Tax-Induced Employment Distortion: Evidence from a Wage Deduction Tax Policy in China

Yi Lu * — Tsinghua
Wei Lv † — DUFE
Jianguo Wang ‡ — NUS
Wanyu Yang § — DUFE

This version: Nov 2019

Abstract

This paper employs China's first wave of economic census in 2004 to analyze the tax-induced employment distortion of a wage deduction limit policy in China. The distribution of the firm-level average per worker monthly wage shows clear evidence of bunching at the deduction limit point. Employing the methodology by Diamond and Persson (2017), we find that firms responded to this policy by increasing capital and substituting skilled labor with unskilled labor. There is no significant evidence of tax evasion by relabeling the labor cost as other deductible terms. By constructing a theoretical model, we find that with the relative wage of skilled to unskilled labor increasing by 1%, firms would decrease the ratio of skilled to unskilled labor by 3.1898%. In addition, firms reported 21% to 53% more nonexistent unskilled labor on average to claim wage deduction from the taxable income. As a result, tax revenue and GDP decreased by 2.74% and 1.51%, respectively.

Keywords: bunching, tax-induced employment distortion, elasticity of substitution between skilled and unskilled labor

JEL Classification:

*School of Economics and Management, Tsinghua University, Beijing, 100084, China (luyi@sem.tsinghua.edu.cn)
†Dongbei University of Finance and Economics, 116000, China
‡School of Economics, National University of Singapore, 117570, Singapore
§Institute for Advanced Economic Research, Dongbei University of Finance and Economics, 116000, China (wanyu.yang@dufe.edu.cn)

1 Introduction

Tax policies are on the top agenda of policy makers. They significantly affect firm behavior, and are critical to economic development. Policy designs, without carefully analyzing firms' responses, may not only fail their policy targets, but also create substantial distortions. In this paper, we examine a wage deduction tax policy that was enacted in early 1990s and lasted for 15 years in China. Specifically, the policy set a statutory province-level up limit that domestic enterprises (DEs) in China could deduct their wage bills from the taxable corporate income. The purpose of the policy was first to discourage high wage rates and second to prevent the erosion of tax base. However, the wage deduction tax schedule introduces a kink point into the firms' budget sets, which favors unskilled relative to skilled labor. As a response, firms may change their employment structures, which could then affect their production efficiency and invalidate the policy intentions.

To identify the tax-induced employment distortion, we draw an empirical analysis on China's first wave of economic census in 2004, which covers basic information and balance sheets for all manufacturing firms. The graphical results clearly demonstrate a sharp jump in the density distribution of average per worker monthly wage for DEs around the wage deduction limit point, suggesting firms' tax avoidance responses to the policy. We do not find the same pattern of bunching for foreign invested enterprises (FIEs) that can fully deduce wages, and also for DEs using the second wave of economic census in 2008 when the wage deduction limit policy was abolished. These confirm that the bunching of affected DEs to the policy is not spurious.

To understand how firms responded to the wage deduction limit, we follow the methodology by Diamond and Persson (2017), which is based on the premise that the relationship between the variables of interests and the average per worker monthly wage would be smooth in the absence of deduction limit. Hence, the relationship characterized in the non-manipulation regions can be used to construct a counterfactual relationship in the manipulation regions around the kink point. We find that firms did not relabel part of the labor cost as unemployment insurance, employment benefits and administration cost (that are fully deductible from the taxable income) in response to the wage deduction limit. Instead, firms increased their capital input and substituted skilled labor with unskilled labor. Finally, the reduced-form analysis shows that the wage deduction limit policy reduces firm productivity (measured by the total factor productivity) by around 6% relative to the mean level.

We then construct a theoretical model to calculate the tax and welfare effects of the wage deduction limit policy. Specifically, firms differ in their relative importance of skilled and unskilled labor to output production, and optimally choose their inputs given the tax policies. Comparing firm behavior under the wage deduction limit and that under a counterfactual tax policy when firms can fully deduct wages, we can derive the mass of firms that choose to bunch at the deduction limit in the former scenario. A crucial part of the model estimation is the estimation of the density distribution of average monthly wage per worker under the counterfactual scenario. To this end, we use two methods: (1) a parametric approach following Chetty et al. (2011) to estimate the counterfactual density from the observed density; and (2) a non-parametric approach using groups of firms that were not subject to the wage deduction limit of 960 as the counterfactuals and combining a difference-in-differences (DD) spirit. We estimate the substitution elasticity between skilled and unskilled labor to be $\hat{\sigma} = 3.6294$. This implies that when the wage ratio of skilled over unskilled labor increases by 1%, firms would decrease the ratio of skilled over unskilled labor by approximately 3.6294%.

Based on the baseline elasticity estimate and density distribution, we then conduct tax and welfare analyses of the wage deduction limit policy. The implementation of the wage deduction limit policy caused total tax revenue to decline by 2.95%. This contrasts the policy intention, which is to increase tax revenues by lowering the deductibles from the corporate income. Meanwhile, the wage deduction limit policy also reduced aggregate GDP by 1.72%, suggesting a significant distortionary loss.

To alleviate the concern that firms may manipulate their reporting of employment (e.g., adding *ghost workers* to inflate the employment) in response to the policy, we extend our model to incorporate this possibility. Specifically, we allow firms to over-report the number of unskilled labor, which includes a new parameter (i.e., the booking manipulation degree) in the model. To simultaneously pin down the two parameters of interests (i.e., the labor substitution elasticity and the booking manipulation degree), we explore the two wage deduction limits implemented in different regions (i.e., 800 RMB and 960 RMB) to construct two moment equations. Estimation results show that firms reported 21% to 53% more nonexistent unskilled labor to claim wage deduction from the taxable income. And the elasticity of substitution declines by around 12% to 3.1898. The changes in tax revenues and GDP are consistent; that is, tax revenues and GDP fall by 2.74% and 1.51%, respectively, in response to the policy.

Our work is connected to the literature on tax policies that target specific groups of workers and hence change firms' employment decisions. Examples include the work opportunity tax credit policy and the Indian employment credit policy in US, the apprenticeship job creation tax credit policy in Canada, and the payroll tax cuts policy for young workers in Sweden. Specifically, Katz (1998) analyzes the US wage subsidy policies for the disadvantaged boosted employment; Huttunen et al. (2013) examine the effect of payroll tax cut for older workers in Finland; Elias (2015) investigates the payroll tax cut policy for older and young workers in Spain; Saez et al. (2019) study the payroll tax cut policies for young workers in Sweden. The departure of our study is that we investigate a tax policy designed without the intention to benefit specific groups of workers, but indirectly distort firms' employment preference. And in addition to the policy effect on the targeted employment, we calculate the welfare consequences of the distortionary effects.

This paper also contributes to the literature on the substitution of skilled and unskilled labor. Most previous studies has tried to explain the rapid increase of college premium in U.S., and hence concentrated in estimating the substitution elasticity between college and high school labor (Katz and Murphy, 1992; Heckman et al., 1998; Krusell et al., 2000; Card and Lemieux, 2001; Acemoglu, 2002; Autor et al., 2008). See Katz and Autor (1999) for a review of this literature. One exception is Angrist (1995), which investigates the substitution between workers with sixteen years of schooling and those with less than twelve years of schooling in Palestine. The contribution of our work focuses on investigating a large emerging economy—China, and estimating the elasticity of substitution between high school and non-high school graduates, which is of more importance than the that between college and non-college labor in the developing country setting.

The present study is related to a recent literature that estimates behavioral responses to discontinuities in incentives created by kinked or notched policies.[1] The recent literature has applied the bunching methodology to explore various types of behavioral responses of firms, including splitting responses of large firms to a value-added tax (VAT) threshold (Onji, 2009), sales size adjusting responses to a threshold of VAT (Harju et al., 2016; Liu et al., 2017), employment size adjusting responses to thresholds in labor laws, and accounting and legal rules (Gourio and Roys, 2014; Garicano et al., 2016), substituting responses between labor and capital to a minimum wage rule (Harasztosi and Lindner, 2019), and R&D investment adjusting responses to a threshold associated with corporate tax cuts (Chen et al., 2019). This paper contributes to the literature by employing the methodology to study the substituting responses of firms between skilled and unskilled labor.

This paper is organized as follows. In Section 2, we describe the policy background, the data, and the bunching evidence motivating this paper. In Section 3, we present the reduced-form analysis of the behavioral responses of firms and the economic consequences. In Section 4, we develop a theoretical model for the welfare analysis of the behavioral responses of firms. In Section 5, we present results for the model estimation and welfare analysis. Section 6 concludes.

2 Background, Data, and Bunching Evidence

2.1 Wage Deduction Limit Policy

In 1978, China began to carry out economic reforms. However, the lack of capital and technology hampered the economic development. To attract foreign investment and introduce new technology, the government adopted a "dual-track" CIT scheme for domestic- and foreign-invested enterprises in the mainland. All DEs were governed by the Provisional Regulations of the People's Republic of China on Corporate Income Tax, which was promulgated in 1993; whereas FIEs were governed by the Income Tax Law of the People's Republic of China for Enterprises with Foreign Investment and Foreign Enterprises, which was promulgated in 1991. This "dual-track" CIT scheme lasted for around 15 years. In 2008, to provide a fair environment for all firms, China abolished the "dual-track' tax scheme and replaced it with a universal tax law—the Law of the People's Republic of China on Corporate Income Tax.

During the "dual-track" CIT scheme period, FIEs can fully deduct wage bill from taxable corporate income, but DEs can only claim up to a statutory wage limit.[2] Specifically, when the average per worker monthly wage was less than or equal to the statutory wage limit, DEs could

[1]The methodology based on kinked policies was developed by Saez (2010) and Chetty et al. (2011); the methodology based on notched policies was developed by Kleven and Waseem (2013). See Kleven (2016) for a review.

[2]DEs and FIEs were also charged with different CIT rates. Specifically, DEs were subject to CIT rates of 18%, 27%, and 33% with the taxable income less than 30,000 RMB, between 30,000 and 100,000 RMB, and more than 100,000 RMB, respectively. However, FIEs were subject to CIT at the reduced tax rate ranging from 15% to 33%, conditional on the location and the main business.

fully deduct wage bill; otherwise, they could only deduct wage bill from the taxable income for the CIT up to the imposed limit level for each employee. The historic changes in the monthly wage deduction limit levels are summarized in Table 1. The limit was initially set at 500 RMB by the Chinese State Administration of Taxation (SAT) in 1994, and was increased to 550 RMB in 1996 and further to 800 RMB in 2000. The local taxation bureaus were given the authority to inflate the limit up to 120% by the SAT, i.e., up to 600 RMB in 1994, 660 RMB in 1996, and 960 RMB in 2000, respectively. In July 2006, the SAT set the limit at 1600 RMB and canceled the adjustment power of local bureaus. After the "dual-track' CIT scheme was abolished in 2008, DEs were allowed to deduct the full wage bill as FIEs.

[Insert Table 1 Here]

Figure 1 shows the distribution of wage deduction limits across regions in 2004, our sample period. Specifically, three municipalities (Beijing, Shanghai, and Tianjin), a sub-provincial city (Qingdao), and six provinces (Fujian, Guangdong, Guizhou, Hunan, Jiangsu, and Zhejiang) adjusted the limit to 960 RMB. One municipality (Chongqing), all other sub-provincial cities, and 18 provinces (Anhui, Gansu, Guangxi, Hebei, Heilongjiang, Henan, Hubei, Hunan, Jiangxi, Jilin, Liaoning, Ningxia, Shaanxi, Shandong, Shanxi, Sichuan, Xinjiang, and Yunnan) implemented the regulated 800 RMB limit. Tibet adopted a wage deduction limit of 1200 RMB, due to the large-scale development strategy in west China. For three provinces (Hainan, Inner Mongolia, and Qinghai), the implemented limits were not documented.

[Insert Figure 1 Here]

2.2 Data and Variables

Our analysis is mainly based on the first Economic Census conducted by the China's National Bureau of Statistics (NBS) in 2004, covering all firms in the secondary and tertiary sectors in China. Basic information for each firm contained in the data include the location code, the industry, the ownership, and the employment by education. Meanwhile, the data also have firms' balance sheets, which record firm output, capital, revenue, profit, total wage bill, total employee benefit, administrative cost, tax, etc.

In this paper, we studies the tax avoidance behavior of domestic manufacturing firms, who were subject to the wage deduction policy in 2004. Specifically, we exclude three regions without information of wage deduction limits (i.e., Hainan, Inner Mongolia, and Qinghai) and Tibet (due to its special treatments). Meanwhile, one complication with the wage deduction limit of 800 RMB is that it also coincides with the monthly exemption threshold for individual income tax during our sample period.[3] To this end, we mainly focus on regions with the wage deduction limit of 960, and use regions with the wage deduction limit of 800 in robustness checks (after properly controlling

[3]According to the law of individual income tax, a certain level of income can been exempted from the monthly taxable income. Specifically, the basic exemption was set at 800 RMB in September 1980, and was increased to 1600 RMB in January 2006, 2000 RMB in March 2008, 3500 RMB in September 2011, and 5000 RMB in October 2010.

for the individual income tax effect). Our analysis sample contains 757,333 manufacturing firms, of which 665,299 were DEs and the rest were FIEs. Given that FIEs were not subject to our focal policy, we use them as the counterfactual control group in the placebo test and the identification.

With the basic information and balance sheet information, we calculate the average monthly wage per worker by dividing the total wage bill by 12 months and by the total employment. To investigate the effect of employment distortion, we classify firms' employment into two categories based on the education level, i.e., skilled labor and unskilled labor. Specifically, the skilled labor includes workers with a high school education or above, whereas the unskilled labor includes workers with a junior secondary school education or below. For the measurement of capital, we use the total capital. We calculate unemployment insurance per worker, employee benefits per worker, and administrative cost per worker as the ratios of the firm's total corresponding spending over the total employment.

Table 2 shows the summary statistics for DEs in our analysis sample. The average per worker monthly wage was about 838 RMB on average. The average number of employee was about 44 and the skilled over unskilled labor ratio was on average 0.98. The logarithm value of total capital was on average 7.21. Meanwhile, on average, DEs spent 315 RMB, 301 RMB, and 2,589 RMB per worker on unemployment insurance, employee benefits, and administrative cost, respectively. Moreover, the distributions for employment, the labor ratio, and the spendings per worker on unemployment insurance, employee benefits, and administrative cost were heavily right skewed.

[Insert Table 2 Here]

2.3 Bunching Responses

According to our focal policy, the tax schedule became nonlinear when a wage deduction limit was imposed. Therefore, DEs with the average monthly wage per worker higher than and closed to the deduction limit may have incentives to manipulate their average monthly wage per worker downwards to avoid paying higher taxes. To identify this potential tax avoidance response, we draw the density distribution of the average per worker monthly wage for DEs in 2004 when the limit was in place.

Figure 2 presents the density distribution of the average monthly wage per worker for DEs with the 960 RMB as the wage deduction limit. The solid curve plots the observed density. Clearly, there is a sharp jump in the distribution around the kink point of 960, implying DEs' strong responses to the wage deduction limit in the implemented nonlinear tax schedule. However, we also observe sharp spikes around other points, such as 900, 920 (corresponding to an annual wage of 11,000 RMB), 1,000, 1,045 (corresponding to an annual wage of 12,500 RMB), and 1,250 (corresponding to an annual wage of 15,000). These spikes may indicate firms' preference to set the monthly or annual wage rate at some integer levels. In the following analyses, we will control for these reference points in the estimation of the nonlinear tax policy effects, following Kleven and Waseem (2013).

[Insert Figure 2 Here]

One concern with Figure 2 is whether the bunching around the kink point reflects firms' behavioral responses to the wage deduction limit or represents a spurious correlation due to other unobserved factors. To alleviate this concern, we conduct several placebo tests; that is, inspect whether there are any bunching behavior at the location whether the wage deduction limit policy is not in effect. First, we explore the distribution pattern for FIEs, who can fully deduct wage bills. Figure 3A shows no spike around the corresponding kink point 960, which is consistent with the linear tax schedule for FIEs. Meanwhile, the observed densities also jump at integer wage levels as in Figure 2, implying the existence of reference points. Second, we study the distribution pattern of DEs located in regions with 800 RMB limit in 2004. As the 960 was not the threshold for wage deduction in these regions, we should not find any bunching at 960 if only our focal policy was in place. As shown in Figure 3B, there is no any jump at 960 in the observed distribution of the average monthly wage per worker for DEs with the 800 RMB as the wage deduction limit. Third, we use the second Economic Census conducted by the China's NBS in 2008, in which all firms (including DEs) were allowed to fully deduct total wage bills from their taxable corporate income. Hence, it is expected to have no bunching at 960 for DEs, which is confirmed in Figure 3C. Combined, these results reinforce our findings that firms respond to the wage deduction limit policy by manipulating their average monthly wage per worker. In the following sections, we investigate how firms manipulate and what are economic consequences of such manipulation.

[Insert Figures 3A-3C Here]

3 Reduced-form Analysis of Manipulation

In this section, we investigate how firms respond to the wage deduction limit policy, and provide a first look at its economic consequences. Specifically, for the reduced form identification, we follow the methodology by Diamond and Persson (2017), which is based on the premise that the relationship between the variables of interests and the average per worker monthly wage would be smooth when no ceiling limit of wage deduction were imposed. Hence, using the domain of firms not responding to the policy, we can estimate the relationship between variables of interests and the average per worker monthly wage. Then, applying this estimated relationship to the region of responding firms, we can back out the counterfactual values of interested variables. And the differences between the counterfactual and observed values identify the policy effect.

3.1 Estimation Framework

We elaborate the estimation methodology in this subsection. First, we group the studied DEs into average wage bins of 5, and estimate the counterfactual outcomes of interest in the manipulation region around the kink point $[w_{lower}, w_{upper}]$, by fitting a third order polynomial to the data of firms outside of the region:

$$y_j = \sum_{i=0}^{3} \theta_i (w_j)^i + \sum_{r \in R, 12R} \theta_r I_{\left\{ \frac{w_j}{r} \in \mathbb{N} \right\}} + \varepsilon_j, \tag{1}$$

where y_j is the average value of the outcome variable in the wage bin j; w_j is the average per worker monthly wage relative to the kink in terms of wage bins; and ε_j denotes the error term. In addition, to contain the reference points effects of integer wage rates, we add $R = \{800, 900, 1000, ..., 1300\}$ to control for monthly wage rounding and $12R = \{8.5K, 9K, 9.5K, 10K, ..., 16K\}$ for yearly wage rounding, where $\mathbb{N}$ is the set of natural numbers. See Kleven and Waseem (2013) for the same practice.

With the estimated coefficients from equation (1), we calculate the counterfactual relationship between the average wage and variables of interest inside the manipulation region $[w_{lower}, w_{upper}]$ as $\hat{y}_j(w_j, \hat{\theta}) = \sum_{i=0}^{3} \hat{\theta}_i (w_j)^i + \sum_{r \in R, 12R} \hat{\theta}_r I_{\{\frac{w_j}{r} \in \mathbb{N}\}}$.

Second, we compute the average counterfactual values of outcome variables inside the region $[w_{lower}, w_{upper}]$ with the following equation

$$\begin{aligned}
&E(y_j(w_j, \theta)|w_j \in [w_{lower}, w_{upper}], \text{no wage deduction limit}) \\
&= \int_{w_{lower}}^{w_{upper}} E(y_j(w_j, \theta)|\text{no wage deduction limit}) \\
&\times \frac{Pr(w_j|\text{no wage deduction limit})}{\int_{w_{lower}}^{w_{upper}} Pr(w_s|\text{no wage deduction limit}) dw_s} dw_j \\
&= \int_{w_{lower}}^{w_{upper}} \hat{y}_j(w_j, \hat{\theta}) \times \frac{\widehat{c}(w_j)}{\int_{w_{lower}}^{w_{upper}} \widehat{c}(w_s) dw_s} dw_j,
\end{aligned}$$

where $\widehat{c}(w_j)$ presents the counterfactual number of firms in the wage bin j when no deduction limit is imposed.

Third, we estimate the Intention to Treat (IIT) effect as follows

$$\begin{aligned}
ITT &= E(y_j(w_j, \theta)|w_j \in [w_{lower}, w_{upper}], \text{with wage deduction limit}) \\
&- E(y_j(w_j, \theta)|w_j \in [w_{lower}, w_{upper}], \text{no wage deduction limit}) \\
&= \frac{\sum_{w_j \in [w_{lower}, w_{upper}]} y_j(w_j, \theta)}{N_{w_j \in [w_{lower}, w_{upper}]}} - \int_{w_{lower}}^{w_{upper}} \hat{y}_j(w_j, \hat{\theta}) \times \frac{\widehat{c}(w_j)}{\int_{w_{lower}}^{w_{upper}} \widehat{c}(w_s) dw_s} dw_j,
\end{aligned} \tag{2}$$

where $N_{w_j \in [w_{lower}, w_{upper}]}$ denotes the total amount of firms with the observed average wage ranging from w_{lower} to w_{upper}.

A crucial element to obtain the ITT estimates is the counterfactual density of the average per worker monthly wage under the linear tax schedule when DEs can fully deduct the wage bill from the taxable income. To this end, we follow the empirical framework by Chetty et al. (2011) to estimate the counterfactual density from the observed density. The estimation methodology relies on the assumptions that the density of the average wage would be smooth in absence of wage deduction limit, and that firms with the average wage much larger than the deduction limit face a very high adjustment cost and hence would not respond to the imposed wage deduction limit. With these assumptions, we estimate the counterfactual density by excluding the observations in the region around the kink point and fitting a polynomial to the observed counts in each bin with the condition that the excess bunching mass equals the missing mass. Specifically, the estimation equation is

$$c(w_j) = \sum_{i=0}^{q} \beta_i (w_j)^i + \sum_{r \in R, 12R} \rho_r I_{\{\frac{w_j}{r} \in \mathbb{N}\}} + \sum_{i=w_{lower}}^{w_{upper}} \gamma_i I_{\{w_j = i\}} + \epsilon_j, \tag{3}$$

where $c(w_j)$ is the number of firms in the wage bin j; q is the order of the polynomial; and $[w_{lower}, w_{upper}]$ denotes the width of the excluded region around the kink point (i.e., the fraction of firms choosing to bunch at the kink point). Following Diamond and Persson (2017), we choose the values of q, w_{lower}, and w_{upper} based on a 5-fold cross-validation method. To address the problem of reference points effects, we add controls for integer wage levels as in equation (1).

After obtaining the estimated coefficients from equation (3), we calculate the counterfactual density distribution as $\hat{c}(w_j) = \sum_{i=0}^{q} \hat{\beta}_i (w_j)^i + \sum_{r \in R, 12R} \hat{\rho}_r I_{\{\frac{w_j}{r} \in \mathbb{N}\}}$. The excess mass of bunching is estimated as $\widehat{B} = \sum_{w_{lower}}^{\bar{w}} (c(w_j) - \widehat{c}(w_j))$, the missing mass is $\widehat{M} = \sum_{\bar{w}+1}^{w_{upper}} (\widehat{c}(w_j) - c(w_j))$, and the normalized bunching mass is defined as $\hat{b} = \widehat{B} / (\sum_{w_j \in [-4,5]} \hat{c}(w_j)/10)$. The red and dotted curve in Figure 2 plots the estimated counterfactual density. Specifically, we set the excluded region as [930, 1155], normalize the average wage with respect to the kink point 960, employ a fourth-order polynomial, and control for the wage rounding points.

With the estimated counterfactual density, we can get the ITT estimates from equation (2). Standard errors are estimated by using a parametric bootstrap procedure. Specifically, following Chetty et al. (2011), we redraw the estimated vector of errors ε_j in equation (1) with replacement to generate a new sample and calculate a new estimate ITT. We repeat this procedure 200 times and get the standard error of the estimate as the standard deviation of the 200 new estimates.

3.2 What to Manipulate

The bunching evidence in Figure 2 suggests that DEs respond to the wage deduction limit policy by adjusting their average monthly wage per worker. We then investigate what variables firms manipulate. One possible manipulation is that firms relabel part of labor wage as unemployment insurance, employee benefits, or administration cost, which are deductible from the CIT taxable income. Hence, the effective payment that workers receive do not change. To investigate this potential manipulation, we estimate the ITT effects of the imposed wage deduction limit on firm expenditure on unemployment insurance, employee benefits, or administration cost, respectively. Results are summarized in columns 1-3 of Table 3. All the three ITT estimates are statistically insignificant with small magnitude, suggesting that firms did not respond to the limit in the relabeling way.

[Insert Table 3 Here]

Second, there are some anecdotal evidence reporting that in response to the wage deduction limit, firms inflate employment by adding "ghost workers"; that is, workers exist on the book but not in the reality. This helps lowering the average monthly wage per worker and achieving the threshold for tax reduction. Without data on the number of *ghost workers*, it is hard for us to directly examine this potential manipulation. Instead, we hypothesize that if firms responded in this way, the real firm operation would not be significantly affected, and hence, key performance indicators would not change. To this end, we examine the policy effect on firm capital. Results are reported in column 4 of Table 3. The ITT estimate for capital is 0.3556, statistically significant at 1% level. This suggests that firms did respond to the wage deduction limit policy by increasing firm capital, alleviating the concern that firms only manipulated the booking.

Third, firms could change their employment structure to lower the average monthly wage per worker. Specifically, they can substitute skilled labor with unskilled ones, as the latter receives lower wage than the former. To examine this possibility, we, in column 5 of Table 3, investigate the policy effect on employment structure, that is, the ratio of unskilled over skilled labor. We find an estimated coefficient of 0.3610 statistically significant at 5% level, confirming that firms changed their employment structure in response to the policy.

In summary, we find that the wage deduction limit policy changes firm operation in some meaningful aspects, i.e., employment structure and capital. But, without direct information, we cannot fully rule out the possibility that firms may manipulate the booking of employment. We will incorporate these evidence into the welfare analysis of the wage deduction limit policy.

3.3 A First Look at Performance

Before moving to the structural estimation of welfare implication, we take a first look at the reduced-form effect of the wage deduction limit policy on firm productivity. Specifically, to calculate firm productivity, we assume a Cobb-Douglas production function with the same share of labor cost in total output in each 2-digit industry, and calculate the TFP in logarithm term as the difference of value-added and production factors in logarithm term.[4] Estimation results are reported in column 6 of Table 2. The estimate is -0.0741 with a statistically significance at 1% level, corresponding to around 6% decline. This result suggests that the manipulation of firm operation in response to the wage deduction limit policy damages firm performance.

4 Theoretical Model

To understand tax and welfare implications of the wage deduction limit policy, we develop a stylized model of employment decisions by firms in response to the policy. Here, we do not consider that firms can inflate the booking number of workers and focus on the optimal adjustment of employment structure. Later, we will extend the model to incorporate the possibility that firms can manipulate the booking by adding some *ghost workers*.

[4]Following Hsieh and Klenow (2009), we estimate the labor cost share in each industry by assuming that the total labor cost is a constant multiple of the wage compensation across all firms and that for the whole market, the aggregate total labor cost makes up 50% of the aggregate value-added.

4.1 Setup

We assume a monopolistic competitive market with the representative consumer's utility function being

$$U = \left(\int_{j \in J} q_j^{\beta} dj \right)^{\frac{1}{\beta}}, \tag{4}$$

where U denotes the utility level; q_j denotes the consumption units of firm j's product; J denotes the full set of firms in the market; and $\beta \in (0, 1)$ is the substitution parameter.

Hence, for firm j, its demand function is $p_j = q_j^{-(1-\beta)} P Q^{1-\beta}$, where $P = \left(\int_{j \in J} p_j^{-\frac{\beta}{1-\beta}} dj \right)^{-\frac{1-\beta}{\beta}}$ denotes the price index; and $Q = \left(\int_{j \in J} q_j^{\beta} dj \right)^{\frac{1}{\beta}}$ denotes the quantity index.

The production function by firm j is assumed to be

$$q_j = K_j^{\alpha_s} \left[(\lambda_j H_j^{\frac{\sigma-1}{\sigma}} + (1-\lambda_j) L_j^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \right]^{1-\alpha_s}, \tag{5}$$

where K_j, H_j, and L_j denote capital, skilled labor, and unskilled labor, respectively; α_s denotes the share of capital in total output in sector s; and λ_j is factor augmenting technology term of skilled labor, varying across firms to generate firm heterogeneity. σ is our parameter of interest, capturing the elasticity of substitution between skilled and unskilled labor.

4.2 Optimal Decision

To explore the effect of the wage deduction limit, we compare firms' optimal choices under implemented nonlinear tax schedule with those under a counterfactual linear tax schedule.

Specifically, we start with the counterfactual linear tax schedule, in which firms can fully deduct wage bills. The profit function is written as

$$(\pi_j)_{ct} = (1-\tau)\left(p_j q_j - w_H H_j - w_L L_j - r K_j\right), \tag{6}$$

where w_H, w_L, and r denotes the input prices of skilled labor, unskilled labor, and capital, respectively; and τ denotes the corporate tax rate. Denote $w_j^r \equiv \frac{w_H H_j + w_L L_j}{H_j + L_j}$ as the monthly average wage level.

Maximizing the profit function, we obtain the optimal decision as

$$\frac{(H_j)^*_{ct}}{(L_j)^*_{ct}} = \left[\frac{w_L}{w_H} \frac{\lambda_j}{1-\lambda_j} \right]^{\sigma}, \tag{7}$$

and

$$(w_j^r)^*_{ct} = \frac{w_H \left[\frac{w_L}{w_H} \frac{\lambda_j}{1-\lambda_j} \right]^{\sigma} + w_L}{\left[\frac{w_L}{w_H} \frac{\lambda_j}{1-\lambda_j} \right]^{\sigma} + 1}. \tag{8}$$

Next, we consider the implemented nonlinear tax schedule under which firms can only deduct wage bill up to a pre-set limit. Let $\bar{w}$ denote the monthly average wage deduction limit. And the deductible wage bill for firm j is

$$DC_j = min\{\bar{w}, w_j^r\}N_j^r, \tag{9}$$

where $N_j^r = H_j + L_j$ denotes the reported total number of employment.

Hence, the profit function becomes

$$\begin{aligned}\pi_j = p_j q_j - (w_H H_j + w_L L_j + rK_j) - \tau(p_j q_j - rK_j - DC_j) \\ - C \times I_{\left\{\frac{(H_j)^*_{ct}}{(L_j)^*_{ct}} > D\right\}} \times (L_j + H_j - (L_j)^*_{ct} - (H_j)^*_{ct}).\end{aligned} \tag{10}$$

To capture that firms with high ratio of skilled over unskilled labor have large costs of manipulating the employment structure, we add a fixed cost of adjustment $C \times I_{\left\{\frac{(H_j)^*_{ct}}{(L_j)^*_{ct}} > D\right\}} \times (L_j + H_j - (L_j)^*_{ct} - (H_j)^*_{ct})$, where $C = \tau\bar{w}$; and $D > (\bar{w} - w_L)/(w_H - \bar{w})$.

For firms with $(w_j)^*_{ct} \le \bar{w}$ (i.e. $\lambda_j \le 1 - 1/\left[1 + \left(\frac{\bar{w}-w_L}{w_H-\bar{w}}\right)^{\frac{1}{\sigma}} \frac{w_H}{w_L}\right] \equiv \lambda_1$), we have $(H_j)^*_{ct}/(L_j)^*_{ct} \le (\bar{w} - w_L)/(w_H - \bar{w}) < D$. Hence, profit functions (6) and (10) are the same. Consequently, firms are unaffected by the imposed wage deduction limit, and choose the same optimal solution as that under the counterfactual linear tax schedule, i.e.,

$$\frac{(H_j)^*}{(L_j)^*} = \frac{(H_j)^*_{ct}}{(L_j)^*_{ct}}, \tag{11}$$

and

$$(w_j^r)^* = (w_j^r)^*_{ct}. \tag{12}$$

For firms with $(w_j)^*_{ct} > \bar{w}$ (i.e. $\lambda_j > \lambda_1$), define $\bar{\lambda} \equiv \frac{w_H D^{\frac{1}{\sigma}}}{w_H D^{\frac{1}{\sigma}} + w_L}$. Given that $D > (\bar{w} - w_L)/(w_H - \bar{w})$, we have that $\lambda_1 < \bar{\lambda}$. Hence, when $\lambda_1 < \lambda_j \le \bar{\lambda}$, we have $(H_j)^*_{ct}/(L_j)^*_{ct} \le D$, and the profit function becomes

$$\pi_j = (1 - \tau)(p_j q_j - w_H H_j - w_L L_j - rK_j) - I_{\{w_j^r > \bar{w}\}} \tau(w_j^r - \bar{w})N_j^r. \tag{13}$$

In this scenario, the firms have two options for the monthly average wage rate. First, set it at $\bar{w}$. Second, solve the maximization of the new profit function (13), which generates

$$\frac{(H_j)^*}{(L_j)^*} = \left[\frac{w_L - \tau\bar{w}}{w_H - \tau\bar{w}} \frac{\lambda_j}{1 - \lambda_j}\right]^\sigma, \tag{14}$$

and

$$(w_j^r)^* = \frac{w_H \left[\frac{w_L - \tau\bar{w}}{w_H - \tau\bar{w}} \frac{\lambda_j}{1-\lambda_j}\right]^\sigma + w_L}{\left[\frac{w_L - \tau\bar{w}}{w_H - \tau\bar{w}} \frac{\lambda_j}{1-\lambda_j}\right]^\sigma + 1}. \tag{15}$$

Comparing the profits from these two options, we have when $\lambda_1 < \lambda \le 1 - 1/\left[1 + \left(\frac{\bar{w}-w_L}{w_H-\bar{w}}\right)^{\frac{1}{\sigma}} \frac{w_H - \tau\bar{w}}{w_L - \tau\bar{w}}\right] \equiv$

$\lambda_2 < \bar{\lambda}$, the firms are affected by the wage deduction limit, and choose to bunch at the corner solution

$$\frac{(H_j)^*}{(L_j)^*} = \frac{\bar{w} - w_L}{w_H - \bar{w}}, \tag{16}$$

and

$$(w_j^r)^* = \bar{w}. \tag{17}$$

When $\lambda_2 < \lambda \leq \bar{\lambda}$, firms are also affected, but do not set monthly average wage per worker at $\bar{w}$. Instead, they choose the new optimal solution described in equations (14) and (15).

Finally, for firms with $\lambda_j > \lambda_3 \equiv \bar{\lambda}$, we have that $(H_j)^*_{ct}/(L_j)^*_{ct} > D$, and hence the profit function is given by

$$\pi_j = (1-\tau)\left(p_j q_j - w_H H_j - w_L L_j - rK_j\right) - \tau(w_j^r - \bar{w})N_j^r - C(L_j + H_j - (L_j)^*_{ct} - (H_j)^*_{ct}). \tag{18}$$

Maximizing this profit function, we have that the optimal solution as

$$\frac{(H_j)^*}{(L_j)^*} = \frac{(H_j)^*_{ct}}{(L_j)^*_{ct}}, \tag{19}$$

and

$$(w_j^r)^* = (w_j^r)^*_{ct}. \tag{20}$$

In other words, in this scenario, the firms are unaffected by the kink introduced into the tax schedule.

In summary, the optimal choices of the average per worker monthly wage under the non-linear tax schedule is

$$(w_j^r)^* = \begin{cases} (w_j^r)^*_{ct} & \text{if } \lambda_j \leq \lambda_1 \\ \bar{w} & \text{if } \lambda_j \in (\lambda_1, \lambda_2] \\ \dfrac{w_H\left[\frac{w_L-\tau\bar{w}}{w_H-\tau\bar{w}}\frac{\lambda_j}{1-\lambda_j}\right]^{\sigma}+w_L}{\left[\frac{w_L-\tau\bar{w}}{w_H-\tau\bar{w}}\frac{\lambda_j}{1-\lambda_j}\right]^{\sigma}+1} & \text{if } \lambda_j \in (\lambda_2, \lambda_3] \\ (w_j^r)^*_{ct} & \text{if } \lambda_j > \lambda_3 \end{cases}. \tag{21}$$

Figure 4 shows the comparison of firms' optimal wage choices under the linear and non-linear tax schedules, respectively.

[Insert Figure 4 Here]

4.3 Implications for Bunching

A firm who bunches at $\bar{w}$ with the lowest λ_j (i.e., $\lambda_j = \lambda_1$) chooses the same skilled over unskilled labor ratio $\frac{(H_j)^*}{(L_j)^*} = \frac{(H_j)^*_{ct}}{(L_j)^*_{ct}} = \frac{\bar{w}-w_L}{w_H-\bar{w}}$ and the same reported average per worker monthly wage $(w_j^r)^* = (w_j^r)^*_{ct} = \bar{w}$ under the implemented nonlinear as those under the counterfactual linear tax schedule. A firm who bunches at $\bar{w}$ with the highest λ_j (i.e., $\lambda_j = \lambda_2$) chooses $\frac{(H_j)^*}{(L_j)^*} = \frac{\bar{w}-w_L}{w_H-\bar{w}}$

and $(w_j^r)^* = \bar{w}$ under the implemented nonlinear policy, and would choose $\frac{(H_j)^*_{ct}}{(L_j)^*_{ct}}$ and $(w_j^r)^*_{ct}$ under the counterfactual linear policy.

Hence, all firms with $\lambda_j \in [\lambda_1, \lambda_2]$ (or $(w_j^r)^*_{ct} \in [\bar{w}, \bar{w} + \Delta w]$) bunch at $\bar{w}$ under the nonlinear tax schedule, where

$$\begin{aligned}\Delta w &\equiv (w^r_{\lambda_2})^*_{ct} - \bar{w} \\ &= \frac{w_H \left[\frac{w_L}{w_H}\frac{w_H - \tau\bar{w}}{w_L - \tau\bar{w}}\right]^\sigma \frac{\bar{w} - w_L}{w_H - \bar{w}} + w_L}{\left[\frac{w_L}{w_H}\frac{w_H - \tau\bar{w}}{w_L - \tau\bar{w}}\right]^\sigma \frac{\bar{w} - w_L}{w_H - \bar{w}} + 1} - \bar{w} \\ &= \varphi(w_L, w_H, \bar{w}, \tau, \sigma).\end{aligned} \tag{22}$$

Therefore, the excess fraction of bunching is given by

$$B = \int_{\lambda_1}^{\lambda_2} f(\lambda_j) d\lambda_j = \int_{\bar{w}}^{\bar{w}+\Delta w} \hat{c}((w_j^r)^*_{ct}) d(w_j^r)^*_{ct} \simeq \hat{c}(\bar{w})\Delta w,$$

where $\hat{c}((w_j^r)^*_{ct})$ denotes the density distribution of the reported monthly average wage under the counterfactual tax schedule; and the second approximation is based on the assumption (as in Saez (2010) and Chetty et al. (2011)) that the counterfactual density $\hat{c}((w_j^r)^*_{ct})$ is uniform around the deduction limit $\bar{w}$. We then have that $\Delta w = B/\hat{c}(\bar{w}) = b$. Hence, we can estimate σ as a function of observable parameters $(w_L, w_H, \bar{w}, \tau, r)$ and the empirically estimable variable $\hat{b}$.

In addition, comparing the ratio H/L for the firms affected by the nonlinear tax schedule, we have that

$$\frac{(H_j)^*_{ct}}{(L_j)^*_{ct}} = \left[\frac{w_L}{w_H}\frac{\lambda_j}{1-\lambda_j}\right]^\sigma > \frac{(H_j)^*}{(L_j)^*} = \left[\frac{w_L - \tau\bar{w}}{w_H - \tau\bar{w}}\frac{\lambda_j}{1-\lambda_j}\right]^\sigma.$$

Hence, the introduction of the wage deduction limit reduces the relative employment of the skilled labor, which is consistent with the reduced-form evidence in Section 3.3.

5 Model Estimation and Welfare Analysis

In this section, we present our estimation procedure of the aforementioned model, and conduct a welfare analysis of the wage deduction limit policy.

5.1 Model Estimation

We estimate the previous model using the 2004 Economic Census. Specifically, we set interest rate $r = 5.58\%$ in 2004 according to the World Bank. Meanwhile, we estimate wage rates w_L and w_H from the census data using the following two conditions:

$$w_L * \sum_j L_j + w_H * \sum_j H_j = \sum_j \text{total\_wage\_bill}_j \tag{23}$$

and

$$Median\left(\frac{H_j}{L_j}\right) = \frac{\bar{w} - w_L}{w_H - \bar{w}} \tag{24}$$

for firms bunching around the kink point 960. We obtain $w_L = 893.31$ and $w_H = 1146.74$. For the corporate tax rate τ, we use the mean effective tax rate of the studied DEs. The counterfactual density distribution of average monthly wage per worker (that is, the distribution when DEs can fully deduce wage bills from their taxable corporate income) is estimated using the empirical framework by Chetty et al. (2011) as elaborated in Section 2.1.

With the estimated counterfactual density distribution and values of key parameter values $(w_L, w_H, \bar{w}, \tau, r)$, we obtain $\widehat{\Delta w} = \hat{b} = 20.2582$. Using equation (22), we obtain our estimated elasticity $\hat{\sigma} = 3.6294$. This implies that with the relative wage of skilled to unskilled labor increasing by 1%, firms decrease the ratio of skilled to unskilled labor by 3.6294%. Results are summarized in the first raw in panel A of Table 4.

[Insert Table 4 Here]

Robustness. A central element in the estimation of σ is the estimation of the counterfactual density distribution, in which we use the empirical framework developed by Chetty et al. (2011). The identification relies on two assumptions: (1) the proper formulation of the polynomial function; and (2) the non-manipulation region is a good counterfactual for the manipulation region. To check the robustness of the estimation of σ, we use an alternative approach to construct the counterfactual density. Specifically, in our research setting, we have groups of firms that were not subject to the wage deduction limit of 960, and their density distribution can then be used to construct the counterfactual density for firms subject to the policy. First, FIEs can fully deduct wage bills, and as shown in Figure 3A, there is no significant jump at the cutoff 960. To improve the comparability, we use FIEs located in the regions with the wage deduction limit of 960 to construct the counterfactual density for DEs located in these regions. Specifically, we choose an arbitrary adjustment degree d, construct a density distribution by multiplying the density of FIEs by d, and calculate the squared sum of differences between the constructed density and the observed density of DEs in all wage bins outside the exclusion region. The counterfactual density is chosen as the one minimizing the squared sum of differences. With the new counterfactual density, we obtain new estimate $\hat{\sigma}$. Results are presented in the second row in panel A of Table 4. We have $\hat{\sigma} = 4.7864$.

Second, DEs located in the regions with wage deduction limit of 800 were not subject to the cutoff 960, and can then be used to construct a counterfactual density distribution. We follow the similar procedure as before, and select the counterfactual density as the one minimizing the difference between the constructed density and the observed density in all wage bins outside the exclusion region. Results are reported in the third raw in panel A of Table 4. We have $\hat{\sigma} = 2.3174$.

One may be concerned that FIEs are different from DEs and different regions are different, and hence, the density estimated in the above two samples of firms may not construct a good counterfactual. To improve the comparability, we use a method in spirit to the DD analysis. That is, we first construct the difference of the density distribution around 960 between DEs located in regions with the wage deduction limit of 960 and DEs in regions with the wage deduction limit of 800, and then compare it with the corresponding difference of FIEs between these regions. This double difference can help us control for the differential distributions between FIEs and DEs, and also between different regions. The results are shown in the fourth row in panel A of Table 4. We obtain $\hat{\sigma} = 4.0458$.

Meanwhile, in the baseline, we estimate w_L and w_H using the model moments. To check the sensitivity of our results to the wage rates, we calculate wage rates from the 2005 Chinese Min Population Census. Specifically, we use the mean wage rates of employees with the highest education lower than senior secondary school and higher than junior secondary school for w_L and w_H, respectively. We have $w_L = 908.66$ and $w_H = 1270.22$. Results using the wage rates calculated from the 2005 Mini Population Census are reported in panel B of Table 4, replicating the analyses in panel A. $\hat{\sigma}$ ranges from 2.0082 to 3.9946.

Comparison with the existing literature. The literature on the estimation of the elasticity of substitution between the skilled and unskilled labor can be traced back to Katz and Murphy (1992). Specifically, they estimate the substitution elasticity between college and high school labor as 1.41, using the March Current Population Surveys (CPS) data from 1967-1987 of U.S. Katz and Autor (1999) include demand shifts as controls, and find that the elasticity ranges from 1 to 3. Krusell et al. (2000) use an alternative definition of skilled-labor as workers with at least some college education, and obtain a moderately higher elasticity estimate of 1.89. Card and Lemieux (2001) include an aggregate relative supply index and age-group specific relative supplies of college workers into the model of Katz and Murphy (1992), and find that the elasticity of substitution between college and high school labor is around 2.2 to 2.5 in both the U.S. and the U.K. Autor et al. (2008) use the 1968-2005 CPS data to re-estimate the elasticity with the model of Katz and Murphy (1992), and show that the elasticity of substitution between college and high school equivalents is 2.43 when no trend break in the annual growth rate in 1992 is added.

The estimated elasticity in this paper is larger than those from the U.S. One potential explanation is that different from most previous studies, our work focuses on the substitution between the high school and non-high school graduates. One study close to us is by Angrist (1995), who uses the Palestinian data and finds the elasticity of substitution between workers with sixteen years of schooling and those with less than twelve years of schooling being close to 2. Another difference between our work and the literature is that we investigate the setting in a big developing country—China, where the technology is developing fast in recent years and favors skilled over unskilled labor in terms of increasing its relative productivity.

5.2 Tax and GDP

Tax. We start with the tax consequences of the wage deduction limit policy. Specifically, we first calculate tax revenues when the policy is in place (referred to as actual tax revenues). Specifically, for DEs who are subject to the wage deduction limit, their profit functions are presented by equation (10), whereas FIEs' profit functions are captured by equation (6). We can solve, for each firm, its optimal K, H, and L as functions of λ. Tax revenue for regions with the wage deduction limit of 960 can, then, be calculated as

$$Tax\ Revenue = \int_0^{\lambda_1} \tau_j I_j d\lambda_j + \int_{\lambda_1}^{1} \tau_j I'_j d\lambda_j. \tag{25}$$

where I_j and I'_j denote the taxable income and are calculated as

$$I_j = \Bigg[\left(K_j(\lambda_j)^{\alpha_s}\left[(\lambda_j H_j(\lambda_j)^{\frac{\sigma-1}{\sigma}} + (1-\lambda_j)L_j(\lambda_j)^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}\right]^{1-\alpha_s}\right)^{\beta}$$
$$-rK_j - w_H H_j - w_L L_j\Bigg]$$

and

$$I'_j = \Bigg[\left(K_j(\lambda_j)^{\alpha_s}\left[(\lambda_j H_j(\lambda_j)^{\frac{\sigma-1}{\sigma}} + (1-\lambda_j)L_j(\lambda_j)^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}\right]^{1-\alpha_s}\right)^{\beta}$$
$$-rK_j - \bar{w}H_j - \bar{w}L_j\Bigg]$$

Next, we calculate tax revenues under the counterfactual linear tax schedule in which all firms can fully deduct wage bills (referred to as counterfactual tax revenues). In this scenario, all firms' profit functions are presented by equation (6). Hence, we can solve a new set of K, H, and L as functions of λ_j (i.e., $K'(\lambda_j)$, $H'(\lambda_j)$, and $L'(\lambda_j)$), respectively. Consequently, we calculate the tax revenues for regions with the wage deduction limit of 960 as

$$Tax\ Revenue' = \int_{\lambda_j} \tau_j I_j d\lambda_j. \tag{26}$$

We set $r = 5.58\%$; $\tau_j = 0.18,\ 0.27,\ or\ 0.33$ for DEs conditional on the firms' taxable income; $\tau_j = 0.15,\ 0.24,\ or\ 0.33$ for FIEs conditional on the firms' registration location and the main business; and $w_L = 893.31$ and $w_H = 1146.74$. The distribution of λ_j can be inversely calculated from the distribution of the average monthly wage per worker given equation (8) and baseline elasticity σ. β is set as 0.75, a central value in the range of estimates used in the previous literature [for a review, see Head and Mayer (2014)].

Finally, we compare $Tax\ Revenue$ with $Tax\ Revenue'$ to obtain the tax consequence of the wage deduction limit policy. Specifically, the percents of differences are calculated as the ratios of the differences between the actual and counterfactual variables over the counterfactual ones, respectively. Table 5 presents the results. As shown in column 8, imposing a wage deduction limit caused the tax revenue to decline by 2.95%, which accounts for around 1.39 billion RMB.[5] This presents a sharp contrast to the policy target, as the policy was designed to increase tax revenues by reducing the deductibles from the corporate income. However, as illustrated in the model, firms respond to the policy by changing their employment structures, which in turn reduces their taxable

[5] According to China Statistical Yearbook (2005) and provincial statistical yearbooks, the total tax revenue for regions with 960 RMB limit is about 47 billion RMB.

income. The fall in taxable corporate income cancels out the reduction in deductibles, causing a fall in total tax revenues.

[Insert Table 5 Here]

GDP. We then investigate the aggregate GDP response of the wage deduction limit policy. Similarly, we first calculate the aggregate GDP when the policy is in place. Specifically, the aggregate GDP for regions with 960 RMB deduction limit is calculated as

$$GDP = \left[\int_{\lambda_j} \left(K_j(\lambda_j)^{\alpha_s} \left[(\lambda_j H_j(\lambda_j)^{\frac{\sigma-1}{\sigma}} + (1-\lambda_j) L_j(\lambda_j)^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \right]^{1-\alpha_s} \right)^{\beta} d\lambda_j \right]^{\frac{1}{\beta}}. \tag{27}$$

α_s is computed by dividing the aggregate capital cost to the aggregate total input costs in each industry sector s, i.e.,

$$\alpha_s = \frac{(1-\tau)rK_s}{(1-\tau)rK_s + p_H H_s + p_L L_s}.$$

where $p_H = (1-\tau)w_H$ and $p_L = (1-\tau)w_L$ for unaffected firms, and $p_H = w_H - \tau\bar{w}$ and $p_L = w_L - \tau\bar{w}$ for affected firms.

Next, we calculate the aggregate GDP under the counterfactual linear tax schedule in which all firms can fully deduct wage bills. Specifically, the counterfactual GDP for regions with the wage deduction limit of 960 is

$$GDP' = \left[\int_{\lambda_j} \left(K'_j(\lambda_j)^{\alpha_s} \left[(\lambda_j H'_j(\lambda_j)^{\frac{\sigma-1}{\sigma}} + (1-\lambda_j) L'_j(\lambda_j)^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \right]^{1-\alpha_s} \right)^{\beta} d\lambda_j \right]^{\frac{1}{\beta}}, \tag{28}$$

where α_s is computed as

$$\alpha_s = \frac{rK_s}{rK + w_H H_s + w_L L_s}.$$

Comparing GDP with GDP', we find that the wage deduction limit policy caused the aggregate GDP to decline by about 1.72%, as shown in column 9 of Table 5. This accounts for 125.16 billion RMB for regions with 960 RMB limit with the reference to the published provincial and city-level GDPs in 2004.[6]

Combined, our analyses suggest that not only the policy did not achieve its intended target (i.e., increasing tax revenues), but also generated significant welfare loss.

5.3 An Extension with Booking Manipulation

In response to the wage deduction limit, firms may manipulate their reported employment level without effectively distorting firm operation. In Section 3.3, we present some reduced-form evidence that firms did have real responses to the policy, e.g., the increase of capital. However, this may not fully rule out the role of booking manipulation in generating the bunching response and hence, possibly bias our estimates in the previous analysis. To this end, we, in this section, extend our

[6] According to China Statistical Yearbook (2005) and provincial statistical yearbooks, the aggregate GDP for regions with 960 RMB limit is about 7,277 billion RMB.

aforementioned theoretical model to incorporate the possibility that firms may inflate the reporting of employment.

Specifically, we change N_j^r in equation (9) to $N_j^r = H_j + L_j + L_j^m$. Without loss of generality and for the simplification of calculation, we define $L_j^m = L_j \times \eta_j$ as firm j's manipulated number of unskilled employment, where η_j being the manipulation degree variable. And $w_j^r = \frac{w_H H_j + w_L L_j}{N_j^r}$.

With this change, the profit function becomes

$$\begin{aligned}\pi_j = p_j q_j - (w_H H_j + w_L L_j + rK_j) - \tau(p_j q_j - rK_j - DC_j) - c(L_j, \eta_j) \\ - C \times I_{\left[\frac{(H_j)^*_{ct}}{(L_j)^*_{ct}} > D\right]} \times (L_j + L_j^m + H_j - (L_j)^*_{ct} - (L_j^m)^*_{ct} - (H_j)^*_{ct}),\end{aligned} \tag{29}$$

where $c(L_j, \eta_j) = \frac{c}{2}\eta_j^2 L_j$ denotes the cost of booking manipulation; and $(L_j^m)^*_{ct}$ denotes the firm j's optimal choice of the manipulated number of unskilled employment under the counterfactual linear tax schedule.

The optimal choices under the implemented nonlinear tax schedule can then be derived as

$$\left((w_j^r)^*, \eta_j^*\right) = \begin{cases} \left((w_j^r)^*_{ct}, 0\right) & \text{if } \lambda_j \le \lambda_1' \\ \left(\bar{w}, [0, \frac{\tau\bar{w}}{c}]\right) & \text{if } \lambda_j \in (\lambda_1', \lambda_2'] \\ \left(\dfrac{w_H\left[\frac{w_L - \tau\bar{w} - \frac{\tau^2\bar{w}^2}{2c}}{w_H - \tau\bar{w}}\frac{\lambda_j}{1-\lambda_j}\right]^\sigma + w_L}{\left[\frac{w_L - \tau\bar{w} - \frac{\tau^2\bar{w}^2}{2c}}{w_H - \tau\bar{w}}\frac{\lambda_j}{1-\lambda_j}\right]^\sigma + 1}, \frac{\tau\bar{w}}{c}\right) & \text{if } \lambda_j \in (\lambda_2', \lambda_3'] \\ \left((w_j^r)^*_{ct}, 0\right) & \text{if } \lambda_j > \lambda_3'. \end{cases} \tag{30}$$

where $\lambda_1' \equiv 1 - 1/\left(1 + \left[\frac{\bar{w} - w_L}{w_H - \bar{w}}\right]^{\frac{1}{\sigma}} \frac{w_H}{w_L}\right)$; $\lambda_2' \equiv 1 - 1/\left(1 + \left[\frac{\bar{w}(1+\frac{\tau\bar{w}}{c}) - w_L}{w_H - \bar{w}}\right]^{\frac{1}{\sigma}} \frac{w_H - \tau\bar{w}}{w_L - \tau\bar{w} - \frac{\tau^2\bar{w}^2}{2c}}\right)$; and $\lambda_3' \equiv \frac{w_H D^{\frac{1}{\sigma}}}{w_H D^{\frac{1}{\sigma}} + w_L}$. The comparison of the firms' optimal choices of the average per worker monthly wage and the manipulation degree are presented in Figure 5.

[Insert Figure 5 Here]

The bunching firm with the lowest $\lambda_j = \lambda_1'$ chooses the same $(w^r_{\lambda_1'})^* = (w^r_{\lambda_1'})^*_{ct} = \bar{w}$ under both the implemented nonlinear and the counterfactual linear tax schedules. The optimal solution of the bunching firm with the highest $\lambda_j = \lambda_2'$ under the counterfactual linear tax schedule is

$$(w^r_{i_2})^*_{ct} = \frac{w_H\left[\frac{w_L}{w_H}\frac{\lambda_2'}{1-\lambda_2'}\right]^\sigma + w_L}{\left[\frac{w_L}{w_H}\frac{\lambda_2'}{1-\lambda_2'}\right]^\sigma + 1}. \tag{31}$$

Hence, all firms with $\lambda_j \in [\lambda_1', \lambda_2']$ (or the counterfactual reported monthly average wage $(w_j^r)^*_{ct} \in [\bar{w}, \bar{w} + \Delta w'])$ bunch around $\bar{w}$ under the nonlinear tax schedule, where

$$\Delta w' = \varphi(w_L, w_H, \bar{w}, \tau, c, \sigma). \tag{32}$$

And the excess fraction of bunching is

$$B = \int_{\lambda_1'}^{\lambda_2'} f(\lambda_j) d\lambda_j = \int_{\bar{w}}^{\bar{w}+\Delta w'} \hat{c}((w_j^r)_{ct}^*) d(w_j^r)_{ct}^* \simeq \hat{c}(\bar{w})\Delta w', \tag{33}$$

Combining equations (32) and (33), we have two unknown parameters c and σ, which requires two empirical moments to pin down them simultaneously. The bunching around 960 of DEs located in regions with the wage deduction limit of 960 provides one moment. For the other, we resort to DEs located in regions with the wage deduction limit of 800.

Estimation results of σ and η are presented in Table 6. The baseline analysis relies on Chetty et al. (2011) method to estimate the counterfactual density distribution of average monthly wage per worker. Results are reported in the first two rows for DEs in regions with the wage deduction limit of 960 and in regions with the wage deduction limit of 800, respectively. $\hat{\eta} = 33\%$ for the former regions and $\hat{\eta} = 21\%$ for the latter regions, implying affected firms with 960 RMB limit and with 800 RMB limit reported 33% and 21% *ghost workers*, respectively. Incorporating the possibility of booking manipulation, we find that the substitution elasticity $\hat{\sigma} = 3.1898$, approximately 12% smaller than the baseline estimate without booking manipulation in Table 4.

[Insert Table 6 Here]

In the rest of the table, we use the alternative estimation methods of the counterfactual density distribution as discussed in Section 5.1, and mimic the analyses in Table 4. Specifically, as the control groups for the counterfactual, we use the corresponding FIEs in the same regions, the DEs in the other regions at the corresponding kink point, and the difference-in-differences design using both of these two samples, respectively. The elasticity estimates $\hat{\sigma}$ remain stable, ranging from 2.7123 to 3.8771.

Based on these estimates, we re-calculate the welfare from the imposed wage deduction limit. The results are presented in Table 7. With the new elasticity estimate, the percents of differences are 2.74% and 1.51% for tax revenue and GDP, respectively. These results imply that the tax revenue and GDP in the regions with 960 RMB limit declined by about 1.29 and 109.88 billion RMB when the wage deduction limit was implemented, respectively.

[Insert Table 7 Here]

6 Conclusion

This article used a bunching method to explore the employment distortion induced by a wage deduction tax policy in China. We first used China's first wave of economic census in 2004 to provide graphical evidence that the density distribution of average per worker monthly wage for DEs bunches at the imposed wage deduction limit point. This pattern was only observed for the firms subject to the deduction limit and the periods when the deduction limit existed. We next employed the methodology by Diamond and Persson (2017) to find that firms responded to the wage deduction limit by increasing their capital input and substituting skilled labor with unskilled labor, instead of relabeling part of the labor cost as other deductible terms. We then constructed a theoretical model and estimated that the substitution elasticity between skilled and unskilled labor is 3.6294. With the elasticity estimate, we calculated that this wage deduction limit policy reduced total tax revenue by 2.95% and aggregate GDP by 1.72%. By extending our model to incorporate the potential reported employment manipulation behavior of firms, we estimated a smaller substitution elasticity estimate , suggesting that with the relative wage of skilled to unskilled labor increasing by 1%, firms would decrease the ratio of skilled to unskilled labor by 3.1898%. As a result, tax revenue and GDP decreased by 2.74% and 1.51%, respectively.

With the extended model, we found that besides substituting skilled labor with unskilled labor, firms tended to report 21% to 53% more nonexistent unskilled labor on average to claim wage deduction from the taxable income. To some extent, this over-reporting behavior contributed to the tax revenue fall of 1.29 billion RMB. Hence, when designing policies, policy makers should not only carefully analyze firms' responses, but also increase the supervising and managing dynamics,

References

Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature*, 40(1):7–72. 2

Angrist, J. D. (1995). The Economic Returns to Schooling in the West Bank and Gaza Strip. *The American Economic Review*, 85(5):1065–1087. 2, 15

Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *The Review of Economics and Statistics*, 90(2):300–323. 2, 15

Card, D. and Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics*, 116(2):705–746. 2, 15

Chen, Z., Liu, Z., Serrato, J. C. S., and Xu, D. Y. (2019). Notching R&D Investment with Corporate Income Tax Cuts in China. 3

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *The Quarterly Journal of Economics*, 126(2):749–804. 1, 3, 7, 8, 13, 14, 19

China Statistical Yearbook (2005). China Statistics Press. 16, 17

Diamond, R. and Persson, P. (2017). The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests. 1, 6, 8, 20

Elias, F. (2015). Labor Demand Elasticities Over the Life Cycle: Evidence from Spain's Payroll Tax Reforms. pages 1–71. 2

Garicano, L., Lelarge, C., and Van Reenen, J. (2016). Firm Size Distortions and the Productivity Distribution: Evidence from France. *The American Economic Review*, 106(11):3439–3479. 3

Gourio, F. and Roys, N. (2014). Size-Dependent Regulations, Firm Size Distribution, and Reallocation. *Quantitative Economics*, 5(2):377–416. 3

Harasztosi, P. and Lindner, A. (2019). Who Pays for the Minimum Wage? *American Economic Review*, 109(8):2693–2727. 3

Harju, J., Matikka, T., and Rauhanen, T. (2016). The Effects of Size-Based Regulation on Small Firms: Evidence from VAT Threshold. 3

Head, K. and Mayer, T. (2014). Gravity Equations: Workhorse, Toolkit, and Cookbook. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*, volume 4, chapter 3, pages 131–195. North-Holland, Poland. 16, 30, 31

Heckman, J. J., Lochner, L., and Taber, C. (1998). Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents. *Review of Economic Dynamics*, 1(1):1–58. 2

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448. 9

Huttunen, K., Pirttilä, J., and Uusitalo, R. (2013). The Employment Effects of Low-Wage Subsidies. *Journal of Public Economics*, 97(1):49–60. 2

Katz, L. F. (1998). Wage Subsidies for the Disadvantaged. In Freeman, R. B. and Gottschalk, P., editors, *Generating Jobs*, pages 21–53. Russell Sage Foundation, New York. 2

Katz, L. F. and Autor, D. H. (1999). Changes in the Wage Structure and Earnings Inequality. In Ashenfeher, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, chapter 26, pages 1463–1555. Elsevier Science B.V. 2, 15

Katz, L. F. and Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78. 2, 15

Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8(1):435–464. 3

Kleven, H. J. and Waseem, M. (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *The Quarterly Journal of Economics*, 128(2):669–723. 3, 5, 7

Krusell, P., Ohanian, L. E., Ríos-Rull, J. V., and Violante, G. L. (2000). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica*, 68(5):1029–1053. 2, 15

Liu, L., Lockwood, B., and Almunia, M. (2017). VAT Notches, Voluntary Registration , and Bunching: Theory and UK Evidence. 3

Onji, K. (2009). The Response of Firms to Eligibility Thresholds: Evidence from the Japanese Value-Added Tax. *Journal of Public Economics*, 93:766–775. 3

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy.*, 2(3):180–212. 3, 13

Saez, E., Schoefer, B., and Seim, D. (2019). Payroll Taxes, Firm Behavior, and Rent Sharing: Evidence from a Young Workers' Tax Cut in Sweden. *The American Economic Review*, 109(5):1717–1763. 2

Figures and Tables

Figure 1

Wage Deduction Limit Policy in 2004

Notes: This figure shows the distribution of implemented wage deduction limits across all municipalities and provinces in China in 2004.

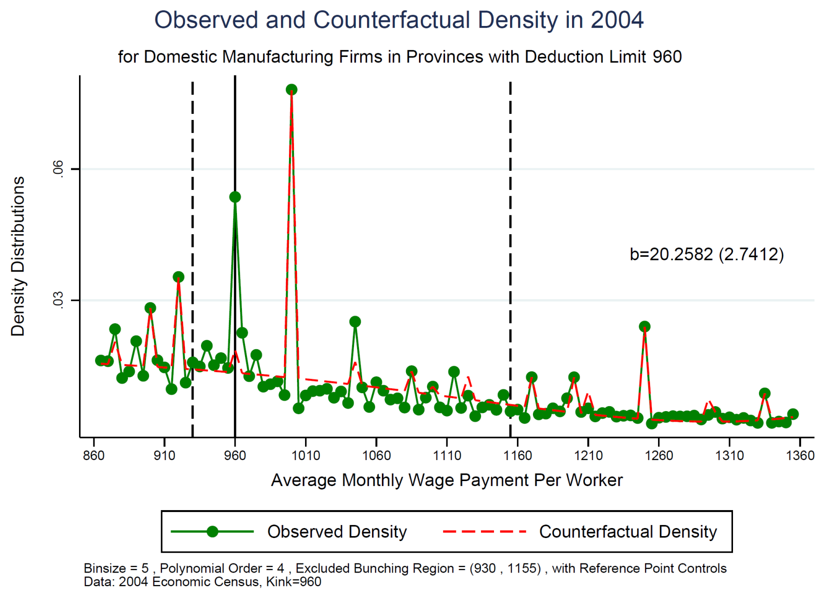
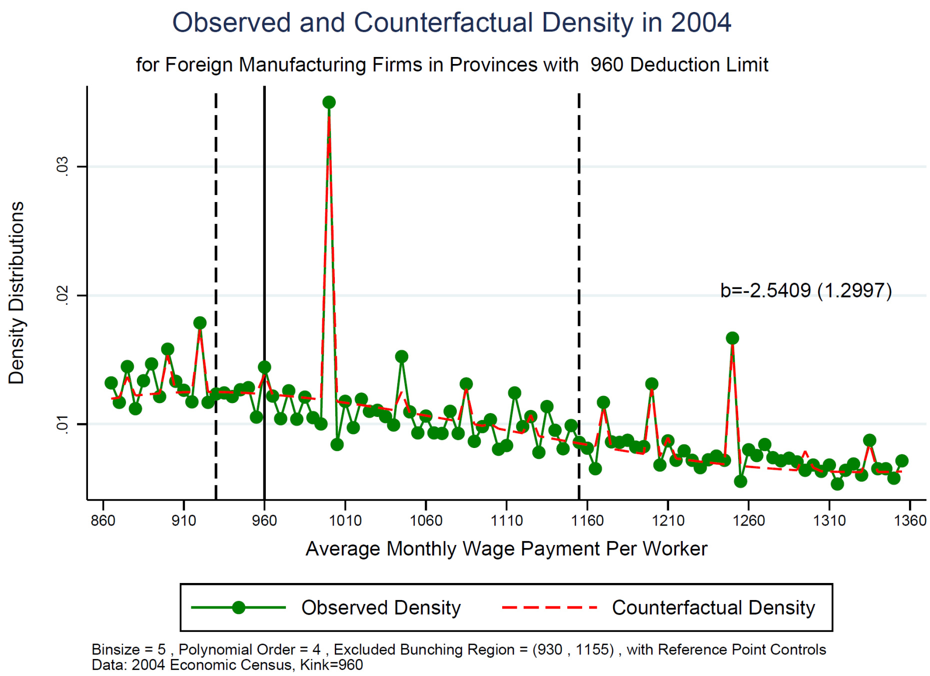


Figure 2

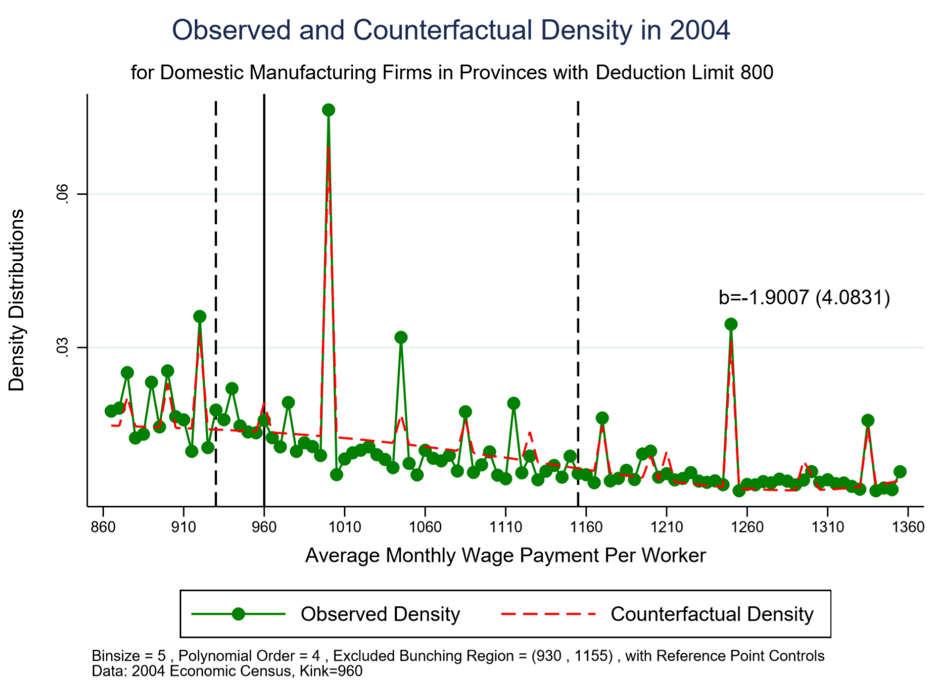
Density Distribution of Monthly Average Per Worker Wage for DEs with Deduction Limit 960

Notes: This figure shows the density distributions of the average per worker monthly wage around the kink point 960 (demarcated by the vertical line) for domestic enterprises (DEs) located in the provinces with 960 RMB wage deduction limit. The solid curve displays the observed density in 5 RMB bins, and the dashed curve displays the counterfactual density by excluding a window of (930, 1155) centered around the kink point, controlling for multiples of 100 RMB for monthly wage, and 500 RMB and 1000 RMB for annual wage, and fitting a polynomial of fourth order to the observed distributions.

Panel A.



Panel B.



Panel C.

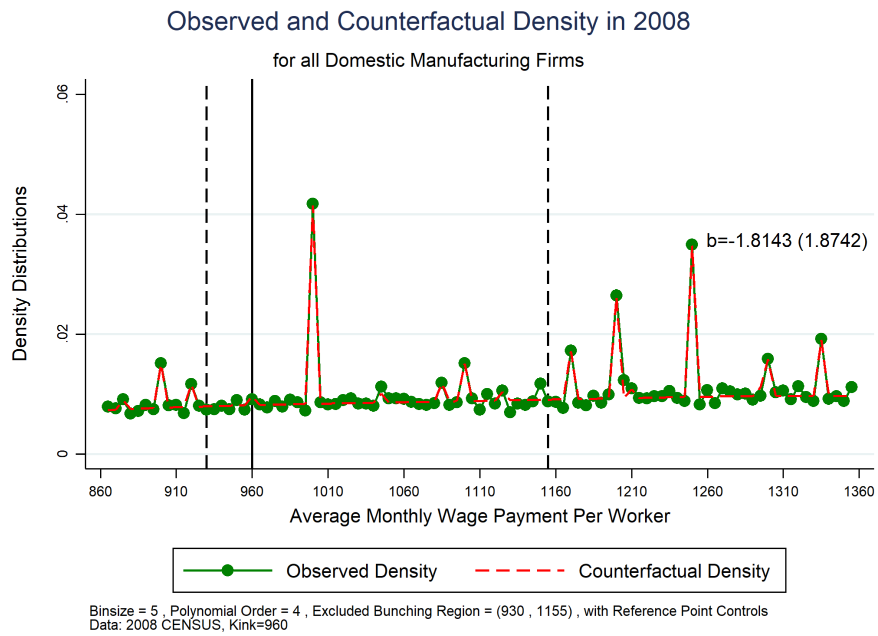


Figure 3

Density Distribution of Monthly Average Per Worker Wage for Control Groups

Notes: These figures show the density distributions of the average per worker monthly wage around the kink point 960 (demarcated by the vertical line) for firms in control groups. Panel A depicts the densities for foreign-invested enterprises (FIE) located in the provinces with 960 RMB wage deduction limit; Panel B shows those for domestic enterprises (DEs) subject to 800 RMB wage deduction limit; Panel C plots those for all DEs in 2008 using the second Economic Census data. The solid curves display the observed densities in 5 RMB bins, and the dashed curves display the counterfactual densities by excluding a window of (930, 1155) around the kink point, controlling for multiples of 100 RMB for monthly wage, and 500 RMB and 1000 RMB for annual wage, and fitting a polynomial of fourth order to the observed distributions in panel A, B, and C, respectively.

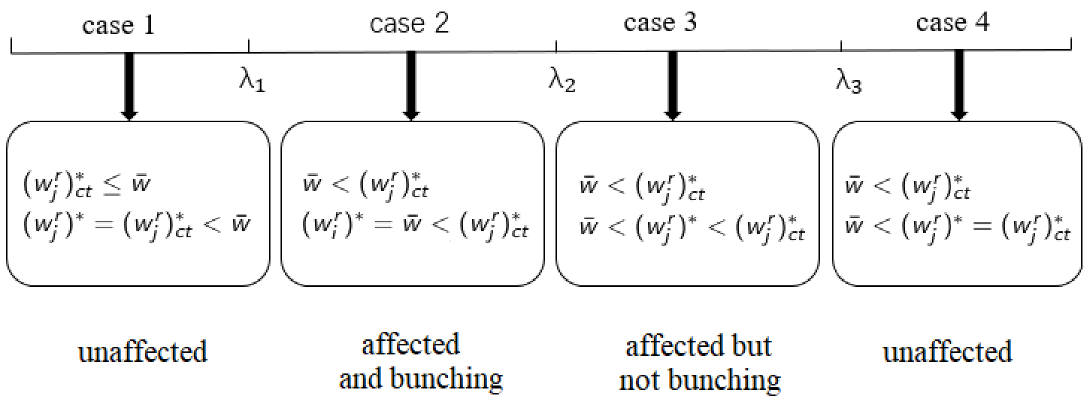


Figure 4

Firms' Responses under the Linear and Non-linear Tax Schedules

Notes: This figure compares the firms' optimal choices of the average per worker monthly wage under the linear and non-linear tax schedules.

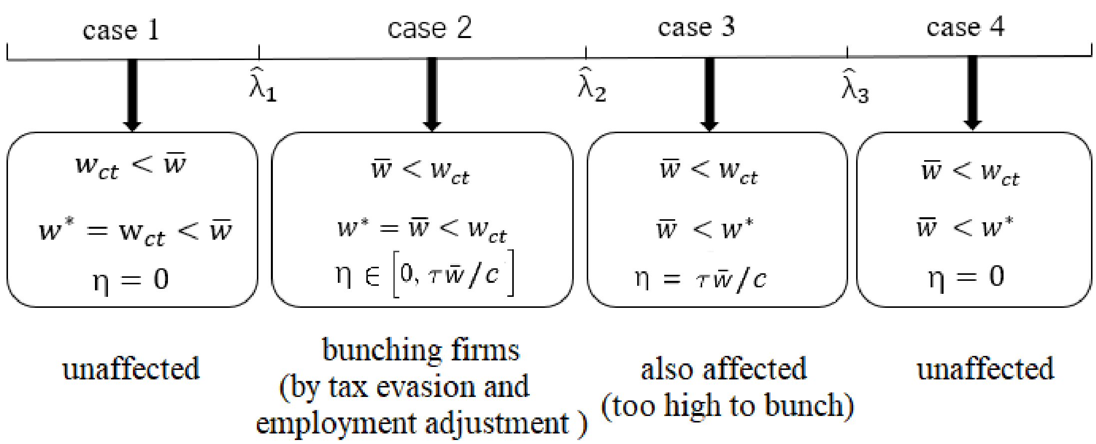


Figure 5

Firms' Responses under the Linear and Non-linear Tax Schedules with the Employment Manipulation

Notes: This figure compares the firms' optimal choices of the average per worker monthly wage and manipulation degree under the linear and non-linear tax schedules.

Table 1. Deduction Limit for the Average Per Worker Monthly Wage in Corporate Income Tax

Period	Domestic Enterprises		Foreign-Invested Enterprises
	National Deduction Limit	Allow for 20% inflation	Deduction limit
1994.5∼1995.12	500	Yes	Fully
1996.1∼1999.12	550	Yes	Fully
2000.1∼2006.6	800	Yes	Fully
2006.7∼2007.12	1600	No	Fully
2008.1∼Now	Fully	No	Fully

Note: This table summarizes deduction limits for the average per worker monthly wage in Corporate Income Tax (CIT) for all firms in China since May 1994.

Table 2. Summary Statistics for Domestic Enterprises in the Main Sample

Variable	(1) N	(2) mean	(3) P25	(4) P50	(5) P75
Average Wage Per Worker (Monthly, RMB)	629,759	837.5624	518.0180	727.2728	947.9167
Employment	665,299	44.4820	8	16	38
Skilled versus Unskilled Labor	587,054	0.9756	0.1111	0.3333	0.8750
Capital (ln)	626,182	7.2088	6.2538	7.0901	8.0790
Unemployment Insurance per Worker (RMB)	629,759	315.4075	0	0	62.7962
Employee Benefits per Worker (RMB)	660,333	301.0581	0	0	0
Administrative Cost per Worker (RMB)	660,333	2588.5920	0	0	0

Note: This table displays the summary statistics for the DEs in the main sample used in this paper, i.e., domestic manufacturing firms in provinces with 960 RMB limit. The average monthly wage per worker is calculated by dividing the total wage bill by 12 months and by the total employment. Workers are classified into two categories based on the education level—the highest education levels are high school or above and junior secondary school or below. The skilled versus unskilled labor ratio is calculated as the ratio of the total employments of these two groups. The total capital is used as the capital measurement. The variables unemployment insurance per worker, employee benefits per worker, and administrative cost per worker are calculated as the ratios of the firm's total corresponding spending to the total employment.

Table 3. Reduced-Form Estimates of Manipulation

Unemployment Insurance per Worker (1)	Employee Benefits per Worker (2)	Administrative Cost per Worker (3)	Capital (ln) (4)	L/H (5)	TFP (6)
0.1700	0.0071	0.7629	0.3556***	0.3610**	-0.0741***
(0.8617)	(0.0346)	(0.7231)	(0.0507)	(0.1726)	(0.0049)

Note: This table shows the reduced-form estimates of firms' manipulation responses, with standard errors in parenthesis. Due to data limitation, unemployment insurance is estimated over small domestic firms with annual sales revenue not above 5 million, employee benefits and administrative cost are estimated over large domestic firms with annual sales revenue above 5 million. The outcome variable Capital is in natural logarithm term. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4. Elasticity of Substitution between Skilled and Unskilled Labor

Samples	(1) Δw	(2) w_H	(3) w_L	(4) τ	(5) σ
Panel A: Δw based on 2004 Economic Census					
w_H and w_L from 2004 Economic Census					
DE_960_2004	20.2582	1146.74	893.31	0.31	3.6294
DE_960_2004 - FIE_960_2004	27.3043	1146.74	893.31	0.31	4.7864
DE_960_2004 - DE_800_2004	12.5880	1146.74	893.31	0.31	2.3174
(DE_960_2004 - DE_800_2004)	22.7656	1146.74	893.31	0.31	4.0458
(FIE_960_2004 - FIE_800_2004)					
Panel B: Δw based on 2004 Economic Census,					
w_H and w_L from 2005 Chinese Population Census					
DE_960_2004	20.2582	1270.22	908.66	0.31	3.0821
DE_960_2004 - FIE_960_2004	27.3043	1270.22	908.66	0.31	3.9946
DE_960_2004 - DE_800_2004	12.5880	1270.22	908.66	0.31	2.0082
(DE_960_2004 - DE_800_2004)	22.7656	1270.22	908.66	0.31	3.4141
(FIE_960_2004 - FIE_800_2004)					

Note: This table shows the estimates for the elasticity of substitution between skilled and unskilled labor for Domestic Enterprises (DE) with 960 RMB limit. Columns (1) presents the normalized bunching size; column (2) and (3) present the monthly wages for skilled and unskilled labor, respectively, and the values are calculated from 2004 economic census with the equations (23) and (24) as conditions; column (4) presents the tax rate, calculated as the mean effective tax rate of the corresponding group of firms; column (5) presents the elasticity estimate. Panel A (B) displays the results using wage rates calculated from 2004 economic census(2005 population census) with the equations (23) and (24) as conditions. The second and third rows of each panel present the results using the density distributions of Foreign-Invested Enterprises (FIE) located in the provinces with 960 RMB limit and of Domestic Enterprises (DE) in the provinces with 800 RMB limit to construct counterfactual densities for DE with 960 RMB limit. The fourth row of each panel presents the result using the difference between the density distributions of FIE located in provinces with 960 RMB and 800 RMB limits to construct the counterfactual for the difference between the densities of DE with 960 RMB and 800 RMB limits.

Table 5. Welfare Analysis with the Baseline Model

β (1)	r (2)	w_h (3)	w_l (4)	$\tau_{domestic}$ (5)	$\tau_{foreign}$ (6)	σ (7)	Percent of Difference in Tax Revenue (8)	Percent of Difference in GDP (9)
0.75	0.0558	1146.74	893.31	0.31	0.19	3.6294	-2.95%	-1.72%

Note: This table shows the welfare analysis result with the baseline model. Columns (1) presents the substitution parameter of CES utility function in equation (4), and the value is set based on the central value in the range of estimates used in the previous literature [for a review, see Head and Mayer (2014)]; column (2) presents the interest rate, and the value is set based on the data from the World Bank; column (3) and (4) present the monthly wages for skilled and unskilled labor, respectively, and the values are calculated from 2004 economic census with the equations (23) and (24) as conditions; column (5) and (6) present the tax rate, calculated as the mean effective tax rate of the corresponding group of firms; column (7) presents the baseline elasticity estimate; column (8) presents the percent of difference in tax revenue, calculated as the ratio of the difference between the real tax revenue under the implemented non-linear tax schedule and the tax revenue under the counterfactual linear tax schedule over the counterfactual tax revenue; column (9) presents the percent of difference in GDP, calculated as the ratio of the difference between the GDP under the implemented non-linear tax schedule and the GDP' under the counterfactual linear tax schedule over GDP'.

Table 6. Elasticity of Substitution and Degree of Booking Manipulation

Samples	Δw (1)	w_H (2)	w_L (3)	τ (4)	σ (5)	η (6)
DE_960_2004	20.2582	1146.74	893.31	0.31	3.1898	33%
DE_960_2004	14.5848	910.20	744.90	0.28	3.1898	21%
DE_960_2004 - FIE_960_2004	27.3043	1146.74	893.31	0.31	3.8771	36%
DE_800_2004 - FIE_800_2004	11.4180	910.20	744.90	0.28	3.8771	25%
DE_960_2004 - DE_800_2004	12.5880	1146.74	893.31	0.31	2.7123	53%
DE_800_2004 - DE_960_2004	11.5062	910.20	744.90	0.28	2.7123	39%
(DE_960_2004 - DE_800_2004) (FIE_960_2004 - FIE_800_2004)	22.7656	1146.74	893.31	0.31	3.3101	30%
(DE_800_2004 - DE_960_2004) - (FIE_800_2004 - FIE_960_2004)	14.9741	910.20	744.90	0.28	3.3101	22%

Note: This table shows the estimates for the elasticity of substitution between skilled and unskilled labor and the degree of booking manipulation. Columns (1) presents the normalized bunching size; column (2) and (3) present the monthly wages for skilled and unskilled labor, respectively, and the values are calculated from 2004 economic census with the equations (23) and (24) as conditions; column (4) presents the tax rate, calculated as the mean effective tax rate of the corresponding group of firms; column (5) presents the elasticity estimate; column (6) presents the booking manipulation estimate.

Table 7. Welfare Analysis with the Extended Model

β (1)	r (2)	w_h (3)	w_l (4)	$\tau_{domestic}$ (5)	$\tau_{foreign}$ (6)	σ (7)	Percent of Difference in Tax Revenue (8)	Percent of Difference in GDP (9)
0.75	0.0558	1146.74	893.31	0.31	0.19	3.1898	-2.74%	-1.51%

Note: This table shows the welfare analysis result with the extended model. Columns (1) presents the substitution parameter of CES utility function in equation (4), and the value is set based on the central value in the range of estimates used in the previous literature [for a review, see Head and Mayer (2014)]; column (2) presents the interest rate, and the value is set based on the data from the World Bank; column (3) and (4) present the monthly wages for skilled and unskilled labor, respectively, and the values are calculated from 2004 economic census with the equations (23) and (24) as conditions; column (5) and (6) present the tax rate, calculated as the mean effective tax rate of the corresponding group of firms; column (7) presents the elasticity estimate from the extended model considering booking manipulation; column (8) presents the percent of difference in tax revenue, calculated as the ratio of the difference between the real tax revenue under the implemented non-linear tax schedule and the tax revenue under the counterfactual linear tax schedule over the counterfactual tax revenue; column (9) presents the percent of difference in GDP, calculated as the ratio of the difference between the GDP under the implemented non-linear tax schedule and the GDP' under the counterfactual linear tax schedule over GDP'.