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

CUHK Structural Estimation Method Workshop

November 12, 2023

Outline

1. Context
2. Descriptive Evidence
3. Descriptive Evidence
4. Model
5. Structural Estimation

Context

TABLE 1—REQUIREMENTS OF THE INNOCOM PROGRAM

Requirement	Before 2008	After 2008
R&D intensity	5 percent	6 percent if sales $<$ 50M 4 percent if 50M $<$ sales $<$ 200M 3 percent if sales $>$ 200M
Sales of high tech products	60 percent of total sales	
Workers with college degree	30 percent of workforce	
R&D workers	10 percent of workforce	
Certifying agency	Local Ministry of Science and Technology	Ministries of Science and Technology, Finance and National Tax Bureau

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

Descriptive Evidence

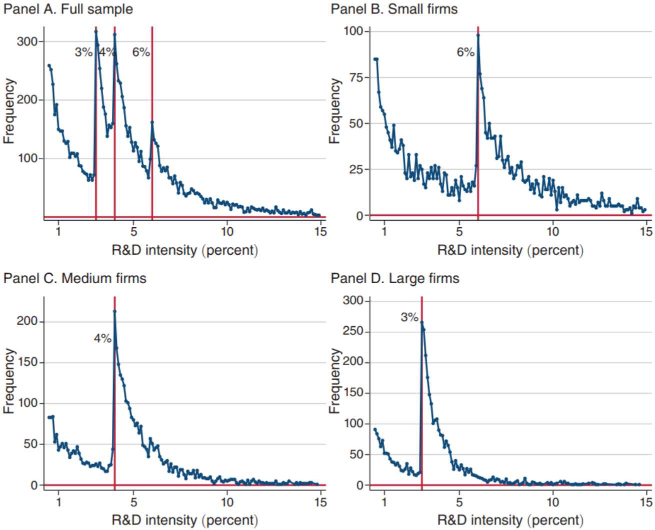

Figure 2. Bunching at Different Thresholds of R&D Intensity, 2011

Descriptive Evidence

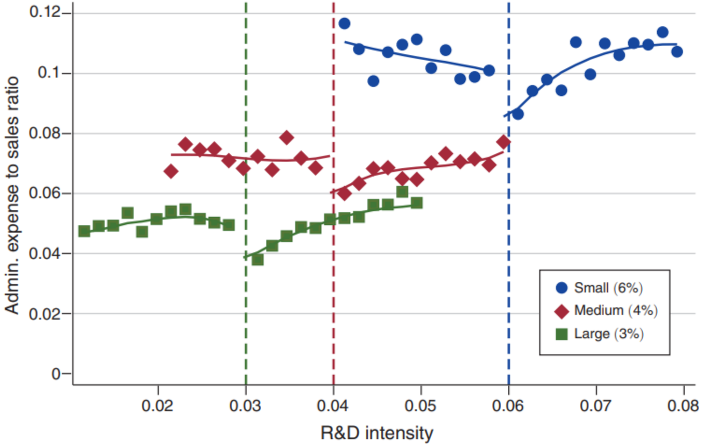

Figure 5. Empirical Evidence of Relabeling

Model

Consider a firm i with a unit cost function $c\left(\phi_{it}, w_t\right) = w_t \exp\left\{-\phi_{it}\right\}$, where w_t is the price of inputs. The term ϕ_{it} is log TFP and has the following law of motion:

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

where $D_{i,t-1}$ is R&D investment and $u_{i,t} \sim$ i.i.d. $N\left(0, \sigma^2\right)$.
We assume that the firm faces a demand function with a constant elasticity: $\theta > 1$. This setup implies that firm sales are given by $\theta\pi_{it}$ and that we can write expected profits as follows:

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

where $\tilde{\pi}_{it} \propto E\left[\exp\left\{(\theta-1)\phi_{it}\right\} \mid \phi_{i,t-1}\right]$ measures the non-R&D expected profitability of the firm.
We model firms' investment decision as a two-period problem.

Model: R&D Choice under a Linear Tax

We first model how R&D investment decisions would respond to a linear income tax:

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

In addition to the direct R&D investment cost D_{i1}, firms pay a cost $g (D_{i1}, \theta\pi_{i1})$ to adjust their R&D. Following the investment literature, we adopt a quadratic formulation for $g (D_{i1}, \theta\pi_{i1}) = b \times (\theta\pi_{1i}/2) [D_{i1}/ (\theta\pi_{1i})]^2$.

Model: R&D Choice under a Linear Tax

The optimal choice of D_{i1}^* is given by

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

$\mathrm{R\&D}$ intensity, defined as the $\mathrm{R\&D}$-to-sales ratio, has an ambiguous relationship with ϕ_{i1}. To see this, we express the firm's FOC in terms of the choice of R&D intensity, $d_{i1} = D_{i1}/\theta\pi_{i1}$, such that

$$\underbrace{-(1-t_1)\,(1+bd_{i1}^*)}_{\text{Increase in Investment Cost}}+\underbrace{\beta\,(1-t_2)\,\varepsilon(\theta-1)d_{i1}^*(\theta-1)\varepsilon-1\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}}_{\text{Productivity Gain from R\&D}}=0$$

Model: A Notch in the Corporate Income Tax

Assume now that the tax in the second period has the following structure, modeled after the incentives in the InnoCom program:

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

where $t_2^{LT} > t_2^{HT}$ and where LT/HT stands for low-tech/high-tech. In practice, firms with high $R\&D$ intensity may not participate in the program if other constraints prevent them from hiring a sufficient number of technical employees, if they do not obtain a significant fraction of their sales from high-tech products, or if the compliance and registration costs are too high. We model these constraints by assuming that firms pay a fixed cost of certification: $c \times \theta\pi_{1i}$, where c varies across firms.

Model: A Notch in the Corporate Income Tax

A firm decides whether to bunch by comparing the value of the firm from bunching, by setting $d_1^* = \alpha$, to the value of the firm at its optimal R&D intensity below the notch. The value-to-sales ratio of the firm conditional on bunching, $\Pi\left(\alpha \mid t_2^{HT}\right) / (\theta\pi_{i1})$, is given by

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

Similarly, the value-to-sales ratio at the interior optimal $d_{i1}^*, \Pi\left(d_{i1}^* \mid t_2^{LT}\right) / (\theta\pi_{i1})$, is

$$\frac{\Pi\left(d_{i1}^* \mid t_2^{LT}\right)}{\theta\pi_{i1}} \equiv (1-t_1)\frac{1}{\theta} + \beta\left(1-t_2^{LT}\right)d_{i1}^*(\theta-1)\varepsilon\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}} - (1-t_1)\,d_{i1}^*\left(1+\frac{bd_{i1}^*}{2}\right)$$

Model: Real and Relabeled R&D Investment under a Tax Notch

Denote a firm's reported level of R&D spending by $\tilde{D}_{i1}$. Firms qualify for the lower tax whenever $\tilde{D}_1 \geq \alpha\theta\pi_1$. We assume that firms face an expected cost of misreporting that is given by $h\left(D_{i1}, \tilde{D}_{i1}\right)$, which represents the likelihood of being caught and the punishment from the tax authority. We further assume that the cost of misreporting is proportional to the reported $\mathrm{R\&D}$ and depends on the percentage of misreported R&D, $\delta_{i1} = \left(\tilde{D}_{i1} - D_{i1}\right) / \tilde{D}_{i1}$, so that

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

where $\tilde{h}$ satisfies $\tilde{h}(0) = 0$ and $\tilde{h}'(\cdot) \geq 0$.

Model: Real and Relabeled R&D Investment under a Tax Notch

Notice first that if a firm decides not to bunch at the level $\alpha\theta\pi_1$, it does not have an incentive to misreport R&D spending, as doing so would not affect total profits or the tax rate. However, a firm might find it optimal to report $\tilde{D}_1 = \alpha\theta\pi_1$ even if it actually invested in a lower level of R&D. Conditional on bunching, the firm's optimal relabeling strategy solves the following problem:

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

Model: Real and Relabeled R&D Investment under a Tax Notch

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

$$0 = \underbrace{-(1-t_1)\left(1+bd_{i1}^{K*}\right)+\tilde{h}'\left(1-\frac{d_{i1}^{K*}}{\alpha}\right)}_{\text{Increase in Investment Cost and Reduction in Relabeling Cost}}$$

$$\underbrace{+\beta\left(1-t_2^{HT}\right)\varepsilon(\theta-1)d_{i1}^{K*(\theta-1)\varepsilon-1}\frac{\tilde{\pi}_{i2}}{(\theta\pi_{i1})^{1-(\theta-1)\varepsilon}}}_{\text{Productivity Gain from Real R\&D}}.$$

Model: Real and Relabeled R&D Investment under a Tax Notch

To gain further intuition, consider the simple case where $b = c = 0$. Using equation (2) to simplify $\Pi\left(d_{i1}^{K*}, \alpha \mid t_2^{HT}\right)$, it follows that firms decide to bunch when the following inequality holds:

$$\underbrace{\left(\frac{d_{i1}^{K*}}{\alpha}\right)^{(\theta-1)\varepsilon}\left(\frac{d_{i1}^{*}}{\alpha}\right)^{1-(\theta-1)\varepsilon}\left(\frac{1-t_2^{HT}}{1-t_2^{LT}}\right)\frac{1}{(\theta-1)\varepsilon}-\frac{d_{i1}^{K*}}{\alpha}}_{\text{Relative Profit from Bunching}}-\underbrace{\frac{\tilde{h}\left(\delta_{i1}^{*}\right)}{\alpha\left(1-t_1\right)}}_{\text{Relabeling Cost}}$$

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

Estimation Framework

- Parameterize the model:
  - Set $\theta = 5$ based on the survey by Head and Mayer (2014)
  - Given a value of θ, the persistence and volatility of log value-added of non-R&D performing firms map directly into ρ and σ^2, which yields the following calibrated values of $\rho = 0.725$ and $\sigma = 0.385$. This process implies a stationary normal distribution for the underlying productivity ϕ_1.
  - Set $\beta = 0.925$

Estimation Framework

- Parameterize the distributions of b and c, which we assume are i.i.d. across firms:
  - Assume b is log-normally distributed, $b \sim \mathcal{LN}\left(\mu_b, \sigma_b^2\right)$
  - c has an exponential distribution, $c \sim \mathcal{EXP}\left(\mu_c\right)$
- Adopt the following functional form for the costs of relabeling: $(\exp\{\eta\delta\} - 1)/\eta$, where δ is the fraction of reported $\mathrm{R\&D}$ corresponding to relabeling

Estimation Framework: Method of Simulated Moments (MSM)

to estimate the parameters $\Omega = \{\varepsilon, \eta, \mu_b, \sigma_b, \mu_c\}$

- We simulate productivity and adjustment and fixed costs for 30,000 firms.
- We determine whether each firm finds it optimal to bunch depending on the firm's optimal R&D investment conditional on not bunching and the optimal relabeling strategy conditional on bunching. Based on these firm-level decisions, we compute data moments that are analogous to those discussed in Section II.
- We obtain the simulated moments by repeating this process 10 times and averaging over these instances.

Estimation Framework: Method of Simulated Moments (MSM)

- Our estimate of Ω minimizes the difference between data moments and moments generated by the distribution of simulated firms as measured by the criterion function:

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

  where W is a bootstrapped covariance weighting matrix, and $m^D(\Omega)$ and $m^B(\Omega)$ are moment conditions based on the descriptive statistics and on the bunching estimator, respectively.

Estimation Framework: Method of Simulated Moments (MSM)

Our initial model relies solely on the moments in $m^D(\Omega)$ to estimate the model. Moment condition $m^D(\Omega)$ includes four types of moments based on the data patterns in Section II.

- The first set of moments uses information from the histogram of R&D intensity. We include the fraction of firms falling in three equally spaced intervals below the 3 percent notch (i.e., [0.003, 0.012], [0.012, 0.021], and [0.021,0.03]) . We summarize the top of the R&D intensity distribution by including moments that measure the fraction of firms falling in three equally spaced intervals between 5 percent and 9 percent (i.e., [0.05,0.063], [0.063,0.076], and $[0.076, 0.09]$).
- Second, we include the average R&D intensity for firms that potentially respond to the InnoCom program (i.e., over the interval $[0.03, 0.05]$).

Estimation Framework: Method of Simulated Moments (MSM)

- Third, we include the average TFP for firms below and above the notch.
- Finally, we include the drop in the administrative cost ratio from Figure 5. This last moment plays an important role in disciplining the costs of relabeling.

Estimation Framework: Method of Simulated Moments (MSM)

For robustness, we show that we obtain similar structural estimates when we also consider additional moments based on the bunching estimator $m^B(\Omega)$. These moments include the following: (i) the lower threshold of the excluded region d^{*-}; (ii) the fraction of firms in the excluded region that do not bunch a^*; and (iii) the percentage increase in R&D intensity over the excluded region Δd. In this case, our model parameters are additionally disciplined by the results from Figure 4.

- these moments jointly inform the three parameters that determine bunching: ε, η, and μ_c, providing additional over-identifying restrictions.

Estimates of Structural Parameters

TABLE 3—STRUCTURAL ESTIMATES

	TFP elasticity of R&D	Relabeling cost	Distribution of adjustment costs		Distribution of fixed costs
	ε	η	μ_b	σ_b	μ_c
Panel A. Point estimates					
Model 1: Excluding bunching moments					
Estimate	0.089	5.900	7.989	2.047	0.687
Standard error	(0.002)	(0.493)	(0.086)	(0.076)	(0.062)
Model 2: All moments					
Estimate	0.091	6.755	8.011	2.014	0.532
Standard Error	(0.002)	(0.449)	(0.075)	(0.073)	(0.012)

Estimates of Structural Parameters

	Data	Simulated: Model 1: excluding bunching	Simulated: Model 2: all moments
Panel B. Simulated versus data moments			
R&D dist. moments: $m^D(\Omega)$			
Below the notch (percent)			
$[0.3, 1.2]$	0.373	0.382	0.379
$[1.2, 2.1]$	0.113	0.157	0.146
$[2.1, 3]$	0.067	0.080	0.069
Above manipulated region (percent)			
$[5, 6.3]$	0.056	0.055	0.057
$[6.3, 7.6]$	0.026	0.037	0.038
$[7.6, 9]$	0.012	0.026	0.027
Mean R&D intensity $[3 \text{ percent}, 5 \text{ percent}]$	0.037	0.035	0.035
Average TFP below notch	−0.015	−0.017	−0.020
Average TFP above notch	0.027	0.023	0.025
Admin. cost ratio break at notch (%)	0.9	0.8	0.7
Bunching moments: $m^B(\Omega)$			
Bunching point d^{*-}	0.009	(0.009)	0.010
Increase in reported R&D: Δd	0.157	(0.124)	0.150
Fraction of firms not bunching	0.641	(0.738)	0.665

Implications

- A 100 percent increase in real R&D would increase TFP by 9 percent.
- Complier firms are, on average, 13.5 percent more productive than firms in the excluded region that do not comply with the policy. They also have idiosyncratic adjustment costs that are 24.3 percent lower than noncompliers, which indicates much better technological opportunities from R&D investment.
- 24.2 percent of the reported R&D investment is due to relabeling.

Code

Example: excluding bunching moments
MatlabStructuralCode/Baseline/Density Moments/Main.m