Online Appendix for "Notching R&D Investment with Corporate Income Tax Cuts in China"
Not For Publication

Zhao Chen
Fudan University

Zhikuo Liu
Fudan University

Juan Carlos Suárez Serrato
Duke University
& NBER

Daniel Yi Xu
Duke University
& NBER

January 2021

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

A Additional Details on the Chinese Corporate Income Tax System

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

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

The Chinese government implemented a major corporate tax reform in 2008 to eliminate the dual-track system based on domestic/foreign ownership and established a common rate of 25%. Some of the existing tax breaks for FOEs were gradually phased out. For instance, FOEs that had previously paid an EIT of 15% paid a tax rate of 18% in 2008, 20% in 2009, 22% in 2010, and

24% in 2011. In contrast, the West Development program will remain in effect through 2020.[1]

B Data Sources

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

The second dataset that we use is the administrative enterprise income tax records from the Chinese State Administration of Tax (SAT). The SAT is the counterpart of the IRS in China and is in charge of tax collection and auditing. In addition, the SAT supervises various tax assistance programs such as the InnoCom program. The SAT keeps its own firm-level records of tax payments as well other financial statement information used in tax-related calculations. We acquire these administrative enterprise income tax records for 2008–2011, which allows us to construct detailed tax rate information for individual manufacturing firms. Our main sample of analysis includes firms with a positive tax liability. We also use these data to construct residualized measures of firm productivity.[2] The scope of the SAT data is slightly different from that of the ASM data, but

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

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

there is a substantial amount of overlap for the firms that conduct R&D. For instance, the share of total R&D that can be matched with ASM records is close to 85% for 2008.

C Estimation of Residual Productivity

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

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

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

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

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

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

Robustness of TFP Estimates

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

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

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

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

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

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

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

E Robustness of Bunching Estimates

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

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

Third, we show that our results are robust to using data from before 2008 for large foreign firms that were not subject to the incentives of the InnoCom program to inform the shape of the density in the excluded region. Panel C of Figure A.5 shows that using these data results in very similar estimates of both the counterfactual density and Δd.[3]

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

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

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

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

F ITT Estimates on Productivity, Relabeling, and Tax Revenue

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

Diamond and Persson (2016) develop an estimator that formalizes this comparison and quantifies the average effect of the program on a given outcome Y:[4]

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

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

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

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

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

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

When Y_t is R&D or administrative costs, we estimate contemporaneous effects, so that $t = t_1$. For the case when Y_t is TFP, we study the effect of the program in time t_1 on future TFP ($t > t_1$). We interpret this estimate as an intent to-treat (ITT).[5] For example, the ITT on $Y = \ln d$ measures Δd, the percentage increase in R&D intensity over the excluded region.

The first quantity in Equation F.2 is the observed average value of a given outcome Y_{it} over the excluded region. The second quantity is a counterfactual average value of Y_{it}. We construct this counterfactual by combining the counterfactual density of R&D intensity that we estimated as part of the bunching analysis ($\hat{f}_0(\cdot)$) with an estimated average value of the outcome conditional on a given value of R&D. We estimate $E[Y_{it}|\widehat{d_{t_1}, \text{No}}\text{ Notch}]$ using a flexible polynomial regression of Y_{it} on R&D intensity over the same excluded region used to estimate $\hat{f}_0(\cdot)$:[6]

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

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

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

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

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

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

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

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

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

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

where we use the same exclusion region as in panel E of Figure 4 (see Appendix D for details), and we use either quadratic or cubic polynomials for each outcome.[7]

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

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

since these patterns may be due to selection. Finally, note that we allow the user cost to have a discontinuous jump in panel C since, in contrast to other outcomes, we would expect participation in the program to have a mechanical effect on the user cost of R&D.[8]

F.2 ITT Estimates

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

To relate our estimates to the existing literature, we obtain estimates of the elasticity of R&D investment to the user cost of capital (UCC). Panel A of Table A.1 shows that the policy lowered the UCC in 2009 by 7.1%.[9] The second panel of Table A.1 presents estimates of UCC elasticities obtained by taking the ratio of the ITT on R&D to the ITT on the UCC, along with bootstrapped confidence intervals. The first row shows that reported R&D increased by 2% for every 1% decrease in the user cost. When we use the approximation above to obtain an estimate of the real increase in R&D, we obtain a user cost elasticity closer to 1.3. Notice that the empirical literature focused on OECD countries (see Hall and Van Reenen, 2000; Becker, 2015) has typically found an elasticity ranging from 0.4 to 1.8 based on direct R&D tax credit programs. Thus, our estimates indicate that, once we correct for the relabeling behavior of Chinese manufacturing firms, their user cost elasticity is comparable to those in more developed economies.

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

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

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

G Lack of Manipulation of Other Expenses

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

H Detailed Model Derivation

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

H.1 Model Setup

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

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

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

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

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

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

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

The firm faces a constant elasticity demand function given by:

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

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

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

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

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

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

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

Head and Mayer (2014) survey estimates of θ from the trade literature. While there is a broad range of estimates, the central estimate is close to a value of 5, which implies a gross markup of

1.25. We also normalize the input cost $w_t \equiv 1$ for the rest of our analysis. We can then write per-period profits as:

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

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

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

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

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

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

H.2 R&D Choice under Linear Tax

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

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

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

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

Notice first that if the tax rate is constant across periods, the corporate income tax does not affect the choice of R&D investment.[10] In the special case of no adjustment costs (i.e., $b=0$), the optimal choice of D_{i1} is given by:

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

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

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

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

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

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

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

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

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

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

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

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

H.3 A Notch in the Corporate Income Tax

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The effects of the InnoCom program are now as follows:

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

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

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

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

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

We again substitute for the expected productivity components of the firm decision, i.e., $\tilde{\pi}_{i2}$ with the interior optimal R&D d_{i1}^* using Equation H.3.

The FOC for d_1^K effectively characterizes the optimal relabeling strategy $\delta_{i1}^* \equiv 1 - \frac{d_{i1}^{K*}}{\alpha}$.

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

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

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

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

I Connecting the Model with Bunching Estimates

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

I.1 Empirical Implications for Bunching on R&D

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

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

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

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

The area B can be computed as follows:

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

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

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

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

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

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

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

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

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

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

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

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

Following Kleven and Waseem (2013), it is possible to estimate the fraction of constrained firms at an R&D intensity α^- such that firms would be willing to jump to the notch even if R&D had no effects on productivity:[11]

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

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

I.3 Connecting the Model with ITT Estimates

Our model generates intuitive predictions for the ITT effects on R&D, relabeling through administrative costs (ADM), and TFP. If some of the reported R&D intensity is real activity, our model would predict that $ITT^{TFP} \geq 0$. According to our model for the evolution of TFP in Equation 1, we would find larger values of ITT^{TFP} for larger values of the parameter ε. We expect to find $ITT^{ADM} < 0$ if a fraction of the reported R&D is due to relabeling of administrative costs. Intuitively, if firms over-report R&D by under-reporting administrative costs, the average ADM over the excluded region would be artificially low. Our model predicts small values of ITT^{ADM}

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

if firms face large costs of relabeling. Finally, consider the case where the outcome of interest is reported R&D intensity. In this case, ITT^d only depends on the counterfactual density $f_0(d)$. Our model predicts a larger fraction of compliers if ε is large or if relabeling costs are low.

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

J Details of Structural Estimation

This appendix provides details on the structural estimation.

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

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

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

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

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

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

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

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

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

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

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

9. Fraction of firms not bunching. We use the observed density $f_1(\cdot)$ and the estimated counterfactual density $\hat{f}_0(\cdot)$ (see Section II and Appendix D) to compute $a^* = \int_{d^*-}^{\alpha} f_1(v)dv / \int_{d^*-}^{\alpha} \hat{f}_0(v)dv$.

We compute the variance of this moment by bootstrapping the counterfactual average over the excluded region.

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

K Robustness of Structural Model Assumptions

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

K.1 Parametric Distribution of Firm Productivity

In our benchmark model, we micro-found the cross-sectional log TFP distribution from a normal AR(1) process. We use the persistence and volatility of log value-added for non-R&D firms to calibrate the persistence parameter $\rho = 0.725$ and variance parameter $\sigma = 0.385$.[12] The assumption of this process requires the cross-sectional distribution of firm TFP $\exp(\phi_1)$ to be log-normal. Since we construct firm-level TFP in our data, we can check this parametric assumption directly with the TFP data.

We use ideas proposed by Kratz and Resnick (1996) and Head et al. (2014) in this robustness

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

check. The basic idea is to compare the distribution of measured productivity with the distribution implied by a given functional form. We first construct the *empirical* CDF of firms' measured TFP as $\hat{F}_i, i = 1, 2, ...N$, with i ranked based on firm TFP.

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

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

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

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

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

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

Figure A.9 shows the predicted TFP from imposing the log-normal CDF on the left panel and the Pareto CDF on the right panel. The predicted TFP from imposing the log-normal CDF tracks the 45 degree line, implying that this assumption is broadly consistent with the data. By contrast, the predicted TFP based on the Pareto CDF does not fit the data well. This graph provides strong evidence that log-normal is a reasonable parametric assumption for the TFP distribution.[13]

K.2 Heterogeneity in the TFP Elasticity: ε

Our benchmark model assumes firms have heterogeneous technological opportunities for R&D investment, driven by the heterogeneity in adjustment costs, b. An alternative way of modeling the heterogeneity in firms' technological opportunities is to allow heterogeneity in ε.[14] As we show

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

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

in this appendix, our average estimates of ε and b do not depend on which can be heterogeneous. However, models where ε is allowed to be heterogeneous produce worse fits of the data. We therefore believe that our benchmark model is superior to models with heterogeneous values of ε.

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

Several findings are worth highlighting. First, several of the parameters are close to our baseline estimates. We estimate average values of ε of 0.0725 and 0.102 in Models A and B, respectively.[15] These values are close to our benchmark estimate of 0.09. Similarly, the average adjustment cost parameter is 7.666 and 8.271 for the two cases, which bracket our benchmark estimate. This is not surprising since adjustment costs are primarily identified by the distribution of R&D intensity away from the notch.

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

An important question is whether these alternative models yield better or worse fits of the data moments. Panel B of Table A.6 shows that both of these models result in significantly lower values for average TFP both below and above the notch. Model A, in particular, results in a smaller

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

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

K.3 Robustness of the Adjustment Cost Function

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

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

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

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

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

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

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

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

K.4 Robustness of the Relabeling Cost Function

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

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

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

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

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

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

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

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

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

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

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

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

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

K.5 Potential Real Responses to Administrative Costs

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

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

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

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

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

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

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

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

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

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

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

K.7 Robustness to Measures of Firm Productivity

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

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

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

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

Let $\omega = \{\phi_1, b, c\}$ denote a firm with random draws of its fundamentals—i.e., productivity, adjustment cost, and fixed cost. We construct moments that match the empirical and simulated counterparts of the ITT estimates:[16]

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

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

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

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

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

L Welfare Analysis

This appendix provides details on the analysis in Section V.C.

L.1 Setup: Firm Optimization

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

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

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

Industry prices are given by:

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

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

Overall profits are given by:

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

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

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

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

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

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

L.2 Case 1: No R&D

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

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

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

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

Consumption is then:

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

The government sets t to maximize:

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

Taking the FOC:

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

implies:

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

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

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

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

Utility is then:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Consumption is then:

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

Utility in the second period is:

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

and overall utility is now:

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

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

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

which is Equation 6 in the text.

L.5 Implementation and Spillovers

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

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

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

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

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

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

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

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

To compute such an equilibrium, it helps to write:

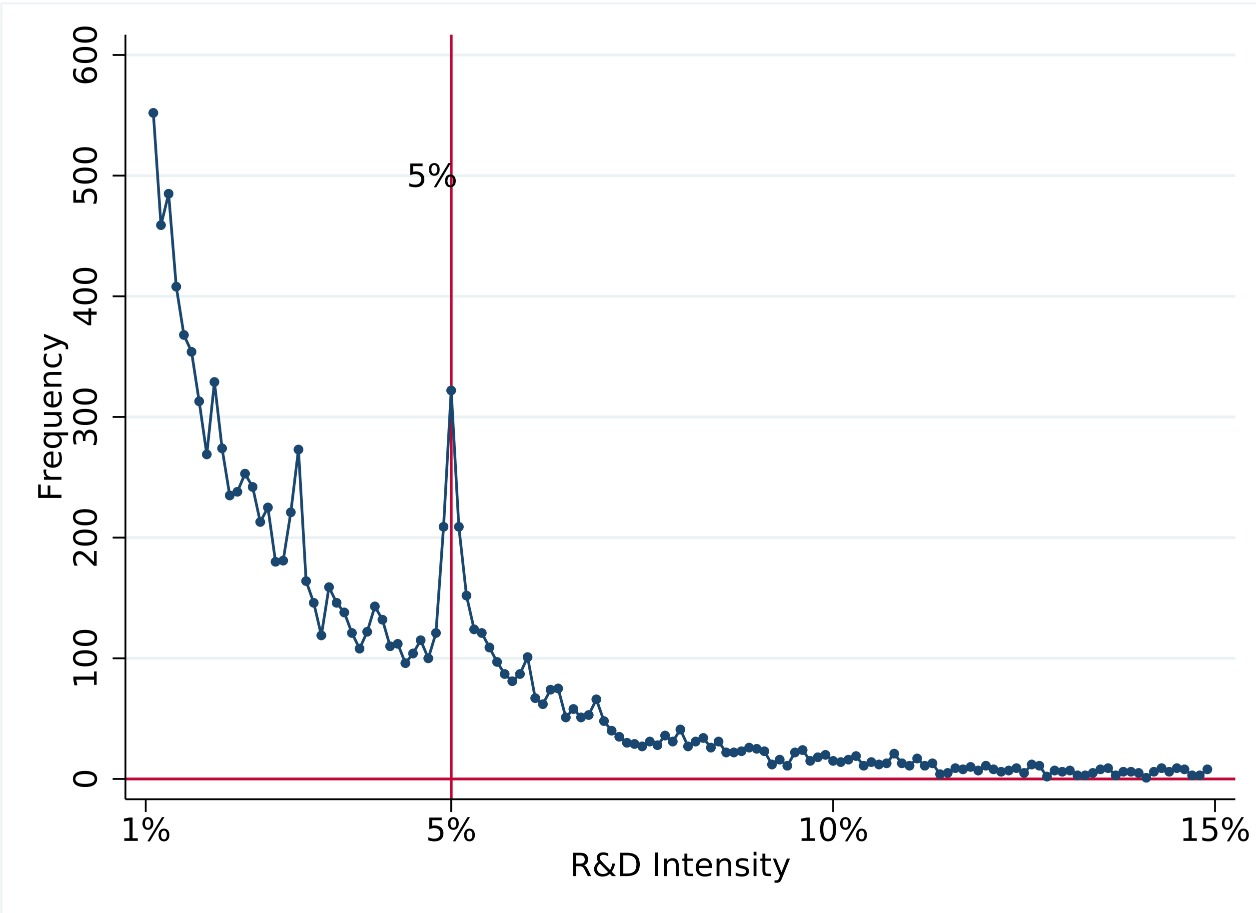
$$\Phi_t^{\theta-1} = (\Phi_t^{\text{R\&D}})^{\theta-1} + (\Phi_t^{\text{Non-R\&D}})^{\theta-1},$$

where $(\Phi_t^{\text{R\&D}})^{\theta-1} = \sum_{i \in N^{\text{R\&D}}} \exp\{\phi_{i,t}\}^{\theta-1}$ and similarly for $\Phi_t^{\text{Non-R\&D}}$. Further, the sales share of

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

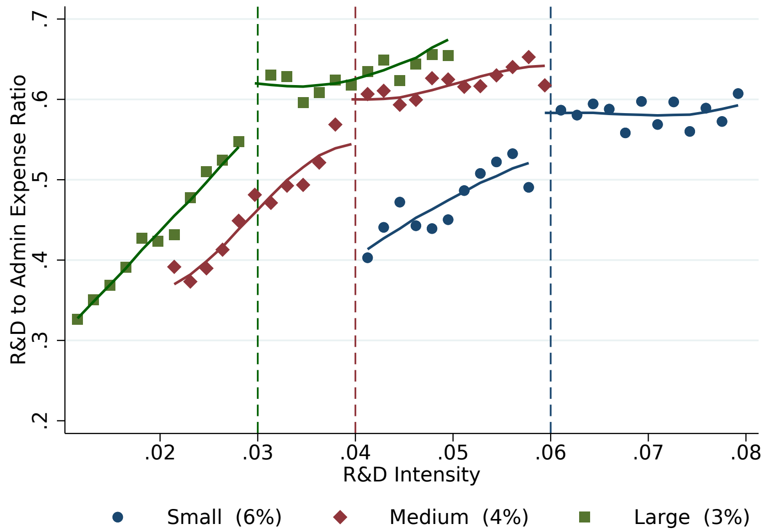
Appendix Graphs

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



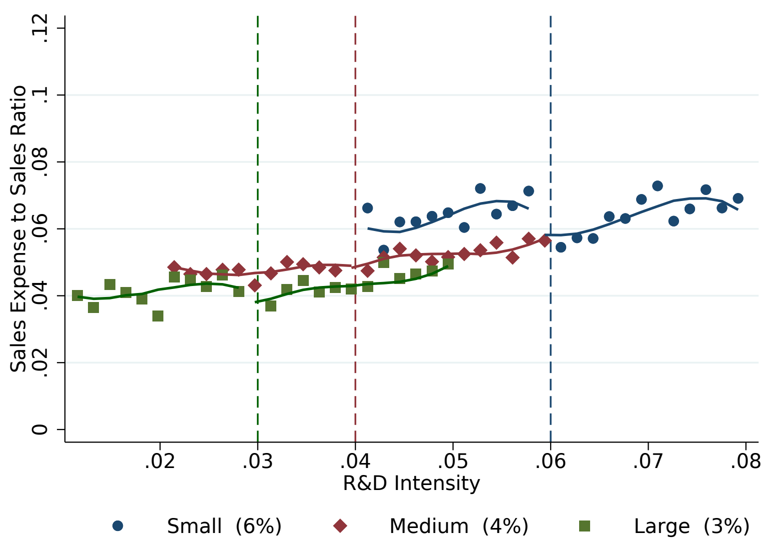
Notes: This figure plots the R&D intensity distribution of manufacturing firms conducting R&D during the period of 2005 to 2007. We include the firms with R&D intensity between 1% and 15%. There is a significant bunching of firms at the 5% threshold. Source: Annual Survey of Manufacturers. See Section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

Figure A.2: Alternative Empirical Evidence of Relabeling



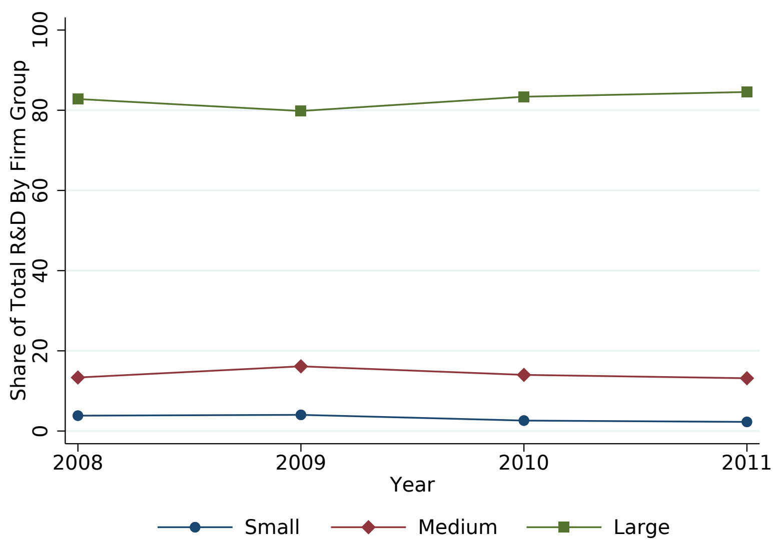
Notes: This figure summarizes the ratio of R&D to administrative expenses for small, medium, and large firms in our sample. The figure shows that this ratio jumps discontinuously across the thresholds of R&D intensity prescribed by the InnoCom program. This suggests that firms manipulate their reported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

Figure A.3: Lack of Manipulation of Sales Expenses



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

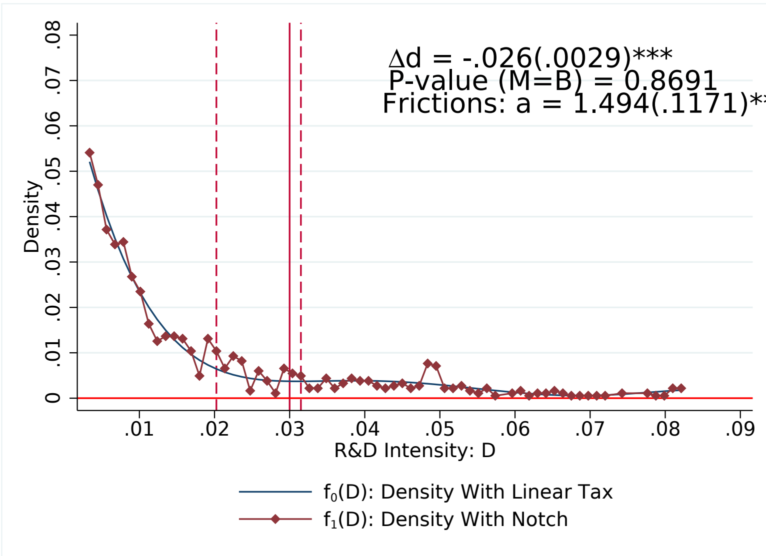
Figure A.4: Aggregate Implications



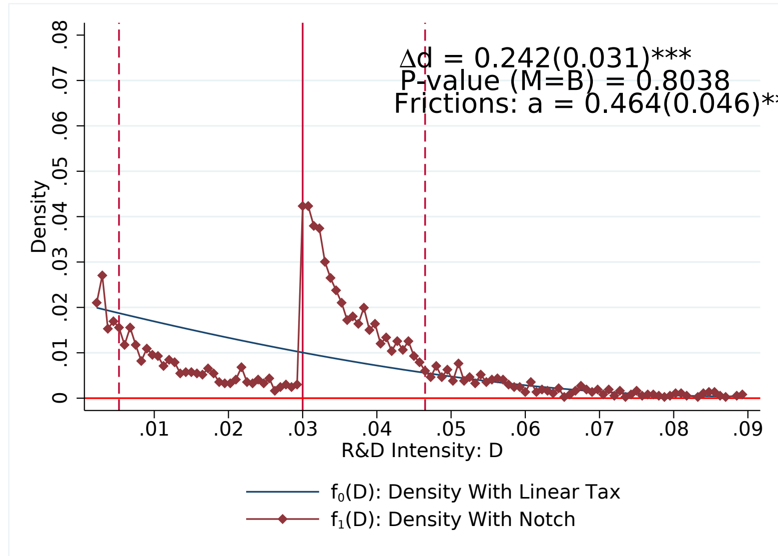
Notes: This figure summarizes the share of total R&D accounted for by the small, medium, and large firms in our sample. As the figure illustrates, the large firms account for more than 80% of total R&D and thus are the most important group for the aggregate implications of the policy.

Figure A.5: Robustness of Bunching Estimates

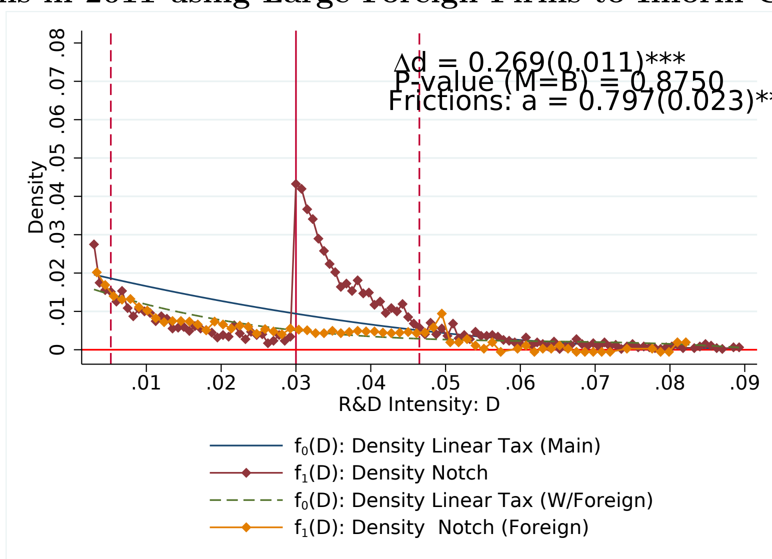
A. Placebo Test: Large Foreign Firms before 2008



B. Large Firms in 2011 (No Extensive Margin)

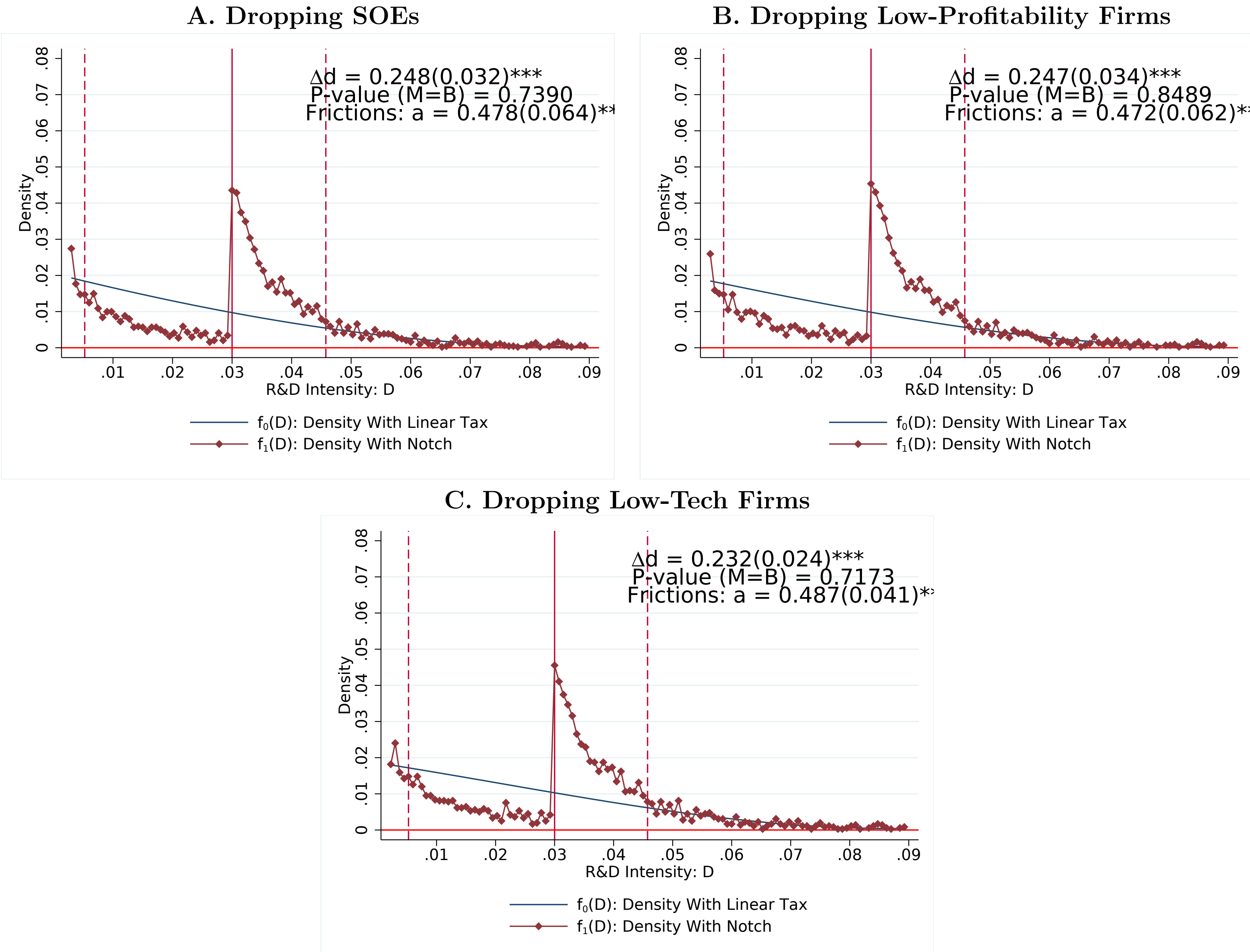


C. Large Firms in 2011 using Large Foreign Firms to Inform Counterfactual



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

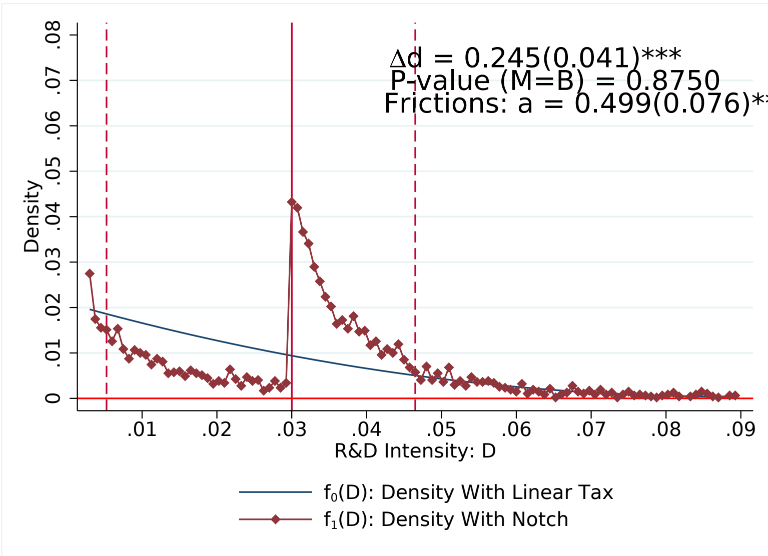
Figure A.6: Robustness of Bunching Estimates to Dropping Groups of Firms



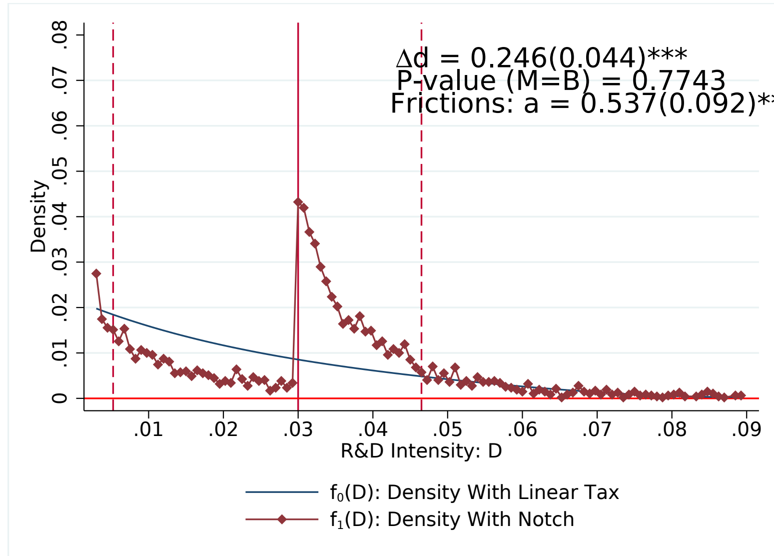
Notes: This figure presents robustness checks of the benchmark bunching analysis for large firms in 2011. In panel A, we drop state-owned enterprises. In panel B, we drop the bottom 20% of firms in terms of profitability. In panel C, we drop all firms not classified in the high-tech industries defined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

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

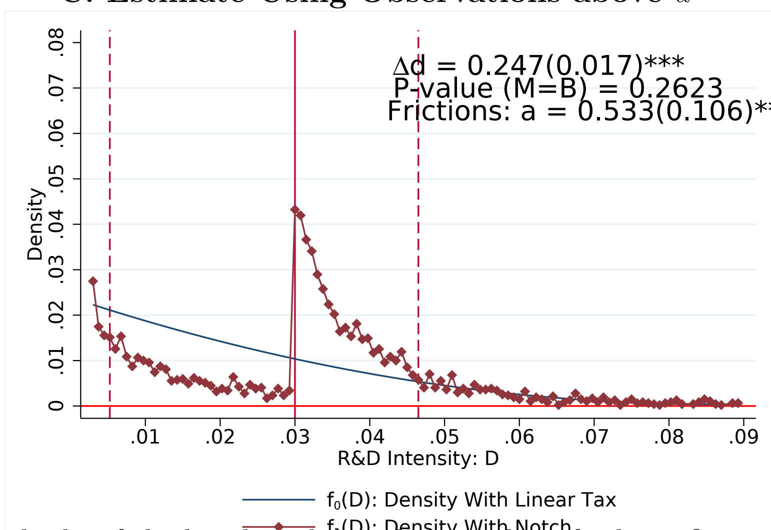
A. Second-Best Choice of Specification (p=3)



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



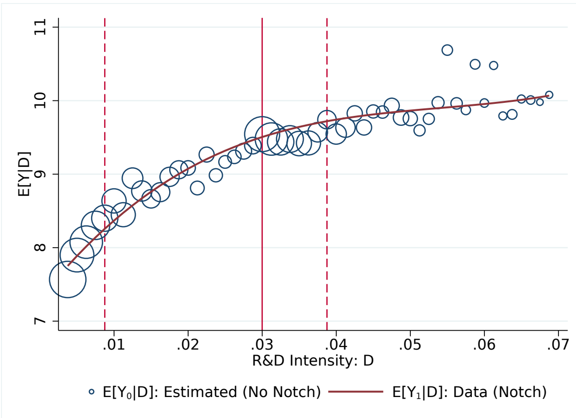
C. Estimate Using Observations above d^{*+}



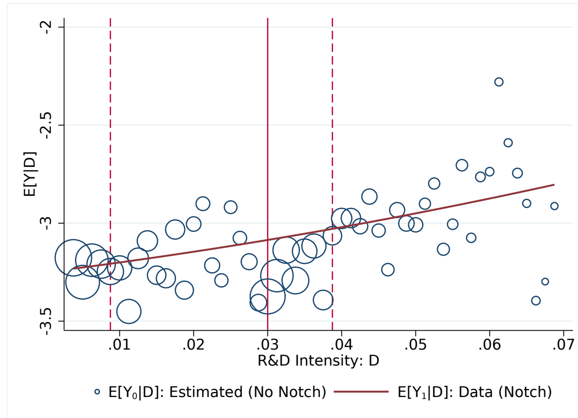
Notes: This figure conducts robustness checks of the benchmark bunching analysis for large firms in 2011. As discussed in Appendix D, we select (p, d^{*-}, d^{*+}) via cross-validation. In panel A, we use the second-best choice for the specification of (p, d^{*-}, d^{*+}). As in our benchmark case, $p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

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

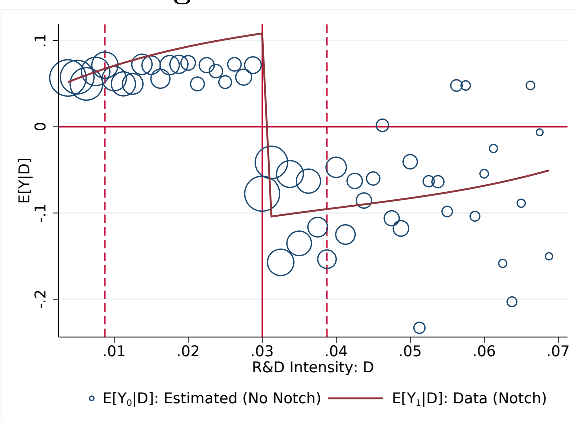
A. Log R&D Intensity in 2009



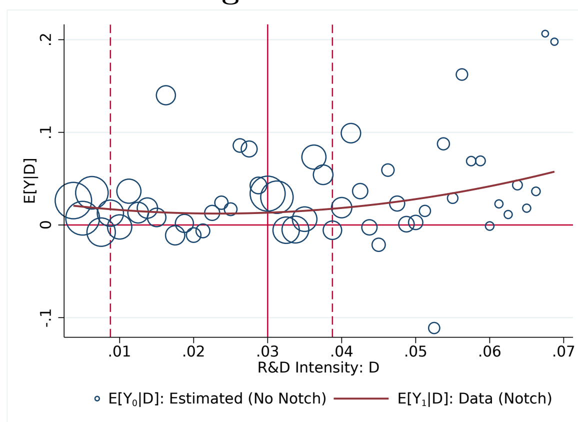
B. Log Administrative Cost-to-Sales Ratio in 2009



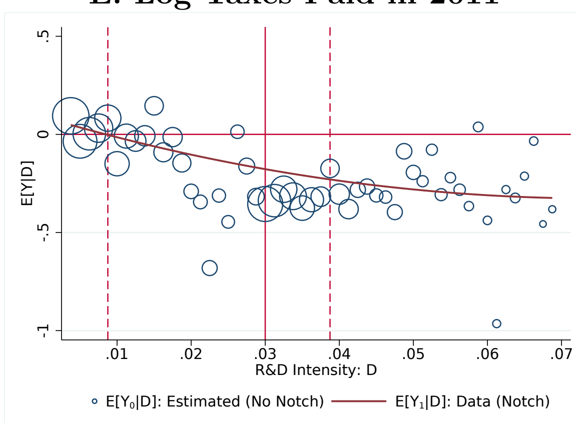
C. Log User Cost in 2009



D. Log TFP in 2011

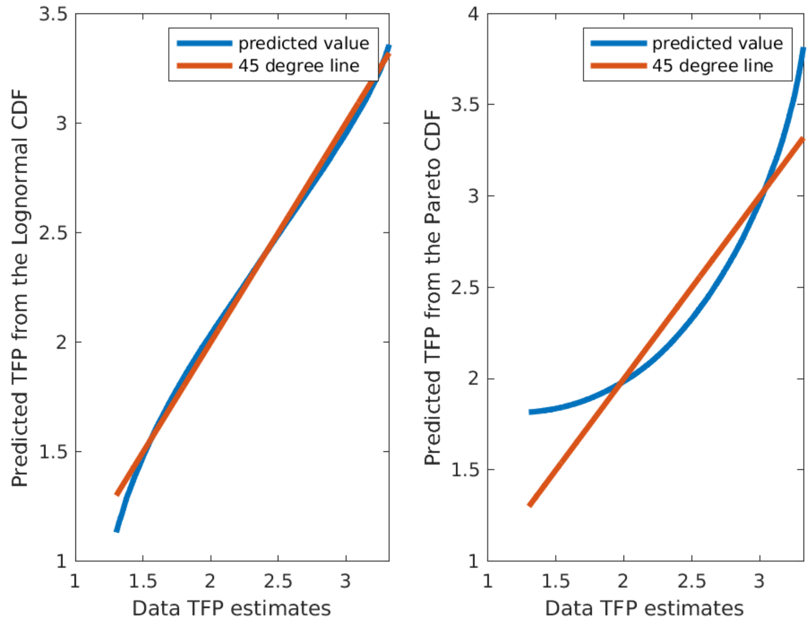


E. Log Taxes Paid in 2011



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

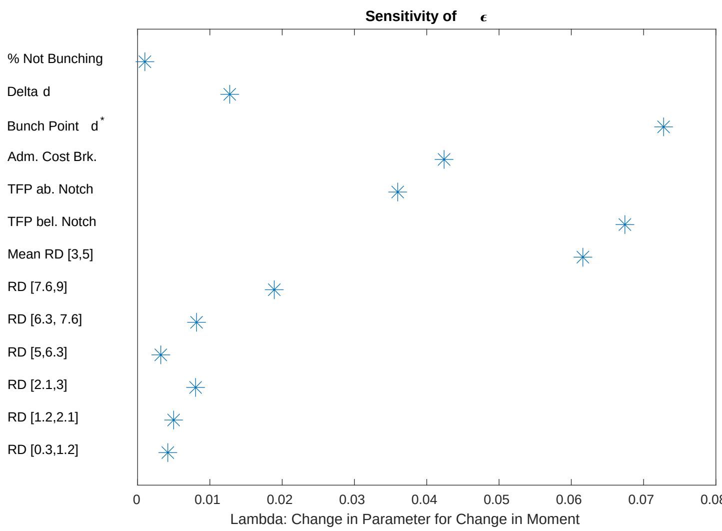
Figure A.9: Observed TFP and Predicted TFP under a Log-Normal vs. Pareto Distribution



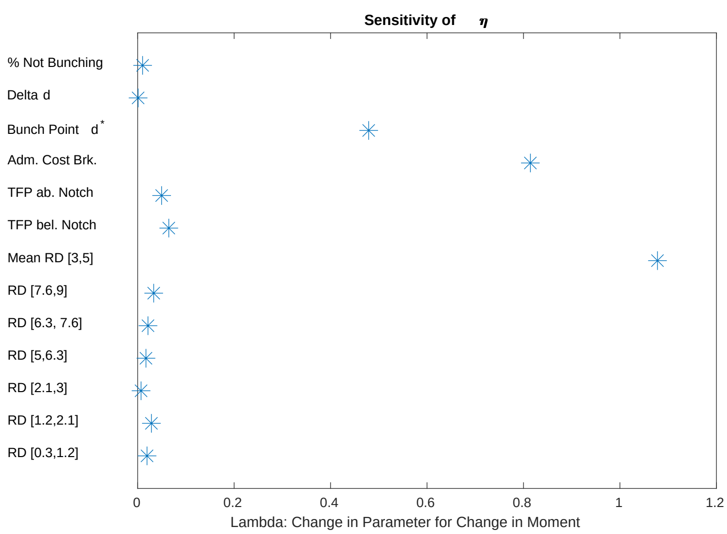
Notes: This figure reports the predicted TFP from imposing a log-normal CDF, a Pareto CDF, and the 45 degree linear line. For both cases, we trim the observed TFP data at 1% and 99%. The figure shows that the predicted TFP from imposing the log-normal CDF tracks observed TFP quite well. It thus provides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

Figure A.10: Sensitivity Analysis

A. Sensitivity Analysis for ε



B. Sensitivity Analysis for η



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

Table A.1: Estimates of Treatment Effects

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

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

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

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

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

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

Appendix Tables

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table A.6: Structural Estimates with Heterogeneous ε

A. Point Estimates

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

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

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

B. Simulated vs. Data Moments

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

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

Table A.7: Robustness of Structural Estimates

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

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

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

Simulated vs. Data Moments

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

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

References

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

_ , 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.

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

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

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.