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

December 2016

Abstract

Governments around the world encourage R&D investment based on the belief that economic growth is highly dependent on innovation. This paper analyzes the effects of a large fiscal incentive for R&D investment using a novel link between administrative and survey data of Chinese firms. The fiscal incentive is part of the InnoCom program, which awards a lower *average* corporate income tax rate to qualifying firms. The program generates a notch, or jump, in after-tax firm values since qualifying firms are required to maintain their ratio of R&D-to-sales above a given threshold. This sharp incentive varies over time and across firm characteristics. We exploit this policy variation to implement a cross-sectional "bunching" estimator that is novel in the R&D literature, to analyze potential evasion responses, and to estimate the effects of R&D on productivity. We find that this program led a large number of firms to locate at the qualifying R&D intensity threshold, implying a tax elasticity of R&D investment between 0.58 and 0.92, depending on firm size. We find that a substantial fraction of this response is due to tax evasion, and that accounting for evasion is crucial when estimating the effect of R&D on productivity.
JEL Codes: D24, O30, H25, H26.

*We are very grateful for comments from Manuel Adelino, Ashish Arora, Michael Best, Wesley Cohen, Dhammika Dharmapala, Michael Devereux, Rebecca Diamond, Damon Jones, Jill Popadak, Adriano Rampini, Nirupama Rao, Leslie Robinson, and Daniel Wilson, as well as seminar participants at Chicago Booth, Duke (Fuqua and Econ), FRB Philadelphia, Hong Kong University, LSE, National School of Development (PKU), NTA, Penn State, Stanford (SIEPR), UCSD, University of Melbourne, 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 often 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 (Ding and Li, 2015).[1] This paper analyzes the effects of one such effort: 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.

This paper analyzes quasi-experimental variation in the InnoCom program to answer three 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, how much do firms value an additional yuan of R&D investment in terms of future profits? Finally, what is the effect of fiscal incentives on firm-level productivity and aggregate productivity growth? 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 a 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. In addition, 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.

We overcome these concerns by leveraging an unusual and large fiscal incentive for R&D investment that is embedded in the Chinese corporate income tax. 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.

We first provide descriptive evidence that the R&D notches have significant effects on R&D intensity. 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 also find that firms adjust their R&D to the changes in the thresholds before and after 2008.

We then we develop a model of firm behavior where R&D investment depends on tax incen-

[1] Figure 1 shows the explosive growth in R&D intensity in China, relative to other selected countries.

tives, the effect of R&D on productivity, as well as heterogeneity in the profitability of R&D investment across the population of firms. As long as the profitability of R&D investment is smoothly distributed across the population, an R&D notch leads to excess bunching at the R&D notch relative to the decisions made under a tax system without a notch. 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. 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). The model then shows that the indifference condition of the marginal buncher is an implicit function for the elasticity of profits to R&D investment, and that this function depends solely on the estimated increase in R&D, and observed tax parameters. We then extend the model to incorporate several realistic features. Most importantly, we allow firms to have evasion motives and to potentially over-report their R&D. Finally, we design an empirical strategy that relies on firm's non-R&D expenditure data to detect and quantify evasion responses.

Our empirical results use the insights from the model to quantify firms' responses to the tax incentives. We first use the bunching estimator to quantify the percentage increase in R&D investment that is due to the tax incentive. We then analyze the potential for this response to be driven by evasion. We find significant evidence that a portion of the observed response is driven by mis-categorization of non-R&D expenses. On average, firms over-report $28\% - 52\%$ of their R&D expenditure, which translates into a large reduction in firm's real R&D responses. For instance, our bunching estimator shows that a marginal firm participating in the InnoCom program increased R&D by 58% to 92%, depending on the size of the firm. Taking the evasion response into account, however, the elasticities of real R&D investment range from 30% to 66%.

One advantage of our setting is that, in contrast to pure administrative data, the Chinese Survey of Manufacturing contains detailed information on factors of production at the firm level. This allows us to relate R&D spending to observable measures of productivity. Using a regression framework of firm level R&D on lagged R&D investment, we find that doubling R&D spending increases firm level productivity in the short-run by 2.8%. We also find that adjusting for misreported R&D is important when estimating the effect of R&D on productivity. In particular, firms that are not mis-reporting R&D have consistently larger effects of R&D on productivity. Using the attenuation in the real effect of R&D on productivity as a second measure of mischaracterization, we find slightly larger estimates of the real increase in R&D investment.

We use a new estimator for causal effects developed by Diamond and Persson (2016) to estimate the effects of the InnoCom policy on profitability and tax revenues. We find persistent and statistically significant effects of the InnoCom program on future productivity and profitability. While increases in profitability lessen the fiscal cost to the government, we find large fiscal costs of increasing productivity. In particular, raising productivity by 1% requires a reduction in corporate tax revenues of 5.3%. Finally, we recover the structural parameters of the model, including the long-run effect of R&D investment on profitability that is implied by the bunching and evasion responses. We find this "revealed preference" estimate of the productivity effects of R&D investment has a similar magnitude to our other estimates.

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 became increasingly available.[2] 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 the R&D tax incentive in a large emerging economy such as China. It is also one of the first few 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, methodologically, 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 and Kleven (2015). However, in contrast to this literature, the incentive generated by the notch targets a particular action, increasing R&D investment, as opposed to simply changing the rate of taxation. We exploit this feature of our setting to estimate treatment effects of the program on R&D investment, tax revenues, and growth in productivity using an estimator recently developed by Diamond and Persson (2016).

Third, 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), 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 firm's R&D bunching and relabeling behaviors.

The rest of the paper is organized as follows. Section 1 provides a description of the fiscal incentive for R&D investment. Section 2 discusses the data and Section 3 provides descriptive evidence of the effects of the tax incentive on R&D investment. 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, the implications for policy analysis, and how accounting for evasion affects estimates of the effects of R&D on firm-level productivity. Section 5 culminates with the estimation of the structural parameters of the model and Section 6 concludes.

[2]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.

[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)).

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%.[4] 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.[5]

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[6] and are part of state-encouraged industries enjoy a preferential tax rate of 15%. However, firms are required to raise at least 70% of their business revenue from this region in order to benefit from this program.[7]

This paper analyzes the "InnoCom" program which targets "high tech" enterprises (HTE). As part of this program, firms that are certified as HTEs qualify for 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 any other tax breaks unless they are located in western China. Prior to 2008, the certification process was administered by the local Ministry of Science and Technology, which establishes a long list of prerequisites. The most important determinants for certification are the following:

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 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.

[4] 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%.

[5] The effective tax rate of the 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 EIT rate for the subsequent 3 years.

[6] West China is defined as the provinces of Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Ningxia, Qinghai, Xinjiang, Inner Mongolia and Guangxi.

[7] 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.

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. Starting in January 2008, the base EIT was set at 25% for both DOEs and FOEs. 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.

In concert with this reform, the Ministry of Science and Technology streamlined the application process of the InnoCom program. The Ministry of Science and Technology teamed up with the Ministry of Finance and National Tax Bureau to modify the criteria of HTE certification process and linked the threshold requirement of R&D intensity to the size of firms in terms of total sales. The new, 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 for our study that allow us to overcome common problems that arise when estimating the effects of fiscal incentives on R&D investment. First, when estimating the effects of fiscal incentives on R&D investment, researchers often lack plausibly exogenous variation in fiscal incentives. One concern is that, as firms with better prospects for innovation are likely to invest more heavily in R&D, comparisons of investment and profitability across different firms will yield upward biases in the value of R&D investment to firms. The key feature of this program that allows us to overcome this problem is the change in the *average tax rate*, which creates a notch in the firm's after-tax value function. This notch generates sharp counterfactual predictions for the distribution of R&D intensity and allows us to use cross-sectional estimation methods such as those of Saez (2010) and Kleven and Waseem (2013) and eschew the concern of endogenous R&D investment. In addition, the 2008 reform generates firm-level variation in the fiscal incentive that can be exploited when estimating panel models.

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.

A final concern is the potential that the reported R&D investment might not represent a real change in investment but instead might be a form of tax avoidance or evasion. This concern is important when interpreting the reported elasticity of R&D investment as real activity and may loom large in the measurement of the effects of R&D investment on productivity. In particular, if reported R&D investment contains non-R&D expenses, the effects of R&D on productivity will be biased toward zero. To our knowledge, the current literature is not able to circumvent this problem.

We address this concern by devising a strategy to test for mis-reporting and to quantify the extent of the evasion response. This strategy leverages details of the institutional setting and of our administrative data. Given that firms must maintain the requirements above for a period of 3 years and that firms often use specialized consulting firms to ensure they satisfy the standards set by the Ministry of Science and Technology, it is unlikely that the bulk of this response will represent tax avoidance. In addition, since China relies on a VAT system with third-party reporting, it is hard for companies to report expenses that are not reported by third-party vendors. However, firms may be able to misreport non-R&D expenses to qualify for the program. The administrative tax data we employ contains detailed information on the breakdown of operating expenses and R&D expenses. This allows us to compare 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, relative to more tangible categories, such as capital investment.

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 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. 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 benchmark cross-sectional results focus on data from 2011, which allows for firms to respond to the phase-in of the tax changes.

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. 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. 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 R&D 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. In particular, we document stark bunching patterns precisely above tax notches. 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 nowhere else. Similarly, for firms with annual sales between 50 million and 200 million RMB, the bunching is at 4%, while for firms with more than 200 million RMB annual sales, the bunching is 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 data sample 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 kicks-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.[8] 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. In the pre-2008 data, unlike the DOEs that show a clear bunching at 5% of R&D intensity level, large FOEs have no clear pattern of bunching. 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.

Combined, these figures provide strong qualitative evidence that firms are responsive to the InnoCom program. Our quantitative analysis will focus on measuring the size of the change in R&D investment, analyzing the degree to which the response is due to evasion, 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 and discusses the empirical implementation of a model of R&D investment where firms respond to notches in the corporate income tax schedule in China. The objective of the model is two-fold. First, the model shows that we can recover a non-parametric estimate of the increase in R&D following the InnoCom program using a bunching estimator as in Saez (2010) and Kleven and Waseem (2013). Second, the model relates the bunching behavior to firm-level parameters of interest, such as the profitability effect of R&D investment.

We start with a simple model and develop extensions to allow for fixed costs of certification, adjustments costs of R&D investment, as well the possibility that the reported R&D response is partly due to evasion. Full details of the model are presented in Appendix A.

[8]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.

4.1 Model Setup

Consider a firm i with a Cobb-Douglas production function given by:

$$q_{it} = \exp\{\phi_{it}\} K_{it}^{\kappa} V_{it}^{1-\kappa},$$

where K_{it} and V_{it} are static inputs with respective prices p_k and p_v, 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. $N(0, \sigma^2)$.[9] 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:

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

Notice first that if the tax rate is constant across periods, the corporate income tax does not affect the choice of R&D investment.[10] 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, will be more responsive to tax incentives.

The choice of R&D depends on the 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:

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

[9] Note that an alternative formulation where we allow for capital to be fixed results in similar conclusions.

[10] This simple model eschews issues related to the user-cost of capital, as in Hall and Jorgenson (1967), and to issues related to the source of funds, as in Auerbach (1984).

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 6 presents two possible scenarios following this incentive. Panel (a) shows the situation where the firm finds it optimal to choose a level of R&D intensity below the threshold. At this choice, the first order condition of the linear tax case holds and the optimal level of R&D is given by Equation 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 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 1 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 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 now characterize the firm that is indifferent between the 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 2, we can manipulate $\Pi(D_{i1}^{*-}|t_2^{LT})$ to obtain:

$$\Pi(D_{i1}^{*-}|t_2^{LT}) = (1 - t_1)(\pi_1 - D_{i1}^{*-}) + \frac{(1 - t_1)}{(\theta - 1)\varepsilon} D_{i1}^{*-}. \tag{3}$$

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 2 to obtain:

$$\Pi(\alpha\theta\pi_1|t^{HT}) = (1 - t_1)(\pi_1 - \alpha\theta\pi_1) + \frac{(1 - t_1)D_{i1}^{*-}}{(\theta - 1)\varepsilon} \left(\frac{\alpha\theta\pi_1}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon} \left(\frac{1 - t_2^{HT}}{1 - t_2^{LT}}\right). \tag{4}$$

Comparing Equations 3 and 4 , we see that Equation 4 shows a larger cost of investment in the first period (since $D_{i1}^{*-} < \alpha\theta\pi_1$) and higher profits in the second period. Profits are higher by a factor of $\left(\frac{\alpha\theta\pi_1}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)$, which combines productivity effects as well as a tax benefit.

We use Equations 3-4 and the indifference condition that defines the marginal bunching firm to obtain a relation between the percentage difference in R&D intensity and $(\theta-1)\varepsilon$. Equating $\Pi(\alpha\theta\pi_1|t_2^{HT})$ and $\Pi(D_1^{*-}|t_2^{LT})$, dividing by $(1-t_1)\alpha\theta\pi_1$, and manipulating we obtain:

$$\frac{(1-\Delta D^*)^{1-(\theta-1)\varepsilon}}{(\theta-1)\varepsilon}\times\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)-1 \;=\; \left(\frac{1}{(\theta-1)\varepsilon}-1\right)(1-\Delta D^*), \tag{5}$$

where we define $\Delta D^* = \frac{\alpha\theta\pi_1 - D^{*-}}{\alpha\theta\pi_1}$ as the percentage increase in R&D spending due to the notch. The right-hand-side of this equation describes the profit from the internal optimum, relative to the after-tax profits in the first period. The left-hand-side describes the relative profits from bunching, which depend on productivity gains and tax gains, but which are lower by the additional cost of investment.

While there is no closed-form solution for Equation 5, this equation describes an implicit function between ΔD^* and $(\theta-1)\varepsilon$. In particular, given observable tax parameters t_2^{HT} and t_2^{LT}, and the empirical quantity ΔD^*, a solution to Equation 5 would result in an estimate of $(\theta-1)\varepsilon$. Figure A3 in the Appendix plots this function. Note that while we are not able to separately identify θ and ε, the quantity $(\theta-1)\varepsilon$ is informative as it represents the effect of R&D investment on future profitability. Before we discuss how we recover ΔD^* from aggregate bunching patterns, we first generalize the model to allow for evasion responses and costs of adjustment.

R&D Choice Under Tax Notch with Evasion

As discussed in Section 1, one mechanism driving the large bunching responses we observe might be manipulation of reported R&D investment. We now assume that firms may misreport their costs and shift non-RD costs to the R&D category.

We choose to focus on this form of evasion since the institutional setting limits other types of evasion. First, cases where all of the change in R&D is driven by evasion are ruled out by the requirement that firms obtain the InnoCom certification. In particular, part of this certification includes an audit of the firm's tax and financial standings. Second, China relies on a value-added-tax (VAT) system based on third-party reporting. As in other settings (e.g., Kleven et al. (2011)), third-party-reporting may limit a firm's ability to avoid taxation by reporting "phantom expenses."

The most likely form of evasion is the mis-categorization of administrative expenses as research expenses. These type of expenses are easily shifted and may be hard to identify in any given audit. Interviews with accountants and controllers of large Chinese firms reveal that this is the most likely source of evasion. 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 this reason, we analyze expense substitution of administrative expenses to R&D.

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:

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

where $\delta = \frac{\tilde{D}_1 - D_1}{\tilde{D}_1}$. We also assume that $\tilde{h}$ satisfies $\tilde{h}(0) = 0$ and $\tilde{h}'(\cdot) \geq 0$. In practice, we parametrize this function with a constant elasticity: $\tilde{h}(\delta) = \delta^\eta / \eta$.

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.

We characterize the firm that is indifferent between bunching and potentially misreporting, and not bunching. Figure 7 describes the intuition behind this choice. The firm that is willing to evade in order to reach the notch now has a lower internal solution that would be preferable to the firm than bunching if evasion were not possible. Because the firm gets positive returns from R&D investment and because increasing actual R&D investment lowers the cost of evasion, the firm increases its real investment to D^{*K}, which is such that $\alpha\theta\pi_1 \geq D^{*K} \geq D^{*-}$. At this point, the firm's choice is characterized by three conditions: the indifference condition, the first order condition of the internal solution, and the first order condition of the extent of evasion.

We now derive these conditions in our model. 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 $\tilde{D}_1$. 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}.

Consider first $\Pi(D_{i1}^{*-}, D_{i1}^{*-} | t_2^{LT})$. Since the firm does not mis-report in this case, Equation 3 still describes the profit in this case. Now manipulate $\Pi(\alpha\theta\pi_1, D_1^{*K} | t^{HT})$ using Equation 1 to obtain:

$$\begin{aligned} \Pi(\alpha\theta\pi_{i1}, D_{i1}^{*K} | t_2^{HT}) \quad &= \quad (1 - t_1)\left(\pi_{i1} - D_{i1}^{*K}\right) - h(D_1^{*K}, \alpha\theta\pi_1) \\ &\quad + \frac{(1 - t_1) D_{i1}^{*-}}{\varepsilon(\theta - 1)} \left(\frac{D_{i1}^{*K}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon} \frac{(1 - t_2^{HT})}{(1 - t_2^{LT})} \end{aligned} \tag{6}$$

As in the case without evasion, we equate and manipulate Equations 3 and 6 to obtain the following indifference condition:

$$\frac{1}{(\theta-1)\varepsilon}\left(\frac{1-\Delta D^*}{1-\delta^*}\right)^{1-(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) - 1 - \frac{(\delta^*)^\eta}{(1-\delta^*)(1-t_1)\eta} = \left(\frac{1}{(\theta-1)\varepsilon} - 1\right)\left(\frac{1-\Delta D^*}{1-\delta^*}\right). \tag{7}$$

Equations 7 and 5 are very similar. If $\delta^* = 0$, such that there is no evasion, these equations are identical. When $\delta^* > 0$ these equations differ by the cost of evasion, and by noting that the real increase in R&D is now given by $\frac{1-\Delta D^*}{1-\delta^*}$. This quantity is the relevant quantity when accounting for the effects of R&D on productivity.

In the case when the firm decides to bunch and evade, we have the additional information that D^K is chosen optimally. The first-order-condition of Equation 6 implies the following condition:

$$\left(\frac{1-\Delta D^*}{1-\delta^*}\right)^{1-(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) = 1 - \frac{(\delta^*)^{\eta-1}}{(1-t_1)}. \tag{8}$$

Equation 8 along with Equation 7 now form a system of two equations that are implicit functions for the parameters η and $(\theta-1)\varepsilon$. As we show in Appendix B, for a given set of tax parameters (t^H, t^L) and response margins $(\Delta D^*, \delta^*)$, Equations 7 and 8 can be rearranged as closed-form functions relating $(\theta-1)\varepsilon$ and η. The intersection of these functions identifies the parameters consistent with the data.

Fixed and Adjustment Costs

In addition to the possibility of evasion, we also allow for the possibility that firms face adjustment costs of investment and fixed costs of certification. We assume that the fixed cost is given by: $c \times \alpha\theta\pi_{1i}$. We allow quadratic adjustment costs governed by: $\frac{b\theta\pi_{1i}}{2}\left[\frac{D_i}{\theta\pi_{1i}}\right]^2$. Appendix A provides technical details and shows that the indifference condition of the marginal firms is now given by:

$$\left(\frac{1-\Delta D^*}{1-\delta^*}\right)^{1-(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) \times \left(\frac{1+\alpha b(1-\Delta D^*)}{(\theta-1)\varepsilon}\right) - 1 - \frac{c}{1-\delta^*} - \frac{\alpha b}{2} - \frac{(\delta^*)^{\eta}}{(1-\delta^*)(1-t_1)\eta} \tag{9}$$

$$= \left(\frac{1-\Delta D^*}{1-\delta^*}\right) \times \left[\left(\frac{1}{(\theta-1)\varepsilon} - 1\right) + \alpha b\left(\frac{1}{(\theta-1)\varepsilon} - \frac{1}{2}\right)(1-\Delta D^*)\right]. \tag{10}$$

Similarly, the condition from the first order condition of evasion is now:

$$\left(\frac{1-\Delta D^*}{1-\delta^*}\right)^{1-(\theta-1)\varepsilon} \times \left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right) = \frac{1+\alpha b(1-\delta^*) - \frac{(\delta^*)^{\eta-1}}{1-t_1}}{1+\alpha b(1-\Delta D^*)}. \tag{11}$$

The parameters $(\theta-1)\varepsilon, \eta, b$, and c are now under-identified by a single pair of data $(\Delta D^*, \delta^*)$. In practice, we calibrate the parameters b and c or impose restrictions across different size groups in order to identify all of these parameters. Appendix B discusses the identification of the structural parameters in this case.

4.2 Empirical Implementation of the Bunching Estimator

We now describe how we obtain estimates of the percentage change in R&D investment ΔD^* from the distributional patterns described in Section 3. Figure 8 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 with homogenous values of $(\theta-1)\varepsilon$. In this case, there is a range of R&D intensity levels that is dominated by the threshold

α. Firms with an internal solution in this rage will opt to bunch at the notch, which generates the bunching patterns. Define this excess mass at the notch, relative to the counterfactual distribution, as B. To understand the empirical content underlying this bunching prediction, recall that firms with higher valuations of R&D investment will be more responsive to the fiscal incentive, and will be more likely to bunch at the notch.

We can now relate the percentage change in R&D investment ΔD^* to the quantities B and $h_0(\alpha)$ by noting that:

$$B = \int_{\frac{D_{i1}^{-*}}{\theta\pi_1}}^{\alpha} h_0\left(\frac{D_{i1}}{\theta\pi_1}\right) d\frac{D_{i1}}{\theta\pi_1} \approx h_0(\alpha)\left(\alpha - \frac{D_{i1}^{-*}}{\theta\pi_1}\right) = h_0(\alpha)\alpha\Delta D^*. \tag{12}$$

The first part of Equation 12 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 $\left(\frac{D_{i1}^{-*}}{\theta\pi_1}, \alpha\right)$. The second part of Equation 12 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:

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

Thus, in order to estimate ΔD^*, and subsequently $(\theta - 1)\varepsilon$, it suffices to have an estimate of the counterfactual density $h_0(\cdot)$, and to use this to recover the quantities B and $h_0(\alpha)$. Prior to detailing the estimation of this counterfactual density, however, we first discuss a generalized version of the model that allows for frictions as well as heterogeneous values of $(\theta - 1)\varepsilon$.

The prediction in Panel (a) of Figure 8 is quite stark in that no firms are expected to locate in the dominated interval. However, it is possible that some of the firms in our sample are not able to increase their R&D investment due to some friction, such as a credit constraint. We follow Kleven and Waseem (2013) in allowing for this possibility. Denote by $a\left(\frac{D_{i1}}{\theta\pi_1}\right)$ the fraction of firms that are constrained in increasing their investment past $\frac{D_{i1}}{\theta\pi_1}$. This fraction can be computed easily by comparing the fraction of firms that are close to the notch but are not bunching to the counterfactual density $h_0(\alpha)$. Recovering ΔD^* with this modification requires a slight modification to Equation 12 by replacing $h_0\left(\frac{D_{i1}}{\theta\pi_1}\right)$ with $\left(1 - a\left(\frac{D_{i1}}{\theta\pi_1}\right)\right) h_0\left(\frac{D_{i1}}{\pi_1}\right)$.

We can further generalize this setup by allowing for heterogenous values of $(\theta - 1)\varepsilon$ across firms. Variation in these parameters will affect the fraction of firms that are constrained as well as the expected change in R&D investment, which we now denote as $a\left(\frac{D_{i1}}{\theta\pi_1}, (\theta - 1)\varepsilon\right)$ and $\Delta D^*_{(\theta-1)\varepsilon}$, respectively. Panel (b) of Figure 8 describes graphically how allowing for this degree of heterogeneity, in addition to frictions, affects the predicted bunching pattern. Equation 12 can now be generalized as follows:

$$
\begin{aligned}
B &= \int \int_{\frac{D_{i1}^{-*}}{\theta\pi_1}}^{\alpha} \left(1 - a\left(\frac{D_{i1}}{\theta\pi_1}, (\theta-1)\varepsilon\right)\right) \tilde{h}_0\left(\frac{D_{i1}}{\theta\pi_1}, (\theta-1)\varepsilon\right) d\frac{D_{i1}}{\theta\pi_1} d(\theta-1)\varepsilon \\
&\approx h_0(\alpha)\alpha\mathbb{E}[\Delta D^*_{(\theta-1)\varepsilon}]\left(1 - \mathbb{E}\left[a\left(\frac{D_{i1}}{\theta\pi_1}, (\theta-1)\varepsilon\right)\right]\right),
\end{aligned}
$$

where $\tilde{h}_0\left(\frac{D}{\theta\pi_1}, (\theta-1)\varepsilon\right)$ is the joint distribution of R&D intensity and the parameters $(\theta-1)\varepsilon$, and where $\mathbb{E}[\Delta D^*_{(\theta-1)\varepsilon}]$ now denotes the average percentage change in R&D intensity.

We now address 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:

$$
c_j = \sum_{k=0}^{p} \beta_k \cdot (d_j)^k + \gamma_j \cdot \mathbf{1}\left[\frac{D_1^{-*}}{\theta\pi_1} \leq d_j \leq \frac{D_1^{+*}}{\theta\pi_1}\right] + \nu_j
$$

where c_j is the count of firms in the bin corresponding to R&D intensity level $\frac{D_{j1}}{\theta\pi_1}$ and where $\left(\frac{D_1^{-*}}{\theta\pi_1}, \frac{D_1^{+*}}{\theta\pi_1}\right)$ 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. We also follow Diamond and Persson (2016) in using a data-based approach to selecting the excluded region (i.e., $\frac{D_1^{-*}}{\theta\pi_1}$ and $\frac{D_1^{+*}}{\theta\pi_1}$) and degree of the polynomial, p. In particular, we use K-fold cross-validation (K=5) to evaluate the fit of a range of values for these three parameters.

An estimate for $h_0\left(\frac{D_{j1}}{\theta\pi_1}\right)$ is now given by $\hat{c}_j = \sum_{i=0}^{p} \hat{\beta}_i \cdot \left(\frac{D_{j1}}{\theta\pi_1}\right)^i$. Similarly, we obtain a counterfactual estimate for $h_0(\alpha)$ and B as follows:

$$
\widehat{h_0(\alpha)} = \sum_{k=0}^{p} \hat{\beta}_k \cdot (\alpha)^k \quad \text{and} \quad \hat{B} = \sum_{d_j = \frac{D_1^{-*}}{\theta\pi_1}}^{\alpha} \sum_{k=0}^{p} \hat{\beta}_k \cdot (d_j)^k.
$$

Finally, an estimate of the fraction of constrained firms is given by:

$$
\mathbb{E}\left[a\widehat{\left(\frac{D_{i1}}{\theta\pi_1}, (\theta-1)\varepsilon\right)}\right] = \frac{\hat{\gamma}_{\alpha^-}}{\sum_{k=0}^{p} \hat{\beta}_k \cdot (\alpha^-)^k},
$$

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.[11] Standard errors for the regression estimates, and for estimates of the counterfactual density can be computed by bootstrapping the residuals and re-estimating the coefficients.

[11] This "money-burning" point is easy to compute given an estimate of θ. In this case, the tax benefit is given by $(t^L - t^H)\pi_1$ and the cost of jumping to the notch is $\theta\pi_1(\alpha - \alpha^-)$, which implies that $\alpha^- = \alpha - (t^L - t^H) \times (1/\theta)$.

4.3 Quantifying Evasion

As mentioned above, the total response ΔD^* may include real investment as well as misreporting. Importantly, since China's SAT also collects value-added tax, it keeps records of transaction invoices between a given firm and its third-party business partners. For this reason, it is very hard, if not impossible, for firms to completely make up "phantom" R&D expenses. In addition, the Chinese State Administration of Tax, together with the Ministry of Science and Technology, conducts regular auditing of the InnoCom HTE firms, which likely eliminates the possibility for all-out evasion.

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.[12] 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. Our empirical strategy exploits additional cost data at the firm level in the SAT dataset in order to quantify the extent of evasion by analyzing the potential for firms to mis-report R&D expenses.

We test for this possibility by studying whether we can uncover a structural break, or discontinuity, in the non-R&D admin expense-to-sales ratio around the notch α. This strategy is similar to those of Hilber and Lyytikäinen (2013) and Bachas and Soto (2015). We implement this strategy by estimating a series regression in the neighborhood to the left and right of the threshold to identify the magnitude of the jump.

To relate this empirical strategy to our model, recall that the fraction of R&D that is evaded is given by:

$$\frac{\alpha\theta\pi_1 - D^{K*}}{\alpha\theta\pi_1} = \delta^*.$$

As our model indicates, if firms indeed over-report part of their R&D $\alpha\theta\pi_1 - D^{K*}$ to reach the threshold for R&D intensity, then one will expect that the non-R&D administrative expense-to-sales ratio will take a downward jump around α. Let $\beta^{Evasion}$ be the estimate of the structural break or jump. If we attribute this jump to evasion, then we can recover

$$\delta^* = \frac{-\beta^{Evasion}}{\alpha}.$$

In Section 5.3 we describe an alternative method of estimating δ^* that relies on the attenuation of the estimated effects of R&D on productivity. Armed with estimates of ΔD^* and δ^*, we can assess the fraction of the bunching response that is due to evasion, and estimate the real effect of the policy on R&D investment. Additionally, we can implement our model and estimate the structural parameters $(\theta - 1)\varepsilon$ and η.

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

5 Effects of R&D Notches on Investment, Evasion, and Productivity

This section discusses the results of our analyses. Section 5.1 estimates the investment response from the bunching estimator. Section 5.2 presents evidence of evasion, and Section 5.3 shows that accounting for evasion results in larger estimates of the effects of R&D on productivity. Section 5.4 discusses reduced-form policy elasticities from these estimates. Section 5.5 presents estimates of treatment effects on productivity and tax revenues, and Section 5.6 uses the moment conditions from the model to recover productivity effects of R&D from firms' revealed investment choices.

5.1 Bunching Estimates of Investment Response

As mentioned above, implementing the bunching estimator requires choosing the degree of the polynomial and selecting the excluded region. Our cross-validation procedure searches over values of $p < 7$, and all possible discrete values of $\frac{D_1^{-*}}{\theta\pi_1} < \alpha$ and $\frac{D_1^{+*}}{\theta\pi_1} > \alpha$ 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.[13] We obtain standard errors by bootstrapping the residuals from the series regression, generating 500 replicates of the data, and re-estimating the parameters.

Figure 9 displays the results of the bunching estimator for the three different notches after 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 implied values of ΔD^*, the fraction of firms that are constrained at the notch point, and the p-value of the test that the missing mass and the excess mass are of the same magnitude.[14]

Panel (a) shows an increase in R&D intensity of 58% as a fraction of the notch. 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 58% 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

[13] 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.

[14] In order to calculate the fraction of firms that is constrained, we use the average of the net profitability ratio in our data of 7%. This implies that firms in the range $(\alpha - .07 \times (t^L - t^H), \alpha)$ are not able to respond to the incentives of the InnoCom program. Table A4 presents alternative estimates that assume $\theta = 5$, which implies a net profitability ratio of 20%.

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. We find similar increases in R&D intensity of 63% and 92%, 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 37% and 33%, 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.

While ΔD^* represents a percentage increase in R&D, we can also obtain an estimate of the increase in R&D as a fraction of sales by multiplying ΔD^* by α. This implies an increase of 2.5% and 2.76% of revenue for medium and large firms, respectively. Finally, 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.

5.2 Estimates of Evasion Response

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 10 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.[15]

In Table 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. Using the formula $\delta^* = \frac{-\beta^{Evasion}}{\alpha}$ to adjust for each group's thresholds 3%, 4%, and 5%, we find that the fraction of R&D misreported δ^* is on average 0.233 for large sales firms, 0.329 for medium sales firms, and 0.269 for small sales firms.

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 discon-

[15]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.

tinuous increases in this ratio at the notch. This is confirmed in Table A1 and in Figure A2. We also explore the degree to which evasion is related to firm liquidity. In Table A2, 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 is larger for small and large firms with high current asset ratios but is not noticeably different for medium firms.

5.3 Productivity

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.

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

Before we estimate the equation above, we describe 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:

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

With these measures of firm-level productivity, we estimate the regression of firm productivity evolution including the relation between log R&D and log TFP.

Table 3 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: $\delta^* = 1 - \frac{\beta_1}{\beta_2}$. 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 A3 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.

5.4 Policy Elasticities

We now combine the estimates from our bunching and evasion analyses to construct reduced-form policy elasticities. From a policy evaluation perspective, policy-makers may be interested in different aspects of these elasticities.

In Table 4, we explore three measures of the response to the InnoCom program: the overall effect of the policy on reported R&D, the real response of the compliers, and the population increase in real R&D intensity. The first two column of Table 4 report the bunching elasticities ΔD^* and fraction of constrained firms a^* from Figure 9. Column (3) combines these estimates to produce the population increase in reported R&D, which results in a substantial decrease from the response by the complies. We can also derive the real percentage increase in R&D by subtracting our two measures of evasion. Columns (4) and (5) report estimates of δ and Columns (6) and (7) report estimates of real increases in R&D investment by complier firms. These estimates are diminished by as much as 52%. Nonetheless, they represent a substantial increase in real investment. Finally, Columns (8) and (9) report the real population increase in R&D intensity. For small and medium firms, this increase is between 0.66 and 1.4 percentage points. For large firms, this increase is between 1.2 and 1.4 percentage points.

Two cautions are warranted when using these numbers for policy analysis. First, while understanding the behavior of firms of different sizes is interesting from an economic point of view, policy makers may be interested in the aggregate increase in R&D across the economy.

Figure A1 shows that the vast majority of R&D is conducted by firms in the large sales category. It thus makes sense to focus on these firms when when mapping these estimates to the patterns in Figure 1. Second, our estimates of evasion from sales may result in a lower bound on the real R&D response to the degree that the structural break is partly due to behavioral responses and not to mis-reporting. From this perspective, the estimates of relabeling form real effects on productivity may be preferable.

5.5 Causal Estimates on Productivity and Tax Collections

We now use an estimator of causal effects developed by Diamond and Persson (2016) to estimate the effects of the InnoCom program on productivity and on fiscal costs. The estimator has three components and takes the form:

$$ITT = \frac{1}{N^{Excluded}} \sum_{i \in (D^{*-}, D^{*+})} Y_i - \int_{D^{*-}}^{D^{*+}} \hat{h}_0(r) \widehat{E[Y|r, \text{No Notch}]} dr \tag{13}$$

The first quantity in this equation is the observed average value of a given outcome Y_i in the excluded region. The estimator recovers a counterfactual value for this average by combining the estimated counterfactual density of R&D intensity $\hat{h}_0(\cdot)$, as in Section 5.1, with an estimated average value of the outcome conditional on a given value of R&D, which is given by $\widehat{E[Y|r, \text{No Notch}]}$.

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

In order to recover $\widehat{E[Y|r, \text{No Notch}]}$, we estimate a flexible polynomial regression of Y on R&D intensity over the same excluded region used to estimate $\hat{h}_0(\cdot)$. Figure 11 presents a visual example. The first panel shows the counterfactual density of R&D in 2009 for firms in the large sales group. The second panel estimates a cubic regression of the profit ratio in 2011 on R&D intensity in 2009 over the same excluded region as the density above. The integral in Equation 13 combines the two blue lines to form an estimate of the counterfactual average of the profit ratio, assuming there is no notch.

Table 5 presents estimates of the ITT of the InnoCom program on several outcomes. This table focuses on large firms and reports estimates of treatment effects for outcomes in 2011, given the the excluded region of R&D intensity in 2009. Following Figure 11, we find an increase in the profit ratio of 1.6% that is statistically significant at the 5% level. We find a similar increase in TFP and a marginally significant increase in the investment to capital ratio. While the statutory decline in tax rates is 10%, we see tax collections decreased only by 8.4%, which is commensurate with an increase in profits of 1.6%.

The second panel of Table 5 presents estimates of ratios of the estimates in the first panel. The first row reports that for a 1% increase in R&D investment, there is also a .06% increase in the profit ratio between 2009 and 2011. 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.[16] From the point of view of the government, it is useful, however, to calculate the fiscal cost of encouraging R&D investment, and increasing productivity. Table 5 shows that doubling R&D investment would cost the government 30% of corporate tax revenues. Similarly, we find that increasing TFP by 1% would cost the government a reduction of 5.3% in corporate tax revenues.

5.6 Implied Structural Parameters

Our final set of results implements the model in Section 4 to map the reduced-form estimates of bunching (ΔD^*) and evasion (δ^*) to the structural parameters $(\theta-1)\varepsilon$, the elasticity of profitability with respect to R&D, and η, the elasticity of evasion costs. Recall that our model describes the optimizing behavior of the firm with two equations: the indifference condition between bunching and not bunching, and the optimal evasion response conditional on bunching.

We implement this model with a method of moments approach. For a given group k, let h_k^B denote the moment formed from the bunching equation, and h_k^E denote the moment condition from the evasion choice.

These two conditions can be expressed as two closed-form solutions relating $(\theta-1)\varepsilon$ and η conditional on data from a given group k: $(\Delta D_k^*, \delta_k^*)$.[17] The intersection of these relations identifies the set of parameters that are identified by the data. Figure A4 plots these relations and Section B discusses the identification of these parameters in detail.

In practice, the simple model without adjustment or fixed costs is not able to fit the data well. For this reason, we use the version of the model that allows for evasion, adjustment costs, and fixed costs of certification. The bunching and evasion equations correspond to Equations 10 and 11, respectively. These equations now depend on four parameters $(\theta-1)\varepsilon, \eta, b$, and c. Section B discusses how these parameters may be identified through restrictions on structural parameters across different groups. Alternatively, our current estimates calibrate the cost parameters b and c in order to fit the data and allow for heterogeneous elasticities across groups.

To implement the method of moments estimator, we form the criterion function:

$$Q((\theta-1)\varepsilon,\eta;b,c) = \begin{bmatrix} h_1^B((\theta-1)\varepsilon,\eta;b,c) \\ h_1^E((\theta-1)\varepsilon,\eta;b,c) \\ \vdots \\ h_K^B((\theta-1)\varepsilon,\eta;b,c) \\ h_K^E((\theta-1)\varepsilon,\eta;b,c) \end{bmatrix}' W \begin{bmatrix} h_1^B((\theta-1)\varepsilon,\eta;b,c) \\ h_1^E((\theta-1)\varepsilon,\eta;b,c) \\ \vdots \\ h_K^B((\theta-1)\varepsilon,\eta;b,c) \\ h_K^E((\theta-1)\varepsilon,\eta;b,c) \end{bmatrix},$$

where W is a weighting matrix. In a first step, we set W equal to the identity matrix and minimize this function with respect to $((\theta-1)\varepsilon,\eta)$ using bootstrapped version of the estimates $\{(\Delta D_{k,b}^*, \delta_{k,b}^*)\}$ for the three groups k and for $b = 1, ...500$. Second, for each bootstrapped version

[16] See Jones (2015) for a useful exposition of the economics of such restrictions. Even from the point of view of the effects of R&D investment, this elasticity would also need to be adjusted for the fact that the program elicits persistent changes in investment, as opposed to the static elasticities that are usually reported in the literature.

[17] We use the estimate of relabeling from the sales analysis in Section 5.2.

of the data, we evaluate the six moment conditions and compute the variance of these empirical moments. Third, we set W equal to the inverse variance of the empirical moments and minimize the function above to obtain estimates of $((\theta-1)\varepsilon, \eta)$. We compute standard errors by repeating the third step on the bootstrapped versions of the data.

Table 6 reports estimates of $((\theta-1)\varepsilon, \eta)$ for calibrated values of $b = 3$ and $c = 0.5$. Panel (a) restricts the parameters to be constant for all groups. Consider the estimate for $(\theta-1)\varepsilon$. Our data suggest an average value of the net profit ratio of 7%, which corresponds to a value of $\theta = 14.5$. The estimate from Panel (a) then implies an estimate of ε of 0.134. Since the InnoCom program requires that firms commit to a permanent increase in R&D, the interpretation of this coefficient is that of a long-run effect. As we mention in Section 4, a transitory elasticity is likely to be closer to $(1-\rho)\varepsilon$, which, for $\rho = 0.74$ as estimated in Table 3, implies a short-run elasticity of R&D on productivity of around 0.035. It is reassuring to find that our model delivers a "revealed preference" value of the effect of R&D on productivity that is quantitatively very close to those reported in Table 3. Panel (b) allows this elasticity to vary across the three groups. The average elasticity is closest to the elasticity for large firms, while smaller firms have a larger elasticity and medium firms a smaller one. Figures A6 and A7 explore the sensitivity of these estimates with respect to the calibrated values b and c. These figures show that our estimate of ε is robust to these calibrations.[18]

6 Conclusions

In this paper, we exploit the tax notch generated by the Chinese InnoCom program to identify the response of firm-level R&D investment to tax incentives. We further estimate the impact of R&D on real economic outcomes, such as productivity, thus connecting the tax incentives in the InnoCom program to the ultimate objective of increasing productivity and economic growth.

The Chinese institutional context provides both rich cross-sectional and panel variation in firms' incentives for R&D investment. Our descriptive results provide strong evidence for the existence of these incentives. Analyses of cross-sectional bunching patterns of small, medium, and large firms, as well changes before and after 2008 are consistent with significant effects of tax incentives on R&D investment. Zooming into particular subgroups of our samples, including foreign owned enterprises and domestic small firms, illustrates that their dynamic behavioral responses are also consistent with the corporate tax reform of 2008.

We then motivate our empirical quantitative analysis using a two-period model of heterogeneous firm R&D investment. As in standard models of knowledge capital accumulation, firms invest to improve their future productivity. However, given the tax notch introduced by the InnoCom program, R&D serves a dual role of moving firms towards a regime with a lower average tax rate on their profit. Our model provides a tight connection between firm level R&D investment, its initial heterogeneity, and the taxes below and above the notch. We show that

[18] However, the estimate of η is more sensitive to assumptions about b and c. This is because our model can rationalize the data by either assuming an increase in evasion costs, which would lower bunching, or an increase in adjustment or fixed costs. Thus, whether a reform would lead to additional evasion or to non-response due to adjustment costs is an open question

our benchmark model motivates a simple bunching estimator that can identify the magnitude of R&D response to InnoCom policy by the marginal firm. We also allow firms to respond to the tax incentive through evasion, such that they may over-report their R&D. We rely on knowledge of the institutional setting obtained through interviews with officials and executives to focus our empirical analysis on the potential for firms to mis-categorize administrative expenses into R&D. This motivates an empirical strategy that tests for a discontinuity in administrative expenses around the bunching threshold to detect evasion.

Our quantitative analyses generate interesting and sensible results. We find that firm's evasion motive is not ignorable: on average, firms over-report up to 52% of their R&D expenditure. This translates into a significant reduction in terms of firms' real responses to the tax incentives. Nonetheless, we find significant effects on real R&D investment that have positive effects on productivity and profitability. Overall, the additional R&D investment following the incentives of the InnoCom program may yield long-term benefits through increased productivity growth. However, we find sizable fiscal costs of increasing R&D investment and productivity, which implies that this benefit comes at the heavy cost to the government.

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, Warwick 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.

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.

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.

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.

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

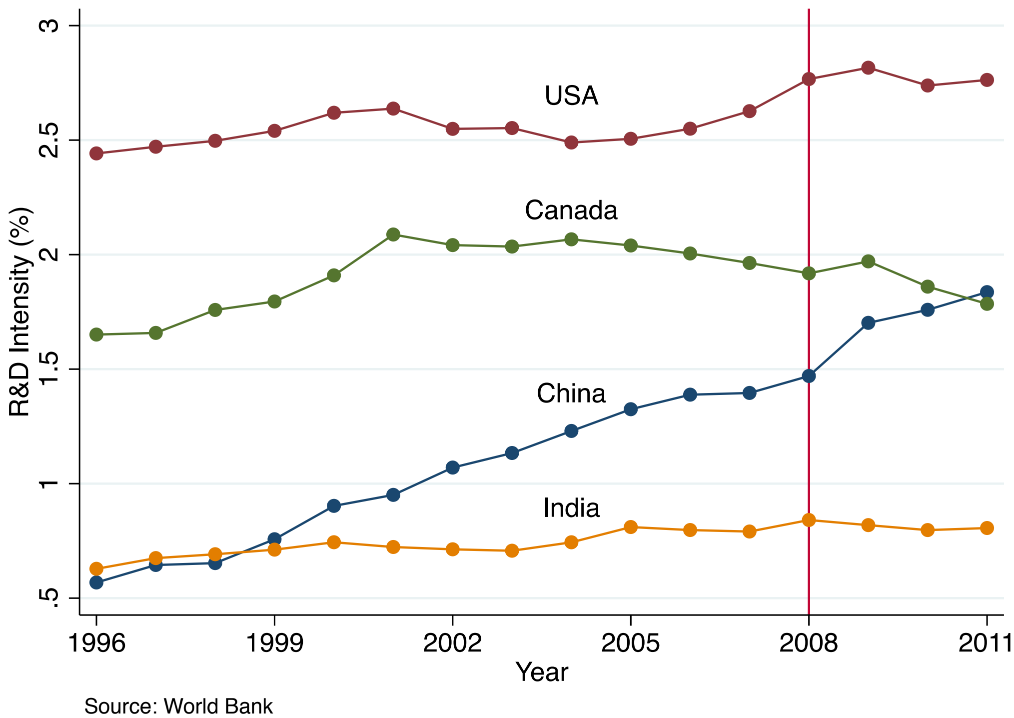
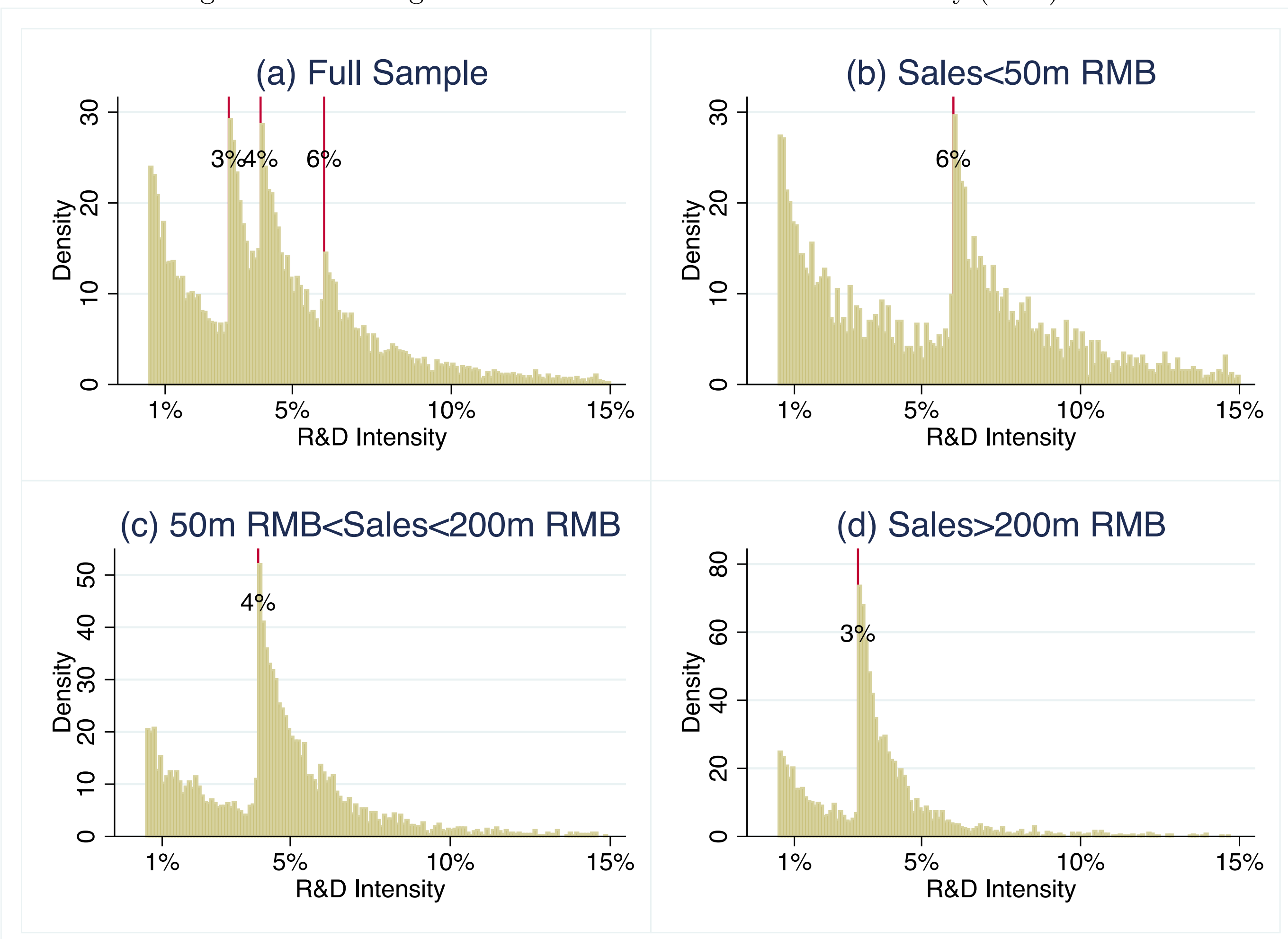
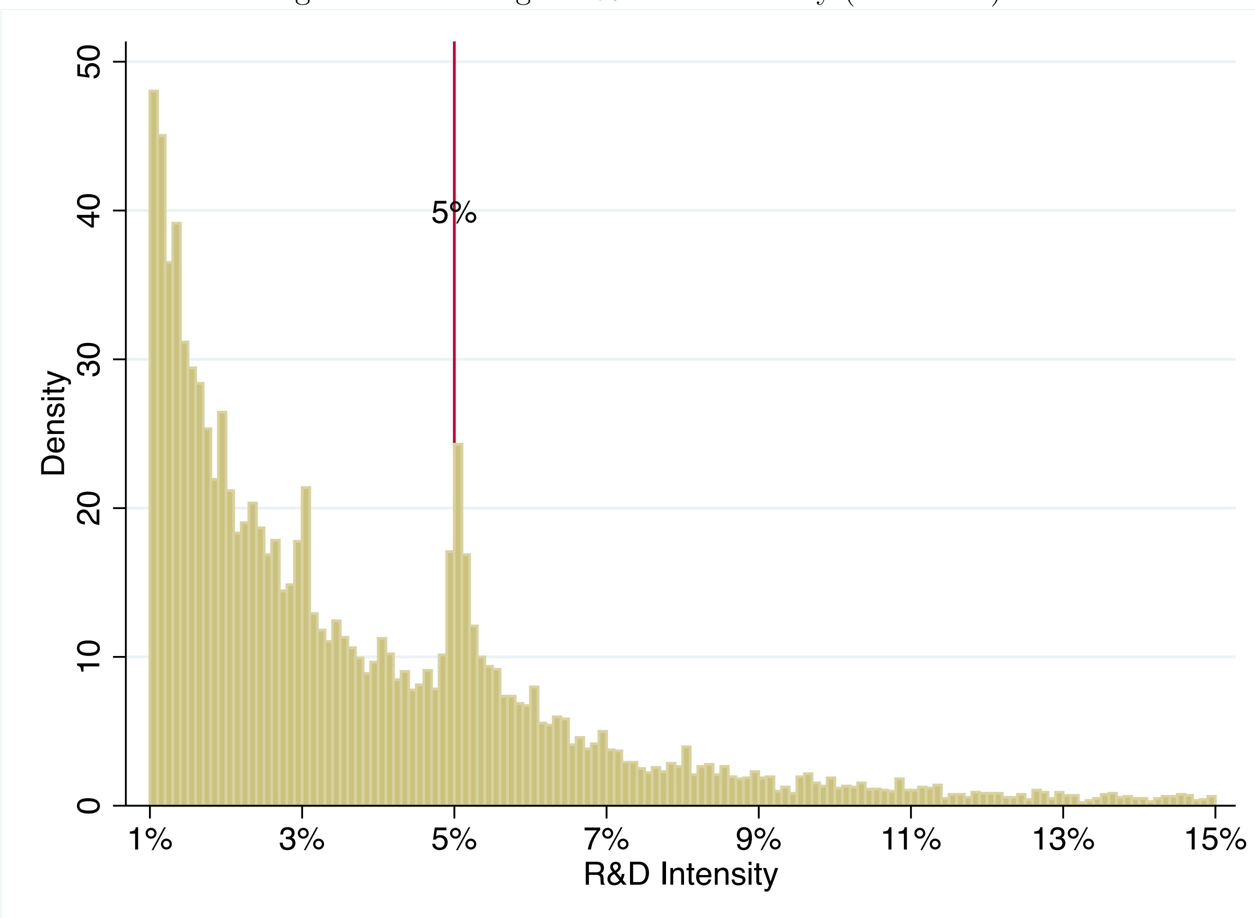


Figure 2: Bunching at Different Thresholds of R&D Intensity (2011)



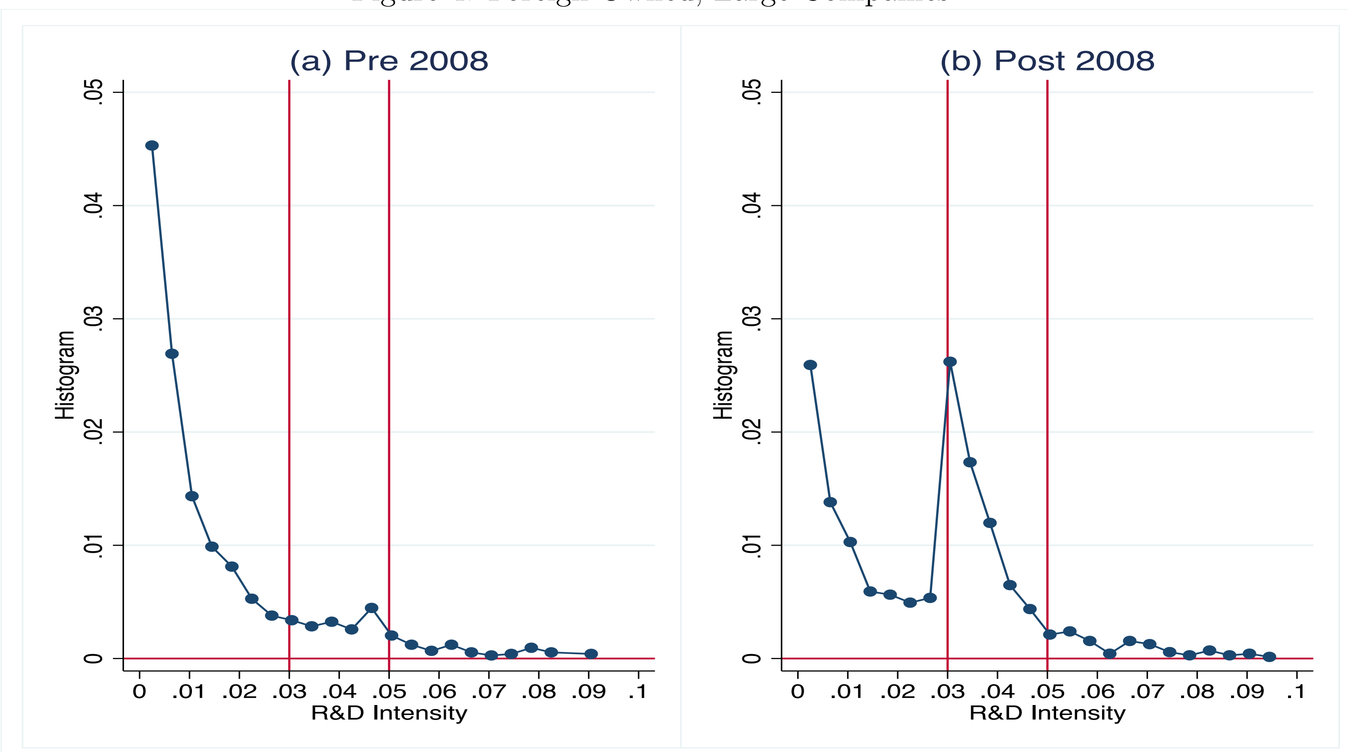
Source: Administrative Tax Return Database. See Section 2 for details.

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



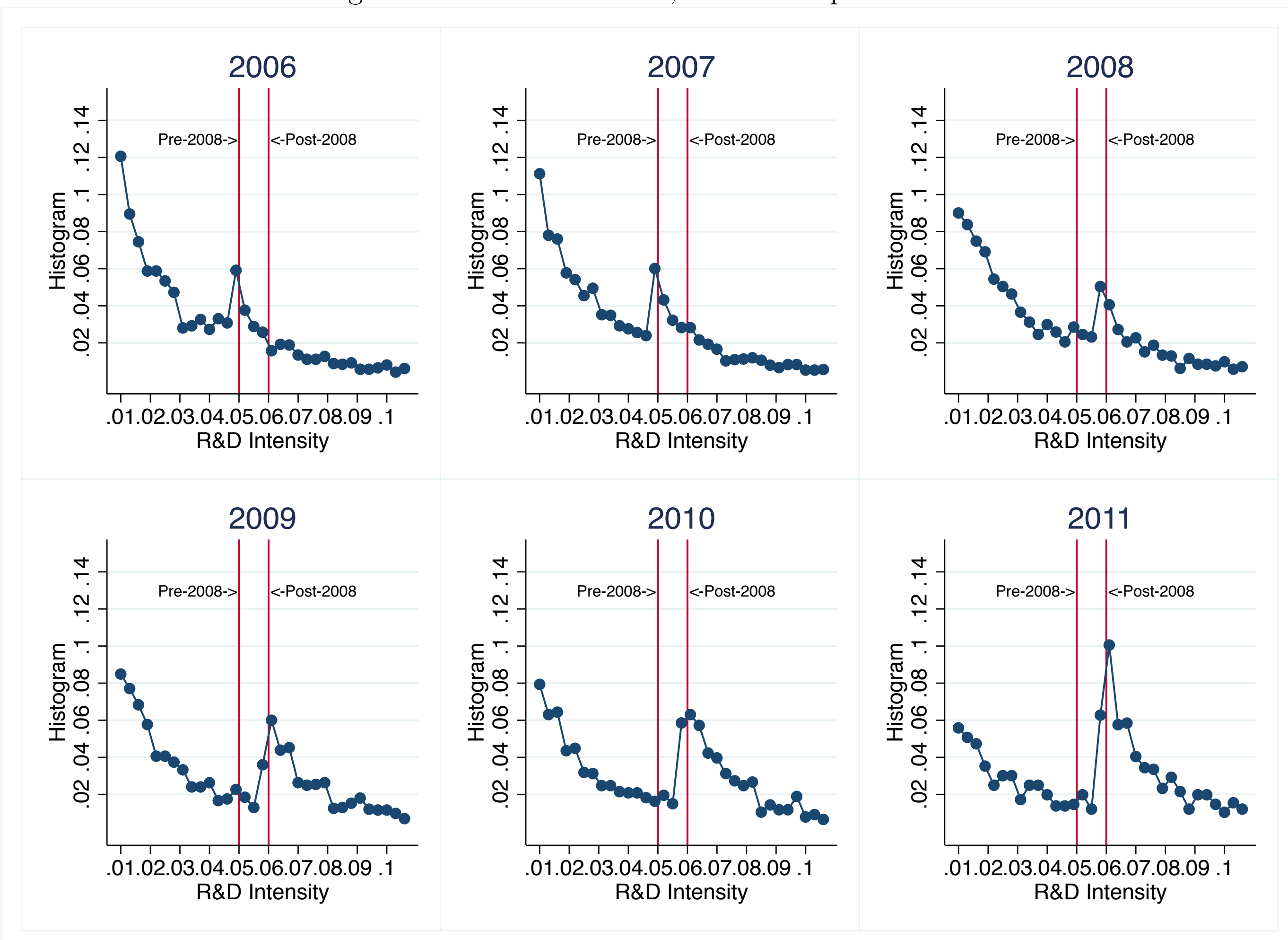
Source: Annual Survey of Manufacturers. See Section 2 for details.

Figure 4: Foreign-Owned, Large Companies



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

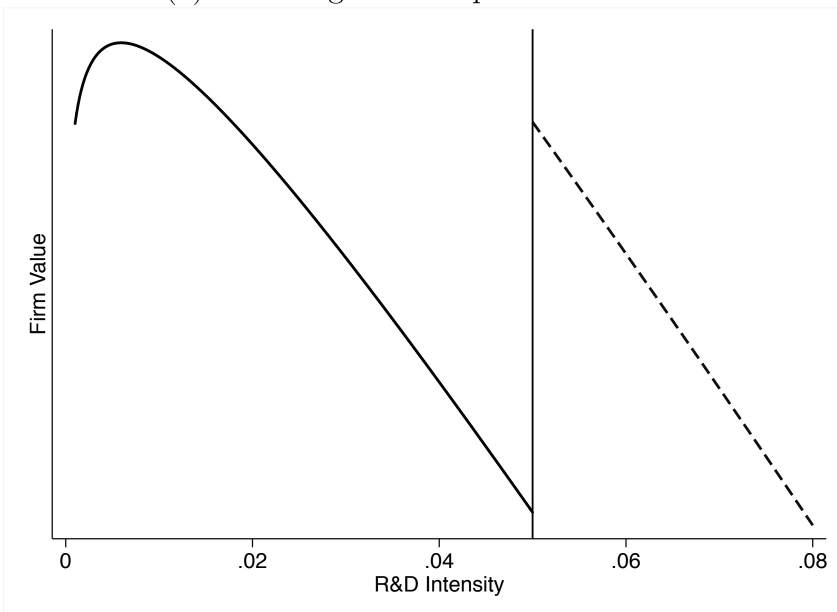
Figure 5: Domestic-Owned, Small Companies



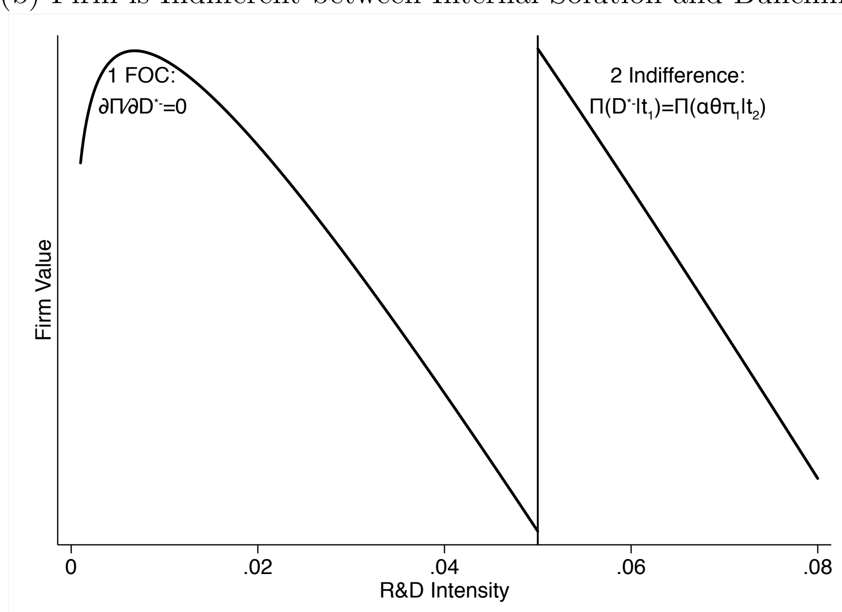
Source: Administrative Tax Return Database and Annual Survey of Manufacturers.
See Section 2 for details.

Figure 6: Induced Notch in Profit Functions

(a) Bunching is Sub-Optimal for Firm

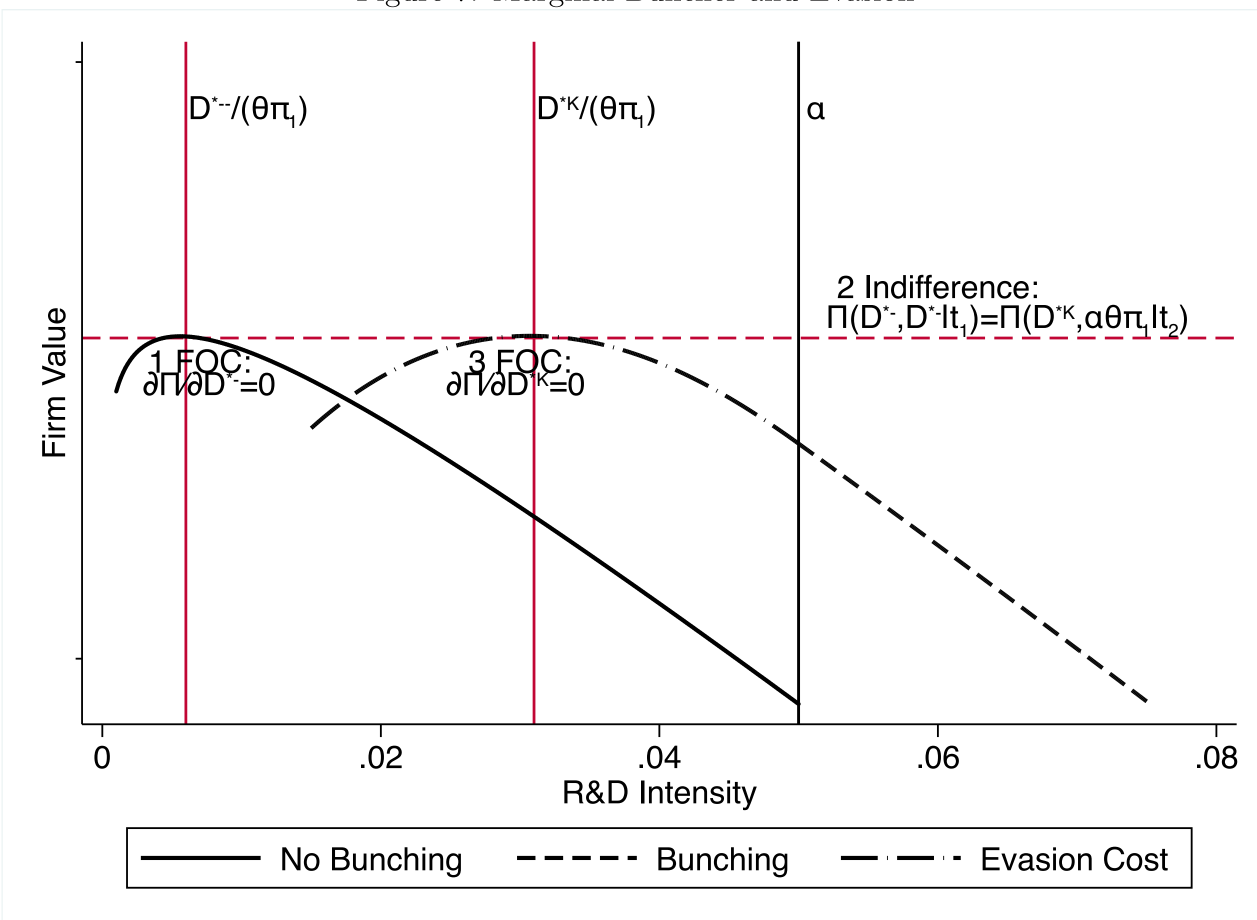


(b) Firm is Indifferent between Internal Solution and Bunching



Notes: See Section 4 for details.

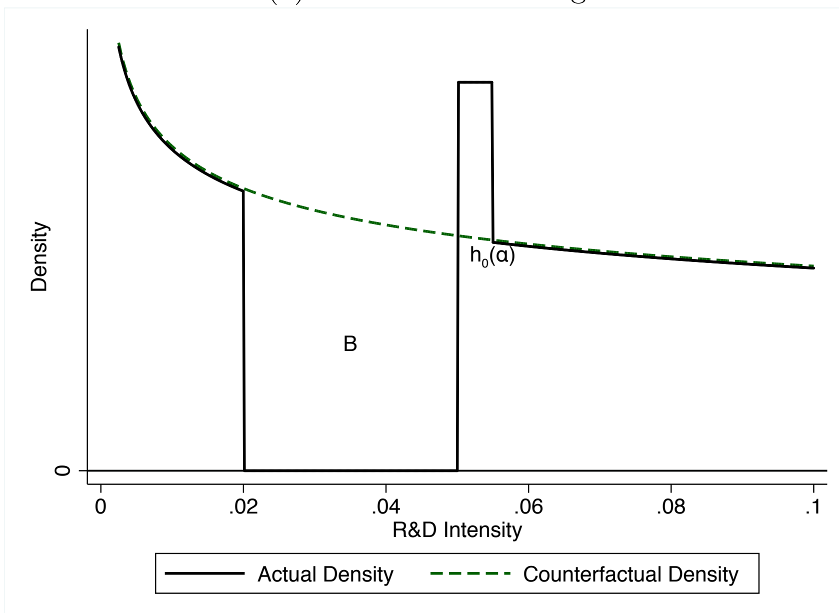
Figure 7: Marginal Buncher and Evasion



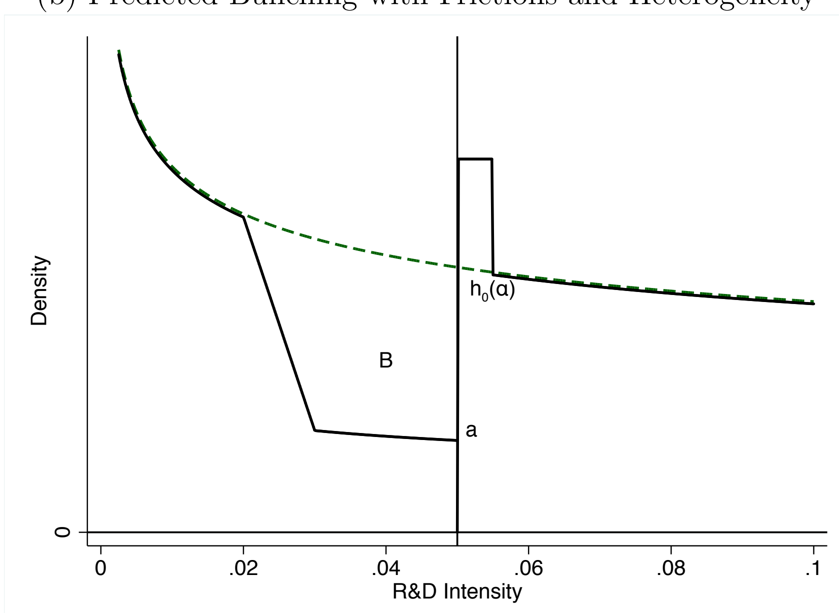
Notes: See Section 4 for details.

Figure 8: Theoretical Predictions of Bunching

(a) Predicted Bunching

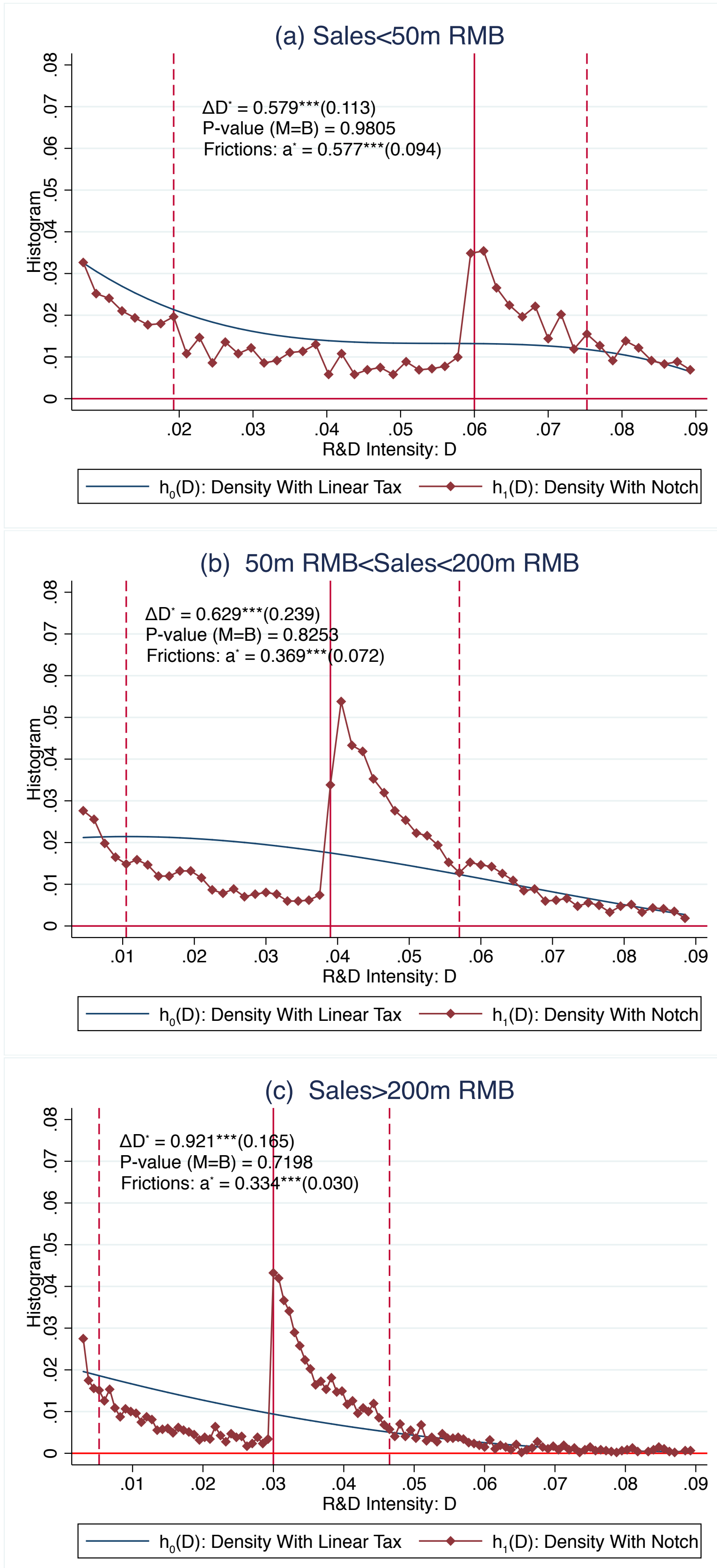


(b) Predicted Bunching with Frictions and Heterogeneity



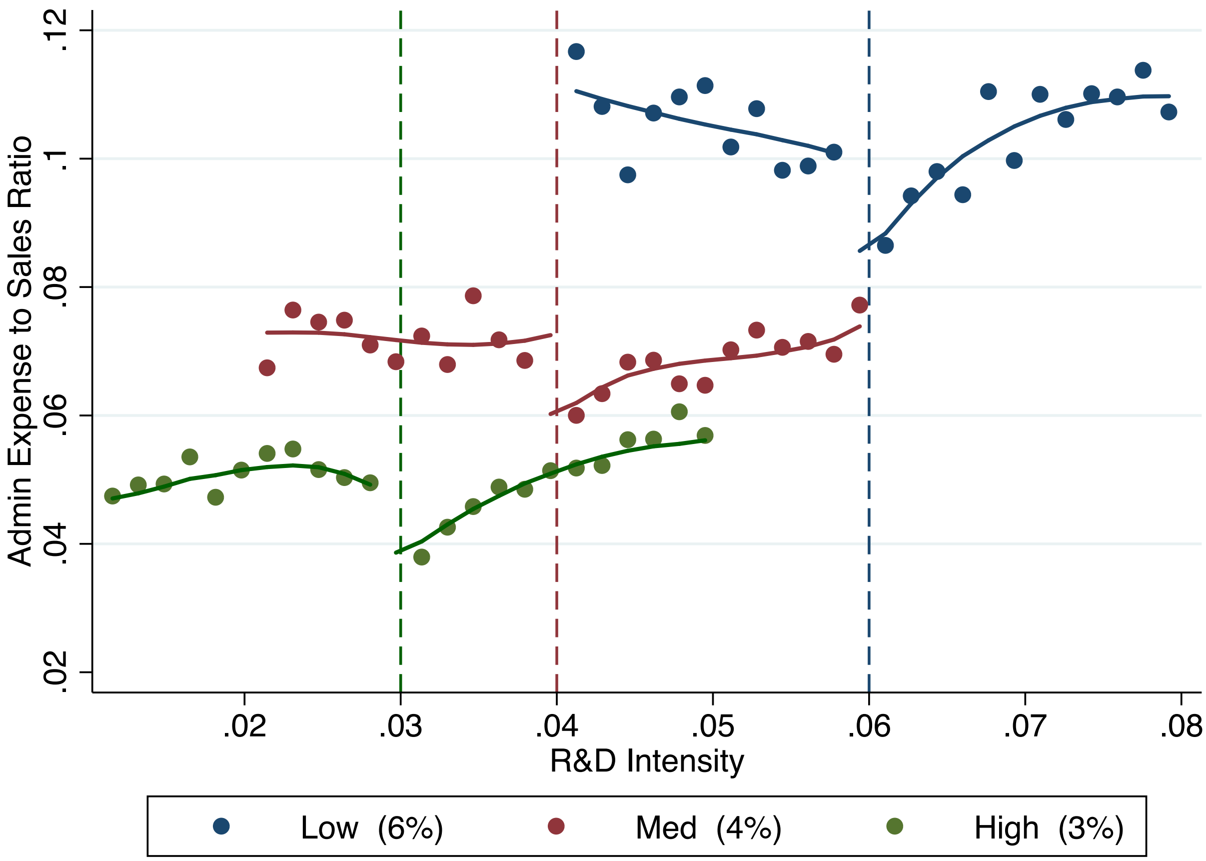
Notes: See Section 4 for details.

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



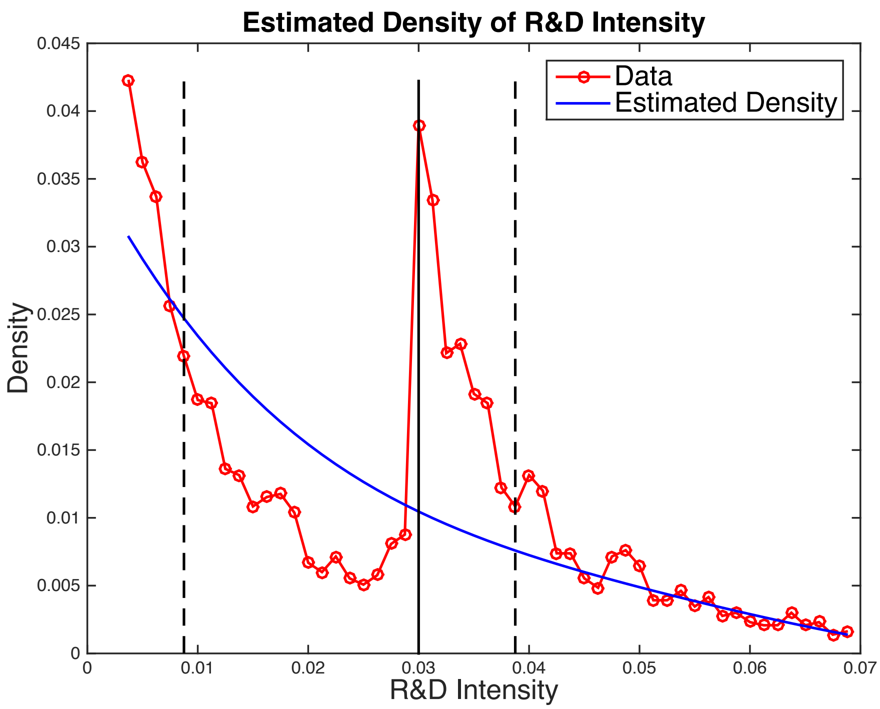
Source: Administrative Tax Return Database. See Section 2 for details on data sources and Section 4 for details on the bunching estimator.

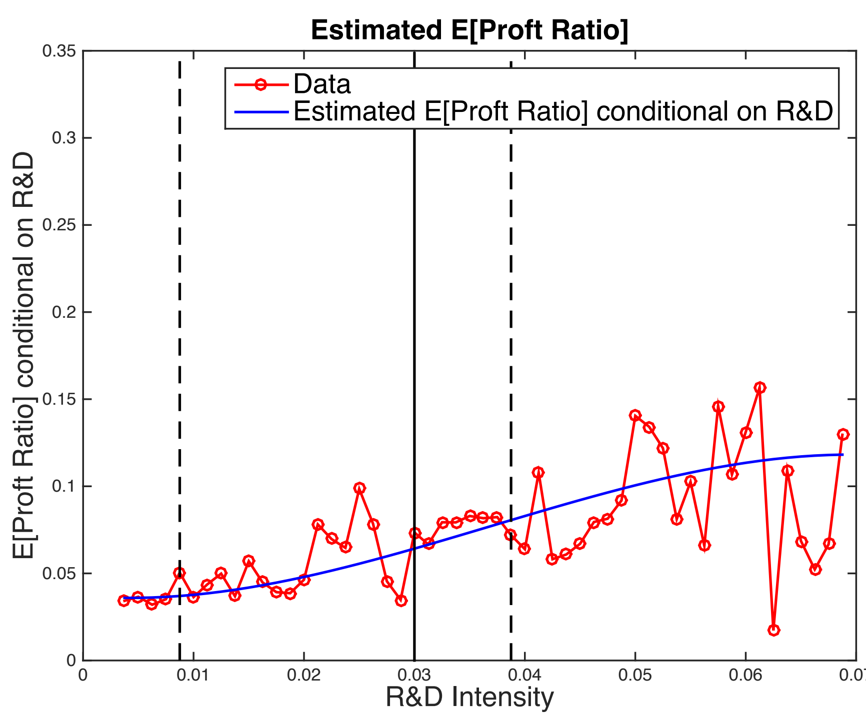
Figure 10: Empirical Evidence of Evasion



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

Figure 11: Estimates of Excess Mass from Bunching at Notch (2009) and ITT on Profit Margin





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

Table 1: Descriptive Statistics

Panel A: State Administration of Tax Data 2008 - 2011						
	Mean	Std	p25	p50	p75	# of Obs.
Sales (mil RMB)	118.2	1394.8	2.6	10.6	42.1	1202257
Fixed Asset (mil RMB)	32.9	390.4	0.4	2.1	10.7	1139038
# of Workers	175	852	17	48	136	1213497
R&D or not	0.08	0.27	–	–	–	1202186
R&D/Sales (if>0)	3.5%	6.9%	0.3%	1.5%	4.3%	98187
Adm Expense/Sales	9.4%	11.8%	2.8%	5.8%	11.1%	1171366
Panel B: Annual Survey of Manufacturing 2006 - 2007						
Sales (mil RMB)	110.8	1066.1	10.7	23.7	59.5	638668
Fixed Asset (mil RMB)	42.7	703.5	1.7	4.5	13.4	634561
# of Workers	238	1170	50	95	200	638668
R&D or not	0.10	0.30	–	–	–	638668
R&D/Sales (if>0)	1.6%	3.2%	0.1%	0.5%	1.7%	65272

Notes: Various sources, see Section 2 for details.

Table 2: Estimates of Mis-categorized R&D

	(1)	(2)	(3)
	Small	Medium	Large
Structural Break	-0.014**	-0.013***	-0.008***
	(0.007)	(0.004)	(0.003)
Observations	5,016	8,336	8,794
Percentage Misreported	.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 3: 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 4: Policy Effects

Sales Group	ΔD^*	Frictions a^*	Population Response $(1-a^*)\Delta D^*$	Misreporting δ^* : Sales	TFP	Real Effect $\Delta D^* - \delta^*$ δ^* : Sales	δ^* : TFP	Real Population R&D Intensity $(1-a^*)\alpha(\Delta D^* - \delta^*)$ δ^* : Sales	δ^* : TFP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Small	0.579***	0.577***	0.239***	0.232**	0.035	0.347**	0.546***	0.898***	1.441***
	(0.113)	(0.094)	(0.040)	(0.111)	(0.072)	(0.164)	(0.130)	(0.330)	(0.293)
Medium	0.629***	0.369***	0.380***	0.325***	0.102***	0.304	0.527**	0.656	1.159***
	(0.239)	(0.072)	(0.062)	(0.101)	(0.029)	(0.254)	(0.240)	(0.460)	(0.380)
Large	0.921***	0.334***	0.618***	0.263***	0.138***	0.658***	0.784***	1.176***	1.397***
	(0.165)	(0.030)	(0.147)	(0.100)	(0.018)	(0.188)	(0.166)	(0.456)	(0.451)

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 in parentheses. The fraction a^* is determined using the average net profit ratio of 0.07. See Appendix Table A4 for alternative calculations that assume a profit ratio of 0.2.

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

Table 5: Estimates of Treatment Effects

(a) Estimates of Intent-to-Treat (ITT) Effects

	ITT	SE	T-Stat	Bootstrap 5th Percentile	Bootstrap 95th Percentile
Profit Ratio	0.016	0.008	2.163	0.005	0.03
I to K Ratio	0.033	0.02	1.612	-0.003	0.063
TFP	0.016	0.005	2.931	0.007	0.025
R&D	0.272	0.033	8.247	0.223	0.33
Tax	-0.084	0.029	-2.91	-0.128	-0.038

(b) Wald Estimates of Treatment Effects

	Wald Estimate	Bootstrap 5th Percentile	Bootstrap 95th Percentile
Profit Ratio to R&D	0.06	0.019	0.11
I to K Ratio to R&D	0.12	-0.01	0.239
TFP to R&D	0.059	0.025	0.096
Tax to R&D	-0.308	-0.494	-0.137
Tax to TFP (1%)	-0.053	-0.134	-0.021

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 6: Structural Estimates

(a) Estimates with Homogeneous $\varepsilon(\theta-1)$

	$\varepsilon(\theta-1)$	η	b	c
Estimate	1.786	0.538	3.000	0.500
SE	0.203	0.288		
T-Stat	8.794	1.868		

(b) Estimates with Heterogeneous $\varepsilon(\theta-1)$

	$\varepsilon(\theta-1)$					
	Small	Medium	Large	η	b	c
Estimate	1.482	1.696	1.868	0.910	3.000	0.500
SE	0.113	0.275	0.240	0.352		
T-Stat	13.135	6.179	7.794	2.585		

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.

Online Appendix: Not For Publication

A Detailed Model Derivation

A.1 Model Setup

Consider a firm i with a Cobb-Douglas production function given by:

$$q_{it} = \exp\{\phi_{it}\} K_{it}^{\kappa} V_{it}^{1-\kappa},$$

where K_{it} and M_{it} are static inputs with respective prices p_k and p_v, 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 $\frac{\varepsilon}{1-\rho}$.

A.2 Cost Function and Profit Function

A.2.1 Cost Function

The cost function for this familiar problem is given by:

$$C(q; \phi_{it}, p_k, p_v) = \frac{q}{\exp\{\phi_{it}\}} \frac{p_k^{\kappa} p_v^{1-\kappa}}{\kappa^{\kappa}(1-\kappa)^{1-\kappa}}$$

Define also the unit cost function:

$$c(\phi_{it}, p_k, p_v) = \frac{C(q; \phi_{it}, p_k, p_v)}{q} = \frac{1}{\exp\{\phi_{it}\}} \frac{p_k^{\kappa} p_v^{1-\kappa}}{\kappa^{\kappa}(1-\kappa)^{1-\kappa}}$$

A.2.2 Profit 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

$$\max_{q_{it}} q_{it}^{1-1/\theta} - qc(\phi_{it}, p_k, p_v).$$

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

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

Revenue is then given by:

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

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 4, which implies a gross-markup around 1.33. Per-period profits are then given by:

$$\pi_{it} = \frac{1}{\theta - 1} q^*_{it} c(\phi_{it}, p_k, p_v) = \frac{(\theta - 1)^{\theta-1}}{\theta^\theta} c(\phi_{it}, p_k, p_v)^{1-\theta}.$$

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

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

where $\mathbb{E}[\pi_{it}|D_{i,t-1} = 0]$ is the expected profit without any R&D investment.

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

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

We also allow for the possibility that firms incur a fixed cost of attaining the InnoCom certification. To model this, we assumer that if firms decide to pursue the certification, they incur a cost of: $c \times D_{i1}$.

A.3 Derivation of Moment Equations

A.3.1 R&D Choice Under Linear Tax

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

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

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:

$$FOC: \; -(1 - t_1) \left(1 + b\left[\frac{D_{i1}}{\theta\pi_{i1}}\right]\right) + \beta(1 - t_2)\varepsilon(\theta - 1)D_{i1}^{(\theta-1)\varepsilon-1}\mathbb{E}[\pi_{i2}|D_{i1} = 0] = 0. \tag{14}$$

Notice first that if the tax rate is constant across periods, the corporate income tax does not affect the choice of R&D investment.[19] In the special case of no adjustment costs (i.e., $b = 0$), the optimal choice of D_{i1} is given by:

$$D_{i1} = \left[\frac{1}{(\theta-1)\varepsilon}\frac{1-t_1}{\beta(1-t_2)}\frac{1}{\mathbb{E}[\pi_{i2}|D_{i1}=0]}\right]^{\frac{1}{(\theta-1)\varepsilon-1}}. \tag{15}$$

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, 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 factors including K_i and ϕ_{i1} that influence $\mathbb{E}[\pi_{i2}|D_{i,t-1}=0]$. An important insight for the proceeding analysis is that we can recover these factors from D_{1i} as follows:

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

A.3.2 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:

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

Using the expression for $\mathbb{E}[\pi_{i2}|D_{i1}=0]$ above, we can re-express this condition for the marginal buncher as:

$$SOC': \; \frac{(1-t_1)}{D^{*-}}\left\{((\theta-1)\varepsilon-1)\left(1+b\left[\frac{D^{*-}}{\theta\pi_{i1}}\right]\right) - b\left[\frac{D^{*-}}{\theta\pi_{i1}}\right]\right\} < 0.$$

Since $\frac{(1-t_1)}{D^{*-}} > 0$ we focus on the term in the brackets and use the definition of ΔD^* to obtain:

$$SOC'': \; ((\theta-1)\varepsilon-1)\left(1+\alpha b(1-\Delta D^*)\right) - \alpha b(1-\Delta D^*) < 0,$$

which holds whenever:

$$\frac{(\theta-1)\varepsilon-1}{2-(\theta-1)\varepsilon}\frac{1}{\alpha(1-\Delta D^*)} < b$$

Since $(\theta-1)\varepsilon-1 \in [1,2]$, this lower bound on b is positive. For sample values $(\theta-1)\varepsilon-1 = 1.9$, $\alpha = .03$, $\Delta D^* = .33$, we have:

$$\frac{(\theta-1)\varepsilon-1}{2-(\theta-1)\varepsilon}\frac{1}{\alpha(1-\Delta D^*)} \approx 429 < b.$$

We check that this equation holds in practice and restrict the structural estimation to parameter values that satisfy this condition.

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

A.3.3 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 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 6 presents two possible scenarios following this incentive. Panel (a) shows the situation where the firm finds it optimal to choose a level of R&D intensity below the threshold. At this choice, the first order condition of the linear tax case holds and the optimal level of R&D is given by Equation 14. 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 14 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 $\mathbb{E}[\pi_{i2}|D_{i,t-1} = 0]$, 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 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 14, we can manipulate $\Pi(D_{i1}^{*-}|t_2^{LT})$ to obtain:

$$\begin{aligned} \Pi(D_{i1}^{*-}|t_2^{LT}) &= (1-t_1)\left(\pi_{i1} - D_{i1}^{*-} - \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]^2\right) + \beta(1-t_2^{LT})(D_{i1}^{*-})^{(\theta-1)\varepsilon}\mathbb{E}[\pi_{i2}|D_{i1}=0] \\ &= (1-t_1)\left(\pi_{i1} + \left(\frac{1}{\varepsilon(\theta-1)} - 1\right)D_{i1}^{*-} + b\theta\pi_{i1}\left(\frac{1}{\varepsilon(\theta-1)} - \frac{1}{2}\right)\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]^2\right), \quad (16) \end{aligned}$$

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 14 to obtain:

$$\begin{aligned}
\Pi(\alpha\pi_1|t_2^{HT}) &= (1-t_1)\left(\pi_{i1}-\alpha\theta\pi_{i1}(1+c)-\frac{b\theta\pi_{i1}}{2}\left[\frac{\alpha\theta\pi_{i1}}{\theta\pi_{i1}}\right]^2\right)+\beta(1-t_2^{HT})(\alpha\theta\pi_{i1})^{(\theta-1)\varepsilon}\mathbb{E}[\pi_{i2}|D_{i1}=0] \\
&= (1-t_1)\left(\pi_{i1}-\alpha\theta\pi_{i1}(1+c)-\frac{\alpha^2 b\theta\pi_{i1}}{2}\right. \\
&\quad \left.+\frac{(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{\alpha\theta\pi_{i1}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon}\left(1+b\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]\right)D_{i1}^{*-}\right). \qquad (17)
\end{aligned}$$

We then use Equations 16 and 17 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:

$$\begin{aligned}
0 &= \left(\frac{1}{\varepsilon(\theta-1)}-1\right)D_{i1}^{*-}+b\theta\pi_{i1}\left(\frac{1}{\varepsilon(\theta-1)}-\frac{1}{2}\right)\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]^2+\alpha\theta\pi_{i1}(1+c)+\frac{\alpha^2 b\theta\pi_{i1}}{2} \\
&- \frac{(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{\alpha\theta\pi_{i1}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon}\left(1+b\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]\right)D_{i1}^{*-} \\
0 &= \left(\frac{1}{\varepsilon(\theta-1)}-1\right)\frac{D_{i1}^{*-}}{\alpha\theta\pi_1}+\alpha b\left(\frac{1}{\varepsilon(\theta-1)}-\frac{1}{2}\right)\left[\frac{D_{i1}^{*-}}{\alpha\theta\pi_{i1}}\right]^2+1+c+\frac{\alpha b}{2} \\
&- \frac{(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{\alpha\theta\pi_{i1}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon-1}\left(1+\alpha b\left[\frac{D_{i1}^{*-}}{\alpha\theta\pi_{i1}}\right]\right) \\
0 &= \left(\frac{1}{\varepsilon(\theta-1)}-1\right)(1-\Delta D^*)+\alpha b\left(\frac{1}{\varepsilon(\theta-1)}-\frac{1}{2}\right)(1-\Delta D^*)^2+1+c+\frac{\alpha b}{2} \\
&- \frac{(1-t_2^{HT})}{(1-t_2^{LT})}\times\frac{(1-\Delta D^*)^{1-(\theta-1)\varepsilon}}{\varepsilon(\theta-1)}\left(1+\alpha b(1-\Delta D^*)\right), \qquad (18)
\end{aligned}$$

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 an estimate of b, c, Equation 18 is an implicit function for $(\theta-1)\varepsilon$. Thus, given observable tax parameters t_2^{HT} and t_2^{LT} and the empirical quantity ΔD^*, which can be estimated from the bunching patterns described in Section 3, it is possible to recover an estimate of the parameters $(\theta-1)\varepsilon$, b, and c from multiple groups of firms with similar structural parameters.

A.3.4 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 inter-temporal 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 16 still describes the profit in this case.

We now we manipulate $\Pi(\alpha\theta\pi_1, D_1^{*K}|t^{HT})$ using the FOC for D_{i1}^{*-} to obtain:

$$\begin{aligned}
\Pi(\alpha\theta\pi_{i1}, D_{i1}^{*K}|t_2^{HT}) \quad &= \quad (1-t_1)\left(\pi_{i1} - D_{i1}^{*K} - \alpha\theta\pi_{i1}c - \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}^{*K}}{\theta\pi_{i1}}\right]^2\right) \\
&\quad +\beta(1-t_2^{HT})(D_{i1}^{*K})^{(\theta-1)\varepsilon}\mathbb{E}[\pi_{i2}|D_{i1}=0] - h(D_1^{*K}, \alpha\theta\pi_1) \\
&= \quad (1-t_1)\left(\pi_{i1} - D_{i1}^{*K} - \alpha\theta\pi_{i1}c - \frac{b\theta\pi_{i1}}{2}\left[\frac{D_{i1}^{*K}}{\theta\pi_{i1}}\right]^2\right) - h(D_1^{*K}, \alpha\theta\pi_1) \\
&\quad + \frac{(1-t_1)(1-t_2^{HT})}{\varepsilon(\theta-1)(1-t_2^{LT})}\left(\frac{D_{i1}^{*K}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon}\left(1 + b\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]\right)D_{i1}^{*-}
\end{aligned} \tag{19}$$

We then use Equations 16 and 19 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 manipulating we obtain:

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

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. We also consider a particular function for $\tilde{h}(\delta)$ given by $\frac{\delta^{1+\eta}}{1+\eta}$. These definitions and assumptions yield the following condition:

$$\begin{aligned}
0 &= 1 + \frac{c}{1-\delta^*} + \frac{\alpha b}{2}(1-\delta^*) + \left(\frac{1-\Delta D^*}{1-\delta^*}\right)\left[\left(\frac{1}{\varepsilon(\theta-1)}-1\right) + \alpha b\left(\frac{1}{\varepsilon(\theta-1)}-\frac{1}{2}\right)(1-\Delta D^*)\right] \\
&- \frac{(1-t_2^{HT})}{(1-t_2^{LT})}\left(\frac{1-\Delta D^*}{1-\delta^*}\right)^{1-(\theta-1)\varepsilon}\left(\frac{1+\alpha b(1-\Delta D^*)}{\varepsilon(\theta-1)}\right) + \frac{(\delta^*)^{\eta+1}(1-\delta^*)^{-1}}{(1-t_1)(\eta+1)} \qquad (20)
\end{aligned}$$

Notice that in the special case with no evasion, when $\delta^* = 0$, Equation 20 is identical to Equation 18.

In the case when the firm decides to bunch and evade, we have the additional information that D^K is chosen optimally. From Equation 19, the firm solves the following problem:

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

with the following FOC:

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

Notice that this equation is equivalent to:

$$\left(\frac{1-\delta^*}{1-\Delta D^*}\right)^{(\theta-1)\varepsilon-1} = \frac{1+\alpha b(1-\delta^*) - \frac{(\delta^*)^\eta}{1-t_1}}{\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)(1+\alpha b(1-\Delta D^*))} \qquad (21)$$

Equation 21 along with Equation 20 now form a system of two equations that are implicit functions for the parameters η and $(\theta-1)\varepsilon$.

A.3.5 Second Order Conditions

Consider again the FOC for the evasion problem:

$$FOC: \; -(1-t_1)\left(1+b\left[\frac{D^{K*}}{\theta\pi_{i1}}\right]\right)+(1-t_1)\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\left(\frac{D_{i1}^{*K}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon-1}\left(1+b\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]\right) + \left(\frac{\alpha\theta\pi_1-D^K}{\alpha\theta\pi_1}\right)^{\eta-1}$$

The SOC is given by:

$$-(1-t_1)b\left[\frac{1}{\theta\pi_{i1}}\right]+\frac{((\theta-1)\varepsilon-1)}{D_{i1}^{*-}}(1-t_1)\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\left(\frac{D_{i1}^{*K}}{D_{i1}^{*-}}\right)^{(\theta-1)\varepsilon-2}\left(1+b\left[\frac{D_{i1}^{*-}}{\theta\pi_{i1}}\right]\right) -(\eta-1)\left(\frac{\alpha\theta\pi_1-D^K}{\alpha\theta\pi_1}\right)^{\eta-2}\frac{1}{\alpha\theta\pi_1}<0$$

Collecting terms and substituting for δ^* and ΔD^* we can rewrite this as:

$$(1-t_1)(1-\Delta D^*)\left\{\frac{((\theta-1)\varepsilon-1)}{(1-\Delta D^*)}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\left(\frac{1-\delta^*}{1-\Delta D^*}\right)^{(\theta-1)\varepsilon-2}(1+\alpha b(1-\Delta D^*))-\alpha b-\frac{(\eta-1)}{(1-t_1)}(\delta^*)^{\eta-2}\right\} < 0$$

B Identification

B.1 Simple Model

Consider first the model without evasion, adjustment, or fixed costs. In this case, Equation 18 defined an implicit function of $(\theta-1)\varepsilon$ as a function of ΔD^*. While there is no closed-form equation for $(\theta-1)\varepsilon$, there is an intuitive relation between ΔD^* and $(\theta-1)\varepsilon$. As more firms choose to bunch at the notch, this would imply that the effect of R&D on profits is larger. It follows that $(\theta-1)\varepsilon$ is increasing in ΔD^*. Figure A3 provides the implied value of Equation 18 for a range of values of ΔD^* and confirms this intuition.

B.2 Model With Evasion

While there are no closed-form expressions for η and $(\theta-1)\varepsilon$, using Equations 20 and 21 we can find a closed-form solution for the effect of R&D on profits as a function of η and b. Solving for $(\theta-1)\varepsilon$ in Equation 21 yields:

$$\varepsilon(\theta-1)=1+\frac{\ln\left(1+\alpha b(1-\delta^*)-\frac{(\delta^*)^\eta}{1-t_1}\right)-\ln\left(\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)(1+\alpha b(1-\Delta D^*))\right)}{\ln\left(\frac{1-\delta^*}{1-\Delta D^*}\right)} \tag{22}$$

Similarly, we can substitute Equation 21 into 20 to obtain the following expression:

$$
\begin{aligned}
0 &= 1+c+\frac{\alpha b}{2}(1-\delta^*)+\left(\frac{1-\Delta D^*}{1-\delta^*}\right)\left[\left(\frac{1}{\varepsilon(\theta-1)}-1\right)+\alpha b\left(\frac{1}{\varepsilon(\theta-1)}-\frac{1}{2}\right)(1-\Delta D^*)\right] \\
&- \frac{1+\alpha b(1-\delta^*)-\frac{(\delta^*)^\eta}{1-t_1}}{\varepsilon(\theta-1)}+\frac{(\delta^*)^{\eta+1}(1-\delta^*)^{-1}}{(1-t_1)(\eta+1)},
\end{aligned}
$$

which is linear in $(\theta-1)\varepsilon$. Solving for $(\theta-1)\varepsilon$, we obtain:

$$
\varepsilon(\theta-1)=\frac{\left(\frac{1-\Delta D^*}{1-\delta^*}\right)(1+\alpha b(1-\Delta D^*))-1-\alpha b(1-\delta^*)+\frac{(\delta^*)^\eta}{1-t_1}}{\left(\frac{1-\Delta D^*}{1-\delta^*}\right)\left(1+\frac{\alpha b}{2}(1-\Delta D^*)\right)-1-\frac{c}{1-\delta^*}-\frac{\alpha b}{2}(1-\delta^*)-\frac{(\delta^*)^{\eta+1}(1-\delta^*)^{-1}}{(1-t_1)(\eta+1)}}. \tag{23}
$$

Figure A4 plots the non-linear relations between η and $(\theta-1)\varepsilon$ that are implied by Equations 22 and 23 while holding $b, c = 0$. Panel (a) explores Equation 22 and shows that for reasonable values of η, $(\theta-1)\varepsilon$ is positive. This figure also shows that, given values of ΔD^* and δ^*, as evasion become more costly (larger η), the value of R&D to the firm also increases. Figure A4 panel (b) explores Equation 23 and shows that, for a given cost and amount of evasion, i.e., η and δ^*, a larger response in terms of reported R&D corresponds to larger values of $(\theta-1)\varepsilon$. This figure plots this relation for different values of η and thus shows how the reduced-form moments δ^* and ΔD^* influence the estimates in the model.

For a given set of empirical estimates ΔD^* and δ^* and values b and c, the structural parameters η and $(\theta-1)\varepsilon$ are identified by the intersection of the graphs in both panels. This intersection will vary as a function of b and c and will generate a set of structural parameters that are compatible with the data. Figure A5 shows the intersection of these functions for multiple values of b, while holding $c = 0$. Th red line represent the locus of parameters that is compatible with a given set of data ΔD^* and δ^*. The parameters $\eta, (\theta-1)\varepsilon, b$, and c are identified through cross-group restrictions that use data on ΔD^* and δ^* for the three groups of firms.

Appendix Graphs

Figure A1: Aggregate Implications

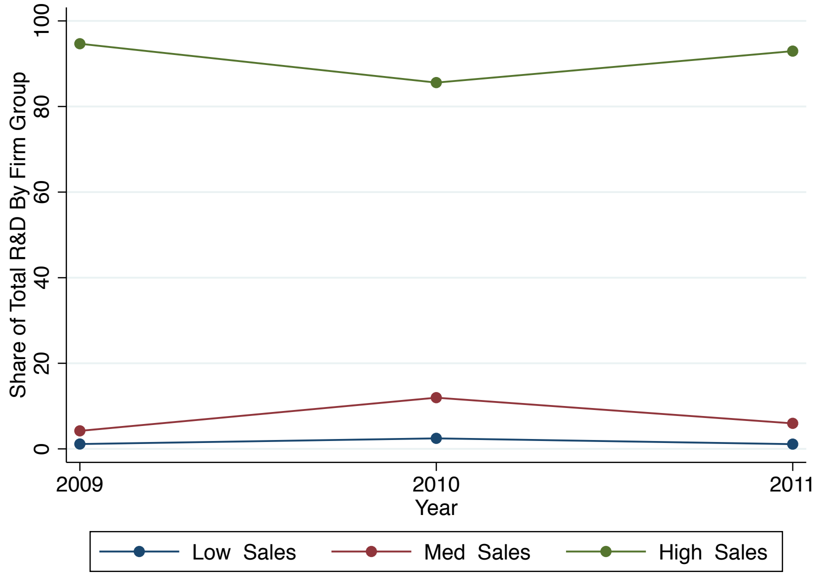


Figure A2: Alternative Empirical Evidence of Evasion

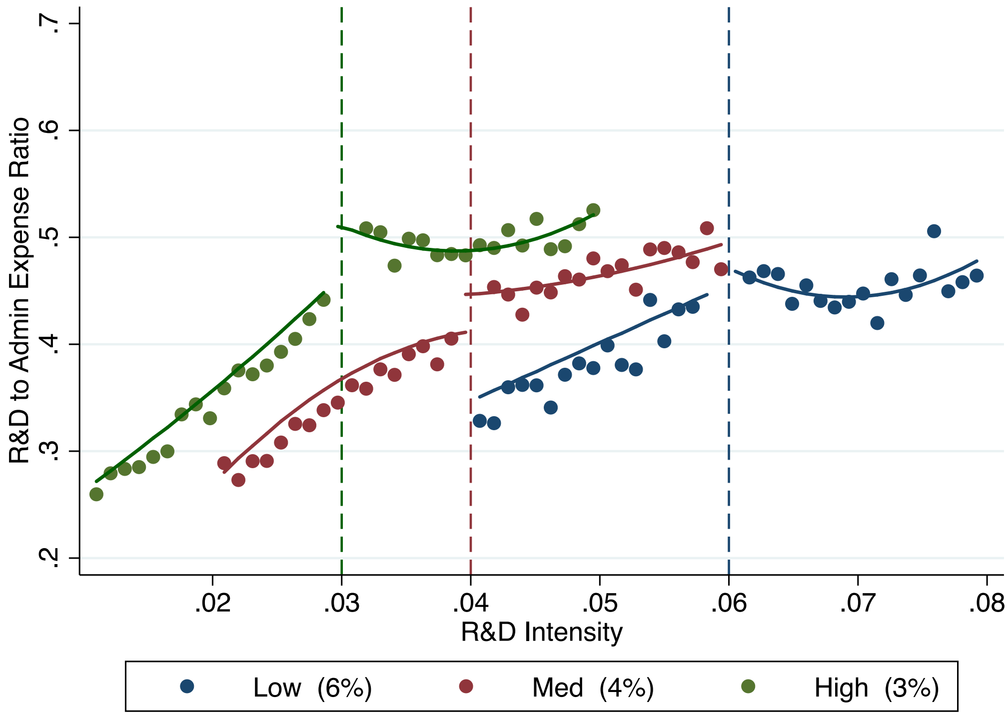


Figure A3: Relation Between $1 - \Delta D^*$ and $(\theta - 1)\varepsilon$ Without Evasion.

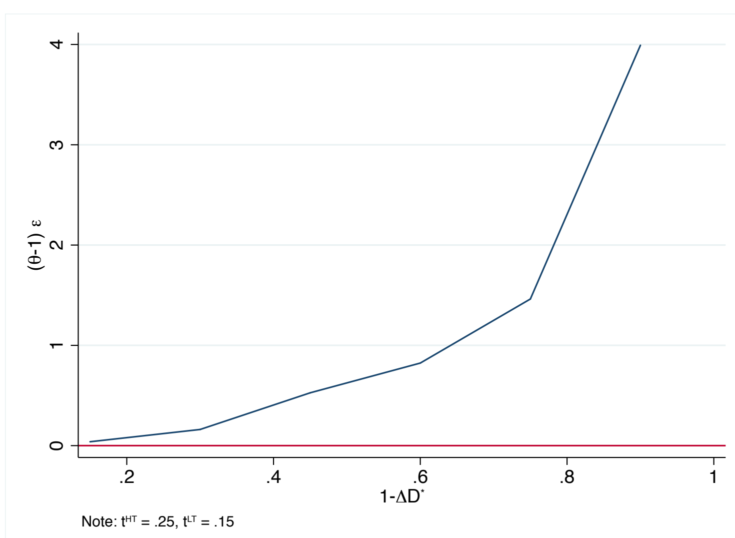
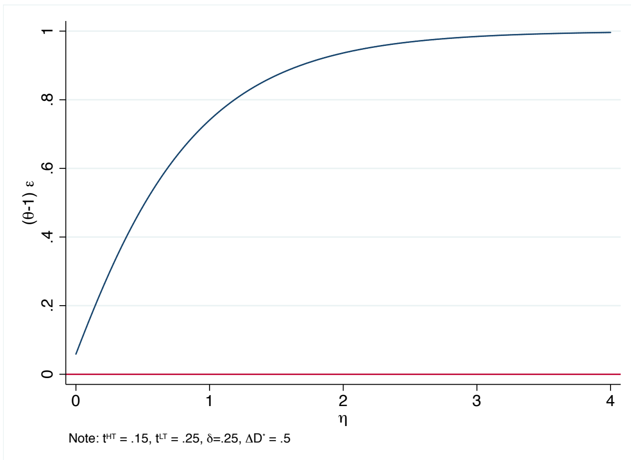


Figure A4: Identification When Evasion is Possible

(a) Relation Between $(\theta - 1)\varepsilon$ and η



(b) Relation Between $(\theta - 1)\varepsilon$ and $1 - \Delta D^*$ for different values of η

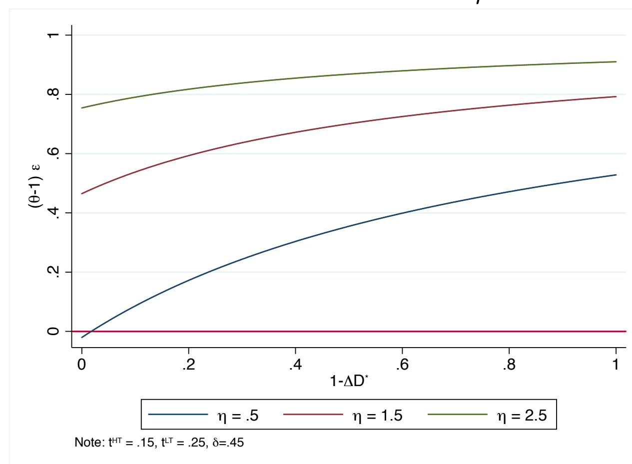


Figure A5: Identification in Full Structural Model

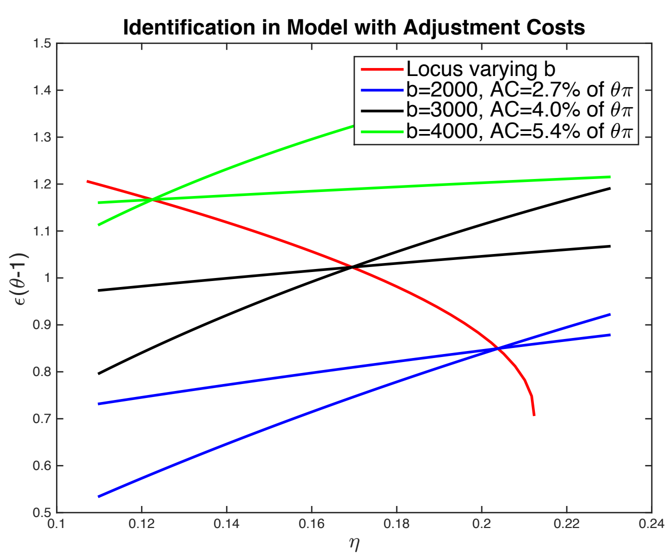


Figure A6: Sensitivity to Calibrated b

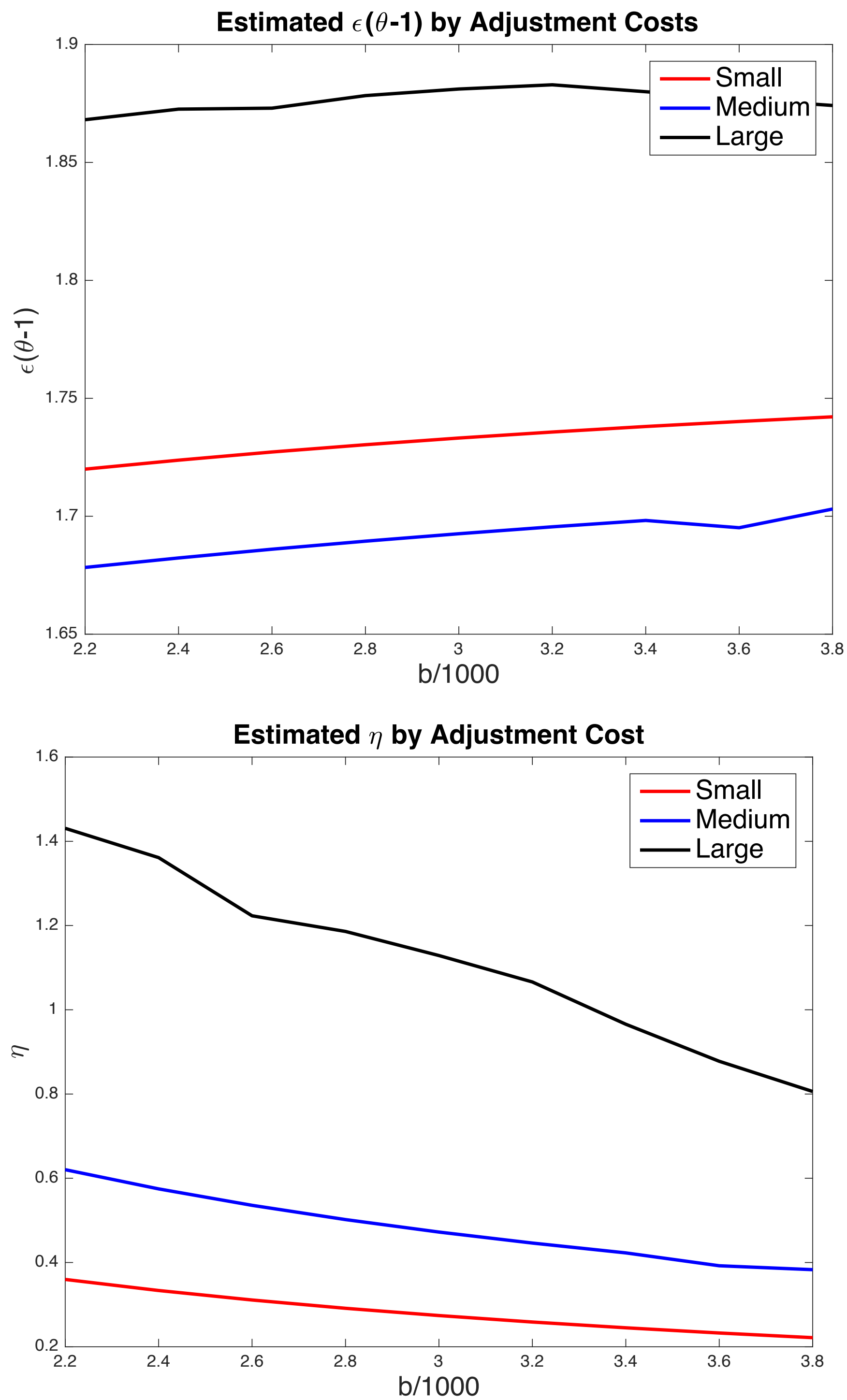
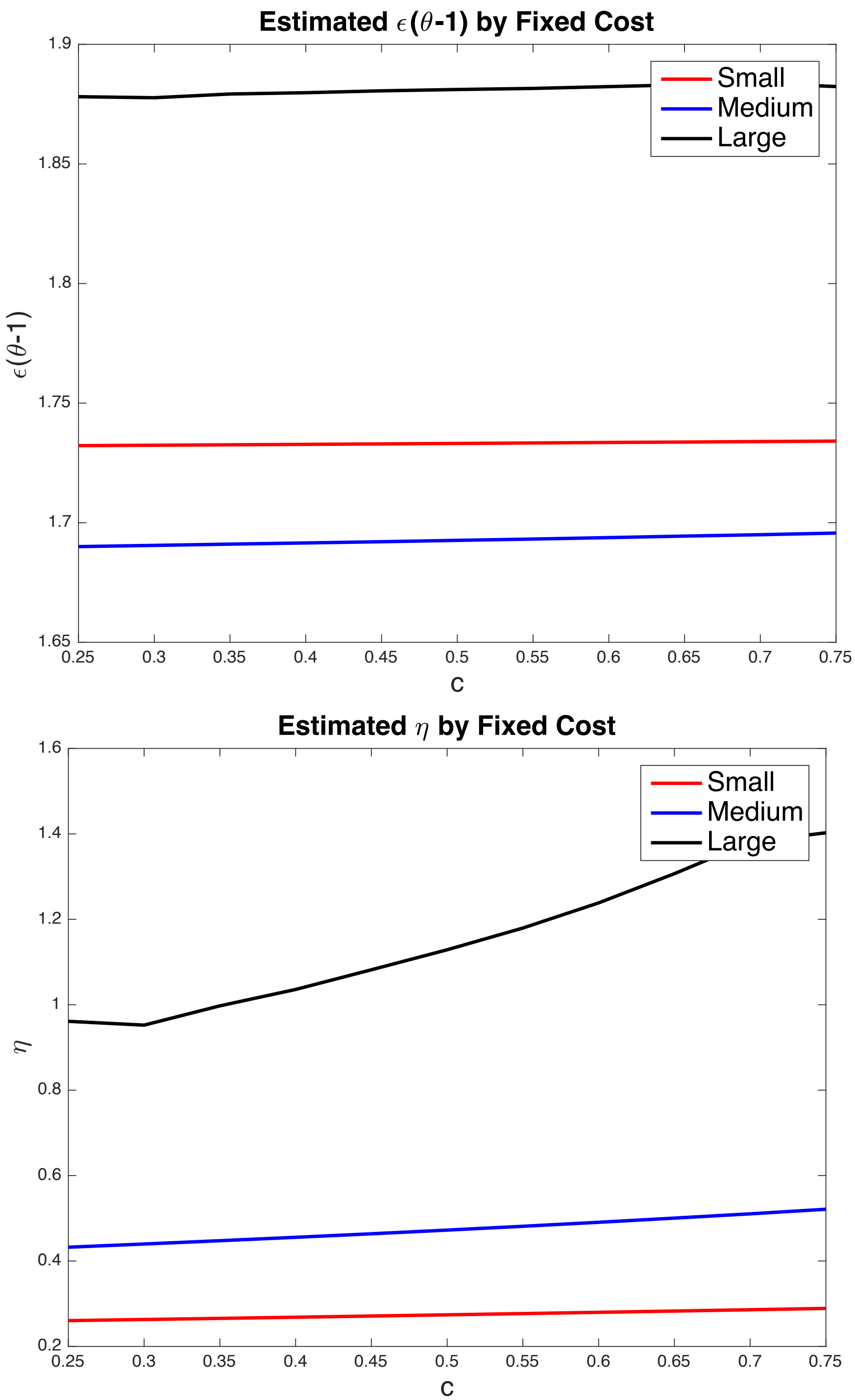


Figure A7: Sensitivity to Calibrated c



Appendix Tables

Table A1: 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 A2: 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 A3: 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.

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

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

Table A4: Alternative Policy Effects

Sales Group	ΔD^*	Frictions a^*	Population Response $(1-a^*)\Delta D^*$	Misreporting δ^* : Sales	Misreporting δ^* : TFP	Real Effect $\Delta D^* - \delta^*$: δ^* : Sales	Real Effect $\Delta D^* - \delta^*$: δ^* : TFP	Real Population R&D Intensity $(1-a^*)\alpha(\Delta D^* - \delta^*)$: δ^* : Sales	Real Population R&D Intensity $(1-a^*)\alpha(\Delta D^* - \delta^*)$: δ^* : TFP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Small	0.533***	0.550***	0.240***	0.232**	0.035	0.301***	0.499***	0.806***	1.350***
	(0.020)	(0.085)	(0.046)	(0.111)	(0.072)	(0.113)	(0.074)	(0.334)	(0.339)
Medium	0.723***	0.434***	0.403***	0.325***	0.102***	0.398**	0.621***	0.878***	1.382***
	(0.180)	(0.039)	(0.062)	(0.101)	(0.029)	(0.202)	(0.182)	(0.362)	(0.266)
Large	1.049***	0.416***	0.620***	0.263***	0.138***	0.786***	0.912***	1.396***	1.617***
	(0.137)	(0.048)	(0.144)	(0.100)	(0.018)	(0.166)	(0.139)	(0.421)	(0.413)

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 in parentheses. The fraction a^* is determined using the average net profit ratio of 0.2. See Table 4 for alternative calculations that assume a profit ratio of 0.07.

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