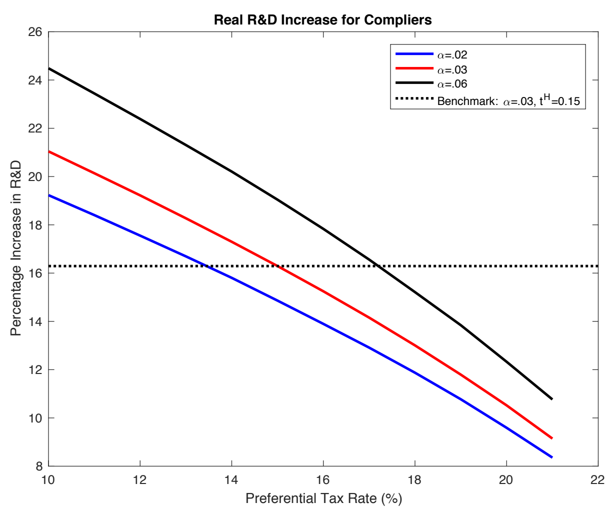
Panel (a) Mean ϕ_1 for Compliers



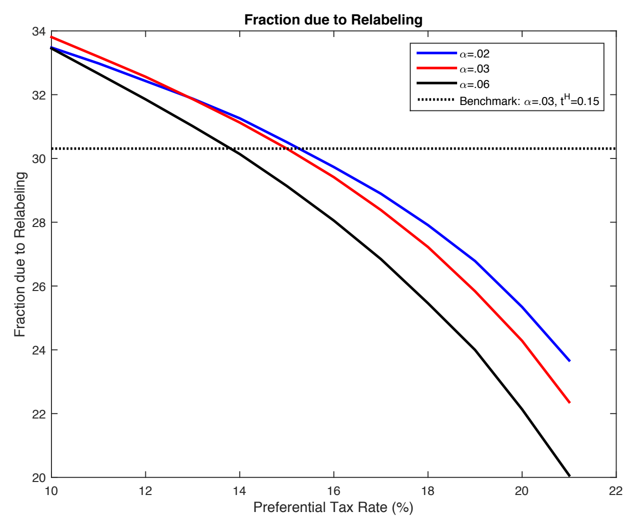
Panel (b) Mean b for Compliers



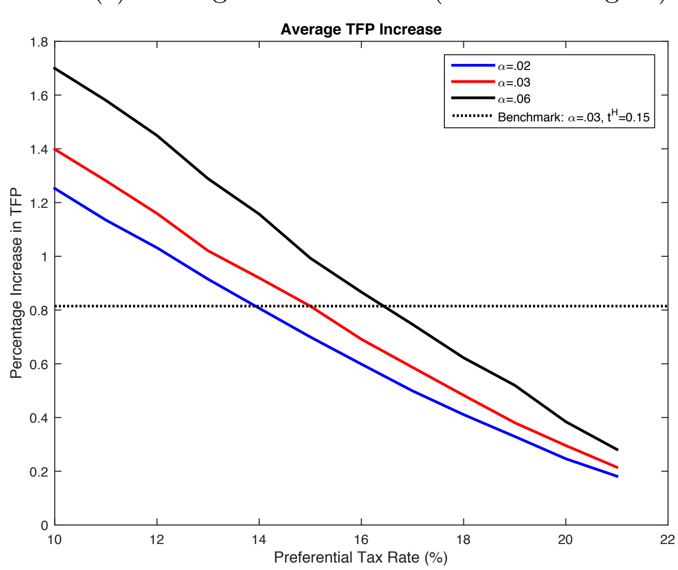
Panel (c) Real R&D Increase for Compliers



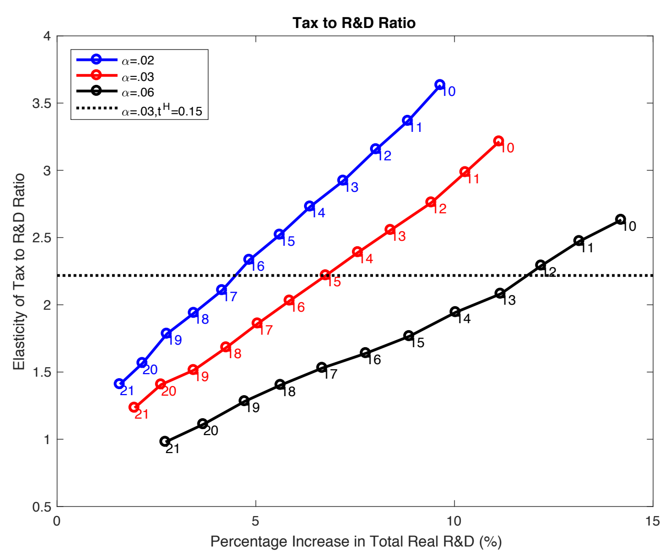
Panel (d) Fraction due to Relabeling for Compliers



Panel (e) Average TFP Increase (Excluded Region)



Panel (f) Tax Revenue Cost of Increasing R&D



Source: Authors calculations using simulated data. See Section 4 for details on the structural model and the simulation.

Table 1: Descriptive Statistics

Panel A: State Administration of Tax Data 2008 - 2011						
	Mean	Std	p25	p50	p75	# of Obs.
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
Adm 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

Panel B: Annual Survey of Manufacturing 2006 - 2007						
	Mean	Std	p25	p50	p75	# of Obs.
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 2 for details.

Table 2: Bunching Estimates of Reported R&D Investment

(a) R&D Investment in 2009

	(1)	(2)	(3)	(4)	(5)
Sales Group	Perc. Inc. in d	Fraction Constrained	Perc. Inc. in d For Compliers	Marginal Buncher Response	R&D Intensity of Marginal Buncher
	Δd	$a^*(\alpha^-)$	$\frac{\Delta d}{1-a^*(\alpha^-)}$	ΔD^*	$\alpha(1-a^*(\alpha^-))\Delta D^*$
Small	0.056	0.739***	0.189***	0.378***	0.556
	(0.093)	(0.283)	(0.046)	(0.092)	(0.748)
Medium	0.133**	0.659***	0.391***	0.782***	1.087**
	(0.067)	(0.027)	(0.150)	(0.299)	(0.541)
Large	0.149***	0.570***	0.347***	0.694***	0.897***
	(0.019)	(0.016)	(0.029)	(0.058)	(0.108)

(a) R&D Investment in 2011

	(1)	(2)	(3)	(4)	(5)
Sales Group	Perc. Inc. in d	Fraction Constrained	Perc. Inc. in d For Compliers	Marginal Buncher Response	R&D Intensity of Marginal Buncher
	Δd	$a^*(\alpha^-)$	$\frac{\Delta d}{1-a^*(\alpha^-)}$	ΔD^*	$\alpha(1-a^*(\alpha^-))\Delta D^*$
Small	0.114*	0.605**	0.289***	0.577***	1.368**
	(0.061)	(0.273)	(0.063)	(0.125)	(0.664)
Medium	0.207***	0.369***	0.327*	0.655*	1.549***
	(0.035)	(0.072)	(0.195)	(0.390)	(0.373)
Large	0.307***	0.334***	0.461***	0.921***	1.856***
	(0.078)	(0.032)	(0.082)	(0.165)	(0.448)

Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 5 for details on the estimation. Standard errors in parentheses.

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

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
Profit Ratio	0.009	0.005	1.888	0.001	0.016

(b) Wald Estimates of Treatment Effects

	Wald Estimate	Bootstrap 5th Perc.	Bootstrap 95th Perc.
2009			
Reported R&D to User Cost	-2.052	-7.919	-0.016
Real R&D to User Cost	-1.272	-4.900	-0.010
2011			
TFP to Reported R&D	0.084	0.002	0.281
Profit Ratio to Reported R&D	0.059	0.000	0.204
Tax to Reported R&D	-0.879	-2.730	-0.458

Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 5 for details on the estimation. Standard errors obtained via bootstrap.

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

Table 4: Structural Estimates

(a) Point Estimates

	TFP Elasticity of R&D	Evasion 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

Note: Estimates based on calibrated values of $\theta = 5$, $\rho = 0.725$, and $\sigma = 0.385$.

(b) Simulated vs. Data Moments

	Simulated	Data
Prob Mass $< d^{-*}$	0.284	0.280
Frac. not Bunch.	0.676	0.675
Prob Mass $> d^{+*}$	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 admin	-0.27%	-0.33%
TFP $< d^{-*}$	-0.032	-0.032
TFP between d^{-*} and d^{+*}	0.003	0.000
TFP $> d^{+*}$	0.056	0.056

The simulation is based on 30,000 firms.

Online Appendix: Not For Publication

This appendix contains multiple additional analyses. Appendix A discusses the estimation of our measure of residualized log-TFP. Appendix B provides additional analyses suggesting that a fraction of the reported R&D activity may be relabeled by contrasting the effect of reported R&D on TFP above and below the notch. Appendix C provides a detailed derivation of the model. Finally, Appendix D provides approximations of bunching implications.

A 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:

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

B Inferring Relabelling from Productivity Effect of R&D

We now investigate the implications of firm bunching and evasion behavior for measured productivity. Our benchmark model assumes the following relationship between R&D and the firm productivity:

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

Our evasion analysis indicates that firms have incentives to over-report their R&D in order to obtain the HTE status.

This measurement problem can result in attenuation bias in the estimated effectiveness of R&D on firm productivity. We overcome this challenge by borrowing from the model intuition that firms do not misreport if they decide to have an R&D intensity below the qualifying threshold. Thus, our empirical specification allows the elasticity of log TFP with respect to log reported R&D, i.e. ε, to depend on whether or not the firm is below or above the respective HTE threshold.

$$\phi_{it} = \rho\phi_{it-1} + \beta_1\mathbb{I}[\text{Above}] \times \ln RD_{i,t-1} + \beta_2\mathbb{I}[\text{Below}] \times \ln RD_{i,t-1} + u_{it}.$$

Table A.5 reports the results of this regression analysis. All specifications include industry-year fixed effects and the standard errors are clustered at the industry level. Overall, the coefficients on lagged log R&D are always highly significant. Column (1) shows that doubling R&D increases firm-level productivity by 2.8%. Comparing columns (1) and (2), we find that separately estimating the R&D elasticity based on a firm's position relative to the notch produces results consistent with the presence of evasion. When a firm's R&D intensity is below the notch, doubling R&D spending improves productivity by 2.8%. However, when a firm's R&D intensity is above the notch, this magnitude is reduced to 2.5%, around ten percent lower than the "no evasion" group. The last row of the table shows that this difference is statistically significant at the 1% level.

Columns (3)-(5) report similar estimates when we estimate this equation separately for small, medium, and large firms. The magnitude of the R&D elasticity varies across these groups, with the effectiveness of R&D improving when firm size is larger. Doubling R&D improves the productivity of a small firm by 1% but improves the productivity of a large firm by 4.4%. We also find evidence of smaller effects of R&D on productivity for firms that are above the notch, and likely misreporting. This difference also grows with firm size and is statistically significant in all cases. The attenuation in the effect of R&D on productivity suggests a second measure of relabeling given by: $1 - \frac{\beta_1}{\beta_2}$. This measure is reported in the last row of the table and is overall lower than that reported in the previous section. A potential concern with this measure is that it represent decreasing returns to scale in R&D investment. Table A.6 assuages this concern by showing that we do not obtain the same pattern of results when we replicate this table at a fake notch that is above the true notch.

C Detailed Model Derivation

C.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:

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

where $D_{i,t-1}$ 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 is depreciated (captured by ρ) and influenced by continuous 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:

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

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:

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

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:

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

where $\mathbb{E}[\pi_{it}|D_{i,t-1} = 0]$ 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:

$$g(D_{i1}, \theta\pi_{i1}) = \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}}{\theta\pi_{i1}}\right]^2.$$

C.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 intertemporal problem is given by:

$$\max_{D_1} (1-t_1)(\pi_{i1} - D_{i1} - g(D_{i1}, \theta\pi_{i1})) + \beta(1-t_2)\mathbb{E}[\pi_{i2}],$$

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:

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

Notice first that if the tax rate is constant across periods, the corporate income tax does not affect the choice of R&D investment.[32] In the special case of no adjustment costs (i.e., $b = 0$), the optimal choice of D_{i1} is given by:

$$D_{i1} = \left[\frac{1}{(\theta-1)\varepsilon}\frac{1-t_1}{\beta(1-t_2)}\frac{1}{\mathbb{E}[\pi_{i2}|D_{i1} = 0]}\right]^{\frac{1}{(\theta-1)\varepsilon-1}}. \tag{C.2}$$

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

This equation shows that the optimal R&D choice has a constant elasticity with respect to the net of tax rate, so that

$$\frac{d\ln D_{i1}}{d\ln(1-t_2)} = \frac{1}{1-(\theta-1)\varepsilon}.$$

In particular, this elasticity suggest that firms that have a higher valuation of R&D, that is when $(\theta-1)\varepsilon$ is greater, the firm will be more responsive to tax incentives.

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 $\mathbb{E}[\pi_{i2}|D_{i,t-1}=0]$. An important insight for the proceeding analysis is that we can recover these factors from D_{1i} as follows:

$$\mathbb{E}[\pi_{i2}|D_{i1}=0] = \frac{(1-t_1)D_{i1}^{1-(\theta-1)\varepsilon}}{\beta(1-t_2)\varepsilon(\theta-1)}\left(1+b\left[\frac{D_{i1}}{\theta\pi_{i1}}\right]\right).$$

Second Order Condition

This problem may feature multiple solutions. To ensure our model results in sensible solutions, we confirm the second order condition at the estimated values. The SOC is given by:

$$SOC:\ -(1-t_1)\left(b\left[\frac{1}{\theta\pi_{i1}}\right]\right)+\beta(1-t_2)\varepsilon(\theta-1)((\theta-1)\varepsilon-1)D_{i1}^{(\theta-1)\varepsilon-2}\mathbb{E}[\pi_{i2}|D_{i1}=0]<0.$$

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

C.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:

$$t_2 = \begin{cases} t_2^{LT} & \text{if } D_1 < \alpha\theta\pi_1 \\ t_2^{HT} & \text{if } D_1 \geq \alpha\theta\pi_1 \end{cases},$$

sales equal $\theta\pi_1$, $t_2^{LT} > t_2^{HT}$ and where α is the R&D intensity required to attain the high-tech certification and LT/HT stands for low-tech/high-tech. In addition, we also introduced a fixed costs of certification c such that firms need to pay $c\times\alpha\theta\pi_{1i}$ 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 $D_1=\alpha\theta\pi_1$. Figure 8 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 C.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 C.1 and by $D_1=\alpha\theta\pi_1$.

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 $\mathbb{E}[\pi_{i2}|D_{i,t-1}=0]$, 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 $\mathbb{E}[\pi_{i2}|D_{i,t-1}=0]$ and $(b,\ c)$ is smoothly distributed around the threshold

α, this incentive will lead a mass of firms to find $D_1 = \alpha\theta\pi_1$ optimal and thus "bunch" at this level. Our analysis proceeds by first identifying the firm that is marginal between both solutions in terms of the R&D intensity and then by using the identity of the marginal firm to relate the amount of bunching at the notch to the firm's valuation of R&D investment $\varepsilon(\theta-1)$.

We start by characterizing the firm that is indifferent between level of R&D given by the notch and a lower level of R&D investment D_{i1}^{*-}. Define $\Pi(\cdot|t)$ as the value function of the firm's inter-temporal maximization problem when facing tax t in period 2. A firm i is a marginal buncher if:

$$\Pi(D_{i1}^{*-}|t_2^{LT}) \quad = \quad \Pi(\alpha\theta\pi_1|t_2^{HT}),$$

where the left-hand side is the profit from an internal solution facing the low-tech tax rate t_2^{LT} and the right hand side is the bunching solution when facing the high-tech tax rate t_2^{HT}. Using the optimal choice for an internal solution in Equation C.1, we can manipulate $\Pi(D_{i1}^{*-}|t_2^{LT})$ to obtain:

$$\begin{aligned}\Pi(D_{i1}^{*-}|t_2^{LT}) &= (1-t_1)\left(\pi_{i1} - D_{i1}^{*-} - \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]^2\right) + \beta(1-t_2^{LT})(D_{i1}^{*-})^{(\theta-1)\varepsilon}\mathbb{E}[\pi_{i2}|D_{i1}=0] \\ &= (1-t_1)\left(\pi_{i1} + \left(\frac{1}{\varepsilon(\theta-1)} - 1\right)D_{i1}^{*-} + b\theta\pi_{i1}\left(\frac{1}{\varepsilon(\theta-1)} - \frac{1}{2}\right)\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]^2\right), \qquad \text{(C.3)}\end{aligned}$$

where we substitute for $\mathbb{E}[\pi_{i2}|D_{i1}=0]$ using the optimality condition above.

Similarly, we manipulate $\Pi(\alpha\theta\pi_1|t^{HT})$ by substituting for the unobserved components of the firm-decision, i.e. $\mathbb{E}[\pi_{i2}|D_{i1}=0]$, using Equation C.1 to obtain:

$$\begin{aligned}\Pi(\alpha\theta\pi_1|t_2^{HT}) &= (1-t_1)\left(\pi_{i1} - \alpha\theta\pi_{i1}(1+c) - \frac{b\theta\pi_{i1}}{2}\left[\frac{\alpha\theta\pi_{i1}}{\theta\pi_{i1}}\right]^2\right) + \beta(1-t_2^{HT})(\alpha\theta\pi_{i1})^{(\theta-1)\varepsilon}\mathbb{E}[\pi_{i2}|D_{i1}=0] \\ &= (1-t_1)\left(\pi_{i1} - \alpha\theta\pi_{i1}(1+c) - \frac{\alpha^2 b\theta\pi_{i1}}{2}\right. \\ &\qquad \left. + \frac{(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{\alpha\theta\pi_{i1}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon}\left(1 + b\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]\right)D_{i1}^{*-}\right). \qquad \text{(C.4)}\end{aligned}$$

We then use Equations C.3 and C.4 and the indifference condition that defines the marginal bunching firm to obtain a relation between the percentage difference in R&D intensity and the parameters of interest: $(\theta-1)\varepsilon$.

Subtracting $\Pi(\alpha\theta\pi_1|t_2^{HT})$ from $\Pi(D_1^{*-}|t_2^{LT})$ and manipulating we obtain:

$$\begin{aligned}
0 &= \left(\frac{1}{\varepsilon(\theta-1)}-1\right)D_{i1}^{*-}+b\theta\pi_{i1}\left(\frac{1}{\varepsilon(\theta-1)}-\frac{1}{2}\right)\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]^2+\alpha\theta\pi_{i1}(1+c)+\frac{\alpha^2 b\theta\pi_{i1}}{2}\\
&- \frac{(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{\alpha\theta\pi_{i1}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon}\left(1+b\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]\right)D_{i1}^{*-}\\
0 &= \left(\frac{1}{\varepsilon(\theta-1)}-1\right)\frac{D_{i1}^{*-}}{\alpha\theta\pi_1}+\alpha b\left(\frac{1}{\varepsilon(\theta-1)}-\frac{1}{2}\right)\left[\frac{D_{i1}^{*-}}{\alpha\theta\pi_{i1}}\right]^2+1+c+\frac{\alpha b}{2}\\
&- \frac{(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{\alpha\theta\pi_{i1}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon-1}\left(1+\alpha b\left[\frac{D_{i1}^{*-}}{\alpha\theta\pi_{i1}}\right]\right)\\
0 &= \left(\frac{1}{\varepsilon(\theta-1)}-1\right)(1-\Delta D^*)+\alpha b\left(\frac{1}{\varepsilon(\theta-1)}-\frac{1}{2}\right)(1-\Delta D^*)^2+1+c+\frac{\alpha b}{2}\\
&- \frac{(1-t_2^{HT})}{(1-t_2^{LT})}\times\frac{(1-\Delta D^*)^{1-(\theta-1)\varepsilon}}{\varepsilon(\theta-1)}\left(1+\alpha b(1-\Delta D^*)\right),
\end{aligned} \tag{C.5}$$

where the first line ignores the common term $(1-t_1)$ in both equations, the second line divides by $\alpha\theta\pi_1$, and the third line defines $\Delta D^* = \frac{\alpha\theta\pi_1 - D^{*-}}{\alpha\theta\pi_1}$ as the percentage increase in R&D spending due to the notch. Given a random draw of b, c, Equation C.5 is an implicit function for $(\theta-1)\varepsilon$, thus the empirical quantity ΔD^* provides a source of identification for the returns to R&D.

C.4 R&D Choice Under Tax Notch with Evasion

Assume now that firms may misreport their costs and shift non-RD costs to the R&D category. Following conversations with CFOs of large Chinese companies, we model evasion as a choice to misreport expenses across R&D and non-RD 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) and Pomeranz (2015)).

Denote a firm's reported level of R&D spending by $\tilde{D}_1$. The expected cost of misreporting to the firm is given by $h(D_1, \tilde{D}_1)$. We assume that the cost of mis-reporting is proportional to the reported R&D, $\tilde{D}_1$, and depends on the percentage of mis-reported R&D, $\frac{\tilde{D}_1 - D_1}{\tilde{D}_1}$, so that:

$$h(D_1, \tilde{D}_1) = \tilde{D}_1 \tilde{h}\left(\frac{\tilde{D}_1 - D_1}{\tilde{D}_1}\right).$$

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 characterize the firm that is indifferent between bunching, and potentially misreporting, and not bunching.

We start by characterizing the firm that is indifferent between level of R&D given by the notch and a lower level of R&D investment D_{i1}^{*-}. Define $\Pi(D_1, \tilde{D}_1|t)$ as the value function of the firm's intertemporal maximization problem when facing tax t in period 2 that spends D_1 on R&D but that declares $\tilde{D}_1$. A firm i is a marginal buncher if:

$$\Pi(D_{i1}^{*-}, D_{i1}^{*-}|t_2^{LT}) \quad = \quad \Pi(\alpha\theta\pi_1, D_1^{*K}|t_2^{HT}),$$

where the left-hand side is the profit from an internal solution facing the low-tech tax rate t_2^{LT} , the right hand side is the bunching solution when facing the high-tech tax rate t_2^{HT}, and where the firm chooses a real R&D level of D^{*K}.

We first consider $\Pi(D_{i1}^{*-}, D_{i1}^{*-}|t_2^{LT})$. Since the firm need not mis-report in this case, Equation C.3 still describes the profit in this case. We then manipulate $\Pi(\alpha\theta\pi_1, D_1^{*K}|t^{HT})$ using the FOC for D_{i1}^{*-} to obtain:

$$\begin{aligned}
\Pi(\alpha\theta\pi_{i1}, D_{i1}^{*K}|t_2^{HT}) &= (1-t_1)\left(\pi_{i1} - D_{i1}^{*K} - \alpha\theta\pi_{i1}c - \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}^{*K}}{\theta\pi_{i1}}\right]^2\right) \\
&\quad +\beta(1-t_2^{HT})(D_{i1}^{*K})^{(\theta-1)\varepsilon}\mathbb{E}[\pi_{i2}|D_{i1}=0] - h(D_1^{*K}, \alpha\theta\pi_1) \\
&= (1-t_1)\left(\pi_{i1} - D_{i1}^{*K} - \alpha\theta\pi_{i1}c - \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}^{*K}}{\theta\pi_{i1}}\right]^2\right) - h(D_1^{*K}, \alpha\theta\pi_1) \\
&\quad +\frac{(1-t_1)(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{D_{i1}^{*K}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon}\left(1+b\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]\right)D_{i1}^{*-} \qquad \text{(C.6)}
\end{aligned}$$

We then use Equations C.3 and C.6 and the indifference condition that defines the marginal bunching firm to obtain a relation between the percentage difference in R&D intensity and the parameters of interest: $(\theta-1)\varepsilon$. Subtracting $\Pi(\alpha\theta\pi_1, D_1^{*K}|t_2^{HT})$ from $\Pi(D_{i1}^{*-}, D_{i1}^{*-}|t_2^{LT})$ and rearranging we obtain:

$$\begin{aligned}
0 &= \left(\frac{1}{\varepsilon(\theta-1)} - 1\right)D_{i1}^{*-} + b\theta\pi_{i1}\left(\frac{1}{\varepsilon(\theta-1)} - \frac{1}{2}\right)\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]^2 + D_{i1}^{*K} + \alpha\theta\pi_{i1}c + \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}^{*K}}{\theta\pi_{i1}}\right]^2 \\
&- \frac{(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{D_{i1}^{*K}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon}\left(1+b\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]\right)D_{i1}^{*-} + \frac{h(D_1^{*K}, \alpha\theta\pi_1)}{(1-t_1)} \\
0 &= \left(\frac{1}{\varepsilon(\theta-1)} - 1\right)\frac{D_{i1}^{*-}}{\alpha\theta\pi_1} + \alpha b\left(\frac{1}{\varepsilon(\theta-1)} - \frac{1}{2}\right)\left[\frac{D_{i1}^{*-}}{\alpha\theta\pi_{i1}}\right]^2 + \frac{D_{i1}^{*K}}{\alpha\theta\pi_{i1}} + c + \frac{\alpha b}{2}\left[\frac{D_{i1}^{*K}}{\alpha\theta\pi_{i1}}\right]^2 \\
&- \frac{(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{D_{i1}^{*K}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon-1}\left(1+\alpha b\left[\frac{D_{i1}^{*-}}{\alpha\theta\pi_{i1}}\right]\right)\frac{D_{i1}^{*K}}{\alpha\theta\pi_{i1}} + \frac{\tilde{h}(D_1^{*K}, \alpha\theta\pi_1)}{(1-t_1)},
\end{aligned}$$

where the first line ignores the common term $(1-t_1)$ and the second line divides by $\alpha\theta\pi_1$. We now use the definitions $\Delta D^* = \frac{\alpha\theta\pi_1 - D^{*-}}{\alpha\theta\pi_1}$ as the percentage increase in R&D spending due to the notch and $\delta = \frac{\tilde{D}_1 - D_1}{\tilde{D}_1}$ as the percentage of misreporting relative to the reported value. These definitions and assumptions yield the following condition:

$$\begin{aligned}
0 &= 1 + \frac{c}{1-\delta^*} + \frac{\alpha b}{2}(1-\delta^*) + \left(\frac{1-\Delta D^*}{1-\delta^*}\right)\left[\left(\frac{1}{\varepsilon(\theta-1)} - 1\right) + \alpha b\left(\frac{1}{\varepsilon(\theta-1)} - \frac{1}{2}\right)(1-\Delta D^*)\right] \\
&- \frac{(1-t_2^{HT})}{(1-t_2^{LT})}\left(\frac{1-\Delta D^*}{1-\delta^*}\right)^{1-(\theta-1)\varepsilon}\left(\frac{1+\alpha b(1-\Delta D^*)}{\varepsilon(\theta-1)}\right) + \frac{\tilde{h}(\delta^*)(1-\delta^*)^{-1}}{(1-t_1)} \qquad \text{(C.7)}
\end{aligned}$$

Notice that in the special case with no evasion, when $\delta^* = 0$, Equation C.7 is identical to Equation C.5.

In the case when the firm decides to bunch and evade, we have the additional information that D^K is chosen optimally. From Equation C.6, the firm solves the following problem:

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

with the following FOC:

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

Notice that this equation is equivalent to:

$$\left(\frac{1-\delta^*}{1-\Delta D^*}\right)^{(\theta-1)\varepsilon-1} = \frac{1+\alpha b(1-\delta^*) - \frac{\tilde{h}'(\delta^*)}{1-t_1}}{\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)(1+\alpha b(1-\Delta D^*))} \tag{C.8}$$

Equation C.8 along with Equation C.7 now form a system of two equations that are implicit functions for ΔD^* and δ^*, given a draw of (b,c).

D Bunching Approximations

This appendix details derivations of expressions that approximate changes in the R&D investment with the estimated density.

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

As in previous papers, (*e.g.*, Kleven and Waseem (2013)), we can use similar approximations to relate the quantities B and $h_0(\alpha)$ to the behavior of the marginal firm. We first consider the special case without frictions, and note that

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

The first part of Equation D.3 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 D.3 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:

$$\Delta D^* \approx \frac{B}{h_0(\alpha)\alpha}. \tag{D.2}$$

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: $\frac{\Delta D^*}{1-\Delta D^*} = \frac{\alpha - d^{*-}}{d^{*-}}$.

In the case of heterogeneous frictions, we may obtain a similar approximation if we assume that the probability of being constrained does not depend on d. This may happen, for instance, if a constant fraction of firms are constrained regardless of 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 D.3, 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:

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

D.2 Average Percentage Increase in R&D Intensity

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

$$\mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})] = \int_{d^{*-}}^{d^{*+}} d h_0(d) dd \approx \underbrace{\frac{\alpha - d^{*-}}{2}}_{\underline{d}} \int_{d^{*-}}^{\alpha} h_0(d) dd + \underbrace{\frac{d^{*+} - \alpha}{2}}_{\bar{d}} \int_{\alpha}^{d^{*+}} h_0(d) dd \qquad \text{(D.3)}$$

$$\mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})] = \int_{d^{*-}}^{d^{*+}} d h_1(d) dd \approx \underbrace{\frac{\alpha - d^{*-}}{2}}_{\underline{d}} \int_{d^{*-}}^{\alpha} h_1(d) dd + \underbrace{\frac{d^{*+} - \alpha}{2}}_{\bar{d}} \int_{\alpha}^{d^{*+}} h_1(d) dd \qquad \text{(D.4)}$$

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

$$\mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})] - \mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})] \approx \bar{d} \underbrace{\int_{\alpha}^{d^{*+}} (h_1(d) - h_0(d))dd}_{B} \tag{D.5}$$

$$+ \quad \underline{d} \underbrace{\int_{d^{*-}}^{\alpha} (h_1(d) - h_0(d))dd}_{-B} \tag{D.6}$$

$$= \quad B(\bar{d} - \underline{d}), \tag{D.7}$$

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

Now, 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:

$$\frac{\mathbb{E}[d|\text{Notch}, d \in (d^{*-}, d^{*+})] - \mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]}{\mathbb{E}[d|\text{No Notch}, d \in (d^{*-}, d^{*+})]} = \frac{B}{2\alpha h_0(\alpha)}. \tag{D.8}$$

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.

D.3 Identification of Intent-to-Treat Effect

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

$$\mathbb{E}[Y|\text{No Notch}, d \in (d^{*-}, d^{*+})] = \underbrace{\int_{d^{*-}}^{\alpha} Yh_0(d) \times \mathbb{P}r(\text{Constrained}|d)dd}_{\text{Never Takers}} \tag{D.9}$$

$$+ \underbrace{\int_{d^{*-}}^{\alpha} Yh_0(d) \times (1 - \mathbb{P}r(\text{Constrained}|d))dd}_{\text{Compliers}} + \underbrace{\int_{\alpha}^{d^{*+}} Yh_0(d)dd}_{\text{Always Takers}}$$

Similarly, we can write

$$\mathbb{E}[Y|\text{Notch}, d \in (d^{*-}, d^{*+})] = \underbrace{\int_{d^{*-}}^{\alpha} Y h_1(d) dd}_{\text{Never Takers}} + \underbrace{\int_{\alpha}^{d^{*+}} Y h_1(d) \times (1 - \mathbb{P}r(\text{Constrained}|d)) \times \mathbb{I}[d_0 \in (d^{*-}, \alpha)] dd}_{\text{Compliers}} + \underbrace{\int_{\alpha}^{d^{*+}} Y h_1(d) \mathbb{I}[d_0 \in (\alpha, d^{*+})] dd,}_{\text{Always Takers}} \tag{D.10}$$

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:

$$ITT^Y = \int_{\alpha}^{d^{*+}} Y h_1(d)(1 - \mathbb{P}r(\text{Constrained}|d))\mathbb{I}[d_0 \in (d^{*-}, \alpha)] dd - \int_{d^{*-}}^{\alpha} Y h_0(d)(1 - \mathbb{P}r(\text{Constrained}|d)) dd, \tag{D.11}$$

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 D.11:

$$ITT^Y \approx B(\bar{Y} - \underline{Y}) \tag{D.12}$$

where $\bar{Y} = \mathbb{E}[Y|d \in (\alpha, d^{*+})]$ and $\underline{Y} = \mathbb{E}[Y|d \in (d^{*-}, \alpha)]$. 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. Combining this equation with Equation D.7 we obtain the Wald estimator as follows:

$$Wald^Y = \frac{ITT^Y}{ITT^d} \approx \frac{\bar{Y} - \underline{Y}}{\bar{d} - \underline{d}},$$

which gives the interpretation of the increase in Y for a given unit increase in d. Note that this interpretation carries the implication that there are no other effects from being certified as an InnoCom firm on Y beyond the effect on d.

Appendix Graphs

Figure A.1: Alternative Empirical Evidence of Evasion

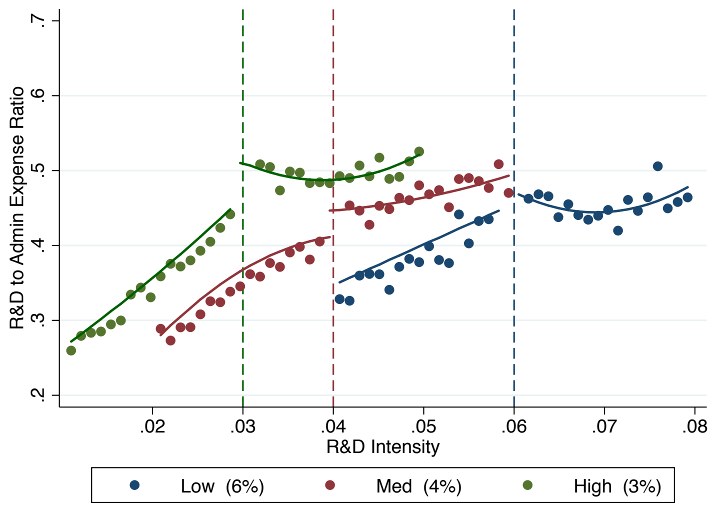


Figure A.2: Aggregate Implications

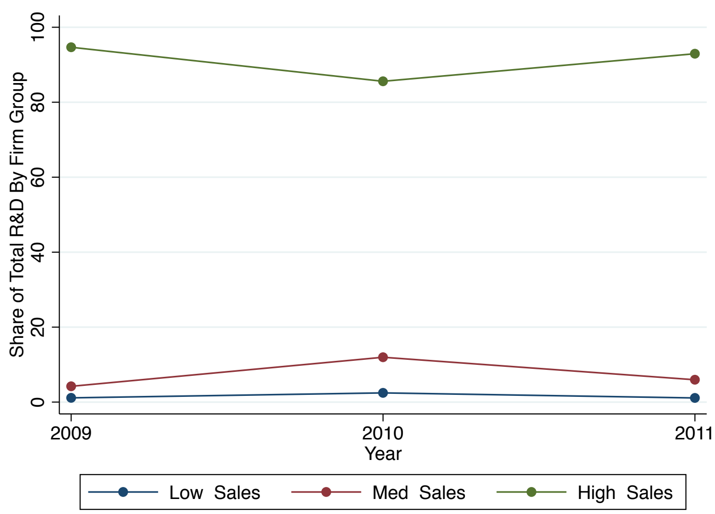
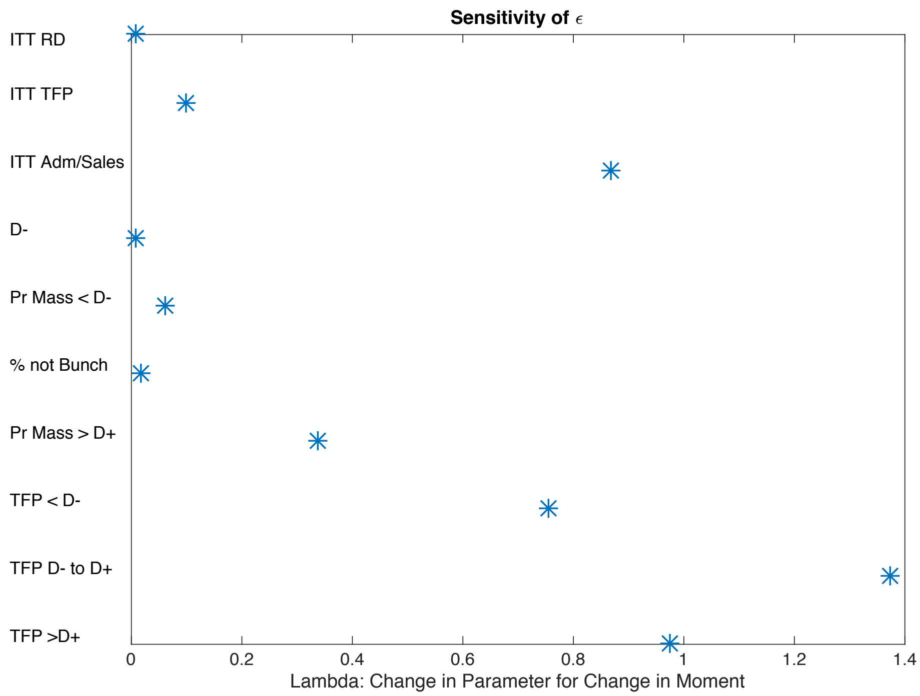
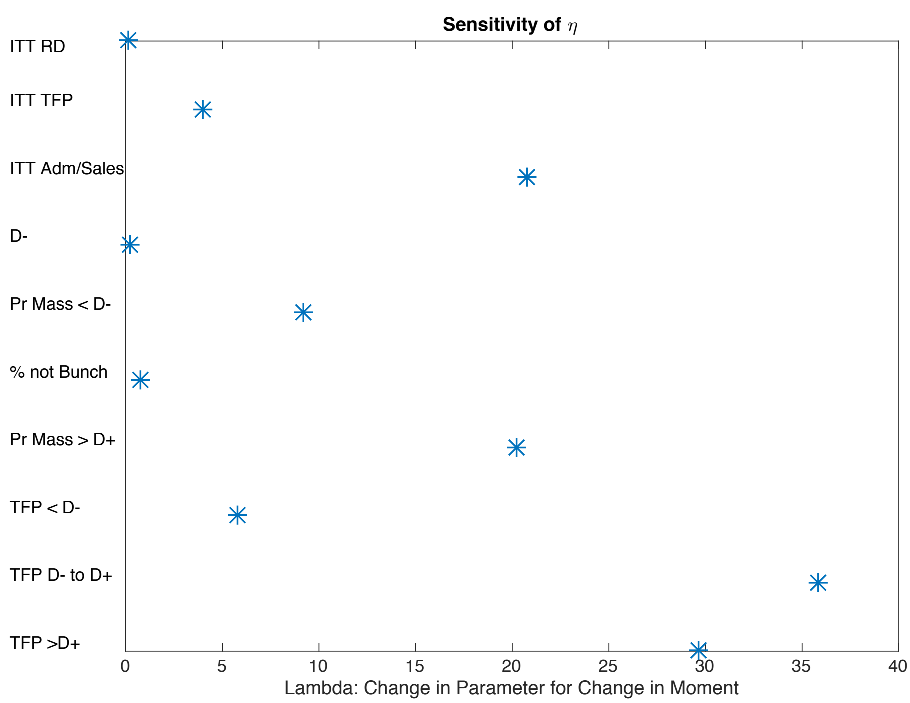


Figure A.3: Sensitivity Analysis

Panel (a) Sensitivity Analysis for ε



Panel (b) Sensitivity Analysis for η



Appendix Tables

Table A.1: Lack of Sales Manipulation around 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
Percentage Misreported			
(SE)			

Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 5 for details on the estimation. Standard errors in parentheses.

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

Table A.2: Estimates of Mis-categorized R&D

	(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
Percentage Misreported Relative to Notch α	.233**	.329***	.269***
(SE)	(.111)	(.093)	(.095)

Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 5 for details on the estimation. Standard errors in parentheses.

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

Table A.3: Alternative Estimates of Mis-categorized R&D

	(1) Low Sales	(2) Medium Sales	(3) High Sales
Structural Break	0.02	0.03**	0.05**
	(0.02)	(0.01)	(0.01)
N	4028	6461	7222
Mean Ratio Above α	0.47	0.45	0.51
Fraction Constrained: a^*	0.87	0.47	0.41
Percentage Evasion: δ^*	0.25	0.15	0.16

Standard errors in parentheses

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

Table A.4: Estimates of Mis-categorized R&D by Current Asset Ratio

	(1) Low	(2) Medium	(3) Large
(a) Low Current Asset Ratio			
Structural Break	-0.017**	-0.013***	-0.004
	(0.007)	(0.004)	(0.002)
Percentage Misreported	.278**	.326***	.117
(SE)	(.111)	(.088)	(.081)
(b) High Current Asset Ratio			
Structural Break	-0.020*	-0.013*	-0.011**
	(0.011)	(0.007)	(0.005)
Percentage Misreported	.328*	.318*	.375**
(SE)	(.181)	(.171)	(.166)

Standard errors in parentheses

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

Table A.5: Effects of R&D on Log TFP

	(1) All	(2) All	(3) Small	(4) Medium	(5) Large
Lagged Log TFP	0.735***	0.735***	0.724***	0.713***	0.738***
	(0.009)	(0.009)	(0.015)	(0.014)	(0.014)
100 X Log R&D	2.779***				
	(0.260)				
100 X Log R&D X Above Notch		2.510***	0.968***	1.503***	3.767***
		(0.232)	(0.355)	(0.320)	(0.397)
100 X Log R&D X Below Notch		2.809***	1.017**	1.681***	4.364***
		(0.263)	(0.408)	(0.373)	(0.454)
Observations	21,052	21,052	6,030	7,662	7,360
Implied $\delta^* = 1 - \frac{\beta_1}{\beta_2}$		.107***	.048	.106***	.137***
		(.008)	(.041)	(.027)	(.017)

Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 5 for details on the estimation. Industry X Year FE, standard errors in parentheses, clustered at Industry level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$$\hat{\phi_{it}} = \rho \hat{\phi_{it-1}} + \beta_1 \mathbb{I}[\text{Above}] \times \ln RD_{t-1} + \beta_2 \mathbb{I}[\text{Below}] \times \ln RD_{t-1} + u_{it}$$

Table A.6: Effects of R&D on Log TFP: Placebo with Fake Notch

	(1) All	(2) All	(3) Small	(4) Medium	(5) Large
Lagged Log TFP	0.716***	0.717***	0.705***	0.688***	0.726***
	(0.014)	(0.014)	(0.027)	(0.021)	(0.017)
100 X Log R&D	3.319***				
	(0.449)				
100 X Log R&D X Above Notch		3.280***	1.514*	3.518***	5.391***
		(0.433)	(0.827)	(0.591)	(0.579)
100 X Log R&D X Below Notch		3.315***	1.370*	3.779***	5.324***
		(0.444)	(0.793)	(0.687)	(0.656)
Observations	9,223	9,223	3,203	3,528	2,492
Implied $\delta^* = 1 - \frac{\beta_1}{\beta_2}$		.011	-.105	.069*	-.013
		(.016)	(.08)	(.041)	(.03)

Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 5 for details on the estimation. Industry X Year FE, standard errors in parentheses, clustered at Industry level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$$\hat{\phi_{it}} = \rho \hat{\phi_{it-1}} + \beta_1 \mathbb{I}[\text{Above}] \times \ln RD_{t-1} + \beta_2 \mathbb{I}[\text{Below}] \times \ln RD_{t-1} + u_{it}$$