Bunching Estimation and Its Theoretical and Empirical Progress[*][*]

SUZUKI Takafumi

Associate Professor, Faculty of Business, Aichi Shukutoku University

Abstract

This paper describes the basic analytical framework of the bunching estimation method, which has been increasingly applied in recent years, especially in the fields of public finance and public economics. Bunching estimation was first established by Saez (2010) as a method for estimating the elasticity of taxable income, but in recent years its applications have spread to other fields. With the expansion of applications, many papers have been published on the extension of analysis and problems in estimation of the method.

In this paper, I (1) describe the basic analytical framework, especially in the context of taxable income elasticity, and (2) introduce the recent development of the method and the problems and issues in estimation that have been pointed out in subsequent studies.

Keywords: Bunching Estimation, Elasticity of Taxable Income
JEL Classification: H00, C01, H24

I. Introduction

The estimation of the elasticity of taxable income with respect to the marginal net tax rate is an important topic in the field of public finance. The elasticity of taxable income is a sufficient statistic for welfare analysis (Feldstein 1999). It is an important parameter in discussions of the effects of tax reform on economic agents and the optimal design of the tax system. For this reason, numerous studies on elasticity estimation have been conducted to date. Currently, there are two main approaches to estimate the elasticity of taxable income: The first approach uses changes in the budget line between different points in time due to changes in the tax system (the tax schedule change approach). This approach uses panel data to estimate the size of elasticity by analyzing the impact of changes in marginal tax rates by institutional changes in the tax schedule on taxable income for each individual[1].

On the other hand, in recent years, many estimates of the elasticity of taxable income have been made using kinks and notches in the budget line. The second approach is to esti-

* This article is based on a study first published in the Financial Review No. 151, pp. 31-52, Takafumi Suzuki, 2023, "The Bunching Estimation Approach and Its Development", written in Japanese. I gratefully acknowledge the financial support from the Japan Society for the Promotion of Science (Grant number 22K13424).

* I would like to express my gratitude to Masayoshi Hayashi (The University of Tokyo) and Naomi Kodama (Meiji Gakuin University) for their invaluable insights and discussions that have greatly enriched this work.

[1] For example, refer to Gruber and Saez (2002), Kleven and Schultz (2014), Jakobsen and Søgaard (2022), or Miyazaki and Ishida (2022). Saez et al. (2012) surveys the literature.

mate the elasticity of taxable income based on the relationship between economic incentives and bunching of economic agents. This is the bunching estimation method I discuss in this paper, and early studies include Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013). These early studies mainly estimated behavioral responses to taxation, but applications are now expanding to other fields such as health, finance, and the environment. In fact, the number of papers using the bunching estimation method has been increasing in recent years; according to Currie et al. (2020), the number of papers including the word "bunching" has doubled in the top five journals[2] and quadrupled in the NBER Working Paper since 2010 when Saez (2010) was published (see Figure 4 on page 45). This trend indicates that it has been established as one of the major research methods.

There are relatively few studies using bunching estimation in Japan, such as Ito and Sallee (2018), Ichikawa et al. (2021), and Kawakubo et al. (2022), despite the fact that many studies using bunching estimation have been accumulated in public finance and other fields[3]. Additionally, it appears that there is a shortage of papers that compiles the various extensions of the analysis and estimation-related problems that have been put forth after Saez (2010) and other pioneering studies. This paper (1) explains the fundamental analytical techniques, particularly in the context of the elasticity of taxable income, and (2) introduces the recent development of the techniques as well as the problems and issues in estimation that have been highlighted in subsequent studies.

This paper is organized as follows. Section II describes the basic theoretical framework and estimation methods of bunching estimation, and Section III introduces the progress of existing research and describes problems and extensions in estimation as well as related discussions. Finally, I conclude in Section IV.

II. Bunching Estimation

In this section, I describe the fundamental framework of bunching estimation, both theoretically and empirically, in order to estimate the income tax elasticity of individuals' taxable income. See Kleven (2016) for a more thorough explanation[4].

II-1. Theoretical Model: The Case of Kink

First, I will discuss the case of a kink in the budget line. Consider a situation in which an individual provides labor and consumes with earned income. Let c denote the individual's

[2] "Top five journals" refer to American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, and Review of Economic Studies.

[3] Ito and Sallee (2018) focus on the vehicle weight threshold with respect to Japan's fuel efficiency regulations, while Ichikawa et al. (2021) and Kawakubo et al. (2022) focus on the VAT exemption threshold in Japan. In addition, although they do not perform bunching estimation, Onji (2009), for example, report bunching in thresholds for a simplified tax system in Japan's VAT, and Yokoyama (2018) and Yokoyama and Kodama (2016) report bunching at thresholds for the spousal deduction in the individual income tax.

[4] This section owes much to Kleven (2016). The study describes the basic analytical framework for bunching estimation and provides a comprehensive review of the relevant literature.

after-tax income and y denote the individual's before-tax income, and the relationship between the two can be expressed as $c = y - T(y)$, where $T(y)$ is the amount of tax on the before-tax income. Let $U(c, y; a)$ denote the utility of an individual. a is a parameter that represents heterogeneity in the earning ability of individuals and is not observable to the analyst. It is assumed that the parameter a follows a density function $f(a)$. Also assume that taxable income is distributed according to a smooth density function $h_0(y)$ and that each person is subject to a proportional tax with a single marginal tax rate (i.e., the taxable amount is $T(y) = \tau \cdot y)$[5].

Suppose that a kink is introduced in the tax schedule, where the marginal tax rate increases from τ to $\tau + \Delta\tau$ for incomes above a certain before-tax income level y^* as a threshold (*where* $\Delta\tau > 0$). That is, the tax amount is written as follows:

$$T(y) = \tau \cdot y + \Delta\tau \cdot (y - y^*) \cdot 1\{y > y^*\},$$

where $1\{\cdot\}$ is an indicator function that takes value of one if the conditions of its argument are met. Below a certain threshold of before-tax income y^*, individuals face a marginal tax rate τ, but above y^* they face a marginal tax rate $(\tau + \Delta\tau)$.

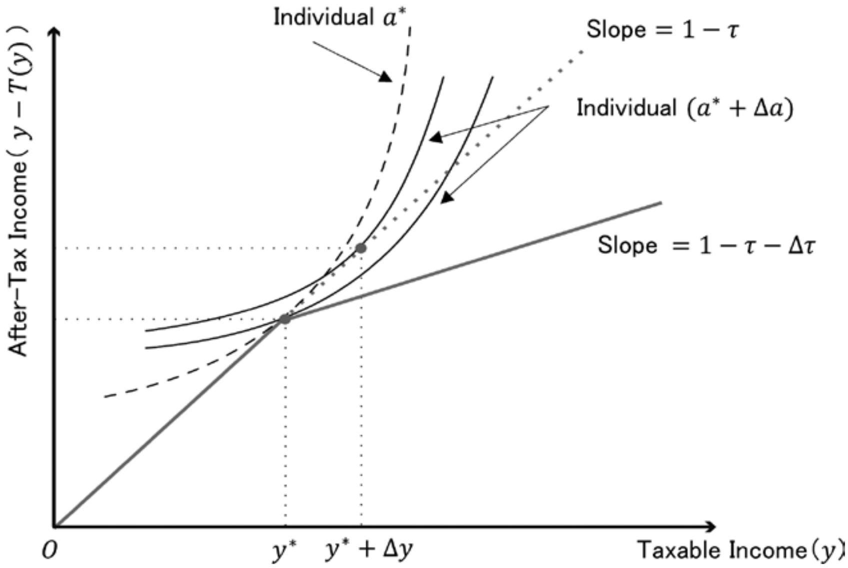
Panel (a) of Figure 1 depicts two budget lines, one for the case where the tax rate is increased by $\Delta\tau$ for taxable income above y^* (the solid line) and one for the case where there is no kink (the dotted line). Panel (b) demonstrates two distributions of individual income under the two budget lines, a real distribution and a counterfactual distribution. Individuals with abilities a^* and $a^* + \Delta a$ will select incomes y^* and $y^* + \Delta y$, respectively, in the absence of a kink. Assuming now that a kink is generated at y^*, those who had previously selected a taxable income of y^* or less will not be affected by the tax increase, whereas those who had selected a taxable income above y^* will be affected. Specifically, individuals who had an indifference curve tangent to the budget line at $y^* + \Delta y$ before the tax increase will have an indifference curve tangent to y^* after the tax increase. These individuals are referred to as marginal bunchers. After the tax increase, all individuals up to y^* who were initially located below $y^* + \Delta y$ will have an indifference curve tangent to y^*. Individuals whose taxable income is reduced to the threshold and who had the highest taxable income prior to the tax increase are marginal bunchers.

An increase in the marginal tax rate above a certain income threshold affects the income distribution of individuals, and the distribution of income before taxes changes to $h(y)$ after the introduction of the kink caused by the tax increase. Panel (b) demonstrates the income distributions $h(y)$ and $h_0(y)$ for each tax schedule in Panel (a).

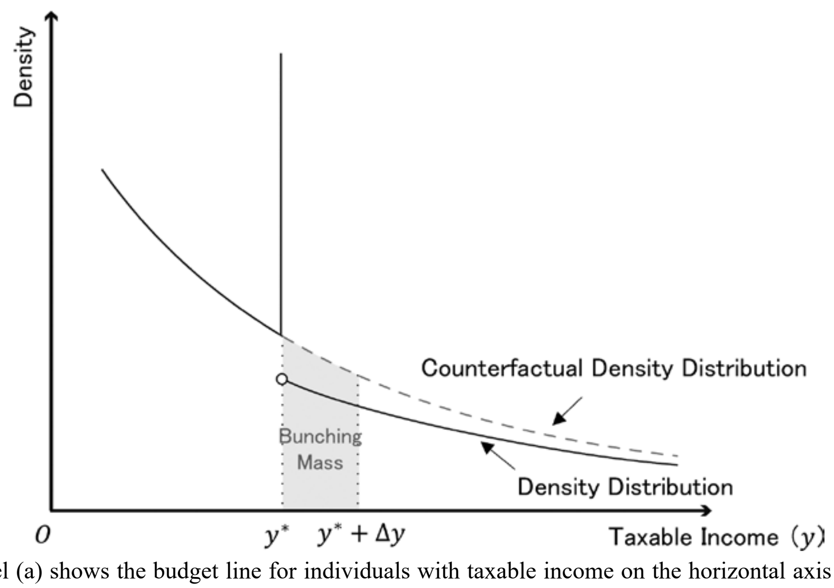
The shape of the distribution is identical for $h_0(y)$ and $h(y)$ in the range $y < y^*$ where the income is less than the threshold, as shown in the figure. For individuals above the marginal bunchers, there is no dip in the distribution above the kink because their taxable income decreases in response to an increase in the marginal tax rate for taxable income above y^*. In other words, the density distribution above y^* has a global leftward shift. If the individual income distribution remains unchanged before and after the tax increase, the shrinkage of

[5] This paper deals with the case where the tax system follows a linear or piecewise linear tax schedule.

Figure 1. The Case of Kink
(a) Budget Line



(b) Taxable Income Density Distribution



Note 1: Panel (a) shows the budget line for individuals with taxable income on the horizontal axis and after-tax income on the vertical axis. It also shows indifference curves for individuals a^* and ($a^* + \Delta a$). The dotted line with slope $(1 - \tau)$ extending from the origin shows the budget line without a kink, and the solid line with slope $(1 - \tau)$ extending from the origin and a kink with slope $(1 - \tau - \Delta\tau)$ at taxable income y^* shows the budget line when kink exists. Individuals who are marginal bunchers ($a^* + \Delta a$) will choose $y^* + \Delta y$ when there is no kink, but will choose y^* when there is a kink.
Note 2: Panel (b) shows the density distribution of taxable income with and without a kink, corresponding to Panel (a). The real density (solid line) is the density distribution when the kink exists, and the counterfactual density (dotted line) is the density distribution when the kink does not exist. The shaded area represents the amount of individuals within the threshold y^* to $y^* + \Delta y$ bunched at y^*.

the density distribution in (y^*, ∞) before the tax increase will compensate for the size of the bunching in y^*.

Assuming that the elasticity of taxable income with respect to the marginal net tax rate, η, is constant with respect to individuals around the threshold y^*, the elasticity η can be represented as follows:

$$\eta = \frac{\Delta y / y^*}{\Delta\tau / (1-\tau)}. \tag{1}$$

Since we are capturing a very localized response of individuals at the threshold, we assume that the income effect is negligible (i.e., Δy is minute). However, it should be noted that the elasticity captured by equation (1) includes the income effect as the extent of kink becomes larger and the range of taxable income $(y^*, y^* + \Delta y)$ occurring behavioral response expand. It should be noted, however, that the elasticity captured by equation (1) includes the income effect as the extent of the kink increases and the range of taxable income $(y^*, y^* + \Delta y)$ occurring behavioral responses expands. Existing studies have avoided this issue by assuming the utility function has a shape in which the income effect does not exist, as described below.

Since we know the threshold y^* and the tax rates τ and $\Delta\tau$, the only information needed to estimate the elasticity η is Δy. Let B be the excess bunching amount at the threshold y^* caused by the tax increase. Then B is written as

$$B = \int_{y^*}^{y^*+\Delta y} h_0(y)\,dy = h_0(\bar{y})\Delta y \simeq h_0(y^*)\Delta y, \tag{2}$$

where $h_0(y)$ is the density function with respect to taxable income in the absence of tax increases (when the tax rate is τ in all regions). From the mean value theorem, the second equality holds for some $\bar{y} \in (y^*, y^* + \Delta y)$. If Δy is sufficiently small, the density $h_0(\bar{y})$ can be approximated by $h_0(y^*)$, and finally the excess bunching B can be approximately expressed as $h_0(y^*)\Delta y$. The specific procedures to be followed in the empirical analysis are explained in the next section. Equation (2) is used to estimate Δy. The specific estimation procedure in the empirical analysis is explained in the next section.

In Saez (2010) and other previous studies of bunching estimation, the utility function (or objective function) is typically specified as an iso-elastic function with no income effect, such as the following:

$$\begin{aligned} U(c,y;a) &= c - \frac{a}{1+1/\eta}\left(\frac{y}{a}\right)^{1+1/\eta} \\ &= y - T(y) - \frac{a}{1+1/\eta}\left(\frac{y}{a}\right)^{1+1/\eta}, \end{aligned} \tag{3}$$

where $a > 0$ and $\eta > 0$. Under equation (3), the individual chooses taxable income $y = a\cdot(1-\tau)^\eta$ from the first-order condition on the utility maximization problem for the tax amount $T(y) = \tau\cdot y$ before the marginal tax rate is raised. That is, the marginal buncher (with ability $a^* + \Delta a$) satisfies two conditions: (1) the first-order condition $y^* + \Delta y = (a^* + \Delta a)\,(1-\tau)^\eta$ when there is *NO* tax rate increase, and (2) the first-order condition $y^* = (a^* + \Delta a)\,(1-\tau-\Delta\tau)^\eta$ when there is a tax rate increase and a kink in the budget line and the indifference curve touches the kink point. These two conditions yield

$$1+\frac{\Delta y}{y^*}=\left(\frac{1}{1-\frac{\Delta\tau}{1-\tau}}\right)^{\eta}.$$

Taking logarithms on both sides, we get

$$\ln\left(1+\frac{\Delta y}{y^*}\right)=-\eta\cdot\ln\left(1-\frac{\Delta\tau}{1-\tau}\right).$$

Thus, elasticity is written as follows:

$$\eta=-\frac{\ln(1+\Delta y/y^*)}{\ln\left(1-\Delta\tau/(1-\tau)\right)}. \tag{4}$$

Again, since y^*, τ and $\Delta\tau$ are known, the elasticity of taxable income can be estimated by observing Δy from the data.

II-2. Theoretical Model: The Case of Notch

Next, I will describe a case in which a notch exists on the budget line. When a notch exists at a certain level of taxable income, the average tax rate changes discontinuously at that level, but the basic analytical framework is the same as in the case of kink. The setting of the individual utility maximization problem is the same as before, with ability a subject to a density function $f(a)$, and the distribution of taxable income in the absence of a notch is given by $h_0(y)$.

We consider the following tax function in the case of tax notch.

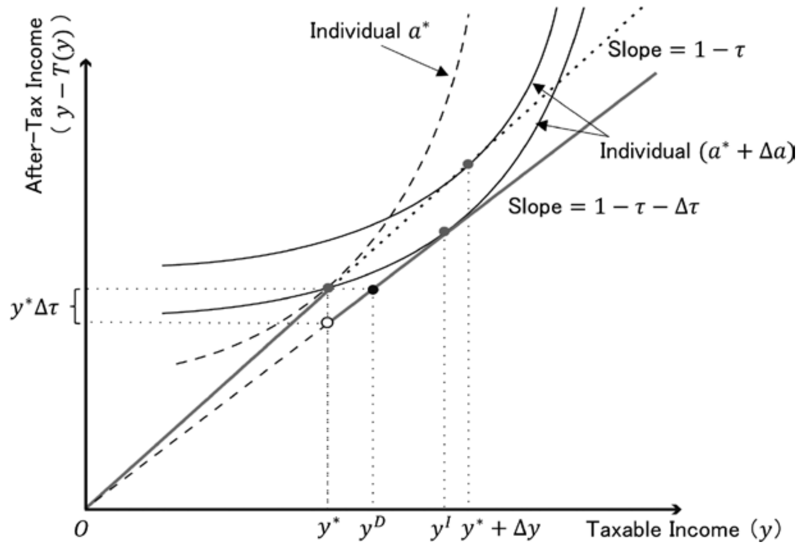
$$T(y)=\tau\cdot y+\Delta\tau\cdot y\cdot 1\{y>y^*\}$$

That is, the average tax rate τ increases to $\tau+\Delta\tau$ discontinuously when taxable income y exceeds the threshold y^*. The notch we are considering here is that both the average and marginal tax rates change discontinuously at the threshold, but the method of analysis would be similar as the case of tax kink (i.e., the marginal tax rate does not change at the threshold and only the average tax rate changes).

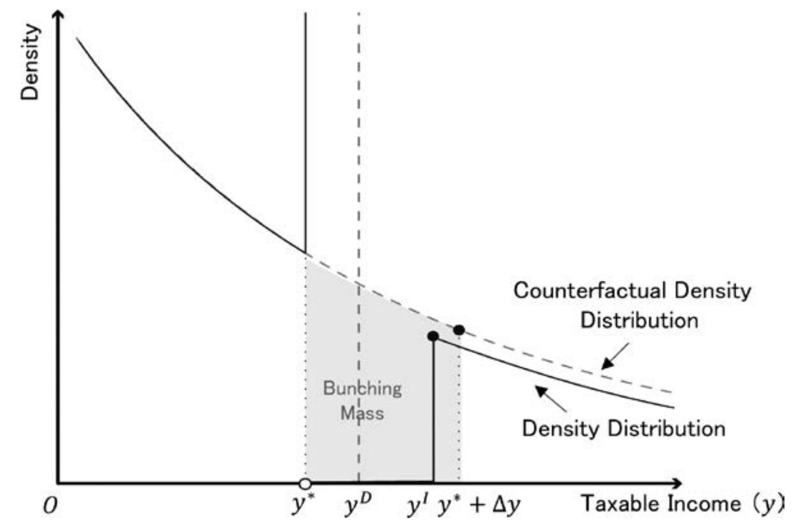
Figure 2 panel (a) shows a budget line (i.e., a straight line with slope $1-\tau$) when tax rate τ is applied for all taxable income y, and a budget line (i.e., a solid line) when tax rate τ is applied below threshold y^* and tax rate $\tau+\Delta\tau$ is applied above threshold y^*. Panel (b) shows their corresponding income distributions.

In the absence of a notch, as shown in panel (a), individuals a^* and $a^*+\Delta a$ choose y^* and $y^*+\Delta y$, respectively, with the indifference curve tangent to the dotted line. However, when there is a notch in the budget line, all individuals who select an income between $(y^*, y^*+\Delta y]$ will bunch at the threshold. Individuals who are at $y^*+\Delta y$, the upper bound of the interval, before the tax increase are marginal bunchers, and their choice of the threshold y^* and their choice of point y^I in the figure are indifferent in the tax schedule where the notch exists. The change in the income distribution resulting from the individual's behavioral response to the tax change is shown in panel (b). As can be seen from the figure, there is a hole in the densi-

Figure 2. The Case of Notch
(a) Budget Line



(b) Taxable Income Density Distribution



Note 1: Panel (a) shows the budget line for individuals with taxable income on the horizontal axis and after-tax income on the vertical axis. It also shows indifference curves for individuals a^* and $(a^* + \Delta a)$. The dotted line with slope $(1 - \tau)$ extending from the origin shows the budget line when there is no notch, and the one extending from the origin with slope $(1 - \tau)$ and moving discontinuously with taxable income y^* to the solid line with slope $(1 - \tau - \Delta\tau)$ shows the budget line when a notch exists. Individuals who are marginal bunchers $(a^* + \Delta a)$ also choose $y^* + \Delta y$ when no notches are present. On the other hand, in the presence of a notch, the choice between y^* and y^I is indifferent. In addition, the range y^* to $y^* + y^D$ is strictly dominated by the choice of the other range of taxable income, even when heterogeneity exists in the elasticity.
Note 2: Panel (b) shows the density distribution of taxable income with and without notches, corresponding to panel (a). The real density (solid line) is the density distribution when a notch is present, and the counterfactual density (dotted line) is the density distribution when a notch is not present. The shaded area between y^* to $y^* + \Delta y$ represents the amount of individuals who bunch at y^*.

ty distribution just above the threshold, indicating that there are no individuals who choose between y^* and y^I. Individuals who choose a taxable income above $y^* + \Delta y$ before the tax change do not lower their income to the threshold y^*, although their income choice is lowered by the tax increase.

Kleven (2016) points out that the major difference between kink and notch is that in the case of a notch, the income range that is strictly dominated by other choices (i.e., here the range $(y^*, \Delta y^D)$ in Figure 2) arises. In this range, no matter what utility function is assumed, it is possible to increase both consumption and leisure by lowering income to the threshold. Since Figure 2 shows the case where elasticity is the same for all individuals, there are no individuals in the range (y^*, y^I) when notch exists[6]. In the case of L-shaped Leontief preferences, the strictly dominated income region corresponds exactly to (y^*, y^D).

To obtain the elasticity of taxable income η in the case of notch, the following conditions are used for the marginal bunchers. Again, we assume the iso-elastic utility function equation (3) as in the case of kink. Since the marginal bunchers (with ability $a^* + \Delta a$) are indifferent between the choice of threshold y^* and the point y^I. The utility at the former is

$$U^* = (1-\tau)y^* - \frac{a^*+\Delta a}{1+1/\eta}\left(\frac{y^*}{a^*+\Delta a}\right)^{1+1/\eta}.$$

On the other hand, the utility of the latter can be written as follows:

$$U^I = (1-\tau-\Delta\tau)y^I - \frac{a^*+\Delta a}{1+1/\eta}\left(\frac{y^I}{a^*+\Delta a}\right)^{1+1/\eta}.$$

U^I can be expressed as

$$U^I = \left(\frac{1}{1+\eta}\right)(a^*+\Delta a)\,(1-\tau-\Delta\tau)^{1+\eta}$$

by using the first-order condition $y^I = (a^* + \Delta a)\,(1 - \tau - \Delta\tau)^\eta$. Using the condition $U^* = U^I$ under which the utility of both is indifferent and the first-order condition $y^* + \Delta y = (a^* + \Delta a)(1 - \tau)^\eta$ for the marginal bunchers under the before-tax schedule, we obtain the following equation:

$$\frac{1}{1+\Delta y/y^*} - \frac{1}{1+1/\eta}\left(\frac{1}{1+\Delta y/y^*}\right)^{1+1/\eta} - \frac{1}{1+\eta}\left(\frac{1-\tau-\Delta\tau}{1-\tau}\right)^{1+\eta} = 0. \tag{5}$$

Equation (5) corresponds to equation (4) in the case of kink. Since y^*, τ and $\Delta\tau$ are known, the elasticity η can be estimated by estimating Δy from the data.

II-3. Procedures for Empirical Analysis

As explained in II-1, the Δy, which we need to estimate the elasticity, depends on the es-

6 See Kleven and Waseem (2013) and Kleven (2016) for a detailed discussion of cases where heterogeneity exists in individual preferences.

timated value of the density distribution $h_0(y)$ when there is no kink or notch. In this section, I present a method for estimating this density distribution based on Chetty et al. (2011) and Kleven and Waseem (2013). The basic strategy for estimating the density distribution is to exclude data in the range around the threshold y^*, then fit a polynomial to the observed income distribution, and extrapolate the distribution around the threshold. Specifically, we draw a histogram of taxable income and estimate the counterfactual distribution $h_0(y)$ using the following regression equation:

$$c_j = \sum_{i=0}^{p} \beta_i (y_j)^i + \sum_{i=y_L}^{y_U} \gamma_i \cdot 1\{y_j = i\} + \epsilon_j, \tag{6}$$

where c_j is the number of individuals (i.e., frequency) in bin j, y_j is the median taxable income in bin j, and ϵ_j is the error term. $[y_L, y_U]$ is the range of income to be excluded near the threshold, and p is the degree of the polynomial. Regarding the choice of the range $[y_L, y_U]$, when estimating bunching at kink, the exclusion range should be selected in the region where excess bunching occurs. In the case of a notch, it is necessary to exclude all areas of the distribution that are affected by the bunching reaction. The detailed procedure for selecting the exclusion area is described below.

The counterfactual density distribution is obtained as the predicted value $\widehat{c_j} = \sum_{i=0}^{p} \widehat{\beta_i} (y_j)^i$ excluding the term $\sum_{i=y_L}^{y_U} \gamma_i \cdot 1\{y_j = i\}$ from equation (6). The excess bunching B can be estimated as $\hat{B} = \sum_{i=y_L}^{y_U} (c_i - \widehat{c_i})$, which is the difference between the actual observed values in the excluded range $[y_L, y_U]$ and the predicted values obtained from the counterfactual distribution.

First, I explain the estimation method in the case of kink. As shown in equation (2), if Δy is sufficiently small, the excess bunching could be expressed by the approximation $B \simeq h_0(y^*)\Delta y$. Theoretically, $h_0(y^*)$ represents the density at threshold y^*, but empirically it corresponds to the density in bins of width W, which can be written as

$$B \simeq \frac{h_0^W(y^*)\Delta y}{W}, \tag{7}$$

where $h_0^W(y^*)$ represents the density distribution with bins of width W. Using this equation (7), the elasticity is expressed as:

$$\eta \simeq \frac{\frac{B}{h_0^W(y^*)}}{\frac{\Delta\tau}{1-\tau} \cdot \frac{y^*}{W}}.$$

Thus, we will estimate the relative excess bunching $b \equiv B / h_0^W(y^*)$, where we normalize the amount of excess bunching B by the counterfactual density distribution at the threshold value. A possible approach to estimate η would be using the above estimate of excess bunching $\hat{B}$, but it has been noted that this estimate tends to overestimate value (Chetty et al. 2011). The reason for this is that it does not take into account the following fact: when a kink occurs, individuals who originally chose to stay above the threshold will lower their income in response to an increase in the marginal tax rate. The actual density above the threshold used for estimation is likely to be underestimated relative to the counterfactual density in the absence of kink. Nevertheless, the estimation of the counterfactual density distribution makes use of the underestimated value, thus introducing a bias in the estimate. Chetty et al. (2011) proposed the following regression as a way to address this issue, obtaining $\widehat{c_j} = \sum_{i=0}^{p} \widehat{\beta_i}(y_j)^i$. The regression is iterated until the estimated value of β_i comprising $\hat{B}$ on the left-hand side is equal to the estimated value of β_i on the right-hand side[7].

$$c_j \cdot \left(1 + 1\{j \geq y_U\} \cdot \frac{\hat{B}}{\sum_{i=y_U}^{\infty} c_j}\right) = \sum_{i=0}^{p} \beta_i (y_j)^i + \sum_{i=y_L}^{y_U} \gamma_i \cdot 1\{y_j = i\} + \epsilon_j$$

Using the $\widehat{c_j}$ obtained from the regression, we can construct a new excess bunching $\tilde{B} = \sum_{i=y_L}^{y_U} (c_i - \widehat{c_i})$. The relative excess bunching estimator can be written as follows:

$$\hat{b} = \frac{\tilde{B}}{(\sum_{i=y_L}^{y_U} \widehat{c_i})/N},$$

where N is the number of bins in the exclusion range. In general, the standard error of $\hat{b}$ is obtained by the bootstrap method. First, a new c_j is generated by performing a random sampling with replacement from the residual sample of the regression. Using this sample, the regression is performed and the relative excess bunching b is calculated from $\widehat{c_j}$. We define the standard error of $\hat{b}$ as the standard deviation of the distribution of $\hat{b}$ obtained by repeating this process a certain number of times. The elasticity estimate $\hat{e}$ and the standard error of $\hat{e}$ can also be obtained using $\hat{b}$ using the delta method.

Next, I explain the estimation method for the case of a notch. In this case, equation (6) is estimated in the same way, but the major difference is that y_U, the upper limit of missing mass, is determined using information on actual density distribution. First, the excess bunching $\widehat{b_E}(y^*)$ in the range of the threshold y^* is calculated from y_L, and the upper limit y_U is determined so that it equals to the missing mass $\widehat{b_M}(y^*)$ that occurs above the threshold value. Here, $\widehat{b_E}(y^*)$ and $\widehat{b_M}(y^*)$ can be defined as follows:

$$\widehat{b_E}(y^*) = \sum_{j=y_L}^{y^*} (c_j - \widehat{c_j}),$$

$$\widehat{b_M}(y^*) = \sum_{j=y>y^*}^{y_U} (\widehat{c_j} - c_j).$$

The specific procedure is to set y_U close to the threshold value and then raise y_U so that $\widehat{b_E}(y^*) = \widehat{b_M}(y^*)$ in the search for the appropriate value. The reason why we determine y_U in

[7] However, Kleven (2016) states that the magnitude of this bias is often negligible for the following two reasons. The first reason is that the response of individuals to kink is very localized, so the change in the distribution tends to be small. The second is that the density above the threshold is close to the counterfactual density distribution unless the slope of the distribution is sufficiently steep. Therefore, the effect of underestimation due to kink is considered to be small.

this manner in the notch case is that, unlike kink, the income range that is not selected by the individual tends to occur just above the threshold, so it is difficult to visually determine the bunching region (Kleven and Waseem 2013). After determining the exclusion range using the method described above, the estimation method is the same as for the case of kink.

In the following, I discuss some points to be considered in the estimation. The first is the selection of the exclusion range $[y_L, y_U]$. Basically, it is common to determine the range by visually checking the shape of the income distribution around the threshold. In the case of kink, y_L and y_U are selected visually. In the case of a notch, y_L is determined visually, and y_U is determined so that the amount of excess mass below the threshold is equal to the amount of missing mass above the threshold, using the procedure described above. In existing studies, robustness checks have been conducted to ensure that the results do not change significantly when the positions of y_L and y_U are moved. A more systematic method for determining the exclusion range has been proposed and will be introduced in the next section.

The second is to deal with the phenomenon of round number bunching. This phenomenon has been reported in many existing studies, and it has been pointed out that it may be caused by factors different from monetary incentives (e.g., a round number becomes a reference point for individuals). The tax threshold at which kink or notch occurs is often set at a rounded number. Ignoring this round number bunching may result in an overestimation of elasticity. A common approach adopted by Kleven and Waseem (2013) and others is to add a dummy variable to the regression equation that takes one for the round number, thereby eliminating the effect. Specifically, $\sum_{r \in \mathbb{R}} \lambda_r \cdot 1 \left\{ \frac{y_j}{r} \in \mathbb{N} \right\}$ is inserted into equation (6) where $\mathbb{R}$ is a set of round numbers such as one million yen or five million yen, $\mathbb{N}$ is a natural number, and λ_r is a parameter. In other words, when taxable income takes a value that is divisible by round numbers, we control such a number in the regression.

Thus, there is a method to cope with rounded number bunching by controlling a dummy variable for the rounded digits, but it may not work well when the noise in the data is large. For example, He et al. (2021), in addition to the usual analysis in which dummy variables are used to control for a round number bunching, conduct an analysis using a sample in which the observation in the round numbers is excluded. Note, however, that the latter method excludes individuals whose incomes are precisely adjusted to the threshold (assumed here to be a round number), which may lead to an underestimation of the elasticity[8].

III. Progress of Estimation Method

As noted above, the basic analytical framework using kinks and notches in the budget line was presented in Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013). Ex-

[8] Another possible approach, conditional on the availability of data, is to use the distribution before the kinks and notches or the distribution of a control group not facing such taxation as a counterfactual distribution. The analysis of Devereux et al. (2014) and Gelber et al. (2020) are examples, although they are not using this to deal with round number bunching.

isting studies have developed a number of discussions on: (i) optimization frictions faced by individuals, (ii) the identifiability of taxable income elasticities, (iii) improvement on the technical aspects in bunching estimation, and (iv) comparisons with other estimation methods. In this section, I discuss the progress of existing studies related to the above-mentioned bunching estimation methods.

III-1. Optimization Frictions

In existing studies in bunching estimation, it has been pointed out that the optimization frictions faced by economic agents prevent bunching responses to tax kinks and notches (e.g., Chetty et al. 2011; Chetty 2012; Chetty et al. 2013; Kleven and Waseem 2013; Aghion et al. 2017; Tazhitdinova 2020). This is because the following facts have been observed in these studies: (1) the bunching tends to be larger when kinks and notches are larger (e.g., Chetty et al. 2011), (2) the bunching tends to be larger when the tax schedule is more stable over time (e.g., Saez 2010), and (3) even if a kink disappears, the bunching at the threshold does not disappear immediately (e.g., Gelber et al. 2020). For (1), the larger a kink or notch, the greater the incentive for economic agents to overcome the cost of adjustment, which can be interpreted as a larger bunching. For (2) and (3), the long-run response is expected to be larger than the short-run response because, although frictions exist in the short-run that prevent adjustment, they dissipate in the long-run (Saez 2010; Saez et al. 2012).

Thus, the existence of optimization frictions has been discussed in the literature from relatively early studies. Of particular importance, however, is the effect of the presence of frictions on estimates of the elasticity of taxable income. Naturally, because optimization frictions affect the size of the excess bunching, the elasticity, a long-run structural parameter that is important for welfare analysis, diverges from the elasticity estimated from short-term fluctuations (Chetty 2012). Various sources of optimization frictions have been considered, including behavioral biases, information asymmetry, monetary and time adjustment costs, and non-standard preferences.

From the perspective of estimating the elasticity of taxable income, it is important to isolate this friction and estimate the structural parameter elasticity. Indeed, Chetty (2012) points out that even if the optimization friction is small, the estimated bound of the structural elasticity is often wide. Quantification of the size of the optimization friction is necessary in conducting welfare analysis (e.g., Chetty 2009; Chetty et al. 2009).

In discussion of optimization frictions, Chetty et al. (2011) - an early work in the literature of bunching estimation - use Danish administrative tax data in their analysis that takes into account the presence of optimization frictions. In the model they present, firms offer job offers with working time constraints and wage workers pay job search costs to search for jobs. Workers face certain costs in optimizing in the model. Given this model, the bunching observed in the data can be explained as follows. That is, (1) the greater the kink, the easier it is for the worker to overcome the job search cost. Therefore, more workers adjust their taxable income to the kink, and the estimated elasticity of taxable income is greater. (2) In a kink faced by more workers, firms and unions adjust wage and hour packages to the aggregate will of workers. This results in larger bunching (i.e., greater elasticities are observed).

In the subsequent literature, two main approaches to estimating structural elasticities are presented: the first is the approach by Chetty (2012) and Kleven and Waseem (2013) to estimate boundary values of structural elasticity parameters. This approach does not identify the magnitude of friction. The second approach, on the other hand, exploits variations in kinks and notches independent of structural elasticity and optimized friction, as well as variations due to institutional changes (e.g., Gelber et al. 2020; Zaresani 2020; He et al. 2021). This method differs from the first approach in that it is characterized by a more parametric specification of the magnitude of the optimizing friction, which is estimated together with the elasticity.

Chetty (2012) and Kleven and Waseem (2013) are representative studies of the former approach. Chetty (2012) assumes a dynamic life-cycle model in which individuals make decisions within certain constraints that are acceptable to the rate of reduction in lifetime welfare relative to adjustments in working hours. I do not go into the details, but Chetty shows that based on this model, the upper and lower limits of structural elasticity, ϵ_U and ϵ_L, can be approximated as a function of the observed elasticity value $\hat{\eta}$, the size of the tax rate change $\Delta\log(1-\tau)$, and the parameter δ which is the rate of decline in an individual's welfare. The following are the elasticities:

$$\epsilon_U = \hat{\eta} + \frac{4\delta}{\left(\Delta\log(1-\tau)\right)^2}(1-\rho)$$

$$\epsilon_L = \hat{\eta} + \frac{4\delta}{\left(\Delta\log(1-\tau)\right)^2}(1+\rho),$$

where $\rho = \left(1 + \frac{1}{2}\frac{\hat{\eta}}{\delta}\left(\Delta\log(1-\tau)\right)^2\right)^{1/2}$. Although δ must be assigned a specific value, Chetty (2012) states that in many cases $\delta = 1\%$ is a reasonable value.

Next, Kleven and Waseem (2013) present a method for estimating the lower and upper bounds of structural elasticity in tax schedules where a notch exists. First, the lower bound is estimated as follows. Let $[y^*, y_D]$ be the domain strictly dominated by individuals' income choices. The fraction of individuals who choose taxable income in this region, r, is defined as

$$r \equiv \frac{\int_{y^*}^{y^*+\Delta y_D} h(y)\,dy}{\int_{y^*}^{y^*+\Delta y_D} h_0(y)\,dy},$$

where $h(\cdot)$ is the density distribution of taxable income as actually observed. The lower limit of structural elasticity can be estimated by using $B/(1-r)$ instead of the excess bunching B used in the usual elasticity estimation. The reason for this can be explained as follows. The closer the distance between the income that would have been chosen in the absence of the notch and the threshold, the stronger the incentive to overcome frictions in the presence of the notch, so individuals adjust their income to the threshold. Conversely, the further away from the threshold, the greater the proportion of individuals who do not adjust. Since r here represents the fraction of individuals who do not adjust to the threshold using only the strictly dominated region, it does not take into account the fraction of individuals who do not adjust in the region above y_D. Therefore, this proportion is underestimated, which means that the magnitude of friction is underestimated. As a result, the hole just above the threshold is evaluated larger (i.e., the bunching range just above the threshold is narrower) and the structural elasticity is evaluated smaller, yielding the lower bound.

Next, the upper bound of elasticity can be determined as Δy, the response of individuals who adjust from the income at the point where the actually observed income distribution and the counterfactual distribution converge to the threshold. This is because among individuals who adjust to the threshold, the individual with the highest income in the absence of a notch is considered to have the highest elasticity.

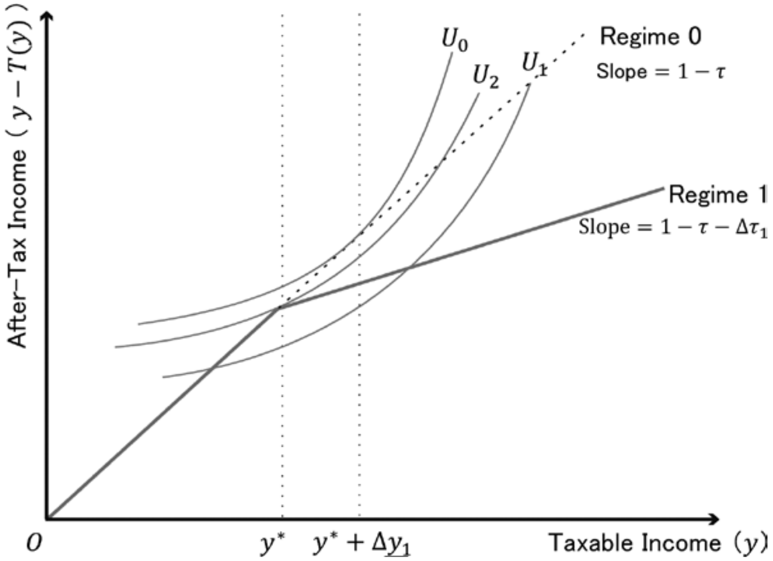
Next, I introduce Gelber et al. (2020), a representative study in the latter approach, which is the first to present a framework for simultaneously estimating the elasticity of taxable income and the optimization frictions. They use the decrease in the marginal tax rate above the kink due to a policy change in the Social Security Annual Earnings Test in the U.S. to analyze workers who are approaching retirement age and have relatively low wages. Specifically, they assume that the magnitudes of elasticity and optimization frictions are the same across kinks before and after the policy change, and estimate the magnitudes of the two unknown values of elasticity and optimization frictions by using the bunching at each kink at each point of time. I will describe the procedure below.

The basic setup is the same as in Section II-1, but the individual's utility function is assumed to be of the form $(c, y; a) - \phi$, with the optimization friction ϕ arising as a fixed cost to utility. The elasticity of an individual's taxable income is assumed to be η, and ϕ and η are assumed to be constant over time. Suppose that a kink is introduced where the marginal tax rate increases from τ to $\tau + \Delta\tau_1$ for incomes above a certain pre-tax income level y^* as a threshold (where $\Delta\tau > 0$). For clarity in the discussion that follows, we refer to the tax regime in this case as "Regime 1" and the tax regime in the absence of the kink as "Regime 0". Figure 3 panel (a) shows the budget lines for Regimes 0 and 1.

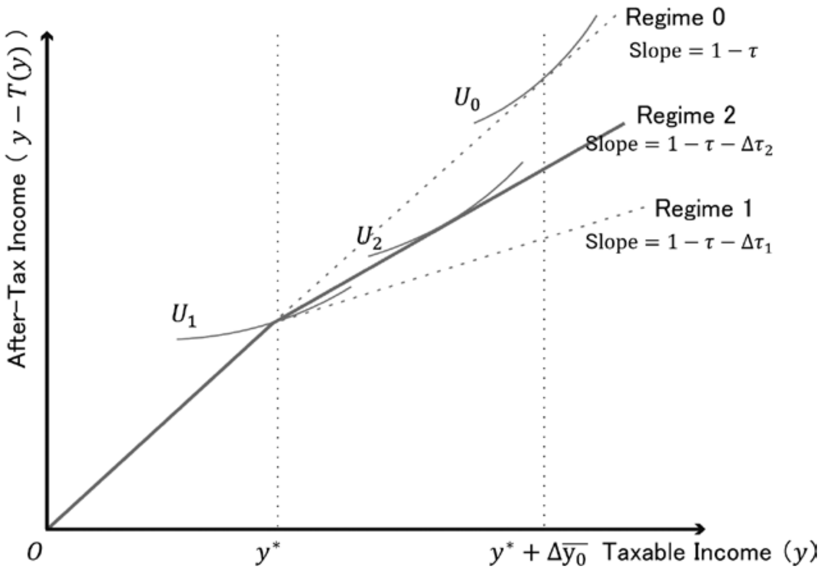
If there are no optimizing frictions ϕ, from the discussion in Section II-1, individuals who had chosen $y \in (y^*, y^* + \Delta y_1]$ in Regime 0 would adjust their income to the threshold y^*. However, in the presence of optimization frictions, individuals who had chosen $y \in (y^*, y^* + \underline{\Delta y_1}]$ in Regime 0 do not adjust their income because the cost of adjusting their taxable income to y^* is large relative to the benefit gained from doing so. Figure 3 panel (a) shows the choices made in Regimes 0 and 1 by individuals with ability a_1 who is indifferent between making and not making income adjustments. Individuals a_1 chooses $y^* + \underline{\Delta y_1}$ under the indifference curve U_0 in Regime 0, while in Regime 1, the utility level obtained by choosing y^* by paying frictional costs ϕ under the indifference curve U_2 and the level of utility obtained by staying at $y^* + \underline{\Delta y_1}$ (i.e., indifference curve U_1) are equal. Then, the following holds:

Figure 3. The Case of Kink: Individuals Face Optimization Frictions

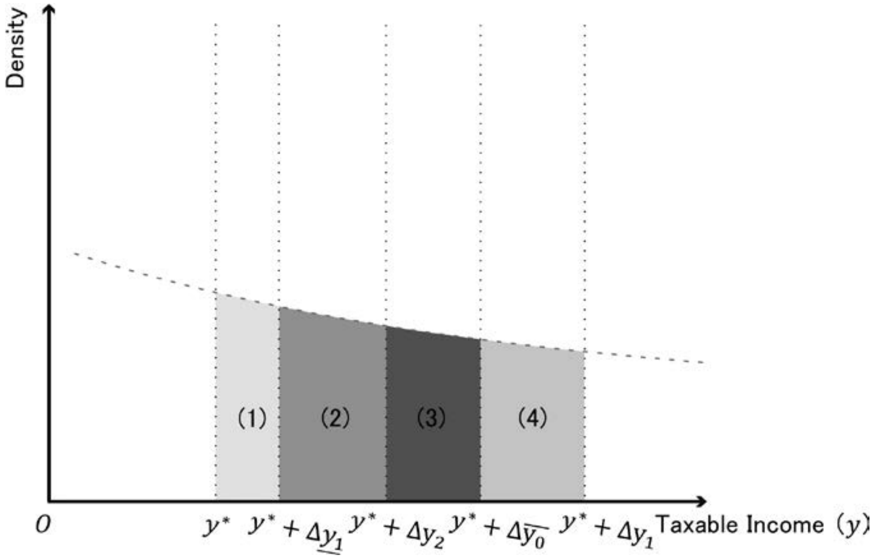
(a) Budget Line: Change from Regime 0 to Regime 1



(b) Budget Line: Change from Regime 1 to Regime 2



(c) Taxable Income Density Distribution



Note 1: Panels (a) and (b) show individual budget lines with taxable income on the horizontal axis and after-tax income on the vertical axis. Panel (a) shows the indifference curves U_0, U_1, and U_2 associated with the change in the budget line from Regime 0 to Regime 1 and the response of individual a_1 to the change. The indifference curve U_0 is tangent to the budget line in Regime 0 at $y^* + \Delta \underline{y_1}$. Panel (b) shows indifference curves U_0, U_1, and U_2 associated with the change in the budget line from Regime 1 to Regime 2 and the response of individual a_2 to the change. The indifference curve U_0 is tangent to the budget line of Regime 0 at $y^* + \Delta \overline{y_0}$, and U_2 is tangent to the budget line of Regime 2.
Note 2: Panel (c) shows the density distribution of taxable income in the absence of kink, corresponding to panels (a) and (b).

$$
\begin{aligned}
&U\left((1-\tau)y^*, y^*; a_1\right) \\
&= U\left((1-\tau-\Delta\tau_1)(y^*+\Delta \underline{y_1}), (y^*+\Delta \underline{y_1}); a_1\right) - \phi,
\end{aligned} \tag{8}
$$

where a_1 is the ability of the individual who had chosen the income $y^* + \Delta y_1$. Note that we assume that $\Delta \underline{y_1} < \Delta y$, so that bunching occurs at the kink even in the presence of optimization frictions. From the above, in Regime 0, since only those individuals who had chosen $(y^* + \Delta \underline{y_1}, y^* + \Delta y_1]$ will move their taxable income to y^*, the excess bunching B_1 can be written as follows:

$$
B_1 = \int_{y^*+\Delta \underline{y_1}}^{y^*+\Delta y_1} h_0(y)\,dy \simeq h_0(y^*)(\Delta y_1 - \Delta \underline{y_1}), \tag{9}
$$

This B_1 corresponds to (2) to (4) in the range of counterfactual distribution shown in Figure 3 panel (c).

Then suppose that the marginal tax rate immediately above the kink is subsequently lowered from $\tau + \Delta\tau_1$ to $\tau + \Delta\tau_2$ (Call this Regime 2, where $\Delta\tau_2 < \Delta\tau_1$. See Figure 3 panel (b).). The case of Regime 2 is also the same as that of Regime 1. For Regime 2, individuals who had chosen taxable income $(y^* + \Delta \underline{y_2}, y^* + \Delta y_2]$ in Regime 0 would have their taxable income adjusted to y^*.

However, since $\varDelta\tau_2 < \varDelta\tau_1$, the incentive to adjust taxable income to the threshold y^* is lower than before the tax cut (Regime 1). As a result, $(y^* + \varDelta y_2) < (y^* + \varDelta y_1)$. At this point, if friction were not present, individuals in the range $(y^* + \varDelta y_2, y^* + \varDelta y_1]$ would not bunch at y^* in Regime 2. However, when the tax schedule changes from Regime 1 to 2 in the presence of frictions, individuals up to the income level at which the adjustment cost exceeds the benefit obtained at the optimal income under Regime 2 will continue to bunch at the threshold y^*. In other words, the individuals in the range of $(y^* + \varDelta \underline{y_1}, y^* + \varDelta \overline{y_0}]$, where $y^* + \varDelta \overline{y_0}$ corresponds to the income for individual a_2 such that the benefit gained from the adjustment from Regime 1 to 2 is exactly equal to the cost of the optimization friction, will continue to bunch at y^*, while individuals in the range of $(y^* + \varDelta \overline{y_0}, y^* + \varDelta y_1]$ would adjust their incomes.

Figure 3 panel (b) shows the indifference curve for individual a_2 on the boundary. As noted above, individuals a_2 choose $y^* + \varDelta \overline{y_0}$ under Regime 0 (see indifference curve U_0), but under Regime 1 they choose the threshold y^* (see indifference curve U_1). And under Regime 2, the choice of $(y^* + \varDelta \overline{y_2})$ (corresponding to the indifference curve U_2) is optimal when there is no optimization friction. However, in the presence of optimization friction, the choice to move from the threshold to $(y^* + \varDelta \overline{y_2})$ and to stay at the threshold is indifferent, so the following equation holds:

$$\begin{aligned} & U\left((1-\tau-\varDelta\tau_2)y^*, y^*; a_2\right) \\ & = U\left((1-\tau-\varDelta\tau_2)(y^*+\varDelta\overline{y_2}), (y^*+\varDelta\overline{y_2}); a_2\right) - \phi . \end{aligned} \tag{10}$$

The excess bunching B_2 in this situation can be written as follows:

$$B_2 = \int_{y^*+\varDelta\underline{y_1}}^{y^*+\varDelta\overline{y_0}} h_0(y)\,dy \simeq h_0(y^*)(\varDelta\overline{y_0} - \varDelta\underline{y_1}) . \tag{11}$$

This B_2 corresponds to (2) and (3) in the distribution shown in Figure 3 panel (c).

In the following discussion, we further assume an iso-elastic utility function, equation (3). Under this utility function, equation (8) derived under Regime 1 becomes:

$$\begin{aligned} & (1-\tau)y^* - \frac{a_1}{1+1/\eta}\left(\frac{y^*}{a_1}\right)^{1+1/\eta} \\ & = (1-\tau-\varDelta\tau_1)(y^*+\varDelta\underline{y_1}) - \frac{a_1}{1+1/\eta}\left(\frac{y^*+\varDelta\underline{y_1}}{a_1}\right)^{1+1/\eta} - \phi . \end{aligned} \tag{12}$$

Since the first-order condition in Regime 1 for this individual is $y^* + \varDelta\underline{y_1} = a_1(1 - \tau - \varDelta\tau_1)^\eta$, we substitute this into equation (12), and then $\varDelta\underline{y_1}$ can be expressed as the following equation of the unknown parameters η and ϕ:

$$\begin{aligned} & (1-\tau)y^*\left[\frac{1}{1+\eta}\left(1-\frac{\varDelta\tau_1}{1-\tau}\right)\left(1+\frac{\varDelta\underline{y_1}}{y^*}\right) + \frac{1}{1+1/\eta}\left(1-\frac{\varDelta\tau_1}{1-\tau}\right)\left(\frac{1}{1+\varDelta\underline{y_1}/y^*}\right)^{1/\eta} - 1\right] \\ & = \phi . \end{aligned} \tag{13}$$

The excess bunching B_1 in equation (9) is derived from the first-order condition on the marginal buncher $y^* + \Delta y_1 = (a + \Delta a)(1-\tau)^\eta$, and $y^* = (a+\Delta a)(1-\tau-\Delta\tau_1)^\eta$ derived from $y_1 = y^*\left(\left(\frac{1-\tau}{1-\tau-\Delta\tau_1}\right)^\eta - 1\right)$. B_1 is expressed as:

$$B_1 \simeq h_0(y^*)\left(y^*\left(\left(\frac{1-\tau}{1-\tau-\Delta\tau_1}\right)^\eta - 1\right) - \Delta\underline{y_1}\right) \tag{14}$$

Therefore, from equations (13) and (14), eliminating $\Delta\underline{y_1}$ yields the equality for elasticity η and optimized friction ϕ.

Equation (10), derived by focusing on the change from Regime 1 to 2, can then be written as:

$$\begin{aligned}&(1-\tau-\Delta\tau_2)y^* - \frac{a_2}{1+1/\eta}\left(\frac{y^*}{a_2}\right)^{1+1/\eta}\\ &= (1-\tau-\Delta\tau_2)(y^*+\Delta\overline{y_2}) - \frac{a_2}{1+1/\eta}\left(\frac{y^*+\Delta\overline{y_2}}{a_2}\right)^{1+1/\eta} - \phi.\end{aligned} \tag{15}$$

The first-order conditions in Regime 0 and Regime 2 for this individual can be written as follows, respectively:

$$y^* + \Delta\overline{y_0} = a_2(1-\tau)^\eta, \tag{16}$$

$$\begin{aligned}y^* + \Delta\overline{y_2} &= a_2(1-\tau-\Delta\tau_2)^\eta\\ &= (y^*+\Delta\overline{y_0})\left(1-\frac{\Delta\tau_2}{1-\tau}\right)^\eta.\end{aligned} \tag{17}$$

Here, the second line of equation (17) uses equation (16). Using these, equation (15) can be rearranged as follows:

$$\begin{aligned}&(1-\tau)(y^*+\Delta\overline{y_0})\left[\left(1-\frac{\Delta\tau_2}{1-\tau}\right)^{1+\eta} + \frac{1}{1+1/\eta}\left(\frac{1}{1+\Delta\overline{y_0}/y^*}\right)^{1+1/\eta}\right.\\ &\left. -\left(1-\frac{\Delta\tau_2}{1-\tau}\right)\left(\frac{1}{1+\Delta\overline{y_0}/y^*}\right) - \frac{1}{1+1/\eta}\left(1-\frac{\Delta\tau_2}{1-\tau}\right)^{1+\eta}\right] = \phi.\end{aligned} \tag{18}$$

Equation (11) for the excess bunching B_2 can be rearranged from equation (14) as follows.

$$B_2 \simeq h_0(y^*)\left(\Delta\overline{y_0} - \left(\frac{1-\tau}{1-\tau-\Delta\tau_1}\right)^\eta + \frac{B_1}{h_0(y^*)}\right) \tag{19}$$

Eliminating $\Delta\overline{y_0}$ from equations (18) and (19), we obtain equations for elasticity η and optimization frictions ϕ.

From the above, we can estimate the unknown parameters η and ϕ using two equations: (i) an equation obtained from (13) and (14) and (ii) an equation obtained from equations (18) and (19).

III-2. The Issue of Non-identification

In the bunching estimation, we estimate the elasticity of taxable income based on the information of tax schedule and data of income distribution at a single point in time. This is in contrast to the analysis using changes in the budget line between different points in time due to changes in the tax schedule, which requires panel data for at least two points in time, before and after the tax schedule change. Bunching estimation may have an advantage in that it can estimate elasticities from distributional data at a cross-section of time points and without individual-level information. However, Blomquist et al. (2021) point out the issue of non-identification in bunching estimation and show that the elasticity parameter η and the distribution of individual ability, $f(a)$, cannot be simultaneously identified from a single income distribution. In the following, I explain this non-identification problem based on the model setup in Section II-1.

As described in II-1, the excess bunching B at the kink point could be written as:

$$B = \int_{y^*}^{y^* + \Delta y} h_0(y)\, dy.$$

The identification problem here is that although the two parameters of the counterfactual density distribution $h_0(y)$ and Δy are unknown, there is only one equation that contains these two parameters. The problem is the same even if the utility function is assumed to have the shape of equation (3), which has been widely used since Saez (2010). From the taxable income equation,

$$y(\tau, \eta, a) = a \cdot (1 - \tau)^{\eta},$$

which is derived from the first-order condition for the utility maximization problem, we obtain the taxable income at the threshold as $y^* = a^*(1 - \tau)^{\eta}$, where a^* is the ability of an individual who is exactly tangent to a point at the threshold of the budget line. On the other hand, the taxable income of a marginal buncher who chooses $y^* + \Delta y$ in the absence of kink, but adjusts to the threshold under the kink, can be written as $y^* + \Delta y = (a^* + \Delta a)\ (1 - \tau)^{\eta}$. That is, the range of ability to adjust income is $[a^*, a^* + \Delta a] = [y^*(1 - \tau)^{-\eta}, (y^* + \Delta y)(1 - \tau)^{-\eta}]$.

Then, the excess bunching B can be written as follows:

$$B = \int_{a^*}^{a^* + \Delta a} f(a)\, da = \int_{y^*(1-\tau)^{-\eta}}^{(y^* + \Delta y)(1-\tau)^{-\eta}} f(a)\, da. \tag{20}$$

It is clear from equation (20) that the excess bunching B can be estimated from the data, but the elasticity parameter η and the density function $f(a)$ are unknown. Since the only equation (20) contains these two unknowns, the elasticities cannot be identified. Blomquist et al. (2021) show more generally that for any given taxable income elasticity η, there is always a distribution of ability, $f(a)$, that matches the actual distribution of taxable income in the model.

To deal with this problem of identifying elasticity, early existing studies made additional assumptions on the distribution so that there is a one-to-one correspondence between the income distribution and the parameters of elasticity. For example, Saez (2010) assumed that the density distribution is linear at each upper and lower point at the kink, while Chetty et al. (2011) assumed that the density can be expressed as a polynomial around the kink. It is common practice in subsequent studies to estimate elasticity by imposing the same assumption as Chetty et al. (2011). It is not possible to extract information (e.g., the elasticity of taxable income) about an individual's tax response behavior from the distribution of taxable income for a single tax schedule without imposing some form of restriction on the distribution.

This non-identification problem arises in the case of notch as well. As in the argument of kink, if the ability of an individual who bunch and is just tangent to a point on the threshold of the budget line is a^*, their taxable income at the threshold can be written as $y^* = a^*(1 - \tau)^\eta$. On the other hand, the marginal bunchers (with ability $a^* + \Delta a$) in the notch are indifferent between choosing taxable income at threshold y^* and taxable income $(y^* + \Delta y)$ at the point of contact with the budget line above the threshold. Referring to equation (5), which describe the response of marginal buncher to the notch, we can express $(y^* + \Delta y)$ as $y^* + \Delta y = k(a^* + \Delta a, \tau, \eta)$ by using some function $k(\cdot, \cdot, \cdot)$, because it depends on the elasticity η. Assuming that the ability of marginal bunchers $a^* + \Delta a$ is expressed as $a^* + \Delta a = l(y^* + \Delta y, \tau, \eta)$ with some function $l(\cdot, \cdot, \cdot)$ using the implicit function theorem, the excess bunching B is written as follows:

$$B = \int_{a^*}^{a^* + \Delta a} f(a)\,da = \int_{y^*(1-\tau)^{-\eta}}^{l(y^* + \Delta y, \tau, \eta)} f(a)\,da. \tag{21}$$

Then, the range of abilities that bunch at the notch is $[a^*, a^* + \Delta a] = [y^*(1 - \tau)^{-\eta}, l(y^* + \Delta y, \tau, \eta)]$. From equation (21), the excess bunching B can be estimated from the data in the notch case as in the kink case. However, since the density function $f(a)$ and the elasticity η are unknown, we see that equation (21) alone does not allow us to identify these parameters.

From the above problems in identification, Blomquist et al. (2021) present a method for estimating the possible range of elasticity by determining upper and lower bounds on the size of the excess bunching by imposing certain constraints on the density distribution[9]. However, an example using similar data to Saez (2010) points out that the possible range of taxable income elasticities can be very wide. This result does not mean that the bunching estimation itself is invalid. The bunching estimation is useful for evaluating the sensitivity of the estimates, such as what assumptions to make about the density distribution and how much weakening of these assumptions changes the possible range of elasticity. Similarly, Bertanha et al. (2021) and Goff (2022), who explain the inability to identify the desired parameters from bunching estimation, also propose partial identification strategies that impose

[9] As noted in footnote 8, Gelber et al. (2020) use the distribution for the sample considered to be the control group as the counterfactual distribution, with the problem of non-identification in mind. This is because if the ability distribution of the control group not facing the kink is equal to the ability distribution of the individuals comprising the income distribution we are observing, then the income distribution observed for the control group serves as a counterfactual distribution.

certain shape restrictions on the density distribution and construct a counterfactual distribution. For example, Bertanha et al. (2021) propose that an econometric method can be applied to bunching estimation when censoring exists in the data and discuss a way to deal with it. More specifically, they show that the upper and lower bounds of the elasticity of taxable income can be identified when there is a restriction on the slope of the distribution, and they propose an approach using censored Tobit regression. However, the proposed estimation method does not allow for optimization frictions, making it impractical for actual data. On the other hand, Aronsson et al. (2021) propose a more parametric approach that combines maximum likelihood estimators with bunching estimators, but this framework allows for unobserved ability heterogeneity and optimization frictions.

III-3. Identification of Bunching Region

Existing studies have used visual inspection to identify the region around the threshold where bunching is occurring. In existing studies, the robustness of the estimates is often ensured based on the fact that the estimates do not change significantly even if we move the visually determined area. However, the choice of the bunching region is a very important procedure because it affects the efficiency and consistency of the estimates. For example, if the range is wider than the true bunching range, it is not efficient because it loses observables that can be used in the estimation of the counterfactual density distribution. Conversely, if the range is narrowed, the counterfactual density distribution is incorrectly estimated, and a bias in the elasticity estimate arises. Based on this problem, Bosch et al. (2020) propose a method to systematically select a range from the data based on a certain information criterion. They show from Monte Carlo simulations that elasticities can be estimated more accurately than existing methods that determine ranges visually[10].

We first set the upper and lower bounds near the threshold, and then set the exclusion region that we do not use to estimate the counterfactual distribution. Using the remaining data points, we perform a polynomial regression to predict the frequency of each bin in the counterfactual, including the confidence interval. We then observe the bins whose actual frequencies lie outside the confidence interval around the threshold, and define the bunching range as the range of these bins that are continuously observed outside the threshold. By performing this operation for any combination of upper and lower bounds, the upper and lower bounds of the bunching range can be obtained from the minimum and maximum values obtained.

III-4. Bunching Estimation vs. Estimation Using Changes of Tax Schedule

Several studies have pointed out that the estimates of the elasticity of taxable income

[10] In Dekker and Schweikert (2021), Monte Carlo simulations compare the method of Bosch et al. (2020) with other methods for determining ranges from data and confirm its superiority.

obtained from bunching estimation are generally smaller than those obtained from the institutional tax schedule change approach. For example, a U.S. study finds that the elasticity estimates for wage workers obtained from the tax schedule change approach range from 0.12 to 0.40 (Saez et al. 2012), while the estimates obtained from the bunching estimates are close to zero (Saez 2010). In the Swedish study, the values obtained from the tax schedule change approach are 0.19 for male wage earners and 1.39 for female wage earners (Blomquist and Selin 2010), while the bunching estimates are near zero (Bastani and Selin 2014). Also in the Danish study, the value obtained from the instrumental variable estimation is 0.2 (Kleven and Schultz 2014), whereas the value obtained from the bunching estimation is 0.01 (Chetty et al. 2011). With regard to this difference in elasticity due to different estimation methods, Chetty (2012) and Kleven and Schultz (2014), among others, argue that it is due to the presence of the optimization friction described in section III-1. According to this argument, although estimated elasticities are different between bunching estimation and estimation using tax schedule changes, the friction should be eliminated, and the difference reduced in the long run.

He et al. (2021) directly compare the results of bunching and tax schedule change approach estimates using Chinese administrative panel data on taxable income and explore the factors that produce the differences. They argue that the estimates from the tax change approach is eight times larger than the bunching estimates which are considered to provide structural elasticities accounting for the presence of optimization frictions. They argue that this difference in estimates is due to the fact that each estimation method captures different behavioral responses of economic agents and present a model to explain the mechanism. In their model, economic agents engage in two types of responses to taxes: temporary and permanent. The behavioral responses of economic agents that the bunching estimator identifies are small because they are due to temporary adjustments in taxable income, while the estimates from the tax change approach are larger because they capture permanent responses. Aronsson et al. (2022) perform Monte Carlo simulations of the bunching estimator and the estimator with the tax schedule change approach in estimating the elasticity of taxable income and evaluate bias (consistency) and precision (efficiency). According to their results, the bunching estimator is more accurate than the estimator using the tax schedule change approach. However, they point out that the former often results in a downward bias, while the latter results in a smaller bias under certain conditions. In addition, they propose a new estimator to improve the quality of estimation and argue that the estimator performs better than other estimators in terms of bias and precision.

Not enough research has been conducted to formulate a view on what causes the differences in estimates between the different estimation methods. There are also few existing studies on the conditions under which each estimation method performs particularly well. However, estimating the elasticity of taxable income is an area that has been actively studied around the world, and it is important to understand the characteristics of the estimates obtained from each estimator.

IV. Conclusion

This paper introduces a basic procedure for bunching estimation, which have been used in a wide range of fields in recent years, including public finance and public economics. In addition, I described the method's recent evolution, criticisms of the method, and related debates.

As long as the cross-sectional distributions around kinks or notches in the economic variables of interest, such as the distribution of taxable income, can be observed, bunching estimation is an excellent analytical method. Nonetheless, it is important to bear in mind that the utility function (or objective function) of economic agents must be assumed to have a particular form, and that assumptions must be made regarding the counterfactual density distribution, among other issues. Because the bunching estimator can be performed with a small amount of information, the accuracy of the distribution and the accuracy of the data are very important. In other words, it is not suitable for analysis of data such as survey data, which is prone to noise from respondents, but is suitable for analysis of data with relatively small noise, such as administrative tax data in developed countries.

In Japan, however, the use of administrative tax data has lagged behind other developed nations, and there have been few studies employing bunching estimation (particularly for individual income tax). In recent years, however, there have been some developments, such as the commencement of joint research with the National Tax Agency and the Center for Research and Education in Policy Evaluation (CREPE) of the University of Tokyo receiving tax data from a number of local governments and conducting research on it. As the use of tax data becomes more widespread in Japan, the bunching estimation method and other cutting-edge techniques are anticipated to become more prevalent.

References

Aghion, P., Akcigit, U., Lequien, M., Stantcheva, S. (2017), "Tax Simplicity and Heterogeneous Learning", *NBER Working Paper,* w24049.

Aronsson, T., Jenderny, K., Lanot, G. (2021), "Maximum Likelihood Bunching Estimators of the ETI", (Available at: http://www.usbe.umu.se/ues/ues987.pdf).

Aronsson, T., Jenderny, K., Lanot, G. (2022), "The Quality of the Estimators of the ETI", *Journal of Public Economics,* 212, 104679.

Bastani, S., Selin, H. (2014), "Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule", *Journal of Public Economics,* 109, pp. 36-49.

Bertanha, M., McCallum, A.H., Seegert, N. (2021), "Better Bunching, Nicer Notching", *arXiv preprint.* arXiv:2101.01170.

Blomquist, S., Newey, W.K., Kumar, A., Liang, C.Y. (2021), "On Bunching and Identification of the Taxable Income Elasticity", *Journal of Political Economy,* 129(8), pp. 2320-2343.

Blomquist, S., Selin, H. (2010), "Hourly Wage Rate and Taxable Labor Income Responsiveness to Changes in Marginal Tax Rates", *Journal of Public Economics,* 94(11-12), pp. 878-889.

Bosch, N., Dekker, V., Strohmaier, K. (2020), "A Data-Driven Procedure to Determine the Bunching Window: An Application to the Netherlands", *International Tax and Public Finance,* 27(4), pp. 951-979.

Chetty, R. (2009), "Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods", *Annual Review of Economics,* 1(1), pp. 451-488.

Chetty, R. (2012), "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply", *Econometrica,* 80(3), pp. 969-1018.

Chetty, R., Friedman, J.N., Olsen, T., Pistaferri, L. (2011), "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records", *Quarterly Journal of Economics,* 126(2), pp. 749-804.

Chetty, R., Friedman, J.N., Saez, E. (2013), "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings", *American Economic Review,* 103(7), pp. 2683-2721.

Chetty, R., Looney, A., Kroft, K. (2009), "Salience and Taxation: Theory and Evidence", *American Economic Review,* 99(4), pp. 1145-1177.

Currie, J., Kleven, H., Zwiers, E. (2020), "Technology and Big Data Are Changing Economics: Mining Text to Track Methods", *AEA Papers and Proceedings,* 110, pp. 42-48.

Dekker, V., Schweikert, K. (2021), "A Comparison of Different Data-driven Procedures to Determine the Bunching Window", *Public Finance Review,* 49(2), pp. 262-293.

Devereux, M.P., Liu, L., Loretz, S. (2014), "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records", *American Economic Journal: Economic Policy,* 6(2), pp. 19-53.

Feldstein, M. (1999), "Tax Avoidance and the Deadweight Loss of the Income Tax", *Review of Economics and Statistics,* 81(4), pp. 674-680.

Gelber, A.M., Jones, D., Sacks, D.W. (2020), "Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test", *American Economic Journal: Applied Economics,* 12(1), pp. 1-31.

Goff, L. (2022), "Treatment Effects in Bunching Designs: The Impact of the Federal Overtime Rule on Hours", *arXiv preprint.* arXiv:2205.10310.

Gruber, J., Saez, E. (2002), "The Elasticity of Taxable Income: Evidence and Implications", *Journal of Public Economics,* 84(1), pp. 1-32.

He, D., Peng, L., Wang, X. (2021), "Understanding the Elasticity of Taxable Income: A Tale of Two Approaches", *Journal of Public Economics,* 197, 104375.

Ichikawa, T., Arudchelvan, M., Onji, K. (2021), "Thresholds and Small Business Bunching in the Consumption Tax: Prospects for a 2019 Tax Increase", *JCER Economic Journal,* 80(2), pp. 25-56. (in Japanese)

Ito, K., Sallee, J.M. (2018), "The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel Economy Standards", *Review of Economics and Statistics,* 100(2), pp. 319-336.

Jakobsen, K.M., Søgaard, J.E. (2022), "Identifying Behavioral Responses to Tax Reforms: New Insights and a New Approach", *Journal of Public Economics,* 212, 104691.

Kawakubo, T., Suzuki, T., Asao, K. (2022), "Tax Avoidance or Compliance Costs Avoidance? Evidence from VAT Reforms in Japan", *PRI Discussion Papers,* 22A(02).

Kleven, H.J. (2016), "Bunching", *Annual Review of Economics,* 8, pp. 435-464.

Kleven, H.J., Schultz, E.A. (2014), "Estimating Taxable Income Responses Using Danish Tax Reforms", *American Economic Journal: Economic Policy,* 6(4), pp. 271-301.

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

Miyazaki, T., Ishida, R. (2022), "Estimating the Elasticity of Taxable Income: Evidence from Top Japanese Taxpayers", *Japan and the World Economy,* 61, 101116.

Onji, K. (2009). "The Response of Firms to Eligibility Thresholds: Evidence from the Japanese Value-Added Tax", *Journal of Public Economics,* 93(5-6), pp. 766-775.

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

Saez, E., Slemrod, J., Giertz, S.H. (2012), "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review", *Journal of Economic Literature,* 50(1), pp. 3-50.

Suzuki, T. (2023), "The Bunching Estimation Approach and Its Development", *Financial Review,* 151, pp. 31-52. (in Japanese)

Tazhitdinova, A. (2020), "Do Only Tax Incentives Matter? Labor Supply and Demand Responses to an Unusually Large and Salient Tax Break", *Journal of Public Economics,* 184, 104162.

Yokoyama, I. (2018), "How the Tax Reform on the Special Exemption for Spouse Affected the Work-Hour Distribution", *Journal of the Japanese and International Economies,* 49, pp. 69-84.

Yokoyama, I., Kodama, N. (2016), "Women's Labor Supply and Taxes: A Data Analysis of the Current Situation", *Special Issue on Tax Reform: Evidence-based Policy Recommendations, Financial Review,* 2, pp. 49-76. (in Japanese)

Zaresani, A. (2020), "Adjustment Costs and Incentives to Work: Evidence from a Disability Insurance Program", *Journal of Public Economics,* 188, 104223.