Identifying the Causal Effect of a Tax Rate Change When There are Multiple Tax Brackets

*Caroline E. Weber**

April 2012

Abstract

Empirical researchers frequently obtain estimates of the behavioral response to a tax change by exploiting variation in the degree to which a tax reform affects different groups of individuals based on their individual characteristics and tax situations. This paper analyzes the conditions under which it is possible to obtain a causal average treatment effect using pre-reform characteristics as instruments for the observed tax rate change, which I term the Fixed-Bracket Average Treatment Effect (FBATE). Previous literature has assumed that only one of these conditions is sufficient to identify a causal parameter. FBATE identifies the average treatment effect for individuals with no incentive to switch tax brackets in response to a tax reform or other shock that affects the bracket in which an individual is located. FBATE is the relevant parameter for welfare analysis if taxpayers whose response is identified by the FBATE estimate will have the same long-run response as the rest of the population. FBATE also highlights new trade-offs between different sources of identification; for example, an oft-touted source of identification—bracket creep—cannot yield a causal estimate. The paper also shows that using an alternative definition of treatment relative to what is usually employed in the literature obtains a causal average treatment effect for a larger subpopulation under weaker assumptions.

A special thanks to my dissertation committee for their invaluable encouragement and comments: Jim Hines, Joel Slemrod (Chair), Jeff Smith, and Kevin Stange. Also, many thanks to David Agrawal, David Cashin, Rob Garlick, Laura Kawano, Andreas Peichl, Sergio Urzúa, Michigan Tax Research Invitational participants, International Institute of Public Finance participants, and University of Michigan Public Finance Seminar participants for helpful comments. All remaining errors are my own.

*Ph.D. Candidate, Department of Economics, University of Michigan: ceweber@umich.edu.

1 Introduction

Empirical researchers frequently obtain estimates of the behavioral response to a tax change by exploiting variation in the degree to which a tax reform affects different groups of individuals based on their individual characteristics and tax situations. Often, the tax schedule examined has multiple brackets and at least part of the identification of the estimates comes from differences in legislated tax rate changes across brackets. Examples include examinations of the responses to the personal income tax schedule, the Earned Income Tax Credit (EITC), and social security contributions, among others. These estimates are important for policy analysis, both in terms of deadweight loss and revenue implications. In most contexts, a theoretical framework has been developed which maps from the estimates obtained to a calculation of deadweight loss.[2]

In general, the empirical literature has gone one of two ways—it constructs a measure of the predicted tax change based on observable characteristics, and then either estimates the response to this predicted change directly, or uses this as an instrument for the actual tax rate change. It is relatively clear how to assess the validity and interpret the parameter when the former approach is employed; however, when the latter approach is employed, it is less straightforward and the existing literature provides no discussion or guidance on this matter. This paper seeks to fill in this gap in the literature. By carefully examining the latter approach, the paper also explains which method may be preferred in a given context.

The challenge of using the actual tax rate as an independent variable in an estimating equation is that we, as researchers, observe a tax rate for all individuals, but the tax rate we observe is systematically wrong for certain subgroups. This is because we only observe the tax rate—the treatment—after individuals have responded, and sometimes individuals face incentives to cross tax bracket lines (thereby altering their observed treatment) as part of their behavioral response. This "treatment mismeasurement" will systematically bias the

[2]For example, Eissa et al. (2008) do this for the EITC, and Feldstein (1999) and Chetty (2009) do this for the elasticity of taxable income (ETI).

estimates unless addressed properly. This is unlike a labor or other classic treatment effect setting in which there may be selection into treatment, but the treatment that determines individuals' responses is observed, and if there was a random assignment mechanism before selection, treatment based on random assignment is also observed. Because of the difference in the point at which researchers studying tax rate changes can observe treatment and the resulting treatment mismeasurement this introduces, the standard analysis and interpretation of the estimates obtained does not apply.

The main contribution of this paper is to derive the conditions under which it is possible to obtain a causal average treatment effect using pre-reform characteristics as instruments, taking treatment mismeasurement into consideration. I call the treatment effect obtained the Fixed-Bracket Average Treatment Effect (FBATE), which will identify the average treatment effect for individuals with no incentive to switch tax brackets in response to a tax reform or other shock that affects the tax bracket in which an individual is located. FBATE provides a standard which can be used to assess possible instruments and sources of identifying variation, interpret existing parameters, and identify conditions under which the response to future anticipated, as well as current, tax changes can be estimated.

Applying FBATE to the existing literature provides a useful interpretation of the estimates. For estimates that identify FBATE, this paper highlights that such estimates exclude particular types of individuals—those with an incentive to deviate across tax bracket lines due a marginal tax rate change—and these individuals may or may not respond in a similar way as other individuals. Therefore, FBATE is the relevant parameter for welfare analysis if taxpayers whose response is identified by the FBATE estimate will have the same long-run response as the rest of the population.

Assessing possible instruments and sources of identifying variation in light of FBATE provides new insights regarding what is ideal. For example, the ETI literature has touted using bracket creep as a source of identifying variation because it changes the marginal tax rate for individuals who are otherwise quite similar(e.g., Saez et al., 2012). The literature

has also noted that individuals may not be aware of such detailed changes in their marginal tax rate, and even if they are, these may not be the most appropriate changes to examine to identify the underlying structural parameter if individuals face substantial optimization frictions (e.g., Saez et al., 2012; Chetty, 2011). However, as will be shown below, using bracket creep as a source of identifying variation will never provide a causal average treatment effect. Additionally, switching to a context in which there is a large marginal tax rate change where individuals will overcome the optimization frictions they face, is not necessarily better, because high optimization frictions will provide greater incentives for individuals to shift tax brackets in response to a tax reform. This suggests that, in reality, there is likely a trade-off between the bias that comes from using a smaller marginal tax rate change, where the estimate is closer to FBATE but further away from the structural parameter desired for welfare calculations (Chetty, 2011), and a larger marginal tax rate change where the opposite is true.

The paper proceeds as follows. Section 2 lays out a framework for causal inference and derives FBATE in this context for panel data under certain assumptions. Section 3 provides several empirical applications of these results. Section 4 discusses broader implications for the literature given the results in Section 2. Section 5 concludes.

2 Framework and Causal Inference

In this section, I lay out a framework for causal inference and derive the conditions under which a causal average treatment effect is obtained. The framework shares some similarities with standard treatment effect settings, but also has a few notable differences due to the fact that, in this context, individuals respond to the treatment they receive and sometimes this response changes the treatment observed by the researcher. When this occurs, the actual and observed treatments no longer coincide. I call this problem "treatment mismeasurement." The problem is similar to the "contamination bias" discussed in Heckman and Robb (1985),

in the sense that we do not observe treatment accurately for all, and if we assigned treatment in the most obvious, observable way, the estimates would be biased. It is very different from the large literature on imperfect compliance with experiments in which the relevant treatment after selection has taken place is observed. I consider both using a proxy measure for treatment and an instrument with treatment defined as the observed tax rate in each period.

I will use the estimation of the ETI as my running example throughout the paper; however, the analytics are written generally for any marginal or average tax rate change, and clearly apply broadly to all cases in which researchers are trying to estimate the causal effect of a tax rate change when there are multiple tax brackets. For estimation of the ETI, the outcome of interest is taxable income and taxable income is also the determinant of the marginal tax rate faced. I assume that the researcher has access to panel data for the derivations, but I discuss how the results apply to repeated-cross-section analyses as well. Subsection 2.1 considers a simplified case, in which some of the complexities of this estimation problem are ignored in order to build intuition. These assumptions are then relaxed in Subsection 2.2. Subsection 2.3 extends the analysis in Subsection 2.2 to consider the estimation of an anticipated tax reform.

2.1 Stylized Example

This subsection uses a stylized example, which strips away some of the additional complexities of the estimation problem in order to build intuition. The results in this subsection are often starker than those in Subsection 2.2 which eliminates the stylized assumption, but the important points and intuition carry through. This section shows that while treatment in this literature has traditionally been determined period by period, a more natural way of thinking about treatment is the treatment determined by the first period. Defining the treatment period by period requires stronger assumptions to obtain a causal average treatment effect, and this treatment effect—FBATE—is identified over a narrower subpopulation.

Additionally, if excluding individuals far away from the treatment cutoff based on the variable that determines treatment status (taxable income), this cutoff should be imposed as a function of income in the first period, not income period by period, to avoid introducing a bias in the estimates. In practice, the latter method is used frequently when repeated-cross-section data is used. Lastly, this section shows that, with the introduction of treatment mismeasurement, rescaling the ITT estimate using a Wald estimator does not necessarily get closer to obtaining the average treatment effect.

The simplifying assumption imposed in this subsection is as follows:

Assumption 1: *Income is fixed, except when it responds to a change in the tax rate; that is, it does not move for secular reasons, including transitory income shocks.*

This means that the only reason income changes is in response to a change in the tax rate. Consider the tax reform depicted in Figure 5. There are two periods, period 1 and period 2. In period 1, the tax rate is the same for all individuals. In period 2, the marginal tax rate is higher for all individuals above the tax kink k.[3] Defining the treatment and comparison group based on period 1 income, the treatment group $t1$ consists of those who are above k in period 1 and the comparison group $c1$ includes those below k. In period 1, there is no tax kink at point k, so that the tax rates are the same for both groups $\tau_{c1} = \tau_{t1}$. In period 2, a tax kink is introduced so that individuals above k face a tax rate $\tau_{t2} > \tau_{c2}$. In order to exploit this potentially attractive quasi-natural experiment, I assume the following:

Assumption 2a: *The change in potential outcome, ΔY, in the treatment and comparison groups is the same, on average.*

This assumption imposes that individuals above and below the tax kink in period 1 would respond in the same way to a tax rate change and, absent a tax rate change, their change in the outcome variable is the same, on average. Such an assumption is pervasive throughout the treatment effects literature. To make this assumption hold, in practice, the analysis

[3]Note that this creates a progressive income tax. If there was a tax decrease above k instead, creating a regressive tax schedule, some of the analysis would be different, as will be made clear at the end of Subsection 2.2.

is often restricted to individuals in a region around the tax kink because it is usually not appropriate to assume, for example, that those making several million dollars would have the same outcome as those making $20,000, absent a tax reform. To introduce that restriction, here, let all individuals in $[\underline{k}(1), \overline{k}(1)]$ be included in the estimation, where $[\underline{k}(1), \overline{k}(1)]$ are the thresholds $[\underline{k}, \overline{k}]$ determined by period 1 income.

We can estimate the causal average treatment effect ε as the difference in the change in taxable income between these two groups:

$$\varepsilon = \mathbb{E}\left[\Delta Y(t1) - \Delta Y(c1)\right], \tag{1}$$

where $\Delta Y(t1)$ is equal to ΔY multiplied by an indicator for being in the treatment group in period 1 and $\Delta Y(c1)$ is equal to ΔY multiplied by an indicator for being in the comparison group in period 1. Note that this is equivalent to the following:

$$\varepsilon = \mathbb{E}\left[Y(t1) - Y(c1) \mid T = 2\right] - \mathbb{E}\left[Y(t1) - Y(c1) \mid T = 1\right], \tag{2}$$

where T is a time indicator. This is also equivalent to defining the treatment and comparison groups in each period to get:

$$\varepsilon = \mathbb{E}\left[Y(t2) - Y(c2) \mid T = 2\right] - \mathbb{E}\left[Y(t1) - Y(c1) \mid T = 1\right], \tag{3}$$

if and only if no individual changes their taxable income, such that they cross k in response to the tax reform. It is a rather trivial statement—they are only equivalent if the categorization based on period 1 and period 2 income is the same—but it is crucially important given that equation (3) is the estimating equation used by the whole of the tax treatment literature that defines treatment as the observed tax rate change. Observe that equation (3) could be equally well implemented in panel and repeated-cross sectional data, and the miscategorization problems are the same for both.

Before addressing whether this is a reasonable assumption, and the bias induced when it fails, it is worth discussing a separate potential source of bias that is driven by the use of different forms of $[\underline{k}, \overline{k}]$. Using $[\underline{k}(1), \overline{k}(1)]$ introduces no bias because these cutoffs are based on pre-treatment income. Alternatively, we may restrict individuals' membership period by period, so that the restriction is still $[\underline{k}(1), \overline{k}(1)]$ in period 1, but in period 2 it becomes $[\underline{k}(2), \overline{k}(2)]$. Now, if there is any heterogeneity in the response to the tax rate, $t2$ will include all individuals who would have been in the sample based on period 1 income plus all individuals above $\overline{k}(1)$ who decreased their income enough in response to the tax rate change to be included based on period 2 income. Therefore, this parameter will be biased upwards relative to the true average treatment effect. In practice, the latter restriction is not implemented in analyses using panel data, but it is when conducting analyses using repeated cross-section data (and it is the only feasible restriction if panel data is not available). Therefore, when a researcher uses repeated cross-section data, an upper (lower) cutoff will yield biased estimates if treatment occurs above (below) the tax kink.[4]

Now, I return to the question of whether the assumption that no individuals cross the tax kink as part of their behavioral response to the tax reform is valid. My working example is a tax reform that introduces a tax kink and makes the tax schedule progressive. Either a tax kink that introduced a regressive tax schedule or a tax notch instead of a tax kink would clearly violate this assumption. If a regressive tax schedule is introduced, the budget set becomes convex and individuals are indifferent between points on both sides of the tax kink. If a tax notch is introduced, there is a discrete decline in the budget set (because there is a discrete increase in tax liability) above the notch providing very strong incentives for individuals near the notch to shift their income below the notch (Slemrod, 2010). However, returning to the working example in this section, the assumption is valid if individuals respond in a perfectly classical way. Classical economic theory would predict that individuals

[4]Note that such a cutoff could be included if an instrument was used that was uncorrelated with these individuals who select into the estimation, but it is unlikely the case in practice, since most instruments used are functions of pre-response income, which is higher for these individuals by definition.

do not cross the tax bracket line in light of a marginal tax rate change because, if they had preferred to be in the other tax bracket, they would have chosen to locate there in the period prior to the tax reform as well. Note that this classical analysis assumes that some individuals will stop earning positive amounts in the higher bracket after the reform, but all of these individuals will choose income $Y = k$; that is, they will all bunch perfectly at the tax kink. Therefore, these individuals are not counted as having changed brackets as long as the upper bracket is defined as $Y \geq k$.

However, it is well accepted in the literature that there are optimization frictions which violate the classical model. For example, a recent paper by Chetty (2009) uses the presence of optimization frictions to provide an explanation of the variation in the ETI estimates across different studies. Empirically, several different types of optimization frictions have been analyzed, including imperfect bunching and occupational switching. For perfect bunching to exist at the tax kink, individuals have to be perfectly attentive to the location of the tax kink each year and perfectly able to manipulate their taxable income precisely. However, everything we know anecdotally and empirically highlights that this is not the case in practice. Rather, bunching is imperfect. For example, Chetty et al. (2011b) find statistically significant bunching in Denmark. While Saez (2010) does not find statistically significant bunching at most tax kinks in the U.S. overall, the results in Chetty et al. (2011a) suggest that this is due to an inability to detect sharp bunching in aggregate, which does not imply that it does not exist within certain responsive subpopulations. Imperfect labor markets may well cause individuals to cross tax bracket lines in the event of a tax reform, as they alter the benefits associated with switching.[5] Note that for this and all other examples of adjustment costs, difference-in-differences is not valid in general, because the true treatment can never be measured—it is always some combination of present and future marginal tax rates. However, in cases where individuals do not switch brackets, the estimates are simply biased downwards because the assigned change in tax treatment between the treatment and

[5]Powell and Shan (2012) find evidence of occupational switching in the 1980's in response to the tax rate changes.

control groups is too large relative to the truth. When individuals do switch brackets as part of their response, the bias is more severe, because it appears that those making the largest tax changes are experiencing a tax rate change of the opposite sign relative to the truth.

With optimization frictions, some individuals who were above k in period 1, will respond to the tax rate change above k, and this response will alter their taxable income such that it is below k in period 2. Now equations (2) and (3) are no longer equivalent. Equation (2) still estimates the same causal average treatment effect because the groups were defined as a function of period 1 income, which was before any selection took place. However, the estimate given by equation (3) is biased towards zero because individuals who faced the high tax rate and responded by moving into the comparison group are now included in $Y(c2)$ instead of $Y(t2)$.

In the context of panel data, there is not a binary treatment representation for equation (3), because the treatment definition changes across periods. In reality, the treatment is a continuous variable—the change in the observed tax rate—but incorporating this measure of treatment into this analysis makes the analysis less transparent. Without loss of generality in this section, I define treatment as if it were treatment determined by period 2, $D(t2)$. It is without loss of generality because, in this simple setup, treatment in period 1 is accurately measured. Defining treatment in this way, we can rewrite equation (3) as:

$$\begin{aligned} \varepsilon &= \mathbb{E}\left[Y(t2) - Y(c2) \mid T = 2\right] - \mathbb{E}\left[Y(t2) - Y(c2) \mid T = 1\right] \\ &= \mathbb{E}\left[\Delta Y(t2) - \Delta Y(c2)\right]. \end{aligned} \tag{4}$$

To see the problem induced by treatment mismeasurement in period 2 and the effects of potential resolutions, I divide individuals i into four principal strata (Frangakis and Rubin, 2002) based on two potential income indicators $S_i(2)$ and $S_i(1)$:[6]

- $HH = \{i | S_i(2) = S_i(1) = 1\}$: individuals who choose income above k without a tax

[6]I define these groups assuming that the tax rate changes for those above k, which is the most common form of tax change; however, the results are equivalent if the groups are instead defined assuming that the tax rate changes below k.

rate change and have no incentive to deviate below k when the tax rate changes.

- $HL = \{i|S_i(2) = 0, S_i(1) = 1\}$: individuals who choose income above k without a tax rate change and face an incentive to deviate below k when the tax rate changes above k.

- $LH = \{i|S_i(2) = 1, S_i(1) = 0\}$: individuals who choose income below k without a tax rate change and face an incentive to deviate above k when the tax rate changes above k.

- $LL = \{i|S_i(2) = S_i(1) = 0\}$: individuals who choose income below k without a tax rate change and face no incentive to deviate above k when the tax rate changes.

The term "incentive to deviate" refers to all individuals who may wish to deviate when the tax rate changes, whether or not they are, in fact, responsive enough to choose to deviate. For example, in the case of imperfect bunchers, all individuals bunching just below the tax kink are potential deviants, whether or not they choose to have income above k after the tax reform. Defining the groups based on their incentive to deviate rather than their actual deviation will enable me to define a parameter that will have substantially more policy relevance.

Assumption 3: *When the tax rate increases above k, there should be no individuals of type LH and when the tax rate decreases above k, there should be no individuals of type HL.*

Assumption 3 requires that all individuals move in the appropriate direction in response to a tax change; that is, when the tax rate rises, no individuals respond by increasing their income. Let $dd = 1$ if an individual chooses to deviate and zero otherwise.

If $dd = 0$ all individuals, I could rewrite equation (4) as:

$$\tilde{\varepsilon} = [\mathbb{P}[HH = 1]\mathbb{E}[\Delta Y(t2)|HH = 1] - \mathbb{P}[LL = 1]\mathbb{E}[\Delta Y(c2)|LL = 1]], \tag{5}$$

and $\tilde{\varepsilon} = \varepsilon$. However, when $dd = 1$ for some individuals, equation (4) can be rewritten as:

$$\begin{aligned}\varepsilon = &[\mathbb{P}[HH = 1]\mathbb{E}[\Delta Y(t2)|HH = 1] + \mathbb{P}[HL = 1, dd = 0]\mathbb{E}[\Delta Y(t2)|HL = 1, dd = 0] \\ &-\mathbb{P}[HL = 1, dd = 1]\mathbb{E}[\Delta Y(c2)|HL = 1, dd = 1] - \mathbb{P}[LL = 1]\mathbb{E}[\Delta Y(c2)|LL = 1]] \,.\end{aligned} \tag{6}$$

The gap between the true average treatment effect and the actual estimand, $\tilde{\varepsilon} - \varepsilon$, which is induced by those who choose to deviate below k, can be quantified as $2\mathbb{P}[HL = 1, dd = 1]\mathbb{E}[\Delta Y(c2)|HL = 1, dd = 1]$. Recall that these are individuals who look as though they are comparison group individuals, but were, in fact, treated; therefore, this term is expected to be non-zero.

There are two ways of addressing the fact that ε does not equal $\tilde{\varepsilon}$: a proxy variable or an instrument. First, consider choosing a proxy variable. In this simplified example, there is a perfect proxy available—treatment status based on period 1 income. This proxy recovers the average treatment effect over the whole population given by equation (2). Observe that if this proxy was used as an instrument to construct estimates using a Wald estimator instead, the estimates would be biased upwards because the numerator of the Wald estimator would be the correctly estimated average treatment effect given by equation (2) and the denominator is not equal to one. It is instead given by:

$$\mathbb{E}[D(t2)|Z = 1] - \mathbb{E}[D(t2)|Z = 0] = 1 - 2\mathbb{P}[HL = 1, dd = 1] < 1, \tag{7}$$

where $Z = D(t1)$ is the instrument. This unusual result that the Wald estimator does worse at revealing the population average treatment effect than the reduced-form estimate is due to the fact that treatment is mismeasured and the Wald estimator is based on the assumption that $D(t2)$ is wrong and Z is right, not the other way around.

Now, consider an intermediate case in which a proxy is available, but it is imperfect.

Assumption 4a: *Suppose Z is an imperfect proxy that identifies an average treatment effect for a subpopulation of interest.*

Then, the following proposition highlights when this imperfect proxy will suffer from the same bias as the perfect proxy $D(t1)$, albeit to a lesser extent.

Proposition 1: *Given Assumptions 1, 2a, 3, and 4a, the reduced-form estimate will underestimate the average treatment effect for a given subpopulation. When*

$$\mathbb{P}[D(t2) \neq D(t1)|Z = 1] - \mathbb{P}[D(t2) \neq D(t1)|Z = 0] > 0, \tag{8}$$

the Wald estimator will overestimate the average treatment effect for the same subpopulation.

Proof: *See Appendix.*

Therefore, when Assumptions 1, 2a, 3, and 4a hold along with equation (8), the reduced-form estimates an intent-to-treat (ITT) effect where Z is an ITT indicator. The Wald estimator will not reveal the average treatment effect as we would like. Instead, it provides an upper bound on this parameter and the ITT estimate provides a lower bound. In words, equation (8) says that the Wald estimator will be biased upwards whenever more selection into the comparison group in period 2 occurs when the ITT measure Z is turned on. When equation (8) is instead less than zero, the Wald estimator also underestimates the average treatment effect.

The Wald estimator is not biased when equation (8) is equal to zero. One assumption that will guarantee this condition holds is given by:

Assumption 4b: *Let Z be a trivial function of the treatment indicator $D(\cdot)$ for each stratum except HH and LL.*

Put another way, Assumption 4b assumes that $D(t2)$ and Z are independent among groups of individuals with an incentive to deviate. The assumptions and proposition that follows examines the average treatment effect that is obtained when Assumption 4b holds.

Assumption 2b: *Let the difference in potential outcomes $\Delta Y(HH) - \Delta Y(LL)$ be the*

same for all individuals in strata HH or LL.

Assumption 2b revises Assumption 2a for this context and requires that all individuals with no incentive to deviate will respond in the same way to these treatments, on average. In the context of a tax reform, this condition requires that individuals below the tax kink would respond the same to the treatment if they were above and vice versa. This assumption is actually stronger than necessary. $\Delta Y(HH) - \Delta Y(LL)$ can vary across individuals with no incentive to deviate, but this variation must be independent of Z. A popular alternative to Assumption 2b is monotonicity.[7] This restriction would generate a LATE-style FBATE parameter, but I do not focus on this restriction, because instruments used in this literature are either not monotonic or grossly violate Assumption 4b.

Assumption 5: *Let the potential outcome $\Delta Y(\cdot)$ and the treatment indicator $D(\cdot)$ be jointly independent of Z for each principal stratum.*

Assumption 5 includes the standard instrument exogeneity condition, which has been a focal point of instrument selection in the tax reform treatment literature.[8] This also imposes the common assumption that the growth rate of Y in the absence of the tax reform must be the same above and below the kink.[9] However, neither of these restrictions are relevant until the next subsection when Assumption 1 is relaxed.

Proposition 2: *Given Assumptions 1, 2b, 3, 4b, and 5, the Fixed-Bracket Average Treatment Effect (FBATE) is obtained from the Wald estimator and is given by:*

$$\varepsilon_{FBATE} = \mathbb{E}[\Delta Y(HH) - \Delta Y(LL) | HH + LL = 1]. \tag{9}$$

Proof: *See Appendix.*

I term this parameter the Fixed-Bracket Average Treatment Effect (FBATE), because it is

[7] Monotonicity is the assumption used by Angrist and Imbens (1994) to obtain the Local Average Treatment Effect (LATE).

[8] In reality, despite the literature's general concern with this condition, many instruments used violate this condition. For example, see Weber (2011).

[9] Alternatively, additional variables could be used to control for the heterogeneous growth rate.

the average treatment effect for individuals with no incentive to cross tax bracket lines in response to a tax reform (e.g. those in strata HH and LL).

The implications of Propositions 1 and 2 also apply to repeated cross-section analysis, because the treatment mismeasurement problem in period 2 that is analyzed here applies equally well to the repeated-cross-section context. The instrument must be independent of the same subpopulations in period 2 regardless of whether the data is panel or repeated-cross-section and a good repeated-cross-section instrument will capture the same subpopulations in both periods. Often, the most substantive concern with repeated cross-section analysis is a change in the composition of the treatment and comparison groups between period 1 and period 2. Proposition 2 highlights that if the instrument is chosen properly to address treatment mismeasurement, the composition bias is also eliminated; that is, in the repeated cross-section context, treatment mismeasurement and composition bias manifest themselves amongst the same subpopulation and addressing the former also addresses the latter.

2.2 Causal Inference with Secular Changes in Tax Rates

This subsection revisits the propositions derived in the last subsection when Assumption 1, which was used to provide a stylized example but does not hold in practice, is relaxed. The key results still hold, but they are more nuanced and require new assumptions to address the additional complexities introduced once Assumption 1 is relaxed.

Without Assumption 1 in place, individuals face transitory income shocks and secular trends in income that will move them across tax bracket lines between periods, regardless of whether there is a tax rate change. These both may induce tax rate changes, but these changes are not expected to be exogenous. Often, these shocks are correlated with the outcome of interest and, in general, individuals always have an incentive to deviate in response to these tax changes. Responsive individuals who face a transitory increase (decrease) in their marginal tax rate this period will shift income out of (into) this period and into (out of) the following period.

For example, suppose there is a marginal tax rate change at \$70,000, the individual's permanent income level is \$65,000, and this individual receives a positive shock of \$6,000 this period, so that this individual's total income is \$71,000. Shifting \$1,000-\$5,000 of income into the next period minimizes tax liability. Only if the individual happens to choose \$1,000 will this response not induce a deviation across tax bracket lines. Note that if the individual's permanent income was instead \$68,000, the individual would no longer be able to avoid the higher marginal tax rate on all their income and the tax minimizing range of shifting across periods would be \$2,000-\$3,000. This analysis assumes the tax bracket is fixed at \$70,000 in both periods and there is no anticipated change in the legislated tax rates in period 2 (so individuals make decisions in period 1 as if there are no legislated tax rate changes in period 2). In this framework, individuals will have an incentive to deviate unless the transitory income shocks they receive do not cause them to move to a different tax bracket if they do not deviate. Because individuals who face changes in their tax rate due to transitory income shocks and secular income trends face an incentive to deviate, they will now be included in the strata HL and LH. If there is an anticipated tax reform, this may alter shifting incentives. This case is discussed in detail in Subsection 2.3.

Period 1 income is no longer a perfect proxy for treatment status, because some individuals face a new marginal tax rate in period 2 due to secular trends and transitory shocks. Therefore, researchers no longer observe true treatment status. Period 1 income is also not exogenous if transitory income shocks are serially correlated and the outcome variable of interest is a function of the variable that determines an individuals' location on the tax schedule (Weber, 2011), so even if this measure is being used as an imperfect proxy for treatment, an instrument is still needed to address its endogeneity. This is clearly true in the context of the elasticity of taxable income, where the change in taxable income is the outcome and the location on the tax schedule is also determined by taxable income.

To examine the causal average treatment effect that can be obtained when period 1 income is used to define treatment, consider the following variation of the four principal

strata considered in the previous subsection, which are now divided based on two potential income indicators $S_i(2)'$ and $S_i(1)'$. This version categorizes individuals exclusively based on incentives to deviate generated by transitory income shocks and secular trends:

- $HH' = \{i|S_i(2)' = S_i(1)' = 1\}$: individuals whose income is above k in period 1 and in period $2'$.
- $HL' = \{i|S_i(2)' = 0, S_i(1)' = 1\}$: individuals whose income is above k in period 1 and below k in period $2'$.
- $LH' = \{i|S_i(2)' = 1, S_i(1)' = 0\}$: individuals whose income is below k in period 1 and above k in period $2'$.
- $LL' = \{i|S_i(2)' = S_i(1)' = 0\}$: individuals whose income is below k in period 1 and in period $2'$.

Period $2'$ indicates income in period 2 excluding any behavioral response to tax rate changes, where the tax changes were either legislated or induced by a transitory income shock or secular trend.

Assumption 4c: *Let Z be a trivial function of D for each strata except HH' and LL'.*

Put another way, Assumption 4c assumes that D and Z are independent among groups of individuals who face transitory income shocks or secular income trends that would induce a tax rate change and thus provide them with an incentive to deviate.

Assumption 2c: *Let the difference in potential outcomes $\Delta Y(HH') - \Delta Y(LL')$ be the same, on average, for all individuals in strata HH' or LL'.*

Assumption 2c revises Assumption 2b for this context. The discussion of Assumption 2b also applies here.

Proposition 3: *Given Assumptions 2c, 4c, and 5, the following treatment effect is*

obtained from a Wald estimator when treatment status is defined by period 1 income:

$$\varepsilon_{p1} = \mathbb{E}[\Delta Y(HH') - \Delta Y(LL')|HH' + LL' = 1]. \tag{10}$$

Proof: *See Appendix.*

Note that ε_{p1} may include all individuals with an incentive to deviate due to the legislated tax rate change in period 2 just as the average treatment effect based on period 1 income (equation 1) did in the last subsection. The difference is that, in this subsection, ε_{p1} is identified for a subpopulation which does not include individuals with an incentive to deviate based on tax changes induced by transitory income shocks or secular income trends. Therefore, defining treatment in this way still has the possibility of identifying the parameter of interest for a larger subpopulation under weaker assumptions, than defining treatment as treatment status based on period by period income, which is considered next.

Treatment status could also be defined by observed income in each period. However, as in the last subsection, I will consider a simpler version of this (treatment based on period 2 income), which is without loss of generality for the results I wish to highlight in this section. To define the causal average treatment effect that can be obtained in this case, consider the following principal strata which combine the previous two sets of strata used to incorporate incentives to deviate from both legislated tax rate changes and tax rate changes due to secular income trends and transitory income shocks:

- $HH'' = \{i|HH_i = HH'_i = 1\}$: individuals whose income is above k in period $2'$ and who face no incentive to deviate below k in period 2.
- $HL'' = \{i|HL_i = 1 \; or \; HL'_i = 1\}$: individuals whose income is above k in period $2'$ and who face an incentive to deviate below k in period 2.
- $LH'' = \{i|LH_i = 1 \; or \; LH'_i = 1\}$: individuals whose income is below k in period $2'$ and who face an incentive to deviate above k in period 2.

- $LL'' = \{i | LL'_i = LL_i = 1\}$: individuals whose income is below k in period $2'$ and who face no incentive to deviate above k in period 2.

Assumption 4d: *Let Z be a trivial function of D for each stratum except HH'' and LL''.*

Put another way, Assumption 4d assumes that D and Z are independent among groups of individuals who face an incentive to deviate when their tax rate changes for any reason, legislated or otherwise.

Assumption 2d: *Let the difference in potential outcomes $\Delta Y(HH'') - \Delta Y(LL'')$ be the same, on average, for all individuals in strata HH'' or LL''.*

Assumption 2d revises Assumption 2b for this context, and the same discussion in that context also applies here.

Proposition 4: *Given Assumptions 2d, 4d and 5, a Fixed-Bracket Average Treatment Effect (FBATE) is obtained from the Wald estimator and is given by:*

$$\varepsilon'_{FBATE} = \mathbb{E}[\Delta Y(HH'') - \Delta Y(LL'') | HH'' + LL'' = 1]. \tag{11}$$

Proof: *See Appendix.*

The interpretation is similar to the Fixed-Bracket Average Treatment Effect obtained in the previous subsection. It is the average treatment effect for individuals with no incentive to cross a tax bracket line in response to a tax reform or tax change brought about by a shock in taxable income or secular income trend.[10]

Corollary 1: *When the assumptions for Propositions 3 and 4 hold simultaneously for a particular instrument Z, $\varepsilon_{p1} = \varepsilon'_{FBATE}$.*

[10]Note that this paper exclusively discusses average treatment effects for notational convenience. However, the results could all easily be applied to elasticities, which are commonly estimated in the literature by replacing the treatment indicators with the log net-of-tax rate faced.

Note that the assumptions required in order to obtain ε'_{FBATE} are stronger than those required to obtain ε_{p1}, in the sense that the instrument Z must be independent of all incentives to deviate, not just those associated with secular income trends and transitory income shocks. Given Corollary 1, one way to test whether the additional assumptions necessary to obtain ε'_{FBATE} hold is to use the same instrument with treatment defined as for ε_{p1}. If the two estimates are not statistically different and the shared assumptions are valid, we cannot reject the null hypothesis that the additional assumptions do, in fact, hold.

The discussion up to this point has assumed that there was no tax kink in period 1 and a progressive income tax in period 2. While this is sometimes accurate, there are also many tax reforms where the tax kink existed before the reform, and there are also occasional examples where part of the tax schedule is regressive. Introducing all these variations has no effect on Proposition 4, although it may change the number of potential deviants in strata HL and LH. Introducing these variations matters for Proposition 3 to the extent that these deviations introduce potential deviants that belong in the strata LH, because for these individuals, they will appear in the comparison group in period 1, but are responding to the tax rate change in the treatment group.[11] This introduces additional treatment mismeasurement into treatment status defined as a function of period 1 income. Leaving the strata defined as before will introduce a downward bias in the estimates if most of the mismeasurement occurs when $Z = 0$ and an upward bias otherwise. Alternatively, the strata can be revised to incorporate these incentives to deviate. This yields an average treatment effect for a subpopulation that is narrower than that originally found in Proposition 3 but still wider than that found by its analogue in Proposition 4.

I will not repeat the discussion in the last subsection, but it is worthwhile noting that, just as in Subsection 2.1, the implications in this subsection also apply to repeated-cross-section analysis. The discussion regarding the choice of the sample thresholds $[\underline{k}, \overline{k}]$ in Subsection

[11] Before the introduction of these variations, there were individuals with incentives to deviate of type HL, but the relevant group for these individuals was the treated group, so period 1 treatment assignment was correct.

2.1 also applies here. However, now there is an additional concern. Suppose the cutoff is a function of period 1 income. Then, around $\overline{k}$, some individuals who would be excluded except that they receive a negative transitory income shock in period 1 are included and some individuals who would be included except that they receive a positive transitory income shock are excluded. The reverse is true around $\underline{k}$. A similar story applies for secular income trends. For these cutoffs to not bias the estimates in the panel context, the instruments must be independent of the selection induced in the outcome of interest by using these cutoffs. When using period-by-period cutoffs with repeated-cross-section data, the same requirement applies, or the cutoff must induce the same bias in both periods (which is then netted out when the two periods are differenced).

2.3 Anticipated Tax Reforms

Anticipated tax reforms have been ignored up to this point and are the focus of this subsection. I discuss the challenges faced when examining anticipated tax reforms assuming that the researcher has decided to estimate a separate parameter which captures the response to the anticipated tax change. The discussion in this subsection applies equally well to a tax reform that is anticipated and an anticipated change in the tax schedule due to something like the loss of a dependent. Except in the most ideal (and likely unrealistic) situations, the anticipation of the tax reform creates additional incentives to deviate; often these incentives to deviate apply to a large portion of the population being analyzed and likely make it impossible to estimate a causal FBATE parameter of the response to the anticipated tax change. This is, unfortunately, the approach used to analyze the response to anticipated tax changes throughout the ETI literature, charitable giving literature, and elsewhere.[12]

As an example, consider the tax reform discussed in Subsection 2.1 and depicted in Figure 5. Suppose in period 1 individuals with taxable income above k learn that their marginal tax rate will decrease in period 2 due to a change in the tax schedule. Let the treatment effect

[12]For example, see Bakija and Heim (2011).

of interest be the change in the outcome between period 0 and period 1 in response to the anticipated tax change that takes place between period 1 and period 2. I consider a simple binary version of treatment, where the treatment variable D_A equals one when the measured anticipated treatment is not zero, and zero otherwise. As in the last subsections, considering this binary version of treatment makes the intuition clearer and the notation cleaner without loss of generality for the points I wish to make. The researcher will simultaneously control for any contemporaneous tax reforms using the methodology discussed in the previous subsections. Assume throughout this subsection that the estimation of that parameter is done correctly, although observe that additional incentives to deviate discussed in this section also introduce additional treatment mismeasurement into the contemporaneous treatment variable. If we conclude that it is not possible to obtain a causal estimate of the anticipated tax change, we will not be able to obtain a causal estimate of the contemporaneous change either.

The true anticipated treatment measured period by period is non-zero either because there is an anticipated change in the legislated tax rate between period 1 and period 2 or because an individual receives a transitory income shock in period 1 or period 2 that makes the tax rate different across the two periods. The former identifies the parameter of interest in this subsection. The latter has already been discussed in the context of Subsection 2.2. Recall from that discussion that all tax changes caused by transitory income shocks provide an incentive to deviate, so the instrument needs to be independent of these changes. The same requirement is needed in this subsection when estimating the effect of the anticipated treatment. Additionally, all individuals with an incentive to deviate either in response to the anticipated or the contemporaneous tax rate change create a treatment mismeasurement problem as before.

The relevant strata are now HH'', HL'', LH'' and LL'', where the membership in strata LH and HL now is also determined by incentives to deviate in response to anticipated legislated tax rate changes. Therefore, if the same conditions are satisfied for these strata,

Proposition 4 applies as before. The rest of this subsection focuses on who is now included in strata HL and LH. Given this, the feasibility of obtaining an FBATE estimate of the anticipated tax change is discussed.

Let R be the amount of taxable income the individual reports in each period and SH be the amount of income that can be shifted across two periods. When $SH = 0$, there will be no treatment mismeasurement because no shifting is possible. However, it makes no sense to estimate the response to D_A if $SH = 0$ because the response will be zero by construction. Therefore, I assume $SH > 0$ throughout this subsection.

Consider individuals that are in the treatment group in both periods absent a tax reform. If these individuals decide to respond in period 1 to the legislated tax change in period 2 depicted in Figure 5, they will attempt to shift as much of their income out of period 2 as possible up to $R = k$ and shift it into period 1. If they can shift smoothly (that is no one shifts to $R < k$), then there is no incentive to deviate. However, if perfect smoothing is not possible, this creates an incentive to deviate. We don't have clear evidence on the degree to which perfect smoothing across periods is possible, but if evidence from the static context, such as imperfect bunching, is any guide, imperfect smoothing exists. Unfortunately, that means that anytime $D_A = 1$, there is an incentive to deviate (at least within a reasonable region around the tax kink), and thus the parameter must be independent of all responses. Therefore, it is not possible to obtain a causal average treatment effect using a period by period measure of treatment.

Even if we assume perfect smoothing, more complicated tax reforms are problematic.[13] As an example, consider a case where there are only two tax brackets and the tax rate remains fixed across periods, but the location of the tax kink moves from \$60,000 in period 1 to \$70,000 in period 2. Let permanent income, absent a tax reform, be \$62,000. Individuals can minimize their tax liability across periods by shifting income into period 2 anywhere in

[13]This is somewhat in contrast to estimation of FBATE in the absence of anticipation, where a more complicated reform may introduce more chances to deviate, but does not eliminate the possibility of obtaining FBATE altogether.

the range \$2,000-\$8,000. Unless the individual chooses to shift exactly \$2,000 this creates treatment mismeasurement; therefore, these individuals have an incentive to deviate. More generally, all individuals with permanent income levels between the old and new tax kink location who can shift their income to avoid the higher tax rate in either period face an incentive to deviate. Therefore, the instrument would need to be independent of all individuals in this region.

As the tax schedule becomes even more complex (i.e. there is more than one kink), the requirements needed to obtain FBATE become even more rigorous. Suppose, for example, that the reform collapses multiple brackets at the top of the income distribution into a single bracket as in the Tax Reform Act of 1986 (TRA86). Let the marginal tax rate in this bracket in period 2 be lower than any of the marginal tax rates in period 1 that were collapsed in this tax bracket. Then, individuals' incentives to shift income into period 1 no longer end at the tax kink of their current brackets, but rather continues all the way down to the new highest tax kink. As a result, all individuals that face an incentive to cross tax bracket lines for this reason face an incentive to deviate. Thus, a valid instrument would have to be independent of all individuals in this region, which effectively rules out estimating a causal anticipation effect for everyone except individuals in the very top tax bracket. Even if an instrument satisfies this constraint, it is likely that Assumption 2d will fail due to the resulting dissimilarities of the two groups who are left that can be compared (i.e the treatment and comparison groups now come from quite different points in the income distribution, and this may lead to a variety of differences between the two groups besides the tax rate change).

With shifting income across periods, defining the anticipated treatment as a function of period 1 income (instead of period by period income) does not resolve the problem because treatment in period 1 is also often mismeasured (because individuals are shifting into or out of period 1). Instead, Proposition 3 would have to be applied to period 0 income. If period 0 income were used to define the anticipated tax rate change, it would also need to be used to

define treatment for the contemporaneous tax rate change because period 1 income is now also mismeasured for the contemporaneous tax change. Depending on the application, this may increase the variance of the estimates too much to be feasible.

3 Empirical Applications

This section applies the results to the existing empirical literature that attempts to estimate the behavioral response to a tax reform. This illustrates how the assumptions discussed in the previous section are applied empirically. It also highlights the likelihood that causal parameters, which can be interpreted as FBATEs, are being obtained in several sizable literatures that attempt to identify the behavioral response to a tax rate change using the Wald estimator.

Empirically, there is some evidence that an FBATE parameter can be obtained in the context of the ETI. For example, Weber (2011) shows that a large number of existing ETI instruments are endogenous as long as transitory income follows an autocorrelated process. She proposes the following related instrument: the predicted tax rate change as a function of income lagged two periods prior to the base year of the difference.[14] Suppose there is no anticipation of the tax reform. Weber (2011) provides evidence that the instrument exogeneity condition holds when the appropriate controls are used.

Verifying Assumption 4d is more difficult. For example, the instrument proposed by Weber (2011) would violate Assumption 4d if the behavior of individuals who bunch around the kink is relatively stationary over time; that is, individuals who were imperfectly bunched below the kink two periods ago are still there today. A similar concern could be raised regarding other optimization frictions. However, Weber (2011) shows that the differences between the estimates obtained using period 1 treatment status and actual treatment status are minimal, suggesting that the instrument is, in fact, doing a relatively good job of obtaining FBATE. Moreover, she finds that using treatment status based on period 1 income is associated with

[14]This instrument will be relevant as long as income two periods ago is indicative of income today.

a substantial reduction in standard errors because there is less treatment mismeasurement. This suggests that when Corollary 1 holds, using this alternative definition of treatment is preferable. Additionally, if the conditions for Proposition 3 are not met because of optimization frictions, it is likely that defining treatment as a function of period 1 income obtains a lower bound because the tax reform examined was primarily a tax decrease, so most of the individuals who suffered from treatment mismeasurement appear in the comparison group. To the extent that Proposition 3 fails, it is likely that most are in the instrument comparison group as well. The weaker form of Assumption 2d—that $\Delta Y(HH'') - \Delta Y(LL'')$ may vary across individuals within strata HH'' and LL'', but is independent of the instrument—is likely to hold in this context, because it is unlikely that income two periods ago predicts an individuals' responsiveness today.

Now consider another prominent empirical literature that estimates the behavioral response to a tax rate change—charitable giving. I consider a recent approach to examining this response, which estimates dynamic responses to contemporaneous and anticipated future changes in the marginal tax rate (Bakija and Heim, 2011). The estimates in this literature are restricted to the intensive margin; that is, individuals who itemize only because of positive charitable giving are excluded because of endogeneity concerns. This literature usually constructs the estimating equation in levels and employs year and individual fixed-effects. The results in this paper apply equally well to both this context and difference-in-differences, but to keep the discussion consistent, I will consider a simple hypothetical example in which there are only two years of data.[15] Crucially, the empirical specification controls for transitory taxable income shocks. The parameters that capture the effect of transitory taxable income shocks can never be properly identified because transitory income shocks are a function of the response; that is, anytime charitable giving changes in response to a transitory income shock, the magnitude of the observed shock changes. But let's set that issue aside.

First, consider the estimation of the response to contemporaneous changes in the marginal

[15]Then difference-in-differences and fixed-effect estimation are equivalent.

tax rate. The instrument used in this context is the change in the tax rate on the first dollar of charitable giving. The instrument exogeneity condition will hold if a secular decision to donate more to charity is independent of the tax rate faced for the first dollar of charitable giving. This is reasonable as long as other components of taxable income do not respond to this decision (which is an odd assumption to make because part of the premise of this estimation is that one expects that charitable giving will respond to shocks in other pieces of taxable income).

Assumption 4d will likely fail. Particularly concerning in this context are individuals who are categorized in strata HL' and LH' because of transitory income shocks. Charitable giving is likely a highly shiftable form of income. To the extent that these individuals use charitable giving and other forms of shiftable income to minimize their tax liability, substantial treatment mismeasurement is introduced. The instrument used is either perfectly correlated with these deviations or is a predictor of the individuals' responsiveness (and thus violates Assumption 2d). In particular, the instrument will exactly mirror the movement in the mismeasured treatment unless, without charitable giving, the marginal tax rate would change; that is to say, it is changes in charitable giving that push the individual over the tax bracket line. However, these individuals are highly responsive by definition, making the instrument a good predictor of the potential outcome $\Delta Y(HH'') - \Delta Y(LL'')$.

Now consider estimating the response to anticipated future tax changes for charitable giving. The instrument for the future tax change used by Bakija and Heim (2011) is tomorrow's tax rate as a function of today's income (i.e. it relies on future pre-announced changes and is not a function of tomorrow's change in taxable income). The tax reforms used to identify this parameter are complex; one of the reforms used is TRA86, which was discussed in Subsection 2.3. This means there are many incentives to deviate for a large portion of the population. Therefore, it is extremely unlikely that FBATE has been obtained for the anticipated response.

4 Discussion

This section discusses a wide range of broader implications of this paper. A wide range of topics are covered, including the degree to which FBATE is a relevant parameter for deadweight loss. I also use the results in Section 2 to highlight that other forms of identifying variation, such as a change in dependent or bracket creep, cannot identify a causal average treatment effect.

Given that the estimates in the literatures that attempt to estimate the causal effect of a tax rate change are often used to calculate deadweight loss, it is important to consider to what extent FBATE—the parameter obtained by Proposition 4—is actually the relevant parameter for policy analysis. Considering the example of the ETI, which applies more generally to many settings, Chetty (2011) shows that the bounds on the structural parameter relevant for welfare analysis are tighter when optimization frictions are low and marginal tax rate changes are high. When the potential outcomes split by the principal strata are homogeneous across all individuals, this parameter will be the relevant structural parameter for welfare analysis as long as the tax reform was large enough to induce individuals to overcome their optimization frictions (Chetty, 2011). However, when they are not homogeneous across all individuals, there are several things to note.

First, if those with an incentive to deviate face higher optimization frictions relative to the average, FBATE will provide tighter bounds on the welfare parameter than a simple average treatment effect. Put another way, if those with an incentive to deviate will eventually respond in the same way as those that do not, FBATE will provide tighter bounds on the welfare parameter than a simple average treatment effect. Second, larger legislated changes in marginal tax rates are more informative regarding the structural welfare parameter, but these same reforms induce more bracket crossing, and are thus less likely to satisfy FBATE. If Assumption 4 fails in a given context, there is a trade-off to consider when selecting the optimal size of the tax rate change. A larger marginal tax rate change will get closer to the structural parameter desired for welfare calculations among individuals that do not violate

Assumption 4, but the bias induced by the increase in individuals that violate Assumption 4 is larger.

Third, note that if the heterogeneity in the potential outcomes is not due to optimization frictions, but rather due to variations in underlying preferences, the elasticity estimates obtained are no longer guaranteed to be relevant for welfare analysis. This is because FBATE is independent of the response of those with an incentive to deviate who are now allowed to respond differently to a change in their marginal tax rate relative to those who are not potential deviants. For example, this would occur if individuals who bunch imperfectly would respond differently, on average, than individuals located further away from the tax kink if the imperfect bunchers found themselves further away from a tax kink.

Proposition 1 highlights new trade-offs between estimating reduced-form ITT estimates and a Wald estimate of the average treatment effect. While the assumptions used to generate Proposition 1 do not hold exactly in practice, the general point still applies. Usually, the purpose of constructing the Wald estimate is to rescale the ITT estimate to recover the average treatment effect. However, Proposition 1 suggests that in this context, if the ITT measure is not independent of the mismeasurement, the Wald estimate will likely not reveal the average treatment effect, and could substantially overstate the truth as the proxy becomes a better and better measure of actual treatment. That said, there are contexts in which it can be interpreted as an upper bound (and the ITT estimate provides a lower bound). Moreover, estimating ITT parameters avoids the relatively strong assumptions required to obtain FBATE. Ultimately, which method is preferred should be informed both by which parameter is expected to be more relevant for deadweight loss and whether picking an instrument that will allow FBATE to be obtained is feasible in a given setting.

Given that the instruments used are often likely correlated with individuals who bunch around the kink, a few more things about this issue should be noted. The degree of imperfect bunching is something that can be tested for using the methodology proposed in Saez (2010) and revised in Weber (2012). In the tax literature, it has become popular to estimate the

degree of bunching as a possible alternative way to estimate the ETI (for example, Saez 2010, Chetty et al. 2011b, or Weber 2012). Once substantial imperfect bunching has been documented, it is not appropriate to proceed with difference-in-differences estimation, unless an instrument is found that is independent of these individuals.[16]

When FBATE fails, the distance between FBATE and the estimate obtained is a function of the portion of the income distribution examined. The advantage of examining a narrow range of the income distribution is that assumptions regarding the similarity of potential outcomes between the treatment and comparison groups are more likely to hold. However, these are also the individuals who are most likely to face an incentive to deviate, because they are near the tax kink, and thus more often face incentives to cross it. As a result, including individuals further away from a given tax kink provides a trade-off when FBATE is not obtained between diluting the effect of violations of Assumption 4d and violating Assumption 5.

This paper has focused on tax reforms as identifying the causal effect of a tax rate change. Other sources of changes in the tax schedule, such as bracket creep[17] or a change in the number of dependents[18] have been touted in the literature as having the following advantage: "...one can compare taxpayers who are very similar both in income and initial marginal tax rate but yet face different prospects for changes in marginal tax rates and hence potentially make a much more convincing case for identification. The main drawback of this strategy is that taxpayers may not be aware of the minute details of the tax code...(Saez

[16] Although, note that it is possible that there is a reasonable degree of bunching, but relative to the whole population being treated by the tax reform, the group of individuals who would find it potentially optimal to bunch at kink points is small. In this case, these individuals will still bias the estimates, but their effect may be negligible relative to the overall estimate. Note that, even in this case, the individuals contributing to the bunching estimates are not the same individuals (hopefully) as those contributing to the estimates in the context of difference-in-differences. Therefore, although the estimates are likely similar, there is nothing to preclude the estimates from these two methods from being entirely different.

[17] In the U.S., the personal income marginal tax rate schedule was fixed in nominal terms until 1985. Saez (2003) uses this source of variation to estimate the ETI during 1979-1981, which was a period of about 10 percent inflation.

[18] This source of variation is used by Looney and Singhal (2006). They argue that the individuals they examine are likely not to respond to the future tax change before it is implemented. However, this identification remains similarly problematic to bracket creep unless individuals respond, but never by shifting their income below the tax bracket line, which obviously cannot be true.

et al., 2012)." From the perspective of this paper, such an identification strategy is even more fundamentally problematic. For example, consider using bracket creep as identifying variation. Now, the treatment is not zero only when an individual moves across the tax bracket line. As a result, individuals who wish to shift their income across time periods to minimize their overall tax burden or those who do not wish to earn income in the next bracket due to their labor-leisure preferences are less likely to be observed as treated. This will create a substantial downward bias in the estimates unless the instrument is independent of these incentives to deviate. But, unfortunately almost everyone faces an incentive to deviate given the narrow window examined on either side of the tax kink, so no causal parameter can be identified.

5 Conclusion

This paper has examined the conditions necessary to obtain a causal average treatment effect for the behavioral response to a tax change when it is identified by exploiting variation in the degree to which a tax reform affects different groups of individuals based on their individual characteristics and tax situations. The analysis has highlighted that more conditions are necessary to obtain a causal average treatment effect than were previously acknowledged by the literature. Satisfying these assumptions can often be relatively restrictive, leading to the identification of a parameter over a certain subpopulations. Even if a causal parameter is identified, researchers must carefully consider whether the parameter obtained is relevant for welfare or other policy analysis.

Choosing an alternative definition of treatment that is a function of base-year income allows the parameter to possibly be estimated over a larger subpopulation under weaker assumptions. In a similar vein, if a researcher has a reasonable measure of intent-to-treat in a given context, the researcher can often be better off using this intent-to-treat measure directly rather than rescaling by the fraction who were treated according to a measure of

observed treatment, even if the latter parameter is the policy relevant one. This result is unusual and exists in this context because treatment cannot be accurately measured for all subpopulations.

These results provide a new set of trade-offs regarding what is ideal. In addition to highlighting the trade-offs between a small and large tax reform, the benefits and drawbacks of different forms of identification, and so forth, they also bring up a more fundamental question. Are there contexts when using an instrumental variables strategy is not ideal for policy analysis? The answer is certainly yes, and this paper highlights many of the important points researchers should consider when asking whether this is the best approach for identifying their parameter of interest.

References

Angrist, J. D. and Imbens, G. W. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Bakija, J. and Heim, B. (2011). How does charitable giving respond to incentives and income? New estimates from panel data. *National Tax Journal*, 64(2, part 2):615–650.

Chetty, R. (2009). Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance. *American Economic Journal: Economic Policy*, 1(2):31–52.

Chetty, R. (2011). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, forthcoming.

Chetty, R., Friedman, J., and Saez, E. (2011a). Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. http://obs.rc.fas.harvard.edu/chetty/eitc_nbhd_slides.pdf.

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011b). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records. *Quarterly Journal of Economics*, 126(2):749–804.

Eissa, N., Kleven, H. J., and Kreiner, C. T. (2008). Evaluation of four tax reforms in the United States: Labor supply and welfare effects for single mothers. *Journal of Public Economics*, 92:795–816.

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

Frangakis, C. and Rubin, D. B. (2002). The defining rolse of 'principal stratification and effects' for comparing treatments adjusted for posttreatment variables: From treatment noncompliance to surrogate endpoints. *Biometrics*, 58:191–199.

Heckman, J. J. and Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30:239–267.

Looney, A. and Singhal, M. (2006). The effect of anticipated tax changes on intertemporal labor supply and the realization of taxable income. NBER Working Papers 12417.

Powell, D. and Shan, H. (2012). Income taxes, compensating differentials, and occupational choice: How taxes distort the wage-amenity decision. *American Economic Journal: Economic Policy*, 4(1):224–247.

Saez, E. (2003). The effect of marginal tax rates on income: A panel study of 'bracket creep'. *Journal of Public Economics*, 87:1231–1258.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.

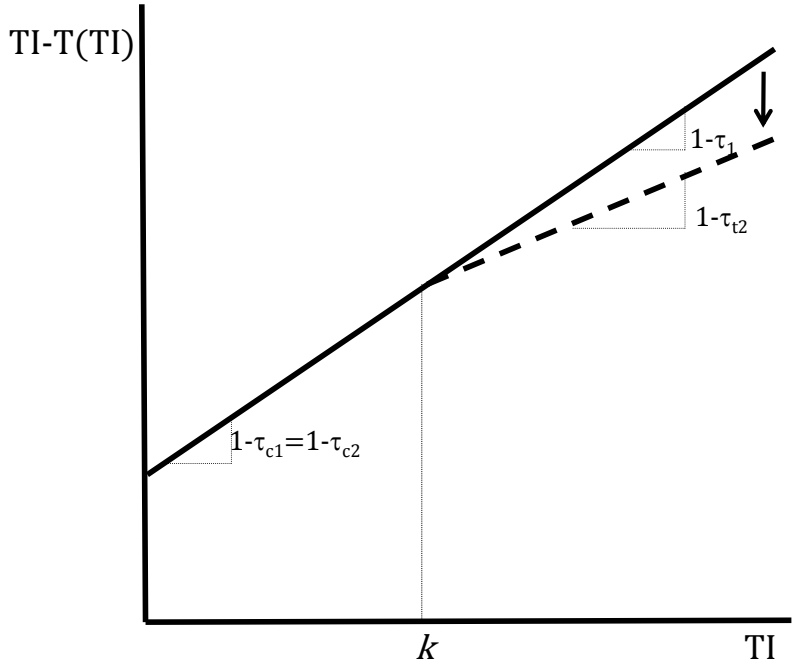
Saez, E., Slemrod, J., and Giertz, S. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, 50(1):3–50.

Slemrod, J. (2010). Buenas notches: Lines and notches in tax system design. In process.

Weber, C. (2011). Obtaining a consistent parameter of the elasticity of taxable income using difference-in-differences. University of Michigan Working Paper.

Weber, C. (2012). Does the Earned Income Tax Credit reduce saving of low-income households? University of Michigan Working Paper.

Figure 1: Tax Reform where Tax Rate Increases above k



Appendix

Proof of Proposition 1:

By definition, if Z is an imperfect proxy for $D(t1)$, the reduced-form estimate will underestimate the average treatment effect because it will miscategorize some individuals relative to their actual treatment status.

The Wald estimator will overestimate the average treatment effect whenever the denominator of the Wald estimator (which is always a fraction) is smaller than it should be based on actual treatment $D(t1)$. Mathematically, this condition is given by:

$$(\mathbb{E}[D(t2)|Z=1] - \mathbb{E}[D(t2)|Z=0]) - (\mathbb{E}[D(t1)|Z=1] - \mathbb{E}[D(t1)|Z=0]) < 0.$$

This can be rewritten as:

$$\begin{aligned}
&(\mathbb{P}[D(t2) = D(t1)|Z=1] \cdot \mathbb{E}[D(t2)|Z=1, D(t2) = D(t1)]] \\
&+ \mathbb{P}[D(t2) \neq D(t1)|Z=1] \cdot \mathbb{E}[D(t2)|Z=1, D(t2) \neq D(t1)] \\
&- \mathbb{P}[D(t2) = D(t1)|Z=0] \cdot \mathbb{E}[D(t2)|Z=0, D(t2) = D(t1)] \\
&-\mathbb{P}[D(t2) \neq D(t1)|Z=0] \cdot \mathbb{E}[D(t2)|Z=0, D(t2) \neq D(t1)]) \\
&- (\mathbb{E}[D(t1)|Z=1] - \mathbb{E}[D(t1)|Z=0]) < 0.
\end{aligned}$$

By Assumption 3, this can be rewritten as:

$$\mathbb{P}[D(t2) \neq D(t1)|Z=1] - \mathbb{P}[D(t2) \neq D(t1)|Z=0] > 0.$$

QED.

Proof of Proposition 2:

I begin by comparing $\mathbb{E}[\Delta Y|Z=z]$ *at* $z=0$ *and* $z=1$ *in period 2. I prove it for the case of a tax increase, but an equivalent proof would apply for a tax decrease. I cheat on notation at the beginning of the proof using* $D(HL,\cdot)$ *to indicate membership in strata* HL *rather than the actual treatment in each period.*[19] *By Assumption 3:*

$$\begin{aligned}
&\mathbb{E}[\Delta Y|Z=1]-\mathbb{E}[\Delta Y|Z=0]\\
&=\mathbb{E}[D(HH,1)\cdot\Delta Y(HH)+D(HL,1)\cdot\Delta Y(HL)+(1-D(HH,1)-D(HL,1))\cdot\Delta Y(LL)|Z=1]\\
&\quad-\mathbb{E}[D(HH,0)\cdot\Delta Y(HH)+D(HL,0)\cdot\Delta Y(HL)+(1-D(HH,0)-D(HL,0))\cdot\Delta Y(LL)|Z=0]
\end{aligned}$$

By Assumptions 4b and 5:

$$\begin{aligned}
&=\mathbb{E}[(D(HH,1)-D(HH,0))\cdot(\Delta Y(HH)-\Delta Y(LL))]\\
&\quad+\mathbb{E}[(D(HL,1)-D(HL,0))\cdot(\Delta Y(HL)-\Delta Y(LL))].
\end{aligned}$$

By Assumption 4b:

$$\begin{aligned}
&=\mathbb{E}[(D(HH,1)-D(HH,0))\cdot(\Delta Y(HH)-\Delta Y(LL))]\\
&=\mathbb{P}[D(HH,1)-D(HH,0)=1]\cdot\mathbb{E}[(\Delta Y(HH)-\Delta Y(LL))|D(HH,1)-D(HH,0)=1]\\
&\quad-\mathbb{P}[D(HH,1)-D(HH,0)=-1]\cdot\mathbb{E}[(\Delta Y(HH)-\Delta Y(LL))|D(HH,1)-D(HH,0)=-1].
\end{aligned}$$

By Assumptions 2b and 4b:

$$=\mathbb{P}[D(HH,1)-D(HH,0)]\cdot\mathbb{E}[\Delta Y(HH)-\Delta Y(LL)|HH+LL=1].$$

Then, it is obvious that the Wald estimator gives:

$$=\mathbb{E}[\Delta Y(HH)-\Delta Y(LL)|HH+LL=1].$$

[19] I do this because it saves notation overall, and by assumption, those in strata HL will drop out during the course of the proof.

Proofs of Propositions 3 and 4:

These proofs are identical to that from Proposition 2, replacing the strata from Proposition 2 with the appropriate strata for Propositions 3 and 4. Therefore, I do not repeat the proofs for these propositions here.