Evaluating Tax Reforms without Elasticities: What Bunching *Can* Identify

Dylan T. Moore*

April 28, 2022

Abstract

I present a new method for evaluating proposed reforms of progressive piecewise linear tax schedules. Typically, estimates of the elasticity of taxable income (ETI) are used to predict taxpayer responses to changes in tax rates and/or tax bracket thresholds. I show that elasticities are not always needed for this task; the "bunching mass" at a bracket threshold (the share of taxpayers locating there) is a sufficient statistic for the revenue effect of behavioral responses to small changes of the threshold. Building on this finding, revenue forecasting and welfare analysis of threshold changes can be conducted using the pre-reform distribution of taxable income alone. I apply these results in an analysis of the Earned Income Tax Credit, an exercise which motivates extensions addressing taxpayer optimization errors, tax rate heterogeneity, large reforms, and income and participation effects. My approach complements existing bunching methods: it avoids key limitations of bunching-based ETI estimation, but addresses a relatively narrower set of policy questions.

*PhD Candidate. Department of Economics, University of Michigan, Ann Arbor, MI, 48109, United States. Email: dtmoore@umich.edu. I am grateful for helpful comments on earlier drafts received from Spencer Bastani, Ashley C. Craig, Enda Hargaden, James R. Hines Jr., Nirupama Rao, Parker Rogers, Nathan Seegert, Joel Slemrod, Ellen Stuart, and Tejaswi Velayudhan. I also thank seminar participants at the University of Michigan as well as conference participants at the Michigan Tax Invitational 2021, the Young Economist Symposium 2021, and the 114th Annual Conference on Taxation for their valuable feedback.

Progressive, piecewise linear income tax schedules are ubiquitous in modern economies. In such schedules, the marginal tax rate is constant within brackets but varies across brackets, increasing discontinuously at thresholds separating the brackets. Major tax reforms in these economies frequently feature both rate and bracket changes. Consider, for example, the Tax Cuts and Jobs Act of 2017 (TCJA), which substantially reformed the US income tax schedule. This reform reduced tax rates across five out of seven brackets of the US federal income tax schedule, and simultaneously changed the location of many tax bracket thresholds.[1] For example, single tax filers in 2017 entered the top tax bracket when their taxable income surpassed \$418,400. The TCJA moved this threshold to \$500,000 in 2018. The TCJA also featured a substantial increase in the standard deduction, which effectively increased all bracket thresholds as function of pre-deduction income.[2]

Accurately forecasting the revenue impact and efficiency consequences of a proposed tax reform like the TCJA requires a prediction of how tax policy changes will alter taxpayer behavior. That is, it requires estimating the behavioral response to taxation. In modern tax analysis, the elasticity of taxable income (ETI) is a key measure of behavioral responses.[3] Estimates of the ETI are commonly employed by government agencies tasked with forecasting the fiscal impact of proposed tax reforms.[4] However, estimates of the ETI are subject to substantial internal and external validity challenges. These estimates vary considerably according to the methodological approach adopted (Aronsson, Jenderny, and Lanot, 2018; Kumar and Liang, 2020; Neisser, 2021; Saez, Slemrod, and Giertz, 2012; Weber, 2014). Furthermore, there is a strong theoretical and empirical rationale for expecting the ETI to vary substantially with features of the tax system and other contextual factors (Jacquet and Lehmann, 2021; Kopczuk, 2005; Neisser, 2021; Slemrod and Kopczuk, 2002).

This paper shows that for some important features of tax policy there may be no need to rely on ETI estimates in revenue forecasting. I present a new approach to predicting the revenue and welfare effects of tax reforms that include changes to brackets. The proposed method takes advantage of a notable feature of taxpayer behavior under a progressive piecewise linear tax schedule: bunching at kinks. Such schedules induce convex kinks in taxpayer budget sets at bracket thresholds, where the marginal benefit of earning an additional dollar falls discontinuously. Consequently, for a large mass of taxpayers it is optimal to "bunch": to earn precisely the amount of income that lands them at the kink point (i.e. at the tax bracket threshold).

The central finding of this paper is that the share of taxpayers who are bunching at a given kink (the "bunching mass") identifies the revenue impact of taxpayer responses to small changes in the

[1] In many countries such as the US, tax bracket thresholds change every year in a bid to account for inflation. When I talk about a reform that changes these thresholds, I mean that it substantially altered their location in terms of real income.

[2] While the tax code defines brackets in terms of *taxable income* after subtracting various deductions, from an economic theory perspective we care about what taxpayer budget constraints look like as a function of total income.

[3] Feldstein (1999) shows that the ETI can be viewed as a sufficient statistic for the efficiency costs (deadweight loss) of taxation.

[4] For example, the Canadian Parliamentary Budget Office (PBO) uses estimates of the ETI to assess the likely revenue effects of actual and hypothetical tax reforms at the request of Members of Parliament and political parties during election campaigns. See, for instance, https://www.pbo-dpb.gc.ca/web/default/files/Documents/Reports/2016/PIT/PIT_EN.pdf.

location of the kink point. This implies that the revenue effect of moving a tax bracket threshold can be identified from the observed pre-reform distribution of taxable income alone, without relying on ETI estimates.[5] After documenting this result, I explore how it might be used to evaluate actual and potential reforms of real-world tax schedules. I also discuss the implications of these findings for the literature on the use of bunching to estimate the ETI.

Section 1 presents the main finding in the case of a simple two-bracket piecewise linear tax schedule. I assume only that taxpayers have continuous, convex preferences and that they make a frictionless choice of income, allowing for arbitrary heterogeneity in taxpayer preferences.[6] In this setting, the bunching mass is a sufficient statistic for the first-order revenue effect of the behavioral responses caused by changing the location of the threshold separating the two brackets. That is, bunching identifies the behavioral response effect of moving a convex kink point.

The formal derivation of this finding is straightforward, but the intuition behind it is somewhat subtle. Suppose a kink point (tax bracket threshold) is reduced by a small amount, raising the marginal tax rate on income in the window between the new kink and the original kink. This reform raises additional revenue from the taxpayers above the kink who pay a higher tax rate on their income within this window but whose marginal tax rate is unchanged (the *mechanical effect*). However, it also causes some changes in taxpayer behavior that in turn alter tax revenue: the *behavioral response effect* of the reform.

Absent income or labor force participation effects, this behavioral response effect has three parts. Holding constant the size of the bunching mass there is a loss in revenue due to the fact this mass is moved to a lower level of income at the new kink. However, the bunching mass will not remain constant. Some individuals are *former bunchers*: those who bunched at the old kink but will choose not to bunch at the new kink, instead locating somewhere in the upper tax bracket below the old kink. Additionally, some individuals are *new bunchers*: those who used to locate in the lower bracket at a spot above the new kink will begin bunching at the new kink point.

I show that the revenue impact of the behavioral responses of the new and former bunchers reform are second-order. This result obtains in spite of the fact that moving a kink point has a first-order effect on the size of the bunching mass. It is a consequence of the fact that—to a first approximation—the decisions of taxpayers who are at the margin of entering or leaving the bunching mass cause no change in tax revenue. Section 1 presents a graphical derivation to build intuition for this result.

Section 2 of the paper explores the implications of this finding for ex ante evaluation of reforms to progressive, piecewise linear income tax schedules. First, I discuss revenue forecasting. When tax reforms change both rates and brackets, estimates of the ETI are still needed to predict the part of the behavioral response effect caused by the rate changes. However, the bunching mass can still be used to predict the part of the behavioral response caused by the bracket changes. By contrast,

[5]Note that, in this paper I say that a parameter is "identified" by the observed distribution of taxable income if it can be obtained as a function of the distribution (under certain specified conditions). This is consistent with the general definition of identification proposed by Lewbel (2019).

[6]That is, I assume that a taxpayer's observed choice of income reflects their optimal choice of income.

commonly employed approaches to tax revenue forecasting which rely on using the ETI to estimate these behavioral responses are biased even if the ETI is correctly calibrated.

Second, I consider this identification result through an optimal tax theory lens. I show how bunching can be used to identify Pareto- or welfare-improving reforms of existing tax schedules. All prior approaches to this task rely on strong functional form assumptions and account for the behavioral response to taxation using estimates of the ETI. By contrast, welfare analysis of tax bracket thresholds only requires knowledge of the observed distribution of taxable income.

Section 3 illustrates the potential practical value of these findings by presenting an application to the first kink in the Earned Income Tax Credit (EITC). Applying my results to real-world income data requires extending the model to incorporate two key features of real-world tax policy. First, in practical policy settings, different taxpayers often face different effective tax schedules at the same level of income. In the case of the EITC, for example, variation in state-level EITC top-ups and welfare benefit phaseouts generates non-trivial differences in the fiscal impact of moving the EITC across different groups of bunching taxpayers.

Second, real-world taxable income data offer little evidence that taxpayers bunch precisely at kink points. Rather, empirical evidence of bunching generally takes the form of a diffuse lump of taxpayers spread around the kink. Following the prior literature on bunching methods, I assume that this departure from the prediction of the frictionless model is caused by optimization error (Bertanha, McCallum, and Seegert, 2021; Cattaneo, Jansson, Ma, and Slemrod, 2018). That is, I assume that taxpayer choices of taxable income are jointly produced by their optimal choice of income and an idiosyncratic random shock.

Section 3 shows how both of these complexities of practical tax policy can be accounted for without compromising the sufficient statistic interpretation of the bunching mass. The results reveal that explicitly accounting for the role of frictions can substantially change the magnitude of the behavioral response effect to moving the kink.

Section 3 also includes a proposed method for better approximating the revenue effect of "large" changes in kink point location, as well as a discussion of income and labor force participation effects. The bunching mass remains an important sufficient statistic in the presence of these additional behavioral responses: it is a sufficient statistic for the *compensated* behavioral response effect of the reform (i.e. the part of the response caused by substitution effects). I discuss how the bunching mass can be combined with other evidence to generate estimates of the total behavioral response effect.

Following this empirical application, section 4 discusses the findings presented in this paper in the context of the existing literature on bunching methodology. The standard approach to bunching methodology first developed in Saez (2010) uses the bunching mass to generate estimates of the ETI. However, recent work on this approach demonstrates that nonparametric identification of the ETI using the bunching mass and other features of the observed distribution of taxable income is not possible (Bertanha, McCallum, and Seegert, 2021; Blomquist, Newey, Kumar, and Liang, 2019). While various remedies to this limitation have been proposed, all these approaches rely on either strong functional form assumptions or only provide bounds on the ETI. Furthermore, even if identification were possible, the ETI parameter that the standard bunching method seeks to identify may or may not provide policy-relevant information when agent preferences are heterogeneous (Blomquist, Newey, Kumar, and Liang, 2019).

Such work left open the question of whether or not the bunching mass has a generally valid policy-relevant interpretation. This paper shows that it does: the bunching mass is directly informative about the behavioral response effect of local movements of tax bracket thresholds. Nonetheless, as I have noted, estimates of the ETI are still required to evaluate the impact of tax rate changes. Thus, the contribution of this paper is to document a new use case for empirical bunching designs that is more robust than bunching-based ETI estimation, but addresses a narrower set of policy questions.

To address this concern, section 4 also includes a discussion of the key factors that determine whether or not the ETI parameter which bunching methods seek to identify has a policy-relevant interpretation. The applicability of this parameter is compromised when the local average ETI is endogenous to the tax rate. However, the bias introduced into policy analysis by a reliance on this parameter may be small if the change in tax rates at the kink is small, or if the variance of taxpayer elasticities is low, and taxpayer preferences are close to isoelastic.

Related Literature This paper builds on prior work characterizing optimal piecewise linear tax schedules. Sheshinski (1989) and Apps, Van Long, and Rees (2014) present necessary conditions for the optimal location of bracket thresholds. The identification result presented in section 1 is implicitly contained in these conditions, but is not discussed in either paper. The link between these optimal tax theory results and the bunching mass heretofore has gone unnoticed.

This paper's connection to bunching methodology is discussed at length in section 4. However, two prior papers in this literature that deserve particular mention do not appear in that discussion and are instead discussed here. Marx (2018) discusses the welfare effect of moving a regulatory notch that generates bunching behavior in a study of charitable organization behavior. His results are not directly relevant to the results discussed here, but are of note because of their focus on a reform that changes the location of the threshold where bunching occurs.

In more directly relevant work, Goff (2021) presents a variety of extensions of standard bunching methodology, illustrated with an application to estimating the impact of overtime pay regulations. Among other results, he shows that bunching of hours at a regulatory threshold can be used to identify the impact that a small change in this threshold has on the average hours of work. While closely related, this result is distinct from that presented in section 1.[7] As well, the additional

[7] His result does not nest mine as a special case because identification of average changes in a choice variable does not imply identification of a revenue effect when price schedules are nonlinear.

identification results presented in sections 2 and 3 have no counterparts in Goff (2021).[8,9]

Finally, this paper is closely connected to the regression discontinuity (RD) literature. The treatment effect parameter identified in a sharp RD design can be interpreted as the partial effect of the treatment cutoff on the average value of the outcome variable.[10] The work of Dong and Lewbel (2015) can be interpreted as showing that higher-order derivatives of the average value of the outcome variable with respect to the treatment cutoff are also identified. Similarly, I show that higher-order derivatives of tax revenue with respect to bracket threshold function are identified by the observed distribution of taxable income (see Section 2 and Appendix A).

1 The Bunching Mass as a Sufficient Statistic

In this section, I present a simple model of taxpayer behavior, discuss key assumptions, and derive the main identification result. As the formal derivation of this result is straightforward, I focus mainly on exploring the intuition behind it. Several graphical illustrations to aid in the explanation.

1.1 Model

Consider an economy populated by agents of different types $\theta \in \Theta$, where the type space Θ is convex and may be multidimensional. Suppose that taxpayer type is continuously distributed according to F, with a corresponding type density $f(\theta)$.

Each type of agent θ chooses taxable income z to solve a type-specific utility maximization problem

$$\max_{z} \{u(c, z; \theta) : c = z - T(z)\}, \tag{1}$$

given some income tax schedule $T(\cdot)$. I derive the main result under very weak assumptions on agent preferences.

Assumption 1 (Convex, Continuous Preferences)**.** *For all types $\theta \in \Theta$, the utility function $u(c, z; \theta)$ is:*

(i) continuous in (c, z);

(ii) strictly increasing and concave in consumption (c), and;

(iii) strictly decreasing and strictly convex in taxable income (z).

---

[8] This includes empirical tests for Pareto efficiency and welfarist optimality, as well as generalizations of the main identification result to account for optimization errors and tax schedule heterogeneity. These latter results are critically important for practical tax policy applications. This also includes the finding that the bunching mass still identifies a policy-relevant parameter in the presence of income and labor force participation effects: the *compensated* behavioral response effect.

[9] Moreover, all the results presented here were developed independently of Goff's results.

[10] Hahn, Todd, and Van der Klaauw (2001) describe this characterization informally.

This assumption ensures that whenever the tax schedule T is (weakly) convex, the maximization problem (1) admits a unique solution for each type θ, which I label $z(\theta)$. Assumption 1 also ensures that this solution changes continuously in response to infinitesimal tax reforms (i.e. that small tax reforms don't induce "jumps" in the taxpayer's location on the budget constraint).[11]

It is worth highlighting the types of preferences that can be accommodated by assumption 1. For example, it allows for the possibility that an agent's choice of taxable income results from adding together multiple income types as well as avoidance or evasion decisions. This assumption also allows for the possibility of a non-differentiable utility function, so that agents may possess reference-dependent preferences.

A Two-Bracket Tax Schedule

Let $T(\cdot)$ be a piecewise linear tax schedule with a marginal tax rate of t_0 on income below $z^\star$, a marginal tax rate of $t_1 > t_0$ on income above $z^\star$, and a demogrant of G (i.e. $T(0) \equiv 0$). Such a schedule can be simply written as follows

$$T(z) \equiv \begin{cases} T_0(z) & \text{if } z \leq z^\star \\ T_1(z) & \text{if } z > z^\star \end{cases}, \tag{2}$$

where

$$T_0(z) \equiv t_0 z - G, \tag{3}$$

and

$$T_1(z) \equiv t_1(z - z^\star) + t_0 z^\star - G. \tag{4}$$

This schedule consists of two linear tax brackets separated by a bracket threshold at $z^\star$. The upper panel of the figure 1 depicts the budget set induced by this type of tax schedule. I will often refer to $z^\star$ as the "kink point" because, as the figure shows, this budget set features a convex kink at $z^\star$.

Taxpayer Choices Under Linear Tax Schedules

To characterize the choices that taxpayers make when facing tax schedule (2), we must first consider the counterfactual choices they would make when facing the linear tax schedules $T_0(z)$ and $T_1(z)$. These two tax schedules differ in both their marginal tax rates (t_0 vs. t_1) and their associated virtual income (G vs. $(t_1 - t_0) z^\star + G$). Let $z(t, V; \theta)$ be the choice of taxable income a type θ taxpayer would make under a linear tax schedule with marginal tax rate $t < 1$ and virtual income

[11] Assumption 1 implies that each taxpayer's taxable income choices will satisfy the strong axiom of revealed preference (SARP) because taxpayer preferences are rational and their choice of taxable income is always unique. This property allows for the simple characterization of taxpayer behavior presented later in equation (1).

$V \in \mathbb{R}$. Furthermore, let $\mathbb{I}\{\cdot\}$ be an indicator function and let

$$H(z;t,V) \equiv \int \mathbb{I}\{z(t,V;\theta) \leq z\} f(\theta)\, \mathrm{d}\theta$$

be the cumulative distribution function (CDF) of taxable income induced by a linear tax schedule with marginal tax rate $t < 1$ and virtual income $V \in \mathbb{R}$.[12] For simplicity of exposition, I require that this CDF is continuously differentiable in z.[13]

Assumption 2 (Regularity Condition)**.** *For any linear tax schedule with marginal tax rate $t < 1$ and virtual income V, the cumulative distribution function of taxable income induced by this schedule ($H(z;t,V)$) is continuously differentiable for all z, with a corresponding density function $h(z;t,V)$.*

In this section I will further assume that differences in the virtual income amount do not affect taxpayer decisions: that is, there are no income effects. Later, I discuss the implications of income effects for my results.

Assumption 3 (No Income Effects on Taxable Income Choices)**.** *For all agent types θ, suppose that the choice of taxable income is invariant to virtual income V,*

$$\frac{\partial z(t,V;\theta)}{\partial V} = 0,$$

for any marginal tax rate (t) and virtual income (V).

Under this assumption, we can simplify notation somewhat, writing a type θ taxpayer's choice under a linear tax schedule with marginal rate t as $z(t;\theta)$, the corresponding CDF of taxable income as $H(z;t)$, and the corresponding density function as $h(z;t)$.[14] Importantly, note that the distribution $H(z;t_1)$ is invariant to changes in the parameter $z^\star$ that affect the virtual income associated with T_1, and that both $H(z;t_1)$ and $H(z;t_0)$ are invariant to changes in the demogrant amount G.

To further simplify notation, let $z_k(\theta) \equiv z(\theta,t_k)$, $H_k(z) \equiv H(z;t_k)$, and $h_k(z) \equiv h(z;t_k)$ for $k \in \{0,1\}$.

[12] Formally, $H(z;t,V) \equiv \int \mathbf{1}\{z(t,V;\theta) \leq z\}\, \mathrm{d}F(\theta)$.

[13] This assumption could be weakened somewhat. Obtaining my key results requires that the CDF of taxable income under a linear tax schedule is continuously differentiable at a tax bracket threshold. However, these results are robust to the possibility that the CDF of taxable income under a linear tax features discontinuities or non-differentiability at income levels away from the bracket threshold. Thus, the result is robust to the possibility that, for example, some taxpayers have reference-dependent preferences that induce bunching at a reference point, as long as this reference point is not located at the tax bracket threshold and is unaffected by changes to the threshold.

[14] This is because absent income effects we have that for any pair of linear tax schedules with the same marginal tax rate t induce the same distribution of taxable income, irrespective of differences in demogrant amounts. That is, for any $t < 1$ and $V, V' \in \mathbb{R}$,

$$H(z;t,V) = H\left(z;t,V'\right)$$

for all z.

Taxpayer Choices under a Two-Bracket Tax Schedule

When facing the two-bracket tax schedule (2), the choice of taxable income for any taxpayer type θ can be written as:[15]

$$z(\theta) = \begin{cases} z_0(\theta) & \text{if } z_0(\theta) < z^\star \\ z_1(\theta) & \text{if } z_1(\theta) > z^\star \\ z^\star & \text{if } z_1(\theta) \leq z^\star \leq z_0(\theta) \end{cases}. \tag{5}$$

The lower panel of figure 1 illustrates this visually, showing how taxpayers make decisions when facing the two-bracket tax schedule (2). All taxpayers will choose to earn income up to the point where the marginal cost of earning exceeds the marginal benefit. For some, this means locating in the interior of the lower bracket and choosing to earn the same amount of income they earn would under T_0. For others, this means locating in the interior of the upper bracket and earning the same amount they would under T_1.

However, not all taxpayers' choices can be characterized this way. The kink in the tax schedule induces a discontinuous drop in the marginal benefit of income at $z^\star$. Consequently, some taxpayers' marginal cost curves pass through the discontinuity, never intersecting the marginal benefit. For such taxpayers, the marginal benefit of earning each dollar of income below $z^\star$ exceeds their marginal cost but for each dollar beyond $z^\star$ the opposite is true.[16] Thus, they will locate at exactly $z^\star$. These taxpayers are *bunchers.*

Figure 2 shows how the choices of a population of different types of taxpayers facing the tax schedule (2) generate a distribution of taxable income which features "bunching" at the kink point. As in figure 1, the upper panel of figure 2 depicts how taxpayers make choices by plotting the marginal benefit of income against the marginal cost curves of different types of taxpayers. The lower panel of the figure shows the CDF of taxable income H that is induced type distribution F and the type-to-choice mapping $z(\cdot)$ (equation 5). Below the kink point, the observed CDF of taxable income coincides with the CDF under the linear tax schedule T_0 (H_0), and above this point it coincides with the CDF under the linear tax schedule T_1 (H_1):

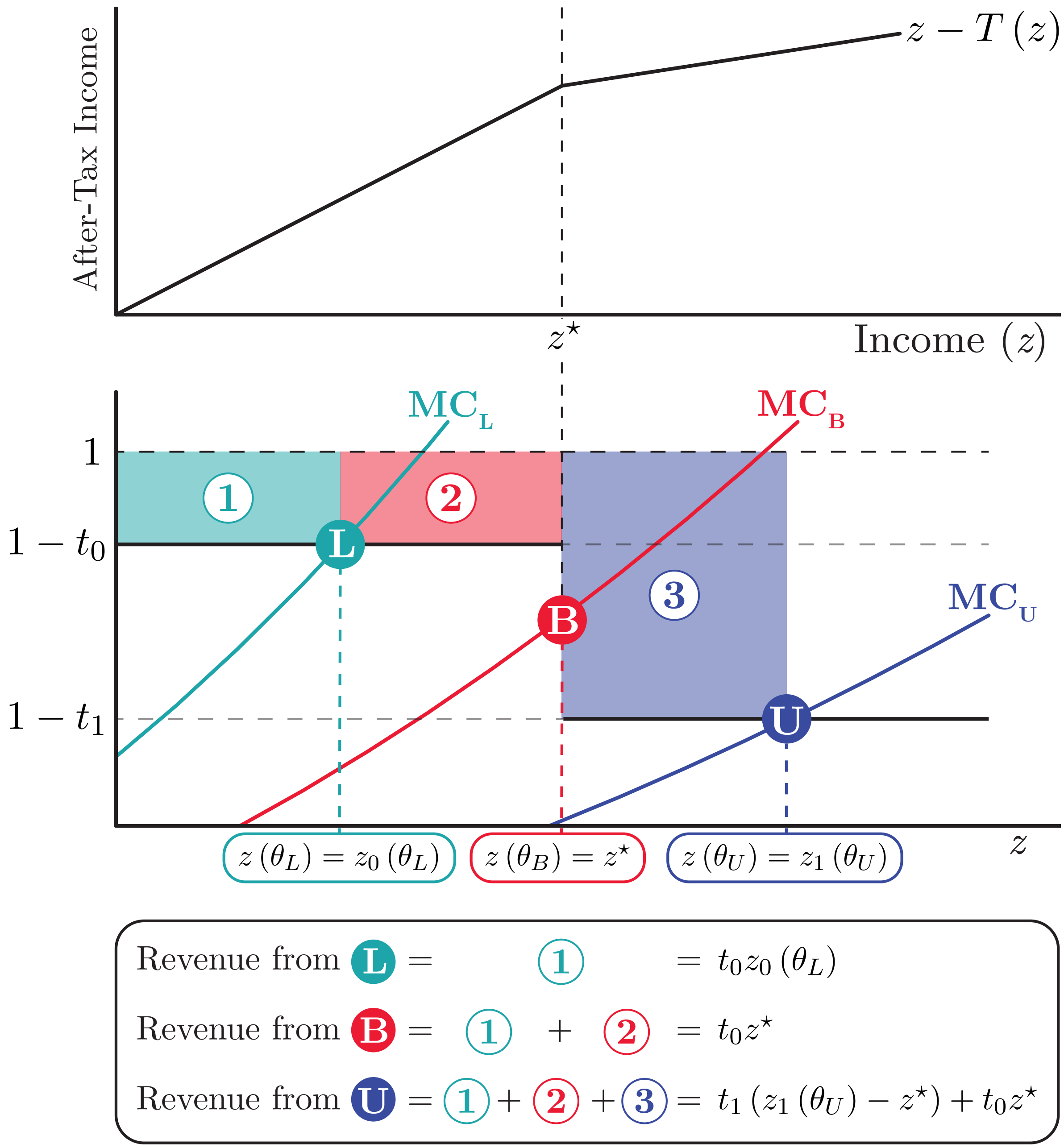
$$H(z) = \begin{cases} H_0(z) & \text{if } z < z^\star \\ H_1(z) & \text{if } z \geq z^\star \end{cases} \tag{6}$$

Importantly, this CDF features a discontinuity in the CDF of taxable income at $z^\star$ (as shown in Figure 2). This is *the bunching mass*: the fraction of agents who would prefer to locate above the

[15] All taxpayers must satisfy one, and only one, of the three conditions in equation (5). The alternative possibility ($z(t_0;\theta) < z^\star < z(t_1;\theta)$) would violate SARP, as $z(t_0;\theta)$ would be affordable under T_1 and $z(t_1;\theta)$ would be affordable under T_0. As Assumption 1 implies that SARP must be satisfied, this is not possible.

[16] Put another way, if such taxpayers faced the linear tax schedule T_0 they would want to locate above the kink, but if they faced a linear tax schedule T_1 they would want to locate below the kink.

Figure 1: Taxpayer Choices and Tax Payments



Notes: The top panel depicts the budget set of taxpayers facing a two-bracket progressive tax schedule. The bottom panel depicts several examples of taxpayer choices and resulting tax payments. The horizontal black curve is the marginal benefit a taxpayer gets from a dollar of income. The marginal benefit falls discontinuously at $z^\star$ due to the kink in the tax schedule. The upward sloping colored curves are three example marginal cost curves. Lower bracket taxpayers (like L) and upper bracket taxpayers (like U) have marginal cost curves that intersect the marginal benefit curve at some point, determining the amount of income they earn. Bunchers, like taxpayer B, have marginal cost curves which run through the discontinuity of the marginal benefit curve. They choose to locate precisely at $z^\star$ because for each dollar earned past this point the marginal cost exceeds the marginal benefit. For each taxpayer, the tax revenue their choice generates is equal to the area of the region between below 1, above the marginal benefit curve, and to the left of their chosen level of taxable income.

kink point under T_0, but below the kink point under T_1:[17]

$$\Pr\left\{z_1\left(\theta\right) < z^\star < z_0\left(\theta\right)\right\} = H_1\left(z^\star\right) - H_0\left(z^\star\right). \tag{7}$$

Figure 2 provides some intuition for why bunching occurs, depicting a set of taxpayers who would all make distinct choices under a linear tax schedule, but many of whom lump together at $z^\star$ when facing a kinked tax schedule.

Note that this presentation ignores important complications that arise in real-world bunching applications due to frictions in taxpayer income choices. Taxable income data are not usually consistent with so-called "sharp" bunching: a large group of taxpayers locating at precisely the kink point. Rather, evidence of bunching usually comes in the form of a diffuse clustering of taxpayers around a kink. Consistent with the existing literature on bunching methodology, I derive my main results by abstracting from such concerns and revisit them in section 3.

1.2 The Bunching Mass as a Sufficient Statistic

Without loss of generality, suppose that $G = 0$. Under assumptions 1–3, tax revenue under the kinked tax schedule (2) is simply:

$$R\left(z^\star\right) \equiv \underbrace{t_0 \int_0^{z^\star} z h_0\left(z\right) \mathrm{d}z}_{\text{revenue below the kink}} + \underbrace{t_0 z^\star \left(H_1\left(z^\star\right) - H_0\left(z^\star\right)\right)}_{\text{revenue from bunchers}} + \underbrace{\int_{z^\star}^{\infty} \left[t_1\left(z - z^\star\right) + t_0 z^\star\right] h_1\left(z\right) \mathrm{d}z}_{\text{revenue above the kink}}. \tag{8}$$

The lower panel of figure 2 depicts tax revenue graphically.

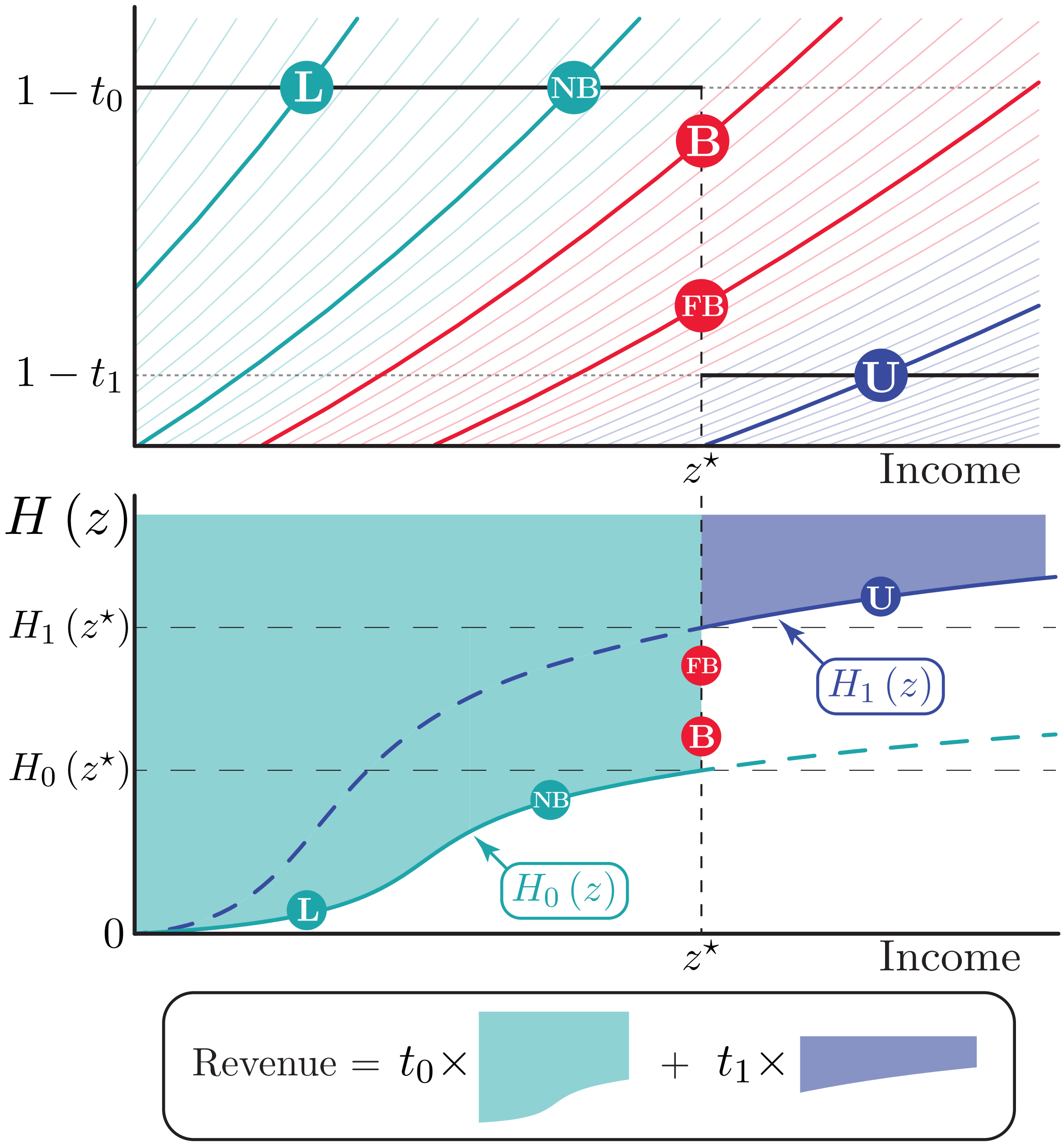
Theorem 1 presents the identification result that provides the foundation of this paper, which can be derived immediately by differentiating the revenue function $R\left(z^\star\right)$. However, this provides little intuition for the result. Here, I present a graphical derivation to provide a deeper understanding of this key finding. Appendix C.1 presents an alternative heuristic derivation.

Now, consider a tax reform displayed in figure 3, which lowers the bracket threshold from $z^\star$ to z'. This reform discretely increases the marginal tax rate in the interval $(z', z^\star]$ while simultaneously reducing the demogrant (virtual income) associated with the second tax bracket. Figure 4 visually depicts how such a reform would change taxpayer choices, and the resulting the revenue effects.

Taxpayers can be broken into five groups that differ in how they are affected by the reform. *Persistent lower bracket taxpayers* are those with pre-reform income below the new kink point ($z_0\left(\theta\right) \leq z^\star$). Their choice remains optimal after the reform so they do not change their behavior in response to it. Similarly, *persistent upper bracket taxpayers*—those with a pre-reform income ($z_1\left(\theta\right) > z'$)—will not respond to the reform if there are no income effects (assumption 3).

[17] The name of this parameter reflects the fact that in the density of taxable income, bunching behavior manifests as a mass point at $z^\star$.

Figure 2: The Pre-Reform Distribution of Taxable Income and Tax Revenue

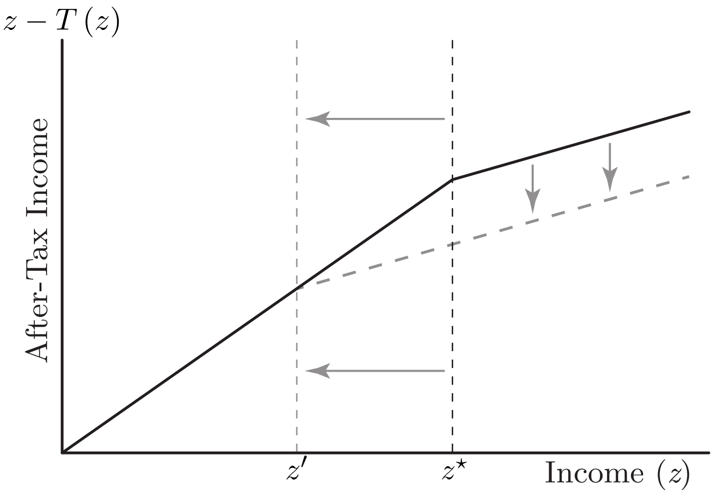


Notes: This figure shows how the choices of a population of taxpayers with differing marginal costs of earning income induce an observed distribution of taxable income with bunching at the kink ($z^\star$), and graphically depicts the resulting tax revenue. The upper panel plots the marginal benefit of earning income alongside a number of representative marginal cost curves. The lower panel shows the resulting CDF of taxable income (H). Below the kink point, H is identical to H_0, the CDF that would be observed under the linear tax schedule T_0. Above the kink, H is identical to H_1, the CDF under the linear tax schedule T_1. At the kink point, H features a positive discontinuity due to fact that a non-zero measure of taxpayers locate at $z^\star$.

The area above H but to the left of $z^\star$ is the average amount of income per taxpayer which is taxed at t_0: $\int_0^{z^\star} z h_0(z)\,\mathrm{d}z + z^\star(H_1(z^\star) - H_0(z^\star)) + z^\star(1 - H_1(z^\star))$. The area above H but to the right of $z^\star$ is the average amount of income per taxpayer which is taxed at t_1: $\int_{z^\star}^{\infty}(z - z^\star)\,h_1(z)\,\mathrm{d}z$. Multiplying each area by its respective tax rate and adding them together yields total tax revenue.

Some taxpayer choices are labeled for reference in later discussion (see figure 4).

Figure 3: Lowering a Tax Bracket Threshold



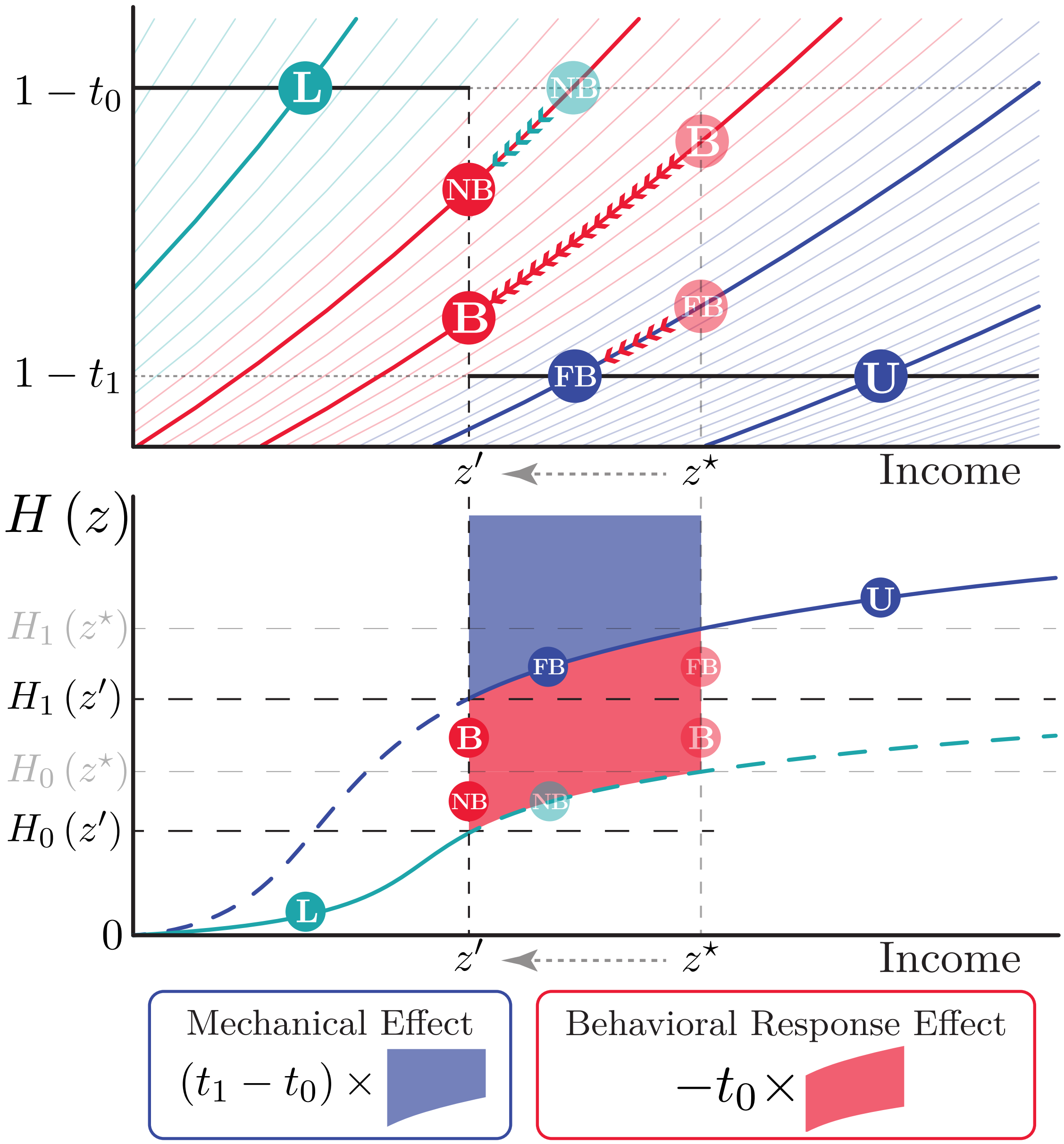
By contrast, taxpayers with pre-reform income in the interval $(z', z^{\star}]$ will respond to the reform. Some are *persistent bunchers*: those who find it optimal to bunch both before and after the reform, reducing their income from $z^{\star}$ to z' ($z_0(\theta) > z^{\star}$ and $z_1(\theta) < z'$). However, not all pre-reform bunchers will choose to continue to bunch. Any pre-reform bunchers with $z_1(\theta) \in (z', z^{\star})$ will find it optimal to leave the bunching mass after the reform, choosing instead to reduce their taxable income from $z^{\star}$ to $z_1(\theta)$. I call this group the *former bunchers*. Finally, there is a group of taxpayers with $z_0(\theta) \in (z', z^{\star})$, who initially have incomes between the new kink point and the old kink. They find it optimal to locate at the new kink point post-reform, reducing their income from $z_0(\theta)$ to z'. I call this group the *new bunchers*.

The *mechanical effect* of the reform is the additional tax revenue that results from taxing incomes in the interval $(z', z^{\star}]$ at an increased marginal tax rate of t_1 (relative to a pre-reform tax rate of t_0). As figure 4 shows, any taxpayers with post-reform incomes above z' contribute to this effect; this group includes both the persistent upper bracket taxpayers and former bunchers. Formally, the mechanical effect can be written as

$$(t_1 - t_0) \underbrace{\left[\left(z^{\star} - z'\right) \left(1 - H_1\left(z^{\star}\right)\right) + \int_{z'}^{z^{\star}} \left(z - z'\right) h_1(z) \, \mathrm{d}z \right]}_{= \text{ area of blue region in figure 4}}. \tag{9}$$

The *behavioral response effect* of the reform is the loss of tax revenue caused by taxpayers reducing their taxable income in response to the increase in the marginal tax rate in the interval $(z', z^{\star}]$. As figure 4 shows, any taxpayers with pre-reform incomes in the interval $(z', z^{\star}]$ contribute to this effect. This includes the persistent bunchers, former bunchers, and new bunchers. Formally, the

Figure 4: Revenue Effects of Decreasing a Kink Point



Notes: This figure visually depicts the effect of reducing the bracket threshold from $z^\star$ to z'. In the top panel, the five labeled taxpayer choices are each representative of a particular class of taxpayers. These labels are also in figure 2 for reference. *Persistent lower bracket taxpayers* (L) and *persistent upper bracket taxpayers* (U); their behavior is unaffected by the reform. However, some taxpayers do respond to the reform. *Persistent bunchers* (B) bunch both before and after the reform, responding by reducing their income from $z^\star$ to z'. *New bunchers* (NB) previously had incomes in the interval $(z', z^\star)$ but reduce this to z', joining the new bunching mass. *Former bunchers* (FB) are induced to exit the bunching mass as a result of the reform, reducing their income from the old kink point $z^\star$ to some value in the interval $(z', z^\star)$.

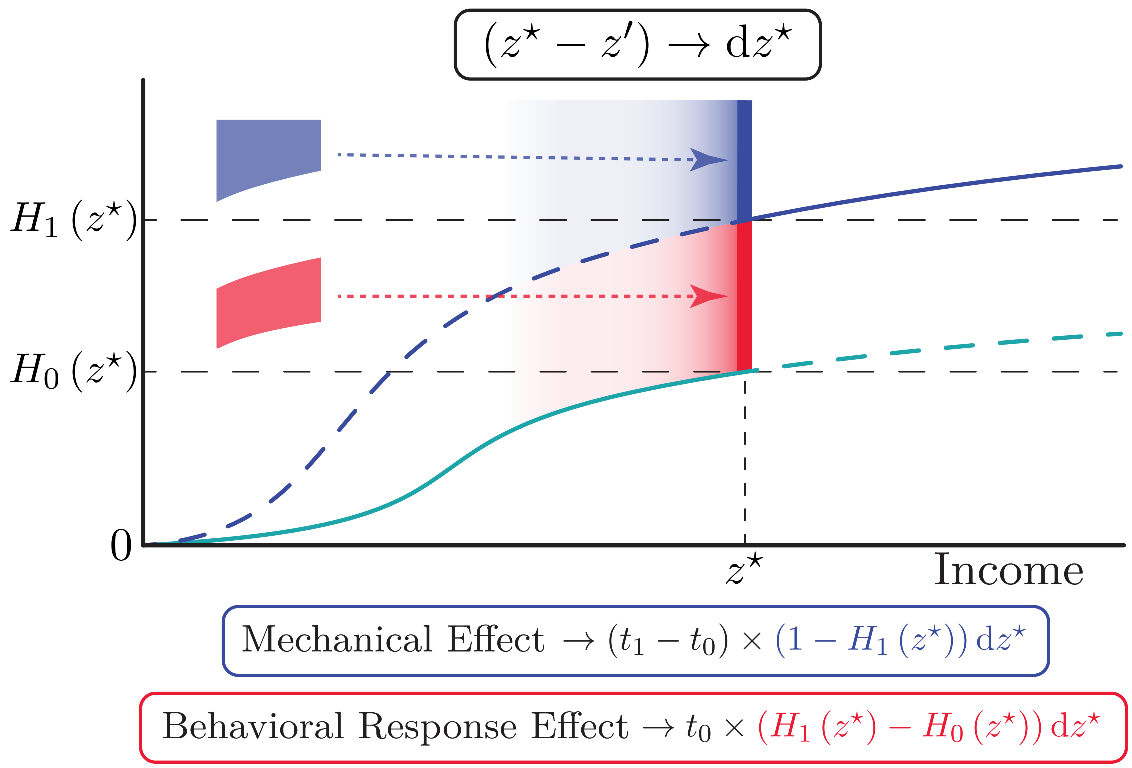
The mechanical effect is the additional revenue that results from taxing income in the interval $(z', z^\star]$ at a higher rate. Taxpayers like U and FB contribute to this effect. The behavioral response effect is the revenue that is lost because some taxpayers in the $(z', z^\star]$ reduce their taxable income. Taxpayers like B, NB, and FB all contribute to this effect.

behavioral response effect can be written as

$$-t_0 \underbrace{\left[\int_{z'}^{z^\star} (z - z')\, h_0(z)\, \mathrm{d}z + (z^\star - z')\left(H_1(z') - H_0(z^\star)\right) + \int_{z'}^{z^\star} (z^\star - z)\, h_1(z)\, \mathrm{d}z\right]}_{=\text{ area of red region in figure 4}}. \tag{10}$$

The mechanical and behavioral response effect of the reform are depicted visually in figure 4. As that figure shows, identification of these effects would require observing H_1—the CDF of taxable income under the linear tax rate t_1—for all incomes within the interval $(z', z^\star)$. However, we only observe H_1 for incomes at or above the original kink point $z^\star$. Thus, in general, these effects are not identified.

Figure 5: Revenue Effects of a "Small" Reform



Notes: For an infinitesimal change in the location of the tax bracket threshold, the blue and red regions from figure 4 simplify to the thin rectangles depicted in this figure.

However, in the limiting case of a "small" (infinitesimal) reform, these effects are identified. This is shown visually in figure 4. As $z^\star - z' \to \mathrm{d}z^\star$, equation (9) reduces to

$$(t_1 - t_0)(1 - H_1(z^\star))\, \mathrm{d}z^\star$$

and equation (9) reduces to

$$-t_0 (H_1(z^\star) - H_0(z^\star))\, \mathrm{d}z^\star.$$

Notice that the fraction of taxpayers with incomes above the kink $(1 - H_1(z^\star))$ and the discontinuity of the CDF of taxable income at the kink $(H_1(z^\star) - H_0(z^\star))$ are features of the observed distribution of taxable income. This leads to the key identification result of this paper.

Theorem 1 (Sufficiency of the Bunching Mass)**.** *Under assumptions 1, 2, and 3, if the tax schedule is piecewise linear with two tax brackets (as defined in equation 2), then the first-order revenue effect of decreasing the kink point $z^\star$ is*

$$-R'(z^\star) = \underbrace{(t_1 - t_0)(1 - H_1(z^\star))}_{\textit{mechanical effect}} \underbrace{-t_0 \overbrace{(H_1(z^\star) - H_0(z^\star))}^{\textit{bunching mass}}}_{\textit{behavioral response effect}}. \tag{11}$$

The bunching mass is a sufficient statistic for the revenue impact of the behavioral response to the reform. The probability of locating above the kink is a sufficient statistic for the mechanical revenue effects of the reform.

2 Evaluation without Elasticities

In this section, I build on Theorem 1 to demonstrate how bunching designs can be used to evaluate proposed reforms to progressive piecewise linear tax schedules. I discuss how Theorem 1 can be used to estimate the revenue impact of a prospective tax reform and compare this approach to existing techniques for predicting the revenue effects of tax reforms, showing that these techniques do not correctly account for the effect of changing tax brackets. Following this discussion, I consider the application of Theorem 1 to welfare analysis. First, I present a novel test for the Pareto efficiency of an observed tax schedule. This amounts to a test of whether a given tax schedule is locally on the wrong side of the Laffer curve, so that a local tax increase (implemented by decreasing a tax bracket threshold) would actually decrease revenue. Next, I characterize the welfare effects of this reform according to standard welfarist objective functions.

2.1 Revenue Forecasting

Consider an economy facing the tax schedule from two-bracket tax schedule (2). Let $(t_0, t_1, z^\star)$ be the parameters of the current tax schedule and consider a reform which changes these parameters to the new values $(t_0 + \Delta t_0, t_1 + \Delta t, z^\star - \Delta z^\star)$. Suppose—without loss of generality—that $\Delta z^\star > 0$ so that this reform results in a lower bracket threshold. One way to estimate the revenue effect of this reform is to use a first-order Taylor approximation. The approximation can then be written as

$$\Delta R \approx \underbrace{-\left(t_1 - t_0\right)\left(1 - H_1\left(z^\star\right)\right)\Delta z^\star}_{\substack{\text{mechanical effect of} \\ \text{threshold change}}} + \underbrace{t_0\left(H_1\left(z^\star\right) - H_0\left(z^\star\right)\right)\Delta z^\star}_{\substack{\text{behavioral response effect} \\ \text{of threshold change}}} + \underbrace{\frac{\partial R}{\partial t_0}\Delta t_0 + \frac{\partial R}{\partial t_1}\Delta t_1}_{\text{rate change effects}} \tag{12}$$

Letting $\varepsilon^c \equiv -\frac{1-t}{z}\frac{\partial z}{\partial(1-t)}$ denote the elasticity of taxable income (ETI), the partial effect of changing the lower bracket tax rate is

$$\frac{\partial R}{\partial t_0} = \underbrace{\mathbb{E}\left[z|z<z^\star\right] H_0\left(z^\star\right) + z^\star\left(1 - H_0\left(z^\star\right)\right)}_{\text{mechanical effect}} \underbrace{-\frac{t_0}{1-t_0}\mathbb{E}\left[z\varepsilon^c|z<z^\star\right] H_0\left(z^\star\right)}_{\text{behavioral response effect}}, \tag{13}$$

and the partial effect of changing the upper bracket tax rate is

$$\frac{\partial R}{\partial t_1} = \underbrace{\mathbb{E}\left[z - z^\star|z>z^\star\right]\left(1 - H_1\left(z^\star\right)\right)}_{\text{mechanical effect}} \underbrace{-\frac{t_1}{1-t_1}\mathbb{E}\left[z\varepsilon^c|z>z^\star\right]\left(1 - H_1\left(z^\star\right)\right)}_{\text{behavioral response effect}}. \tag{14}$$

Equation (12) makes clear that not every aspect of tax reform evaluation requires knowledge of taxpayer elasticities. The part of the reform effect which is caused by tax bracket changes can be estimated relying only on features of the pre-reform distribution of taxable income. On the other hand, equation (12) also shows that predicting the revenue effect of changing rates within a bracket does require having some knowledge of the taxpayer ETIs, as equations (13) and (14) contain terms which depend on these elasticities.

To understand the value of this insight, it is helpful to compare equation (12) to current practices for predicting the effect of tax policy reforms. Analysts in some government agencies and think tanks have adopted the practice of estimating the behavioral response effect of any proposed tax reform using an elasticity-based approach, as follows:[18]

$$\text{behavioral response effect} \approx -\sum_i \Delta EMTR_i \cdot \frac{EMTR_i}{1 - EMTR_i} \cdot z_i \cdot \varepsilon_i^c$$

where $EMTR_i$ is the effective marginal tax rate of taxpayer i, $\Delta EMTR_i$ is the change in their effective marginal rate induced by the reform, z_i is pre-reform taxable income, and ε_i^c is the assumed ETI of taxpayer i. Applying this approach to approximate the revenue effect of a tax reform would yield the following expression

[18]For example, the Government of Canada's Parliamentary Budget Office appears to follow this practice when estimating revenue effects of proposed tax reforms, including reforms that induce bracket changes (http://www.pbo-dpb.gc.ca/web/default/files/files/files/Fiscal_Impact_and_Incidence_EN.pdf). More generally, I have found that organizations that report the results of similar revenue forecasting exercises do not provide precise information about the methodological approach used, though many such reports provide vague descriptions of "elasticity-based estimation".

$$\Delta R \approx \underbrace{\frac{\partial R}{\partial t_0}\Delta t_0 + \frac{\partial R}{\partial t_1}\Delta t_1}_{\text{rate change effects}} \quad \underbrace{-\left(t_1 - t_0\right)\left(1 - H_1\left(z^\star\right)\right)\Delta z^\star}_{\text{mechanical effect of threshold change}}$$

$$+ \underbrace{\frac{t_0}{1-t_0}\left(t_1 - t_0\right)\mathbb{E}\left[z\varepsilon^c | z \in \left[z^\star - \Delta z^\star, z^\star\right]\right]\left(H_1\left(z^\star\right) - H_0\left(z^\star - \Delta z^\star\right)\right)}_{\text{behavioral response effect of threshold change}} \quad (15)$$

The elasticity-based approach approximates the effect of rate changes the same way equation (12) does, but differs markedly in how it accounts for the impact of bracket changes. The elasticity-based approach treats bracket changes as simply another way of changing marginal tax rates over some range of income, and attempts to use elasticities to approximate the behavioral response to these rate changes.[19]

My results show that this reliance on elasticities is unnecessary. Given the internal and external validity challenges associated with ETI estimates, this finding is potentially of considerable policy-relevance. Moreover, the elasticity-based approximation of the behavioral response effect of a bracket change is biased. Even if the correct elasticities were plugged into equation (15), the resulting prediction of the behavioral response effect of the bracket change should not be expected to coincide with a prediction based on equation (12).[20]

An important caveat to the application of equation (12) is that it is only a first-order approximation, and so its predictions about the impact of discrete policy changes will have errors. Appendix A discusses these errors in detail, as well as possible methods for improving on the first-order approximation. For example, higher-order Taylor approximations of the revenue effects of moving a bracket threshold are also identified by the observed distribution of taxable income.[21] Appendix A also discusses the possibility of a partial identification approach, suggesting one way to bound discrete reform revenue effects. Finally, in the empirical application in section 3.3, I present estimated

[19]In particular, this approach estimates the behavioral responses of taxpayers by first calculating how the reform will alter the marginal tax rates of taxpayers who fall between the old bracket threshold and the new threshold and then multiplying this change by $\frac{\partial z}{\partial t} = -\frac{z\varepsilon^c}{1-t_0}$. These estimated responses are multiplied by the pre-reform marginal tax rate t_0 to obtain the estimated revenue effect.

[20]To be more precise about the "bias" inherent in the elasticity-based approach of equation (15), note that a valid first-order approximation to the effect of a reform should be exact in the limit of a small reform. This is true for the approximation presented in equation (12) by construction. Consequently, we can assess the validity of the approximation of equation (15) by checking to see whether it delivers the same predicted revenue effect in the limiting case of a small bracket change. This would require that

$$\lim_{\Delta z^\star \to 0} \frac{-\frac{t_0}{1-t_0}\left(t_1 - t_0\right)\mathbb{E}\left[z\varepsilon^c | z \in \left[z^\star - \Delta z^\star, z^\star\right]\right]\left(H_1\left(z^\star\right) - H_0\left(z^\star - \Delta z^\star\right)\right)}{\Delta z^\star} = -t_0\left(H_1\left(z^\star\right) - H_0\left(z^\star\right)\right)$$

However, for any $t_0 \neq 0$, the limit on the left-hand side of the equation above diverges to negative infinity, so this condition cannot hold. Thus, equation (15) is not a valid first-order approximation to the revenue effect of a tax reform that changes tax brackets.

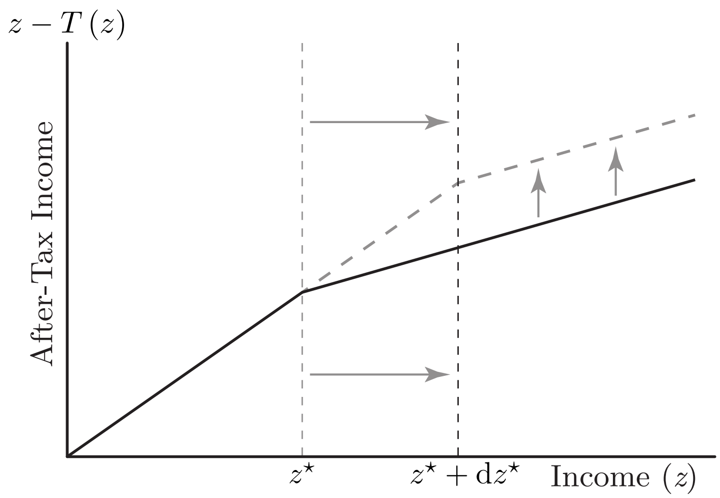
[21]This result parallels some results from the literature on regression discontinuity designs (RDD). The usual parameter of interest in such designs is the average treatment effect at the cutoff (ATEC), which is well known to be nonparametrically identified under standard conditions. Dong and Lewbel (2015) show that partial derivatives of the ATEC with respect to the location of the treatment cutoff are also identified.

revenue effects of a discrete reform based on a simple polynomial extrapolation method.

2.2 Testing Pareto Efficiency

Theorem 1 can also be used be used to derive a simple test for the Pareto efficiency of an observed tax schedule. Consider the tax reform that increases the tax bracket threshold $z^\star$. Figure 6 depicts such a reform visually. This reform expands the budget set and so, by revealed preference, all taxpayers weakly prefer the post-reform tax schedule. If it also causes a net increase in tax revenue, then this would constitute a Pareto improvement.

Figure 6: Infinitesimally Raising a Kink Point



Theorem 2 (Test for Pareto Efficiency)**.** *Under assumptions 1, 2, and 3, then a piecewise linear tax schedule with two tax brackets (as defined in equation 2) which is Pareto efficient must satisfy*

$$R'(z^\star) \leq 0. \tag{16}$$

If $R'(z^\star) = 0$, then such a tax schedule must also satisfy the second-order condition

$$R''(z^\star) = t_1 h_1(z^\star) - t_0 h_0(z^\star) < 0. \tag{17}$$

These conditions are empirically testable using the observed (pre-reform) distribution of taxable income.

Proof. If either $R'(z^\star) > 0$, or if $R'(z^\star) = 0$ and $R''(z^\star) \geq 0$, then there exists some (sufficiently small, but finite) increase in the location of the kink point which increases tax revenue.[22] As noted

[22] In the latter case, this is true because $z^\star$ is a local minimum of $R(\cdot)$.

above, this would constitute a Pareto-improving reform. Theorem 1 shows that $R'(z^{\star})$ can be identified from the observed distribution of taxable income. It emerges that $R''(z^{\star})$ can also be identified because the density terms that appear on the right-hand side of equation (17)—while not directly observable—are limit points of the observed income density.[23] □

Theorem 2's test of Pareto efficiency is an analogue of the test for local Laffer effects initially proposed by Werning (2007). However, no prior implementations of such tests can account for the behavioral response to taxation without committing to a specific model of taxpayer behavior and calibrating model parameters. By contrast, this test accounts for the behavioral response to changing tax bracket thresholds without using any information beyond the observed distribution of taxable income. The robustness of this test comes at the cost of the scope of inquiry, as the test is uninformative about other tax reforms of interest.

In appendix B, I expand on the above Pareto efficiency test, constructing additional testable conditions for efficiency using an analogue of the two-bracket reforms discussed by Bierbrauer, Boyer, and Hansen (2020).

2.3 Welfare Effects of Moving Kinks

The efficiency results above demonstrate that bunching can be used to provide a nonparametric test of the Pareto principle, but efficiency is a very weak normative criterion. It is also of interest to consider welfare effects under a stronger definition of welfare.

The sharpest welfare effect results are obtained by assuming a Rawlsian (revenue-maximizing) objective. A Rawlsian planner has no reason to forgo any revenue gains, irrespective of whether these come at the cost of some agents' private welfare.[24]

[23] Specifically, $h_0(z^{\star})$ can be identified as a limit point of the observed density from below the kink

$$\lim_{z \to z^{\star}-} h(z) = \lim_{z \to z^{\star}} h_0(z) = h_0(z^{\star})$$

and $h_1(z^{\star})$ as a limit point of the observed density from above the kink

$$\lim_{z \to z^{\star}+} h(z) = \lim_{z \to z^{\star}} h_1(z) = h_1(z^{\star}).$$

[24] Here, I am adopting the standard optimal tax theory convention of labeling the maximin objective

$$\max \left\{ \min_{\theta \in \Theta} u(c(\theta), z(\theta); \theta) \right\}$$

as a "Rawlsian" welfare maximization problem. I also adopt the standard assumption that this maximization problem is equivalent to

$$\max \left\{ \min_{\theta \in \Theta} c(\theta) \right\},$$

so that the agent with the lowest level of utility is the agent with the lowest level of consumption. Further assuming that the marginal dollar of tax revenue finances increases in the demogrant G, the social planner's problem reduces to one of revenue maximization.

Theorem 3 (Test for Rawlsian Optimality). *Under assumptions 1, 2, and 3, a piecewise linear tax schedule with two tax brackets (as defined in equation 2) which is Rawlsian-optimal (revenue-maximizing) must satisfy*

$$R'(z^\star) = 0, \tag{18}$$

and

$$R''(z^\star) = t_1 h_1(z^\star) - t_0 h_0(z^\star) < 0. \tag{19}$$

This theorem simply states that the location of a kink point in a revenue-maximizing tax schedule must satisfy standard necessary and sufficient conditions for a local maximum.[25] As in theorem 2, the second-order condition (19) ensures that the kink is not located at a local minimum.

2.3.1 Bergsonian Welfare Functions

Next, consider the case where the planner may have some reason to forgo revenue-increasing tax reforms. For this discussion, it will be helpful to first reframe the finding presented in Theorem 1 in the language of fiscal externalities. The fiscal externality of a given tax reform is the ratio of the behavior response effect to the mechanical effect of the reform. This metric provides an intuitive summary of the efficiency costs of a reform, reflecting the revenue loss due to the behavioral response per dollar raised through the mechanical effect of the reform.

Corollary 1 (Fiscal Externality is Identified). *Under assumptions 1, 2, and 3, if the tax schedule is piecewise linear with two tax brackets (as defined in equation 2), then the fiscal externality (FE) of marginally decreasing the kink point $z^\star$ is*

$$FE(z^\star) \equiv -\frac{t_0 (H_1(z^\star) - H_0(z^\star))}{(t_1 - t_0)(1 - H_1(z^\star))}, \tag{20}$$

and is identified by the observed (pre-reform) distribution of taxable income.

Suppose we adopt a standard, additively separable Bergsonian welfare function as our measure of social welfare in a society with a two bracket tax schedule like that defined in equation (2). That is, suppose that social welfare when the kink point is located at $z^\star$ is

$$W(z^\star) \equiv \frac{\int \omega(v(\theta; z^\star)) \, \mathrm{d}F(\theta)}{\lambda} + R(z^\star), \tag{21}$$

where $v(\theta; z^\star)$ is the indirect utility function of a type θ agent, λ is the marginal social value of government revenue, and $\omega(\cdot)$ is some weakly increasing, weakly convex function of utility.

[25] Note, theorem 3 does not provide a test for whether the kink point is at a globally revenue-maximizing location, as $R(z^\star)$ need not be a strictly concave function in general.

I define the marginal rate of substitution for a type θ agent earning income z as

$$MRS\left(z,\theta\right) \equiv -\frac{\partial u(c,z;\theta)/\partial z}{\partial u(c,z;\theta)/\partial c},$$

where consumption $c \equiv z - T\left(z\right)$. Further, I shall denote the marginal social welfare weight of a type θ agent as

$$g\left(\theta\right) \equiv \frac{\omega'\left(v\left(\theta;z^{\star}\right)\right)}{\lambda}\frac{\partial u\left(c,z;\theta\right)}{\partial c}.$$

The welfare weight reflects the marginal social value of a dollar of private consumption for a type θ agent.

I assume that the marginal dollar of government revenue is used to finance increases in the demogrant (or something with equivalent marginal social value). Thus, given the above definition, the average marginal social welfare weight is

$$\int g\left(\theta\right)\mathrm{d}F\left(\theta\right) = 1.$$

That is to say the marginal social value of a dollar of government revenue is one.[26]

The first-order welfare effect of decreasing the location of the kink point can then be written as

$$-W'\left(z^{\star}\right) = \underbrace{\left(t_1 - t_0\right)\int_{z^{\star}}^{\infty}\left(1-\bar{g}\left(z\right)\right)h_1\left(z\right)\mathrm{d}z}_{\text{mechanical effect}} \quad \underbrace{-t_0\left(H_1\left(z^{\star}\right)-H_0\left(z^{\star}\right)\right)}_{\text{revenue effect of behavioral response}}$$
$$\underbrace{-\int_{\{\theta:z_1(\theta)<z^{\star}<z_0(\theta)\}} g\left(\theta\right)\left(1-t_0-MRS\left(z^{\star},\theta\right)\right)\mathrm{d}F\left(\theta\right)}_{\text{utility effect of behavioral response}} \quad (22)$$

where $\bar{g}\left(z\right) \equiv \mathbb{E}\left[g\left(\theta\right)|z\left(\theta\right)=z\right]$ is the average marginal social welfare weight of agents with an income of z under the current tax schedule.

Notice that the first-order welfare effect of decreasing the location of a kink point differs from most similar expressions found in the optimal tax literature, because the behavioral response to this reform has a first-order effect on welfare for individuals in the bunching mass. To understand why, recall that the absence of such first-order welfare effects depends critically on the envelope theorem. But for bunchers, the envelope condition need not be satisfied.[27] In particular, for at least some types of agents who choose to bunch, we have

$$1 - t_0 - MRS\left(z^{\star},\theta\right) > 0. \quad (23)$$

Thus, reducing the location of the kink point has a first-order welfare effect because it has a

[26] The marginal cost of increasing the demogrant is 1 and the marginal social benefit is $\int g\left(\theta\right)\mathrm{d}F\left(\theta\right)$. They must be equalized in the optimal tax system.

[27] Figure 1 shows this visually. The marginal cost curve of a typical buncher does not intersect the marginal benefit function; rather, it runs through the discontinuity in this function at $z^{\star}$. Thus, their marginal benefit of earning exceeds their marginal cost at $z^{\star}$: the envelope condition fails.

first-order effect on the private welfare of bunchers, decreasing utility by

$$\left(1 - t_0 - MRS\left(z^\star, \theta\right)\right) \frac{\partial u}{\partial c}$$

for any type θ who is currently bunching and continues to bunch following the reform. The third term of equation (22) multiplies this utility effect by $G'\left(V\left(\theta; z^\star\right)\right)$ and integrates over all agents who are bunching to obtain the total first-order welfare effect.[28]

The exception to the general failure of the envelope theorem are the so-called *marginal bunchers* that Saez (2010) used to motivate his original bunching method. These are taxpayers who choose to locate at $z^\star$ under both the observed tax schedule $\left(z\left(\theta\right) = z^\star\right)$ and the counterfactual linear tax schedule with tax rate t_0 $\left(z_0\left(\theta\right) = z^\star\right)$. For such individuals, the envelope condition holds

$$1 - t_0 - MRS\left(z^\star, \theta\right) = 0.$$

Notice that these marginal bunchers can also be thought of as representing the *new bunchers* who will enter the bunching mass following an infinitesimal decrease in the location of the kink point. This explains why equation (22) does not include any terms accounting for the welfare implications of the *new buncher effect*: for an infinitesimal change, there are no such effects to account for.

By contrast, the *former buncher effect* does have welfare implications, and these are accounted for in equation (22). Former bunchers represent a second type of marginal buncher. For an infinitesimal reform, the former bunchers are taxpayers who choose to locate at $z^\star$ under both the observed tax schedule $\left(z\left(\theta\right) = z^\star\right)$ and the counterfactual linear tax schedule with tax rate t_1 $\left(z_1\left(\theta\right) = z^\star\right)$. This implies that the indifference curves of the former bunchers are tangent to the slope of the higher tax bracket before the reform,

$$1 - t_1 - MRS\left(z^\star, \theta\right) = 0.$$

Thus, the first-order effect of the reform on their utility is the same as for taxpayers located above $z^\star$:

$$\left(1 - t_0 - MRS\left(z^\star, \theta\right)\right) \frac{\partial u}{\partial c} = \left(t_1 - t_0\right) \frac{\partial u}{\partial c}.$$

Building on these results, below I present an empirical test of welfarist optimality of the observed tax schedule given a known function describing the average marginal social welfare weight at each income level $\bar{g}\left(\cdot\right)$. Before doing so, I will introduce one additional piece of helpful notation. Let

$$\begin{aligned} \hat{g}_+\left(z\right) &\equiv \mathbb{E}\left[g\left(\theta\right) | z\left(\theta\right) > z\right] \\ &= \int_z^\infty \bar{g}\left(z\right) h\left(z\right) \mathrm{d}z \end{aligned}$$

be the average marginal social welfare weight of taxpayers with income above the kink point.

[28]In the optimal tax theory literature, this type of first-order welfare effect usually only appears in the presence of some kind of market failure. For example, externalities, labor market frictions, and behavioral biases can induce first-order welfare effects to behavioral responses to tax reforms.

Theorem 4 (Conditions for Welfarist-Optimality)**.** *Under assumptions 1, 2, and 3, a piecewise linear tax schedule with two tax brackets (as defined in equation 2) which is welfare-maximizing according to (21) must satisfy*

$$\underbrace{1-\hat{g}_{+}\left(z^{\star}\right)}_{\textit{mechanical welfare effect}}+\underbrace{\left(1+\frac{t_1-t_0}{t_0}k\bar{g}\left(z^{\star}\right)\right)FE\left(z^{\star}\right)}_{\textit{behavioral response welfare effect}}=0 \tag{24}$$

where

$$k\equiv\frac{\mathbb{E}\left[g\left(\theta\right)\left(1-t_0-MRS\left(z^{\star},\theta\right)\right)|z\left(\theta\right)=z^{\star}\right]}{\bar{g}\left(z^{\star}\right)\left(t_1-t_0\right)} \tag{25}$$

is the ratio of the true first-order welfare effect of the reform caused by the behavioral response of bunchers relative to the upper bound of this effect: $\bar{g}\left(z^{\star}\right)\left(t_1-t_0\right)$.
If the left-hand side of equation (4) is positive (negative), then there exists a welfare-improving decrease (increase) of the kink point.

To gain some intuition about Theorem 4, consider what equation (24) would look like if the welfare of the bunchers was irrelevant: $\bar{g}\left(z^{\star}\right)=0$. In this case, the condition simply states that at the optimum the fiscal externality of the reform must be equal to the average gain in social welfare caused by the mechanical effect of the reform. This average gain is the difference between the marginal social value of a dollar of government revenue[29] and the average marginal social value of a dollar of private consumption for the agents who are impacted by the mechanical effects (which is $\hat{g}_{+}\left(z^{\star}\right)$). Thus, unlike in the Rawlsian case, at the optimum the social planner will sometimes choose to forgo feasible revenue-increasing tax reforms because the redistributive benefits of the reform may be insufficient to justify the welfare loss causes by the behavioral response to the reform (as measured by the fiscal externality).

Returning to the general case, where $\bar{g}\left(z^{\star}\right)\neq 0$, the fact that moving a kink point causes behavioral responses which have a first-order welfare effect simply inflates the welfare implications of the fiscal externality. This reflects the fact that the welfare loss caused by the behavioral response includes both lost revenue *and* these first-order welfare impacts.

In general, the value of k is unidentified, so theorem 4 cannot be used as the basis for an empirical test of the optimality of a tax schedule. However, given known values for the average welfare weight parameters $\hat{g}_{+}\left(z^{\star}\right)$ and $\bar{g}\left(z^{\star}\right)$, we can obtain an empirical test motivated by a partial identification approach to the problem.

[29] Recall that this is equal to one at the optimum.

Corollary 2 (Test for Welfarist-Optimality)**.** *Under assumptions 1, 2, and 3, a piecewise linear tax schedule with two tax brackets (as defined in equation 2) which is welfare-maximizing according to (21) must satisfy*

$$1+\left(1+\left(\frac{t_1-t_0}{t_0}\right)\bar{g}\left(z^\star\right)\right)FE\left(z^\star\right)<\hat{g}_+\left(z^\star\right)<1+FE\left(z^\star\right). \tag{26}$$

On the other hand, if

$$1+\left(1+\left(\frac{t_1-t_0}{t_0}\right)\bar{g}\left(z^\star\right)\right)FE\left(z^\star\right)\geq\hat{g}_+\left(z^\star\right)$$

then there exists a welfare-improving decrease of the kink point. If

$$1+FE\left(z^\star\right)\leq\hat{g}_+\left(z^\star\right)$$

then there exists a welfare-improving increase of the kink point.

This test simply checks to make sure that condition (24) from theorem 4 is satisfied for at least one $k\in(0,1)$. If there is no such value of k, then we can conclude that a welfare-improving movement of the kink point exists. Importantly, a condition (24) is a necessary but insufficient condition for welfare maximization.

3 Empirical Application

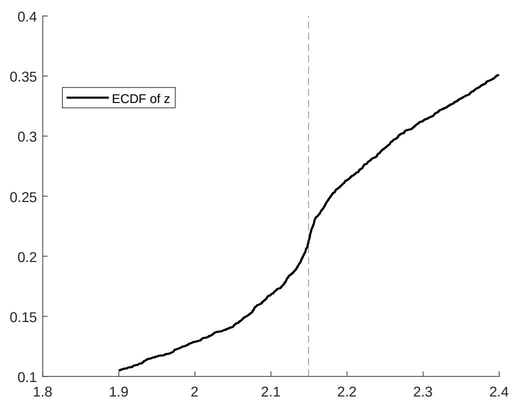
This section applies the ideas discussed above to revisit an empirical application from Saez (2010). Saez applies his bunching methodology to estimate the ETI using a sample of US personal income tax returns: the IRS Individual Public Use Tax Files. I use this same sample to revisit his investigation of bunching at the first kink point of the Earned Income Tax Credit (EITC) schedule between 1995 and 2004.

During this period, the EITC provided a 34% subsidy on the marginal dollar of family earnings in a single child household below \$8580 (in 2008 \$). The marginal subsidy for such a household fell to zero at levels of earnings above these thresholds, introducing a convex kink in their budget constraint. These households also faced a second convex kink point at earnings of \$15740 (in 2008 \$), above which the marginal subsidy became a marginal tax of 16%.

3.1 Filtering the Taxable Income Data

Before we can proceed with applying the results of section 2 to this dataset, we need to address the matter of frictions in taxpayer choices. Recall that the results presented in the preceding sections apply to a scenario where labor supply choices are frictionless. However, real-world taxable income distributions do not support the assumption of a frictionless model. As Saez (2010) noted, some groups of EITC recipients do exhibit evidence of bunching at the first kink point, but this is in the form of a diffuse lump in the density of taxable income at the kink. That is, the CDF of observed taxable income choices does not exhibit a discontinuity at the kink point $z^{\star}$. Figure 7 demonstrates this phenomenon, plotting the empirical CDF of log-income for the subgroup of self-employed, unmarried taxpayers in my sample within a local neighborhood of the kink.

Figure 7: Empirical CDF of Taxable Income



For the purpose of this paper, I assume the observed taxable income of an agent (z) is

$$z \equiv y\xi,$$

a function of both their optimal choice of income (y) and a multiplicative optimization error (ξ). Alternatively, the optimization error is assumed to be additive in log-space. Let $\tilde{x} \equiv \log(x)$ for any variable x, so that the natural log of taxable income is

$$\tilde{z} = \tilde{y} + \tilde{\xi}.$$

Let $\tilde{\xi}$ be independently and identically distributed across taxpayers with a CDF $\tilde{F}_{\xi}$. Furthermore, let $\tilde{H}_y$ be the CDF of $\tilde{y}$, and let $\tilde{H}_z$ be the CDF of $\tilde{z}$.

As Cattaneo, Jansson, Ma, and Slemrod (2018) point out, a the observed distribution of taxable income $(\tilde{H}_z)$ is insufficient to nonparametrically identify the latent bunching mass (the discontinuity in $\tilde{H}_y$ at $\log(z^{\star})$). Following Cattaneo, Jansson, Ma, and Slemrod (2018), I instead adopt a semiparametric identification strategy, assuming that $\tilde{\xi}$ is mean-zero normally distributed with variance σ^2 (that is, $\tilde{F}_{\xi}\left(\tilde{\xi}\right) \equiv \Phi\left(\frac{\tilde{\xi}}{\sigma}\right)$). They show that if the distribution of optimal log-income $(\tilde{H}_y)$ features a non-zero bunching mass at the kink point, then both $\tilde{H}_y$ and σ^2 are semiparametrically identified by $\tilde{H}_z$.[30] For reference, appendix D.1 replicates their proof of this result.

Building on this identification strategy, Cattaneo, Jansson, Ma, and Slemrod (2018) propose a method for estimating $\tilde{H}_y$ and σ^2 based on approximating the latent optimal income density function $\tilde{h}_y$ using histograms. I propose an alternative method of estimation based approximating the latent optimal income CDF ($\tilde{H}_y$) using two flexible polynomial functions. This method is briefly outlined below. For a full description, see appendix D.2.

Let $\tilde{H}_y^0$ be the distribution of $\log(y)$ under a linear tax schedule with tax rate t_0 and let $\tilde{H}_y^1$ be the distribution of $\log(y)$ under a linear tax schedule with tax rate t_1. As shown in section 1.1, under assumption 1 the distribution of $\log(y)$ under the two-bracket tax schedule is

$$\tilde{H}_y(\tilde{y}) \equiv \begin{cases} \tilde{H}_y^0(\tilde{y}) & \text{if } \tilde{y} \leq \log(z^\star) \\ \tilde{H}_y^1(\tilde{y}) & \text{if } \tilde{y} > \log(z^\star) \end{cases}. \tag{27}$$

The observed distribution of log-income is then given by

$$\tilde{H}_z(\tilde{z}) \equiv \int_{\tilde{z}-\log(z^\star)}^{\infty} \tilde{H}_y^0\left(\tilde{z}-\tilde{\xi}\right) \mathrm{d}\Phi\left(\frac{\tilde{\xi}}{\sigma}\right) + \int_{-\infty}^{\tilde{z}-\log(z^\star)} \tilde{H}_y^1\left(\tilde{z}-\tilde{\xi}\right) \mathrm{d}\Phi\left(\frac{\tilde{\xi}}{\sigma}\right).$$

My estimation method approximates $\tilde{H}_y^0(\tilde{y})$ and $\tilde{H}_y^1(\tilde{y})$ in the expression above using 8th-order polynomial functions. Using a sample of taxpayers in a neighborhood of the kink point, I estimate the coefficients of these polynomials via constrained minimization of simulated least squares. See appendix D.2 for complete details.

I apply this filtering method independently for each of four subgroups of EITC eligible taxpayers: unmarried employees, married employees, unmarried self-employed people, and married self-employed people. Figure 8 shows the resulting estimate of the CDF $\tilde{H}_y(\tilde{y})$ for the subgroup of unmarried self-employed taxpayers in my sample. This function is presented alongside the empirical CDF of log-income ($\tilde{H}_z(\tilde{z})$) and a simulated CDF of log-income based on the estimate function $\tilde{H}_y(\tilde{y})$ and the estimated variance of errors σ^2. In this case, the estimated model provides a good fit for the data, with the simulated CDF closely tracking the empirical CDF.

Table 1: Estimated Bunching Mass & Share Above Kink by Subgroup

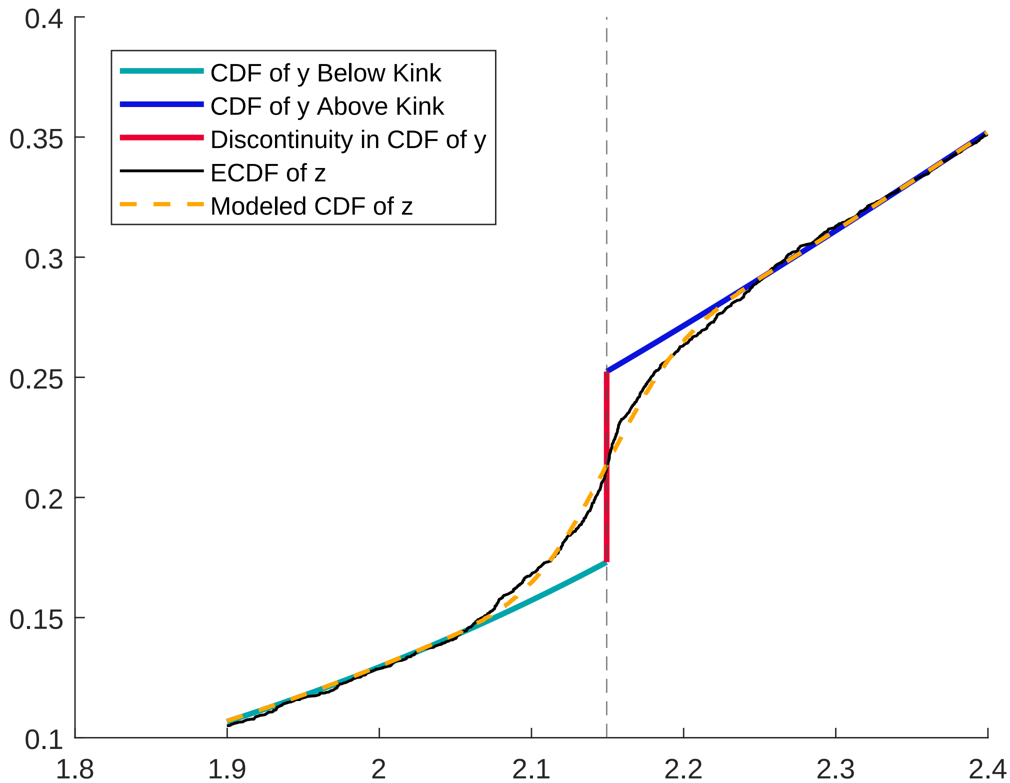
	Unmarried Employees	Unmarried & Self-Employed	Married Employees	Married & Self-Employed
$H_1(z^\star) - H_0(z^\star)$	0.004	0.079	0.002	0.003
$1 - H_1(z^\star)$	0.8543	0.7476	0.9732	0.9513

Table 1 displays estimates of the bunching mass and the share of taxpayers above the kink for each

[30] In fact, Cattaneo, Jansson, Ma, and Slemrod (2018) present this finding under the assumption of normally distributed optimization error that is additive in *levels* rather than in *logs*, but their identification result holds in either case.

of the four subgroups. The estimated size of the bunching mass is less than 1% of taxpayers in three of the four subgroups, but for the sample of unmarried, self-employed taxpayers it is much larger: 7.93% of these taxpayers are estimated to be bunchers. In subsequent sections, I use the estimates from Table 1 as inputs for my analysis of the EITC.

Figure 8: Estimated CDF of Optimal Income Choices



3.2 Accounting for Heterogeneous Tax Schedules

An ideal analysis of the fiscal externality of moving the EITC kink would account for all the ways that the taxable income choices of EITC taxpayers can influence government revenues/expenditure. Thus, the statutory marginal tax rates from the EITC schedule are not the relevant marginal rates for our analysis. Instead, I will employ measures of marginal rates which incorporate state income taxes, payroll taxes, and the phaseout of two major welfare programs (AFDC/TANF and SNAP).

I use the NBER's Taxsim tax calculator to determine the effective marginal tax rate on the wage income for each taxpayer in the dataset. This measure accounts for the impact of other federal tax credits, state tax policies, and FICA. The state income tax calculations include state-level EITC top-up programs which are usually specified as a percentage of federal EITC payments, and therefore have kink points which almost always line up with the federal kink points. I calculate AFDC/TANF payments using a state-level benefit calculator created by Kroft, Kucko, Lehmann, and Schmieder (2020). I use a crude approximation to the impact of SNAP (food stamps) which assumes that a constant phase out rate of 20% applies throughout the relevant range of income, following Kleven (2021).[31]

Importantly, effective marginal tax rates on either side of the first EITC kink may differ substantially across the taxpayers in my analysis, as tax/benefit policy is not constant across states nor within states across years. A valid empirical analysis therefore requires extending the baseline results presented in sections 1 and 2. For simplicity, I will present a two bracket extension, though this can be easily generalized to the multibracket case.

Suppose that type θ agents face the following two bracket tax schedule

$$T\left(z, z^{\star}; \theta\right) \equiv \begin{cases} t_0\left(\theta\right) z & \text{if } z \leq z^{\star} \\ t_1\left(\theta\right) z + \left[t_0\left(\theta\right) - t_1\left(\theta\right)\right] z^{\star} & \text{if } z > z^{\star} \end{cases},$$

where $t_0\left(\theta\right)$, and $t_1\left(\theta\right)$ are type-θ-specific tax schedule parameters.[32] The first-order revenue effect a tax reform which infinitesimally reduces $z^{\star}$ is[33]

$$\begin{aligned} -R'\left(z^{\star}\right) = &\underbrace{\mathbb{E}\left[t_1\left(\theta\right) - t_0\left(\theta\right) | z\left(\theta\right) > z^{\star}\right]\left(1 - H_1\left(z^{\star}\right)\right)}_{\text{mechanical effect}} \\ &\underbrace{-\mathbb{E}\left[t_0\left(\theta\right) | z_1\left(\theta\right) < z^{\star} < z_0\left(\theta\right)\right]\left(H_1\left(z^{\star}\right) - H_0\left(z^{\star}\right)\right)}_{\text{behavioral response effect}}. \end{aligned} \tag{29}$$

Results

The filtering method discussed in the previous section provides estimates of $1 - H_1\left(z^{\star}\right)$ and $H_1\left(z^{\star}\right) - H_0\left(z^{\star}\right)$. I estimate the expected tax rates on either side of the kink point—$\mathbb{E}\left[t_0\left(\theta\right) | z_0\left(\theta\right) = z^{\star}\right]$ and $\mathbb{E}\left[t_1\left(\theta\right) | z_1\left(\theta\right) = z^{\star}\right]$—via local linear regression. These estimates are presented in Table 2 for each of the four subgroups discussed in 3.1. Contrary to what statutory federal income tax rates would imply, effective marginal rates on either side of the EITC kink are positive for the average taxpayer in my sample. This is consistent with the observation made by Bierbrauer, Boyer, and Hansen (2020), who note that the negative marginal rates of the EITC are offset by the large positive marginal rates induced by the phaseout of welfare benefits throughout most the relevant range of income. Only in the case of unmarried, self-employed taxpayers is the marginal tax rate just below the kink negative. The estimated size of the kink is also smaller than what statutory EITC rates would imply for all subgroups.

[31] As in Kleven (2021) I assume a 54% take-up rate for both AFDC/TANF and SNAP.

[32] This result can also be extended to cases where the kink point is a type-specific parameter.

[33] The fiscal externality of this reform can be written as

$$FE \equiv -\frac{\mathbb{E}\left[t_0\left(\theta\right) | z_1\left(\theta\right) < z^{\star} < z_0\left(\theta\right)\right]}{\mathbb{E}\left[t_1\left(\theta\right) - t_0\left(\theta\right) | z\left(\theta\right) > z^{\star}\right]} \frac{H_1\left(z^{\star}\right) - H_0\left(z^{\star}\right)}{1 - H_1\left(z^{\star}\right)}. \tag{28}$$

Table 2: Estimated Local Average Tax Rates by Subgroup

	Unmarried Employees	Unmarried & Self-Employed	Married Employees	Married & Self-Employed
$\mathbb{E}\left[t_0\left(\theta\right)\left|z\left(\theta\right)=z^{\star}\right.\right]$	0.008	-0.007	0.164	0.059
$\mathbb{E}\left[t_1\left(\theta\right)\left|z\left(\theta\right)=z^{\star}\right.\right]$	0.3136	0.283	0.342	0.276

It is important to note, however, that the estimated tax rates in Table 2 cannot be used to estimate equation (29) without imposing an additional simplifying assumption. Let G be denote one of the four subgroups. I assume that for each G,

$$\mathbb{E}\left[t_0\left(\theta\right)\left|z_1\left(\theta\right)<z^{\star}<z_0\left(\theta\right),\theta\in G\right.\right]=\mathbb{E}\left[t_0\left(\theta\right)\left|z_0\left(\theta\right)=z^{\star},\theta\in G\right.\right].$$

I further assume that

$$\mathbb{E}\left[t_0\left(\theta\right)\left|z\left(\theta\right)=z^{\star},\theta\in G\right.\right]=\mathbb{E}\left[t_0\left(\theta\right)\left|z\left(\theta\right)>z^{\star},\theta\in G\right.\right]$$

and

$$\mathbb{E}\left[t_1\left(\theta\right)\left|z\left(\theta\right)=z^{\star},\theta\in G\right.\right]=\mathbb{E}\left[t_1\left(\theta\right)\left|z\left(\theta\right)>z^{\star},\theta\in G\right.\right].$$

That is to say, I assume that within each subgroup the average marginal rate just below the kink for *bunchers* is the same as the observed average marginal rate just below the kink. I also assume that the marginal rate in each bracket for taxpayers who locate near the kink is, on average, the same as those for taxpayers who locate away from the kink point.[34]

Table 3 presents a first-order approximation to the change in average annual revenue that would have resulted from reducing the first EITC kink by $100 (in 2008 $) over the sample period.[35] This effect is broken down into the part due to the mechanical effect of the reform and the part due to the behavioral response effect. The last row of the table also reports the estimated fiscal externality of reducing the location of the kink. Note that Table 3 presents estimated effects for the full group of EITC eligible taxpayers, as well as three subgroups of interest: umarried taxpayers, self-employed taxpayers, and taxpayers who are both unmarried & self employed.[36]

The behavioral response effect is quite small relative to the mechanical effect in every case. This can be seen most clearly in the small fiscal externality estimates, which suggest that for each dollar raised through the mechanical effect of reducing the first EITC kink, about a tenth of a cent would

[34] A more exact approach to estimating equation (29) is possible, as Taxsim could be used to estimate the marginal tax rates on either side of the kink for all taxpayers in the sample. I defer this more comprehensive analysis to future work.

[35] These estimates rely on a variant of the approximation presented in equation (12) which accounts for heterogeneous tax rates following the same approach as equation (29). Plugging in the estimates from Table (2) and setting $\Delta z^{\star}=100$ produces an estimate of the change in average tax revenue per taxpayer per year as a result of the reform. To calculate the total annual revenue effect numbers estimates, I simply scale this estimate by the average number of taxpayers in a per year in the sample.

[36] These are different subgroups than those used for filtering the data and estimating local average tax rates.

Table 3: Revenue Effects of Reducing the First EITC Kink (Single Child Households, 1995-2004)

	All Taxpayers	Unmarried	Self-Employed	Unmarried & Self-Employed
Mechanical Effect (2008 \$)	$38,389,708$	$12,305,422$	$13,497,845$	$2,228,590$
Behav. Response Effect (2008 \$)	$-33,569$	$4,409$	$-3,753$	$5,614$
Total Revenue Effect (2008 \$)	$38,356,139$	$12,309,832$	$13,494,092$	$2,234,204$
Fiscal Externality	-0.0009	0.0004	-0.0003	0.0025

Notes: The revenue effects reported in the first three rows of this table are first-order approximations to the average additional revenue per year that would result from reducing the first EITC kink by \$100 (2008 \$) based on equation (8).

be lost to this behavioral response. Note as well that the behavioral response effect is actually positive for unmarried taxpayers. and those taxpayers who are both unmarried and self-employed. This happens because within these subgroups the average marginal tax rate just below the kink is negative. Consequently, reducing the location of the first EITC kink causes both a mechanical increase in revenue from all taxpayers above the kink who now receive a lower total subsidy amount and an increase in revenue due to the behavioral response effect because moving the bunching taxpayers to a lower taxable income reduces the subsidy they receive.

Theorem 2 provided conditions which must be satisfied in order for the observed tax schedule to be plausibly welfare-optimal (ignoring income effects and extensive margin responses). One of these conditions states that $1 + FE\left(z^{\star}\right)$ provides an upper bound on the average welfare weight above the kink point $\left(\hat{g}_{+}\left(z^{\star}\right)\right)$. Thus, these results imply that we must have $\hat{g}_{+}\left(z^{\star}\right) < 1$ to rationalize observed tax policy as welfare-maximizing for the full group of EITC eligible taxpayers.

3.3 Discrete Reforms

One possible issue with the results presented in table 3 is that they are based on a first-order approximation to the effect of a discrete reform that reduces the kink point by \$100 (2008 \$). As noted in section 1.2, identification of discrete revenue effects of lowering a kink requires knowledge of the unobserved values of $H_1\left(z\right)$ between the old and new kink points. Appendix A discusses this issue at length, as well as the similar issues that arise for a discrete reform which increases the kink point.

One way to more precisely estimate of revenue effects is to estimate the unobserved values of $H_1\left(z\right)$ using the polynomial approximation obtained via the filtering procedure discussed in section 3.1. Table 4 presents estimates based on this extrapolation method. The estimated revenue effects are quite close to those presented in table 3. However, is important to note that while the filtering procedure simply relies on using flexible polynomials as a way of implementing a semiparametric identification strategy, this extrapolation exercise is not justified by that identification argument.

Appendix A considers alternatives ways to estimate the effect of discrete reforms including estimating

Table 4: Revenue Effects for Discrete Reforms

	All Taxpayers	Unmarried	Self-Employed	Unmarried & Self-Employed
Mechanical Effect (2008 \$)	$38,644,215$	$12,480,466$	$13,625,249$	$2,294,772$
Behav. Response Effect (2008 \$)	$-31,939$	$3,942$	$-1,674$	$5,408$
Total Revenue Effect (2008 \$)	$38,612,277$	$12,484,408$	$13,623,575$	$2,300,180$

Notes: The revenue effects reported in the first three rows of this table estimate the exact effect on average annual revenue that would result from reducing the first EITC kink by \$100 (2008 \$), based on extrapolations of the unobserved CDF $\tilde{H}_y^1(\tilde{y})$.

higher-order Taylor approximations of the revenue change or estimating bounds on the effect using a partial identification strategy.

3.4 Incorporating Optimization Error into Policy Analysis

In the prior sections, I used the estimated distribution of optimal income obtained in section 3.1 to estimate revenue effects of tax reforms. However, it must be noted that these estimated effects pertain to changes the hypothetical tax revenue that would result from taxing these unobserved optimal income choices. Because actual taxes are levied on observed taxable income, which includes the effect of any optimization error, these estimated effects may be substantially biased.

As an explanation, note that when optimization error is present, bunching taxpayers are not exclusively located at the kink point; rather, they are scattered around at many different levels of taxable income, with some taxpayers locating in the lower bracket and others in the top bracket. Moving a kink point thus affects the tax revenue obtained from different bunchers in different ways. Obtaining the true behavioral response effect of the reform requires weighting the behavioral responses of bunchers above and below the kink by their corresponding marginal tax rates.

In appendix C.3 I show that, in the model of optimization error described in section 3.1, the first-order revenue effect of reducing a kink point is

$$
\begin{aligned}
-R'(z^\star) = &\underbrace{(t_1 - t_0)\left(1 - H_z(z^\star)\right)}_{\text{mechanical effect}} \\
&- \underbrace{\overbrace{\left(t_0 F_\xi(1)\,\mathbb{E}[\xi|\xi \le 1] + t_1\left(1 - F_\xi(1)\right)\mathbb{E}[\xi|\xi > 1]\right)}^{\text{effective marginal tax rate}} \overbrace{\left(H_y(z^\star; t_1) - H_y(z^\star; t_0)\right)}^{\text{bunching mass}}}_{\text{behavioral response effect}} \qquad (30)
\end{aligned}
$$

where $H_y(z;t)$ is the distribution of optimal income under the tax rate t, $H_z(z)$ is the observed distribution of taxable income, and $F_\xi(1)$ is the fraction of taxpayers induced to work below their optimum point by optimization errors.

There are two important differences between this version of the revenue effect and the version based on frictionless income choices. First, the share of taxpayers above the kink in the frictionless distribution is higher than in the observed distribution:

$$1 - H_y\left(z^\star; t_1\right) < 1 - H_z\left(z^\star\right).$$

Consequently, failing to explicitly account for optimization errors attenuates the mechanical effect of the reform externality. Second, the behavioral response effect also changes because the effective marginal tax rate for bunchers is now

$$t_0 F_\xi\left(1\right) \underbrace{\mathbb{E}\left[\xi | \xi \leq 1\right]}_{\in(0,1]} + t_1\left(1 - F_\xi\left(1\right)\right) \underbrace{\mathbb{E}\left[\xi | \xi > 1\right]}_{>1}.$$

This reflects the variation in the marginal tax rate of bunchers due to the variation in their location. Notice that if $\mathbb{E}\left[\xi\right] = 1$ this expression is just a weighted average of t_0 and t_1 so it will be greater than t_0.

Empirical Results

It is possible to use equation (30) to obtain a first-order approximation of the revenue effect of moving the kink point. Recall that the filtering method introduced in section 3.1 assumes that optimization errors are log-normally distributed. Applying that same assumption here we can infer the components of equation (30).This assumption implies that $F_\xi\left(1\right) = 0.5$ and that estimated variance of optimization errors (σ^2) obtained in that section can be used to estimate $\mathbb{E}\left[\xi | \xi \leq 1\right]$ and $\mathbb{E}\left[\xi | \xi > 1\right]$.[37]. Finally, I use the empirically estimated average marginal tax rates from section 3.2 in place of the fixed tax rates t_0 and t_1 in equation (72).[38]

Table 5 presents approximations of the revenue effects and fiscal externality of reducing the first EITC kink by \$100 (2008 \$) that explicitly account for optimization errors. Relative to the results presented in table 2, the behavioral response effect of the reform is substantially larger for all groups of taxpayers. For example, the fiscal externality estimate for the full set of EITC eligible taxpayers implies that for each dollar of mechanical revenue raised by decreasing the kink point, 0.55 cents are lost to the behavior response effect. While this number remains relatively small, it is five times higher than the estimate from Table 2. For other groups, we also observe larger magnitude estimates than in the frictionless case.

[37]If ξ is log-normally distributed, as I assume in the filtering procedure described in section 3.1, we have

$$\begin{aligned} t_0 F_\xi\left(1\right) \mathbb{E}\left[\xi | \xi \leq 1\right] + t_1\left(1 - F_\xi\left(1\right)\right) \mathbb{E}\left[\xi | \xi > 1\right] &= \exp\left\{\frac{\sigma^2}{2}\right\}\left[t_0\left(1 - \Phi\left(\sigma\right)\right) + t_1 \Phi\left(\sigma\right)\right] \\ &> t_0, \end{aligned}$$

so the effective marginal tax rate for the bunchers is higher than the lower bracket rate.

[38]This requires assuming that average marginal tax rates are locally constant on either side of the bracket threshold.

The fiscal externalities for unmarried taxpayers and for those taxpayers who are both unmarried and self-employed have also changed sign relative to table 2: they are now negative. The latter group has the largest magnitude fiscal externality: the estimate suggests that for each dollar of mechanical revenue raised from this group, their behavioral response to the reform would result in a 5.6 cent reduction in revenue.

These large differences from the results presented in table 2 primarily result from the fact that, in this application, the effective marginal tax rate of bunchers is always negative and is much larger in magnitude than the marginal tax rate just below the kink (which is relatively low). This effect is especially large because the estimated average marginal rates below the kink are so small for most subgroups.

Table 5: Revenue Effects Incorporating Frictions

	All Taxpayers	Unmarried	Self-Employed	Unmarried & Self-Employed
Mechanical Effect (2008 \$)	$38,607,280$	$12,416,703$	$13,896,830$	$2,489,831$
Behav. Response Effect (2008 \$)	$-211,534$	$-139,976$	$-143,004$	$-116,166$
Total Revenue Effect (2008 \$)	$38,395,745$	$12,276,727$	$13,753,826$	$2,373,665$
Fiscal Externality	-0.0055	-0.0113	-0.0103	-0.0467

Notes: The revenue effects reported in the first three rows of this table are first-order approximations to the average additional revenue per year that would result from reducing the first EITC kink by \$100 (2008 \$) based on equation (30).

3.5 What Doesn't Bunching Identify? Income and Participation Effects

Excluded from the analysis above are two potentially important components of the behavioral response to tax reforms: income effects and labor force participation (extensive margin) responses. The latter are of particular interest, as empirical work on the EITC suggests it may have substantial labor force participation effects. Here, I summarize results presented in appendix C.4 which show how these effects alter the revenue impact and fiscal externality of moving a kink point. Importantly, the bunching mass remains a sufficient statistic for that part of the fiscal externality of the tax reform which is attributable to substitution effects.

3.5.1 Income Effects

Suppose that assumption 3 does not hold because some agents have non-zero income effects. Consider once again the tax reform presented in figure 3 which infinitesimally reduces the location of the kink point $z^{\star}$ and notice that upper bracket taxpayers experience a reduction in their virtual income as a result of the reform. If leisure is a normal good, these taxpayers should be expected to increase their taxable income to partially offset this loss in income. Thus, moving the kink point changes the mapping from the distribution of types into the distribution of taxable income under tax schedule T_1 and, consequently, changes the observed distribution of taxable income above the kink point.

3.5.2 Labor Force Participation Effects

Assumption 1 together with convexity of the tax schedule rules out extensive margin responses by excluding the possibility that an agent's indifference curve may be tangent to the budget constraint at more than one point. One way to relax this assumption is to assume that agents face a fixed cost of labor force participation and that they will only enter the labor force (i.e. have positive income) if the gains from doing so exceed this cost. In this scenario, changes in the location of the kink point can induce workers to enter or exit the labor force by changing the utility they receive if they choose to work.

Returning to the tax reform presented in figure 3, there are two groups of taxpayers who experience such participation effects. First, as noted above, a reduction in $z^\star$ lowers the virtual income associated with the upper tax bracket, decreasing the consumption of taxpayers who choose to enter the labor force and locate in the upper bracket. For taxpayers at the margin of participating—those whose fixed cost of work is exactly offset by the benefits of work—this reduced consumption will induce them to exit the labor force. Second, recall that changing the location of a kink point has a first-order effect on the utility of agents located at the kink (as I discussed in section 2.3). Again, for taxpayers at the margin of participating, this will induce them to exit the labor force. Both these types of participation effects cause a reduction in tax revenue.

3.5.3 Bringing It All Together

Although income and participation effects alters the revenue impact of moving a kink point, the bunching mass still identifies the component of this impact attributable to the substitution effect. In particular, the fiscal externality of moving a kink point becomes

$$FE\left(z^\star\right) = FE_{sub}\left(z^\star\right) + FE_{inc}\left(z^\star\right) + FE_{part}\left(z^\star\right),$$

where

$$FE_{sub}\left(z^\star\right) \equiv -\frac{t_0\left(H_1\left(z^\star\right) - H_0\left(z^\star\right)\right)}{\left(t_1 - t_0\right)\left(1 - H_1\left(z^\star\right)\right)} \leq 0,$$

is the part of fiscal externality caused by the substitution effects of the reform, and is identified by the observed (pre-reform) distribution of taxable income. $FE_{inc}\left(z^\star\right)$ and $FE_{part}\left(z^\star\right)$ are the parts of the fiscal externality caused by income and participation effects (respectively). These terms are defined in detail in Appendix C.4. Importantly, neither is identified by the observed distribution of taxable income. Accounting for these effects in an ex ante evaluation of a proposed tax reform requires relying on external evidence about the probable magnitude of these effects.

Other Extensive Margin Responses Participation effects are not the only extensive margin responses that might be relevant to my analysis of the EITC. Small tax reforms can also cause taxpayers to "jump" from one level of taxable income to another whenever the tax schedule induces a non-convex budget set. Even under assumption 1, such non-convexity introduces the possibility that taxpayer indifference curves can be tangent to the budget constraint at more than one point—that a taxpayer may be at the margin between two distinct choices of positive income. We might therefore be concerned that moving a convex kink in a tax schedule which also includes an non-convex kink can potentially induce such taxpayers to jump from one income level to another. In principle, these effects can be incorporated into the fiscal externality of the reform as I have done above with income and participation effects, but I do not discuss this here. Bergstrom and Dodds (2021) provide an extensive discussion of how these responses can affect optimal nonlinear income taxation.

3.5.4 Application to the EITC

The discussion above shows that the size of the bunching mass is informative about one component of the behavioral response effect of moving the first EITC kink, but it may be important to account for the other aspects of the behavioral response effect. Reducing the kink point might reduce revenue due to taxpayers exiting the labor force at or above the kink point. Income effects would generate an increase in revenue as some taxpayers increase their incomes. It is possible to supplement my empirical analysis to account for these responses using income and participation elasticity estimates derived from the literature. However, correctly accounting for the fiscal externality of participation effects requires more detailed information about the change in total net government payments that occurs when an individual above the kink moves to zero income. Such an analysis is complex and beyond the scope of this paper.[39]

4 What is the Bunching Elasticity?

This section revisits standard bunching methodology in light of the results presented above. Bunching-based elasticity estimates are closely connected to the behavioral response to moving a kink point, but their application to other tax reforms of interest faces external validity challenges. In this section, the conditions that give rise to these challenges are characterized analytically. Several examples are presented which concretely demonstrate how particular behavioral assumptions can compromise the external validity of the bunching elasticity.

[39]Bierbrauer, Boyer, and Hansen (2020) conduct an analysis of the efficiency consequences of the introduction of the EITC which accounts for these effects but confine their analysis to California due to the difficulty associated with conducting a full analysis.

4.1 Relating the Bunching Mass to Elasticities

Discussion of the standard bunching method requires a special case of the modeling framework that I introduced in section 1.1. In particular, suppose that each agent's type consists of a wage rate $w \in \mathbb{R}_{++}$ and a vector of other characteristics $\psi \in \Psi$ where Ψ is a convex (possibly multidimensional) parameter space. That is to say, the agent's type is $\theta = (w, \psi)$ and the type space is $\Theta = \mathbb{R}_{++} \times \Psi$. For convenience, I will refer to a given value of ψ as a *group*.

Let $F(w, \psi)$ represent the joint distribution of wages and group membership. For simplicity, I maintain the assumption that there are no income effects (assumption 3) but the results below can be easily extended to incorporate income effects. Thus, I can define $z(t; w, \psi)$ as the taxable income a type (w, ψ) agent chooses when facing a linear income tax rate $t < 1$ (irrespective of the intercept of their budget constraint). Let $H(z; t)$ be the distribution of taxable income under a tax rate of t, as induced by the function $z(t; w, \psi)$ and the parameter distribution F.

In addition to assumptions 1, 2, and 3, in this section I assume that within each group ψ agent preferences satisfy the single-crossing condition with respect to the wage rate.

Assumption 4 (Single Crossing Condition)**.** *For all types $(w, \psi) \in \mathbb{R}_{++} \times \Psi$ and any tax rate $t < 1$, taxable income is strictly increasing in the wage rate $\frac{\partial z(t;w,\psi)}{\partial w} > 0$.*

With this additional assumption, it is possible to write the bunching mass as a function of compensated elasticities of taxable income. Let

$$\varepsilon^c(t; w, \psi) \equiv -\frac{1-t}{z^\star}\frac{\partial z(t; w, \psi)}{\partial t} \tag{31}$$

denote the compensated elasticity of taxable income for type (w, ψ) agents facing a linear tax rate of t.

Theorem 5 (Bunching Mass in Terms of Elasticities)**.** *Under assumptions 1, 2, 3, and 4, and given a piecewise linear tax schedule with two tax brackets (as defined in equation 2), the bunching mass at the kink point can be written as:*

$$H(z^\star; t_1) - H(z^\star; t_0) = \int_{t_0}^{t_1} \bar{\varepsilon}^c(z^\star; t) \cdot \frac{z^\star h(z^\star; t)}{1-t}\, dt, \tag{32}$$

where $\bar{\varepsilon}^c(z^\star; t)$ is the average (compensated) elasticity of taxable income among agents who would choose to locate at $z^\star$ when facing a counterfactual linear tax schedule with a tax rate of t,

$$\bar{\varepsilon}^c(z^\star; t) \equiv \mathbb{E}\left[\varepsilon^c(t; w, \psi)\,|z(t; w, \psi) = z^\star\right]. \tag{33}$$

Proof. The bunching mass at $z^\star$ can be written as

$$
\begin{aligned}
H\left(z^\star;t_1\right)-H\left(z^\star;t_0\right) &= \int_\psi\left[H\left(z^\star|\psi;t_1\right)-\tilde{H}\left(z^\star|\psi;t_0\right)\right]f_\psi\left(\psi\right)\mathrm{d}\psi \\
&= \int_\psi\int_{t_0}^{t_1}\frac{\partial H\left(z^\star|\psi;t\right)}{\partial t}f_\psi\left(\psi\right)\mathrm{d}\psi
\end{aligned}
$$

where the second equality follows from the fundamental theorem of calculus.

Note that the single crossing condition (assumption 4) ensures that for any linear tax rate t and within any group ψ there exists a wage rate $w^\star\left(t,\psi\right)$ at which group members would choose to locate at the kink point. That is to say, there exists a function $w^\star\left(t,\psi\right)$ satisfying $z\left(t;w^\star\left(t,\psi\right),\psi\right)=z^\star$ for all $t<1$ and $\psi\in\Psi$. Furthermore, the single crossing condition ensures that there is a straightforward connection between the conditional distribution of taxable income and the conditional distribution of wages:

$$
H\left(z\left(t;w,\psi\right)|\psi;t\right)=F\left(w|\psi\right).
$$

Differentiating both sides of this expression with respect to t we obtain

$$
\frac{\partial H\left(z\left(t;w,\psi\right)|\psi;t\right)}{\partial t}=-\frac{\partial z\left(t;w,\psi\right)}{\partial t}h\left(z\left(t;w,\psi\right)|\psi;t\right)
$$

and therefore the bunching mass can be expressed as

$$
\begin{aligned}
H\left(z^\star;t_1\right)-H\left(z^\star;t_0\right) &= -\int_\psi\int_{t_0}^{t_1}\frac{\partial z\left(t;w^\star\left(t,\psi\right),\psi\right)}{\partial t}h\left(z^\star|\psi;t\right)f_\psi\left(\psi\right)\mathrm{d}t\mathrm{d}\psi \\
&= -\int_{t_0}^{t_1}\left\{\int_\psi\frac{\partial z\left(t;w^\star\left(t,\psi\right),\psi\right)}{\partial t}\frac{h\left(z^\star|\psi;t\right)f_\psi\left(\psi\right)}{h\left(z^\star;t\right)}\mathrm{d}\psi\right\}h\left(z^\star;t\right)\mathrm{d}t
\end{aligned}
$$

The result then follows from using equation (31) to simplify the interior integral above. □

Theorem 5 is simply a restatement of a result derived by Blomquist, Newey, Kumar, and Liang (2019) and represents the most general formulation of the relationship between the bunching mass and taxpayer elasticities. Importantly, equation (32) shows that the bunching mass is a function of both agents' elasticities and unobserved densities of taxable income. Thus, the bunching mass alone cannot identify the ETI.

In order to obtain a weighted average of elasticities from an estimate of the bunching mass, it is necessary to additionally an have an estimate of the normalization term

$$
\mathcal{H}\equiv z^\star\int_{t_0}^{t_1}\frac{h\left(z^\star;t\right)}{1-t}\mathrm{d}t. \tag{34}
$$

By dividing the bunching mass by $\mathcal{H}$, it is possible to identify the *bunching elasticity*

$$
\bar{\varepsilon}^c_B\equiv\int_{t_0}^{t_1}\bar{\varepsilon}^c\left(z^\star;t\right)\rho\left(t\right)\mathrm{d}t, \tag{35}
$$

which is a weighted average elasticity which places some positive weight

$$\rho(t) \equiv \frac{\frac{h(z^\star;t)}{1-t}}{\int_{t_0}^{t_1} \frac{h(z^\star;t')}{1-t'}\mathrm{d}t'}, \tag{36}$$

on the local average elasticity at the kink point under a linear tax rate t ($\bar{\varepsilon}^c(z^\star;t)$) for each tax rate $t \in [t_0, t_1]$.

Corollary 3. *Under assumptions 1, 2, 3 and 4, and given a piecewise linear tax schedule with two tax brackets (as defined in equation 2), the bunching mass at the kink point together with the normalizing term $\mathcal{H}$ identifies the bunching elasticity:*

$$\bar{\varepsilon}^c_B = \frac{H(z^\star;t_1) - H(z^\star;t_0)}{\mathcal{H}} \tag{37}$$

Recent work in the bunching literature has called attention to the fact that the observed distribution of taxable income cannot nonparametrically identify the bunching elasticity because it cannot identify the normalization term $\mathcal{H}$ (Bertanha, McCallum, and Seegert, 2021; Blomquist, Newey, Kumar, and Liang, 2019). Saez (2010) and Chetty, Friedman, Olsen, and Pistaferri (2011) proposed identification strategies that relied on approximating $\mathcal{H}$. Blomquist, Newey, Kumar, and Liang (2019) and Bertanha, McCallum, and Seegert (2021) propose alternative approaches to obtaining identification or partial identification of the bunching elasticity. These papers also show that estimated elasticities can be quite sensitive to the choice of identification strategy. All these identification strategies require imposing assumptions about how the density of taxable income at $z^\star$ ($h(z^\star;t)$) varies across different linear tax schedules, and in most cases further relies on the assumption that agents have quasi-linear, isoelastic preferences and a constant elasticity.

Blomquist, Newey, Kumar, and Liang (2019) present a version of Corollary 3, and note that the bunching elasticity ($\bar{\varepsilon}^c_B$) is subject to "issues of external validity" because it provides information about a very specific convex combination of elasticities.[40] In the next subsection, I build on their work by more precisely characterizing these issues.

[40] It is important to note that this weighted average is not, in general, interpretable as the average elasticity of bunchers. For example, in the special case where taxpayers have quasi-linear, isoelastic preferences it can be shown that the bunching elasticity is the inverse of the expected inverse elasticity of the bunchers:

$$\bar{\varepsilon}^c_B = \left(\mathbb{E}\left[\frac{1}{\varepsilon^c} | z_1(\theta) < z^\star < z_0(\theta)\right]\right)^{-1}.$$

Therefore, by Jensen's inequality, this implies that the bunching elasticity is in fact strictly smaller than the average elasticity of bunchers in this case.

4.2 Policy Analysis with the Bunching Elasticity

Corollary 1 already provides one unambiguously valid interpretation of the bunching elasticity. Using equation (37), the fiscal externality of moving a kink point can be rewritten in terms of the bunching elasticity:[41]

$$FE\left(z^{\star}\right) \equiv -\frac{t_0}{t_1 - t_0} \cdot \frac{\mathcal{H}}{1 - H_1\left(z^{\star}\right)} \cdot \bar{\varepsilon}_B^c$$

However, Corollary 1 also makes it clear that decomposing the bunching mass into the bunching elasticity ($\bar{\varepsilon}_B^c$) and the normalizing constant ($\mathcal{H}$) is unnecessary for the purpose of identifying the fiscal externality of moving a kink point. Thus, the value of such a decomposition depends on whether the bunching elasticity is useful for analyzing other tax reforms of interest.

A natural candidate for a tax reform of interest is the effect of an infinitesimal change in the marginal income tax rate either immediately below or immediately above the kink. In particular, consider a tax schedule perturbation in the style of Saez (2001) which infinitesimally increases the marginal tax rate just below $z^{\star}$ and leaves it unchanged everywhere else. Applying results from the literature on the perturbation approach to optimal nonlinear taxation, the fiscal externality of such a reform is[42]

$$FE_t^{-}\left(z^{\star}\right) \equiv -\frac{t_0}{1 - t_0} \cdot \frac{z^{\star} h\left(z^{\star}; t_0\right)}{1 - H\left(z^{\star}; t_0\right)} \cdot \bar{\varepsilon}^c\left(z^{\star}; t_0\right). \tag{38}$$

Similarly, the fiscal externality of a tax reform which infinitesimally increases the marginal tax rate just above $z^{\star}$ is

$$FE_t^{+}\left(z^{\star}\right) \equiv -\frac{t_1}{1 - t_1} \cdot \frac{z^{\star} h\left(z^{\star}; t_1\right)}{1 - H\left(z^{\star}; t_1\right)} \cdot \bar{\varepsilon}^c\left(z^{\star}; t_1\right). \tag{39}$$

The fiscal externalities of these two local tax reforms can be estimated using the average elasticity of two different types of "marginal bunchers". In particular, $FE_t^{-}\left(z^{\star}\right)$ depends on the average elasticity of agents who would locate at the kink $z^{\star}$ under the linear tax rate t_0 ($\bar{\varepsilon}^c\left(z^{\star}; t_0\right)$), and $FE_t^{+}\left(z^{\star}\right)$ depends on the average elasticity of agents who would locate at the kink $z^{\star}$ under the linear tax rate t_1 ($\bar{\varepsilon}^c\left(z^{\star}; t_1\right)$).

Precisely defining two local tax reforms of interest clarifies the elasticities needed to identify their revenue consequences. This sheds light on two key issues: the external validity of the bunching elasticity, and the differences between the revenue effects of moving a kink point and of infinitesimally changing marginal rates near the kink point.

[41] Or, equivalently, the fiscal externality of increasing the tax rate just below the kink point from t_0 to t_1.

[42] These results are usually framed in terms of welfare effects rather than fiscal externalities. The perturbation approach to optimal taxation began with Saez (2001) and has subsequently been formalized and extended to the case of multidimensional heterogeneity by Gerritsen (2016), Golosov, Tsyvinski, and Werquin (2014), and Jacquet and Lehmann (2021). Bergstrom and Dodds (2021) extend these results to reforms of tax schedules which are almost everywhere continuously differentiable (i.e. tax schedules that may contain kink points) so the expressions below follow from their results.

4.2.1 Determinants of External Validity

The definition of the bunching elasticity (equation 35) makes it clear that neither of the two policy-relevant local average elasticities contained in equations (38) and (39) will generally be identified by the standard bunching method. Although the bunching elasticity does assign positive weight to both $\bar{\varepsilon}^c(z^\star; t_0)$ and $\bar{\varepsilon}^c(z^\star; t_1)$, in general, a given value of $\bar{\varepsilon}^c_B$ is consistent with any finite, positive values of $\bar{\varepsilon}^c(z^\star; t_0)$ and $\bar{\varepsilon}^c(z^\star; t_1)$. It is thus important to clarify under what circumstances we should expect $\bar{\varepsilon}^c_B$ to differ substantially from $\bar{\varepsilon}^c(z^\star; t_0)$ and $\bar{\varepsilon}^c(z^\star; t_1)$.

Equation (35) also makes it clear what the source of any such differences would be: endogeneity of the local average ETI at the kink point ($\bar{\varepsilon}^c(z^\star; t)$) to the tax rate. It can be shown that such endogeneity results from two sources:[43]

$$\frac{\mathrm{d}\bar{\varepsilon}^c(z^\star; t)}{\mathrm{d}t} = \underbrace{-\frac{1}{1-t}\mathrm{Cov}\left(\varepsilon^c, (1+\alpha)\,\varepsilon^c | z = z^\star\right)}_{\text{composition effect}} \\ + \underbrace{\mathbb{E}\left[\frac{\partial \varepsilon^c}{\partial t} | z = z^\star\right] + \mathbb{E}\left[\left(2 - \frac{\bar{\varepsilon}^c(z^\star; t)}{\varepsilon^c}\right)\frac{\partial \varepsilon^c}{\partial w}\frac{\partial w^\star(t;\psi)}{\partial t} | z = z^\star\right]}_{\text{misspecification effect}}, \quad (40)$$

where $\alpha(z; t, \psi) \equiv -\left(1 + \frac{zh'(z|\psi;t)}{h(z|\psi;t)}\right)$ is the (second-order) Pareto parameter of the group-ψ-specific distribution of taxable income at a given level of income z under tax rate t.[44] Note that I have suppressed the arguments of various functions in equation (40) for simplicity, and will continue to do so below.[45]

To understand the first term in equation (40), note that, when agents exhibit multidimensional heterogeneity (i.e. when there is more than one group ψ), changing the tax rate influences the relative size of different groups among agents located at a given value of taxable income (in this case, $z^\star$). This generates what Jacquet and Lehmann (2021) label a *composition effect*.[46] We can

[43] See appendix C.5 for the derivation.

[44] Note, for a continuous random variable X with distribution F_X the function

$$\frac{xf_X(x)}{1 - F_X(x)}$$

is sometimes described as the Pareto parameter at x in the public finance literature, because if F_X has a Pareto tail, this measure converges to the Pareto parameter as $x \to \infty$. This claim is also true for the function

$$-\left(1 + \frac{xf'_X(x)}{f_X(x)}\right),$$

and this measure has also sometimes been described in the public finance literature as the "Pareto parameter" (for example, see Hendren (2020)). Because this paper includes references to both these measures, I distinguish the second measure by labeling it as the *second-order Pareto parameter at x*.

[45] In particular, I have employed the convention that $\varepsilon^c \equiv \varepsilon^c(t; w, \psi)$ and $\alpha \equiv \alpha(z; t, \psi)$, and $z \equiv z(t; w, \psi)$.

[46] They discuss the implications of such effects for attempts to use sufficient statistics pertaining to an observed tax system for the purpose of inferring the optimal tax system.

break composition effects down further:

$$\operatorname{Cov}\left(\varepsilon^c,(1+\alpha)\,\varepsilon^c|z=z^\star\right)=\underbrace{\left(1+\mathbb{E}\left[\alpha|z=z^\star\right]\right)\operatorname{Var}\left(\varepsilon^c|z=z^\star\right)}_{\text{elasticity heterogeneity across groups}}+\underbrace{\operatorname{Cov}\left(\varepsilon^c\left(\varepsilon^c-\bar{\varepsilon}^c\left(z^\star;t\right)\right),\alpha|z=z^\star\right)}_{\text{heterogeneity of Pareto parameter across groups}} \tag{41}$$

The first term in this expression captures the fact that, all else equal, greater variance of elasticities among agents at $z^\star$ implies a larger magnitude composition effect. This variance effect is proportional to the average group-specific Pareto parameter at $z^\star$. The second term is non-zero whenever group-specific Pareto parameters are correlated with a particular function of group-specific elasticities. Note that both of these types of composition effects are present if agents have isoelastic preferences so that elasticities are constant within groups; it is variation in elasticities across groups that generate composition effects.

The second term in equation (40) reflects another mechanism through which the tax rate can influence the local average ETI $\bar{\varepsilon}^c\left(z^\star;t\right)$: if agent preferences are not isoelastic. I call this the *misspecification effect*, in reference to the ubiquity of isoelastic function form assumptions in the bunching literature (and in empirical public finance generally). As with the composition effect, the misspecification effect can be broken down into two parts:

$$\underbrace{\mathbb{E}\left[\frac{\partial\varepsilon^c}{\partial t}|z=z^\star\right]}_{\text{direct effect of tax rate}}+\underbrace{\mathbb{E}\left[\left(2-\frac{\bar{\varepsilon}^c\left(z^\star;t\right)}{\varepsilon^c}\right)\frac{\partial\varepsilon^c}{\partial w}\frac{\partial w^\star\left(t;\psi\right)}{\partial t}|z=z^\star\right]}_{\text{elasticity heterogeneity within groups}}. \tag{42}$$

The first term in this expression reflects the fact that when agents do not have isoelastic preferences, the tax rate can directly influence the elasticity of taxable income. The second term will be present whenever agents within the same group can have different elasticities (i.e. the group-specific elasticity varies with the wage rate). The interpretation of this term is somewhat complex; intuitively, it captures the fact that changes in the tax rate will generally change which members of a given group choose to locate at a particular level of taxable income, and so if elasticities vary within groups such changes can generate an indirect effect on the local average ETI. Note that both of these effects would be present even in a model with unidimensional heterogeneity (i.e. if there were only one group ψ).

To further clarify the nature of composition and misspecification effects, let us consider some motivating examples.

Example 1 (Isoelastic Preferences with Constant Pareto Parameter)**.** Suppose that all taxpayers have quasi-linear, isoelastic preferences of the form

$$u\left(z;w,\psi\right)\equiv z-T\left(z\right)-\frac{w}{1+\frac{1}{\psi}}\left(\frac{z}{w}\right)^{1+\frac{1}{\psi}},$$

so that

$$z\left(t;w,\psi\right)=w\left(1-t\right)^{\psi}.$$

With this specification, agents within a given group all have the same elasticity: ψ. Thus, there are no misspecification effects. Further suppose that all groups have wages are distributed according to F, a distribution which is Pareto Type I above some threshold wage w_0. That is to say, for any $w \geq w_0$

$$F_w(w) = F_w(w_0) + (1 - F_w(w_0))\left(1 - \left(\frac{w}{w_0}\right)^{-a}\right)$$

for some Pareto slope parameter $a > 1$. Given these assumptions, it can be shown that for any taxable income $z > w_0$, all groups ψ will have the same group-specific Pareto parameter: $\alpha(z; t, \psi) = a$ for all ψ and $t < 1$. This implies that the second term in equation (41) is zero. The effect of the tax rate on the local average ETI can thus simply be written as

$$\frac{\mathrm{d}\bar{\varepsilon}^c(z^\star; t)}{\mathrm{d}t} = -\frac{1}{1-t} \cdot (1+a) \cdot \mathrm{Var}(\psi | z = z^\star).$$

Example 2 (Isoelastic Preferences with Non-Constant Pareto Parameter)**.** Suppose that all taxpayers have quasi-linear, isoelastic preferences but of a slightly different form than that discussed above:

$$u(z; w, \psi) \equiv z - T(z) - \frac{1}{1+\frac{1}{\psi}}\left(\frac{z}{w}\right)^{1+\frac{1}{\psi}}.$$

This specification implies a different taxable income function

$$z(t; w, \psi) = w^{1+\psi}(1-t)^{\psi}.$$

As in the previous example, within a given group all agents have the same ETI (ψ) and so there are no misspecification effects. Again, suppose that wages are identically distributed across groups according to F, a distribution which is Pareto Type I with slope a at any wage above w_0. It can be shown that for any taxable income $z > w_0$, different groups have different local Pareto parameters. In particular, $\alpha(z; t, \psi) = \frac{a}{1+\psi}$ for all ψ and $t < 1$. This implies that the second term in equation (41) is non-zero. The effect of the tax rate on the local average ETI can thus simply be written as

$$\frac{\mathrm{d}\bar{\varepsilon}^c(z^\star; t)}{\mathrm{d}t} = -\frac{1}{1-t}\left[\mathrm{Var}(\psi | z = z^\star) + a\mathrm{Cov}\left(\psi, \frac{\psi}{1+\psi} | z = z^\star\right)\right].$$

Example 3 (Unidimensional Heterogeneity with Multiple Incomes)**.** Suppose that all taxpayers have two sources of labor income (z_1 and z_2) and that their total taxable income is simply the sum of these two incomes. For simplicity, I assume that both labor activities have the same wage rate. Additionally, suppose that there is only one group ψ, so that agents differ only in their wage rate w. Under this assumption, there are no composition effects but there is a misspecification effect. Finally, suppose that agents have a separable utility function, with the disutility of income for each source taking a quasi-linear, isoelastic form. In particular, suppose that an agent with wage rate w

chooses values of z_1 and z_2 to solve the problem

$$\max_{z_1,z_2>0}\left\{z_1+z_2-T\left(z_1+z_2\right)-\frac{w}{1+\frac{1}{e_1}}\left(\frac{z_1}{w}\right)^{1+\frac{1}{e_1}}-\frac{w}{1+\frac{1}{e_2}}\left(\frac{z_2}{w}\right)^{1+\frac{1}{e_2}}\right\}.$$

Such an agent facing a linear tax rate t will choose to generate $z_1\left(t;w\right)\equiv w\left(1-t\right)^{e_1}$ dollars of income from the first source and $z_2\left(t;w\right)\equiv w\left(1-t\right)^{e_2}$ dollars from the second source. Total taxable income for such an agent is

$$\begin{aligned} z\left(t;w\right) &\equiv z_1\left(t;w\right)+z_2\left(t;w\right)\\ &= w\left[\left(1-t\right)^{e_1}+\left(1-t\right)^{e_2}\right]. \end{aligned}$$

Notice that the elasticities of z_1 and z_2 with respect to the net-of-tax rate are e_1 and e_2. These elasticities are constant across all agents and across different tax rates. However, the elasticity of taxable income is a convex combination of the elasticities of each of the two income sources

$$\begin{aligned} \varepsilon^c\left(t\right) &\equiv \frac{z_1\left(t;w\right)}{z\left(t;w\right)}\cdot e_1+\frac{z_2\left(t;w\right)}{z\left(t;w\right)}\cdot e_2\\ &= \left(\frac{\left(1-t\right)^{e_1}}{\left(1-t\right)^{e_1}+\left(1-t\right)^{e_2}}\right)e_1+\left(\frac{\left(1-t\right)^{e_2}}{\left(1-t\right)^{e_1}+\left(1-t\right)^{e_2}}\right)e_2, \end{aligned}$$

where the weights on the elasticity of each income source correspond to the share of the income source in the agent's total income. As shown above, these income shares are independent of taxpayer wage rate, so the misspecification effect in this example is entirely a direct effect of taxation on this elasticity. This direct effect stems from the fact that, as long as $e_1\neq e_2$, a change in the tax rate changes the share of each type of income in total taxable income. Specifically, the misspecification effect is

$$\frac{\mathrm{d}\varepsilon^c\left(t\right)}{\mathrm{d}t}=-\frac{\left(e_1-e_2\right)^2}{1-t}\cdot\frac{z_1\left(t;w\right)}{z\left(t;w\right)}\cdot\frac{z_2\left(t;w\right)}{z\left(t;w\right)}.$$

4.2.2 Infinitesimal Rate Changes vs. Moving Kinks

The external validity concerns associated with use of the bunching elasticity might seem unimportant in light of the earlier results of this paper. As noted above, moving a kink point provides a mechanism for implementing a specific local rate change. However, the impact of moving a kink point may differ in important ways from the impact of other local tax reforms of interest. By comparing (38) and (39) with the fiscal externality of moving a kink point, we can gain insight into these differences.

Inserting equation (32) into equation (20), we obtain an expression for the fiscal externality of moving a kink point as a function of the average elasticity of taxable income at $z^\star$ under various counterfactual linear tax rates between t_0 and t_1:

$$FE\left(z^{\star}\right)=-\frac{t_0\int_{t_0}^{t_1}\frac{z^{\star}h\left(z^{\star};t\right)}{1-t}\bar{\varepsilon}^c\left(z^{\star};t\right)\mathrm{d}t}{\left(t_1-t_0\right)\left(1-H\left(z^{\star};t_1\right)\right)} \tag{43}$$

This expression bears some superficial similarities in structure to the expressions describing the fiscal externalities of the other local tax reforms introduced above (equations 38 and 39). To make the connection more precise, we can rewrite $FE\left(z^{\star}\right)$ in terms of the fiscal externalities of these other local reforms. In each case, the fiscal externality of moving a kink point can be written as the sum of a re-scaled version of the fiscal externality of the local tax reform and an adjustment term:

$$FE\left(z^{\star}\right)=\left(\frac{1-H\left(z^{\star};t_0\right)}{1-H\left(z^{\star};t_1\right)}\right)FE_t^{-}\left(z^{\star}\right)-\frac{t_0z^{\star}\int_{t_0}^{t_1}\int_{t_0}^{t}\left\{\frac{\mathrm{d}}{\mathrm{d}t'}\left[\frac{h\left(z^{\star};t'\right)}{1-t'}\bar{\varepsilon}^c\left(z^{\star};t'\right)\right]\right\}\mathrm{d}t'\mathrm{d}t}{\left(t_1-t_0\right)\left(1-H\left(z^{\star};t_1\right)\right)}, \tag{44}$$

and

$$FE\left(z^{\star}\right)=\left(\frac{t_0}{t_1}\right)FE_t^{+}\left(z^{\star}\right)+\frac{t_0z^{\star}\int_{t_0}^{t_1}\int_{t}^{t_1}\left\{\frac{\mathrm{d}}{\mathrm{d}t'}\left[\frac{h\left(z^{\star};t'\right)}{1-t'}\bar{\varepsilon}^c\left(z^{\star};t'\right)\right]\right\}\mathrm{d}t'\mathrm{d}t}{\left(t_1-t_0\right)\left(1-H\left(z^{\star};t_1\right)\right)}. \tag{45}$$

Note that the adjustment terms depend on how $\frac{h(z^{\star};t)}{1-t}\bar{\varepsilon}^c\left(z^{\star};t\right)$ varies over the interval $[t_0,t_1]$. Evaluating the limits of equations (44) and (45) as $t_1\to t_0$, we can see that the fiscal externalities of all three reforms converge for an arbitrarily small kink size. But for a finite sized kink, they will differ from each other.

To better understand these differences, consider the adjustment terms in equations (44) and (45). Note that the integrand in these terms can be decomposed into three effects:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{h\left(z^{\star};t\right)}{1-t}\bar{\varepsilon}^c\left(z^{\star};t\right)\right]&=\underbrace{\frac{h\left(z^{\star};t\right)}{\left(1-t\right)^2}\left(\bar{\varepsilon}^c\left(z^{\star};t\right)-\left(1+\mathbb{E}\left[\alpha|z=z^{\star}\right]\right)\bar{\varepsilon}^c\left(z^{\star};t\right)^2\right)}_{\text{baseline effect}}\\&\quad\underbrace{-\frac{h\left(z^{\star};t\right)}{\left(1-t\right)^2}\left(\left(1+\mathbb{E}\left[\alpha|z=z^{\star}\right]\right)\mathrm{Var}\left(\varepsilon^c|z=z^{\star}\right)+\mathrm{Cov}\left(\left(\varepsilon^c\right)^2,\alpha|z=z^{\star}\right)\right)}_{\text{composition effect}}\\&\quad+\underbrace{\frac{h\left(z^{\star};t\right)}{1-t}\left(\mathbb{E}\left[\frac{\partial\varepsilon^c}{\partial t}|z=z^{\star}\right]+2\mathbb{E}\left[\frac{\partial\varepsilon^c}{\partial w}\frac{\partial w^{\star}\left(t;\psi\right)}{\partial t}|z=z^{\star}\right]\right)}_{\text{misspecification effect}}.\end{aligned} \tag{46}$$

As with equation (40) the *composition effect* in equation (46) is non-zero if there is heterogeneity in agent elasticities across groups and the *misspecification effect* is non-zero if agent preferences are non-isoelastic. Though these effects result from the same mechanisms as their counterparts in equation (40) their precise nature differs somewhat. In addition, equation (46) includes a *baseline effect*, which reflects that fact that even if taxpayers have isoelastic preferences and homogeneous elasticities, $\frac{h(z^{\star};t)}{1-t}$ remains endogenous to the tax rate.

If $\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{h(z^{\star};t)}{1-t}\bar{\varepsilon}^c\left(z^{\star};t\right)\right]>0$ for all $t\in[t_0,t_1]$, then equations (44) and (45) imply that the fiscal externality of moving a kink point always falls between the fiscal externalities of the two local rate changes:

$$0 > FE_t^-(z^\star) > FE(z^\star) > FE_t^+(z^\star).$$

However, the sign of $\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{h(z^\star;t)}{1-t}\bar{\varepsilon}^c(z^\star;t)\right]$ is ambiguous, so in principle it is possible that the fiscal externality of moving a kink point $FE(z^\star)$ could also be above or below both $FE_t^-(z^\star)$ and $FE_t^+(z^\star)$.[47] This implies that there can exist Pareto- or welfare-improving movements of a kink point even if such improvements cannot be obtained via infinitesimal local rate changes near the kink point. On the other hand, the converse also holds true: in some cases, an infinitesimal local rate change might be efficiency- or welfare-enhancing even when moving a kink point is not. Indeed, this latter point holds irrespective of the sign of $\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{h(z^\star;t)}{1-t}\bar{\varepsilon}^c(z^\star;t)\right]$.

4.3 Discussion

In comparing my proposed bunching methodology to the standard approach, two salient considerations emerge. First, my method remains valid in the presence of multidimensional heterogeneity and/or departures from isoelastic functional form assumptions, whereas the standard bunching method—even in an ideal application where $\mathcal{H}$ is known—does not. Standard bunching estimates of the ETI can thus generate misleading policy implications in circumstances where my method will generate accurate policy implications. Second, policy analysis which focuses only on moving kink points to the exclusion of other local tax reforms can miss on beneficial reform opportunities.

A strength of the new bunching method proposed in this paper is its robustness to any realization of the unknown factors that determine how close a bunching-based elasticity ($\bar{\varepsilon}_B^c$) will be to the policy-relevant local average elasticities $\bar{\varepsilon}^c(z^\star;t_0)$ and $\bar{\varepsilon}^c(z^\star;t_1)$. However, the same ambiguity about these unknown factors also implies that it is impossible to know to what extent other beneficial local tax reforms might be overlooked by a policy analysis that focuses exclusively on moving kink points.

Other Challenges to Bunching-Based ETI Estimation

While I have discussed the some of the major challenges to identification and interpretation of bunching elasticities above, two further challenges are worth noting.

First, recall that bunching can be used to identify the revenue effect of moving a bracket threshold even when tax schedules are heterogeneous across groups, as shown in section 3.2 of this paper. For

[47] To see that the sign of $\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{h(z^\star;t)}{1-t}\bar{\varepsilon}^c(z^\star;t)\right]$ is ambiguous note because all three effects in equation (46) can be positive or negative. The baseline effect is positive if and only if

$$\bar{\varepsilon}^c(z^\star;t) < \frac{1}{1+\mathbb{E}[\alpha|z=z^\star]}.$$

The composition effect has an ambiguous sign because $\mathbb{E}[\alpha|z=z^\star]$ can be above or below -1 and $\mathrm{Cov}\left((\varepsilon^c)^2,\alpha|z=z^\star\right)$ can be positive or negative. Finally, both components of the misspecification effect have ambiguous signs.

the purpose of bunching-based ETI estimation on the other hand, this heterogeneity presents another challenge to the identification and interpretation of the bunching elasticity. In particular, note that the definition of the bunching elasticity given in equation (35) has fixed bounds of integration across taxpayers. While this expression can be extended to accommodate heterogeneous bounds of integration, tax schedule heterogeneity is another limitation of bunching-based ETI estimation methods that is relevant any time the data being used in the analysis pools taxpayers with differing tax schedules together.

Second, Mortenson and Whitten (2020) show that in the case of the EITC, it appears that observed bunching behavior is driven by taxpayers who simply report the level of income that maximizes their tax refund. This refund-maximizing level of income often coincides with the first EITC kink (or, in some years, other tax credit-induced kinks). They suggest that when this type of behavior is driving observed bunching behavior, the bunching elasticity may be of limited policy-relevance. Observed bunching in this case is primarily driven by the behavior of high (or even infinite) elasticity taxpayers, whereas the population of taxpayers elsewhere will have much lower elasticities.

However, it should be noted that the use of bunching to analyze kink point movements is robust to the presence of refund-maximizers. If a given kink point is the refund-maximizing point then moving the kink results in refund-maximizing agents simply relocating to the new kink point. This is precisely what the first-order approximation based on equation (11) assumes that all bunching taxpayers will do after the reform, so this estimation method remains valid.

5 Conclusion

The concept of the elasticity of taxable income as a sufficient statistic for revenue and welfare effects of taxation is powerful. However, as I show in this paper, it is not always necessary to have an estimate of the ETI in order to predict the impact of a tax reform of interest. Under basic theoretical restrictions, the bunching mass at a tax bracket threshold is itself a sufficient statistic for the (compensated) behavioral response effect of local movements of the threshold. This use of the bunching mass is fully nonparametric and requires no information about taxpayer preferences beyond the observed distribution of taxable income generated by their choices.

Building on this result, I develop a new approach to revenue forecasting, as well as empirical tests for Pareto efficiency and welfarist-optimality of piecewise linear tax schedules. An empirical application to the EITC motivates extensions of these results that incorporate important features of real-world policy settings. The results show that correctly incorporating optimization errors into revenue analysis can have a meaningful impact on the estimated revenue effects of moving a tax bracket threshold.

However, reforms that change bracket thresholds are not the only reforms of interest. Thus, the method proposed in this paper is best viewed as a complement to—rather than a substitute for—bunching-based elasticity estimation. By presenting a precise nonparametric interpretation of the bunching mass, this paper clarifies the circumstances under which the bunching elasticity provides a useful tool for predicting the behavioral response effect of other types of tax reforms. I show that the discrepancy between the bunching elasticity and local policy-relevant elasticities near the kink point depends on factors which are not currently well-documented empirically, such as the variance of the ETI. Gathering evidence about these factors could therefore be of great value for providing bunching practitioners with guidance on the external validity of their elasticity estimates.

Saez (2010) presented a tantalizing idea: that bunching in the observed distribution of taxable income can teach us about the response to taxation even absent any variation in the tax schedule over time or across different groups of taxpayers. Recent work has called this idea into question, documenting substantial challenges to identification and interpretation. However, this paper shows that by shifting our focus away from the ETI it is possible to deliver on the original promise of bunching methodology by using the bunching mass to identify a different policy-relevant parameter with a clear interpretation. The focus of prior bunching literature on the ETI makes sense given the importance of this parameter in the broader public economics literature. However, this focus has come at the cost of overlooking another valuable application of the bunching mass. Bunching methodology may therefore offer a case study in the value of ensuring that policy-motivated empirical work is tightly integrated with a formal welfare analysis of policy reforms of interest.

References

Apps, P., Van Long, N., & Rees, R. (2014). Optimal piecewise linear income taxation. *Journal of Public Economic Theory*, *16*(4), 523–545.

Aronsson, T., Jenderny, K., & Lanot, G. (2018). The quality of the estimators of the ETI. *Proceedings of the Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, *111*, 1–59.

Bergstrom, K., & Dodds, W. (2021). Optimal taxation with multiple dimensions of heterogeneity. *Journal of Public Economics*, *200*, 104442.

Bertanha, M., McCallum, A. H., & Seegert, N. (2021). Better bunching, nicer notching. *arXiv preprint arXiv:2101.01170*.

Bierbrauer, F., Boyer, P., & Hansen, E. (2020). Pareto-improving tax reforms and the Earned Income Tax Credit.

Blomquist, S., Newey, W. K., Kumar, A., & Liang, C.-Y. (2019). The bunching estimator cannot identify the taxable income elasticity.

Cattaneo, M., Jansson, M., Ma, X., & Slemrod, J. (2018). Bunching Designs: Estimation and Inference.

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. *The Quarterly Journal of Economics*, *126*(2), 749–804.

Dong, Y., & Lewbel, A. (2015). Identifying the effect of changing the policy threshold in regression discontinuity models. *Review of Economics and Statistics*, *97*(5), 1081–1092.

Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics*, *81*(4), 674–680.

Gerritsen, A. (2016). Optimal nonlinear taxation: The dual approach.

Goff, L. (2021). Treatment Effects in Bunching Designs: The Impact of the Federal Overtime Rule on Hours.

Golosov, M., Tsyvinski, A., & Werquin, N. (2014). A variational approach to the analysis of tax systems.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209.

Hendren, N. (2020). Measuring economic efficiency using inverse-optimum weights. *Journal of Public Economics*, *187*, 104198.

Jacquet, L., & Lehmann, E. (2021). Optimal income taxation with composition effects. *Journal of the European Economic Association*, *19*(2), 1299–1341.

Kleven, H. (2021). *The eitc and the extensive margin: A reappraisal* (Working Paper No. 26405). National Bureau of Economic Research. https://doi.org/10.3386/w26405

Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics*, *89*(11-12), 2093–2119.

Kroft, K., Kucko, K., Lehmann, E., & Schmieder, J. (2020). Optimal income taxation with unemployment and wage responses: A sufficient statistics approach. *American Economic Journal: Economic Policy*, *12*(1), 254–92.

Kumar, A., & Liang, C.-Y. (2020). Estimating taxable income responses with elasticity heterogeneity. *Journal of Public Economics*, *188*, 104209.

Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, *57*(4), 835–903.

Marx, B. M. (2018). The Cost of Requiring Charities to Report Financial Information.

Mortenson, J. A., & Whitten, A. (2020). Bunching to maximize tax credits: Evidence from kinks in the us tax schedule. *American Economic Journal: Economic Policy*, *12*(3), 402–32.

Neisser, C. (2021). The Elasticity of Taxable Income: A Meta-Regression Analysis.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies*, *68*(1), 205–229.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic policy*, *2*(3), 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), 3–50.

Sheshinski, E. (1989). Note on the shape of the optimum income tax schedule. *Journal of Public Economics*, *40*(2), 201–215.

Slemrod, J., & Kopczuk, W. (2002). The optimal elasticity of taxable income. *Journal of Public Economics*, *84*(1), 91–112.

Weber, C. E. (2014). Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences. *Journal of Public Economics*, *117*, 90–103.

Werning, I. (2007). Pareto Efficient Income Taxation.

A Discrete Kink Point Movements

The headline result of this paper pertains to infinitesimal movements of kink points. Since any real-world tax reform is necessarily discrete in nature, it may be insightful to consider what can (and cannot) be learned about the revenue effect of a discrete movement in a kink point using the bunching mass. In particular, consider decrease in the location of the kink point from some initial $z^{\star}$ to some new point $z^{\star\star} < z^{\star}$.

This reform has mechanical effects, which stem from the fact that any taxpayer initially located above the kink point ($z_1(\theta) > z^{\star}$) faces an increase in their tax liability of $(t_1 - t_0)(z^{\star} - z^{\star\star})$. Assuming there are no income effects, these agents will not change their choice of taxable income in response to this reform, and therefore the mechanical effect of the reform on tax revenue is

$$(t_1 - t_0)\left(1 - H_1\left(z^{\star}\right)\right)\left(z^{\star} - z^{\star\star}\right). \tag{47}$$

Turning to the behavioral response, as in the infinitesimal case we can split it into three components.

First, we have the relocation effect. Holding constant the size of the bunching mass, bunchers used to pay $t_0 z^{\star}$ in taxes, but now instead pay $t_0 z^{\star\star}$, changing revenue by

$$-t_0\left(z^{\star} - z^{\star\star}\right)\left(H_1\left(z^{\star}\right) - H_0\left(z^{\star}\right)\right), \tag{48}$$

Adding (48) and (47) together we obtain the first-order approximation of the revenue effect of a discrete change in the kink point (equation 11).

This approximation ignores the two additional components of the behavioral response: the new buncher effect, and the former buncher effect. some taxpayers who previously located below the kink will become bunchers. Specifically, those taxpayers whose choices under T_0 and T_1 satisfy,

$$z_1(\theta) < z^{\star\star} < z_0(\theta) < z^{\star}.$$

The first-order approximation implicitly assumes these taxpayers continue to pay $t_0 z_0(\theta)$ in taxes, but in reality, they now pay the lower amount $t_0 z^{\star\star}$. Thus, correcting the first-order approximation to account for the new buncher effect reduces tax revenue by

$$-t_0 \int_{z^{\star\star}}^{z^{\star}} \left(z - z^{\star\star}\right) h_0(z)\, \mathrm{d}z. \tag{49}$$

On the other hand, some taxpayers who previously chose to bunch will stop bunching and locate somewhere above the new kink point. These taxpayers are those whose choices under T_0 and T_1 satisfy

$$z^{\star\star} < z_1(\theta) < z^{\star} < z_0(\theta).$$

Approximating the behavioral response effect using the relocation effect implicitly assumes that these taxpayers remain as bunchers but move to the new kink point, paying $t_0 z^{\star\star}$ in taxes. But in fact, they pay the larger amount of $t_1 \left(z_1(\theta) - z^{\star\star}\right) + t_0 z^{\star\star}$. Thus, correcting the first-order approximation for the former buncher effect increases tax revenue by

$$t_1 \int_{z^{\star\star}}^{z^{\star}} \left(z - z^{\star\star}\right) h_1(z)\, \mathrm{d}z. \tag{50}$$

Adding (47), (48), (49), and (50) together we obtain the total effect of a discrete reform on tax revenue:[48]

$$\begin{aligned} R\left(z^{\star\star}\right) - R\left(z^{\star}\right) = & \underbrace{\left(t_1 - t_0\right)\left(z^{\star} - z^{\star\star}\right)\left(1 - H_1\left(z^{\star}\right)\right)}_{\text{mechanical effect}} \quad \underbrace{-t_0\left(z^{\star} - z^{\star\star}\right)\left(H_1\left(z^{\star}\right) - H_0\left(z^{\star}\right)\right)}_{\text{relocation effect}} \\ & - \underbrace{t_0 \int_{z^{\star\star}}^{z^{\star}} \left(z - z^{\star\star}\right) h_0(z)\, \mathrm{d}z}_{\text{new buncher correction}} + \underbrace{t_1 \int_{z^{\star\star}}^{z^{\star}} \left(z - z^{\star\star}\right) h_1(z)\, \mathrm{d}z}_{\text{former buncher correction}} \end{aligned} \tag{51}$$

Equation (51) provides some intuition for my main result. The bunching mass is a sufficient statistic for the behavioral response to a small change in the location of the kink point because, for sufficiently small reforms, the net impact of the new and former buncher effects is negligible.

Equation (51) presents the revenue effect of a discrete decrease in the location of a kink point. The same equation can be used to describe the effect of increasing the location of a kink point from $z^{\star}$ to $z^{\star\star} > z^{\star}$, but it is helpful to provide an explicit reinterpretation. The revenue effect of this change can be written as

$$\begin{aligned} R\left(z^{\star\star}\right) - R\left(z^{\star}\right) = & \underbrace{-\left(t_1 - t_0\right)\left(z^{\star\star} - z^{\star}\right)\left(1 - H_1\left(z^{\star}\right)\right)}_{\text{first-order approximation of mechanical effect}} + \underbrace{t_0\left(z^{\star\star} - z^{\star}\right)\left(H_1\left(z^{\star}\right) - H_0\left(z^{\star}\right)\right)}_{\text{first-order approximation of behavioral response}} \\ & \cdots - \underbrace{t_0 \int_{z^{\star}}^{z^{\star\star}} \left(z^{\star\star} - z\right) h_0(z)\, \mathrm{d}z}_{\text{former buncher correction}} + \underbrace{t_1 \int_{z^{\star}}^{z^{\star\star}} \left(z^{\star\star} - z\right) h_1(z)\, \mathrm{d}z}_{\text{new buncher correction}} \end{aligned} \tag{52}$$

Recall, the first two terms in this expression provide a first-order approximation to the effect of this reform. This approximation implicitly assumes that the revenue effect of the reform consists of a reduction in the tax liability of all taxpayers initially located above the original kink point $z^{\star}$, and an increase in revenue resulting from moving a fixed bunching mass to a higher locations. The interpretation of errors this approximation introduces differs slightly from the case of a kink point decrease however, where the approximation only introduced error into the behavioral response to

[48]The observant reader will notice that the derivation of equation (51) presented above abstracts from an additional form of behavioral response to discretely decreasing a kink point: that some taxpayers will move from an interior solution in the lower tax bracket to an interior solution in the higher tax bracket. However, this is done only for expositional simplicity; equation (51) can also be derived using the revenue function (8). The heuristic derivation works because the new buncher effect accounts for the removal of these individuals from the lower bracket and the former buncher effect accounts for their appearance in the higher bracket.

the reform. Here, it includes errors in both the behavioral response and the mechanical effect.

To see this, note that the first-order approximation implicit assumes that taxpayers initially located in the upper tax bracket between the original kink point and the new kink point will continue to choose $z_1(\theta)$ and will pay taxes of

$$t_1\left(z_1(\theta)-z^{\star}\right)+t_0 z^{\star}-\underbrace{\left(t_1-t_0\right)\left(z^{\star\star}-z^{\star}\right)}_{\text{mechanical effect}}.$$

Instead, these individuals choose to become bunchers at the new kink point $z^{\star\star}$, paying taxes of $t_0 z^{\star\star}$. The correction for this error take a simple form: the new buncher effect in equation (52).

There is also once again a former buncher effect, which corrects the error arising in the first-order approximation due to the fact that some taxpayers exit the bunching mass after this reform. Following an increase in the kink point, these taxpayers now locate somewhere on the lower tax bracket, paying taxes of $t_0 z_0(\theta)$ rather than the $t_0 z^{\star\star}$ implicitly assumed when using the first-approximation to the behavioral response to a reform.

Partial Identification and Approximations

Together, equations (51) and (52) provide a complete guide to the revenue effects of moving the location of a kink point.

If the pre-reform distribution of taxable income is observed, it is easy to see that three out of the four terms contained in equations (51) and (52) are identified. In both cases, the new buncher effect is identified by the observed income distribution (in addition to two first-order approximation terms). This is because the new buncher effect corrects for the first-order approximation's implicit assumption that they will continue to make the same choices that we currently observed them making, when in fact they will end up at a known counterfactual bunching location $z^{\star\star}$.

By contrast, the former buncher effect is not identified by observed data, because it serves to correct the first-order approximation's implicit assumption that some taxpayers will locate at the new bunching location $z^{\star\star}$ when in fact, they will end up choosing to locate some point on the new budget constraint which is infeasible on the pre-reform budget constraint. As we don't observe these choices in pre-reform data, we cannot identify this effect. Notice, this non-identification result stems from the same fundamental problem that underlies recent non-identification results regarding the standard bunching method: counterfactual decisions of bunchers are not observed.[49]

While this is a valuable result in itself, it is possible to expand on the practical value of the result by adopting a partial identification lens. For example, without any additional assumptions, we can identify a lower bound on the revenue effect of a discrete decrease in the kink point by evaluating

[49] As I discuss in the previous section, this is what causes non-identification in the standard bunching method when preferences are quasi-linear and isoelastic, and the ETI is homogeneous. Outside this special case, knowledge of counterfactual decisions of bunchers is insufficient.

equation (51) with the former buncher effect set to zero. This works because, in the case of a decrease in the kink point, the former buncher effect is always positive, so ignoring it will lead to an underestimate of the effect. A similar result holds for the case of a discrete increase in the kink point, but in this case evaluating equation (52) with the former buncher effect set to zero yields an upper bound, because for such reforms the former buncher effect is always negative.

Rather than simply assuming the former buncher effect is zero, we could use a trapezoidal approximation to the former buncher effect. It turns out that such an approximation is identified by the pre-reform distribution of taxable. In the case of a discrete decrease in a kink point we have

$$t_1 \int_{z^{\star\star}}^{z^{\star}} (z - z^{\star\star})\, h_1(z)\, \mathrm{d}z \approx \frac{(z^{\star} - z^{\star\star})^2}{2} t_1 h_1(z^{\star}) \tag{53}$$

and in the case of a discrete increase we have

$$-t_0 \int_{z^{\star}}^{z^{\star\star}} (z^{\star\star} - z)\, h_0(z)\, \mathrm{d}z \approx -\frac{(z^{\star} - z^{\star\star})^2}{2} t_0 h_0(z^{\star}). \tag{54}$$

Note, similar strategies have been employed in the traditional bunching literature from its inception. Saez (2010) proposes the use of trapezoidal approximation for counterfactual distribution of income for bunchers. However, in the standard bunching literature, identification of these approximations themselves requires much stronger assumptions about agent behavior than those presented here, because they include a counterfactual unobserved density. Equations (53) and (54) do not include any such counterfactual densities because the error the former buncher effect is largest in the region of taxable income close to the original kink point $z^{\star}$, but shrinks to zero close to the new kink point $z^{\star\star}$.[50]

Imposing weak additional restrictions on the density of taxable income allows for a reinterpretation of approximations (53) and (54) as bounds on the former buncher effect. In the case of a kink point decrease, if $h_1(z)$ is weakly decreasing for all $z \in [z^{\star\star}, z^{\star}]$, then (53) provides a lower bound on the former bunching effect. Conversely, if $h_0(z)$ is weakly decreasing over this interval, then we obtain an upper bound. Similarly, in the case of a kink point increase, if $h_0(z)$ is weakly decreasing (increasing) for all $z \in [z^{\star}, z^{\star\star}]$, then (54) provides an upper (lower) bound on the former bunching effect. Thus, when density functions are assumed to be monotone, sharper bounds on the revenue effect of discrete changes are possible.

[50] Using higher derivatives of of the distribution of taxable income, it is possible to extend these approximations even further. In fact, for any $n \in \mathbb{N}$ an nth-order Taylor approximation to the former buncher effect is identified by the observed distribution, though the practical value of this extension is limited by the feasibility of estimating such derivatives.

B Multiple Tax Brackets and Joint Reforms

My method can be easily extended to incorporate convex, piecewise linear tax schedules with more than two brackets.[51]

Suppose the tax schedule is

$$T(z) \equiv \max_{i \in \{0,1,\ldots,n\}} T_i(z) \tag{55}$$

where for each $i \in \{0, 1, \ldots, n\}$,

$$T_i(z) \equiv t_i z - V_i. \tag{56}$$

The tax schedule is thus fully characterized by a vector of bracket-specific marginal tax rates $\mathbf{t} = (t_0, t_1, \ldots, t_n)$ and an accompanying vector of bracket-specific virtual incomes $\mathbf{V} = (V_0, V_1, \ldots, V_n)$. Suppose that $(\mathbf{V}, \mathbf{t})$ satisfies $V_i > V_{i-1}$ and $t_i > t_{i-1}$ for all $i \in \{1, 2, \ldots, n\}$. Under these conditions, equations (55) and (56) describe a convex, piecewise linear tax schedule.

Further suppose that

$$\frac{V_{i+1} - V_i}{t_{i+1} - t_i} > \frac{V_i - V_{i-1}}{t_i - t_{i-1}}$$

for all $i \in \{1, 2, \ldots, n-1\}$. Then this tax schedule has $n+1$ different tax brackets and features n different convex tax bracket thresholds (kink points) whose locations are implicitly defined by the vectors of tax rates and virtual incomes $(\mathbf{V}, \mathbf{t})$. A taxpayer in the i-th bracket who is not bunching at a kink point has income some income $z \in \left(\frac{V_i - V_{i-1}}{t_i - t_{i-1}}, \frac{V_{i+1} - V_i}{t_{i+1} - t_i}\right)$. The kink point $z_i^\star \equiv \frac{V_{i+1} - V_i}{t_{i+1} - t_i}$ serves as the the ith bracket and the $(i+1)$-th bracket.

As in section 1, the actual tax schedule is thus composed of many constituent linear tax schedules of the form given by (56). Let $z_i(\theta)$ be the choice of taxable income a type θ taxpayer would make if they faced a particular linear tax $T_i(z)$ and let $H(z; t_i)$ be the resulting distribution of taxable income. Under assumptions 1 and 2, the corresponding density of taxable income under this linear schedule $(h_i(z))$ is continuous. Furthermore, we can describe the observed choices of taxpayers under the piecewise linear tax schedule (55) as

$$z(\theta) = \begin{cases} z_i(\theta) & \text{if } z_i(\theta) \in \left(z_{i-1}^\star, z_i^\star\right) \\ z_i^\star & \text{if } z_{i+1}(\theta) < z_i^\star < z_i(\theta) \end{cases} \tag{57}$$

That is to say, if a taxpayer's choice on a particular counterfactual linear tax schedule $T_i(z)$ is located in the tax bracket where the actual tax schedule (55) coincides with $T_i(z)$, then the taxpayer will choose an interior solution. On the other hand, the taxpayer will locate at a given kink point

[51] Indeed, it can actually be extended to include any budget set formed by the intersection of some finite set of smooth, convex budget sets. That is to say, the tax schedule within each tax bracket can be nonlinear, as long as it is convex and continuously differentiable. Thus, for example, my method could be readily be applied to analyze any bunching observed at the first kink point in the German income tax schedule, where the marginal tax rate rises discontinuously from 0 to 14% and thereafter increases continuously until reaching 42% (https://taxsummaries.pwc.com/germany/individual/taxes-on-personal-income). Such an extension is straightforward, and for simplicity I will focus here on only the piecewise linear case.

$z_i^\star$ which divides the i-th and $(i+1)$-th tax brackets if their choice under the counterfactual linear tax schedule $T_i(z)$ is located above the kink point and their choice under $T_{i+1}(z)$ is located below the kink point. By the strong axiom of revealed preference, all taxpayers will satisfy one and only one of these possible conditions.[52]

It follows from equation (5), that the observed distribution of taxable income will be

$$H(z) = \begin{cases} H(z; t_i) & \text{if } z \in \left(z_{i-1}^\star, z_i^\star\right) \\ H(z_i^\star; t_{i+1}) - H(z_i^\star; t_i) & \text{if } z = z_i^\star \end{cases} \tag{58}$$

Tax revenue under the tax schedule (55) can be written as a

$$R(\mathbf{z}^\star) \equiv \sum_{i=0}^{n} \left(\int_{z_{i-1}^\star}^{z_i^\star} T_i(z)\, h(z; t_i)\, \mathrm{d}z \right) + \sum_{i=1}^{n} t_i z_i^\star \left(H(z_i^\star; t_{i+1}) - H(z_i^\star; t_i) \right) \tag{59}$$

where $T_i(z; \mathbf{z}^\star)$ refers to the counterfactual linear tax schedules of equation (56). I have adjusted the notation here to make clear that, in general, the virtual income for each of these counterfactual tax schedules depends on multiple kink point locations. Note as well, that for equation 59 to provide a full description of tax revenue requires adopting the convention that $V_{-1} = V_0$ so that $z_{-1}^\star = 0$.

If we continue to rule out income effects (assumption 3), the first-order revenue effect of equally decreasing the virtual income level for tax brackets with an index $j > i$ is

$$-\frac{\partial R(\mathbf{z}^\star)}{\partial \kappa} = 1 - H(z_i^\star; t_{i+1}) - \frac{t_i}{t_{i+1} - t_i} \left(H(z_i^\star; t_{i+1}) - H(z_i^\star; t_i) \right). \tag{60}$$

But notice, this expression is proportional to the first revenue effect of decreasing a kink because that is exactly what such a reform does: it moves the kink point by adjust virtual income levels/Consequently the first-order welfare effects are also the same. Thus, Theorems 1–4 continue to apply in the case of multiple tax brackets. The fiscal externality of decreasing all V_j for $j > i$ is

$$FE(z_i^\star) \equiv -\frac{t_i \left(H(z_i^\star; t_{i+1}) - H(z_i^\star; t_i) \right)}{(t_{i+1} - t_i)\left(1 - H(z_i^\star; t_{i+1})\right)}. \tag{61}$$

[52] That at least one such condition must hold follows from the fact that the agent must choose some location on the budget set. To see why only one condition can hold, note that the budget set induced by the actual tax schedule (55) is simply the intersection of the budget sets induced by our $n+1$ counterfactual linear tax schedules. This means that if for some type θ we have $z_i(\theta) \in (z_i^\star, z_{i+1}^\star)$ and $z_j(\theta) \in \left(z_j^\star, z_{j+1}^\star\right)$ for $i \neq j$, then $z_i(\theta) \succ z_j(\theta)$ by SARP because $z_j(\theta)$ is affordable under the linear schedule T_i. Similarly, $z_j(\theta) \succ z_i(\theta)$ because $z_i(\theta)$ is affordable under the linear schedule T_j, a contradiction. Similarly, suppose that for type θ, $z_{i+1}(\theta) < z_{i+1}^\star < z_i(\theta)$ and $z_{j+1}(\theta) < z_{j+1}^\star < z_j(\theta)$ for some $i \neq j$. Suppose, without loss of generality, that $i > j$. This implies that $z_i(\theta)$ is affordable under T_{j+1} and that $z_{j+1}(\theta)$ is affordable under T_i. Thus, by SARP, we have $z_i(\theta) \succ z_{j+1}(\theta)$ and $z_{j+1}(\theta) \succ z_i(\theta)$. Similar reasoning completes the proof.

B.1 Moving Two Kinks

Suppose instead of moving a full set of virtual incomes we only alter decrease the virtual income V_i in the ith bracket. Such a reform has the effect of implicitly changing the location of both kink points that border the bracket, decreasing the lower kink point ($z^\star_{i-1}$) and increasing the upper kink point ($z^\star_i$). By considering this reform, we can develop tests of Pareto efficiency and welfarist optimality that go beyond the tests developed in section 1.

Consider a tax reform, indexed by $\kappa \in \mathbb{R}$, that simultaneously moves the locations of two kinks points $z^\star_i$ and $z^\star_j$, satisfying $z^\star_j > z^\star_i$. In particular, the reform decreases the location of $z^\star_i$ and increases the location of $z^\star_j$ as follows

$$\frac{\partial z^\star_i}{\partial \kappa} = -1 \qquad \frac{\partial z^\star_j}{\partial \kappa} = \frac{t_i - t_{i-1}}{t_j - t_{j-1}}, \tag{62}$$

while leaving the locations of all other kink points constant. Notice, independently decreasing the location of the lower kink point $z^\star_i$ would result in a mechanical effect that collects additional revenue from all taxpayers above this kink. But by increasing the location of the higher kink point $z^\star_j$ in just the right proportion, this joint reform offsets this effect for any taxpayers above $z^\star_j$, keeping the mechanical effect confined to only those taxpayers between the two kink points under consideration ($z^\star_i$ and $z^\star_j$).

Thus, absent income effects, the first-order effect of this reform on tax revenue is

$$\begin{aligned} -\frac{\mathrm{d}R(\mathbf{z}^\star)}{\mathrm{d}V_i} = {} & \underbrace{H(z^\star_i;t_i) - H\left(z^\star_{i-1};t_i\right)}_{\text{mechanical effect}} \\ & \underbrace{-\left(\frac{t_{i-1}}{t_i - t_{i-1}}\right)\overbrace{\left(H\left(z^\star_{i-1};t_i\right) - H\left(z^\star_{i-1};t_{i-1}\right)\right)}^{\text{bunching mass at } z^\star_j} + \frac{t_i}{t_{i+1}-t_i}\overbrace{\left(H(z^\star_i;t_{i+1}) - H(z^\star_i;t_i)\right)}^{\text{bunching mass at } z^\star_i}}_{\text{revenue effect of behavioral response}}. \end{aligned} \tag{63}$$

As with single kink point reforms, this effect is identified from the observed distribution of taxable income, though it now requires using the size of two bunching masses (one for each kink point). However, this joint reform differs from single kink reforms in a key respect: the behavioral response consists of both a positive and negative component. On the one hand, the reform decreases taxable income for the bunchers at the lower kink point, a behavioral response that decreases tax revenue. On the other hand, the reform increases taxable income for bunchers at the higher kink point, a behavioral response that *increases tax revenue.*

The fiscal externality of this tax reform is

$$FE^{V_i} \equiv -\frac{\frac{t_{i-1}}{t_i - t_{i-1}}\left(H\left(z^\star_{i-1};t_i\right) - H\left(z^\star_{i-1};t_{i-1}\right)\right) - \frac{t_i}{t_{i+1}-t_i}\left(H(z^\star_i;t_{i+1}) - H(z^\star_i;t_i)\right)}{\left(H(z^\star_i;t_i) - H\left(z^\star_{i-1};t_i\right)\right)}. \tag{64}$$

We can re-write this as a function of the fiscal externalities of decreasing the two kink points by themselves (using equation 61):

$$FE^{V_i} = \left(\frac{1 - H\left(z_{i-1}^{\star}; t_i\right)}{H\left(z_i^{\star}; t_i\right) - H\left(z_{i-1}^{\star}; t_i\right)} \right) FE\left(z_i^{\star}\right) - \left(\frac{1 - H\left(z_i^{\star}; t_{i+1}\right)}{H\left(z_i^{\star}; t_i\right) - H\left(z_{i-1}^{\star}; t_i\right)} \right) FE\left(z_{i+1}^{\star}\right). \tag{65}$$

This expression makes it clear that, in principle, it is possible for the fiscal externality of this type of joint reform to be positive.

Theorem 6 (Test for Pareto Efficiency (Multibracket))**.** *Under assumptions 1, 2, and 3, then a piecewise linear tax schedule which is Pareto efficient must satisfy*

$$FE\left(z_i^{\star}\right) \geq -1 \tag{66}$$

for every kink point $z_i^{\star}$*, and*

$$FE^{V_i} \geq -1 \tag{67}$$

for every virtual income level V_i*.*
Moreover, when a tax schedule satisfies both of these conditions (and the relevant second-order conditions in edge cases where $FE\left(z_i^{\star}\right) = -1$ *or* $FE^{V_i} = -1$*) no Pareto-improvement can be obtained through any infinitesimal tax reform that jointly moves any set of kink points.*

The second part of this theorem is analogous to the result presented in Bierbrauer, Boyer, and Hansen (2020), which shows that if there are no Pareto-improving "single-bracket" or "two-bracket" reforms to marginal tax rates, then there are no Pareto-improving reforms.[53] Theorem 6 states that if the are no Pareto-improving moves of a single kink point, and no Pareto-improving joint reforms of the kind described above, then there are no other Pareto-improvements which can be obtained through jointly moving kink points (infinitesimally).

Sketch of a proof: This result can be derived by first demonstrating that the fiscal externality of any joint movement of a collection of kink points can be written as a linear combination of the fiscal externalities of moving individual kink points (as equation 65 does for the particular joint reform discussed above). It is not difficult to then show that any the existence of a Pareto-improving joint movement of three kink points would imply that either condition 66 or condition 67 must fail.

[53]Note, in Bierbrauer, Boyer, and Hansen (2020), "single-bracket" and "two-bracket" reforms do not refer to changes in tax rates within different tax brackets in a piecewise linear tax schedule but rather describe local reforms at a single point in the tax schedule which can be intuitively described as "single-bracket" and "two-bracket".

C Derivations

C.1 Alternative Heuristic Derivation of Theorem

Here I present an alternative approach to building intuition for the result in Theorem 1, one based on showing that the errors that result from using equation (11) to approximate revenue effects are second-order.

Consider a tax reform which infinitesimally lowers the kink point by $\mathrm{d}z^\star$ (moving it from $z^\star$ to $z^\star - \mathrm{d}z^\star$). This reform discretely increases the marginal tax rate within a small window of taxable income immediately below $z^\star$ while simultaneously reducing the demogrant (virtual income) associated with the second tax bracket.

A seemingly naive approximation of the effect of this reform consists of summing together the following two components:

1. The *approximate mechanical effect* is the additional tax revenue generated by taxpayers with incomes in the upper bracket before the reform. Such agents pay a higher marginal rate on their earnings in the interval $(z^\star - \mathrm{d}z^\star, z^\star]$, so that each such agent's tax liability increases by by $(t_1 - t_0)\,\mathrm{d}z^\star$. This group of agents has a mass of $1 - H_1(z^\star)$, resulting in a total revenue effect of
$$(t_1 - t_0)\,(1 - H_1(z^\star))\,\mathrm{d}z^\star.$$
2. The *relocation effect* is the loss in tax revenue that results from simply moving all bunchers from $z^\star$ to $z^\star - \mathrm{d}z^\star$, holding constant the size of the bunching mass. For each such agent, tax liability falls by $-t_0\mathrm{d}z^\star$. These agents a pre-reform mass of $H_1(z^\star) - H_0(z^\star)$, resulting in a total revenue effect of:
$$-t_0\,(H_1(z^\star) - H_0(z^\star))\,\mathrm{d}z^\star.$$

Note, in contrast to the assumptions of the relocation effect, the size of the bunching mass will not in fact remain constant because some agents are at at the margin between bunching and not bunching pre-reform. For example, the group of agents with $z_0(\theta) \in (z^\star - \mathrm{d}z^\star, z^\star)$ are initially in the lower tax bracket but after the tax reform, these agents will become bunchers. As a result their taxable income will fall from $z_0(\theta)$ to $z^\star - \mathrm{d}z^\star$, inducing a reduction in tax revenue by $t_0\,(z_0(\theta) - z^\star + \mathrm{d}z^\star)$. For the case of an infinitesimal kink point shift, the mass of these agents is $h_0(z^\star)\,\mathrm{d}z^\star$ and their pre-reform taxable income is $z_0(\theta) = z^\star$. Thus, we can account for this additional loss in revenue by adding the *new buncher correction*,

$$-t_0 h_0(z^\star)\,(\mathrm{d}z^\star)^2\,,$$

to our naive approximation.

There is also be a group of agents with $z_1(\theta) \in (z^\star - \mathrm{d}z^\star, z^\star)$ who are bunching in the pre-reform period but will leave the bunching mass after the reform, choosing a post-reform level of taxable income $z_1(\theta)$, in the upper tax bracket. The relocation effect implicitly assigns these taxpayers a post-reform income of $z^\star - \mathrm{d}z^\star$, underestimating the post-reform tax revenue they will generate by $t_1(z_1(\theta) - z^\star + \mathrm{d}z^\star)$. For the case of an infinitesimal kink point shift, the mass of these agents is $h_1(z^\star)\,\mathrm{d}z^\star$ and their post-reform taxable income is $z_1(\theta) = z^\star$. Thus, we can account for this additional gain in revenue by adding the *former buncher correction*,

$$t_1 h_1(z^\star)(\mathrm{d}z^\star)^2,$$

to our naive approximation.

Theorem 1 follows from the fact that these new and former buncher corrections are both second-order.[54]

C.2 Heterogeneous Tax Rates

Suppose that type θ agents face the following two bracket tax schedule

$$T(z;\theta) \equiv \begin{cases} t_0(\theta)\,z & \text{if } z \leq z^\star(\theta) \\ t_1(\theta)\,z + [t_0(\theta) - t_1(\theta)]\,z^\star(\theta) & \text{if } z > z^\star(\theta) \end{cases},$$

where $z^\star(\theta)$, $t_0(\theta)$, and $t_1(\theta)$ are type-θ-specific tax schedule parameters. Let tax revenue from a type θ taxpayer be denoted as

$$\mathcal{T}(\theta) \equiv T(z(\theta);\theta).$$

Total tax revenue can then be written as

$$R \equiv \int \mathcal{T}(\theta)\,f(\theta)\,\mathrm{d}\theta.$$

Now consider a tax reform indexed by κ which infinitesimally changes the location of the kink point for type θ taxpayers by $\frac{\mathrm{d}z^\star(\theta)}{\mathrm{d}\kappa}$. The revenue effect of this reform is

$$\frac{\mathrm{d}R}{\mathrm{d}\kappa} = \int \frac{\mathrm{d}\mathcal{T}(\theta)}{\mathrm{d}z^\star(\theta)}\frac{\mathrm{d}z^\star(\theta)}{\mathrm{d}\kappa} f(\theta)\,\mathrm{d}\theta$$

where

$$\frac{\mathrm{d}\mathcal{T}(\theta)}{\mathrm{d}z^\star(\theta)} \equiv \begin{cases} 0 & \text{if } z_0(\theta) \leq z^\star(\theta) \\ t_0(\theta) & \text{if } z_1(\theta) < z^\star(\theta) < z_0(\theta) \\ t_0(\theta) - t_1(\theta) & \text{if } z_1(\theta) \geq z^\star(\theta) \end{cases}.$$

[54] That is, as $\mathrm{d}z^\star \to 0$, an effect which is proportional to$(\mathrm{d}z^\star)^2$ shrinks faster than an effect which is proportional to $\mathrm{d}z^\star$.

As in the standard case, some taxpayers are below the kink, and face no change in tax payments, some are above the kink and face a mechanical change in tax payments, and some are (non-marginal) bunchers and face a change in tax payments due to their behavioral response.The difference is that now the kink point locations and marginal rates are variable across individuals. For a general reform we have

$$\frac{\mathrm{d}R}{\mathrm{d}\kappa} = \int_{\theta\in\Theta_1} \left[t_1(\theta) - t_0(\theta)\right] \frac{\mathrm{d}z^\star(\theta)}{\mathrm{d}\kappa} f(\theta)\,\mathrm{d}\theta - \int_{\theta\in\Theta_B} t_0(\theta) \frac{\mathrm{d}z^\star(\theta)}{\mathrm{d}\kappa} f(\theta)\,\mathrm{d}\theta \tag{68}$$

where $\Pr\{z_1(\theta) < z^\star(\theta) < z_0(\theta)\}$ is the generalized bunching mass—the fraction of taxpayers who are bunching at their own kink point—and $\Pr\{z_1(\theta) > z^\star(\theta)\}$ is the fraction of taxpayers who are located above their own kink point.

In the main text, I assume that all taxpayers have the same kink point location so that $z(\theta) = z^\star$ for all θ so that

$$\Pr\{z_1(\theta) < z^\star(\theta) < z_0(\theta)\} = H_1(z^\star) - H_0(z^\star)$$

and $\Pr\{z_1(\theta) > z^\star(\theta)\} = 1 - H_1(z^\star)$. Letting $\frac{\mathrm{d}z^\star(\theta)}{\mathrm{d}\kappa} = -1$, equation (68) simplifies to equation (29).

C.3 Frictions

Suppose that the observed value of taxable income reflects a combination of the taxpayer's optimal choice and a random optimization error. In particular, let observed taxable income for a type θ agent be

$$z(\theta,\xi) = y(\theta)\,\xi$$

where $y(\theta)$ is the optimal choice of income for this taxpayer and $\xi \sim F_\xi$ is a multiplicative optimization error. I assume these errors are iid, and that the support of F_ξ is a convex subset of $\mathbb{R}_{++}$.[55] Suppose that assumptions 1—3 hold. Let $H_y(\cdot;t)$ be the distribution of taxpayer optimal choices under a linear tax schedule with marginal tax rate $t < 1$, and let $h_y(\cdot;t)$ be the corresponding density of optimal choices.

Total tax revenue derived from this tax system is

$$R(z^\star) \equiv \int_0^{z^\star} \hat{R}(y;z^\star)\,h_y(y;t_0)\,\mathrm{d}y + \hat{R}(z^\star;z^\star)\left(H_y(z^\star;t_1) - H_y(z^\star;t_0)\right) + \int_{z^\star}^{\infty} \hat{R}(y;z^\star)\,h_y(y;t_1)\,\mathrm{d}y \tag{69}$$

where

$$\hat{R}(y;z^\star) \equiv t_0 \int_0^{z^\star/y} y\xi f_\xi(\xi)\,\mathrm{d}\xi + \int_{z^\star/y}^{\infty} \left[t_1(y\xi - z^\star) + t_0 z^\star\right] f_\xi(\xi)\,\mathrm{d}\xi \tag{70}$$

is the average tax revenue generated by taxpayers with an optimal income of y. Note, the revenue

[55] This allows for the possibility of a bounded or unbounded error distribution and ensures that $z(\theta,\xi) > 0 \iff y(\theta) > 0$.

function without frictions (equation 8) can be obtained as a special case of equation (69) when F_ξ is a degenerate distribution such that $\mathbb{P}\{\xi = 1\} = 1$.

The derivative of equation (69) can be written as

$$R'(z^\star) = \underbrace{\hat{R}(z^\star; z^\star) h_y(z^\star; t_0) + \hat{R}(z^\star; z^\star)(h_y(z^\star; t_1) - h_y(z^\star; t_0)) - \hat{R}(z^\star; z^\star) h_y(z^\star; t_1)}_{=0 \text{ (new and former buncher effects)}}$$
$$+ \int_0^{z^\star} \frac{\partial \hat{R}(y; z^\star)}{\partial z^\star} h_y(y; t_0)\,\mathrm{d}y + \int_{z^\star}^{\infty} \frac{\partial \hat{R}(y; z^\star)}{\partial z^\star} h_y(y; t_1)\,\mathrm{d}y$$
$$+ \left[\left.\frac{\partial \hat{R}(y; z^\star)}{\partial y}\right|_{y=z^\star} + \frac{\partial \hat{R}(z^\star; z^\star)}{\partial z^\star} \right] (H_y(z^\star; t_1) - H_y(z^\star; t_0)) \quad (71)$$

Note that, as in the frictionless case, terms which are related to the new and former buncher effects cancel out. This is a consequence of the fact that changing the kink point induces the same changes in taxpayer optimal choices as in the frictionless case. The difference in revenue effect in stems from the fact that taxpayer optimal choices are differently aggregated into changes in revenue when optimization errors are present.

To further simplify equation (71), we can substitute in expressions for $\frac{\partial \hat{R}(y;z^\star)}{\partial z^\star}$ and $\left.\frac{\partial \hat{R}(y;z^\star)}{\partial y}\right|_{y=z^\star}$. The former partial derivative captures changes average revenue from a taxpayer who chooses optimal income y as a result of changing the kink point:

$$\frac{\partial \hat{R}(y; z^\star)}{\partial z^\star} = -(t_1 - t_0)\left(1 - F_\xi\left(\frac{z^\star}{y}\right)\right).$$

Following an increase of the kink point, all such taxpayers will pay slightly less in taxes any time they taxable income falls above $z^\star$, due to the mechanical effect of the reform. The second partial derivative we need captures the behavioral response effect caused by the change in the optimal choice of bunchers:[56]

$$\begin{aligned}\left.\frac{\partial \hat{R}(y; z^\star)}{\partial y}\right|_{y=z^\star} &= t_0 \int_0^1 \xi f_\xi(\xi)\,\mathrm{d}\xi + t_1 \int_1^\infty \xi f_\xi(\xi)\,\mathrm{d}\xi \\ &= t_0 F_\xi(1)\,\mathbb{E}[\xi|\xi \le 1] + t_1(1 - F_\xi(1))\,\mathbb{E}[\xi|\xi > 1].\end{aligned}$$

[56] Here, we define bunchers as taxpayers whose optimal choice is to locate at the kink ($y(\theta) = z^\star$) even though they may not locate at the kink due to optimization error.

Substituting these into equation (71) and applying integration by parts we get

$$R'(z^\star) = \underbrace{-(t_1 - t_0)(1 - H_z(z^\star; z^\star))}_{\text{mechanical effect}}$$
$$+ \underbrace{(t_0 F_\xi(1)\,\mathbb{E}[\xi|\xi \le 1] + t_1(1 - F_\xi(1))\,\mathbb{E}[\xi|\xi > 1])(H_y(z^\star; t_1) - H_y(z^\star; t_0))}_{\text{behavioral response effect}}.$$

The fiscal externality of moving a kink in this model can be written as

$$FE(z^\star) = -\left(\frac{t_0 F_\xi(1)\,\mathbb{E}[\xi|\xi \le 1] + t_1(1 - F_\xi(1))\,\mathbb{E}[\xi|\xi > 1]}{t_1 - t_0}\right)\left(\frac{H_y(z^\star; t_1) - H_y(z^\star; t_0)}{1 - H_z(z^\star; z^\star)}\right). \quad (72)$$

The second-order revenue effect is

$$R''(z^\star) = (t_1 - t_0)\frac{\mathrm{d}H_z(z^\star; z^\star)}{\mathrm{d}z^\star}$$
$$+ (t_0 F_\xi(1)\,\mathbb{E}[\xi|\xi \le 1] + t_1(1 - F_\xi(1))\,\mathbb{E}[\xi|\xi > 1])(h_y(z^\star; t_1) - h_y(z^\star; t_0))$$

where

$$\begin{aligned}\frac{\mathrm{d}H_z(z^\star; z^\star)}{\mathrm{d}z^\star} &= \left.\frac{\partial H_z(z; z^\star)}{\partial z}\right|_{z=z^\star} + \left.\frac{\partial H_z(z; z^\star)}{\partial z^\star}\right|_{z=z^\star} \\ &= \int_0^1 h_y(z^\star\xi; t_0)\, f_\xi(\xi)\,\mathrm{d}\xi + \int_1^\infty h_y(z^\star\xi; t_1)\, f_\xi(\xi)\,\mathrm{d}\xi,\end{aligned}$$

because

$$H_z(z; z^\star) \equiv \int_0^{z^\star/z} H_y(z\xi; t_0)\, f_\xi(\xi)\,\mathrm{d}\xi + \int_{z^\star/z}^\infty H_y(z\xi; t_1)\, f_\xi(\xi)\,\mathrm{d}\xi,$$

$$\left.\frac{\partial H_z(z; z^\star)}{\partial z}\right|_{z=z^\star} = \int_0^1 h_y(z^\star\xi; t_0)\, f_\xi(\xi)\,\mathrm{d}\xi + \int_1^\infty h_y(z^\star\xi; t_1)\, f_\xi(\xi)\,\mathrm{d}\xi + \frac{1}{z^\star} f_\xi(1)\left[H_y(z^\star; t_1) - H_y(z^\star; t_0)\right],$$

and

$$\left.\frac{\partial H_z(z; z^\star)}{\partial z^\star}\right|_{z=z^\star} = -\frac{1}{z^\star} f_\xi(1)\left(H_y(z^\star; t_1) - H_y(z^\star; t_0)\right).$$

C.4 Income and Participation Effects

Suppose that taxpayers face the two-bracket tax schedule (2). For each type of taxpayer θ, I assume that they make two choices:

1. Whether or not to enter the labor force.
2. Conditional on entering the labor force, how much income to earn.

Allowing for the possibility of income effects, a type θ taxpayer who enters the labor force will choose a value of taxable income that solves the problem utility maximization (1) where their utility function is assumed to satisfy Assumption 1. Recall, $z(t, V; \theta)$ is the choice of taxable income a type θ agent would make if facing a linear tax schedule with marginal tax rate t and virtual income V. Conditional on entering the labor force, the choice of taxable income a type θ agent makes when facing the kinked tax schedule (2) can be written as

$$z(\theta) = \begin{cases} z(t_0, V_0; \theta) & \text{if } z(t_0, V_0; \theta) < z^\star \\ z(t_1, V_1; \theta) & \text{if } z(t_1, V_1; \theta) > z^\star \\ z^\star & \text{if } z(t_1, V_1; \theta) \leq z^\star \leq z(t_0, V_0; \theta) \end{cases},$$

where $V_0 \equiv G$ and $V_1 \equiv (t_1 - t_0) z^\star + G$. As in section (1.1), it remains that case that all taxpayers must satisfy one, and only one of the three conditions in equation listed above. The alternative possibility $(z(t_0; \theta) < z^\star < z(t_1; \theta))$ would violate SARP, since the $z(t_0; \theta)$ would be affordable under T_1 and $z(t_1; \theta)$ would be affordable under T_0.

Let $v_E(\theta)$ be the indirect utility associated with a type θ agent's utility maximization problem:

$$v_E(\theta) \equiv u(z(\theta) - T(z(\theta)), z(\theta); \theta).$$

On the other hand, let $v_U(\theta)$ be the utility a type θ agent would receive if unemployed:

$$v_U(\theta) \equiv u(T(0), 0; \theta).$$

I assume that a type θ taxpayer only chooses to enter the labor force if the private benefit of doing so exceeds some fixed cost of work ξ. Assuming that $\xi \sim Q(\cdot|\theta)$ for all type θ taxpayers, the probability that such a taxpayer enters the labor force is

$$Q(\theta) \equiv Q(v_E(\theta) - v_U(\theta) | \theta).$$

Total tax revenue in this economy can constructed by integrating over the type distribution:

$$\begin{aligned} R(z^\star) \equiv & \underbrace{t_0 \int_{\theta \in \Theta_0(z^\star)} z(t_0, V_0; \theta) Q(\theta) f(\theta) \, d\theta}_{\text{lower bracket revenue}} + \underbrace{t_0 z^\star \int_{\theta \in \Theta_B(z^\star)} Q(\theta) f(\theta) \, d\theta}_{\text{buncher revenue}} \\ & + \underbrace{\int_{\theta \in \Theta_1(z^\star)} [t_1 (z(t_1, V_1; \theta) - z^\star) + t_0 z^\star] Q(\theta) f(\theta) \, d\theta}_{\text{upper bracket revenue}} \end{aligned} \quad (73)$$

where

$$\Theta_0(z^\star) = \{\theta \in \Theta : z(t_0, V_0; \theta) < z^\star\}$$

is the set of types who choose to locate at an interior point the lower tax bracket if they enter the labor force,

$$\Theta_1(z^\star) = \{\theta \in \Theta : z(t_1, V_1; \theta) > z^\star\}$$

is the set of types who choose to locate in the upper tax bracket if they enter the labor force, and

$$\Theta_B(z^\star) = \{\theta \in \Theta : z(t_1, V_1; \theta) \leq z^\star \leq z(t_0, V_0; \theta)\}$$

is the set of types who choose to bunch if they enter the labor force.

Differentiating equation (73), and applying the multidimensional version of Leibniz rule, we obtain

$$\begin{aligned}
-R'(z^\star) = {} & \underbrace{(t_1 - t_0)\int_{\theta\in\Theta_1} Q(\theta) f(\theta)\,\mathrm{d}\theta}_{\text{mechanical effect}} \underbrace{-t_0 \int_{\theta\in\Theta_B} Q(\theta) f(\theta)\,\mathrm{d}\theta}_{\text{substitution effect}} \\
& \underbrace{-t_1(t_1 - t_0)\int_{\theta\in\Theta_1} \frac{\partial z(t_1, V_1; \theta)}{\partial V_1} Q(\theta) f(\theta)\,\mathrm{d}\theta}_{\text{income effect}} \\
& \underbrace{-t_0 z^\star \int_{\theta\in\Theta_B} (1 - t_0 - MRS(z^\star, \theta)) \frac{\partial u(\theta)}{\partial c} q(\theta) f(\theta)\,\mathrm{d}\theta}_{\text{participation effect (bunchers)}} \\
& \underbrace{-\int_{\theta\in\Theta_1} [t_1(z(t_1, V_1; \theta) - z^\star) + t_0 z^\star](t_1 - t_0) \frac{\partial u(\theta)}{\partial c} q(\theta) f(\theta)\,\mathrm{d}\theta}_{\text{participation effect (upper bracket taxpayers)}} \qquad (74)
\end{aligned}$$

where $q(\theta) \equiv Q'(v_E(\theta) - v_U(\theta)|\theta)$ is the density of the fixed cost of working for type θ agents evaluated at $v_E(\theta) - v_U(\theta)$.

The first two terms in equation (74) are familiar from the discussion in section (1.2): the mechanical effect of the reform and the relocation effect. The *income effect* captures the revenue consequences of the behavioral responses of taxpayers located above the kink point who may choose to respond to their loss of income by increasing their earnings. For an infinitesimal reduction in the location of the kink point, these taxpayers experienced an income loss of $t_1 - t_0$, and their response is proportional to the size of this loss. If leisure is a normal good then $\frac{\partial z(t_1,V_1;\theta)}{\partial V_1} \leq 0$ for all types θ, so this effect will be positive as long as $t_1 > 0$.

The *participation effects* capture the fact that some taxpayers will choose to exit the labor force after the reform. Specifically, these are taxpayers who were at the margin between participating and not participating before the reform. Because the utility these agents derive from working declines following the reform, they will no longer choose to do so. For bunchers, the resulting change in the participation rate is proportional to the first-order reduction in utility caused by their behavioral response to the reform (as discussed in section (2.3)). For upper bracket taxpayers, the resulting change in the participation rate is proportional to the utility loss they experience due to their reduction in virtual income.

Let

$$\pi(\theta) \equiv \frac{1}{Q(\theta)} \frac{\partial u(\theta)}{\partial c} q(\theta)$$

be the semi-elasticity of labor force participation rate of type θ agents with respect to their consumption $c(\theta)$ and let

$$\eta(\theta) \equiv -\frac{\partial z_1(\theta z(t_1, V_1; \theta))}{\partial V_1} \geq 0$$

be the magnitude of the income effect for type θ agents (conditional on labor force participation). We can rewrite equation (74) as

$$\begin{aligned} -R'(z^\star) = & \underbrace{(t_1 - t_0)(1 - H_1(z^\star))}_{\text{mechanical effect}} \underbrace{-t_0(H_1(z^\star) - H_0(z^\star))}_{\text{substitution effect}} \\ & + \underbrace{t_1(t_1 - t_0)(1 - H_1(z^\star))\hat{\eta}^+(z^\star)}_{\text{income effect}} \\ & \underbrace{-t_0 z^\star (H_1(z^\star) - H_0(z^\star)) \mathbb{E}[k(\theta)\pi(\theta) | z(\theta) = z^\star]}_{\text{participation effect (bunchers)}} \\ & \underbrace{-t_1(t_1 - t_0)(1 - H_1(z^\star)) \mathbb{E}[(t_1(z(\theta) - z^\star) + t_0 z^\star)\pi(\theta) | z(\theta) > z^\star]}_{\text{participation effect (upper bracket taxpayers)}} \end{aligned} \tag{75}$$

where

$$\hat{\eta}^+(z^\star) \equiv \mathbb{E}[\eta(\theta) | z(\theta) > z^\star]$$

is the average income effect of the kink point and

$$k(\theta) \equiv 1 - t_0 - MRS(z^\star, \theta).$$

The total fiscal externality of moving the kink point is

$$FE(z^\star) = FE_{sub}(z^\star) + FE_{inc}(z^\star) + FE_{part}(z^\star),$$

where the portion attributable to the substitution effect is

$$FE_{sub}(z^\star) \equiv -\frac{t_0(H_1(z^\star) - H_0(z^\star))}{(t_1 - t_0)(1 - H_1(z^\star))} \leq 0.$$

The parts of $FE(z^\star)$ associated with the income and participation effects can be obtain by normalizing these effects (as they appear in equation (75)) by the mechanical effect:

$$FE_{inc}(z^\star) \equiv t_1 \hat{\eta}^+(z^\star)$$

and

$$FE_{part}\left(z^{\star}\right)=\underbrace{z^{\star}\mathbb{E}\left[k\left(\theta\right)\pi\left(\theta\right)|z\left(\theta\right)=z^{\star}\right]FE_{sub}\left(z^{\star}\right)}_{\text{buncher participation effect}}\underbrace{-t_{1}\mathbb{E}\left[\left(t_{1}\left(z\left(\theta\right)-z^{\star}\right)+t_{0}z^{\star}\right)\pi\left(\theta\right)|z\left(\theta\right)>z^{\star}\right]}_{\text{upper bracket participation effect}}$$

Interestingly, the fiscal externality of the buncher participation effect is proportional to the fiscal externality caused by the substitution effect. Nonetheless, because $\mathbb{E}\left[k\left(\theta\right)\pi\left(\theta\right)|z\left(\theta\right)=z^{\star}\right]$ is unknown, this fiscal externality is not identified by the observed distribution of taxable income.

C.5 Effect of Tax Rate on Local Average ETI

Note that the local average ETI at $z^{\star}$ can be written as

$$\bar{\varepsilon}^{c}\left(z^{\star};t\right)=\frac{\int_{\psi}\varepsilon^{c}\left(t;w^{\star}\left(t;\psi\right),\psi\right)h\left(z^{\star}|\psi;t\right)f_{\psi}\left(\psi\right)\mathrm{d}\psi}{\int_{\psi}h\left(z^{\star}|\psi;t\right)f_{\psi}\left(\psi\right)\mathrm{d}\psi}.$$

Differentiating this expression, we obtain

$$\begin{aligned}\frac{\mathrm{d}\bar{\varepsilon}^{c}\left(z^{\star};t\right)}{\mathrm{d}t}&=\frac{\int_{\psi}\varepsilon^{c}\left(t;w^{\star}\left(t;\psi\right),\psi\right)\frac{\partial h\left(z^{\star}|\psi;t\right)}{\partial t}f_{\psi}\left(\psi\right)\mathrm{d}\psi}{\int_{\psi}h\left(z^{\star}|\psi;t\right)f_{\psi}\left(\psi\right)\mathrm{d}\psi}-\bar{\varepsilon}^{c}\left(z^{\star};t\right)\frac{\int_{\psi}\frac{\partial h\left(z^{\star}|\psi;t\right)}{\partial t}f_{\psi}\left(\psi\right)\mathrm{d}\psi}{\int_{\psi}h\left(z^{\star}|\psi;t\right)f_{\psi}\left(\psi\right)\mathrm{d}\psi}\\&+\int_{\psi}\left[\frac{\partial\varepsilon^{c}\left(t;w^{\star}\left(t;\psi\right),\psi\right)}{\partial t}+\frac{\partial\varepsilon^{c}\left(t;w^{\star}\left(t;\psi\right),\psi\right)}{\partial w}\frac{\partial w^{\star}\left(t;\psi\right)}{\partial t}\right]\frac{h\left(z^{\star}|\psi;t\right)f_{\psi}\left(\psi\right)}{h\left(z^{\star};t\right)}\mathrm{d}\psi.\end{aligned}\quad(76)$$

To further simplify this expression, note that by differentiating the equality $H\left(z\left(t;w,\psi\right)|\psi;t\right)=F\left(w|\psi\right)$ we can obtain

$$\frac{\partial z}{\partial w}h\left(z\left(t;w,\psi\right)|\psi;t\right)=f\left(w|\psi\right),$$

and by further differentiating this expression we get

$$\frac{\partial h\left(z\left(t;w,\psi\right)|\psi;t\right)}{\partial t}=-\left(\frac{\partial^{2}z}{\partial w\partial t}/\frac{\partial z}{\partial w}\right)h\left(z\left(t;w,\psi\right)|\psi;t\right)-\frac{\partial z}{\partial t}h'\left(z\left(t;w,\psi\right)|\psi;t\right).\quad(77)$$

Further note that by differentiating both sides of the equality $z\left(t;w^{\star}\left(t;\psi\right),\psi\right)=z^{\star}$ we obtain

$$\frac{\partial w^{\star}\left(t;\psi\right)}{\partial t}=-\frac{\partial z\left(t;w^{\star}\left(t;\psi\right),\psi\right)/\partial t}{\partial z\left(t;w^{\star}\left(t;\psi\right),\psi\right)/\partial w}.$$

Combining this expression with equation (77), together with the definitions of $\varepsilon^{c}\left(t;w^{\star}\left(t;\psi\right),\psi\right)$ and $\alpha\left(z^{\star}|\psi;t\right)$ we get

$$\frac{\partial h\left(z^{\star}|\psi;t\right)}{\partial t}=\frac{\frac{\partial\varepsilon^{c}\left(t;w^{\star}\left(t;\psi\right),\psi\right)}{\partial w}\frac{\partial w^{\star}\left(t;\psi\right)}{\partial t}}{\varepsilon^{c}\left(t;w^{\star}\left(t;\psi\right),\psi\right)}h\left(z^{\star}|\psi;t\right)-\frac{1}{1-t}\varepsilon^{c}\left(t;w^{\star}\left(t;\psi\right),\psi\right)\left(1+\alpha\left(z^{\star}|\psi;t\right)\right)h\left(z^{\star}|\psi;t\right),\quad(78)$$

and substituting this into equation (76), we obtain equation (40).

$$\frac{\mathrm{d}\bar{\varepsilon}^c\left(z^\star;t\right)}{\mathrm{d}t}=-\frac{1}{1-t}\mathrm{Cov}\left(\varepsilon^c,\left(1+\alpha\right)\varepsilon^c|z=z^\star\right)+\mathbb{E}\left[\frac{\partial\varepsilon^c}{\partial t}|z=z^\star\right]+\mathbb{E}\left[\left(2-\frac{\bar{\varepsilon}^c\left(z^\star;t\right)}{\varepsilon^c}\right)\frac{\partial\varepsilon^c}{\partial w}\frac{\partial w^\star\left(t;\psi\right)}{\partial t}|z=z^\star\right]$$

C.6 Connecting Fiscal Externalities of Different Reforms

First, note that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{h\left(z^\star;t\right)}{1-t}\bar{\varepsilon}^c\left(z^\star;t\right)\right]=\frac{h\left(z^\star;t\right)}{1-t}\frac{\partial\bar{\varepsilon}^c\left(z^\star;t\right)}{\partial t}+\left(\frac{h\left(z^\star;t\right)}{\left(1-t\right)^2}+\frac{1}{1-t}\frac{\partial h\left(z^\star;t\right)}{\partial t}\right)\bar{\varepsilon}^c\left(z^\star;t\right). \tag{79}$$

Integrating over equation (78) we obtain $\frac{\partial h(z^\star;t)}{\partial t}$ and by combining this with equation (40) we can re-write equation (79) as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{h\left(z^\star;t\right)}{1-t}\bar{\varepsilon}^c\left(z^\star;t\right)\right]=\frac{h\left(z^\star;t\right)}{1-t}&\left[\frac{1}{1-t}\left(\bar{\varepsilon}^c\left(z^\star;t\right)-\mathbb{E}\left[\left(1+\alpha\right)\left(\varepsilon^c\right)^2|z=z^\star\right]\right)\right.\\&\left.+\mathbb{E}\left[\frac{\partial\varepsilon^c}{\partial t}|z=z^\star\right]+2\mathbb{E}\left[\frac{\partial\varepsilon^c}{\partial w}\frac{\partial w^\star\left(t;\psi\right)}{\partial t}|z=z^\star\right]\right]\end{aligned}$$

Rearranging this expression gives equation (46).

D Filtering Method

D.1 Identification

The identifications strategy underlying my filtering method comes from Cattaneo, Jansson, Ma, and Slemrod (2018). As noted in section 3.1, I assume that observed taxable income choices satisfy

$$\tilde{z}=\tilde{y}+\tilde{\xi}$$

where $\tilde{x}\equiv\log\left(x\right)$ for any variable x, y is the (latent) optimal choice of taxable income, and $\tilde{\xi}$ is a mean-zero normally distributed optimization error. Cattaneo, Jansson, Ma, and Slemrod (2018) show that if the distribution of optimal log-income ($\tilde{H}_y$) is assumed to feature a non-zero bunching mass at the kink point, then both $\tilde{H}_y$ and the variance of the optimization errors σ^2 are semiparametrically identified by $\tilde{H}_z$.[57]

For the reader's convenience, I replicate the identification proof from Cattaneo, Jansson, Ma, and Slemrod (2018) here.

[57] In fact, their modeling assumptions differ from mine, as they assume that the optimization error is additive in levels rather than in logs. While this difference could have an impact on resulting estimates, it does not change the semiparametric identification result that underlies this approach to filtering.

First, let

$$\tilde{H}_z\left(\tilde{z};\tilde{H}_y,\sigma^2\right)\equiv\int_{-\infty}^{\log(z^\star)}\tilde{H}_y\left(\tilde{y};t_0\right)\frac{1}{\sigma}\phi\left(\frac{\tilde{z}-\tilde{y}}{\sigma}\right)\mathrm{d}\tilde{y}+\int_{\log(z^\star)}^{\infty}\tilde{H}_y\left(\tilde{y};t_1\right)\frac{1}{\sigma}\phi\left(\frac{\tilde{z}-\tilde{y}}{\sigma}\right)\mathrm{d}\tilde{y}$$

be the CDF of observed log-income that would result for a given particular distribution of optimal log-income ($\tilde{H}_y$) and the variance of optimization errors (σ^2). Semiparametric identification can be demonstrated by showing that

$$\tilde{H}_z\left(\tilde{z};\tilde{H}_y,\sigma^2\right)=\tilde{H}_z\left(\tilde{z};\hat{H}_y,\hat{\sigma}^2\right)\quad\forall\tilde{z}\qquad\Longleftrightarrow\left(\tilde{H}_y,\sigma^2\right)=\left(\hat{H}_y,\hat{\sigma}^2\right).$$

Suppose not, so that there exists some $\left(\tilde{H}_y,\sigma^2\right)$ and $\left(\hat{H}_y,\hat{\sigma}^2\right)$ such that $\left(\tilde{H}_y,\sigma^2\right)\neq\left(\hat{H}_y,\hat{\sigma}^2\right)$ and

$$\tilde{H}_z\left(\tilde{z};\tilde{H}_y,\sigma^2\right)=\tilde{H}_z\left(\tilde{z};\hat{H}_y,\hat{\sigma}^2\right)\qquad\forall\tilde{z}.$$

Note, the density of log-income can be written as

$$\begin{aligned}\tilde{h}_z\left(\tilde{z};\tilde{H}_y,\sigma^2\right)=&\int_{-\infty}^{\log(z^\star)}\tilde{h}_y\left(\tilde{y};t_0\right)\frac{1}{\sigma}\phi\left(\frac{\tilde{z}-\tilde{y}}{\sigma}\right)\mathrm{d}\tilde{y}+\int_{\log(z^\star)}^{\infty}\tilde{h}_y\left(\tilde{y};t_1\right)\frac{1}{\sigma}\phi\left(\frac{\tilde{z}-\tilde{y}}{\sigma}\right)\mathrm{d}\tilde{y}\\&+\left[\tilde{H}_y\left(\log\left(z^\star\right);t_1\right)-\tilde{H}_y\left(\log\left(z^\star\right);t_0\right)\right]\frac{1}{\sigma}\phi\left(\frac{\tilde{z}-\log\left(z^\star\right)}{\sigma}\right).\end{aligned}$$

The characteristic function of observed log-income given the parametrization $\left(\tilde{H}_y,\sigma^2\right)$ is

$$\begin{aligned}\mathcal{F}\left(\tilde{h}_z\left(\cdot;\tilde{H}_y,\sigma^2\right)\right)&\equiv\int\exp\left\{-i\tilde{z}t\right\}\tilde{h}_z\left(\tilde{z};\tilde{H}_y,\sigma^2\right)\mathrm{d}\tilde{z}\\&=\exp\left\{-\frac{1}{2}\sigma^2t^2\right\}\cdot\left\{\int_{-\infty}^{\log(z^\star)}\exp\left\{-i\tilde{y}t\right\}\tilde{h}_y\left(\tilde{y};t_0\right)\mathrm{d}\tilde{y}+\int_{\log(z^\star)}^{\infty}\exp\left\{-i\tilde{y}t\right\}\tilde{h}_y\left(\tilde{y};t_1\right)\mathrm{d}\tilde{y}\right.\\&\qquad\left.+\left[\tilde{H}_y\left(\log\left(z^\star\right);t_1\right)-\tilde{H}_y\left(\log\left(z^\star\right);t_0\right)\right]\exp\left\{-i\log\left(z^\star\right)t\right\}\right\}\\&=\exp\left\{-\frac{1}{2}\sigma^2t^2\right\}\cdot\mathcal{F}\left(\tilde{h}_y\left(\tilde{y}\right)\right)\end{aligned}$$

where $\mathcal{F}\left(\tilde{h}_y\left(\cdot\right)\right)$ is the characteristic function of optimal log-income given the distribution $\tilde{H}_y$.

Note that the characteristic function of observed log-income should be the same under both parametrizations:

$$\mathcal{F}\left(\tilde{h}_z\left(\cdot;\tilde{H}_y,\sigma^2\right)\right)=\mathcal{F}\left(\tilde{h}_z\left(\cdot;\hat{H}_y,\hat{\sigma}^2\right)\right).$$

Without loss of generality, suppose that $\hat{\sigma}>\sigma$. Dividing both sides of the above equation by $\exp\left\{-\frac{1}{2}\sigma^2t^2\right\}$, we get

$$\mathcal{F}\left(\tilde{h}_y\left(\cdot\right)\right)=\mathcal{F}\left(\tilde{h}_z\left(\cdot;\hat{H}_y,\hat{\sigma}^2-\sigma^2\right)\right)$$

which implies that

$$\tilde{H}_y(\tilde{z}) = \tilde{H}_z\left(\tilde{z}; \hat{H}_y, \hat{\sigma}^2 - \sigma^2\right) \qquad \forall \tilde{z}.$$

But if there is a discontinuity in $\tilde{H}_y(\cdot)$ this is not possible since the CDF on the right hand side of this equation is continuous. Thus, under the assumption that $\tilde{H}_y(\cdot)$ features a discontinuity there exists a unique pair $\left(\tilde{H}_y, \sigma^2\right)$ which induces the distribution of observed log-income $\tilde{H}_z$.

D.2 Estimation

My estimation method approximates $\tilde{H}_y^0(\tilde{y})$ and $\tilde{H}_y^1(\tilde{y})$ in the expression above using higher-order polynomials:

$$\tilde{H}_y^0(\tilde{y}) \approx \sum_{k=0}^{p} \beta_k \tilde{y}^k \qquad \text{and} \qquad \tilde{H}_y^1(\tilde{y}) \approx \sum_{k=0}^{p} \alpha_k \tilde{y}^k.$$

For any given parametrization of these polynomials, and standard deviation of errors σ, we can simulate the observed CDF of log-income at $\log(z_i)$ as

$$\begin{aligned}\tilde{H}_z^{poly}(\log(z_i); \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma) \equiv \frac{1}{S}\sum_{s=1}^{S}\Bigg[&\mathbf{1}\{\sigma\epsilon_{i,s} > \log(z_i) - \log(z^\star)\}\sum_{k=0}^{p}\beta_k(\log(z_i) - \sigma\epsilon_{i,s})^k \\ &+\mathbf{1}\{\sigma\epsilon_{i,s} \leq \log(z_i) - \log(z^\star)\}\sum_{k=0}^{p}\alpha_k(\log(z_i) - \sigma\epsilon_{i,s})^k\Bigg]\end{aligned}$$

Note, for any value of $\log(z_i)$, the value of $\tilde{H}_z^{poly}\left(\log(z_i); \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2\right)$ is numerically simulated using S draws from a standard normal distribution. The sth draw for observation i is denoted $\epsilon_{i,s}$.

Let $\tilde{H}_z^{Emp}(\cdot)$ be the empirical CDF of taxable income.

Using a sample of taxpayers in a neighborhood of the kink point, $\mathcal{N}(z^\star)$, I estimate the coefficients of these polynomial functions $(\boldsymbol{\beta}, \boldsymbol{\alpha})$, along with the error standard deviation σ, via constrained minimization of simulated least squares:

$$\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\sigma}\right) \equiv \arg\min_{(\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma)} \sum_{i: z_i \in \mathcal{N}(z^\star)} \left(\tilde{H}_z^{poly}(\log(z_i); \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma) - \tilde{H}_z^{Emp}(\log(z_i))\right)^2 \tag{80}$$

subject to

$$0 < \sum_{k=0}^{p} \beta_k \tilde{y}^k < \sum_{k=0}^{p} \alpha_k \tilde{y}^k < 1 \qquad \forall \tilde{y} \in \mathcal{N}(z^\star), \tag{81}$$

$$\sum_{k=1}^{p} k\beta_k \tilde{y}^{k-1} > 0 \qquad \forall \tilde{y} \in \mathcal{N}(z^\star), \tag{82}$$

and

$$\sum_{k=1}^{p} k\alpha_k \tilde{y}^{k-1} > 0 \qquad \forall \tilde{y} \in \mathcal{N}(z^\star). \tag{83}$$

These constraints ensure that the polynomial approximations of $\tilde{H}_y^0(\tilde{y})$ and $\tilde{H}_y^1(\tilde{y})$ behave like valid CDFs within the estimation window $\mathcal{N}(z^\star)$: constraining them to remain in the $[0,1]$ interval and to have positive first derivatives. The first constraint also ensures that these functions satisfy the theoretical requirement that $\tilde{H}_y^0(\tilde{y}) < \tilde{H}_y^1(\tilde{y})$ for all $\tilde{y}$.[58] Note, this constraint also implies that there must be a non-zero bunching mass in the latent distribution $\tilde{H}_y(\tilde{y})$ that would be derived from these polynomial functions, thus imposing the assumption needed to semiparametrically identify the $\tilde{H}_y^0$, $\tilde{H}_y^1$, and σ (as Cattaneo, Jansson, Ma, and Slemrod (2018) show).

[58]This should be expected as long as some taxpayers at every level of income are at least a little responsive to taxation.