Three Essays in Taxation

by

Caroline E. Weber

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Economics) in The University of Michigan 2012

Doctoral Committee:

Professor Joel B. Slemrod, Chair
Professor James R. Hines, Jr.
Professor Jeffrey Andrew Smith
Assistant Professor Kevin Michael Stange
To Matthew my parents
ACKNOWLEDGEMENTS

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, Gerald Auten, Dora Benedek, David Cashin, Rob Garlick, Alexander Gelber, Susan Godlonton, Laura Kawano, Andreas Peichl, Emmanuel Saez, Hakan Selin, Sergio Urzua, Daniel Wilson, two anonymous referees, National Tax Association conference participants, International Institute of Public Finance conference participants, Michigan Tax Research Invitational participants, and numerous seminar participants for helpful comments. Also, thanks to Dan Feenberg for his help with the TAXSIM calculator. And, a special thanks to Emmanuel Saez for giving me the code used in Gruber and Saez (2002). Thank you to the Lincoln Institute for providing financial support. All errors are my own.
## TABLE OF CONTENTS

**DEDICATION** ................................................................. ii

**ACKNOWLEDGEMENTS** ...................................................... iii

**LIST OF FIGURES** ........................................................... vi

**LIST OF TABLES** ............................................................. vii

**CHAPTER**

I. Introduction ................................................................. 1

II. Identifying the Causal Effect of a Tax Rate Change When There are Multiple Tax Brackets ................................................................. 3

2.1 Introduction ................................................................. 3

2.2 Framework and Causal Inference ......................................... 5

2.2.1 Stylized Example ...................................................... 6

2.2.2 Causal Inference with Secular Changes in Tax Rates .......... 13

2.2.3 Anticipated Tax Reforms ............................................ 18

2.3 Empirical Applications .................................................. 21

2.4 Discussion ...................................................................... 23

2.5 Conclusion ..................................................................... 26

III. Does the Earned Income Tax Credit Reduce Saving by Low-Income Households? ................................................................. 31

3.1 Introduction ................................................................. 31

3.2 Background ................................................................... 32

3.2.1 EITC Details ............................................................. 32

3.2.2 Related Literature ..................................................... 34

3.3 Empirical Strategy .......................................................... 37

3.3.1 Data ...................................................................... 37

3.3.2 Research Design ....................................................... 39

3.3.3 Awareness .............................................................. 41

3.4 Estimating the Response at the Second Tax Kink ................. 42

3.4.1 Bunching ................................................................. 43

3.4.2 Changes in Average Income ....................................... 47

3.5 Difference-in-Differences Estimates ..................................... 51

3.5.1 Separating the Income and Substitution Effects ............... 59

3.6 Conclusion ..................................................................... 61
IV. Obtaining a Consistent Estimate of the Elasticity of Taxable Income Using Difference-in-Differences ...................................................... 82

4.1 Introduction ............................................................... 82
4.2 Background ............................................................ 84
4.3 Model Setup ............................................................ 86
4.4 Data and Estimation .................................................... 88
  4.4.1 Data ................................................................. 88
  4.4.2 Instrument Selection .............................................. 90
  4.4.3 Heterogeneous Growth Rates and Time Trends ............. 105
4.5 Conclusion ............................................................ 109

V. Conclusion ............................................................... 115

BIBLIOGRAPHY ............................................................. 117
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Tax Reform where Tax Rate Increases above $k$</td>
<td>27</td>
</tr>
<tr>
<td>3.1</td>
<td>EITC Schedule</td>
<td>63</td>
</tr>
<tr>
<td>3.2</td>
<td>EITC Schedule with Unearned Income</td>
<td>64</td>
</tr>
<tr>
<td>3.3</td>
<td>EITC Calculation Form in 2006</td>
<td>65</td>
</tr>
<tr>
<td>3.4</td>
<td>Estimating Elasticity from Bunching Around a Tax Kink</td>
<td>66</td>
</tr>
<tr>
<td>3.5</td>
<td>Estimated Density Around the Second EITC Kink for Self-Employed Individuals</td>
<td>67</td>
</tr>
<tr>
<td>3.6</td>
<td>Estimated Density Around the Second EITC Kink for Wage-Earning Individuals</td>
<td>68</td>
</tr>
<tr>
<td>3.7</td>
<td>Estimating Elasticity from Change in Average Income Around the Tax Kink</td>
<td>69</td>
</tr>
<tr>
<td>3.8</td>
<td>Estimated Average Earned and Unearned Income Around the Second EITC Kink</td>
<td>70</td>
</tr>
<tr>
<td>3.9</td>
<td>Estimated Average Income Around the Second EITC Kink</td>
<td>71</td>
</tr>
<tr>
<td>3.10</td>
<td>Variation in the Marginal Net-of-Tax Rate by Number of Dependents</td>
<td>72</td>
</tr>
<tr>
<td>3.11</td>
<td>Likelihood Individuals have Investment Income by Year and Number of Dependents</td>
<td>72</td>
</tr>
<tr>
<td>3.12</td>
<td>Individuals’ Responses to the Informational Reporting Threshold</td>
<td>73</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>EITC Schedule Parameters</td>
<td>74</td>
</tr>
<tr>
<td>3.2</td>
<td>Descriptive Statistics for EITC Eligible Taxpayers</td>
<td>76</td>
</tr>
<tr>
<td>3.3</td>
<td>Elasticities at Second EITC Kink</td>
<td>77</td>
</tr>
<tr>
<td>3.4</td>
<td>Repeated-Cross-Section Descriptive Statistics</td>
<td>78</td>
</tr>
<tr>
<td>3.5</td>
<td>Baseline Estimates</td>
<td>79</td>
</tr>
<tr>
<td>3.6</td>
<td>Robustness Checks</td>
<td>80</td>
</tr>
<tr>
<td>3.7</td>
<td>Substitution and Income Effects</td>
<td>81</td>
</tr>
<tr>
<td>4.1</td>
<td>Descriptive Statistics</td>
<td>111</td>
</tr>
<tr>
<td>4.2</td>
<td>Baseline Results</td>
<td>112</td>
</tr>
<tr>
<td>4.3</td>
<td>Alternative Definition of Treatment</td>
<td>113</td>
</tr>
<tr>
<td>4.4</td>
<td>Heterogeneous Income Trends</td>
<td>114</td>
</tr>
</tbody>
</table>
CHAPTER I

Introduction

Tax researchers often examine how individuals respond to marginal tax rate changes in order to assess the deadweight loss of a particular tax regime. However, the conditions under which we could estimate a causal parameter in this context, where we estimate the parameter by exploiting variation in the degree to which a tax reform affects different groups of individuals based on their individual characteristics and tax situations, has not been well understood.

Chapter 2 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). 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.

Chapters 3 and 4 use insights gained in Chapter 2 to empirically analyze the effect of taxation on two outcomes of interest—saving and taxable income. Chapter 3 analyzes the effect of the Earned Income Tax Credit (EITC) on investment income. Policy-makers have devoted substantial time and resources toward increasing the saving rate of low-income households with programs like the Saver’s Credit and Individual Development Accounts. Yet the EITC—the largest federal cash transfer program in the U.S.—provides a substantial disincentive for individuals to save and realize investment income because EITC benefits decline as investment income rises over certain income
ranges. I find that nearly 40 percent of the decline over the last two decades in the fraction of EITC recipients with savings in income-bearing accounts can be explained by changing EITC incentives.

Chapter 4 estimates the elasticity of taxable income (ETI). This paper shows that most estimators employed in the literature fail to obtain consistent estimates of the ETI. A new method is proposed that will provide consistent estimates under testable assumptions regarding the degree of serial correlation in the error term. Using this procedure, I estimate an ETI of 1.046, which is more than twice as large as the estimates found in the most frequently cited paper on this subject (Gruber and Saez, 2002). I also consider an alternative definition of treatment, which was proposed in Weber (2012b). It has a minimal effect on the estimates, but decreases standard errors by 16 percent.
CHAPTER II

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

2.1 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.¹

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

¹For example, Eissa et al. (2008) do this for the EITC, and Feldstein (1999) and Chetty (2009a) do this for the elasticity of taxable income (ETI).
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 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 2.2 lays out a framework for causal inference and derives FBATE in this context for panel data under certain assumptions. Section 2.3 provides several empirical applications of these results. Section 2.4 discusses broader implications for the literature given the results in Section 2.2. Section 2.5 concludes.

2.2 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 2.2.1 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 2.2.2. Subsection 2.2.3 extends the analysis in Subsection 2.2.2 to consider the estimation of an anticipated tax reform.

2.2.1 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 2.2.2 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 2.1. 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 \( k \).\(^2\) 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 \( \tau_{t1} = \tau_{c1} \). 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:

\(^2\)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 2.2.2.
Assumption 2a: The change in potential outcome, $\Delta Y$, 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 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 $[k(1), \bar{k}(1)]$ be included in the estimation, where $[k(1), \bar{k}(1)]$ are the thresholds $[k, \bar{k}]$ determined by period 1 income.

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

\[
\varepsilon = \mathbb{E} [\Delta Y(t1) - \Delta Y(c1)],
\]

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

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

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

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

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 (2.3) 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 (2.3) 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 \([k, \bar{k}]\). Using \([k(1), \bar{k}(1)]\) 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 \([k(1), \bar{k}(1)]\) in period 1, but in period 2 it becomes \([k(2), \bar{k}(2)]\). 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 \(k(1)\) 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.\(^3\)

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 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 (2009a) uses the presence of optimization frictions to provide an explanation of the variation in the ETI estimates across different studies.

\(^3\)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.
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.\(^4\) 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 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 (2.2) and (2.3) are no longer equivalent. Equation (2.2) 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 (2.3) 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 (2.3), 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

\(^4\)Powell and Shan (2012) find evidence of occupational switching in the 1980’s in response to the tax rate changes.
this way, we can rewrite equation (2.3) as:

$$
\varepsilon = E[Y(t2) - Y(c2) \mid T = 2] - E[Y(t2) - Y(c2) \mid T = 1]
$$

(2.4)

$$
= E[\Delta Y(t2) - \Delta Y(c2)]
$$

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

- $HH = \{i | S_i(2) = S_i(1) = 1\}$: individuals who choose income above $k$ without a tax 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 (2.4) as:

$$
\tilde{\varepsilon} = [P[HH = 1]E[\Delta Y(t2) \mid HH = 1] - P[LL = 1]E[\Delta Y(c2) \mid LL = 1]]
$$

(2.5)

5I 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 $k$. 

10
and $\bar{\varepsilon} = \varepsilon$. However, when $dd = 1$ for some individuals, equation (2.4) can be rewritten as:

$$
(2.6) \quad \varepsilon = [P[HH = 1]E[\Delta Y(t2)|HH = 1] + P[HL = 1, dd = 0]E[\Delta Y(t2)|HL = 1, dd = 0]
\quad - P[HL = 1, dd = 1]E[\Delta Y(c2)|HL = 1, dd = 1] - P[LL = 1]E[\Delta Y(c2)|LL = 1]].
$$

The gap between the true average treatment effect and the actual estimand, $\bar{\varepsilon} - \varepsilon$, which is induced by those who choose to deviate below $k$, can be quantified as $2P[HL = 1, dd = 1]E[\Delta Y(c2)|HL = 1, dd = 1]$. 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 $\varepsilon$ does not equal $\bar{\varepsilon}$: 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 (2.2). 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 (2.2) and the denominator is not equal to one. It is instead given by:

$$
(2.7) \quad E[D(t2)|Z = 1] - E[D(t2)|Z = 0] = 1 - 2P[HL = 1, dd = 1] < 1,
$$

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

$$
(2.8) \quad P[D(t2) \neq D(t1)|Z = 1] - P[D(t2) \neq D(t1)|Z = 0] > 0,
$$
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 (2.8), 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 (2.8) 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 (2.8) is instead less than zero, the Wald estimator also underestimates the average treatment effect.

The Wald estimator is not biased when equation (2.8) 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 \( \Delta Y(\text{HH}) - \Delta Y(\text{LL}) \) be the same for all individuals in strata \( \text{HH} \) or \( \text{LL} \).

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. \( \Delta Y(\text{HH}) - \Delta Y(\text{LL}) \) 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.\(^6\) 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.

---

\(^6\) Monotonicity is the assumption used by Angrist and Imbens (1994) to obtain the Local Average Treatment Effect (LATE).
Assumption 5 includes the standard instrument exogeneity condition, which has been a focal point of instrument selection in the tax reform treatment literature. 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. 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:

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

**Proof:** See Appendix.

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

### 2.2.2 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.

---

7In reality, despite the literature’s general concern with this condition, many instruments used violate this condition. For example, see Weber (2011).

8Alternatively, additional variables could be used to control for the heterogeneous growth rate.
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 2.2.3.

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
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:

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

**Proof:** See Appendix.

Note that $\varepsilon_{p1}$ 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 2.1) did in the last subsection. The difference is that, in this subsection, $\varepsilon_{p1}$ 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.
- \( HL'' = \{ i | HL_i = 1 \text{ or } HL'_i = 1 \} \): individuals whose income is above \( k \) in period 2' and who
  face an incentive to deviate below \( k \) in period 2.
- \( LH'' = \{ i | LH_i = 1 \text{ or } LH'_i = 1 \} \): individuals whose income is below \( k \) in period 2' and who
  face an incentive to deviate above \( k \) in period 2.
- \( LL'' = \{ i | LL_i = 1 \text{ or } LL'_i = 1 \} \): 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:

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

(2.11)

**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.\(^9\)

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

Note that the assumptions required in order to obtain \(\varepsilon_{FBATE}'\) are stronger than those required to obtain \(\varepsilon_{p1}\), 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 \(\varepsilon_{FBATE}'\) hold is to use the same instrument with treatment defined as for \(\varepsilon_{p1}\). 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.\(^{10}\) 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

---

\(^9\)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.

\(^{10}\)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.
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 2.2.1, the implications in this subsection also apply to repeated-cross-section analysis. The discussion regarding the choice of the sample thresholds $[\bar{k}, \overline{k}]$ in Subsection 2.2.1 also applies here. However, now there is an additional concern. Suppose the cutoff is a function of period 1 income. Then, around $\bar{k}$, 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 $\overline{k}$. 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).

### 2.2.3 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.\(^{11}\)

As an example, consider the tax reform discussed in Subsection 2.2.1 and depicted in Figure 2.1. 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 of interest be the change in the outcome between period 0 and period 1 in response to the anticipated tax change that

---

\(^{11}\)For example, see Bakija and Heim (2011).
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 2.2.2. 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 2.1, 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. 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 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

---

12This 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.
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.

2.3 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.\(^\text{13}\) 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 a substantial reduction in standard errors because there

\(^{13}\)This instrument will be relevant as long as income two periods ago is indicative of income today.
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.\footnote{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 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 2.2.3. 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.

### 2.4 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 2.2 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.
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 (2012a). 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 2012a). 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.  

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 or a change in the number of dependents 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 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

15Although, 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.

16In 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.

17This 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.
narrow window examined on either side of the tax kink, so no causal parameter can be identified.

2.5 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.
Figure 2.1: Tax Reform where Tax Rate Increases above $k$
Appendix

Proof of Proposition 1:

By definition, if $Z$ is an imperfect proxy for $D(t_1)$, 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 $D(t_1)$. Mathematically, this condition is given by:

$$(\mathbb{E}[D(t_2)|Z = 1] - \mathbb{E}[D(t_2)|Z = 0]) - (\mathbb{E}[D(t_1)|Z = 1] - \mathbb{E}[D(t_1)|Z = 0]) < 0.$$  

This can be rewritten as:

$$(\mathbb{P}[D(t_2) = D(t_1)|Z = 1] \cdot \mathbb{E}[D(t_2)|Z = 1, D(t_2) = D(t_1)]) + \mathbb{P}[D(t_2) \neq D(t_1)|Z = 1] \cdot \mathbb{E}[D(t_2)|Z = 1, D(t_2) \neq D(t_1)] - \mathbb{P}[D(t_2) = D(t_1)|Z = 0] \cdot \mathbb{E}[D(t_2)|Z = 0, D(t_2) = D(t_1)] - \mathbb{P}[D(t_2) \neq D(t_1)|Z = 0] \cdot \mathbb{E}[D(t_2)|Z = 0, D(t_2) \neq D(t_1)]) - (\mathbb{E}[D(t_1)|Z = 1] - \mathbb{E}[D(t_1)|Z = 0]) < 0.$$  

By Assumption 3, this can be rewritten as:

$$\mathbb{P}[D(t_2) \neq D(t_1)|Z = 1] - \mathbb{P}[D(t_2) \neq D(t_1)|Z = 0] > 0.$$  

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.\(^{18}\) By Assumption 3:

$$\mathbb{E}[\Delta Y | Z = 1] - \mathbb{E}[\Delta Y | Z = 0]$$

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

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

By Assumptions 4b and 5:

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

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

By Assumption 4b:

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

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

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

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:

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

\(^{18}\)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.
CHAPTER III

Does the Earned Income Tax Credit Reduce Saving by Low-Income Households?

3.1 Introduction

Policy-makers have devoted substantial time and resources toward increasing the saving rate of low-income households with programs like the Saver’s Credit and Individual Development Accounts. Yet the Earned Income Tax Credit (EITC)—the largest federal cash transfer program in the U.S.—provides a substantial disincentive for individuals to save and realize investment income because EITC benefits decline as investment income rises over certain income ranges. Over the last two decades, an average of 17.6 percent of low-income individuals that claim the EITC have some dividend and interest income, but strikingly, the fraction has declined by more than 50 percent over time, from 26.2 percent in 1988 to just 12.3 percent in 2006. In this paper, I determine the extent to which the disincentive to save created by the EITC has contributed to the decline in saving in income-bearing accounts. I use Individual Public Use Tax Files for 1988-2006, which contain data on a random stratified cross-sectional sample of individual income tax returns.

This paper uses two different sources of variation in the implicit EITC tax rate in the region in which the EITC benefit is declining (known as the phase-out tax rate) to examine the degree to which investment income falls as the implicit phase-out tax rate rises. The first identification strategy examines a narrow region around the second EITC tax kink. When an individual is below this tax kink there is no implicit tax on a marginal dollar of income whereas the phase-out tax rate applies on the margin to all types of adjusted gross income (AGI)—earned income, non-labor income, and deductions—above the second tax kink for most individuals. The most prevalent type of non-labor income is investment income. I find that AGI responds to the phase-out tax rate, and that this response is statistically significant for both wage-earners and the self-employed. For wage-earners, this response is concentrated in non-labor income, particularly investment income.
Investment income declines substantially above the tax kink, where the after-tax return to saving is lower. The implied elasticity of investment income with respect to the after-tax return to saving, which bounds the true elasticity from below, is 1.88.

The main estimates come from a difference-in-differences specification, which exploits differences in tax rate changes between taxpayers with differing numbers of dependents for identification. The extensive margin estimates imply that nearly 40 percent of the decline over the last two decades in the fraction of EITC recipients with savings in income-bearing accounts can be explained by changing EITC incentives. The elasticity of investment income is 3.05; that is, a one percent increase in the after-tax return to saving increases the amount of investment income by 3.05 percent. Given the increase in the phase-out tax rate over this period, the elasticity estimates imply a $1,776 decline at the mean in the annual amount in savings for those that held positive investment income in 1988-1990. This calculation assumes a five percent rate of return on saving, on average; if it is lower, then the decline will be even higher. A key advantage of exploiting both types of variation is that the assumptions required in order to obtain consistent estimates are roughly orthogonal in the two cases. The fact that the results from both are similar provides additional credibility to the estimates obtained.

The paper proceeds as follows. Section 3.2 provides relevant background information regarding the EITC as well as a brief literature review. Section 3.3 discusses the data and provides an overview of the empirical strategies. Section 3.4 conducts the estimation around the second tax kink in the EITC schedule. Section 3.5 conducts the difference-in-differences estimation, which provides the main estimates in this paper. Section 3.6 concludes.

3.2 Background

3.2.1 EITC Details

The EITC, which was introduced in 1975, has expanded over time and is now the largest federal cash transfer program in the U.S. The EITC is administered through the tax system, which is not true of other major cash transfer programs in the U.S. This means that the credit is received as a lump sum annually and is based on annual income, whereas other transfer programs are administered at a monthly frequency.\(^1\) The federal EITC is refundable, which allows individuals to receive the benefit even if it is larger than the taxes they owe. Some states also offer a credit, which usually

\(^1\)Individuals can receive their benefits at a monthly frequency using the Advance EITC, but very few people choose this option.
takes the form of a “piggy-back” credit (meaning that it is calculated as a certain percentage of the federal credit). The state credit is usually refundable.

When determining the EITC amount, several types of income are relevant. Earned income is defined as wage and salary income plus self-employment income (minus one-half the self-employment tax in the years 1990-2006), where self-employment income includes income from a sole proprietorship or a farm.\(^2\) Non-labor income includes categories such as investment income and unemployment insurance receipts; for a complete list, see Section 3.3.1. Deductions have varied over time; examples include alimony paid, moving expenses, and IRA contributions.\(^3\) I will refer to non-labor income minus deductions as unearned income. AGI, which is a term used on the tax form, is the sum of earned and unearned income.

The EITC schedule consists of three regions—the phase-in region, the plateau region, and the phase-out region—and is depicted in Figure 1. The example shown in Figure 3.1 is for taxpayers with two children, and reflects the tax schedule in real dollars for unmarried taxpayers in 1996-2006 (it also applies for married filers in 1996-2001).\(^4\) When an individual has no earned income, the EITC credit is zero. In the phase-in region, the implicit marginal tax rate is negative (the benefit is increasing); in the plateau region, the implicit marginal tax rate is zero (the benefit is constant); in the phase-out region, the implicit marginal tax rate is positive (the benefit is declining). When the end of the phase-out region is reached, the credit is again zero. The solid line is the EITC schedule as depicted in the existing EITC literature. It applies anytime an individual has no or negative unearned income, which is true for 75 percent of two-child EITC filers in 1996-2006.

When an individual has positive unearned income, it is included as income for the purposes of calculating the EITC anytime AGI is above the second kink (i.e. the individual is in the phase-out region). When AGI is above the second kink, a filer’s EITC is the minimum of the benefit as determined by their earned income and the benefit given by their AGI. In this two-dimensional graph, this can be depicted by shifting the solid phase-out region line towards the dashed line (or even past it). The phase-out region is shifted in for 25 percent of two-child filers in 1996-2006; the median amount of unearned income for these taxpayers is $1,227. Approximately one-fifth of these individuals have unearned income that is sizable enough such that they face no plateau region at all, facing a spike instead. Note that the location of the second kink appears to be person-specific only

\(^2\)There are a few minor adjustments to the measure of self-employment income for a small number of individuals, but these cannot be taken into account in this paper because the tax return data do not provide these details.

\(^3\)The term deduction refers to “above-the-line” deductions; that is, it does not include those only available to individuals that itemize. Also, since one-half the self-employment tax is included in earned income, it will not be counted as a deduction in this paper (otherwise it will be counted twice, once in earned income and once in unearned income).

\(^4\)The tax schedule is inflation-indexed in this period.
because the horizontal axis in Figure 3.1 is earned income, while AGI is the variable determining the EITC benefit received for most individuals. Figure 3.2 further illuminates the interaction between earned and unearned income in determining an individual’s EITC by plotting a three-dimensional graph in earned income, unearned income, and EITC benefit. As earned income rises, the EITC first rises and then falls. As unearned income rises, it first has no effect on the EITC and then causes it to decline at a certain threshold, where the threshold varies depending on the individual’s earned income amount.

Table 3.1 lists historic federal EITC parameters for taxpayers with at least one qualifying child.\footnote{The state “piggy-back” credits followed a similar expansionary trend. Prior to 1997, at most six states offered an EITC credit in any given year. The state EITC expanded in terms of generosity and the number of states offering the credit in the years 1997-2006. In 2006, 19 states offered a state EITC for families with children (for most states the same EITC applies for those without children) and the average rate across states was 15.5 percent. Fifteen of these states had a refundable credit. Conditional on having at least one child, only Wisconsin varies the percentage of the federal EITC benefit given depending on the number of dependents.}

Households without children became eligible for the EITC after 1994, but the phase-out region occurs at a much lower income level ($6,710 in 2006 dollars), so I do not anticipate these individuals to have a substantial investment income response (or to have substantial investment income). They also are a relatively poor comparison group for families with children. Therefore, these individuals are not part of the analysis that follows.

### 3.2.2 Related Literature

There is a substantial literature on low-income saving decisions, but the effect of the EITC on saving behavior has never been examined beyond the role played by the lump sum nature of its distribution. This section provides a brief overview of existing work on the EITC and saving decisions of low-income households as it relates to this paper.

The EITC literature has focused almost entirely on the effect of the EITC on labor supply, the source of true earned income.\footnote{There is, though, an emerging literature that has looked at other outcomes. Examples include the EITC’s effect on fertility (Baughman and Dickert-Conlin, 2003), time use (Gelber and Mitchell, 2012), and children (Dahl and Lochner, 2010).} The overall effect of the EITC on labor supply is theoretically ambiguous because the phase-in region encourages work and the phase-out region discourages work on the margin. The literature generally finds that the EITC increases work, mostly on the extensive margin, for single individuals.\footnote{See Hotz and Scholz (2003) for an extensive review of the EITC literature through 2002.} True earned income may not equal earned income reported to the tax authorities. Studies that make use of tax return data distinguish between wage-earning and self-employed individuals. These studies find that, while the intensive margin response by wage-earning individuals is approximately zero, there is a substantial response among the self-employed...
and there is a larger response to the phase-in region (LaLumia, 2009; Saez, 2010). This response is a combination of labor supply, tax avoidance, and tax evasion responses.

The literature that uses audit data to examine the likelihood that qualifying individuals claim the EITC and the literature on tax evasion as it pertains to the EITC both have some relevance for interpreting the results in this paper given that tax return data are used. Blumenthal et al. (2005) use data from the Taxpayer Compliance Measurement Program (TCMP) on both filers and non-filers in 1988, which contain audit data for a stratified random sample of these individuals. The authors estimate that 89 percent of EITC individuals above the filing threshold claimed the EITC credit if eligible.\(^8\) Thus, most EITC-eligible individuals in the region I examine will be observed in the tax return data.

McCubbin (2000) finds that claiming an extra dependent to which the taxpayer is not entitled increases with EITC generosity (a one percent increase in EITC generosity increases the likelihood that a taxpayer reports an extra child by about four percent) and decreases with income using IRS Criminal Investigations Division random audit data for 1994. Given this, I will consider possible biases induced in the estimates in Section 3.5 if individuals manipulate the number of children they report.

There is not much evidence that wage-earners evade by adjusting their earned income, but there is evidence that self-employed individuals do. Joulfaian and Rider (1996) examine the responsiveness of income reported to the EITC marginal tax rate using 1988 audit data from the TCMP. They find that 94.6 percent of wage-earning individuals accurately report their earned income, and the EITC rate has no effect on the decision to report accurately. For sole proprietors, 19 percent report accurately and they find that a one percent higher EITC rate increases the probability of underreporting by 0.3 percent.

There is a literature on the saving decisions of low-income households, but not about the role played by the EITC.\(^9\) One strand of this literature examines the effects of other need-based programs, such as AFDC/TANF and Medicaid, on savings. Hubbard et al. (1995) show that these programs affect savings decisions both because they decrease the need for precautionary savings in case of a negative income shock and the asset limits imposed by these programs create a high implicit tax

---

8 The fraction was substantially lower if the taxpayer was below the filing threshold. The second EITC kink is near (within $1,000) or above the filing threshold for all years.

9 The only evidence of the effect of the EITC on saving decisions is ethnographic and it is interested not in the effect of the EITC schedule, but rather in the lump sum nature of its distribution. Romich and Weisner (2000) conduct extensive interview-based analysis of 42 EITC recipient families in Wisconsin. They find that individuals prefer to receive the credit as one lump sum, rather than using the advance payment option. They find that 32 percent of the individuals in their sample had saved some of their EITC credit in a bank account and kept it in the account at least two months after the receiving their EITC credit.
rate when binding. Gruber and Yelowitz (1999) find empirical evidence supporting these theoretical results in the context of Medicaid. In contrast, recent papers examining the effect of AFDC/TANF asset limits (Hurst and Ziliak, 2006; Sullivan, 2006) on liquid assets, such as saving accounts, find no substantial effect. The EITC changes the after-tax return on savings for individuals in the phase-out region, providing a mechanism to decrease savings in addition to asset limits or precautionary savings. This mechanism will be the primary focus of this paper, but precautionary savings motives will be discussed when relevant in Section 3.5. The asset limits are much more likely to be binding in the case of TANF or Medicaid because the limits are effectively much lower for TANF and Medicaid than the limit introduced for the EITC in 1996.

A separate literature examines the effectiveness of programs designed to induce low-income individuals to save more. In addition to providing evidence on the degree to which individuals can be encouraged to save more, this literature also highlights a recent policy focus on increasing low-income saving. One popular program is Individual Development Accounts (IDA), which encourage individuals to save by providing matching funds (up to a limit each year) for income saved in these accounts. The individuals receive the matching funds when they withdraw their savings for qualified purposes (e.g. buying a house or paying for college). Over 400 IDA programs existed by 2006, which were privately funded or funded by states (Mills et al., 2008). Mills et al. (2008) examine a particular IDA program in Oklahoma with a randomized design; they find a large IDA take-up rate among the treated group (almost 90 percent). Those with IDA accounts are more likely to purchase a home, but this is somewhat offset by a decrease in other forms of asset accumulation (this IDA provided a larger match when the funds were withdrawn for the purpose of buying a home).

Another program, the Saver’s Credit, was implemented through the tax system in 2002 to encourage saving for retirement among low-income individuals. The literature has found little impact on saving of this credit; for example, see Ramnath (2010). Duflo et al. (2006) conduct a randomized field experiment of a program similar in spirit to the Saver’s Credit, and find that they can indeed induce individuals to save more. They cite differences between their experiment and the Saver’s Credit, such as the complexity and non-refundability of the Saver’s Credit, as likely to be driving the difference in findings between their experiment and the studies that examine the Saver’s Credit.

The existing EITC literature has found little effect of the EITC on labor supply decisions along the intensive margin (except perhaps for self-employed individuals). My paper examines the extent to which it has an effect on another margin—non-labor income, particularly investment income—which has not been previously examined. Given the existing empirical literature on low-income saving, the degree to which the EITC will affect low-income saving is not obvious, since the effects
of other programs on low-income saving is mixed.

3.3 Empirical Strategy

This section provides an overview of the empirical strategy used in this paper. Because the methods used in each section are quite different, the details can be found in their respective sections. Subsection 3.3.1 provides a detailed description of the data. Subsection 3.3.2 explains the focus of this paper, the degree to which a change in saving in an income-bearing account reflects an actual change in savings, and the different forms of variation used for identification in each section. Subsection 3.3.3 discusses the central role of awareness in studying the impact of the EITC on saving, which is key to interpreting the estimates.

3.3.1 Data

This paper uses the Individual Public Use Tax Files for 1988-2006, which were created by the Statistics of Income Division of the Internal Revenue Service. The sample is stratified, so all estimates are population weighted. Some components of AGI—wage and salary income, alimony paid, and alimony received—are blurred in the tax return data for low-income individuals in the years 1996-2006. Alimony income was blurred at the national level. Wage and salary income was blurred at the national level in 1996, at the state level for years 1997-1999, and at the state x married level for years 2000-2006. The blurring was done as follows: “for every three records, in descending order, the average...[of a given income type to be blurred] has been determined and that value has been placed in the...field for each of the three records.” (Weber, 2001). This means that earned income for all individuals that have positive wage earnings is measured with noise in the years in which blurring occurred. An alternative measure of earnings can be constructed by backing out earned income from AGI. This measure is also imperfect because most, but not all, types of unearned income are reported in the tax return data, and those that are also blurred must be excluded. I use the latter measure for all years throughout the paper, so the accuracy of my imputation method can be precisely determined. In the years before blurring, the latter measure is within $15 of the actual wage earnings for 99.5 percent of taxpayers for the sample used in Section 3.5. In the years after blurring, the blurred wage earnings is within $15 of the latter measure 65 percent of the time, indicating that there is substantial noise in the blurred wage earnings. Still, the estimates are robust to the use of either measure; for example, the baseline estimates in Section 3.5 decline by about 2

---

10 Alimony paid and received were blurred beginning in 1995.
11 The latter affects a relatively small number of individuals in the sample used in this paper.
percent when the blurred wage earnings is used instead.

Table 3.2 provides descriptive statistics for all EITC-eligible individuals for 1988-2006. The first two rows summarize the components of earned income. Almost 16 percent of individuals have some self-employment income, but it is the sole source of earned income for only 1.6 percent of taxpayers. The third row provides summary statistics for total unearned income, and the next nine rows break out unearned income by type.\textsuperscript{12} Note that 33.6 percent of individuals have some unearned income, with a mean of $571.66. The most frequent type of unearned income is dividend and interest income (18.6 percent). The next three rows—capital gains or losses, partnership income, and other gains or losses—reflect other types of income (or loss), but less than five percent of individuals have these types of income. Taxable refunds only apply to those who itemize deductions and are reported by a small fraction of individuals (3.2 percent). Taxable pension, annuity and IRA distributions are only received by about five percent of the sample and, unlike the previous categories, are unlikely to respond unless individuals adjust the amount received each year, decreasing it if they are temporarily EITC-eligible. Almost no one has taxable Social Security benefits, which also are not expected to be adjusted in response to EITC schedule changes. A substantial number of individuals have unemployment insurance (10.7 percent). Individuals will not adjust this form of income directly in response to the EITC schedule (because it does not pay to give up a dollar of unemployment insurance to obtain 10-20 cents of EITC benefits), but it may be adjusted indirectly if individuals change the number of weeks worked in response to the EITC. About five percent of individuals have deductions and the mean is $103.94; deductions are likely responsive to the EITC schedule. About 30 percent of the sample is married, and over 60 percent use a tax preparer. Conditional on filing a tax form, approximately 92 percent of EITC-eligible individuals claim the EITC.\textsuperscript{13} As noted in Subsection 3.2.1, I exclude those with no dependents. For most of the analysis in Section 3.5, I restrict the sample to those with one and two dependents; for comparability, I impose the same restriction throughout the paper. The mean number of dependents is 1.41.\textsuperscript{14}

The analysis of tax return data has several advantages. Most importantly, it provides a precise (or near precise) measure of the reported amount of each type of income, and is not subject to

\textsuperscript{12}Note that unearned income does not include blurred sources (alimony) and unreported sources (other income and a few deductions). Less than one percent of individuals in this sample have alimony income. In the years before the blurring of low-income wage data was instituted, I can back out unearned income directly by subtracting earned income from AGI. The likelihood that individuals have unearned income in these years (1988-1995) is 1.4 percentage points higher using this measure.

\textsuperscript{13}The actual number is likely slightly higher, because I cannot perfectly measure eligibility with the data available.

\textsuperscript{14}Dependents do not always correspond exactly to the number of children claimed for EITC purposes; the latter is not in this data set for all years. For years in which both are available, the actual measure is about one-tenth of a child lower. In addition to data availability, the advantage of using dependents is that it avoids additional manipulation of claiming children only for the purpose of maximizing the EITC. For years in which I have both measures, the baseline estimates decline slightly when the actual measure is used, but the estimates are not statistically different.
survey non-response, rounding, and other approximations. The main disadvantage of tax return data is that one cannot observe where individuals place the money that they would have held in an income-bearing account in the absence of a positive tax rate or those that do not file. For welfare, both moving the money out of the income-bearing account (assuming this is not costless to the individuals) and a real decline in saving are relevant; this data can only address the former. Tax evasion on investment income (i.e. not reporting some or all investment income) is also a potential drawback, although I will be able to provide some evidence below that evasion is likely not a key mechanism behind the results found.

3.3.2 Research Design

The key research question examined in this paper is whether the design of the EITC has the unintended consequence of distorting non-labor-income decisions in the phase-out region, particularly the decision to save in an income-bearing account. This paper chooses to focus mostly on the decision to save in an interest- or dividend-bearing account. I also analyze a broader measure of investment income, which includes capital gains and losses, partnership income, other gains and losses, and IRA contributions (IRA contributions are a deduction). I focus on a particular part of unearned income because unearned income contains a broad range of different types of income. As a result, the coefficient on unearned income does not have much policy relevance outside of choosing an optimal EITC design. In contrast, examining the effect of the EITC on saving of low-income households fits into the substantial literature on the effect of mean-tested programs on savings. Furthermore, unearned income contains both types of income one would expect to be responsive to the tax rate (e.g. dividend and interest income) and types that one would not (e.g. Social Security benefits).

The decision to save in an interest- or dividend-bearing account is related, but not identical, to the decision to save, as the EITC could induce individuals to stop saving in interest- or dividend-bearing accounts, instead placing their money in an interest-free account. This paper has nothing to say about this mechanism, but this response is distortionary and presumably not a behavior policymakers would like to encourage. There are several other reasons why the measured response may not be identical to the decision to save. First, if two taxpayers are cohabiting, one could save while the other claims the child and gets the EITC. This paper cannot provide evidence of the degree

---

15Capital gains are not included in the primary measure of the return on saving in an income-bearing account because they are a very lumpy measure of asset accumulation, and there is a large capacity to re-time the sale of an asset to a period in which the tax rate is lower. In reality, very few individuals in this range of the income distribution have capital gains, and I will show that the estimates are not affected by the inclusion of capital gains.
to which the measured response is due to this mechanism, because I cannot observe tax returns of
two separate cohabitators. Second, individuals could save in an interest- or dividend-bearing account,
but choose not to report it on their tax return—tax evasion. I argue that tax evasion is unlikely
because there is information reporting by financial institutions on interest and dividend income
above $10 per account, but it is possible, particularly for those with dividend or interest income
less than $10. I will return to the issue of information reporting and tax evasion in Section 3.5,
where I provide evidence that, under reasonable assumptions, any evasion of taxes on investment
income has a minimal effect on the results. Going forward, much of the discussion will proceed as
though individuals are altering their real savings decisions, but readers should keep in mind that
the estimates in this paper are reduced-form.

To identify the effect of the EITC phase-out tax rate on the decision to save in an income-bearing
account, I exploit two different forms of variation in the tax rate faced. Section 3.4 uses variation in
the marginal EITC tax rate above the second EITC kink, relative to the marginal EITC rate below.
The evidence in this section comes from a narrow region in the vicinity of the tax kink. It provides
evidence on the overall response in earned and unearned income above the second kink, as well as
which types of income are most responsive to the higher tax rate above the second kink, and thus
driving the overall response. The second source of variation used in this paper exploits the feature
that the phase-out tax rate has varied depending on the number of dependents an individual claims
since 1991. Section 3.5 exploits this variation using a difference-in-differences estimator.

An advantage of these two different forms of identification is that they require roughly orthogonal
identifying assumptions. The key identification assumption for estimating the response of investment
income in Section 3.4 is that investment income as a fraction of AGI would increase at a constant
rate, on average, in the region around the second EITC tax kink if no tax kink existed in this region.
Broadly speaking, the estimation strategy in Section 3.5 requires that the composition or secular
behavior of those observed with one or two dependents cannot change differentially for one relative
to two dependents in a way that is correlated with changes in the tax rate and outcomes. The fact
that the results from both are similar provides additional credibility to the estimates obtained.

Individuals receiving the EITC are also subject to the regular tax schedule and may also be
receiving subsidies from other cash transfer programs, such as AFDC/TANF, Medicaid and food-
stamps/SNAP. These interactions will matter only if changes in these programs coincide with changes
in the EITC tax rate and also discourage saving. Often, individuals in the phase-out region are not
eligible to receive benefits from these other need-based programs. Moreover, the disincentives to
save from these programs declined (while the disincentives from the EITC increased) over the pe-
period I examine and the decline in disincentives did not vary by number of children (Hurst and Ziliak, 2006; Gruber and Yelowitz, 1999), which means that these programs are unlikely to bias my results. Regarding the regular tax schedule, interactions between the EITC schedule and the regular tax schedule will be discussed below when they are important for the analysis.

3.3.3 Awareness

Awareness of the tax system is important in studies where the treatment is the marginal tax rate faced, because individuals must know and respond to the tax rate they will face in advance; this is in contrast to more standard treatment effect studies (e.g. the effect of a job training program) in which individuals necessarily know the treatment they receive before responding. When individuals are filing their tax return, it is too late to respond, except by engaging in tax evasion or making an IRA contribution.¹⁶ A small literature finds that awareness of detailed features of the EITC schedule is limited (Liebman, 1998; Romich and Weisner, 2000). But, there is no reason to think that awareness of the EITC is different than that for other features of the tax schedule or other cash transfer programs.

Awareness matters for several reasons. First, individuals must be aware that their unearned income is taxed if they are in the phase-out region. The fact that unearned income is taxed in the phase-out region has often been ignored in the EITC literature. However, this feature of the EITC is prominent when taxpayers fill out the form to calculate their EITC.¹⁷ A picture of this form for 2006 is displayed in Figure 3.3. If (and only if) individuals’ AGI places them in the phase-out region, they are asked to calculate their credit based on their earned income, and then again based on their AGI. They are then required to take the minimum of the two credit amounts. If individuals’ AGI makes them ineligible for the credit, they are stopped before they start to fill in the form. Thus, individuals who are filing for the first time, or even are in the phase-out region for the first time, may not know about this feature until they fill out the tax form. This is too late for the individuals to respond, unless the individuals decide to engage in tax evasion or make IRA contributions.

Second, individuals must be aware of changes in the phase-out tax rate over time. Third, conditional on knowing that their income will be taxed at a higher rate in the phase-out region, individuals must also know whether they will be in the phase-out region, and thus face the phase-out tax rate.

¹⁶IRA contributions can be made up to April 15 and be applied as a deduction to the previous year. However, very few EITC individuals make IRA contributions.

¹⁷These features are less prevalent when individuals use TurboTax, but this type of electronic filing was not too common in the years examined. Only 13 percent of taxpayers filed their returns from home using tax preparation software in 2005 and this was likely not concentrated among low-income households. Free tax preparation software for low-income households did not exist until 2006.
in advance. Before 1987, the EITC schedule was not inflation-indexed, which meant that the location of the tax kink remained the same in nominal dollars year after year, unless there was a tax reform. After 1987, the EITC schedule was inflation-indexed with a lag. For example, the inflation parameter which generated the 2006 tax schedule was the CPI for September 2004-August 2005. Rounding is then used (usually the values are rounded down to the nearest $50). For this reason, it is somewhat unlikely that individuals’ income will naturally move in lock-step with the tax kink after inflation-indexing was introduced. An individual can know at the beginning of the tax year what the tax parameters will be and respond accordingly, but this requires them to look up the location of the tax kink each year. As individuals’ income gets further away from the tax kink, the amount of awareness necessary for them to realize the treatment they face and respond accordingly decreases, because they can more easily guess whether or not they will remain in the phase-out region in the following year.

The baseline estimates in Sections 3.4 and 3.5 ignore the issue of awareness. The estimates in Section 3.4 are expected to be lower than those in Section 3.5 precisely because it requires a greater degree of awareness to respond each year in a narrow region around the tax kink, than somewhere in the phase-out region. However, several additional specifications in Section 3.5 will analyze ways in which individuals become more or less aware of EITC-related incentives, including their use of a paid preparer and having unemployment insurance (where unemployment insurance is considered a rough proxy for whether the individual experiences a transitory shock in earnings).

3.4 Estimating the Response at the Second Tax Kink

This section examines the behavior of people whose income places them in the income region around the second tax kink, exploiting the change in the EITC marginal tax rate at the kink. The section first estimates the overall response in AGI to the change in the tax rate using a modification of the method proposed by Saez (2010), which relies on estimates of bunching in the region around the tax kink. I find small, but statistically significant elasticities at the second EITC kink for both wage-earning and self-employed individuals. I then implement a procedure that provides evidence regarding which types of income are most responsive to the change in the EITC rate at the kink; that is, which types decline significantly above the kink. The response for wage-earners is driven by changes in unearned income, particularly investment income.
3.4.1 Bunching

A classical model predicts bunching at tax kinks when the budget set is convex, because individuals above the tax kink will wish to decrease their income as the tax rate above the kink rises given their preferences (labor-leisure preferences in the context of earned income, saving-consumption preferences in the context of dividend and interest income, and so forth). However, once they reach the tax kink, they have no incentive to decrease their income further. This is because the tax rate below the kink did not change and they were already willing to earn income at this tax rate before the tax rate increased above the kink. If individuals were able to perfectly adjust their AGI, all these individuals would be located at a point mass at the kink; however, adjustment costs and income volatility generate imperfect bunching around the tax kink. Saez (2010) shows that the amount of bunching in the region near a tax kink can be used to calculate an elasticity of AGI with respect to the marginal net-of tax rate using the following formula:

\[ \hat{\varepsilon} \approx \frac{\hat{B}}{z^* \ln \left( \frac{1-\tau_l}{1-\tau_h} \right)} \]

where \( z^* \) is the point in the income distribution at which the tax kink is located, \( \tau_l \) and \( \tau_h \) are the marginal tax rates on either side of \( z^* \), such that \( \tau_l < \tau_h \). Let there be a region \([z^* - \delta_b, z^* + \delta_b]\) in which all imperfect bunching occurs. Then, \( \hat{B} \) is the estimated excess mass and \( \hat{h}_0(z^*) \) is the estimated counterfactual density of taxpayers in this region; that is, the density of taxpayers if there was no tax kink. These parameters are depicted graphically in Figure 3.4. Saez (2010) implements this method to estimate EITC elasticities at each kink; however, his analysis ignored that individuals’ location relative to the second kink is usually determined by AGI, not earned income, as stressed in this paper.

To estimate \( \hat{B} \) and \( \hat{h}_0(z^*) \), I first estimate the smoothed density \( \hat{h} \) using a local linear regression within $12,000 of the tax kink.\textsuperscript{18} I use the smoothed density estimates to construct the elasticities, because this method provides substantial efficiency gains, relative to constructing the estimates using unsmoothed estimates, which is the method used by Saez (2010) and Kleven and Waseem (2011). I use an automatic bin and bandwidth selection criteria\textsuperscript{19} to select the bin and bandwidth size within

\textsuperscript{18} The figures are trimmed from -$9,000-$9,000 to exclude the edges which are noisy because the estimates are smoothed over a small number of observation at the edges.

\textsuperscript{19} The bandwidth is based on the rule-of-thumb approach given in Fan and Gijbels (1996). This is the method used in McCrary (2008), whose estimation procedure is the same as is used here, except I do not allow for a break at \( z^* \). Since I am examining tax kinks, not tax notches, imperfect bunching should take place on both sides of the tax kink. Therefore, to gain efficiency in the region of interest—just on either side of the tax kink—I do not allow for a break in the density at zero because one is not expected. The results when I do and do not allow a break are qualitatively similar, but the latter provides substantial efficiency gains.
As long as the bin size chosen is small relative to the selected bandwidth, the exact choice of bin size is usually not important. However, Saez (2010) effectively over-smooths the data by using a bin size ($500$) that is an order of magnitude larger than the bin size used here for his smoothed density figures (it is equivalent to using a uniform kernel at all points within the bin size). Over-smoothing the data would induce a downward bias in my elasticity estimates; for example, using the baseline bandwidth reported in this paper for self-employed individuals for all years, but replacing the bin size with $500$, biases the estimates downward by more than 10 percent, with a minimal decrease in variance.

Let $\hat{h}^*_0$ be the estimated mean density in the region $[z^* - \delta_b - \delta_c, z^* - \delta_b]$, $\hat{h}^*$ be the estimated mean density in the region $[z^* - \delta_b, z^* + \delta_b]$, and $\hat{h}^*_0$ be the estimated mean density in the region $[z^* + \delta_b, z^* + \delta_b + \delta_c]$, where $\delta_c$ is the width of the counterfactual region on either side of the imperfect bunching region. Let $\hat{H}^*_0 = \delta_c \hat{h}^*_0$, $\hat{H}^* = 2\delta_b \hat{h}^*$, $\hat{H}^*_0 = \delta_c \hat{h}^*_0$, where these variables denote the cumulative density in their respective regions. Ideally, $\hat{h}^*_0$ would be the counterfactual density, absent a tax kink, in the region $[z^* + \delta_b, z^* + \delta_b + \delta_c]$. However, when $\tau_h$ is the tax rate above the tax kink, it overestimates the counterfactual if individuals above $z^*$ decrease their income by a certain percentage in response to the higher tax rate (as opposed to decreasing their income by a fixed amount); this is consistent with what our models predict and the parameter our estimation strategies are designed to uncover. It overestimates the counterfactual in the region of the tax kink because the counterfactual distribution gets compressed as each individual decreases their income by a percent of their total income. Therefore, the elasticity estimates of a particular income type with respect to the marginal net-of-tax rate $(1 - \tau)$ presented in this section can be interpreted as lower bounds on the truth.

The counterfactual density is the mean of the densities above and below the region of imperfect bunching: $\hat{h}_0(z^*) = \frac{1}{2}(\hat{h}^*_0 + \hat{h}^*_0)$. This method assumes that the densities on either side of the region of imperfect bunching are a good approximation for the counterfactual density in the imperfect bunching region. The excess mass is constructed by subtracting the counterfactual mass from the

---

20 A bin size and bandwidth chosen over a larger region would not be appropriate to efficiently measure the density in the imperfect bunching region.

21 If $\tau_l$ is the tax rate above the tax kink, this discussion applies to $\hat{h}^*_0$ instead of $\hat{h}^*_0$.

22 Previous literature has either ignored this issue (Saez, 2010; Kleven and Waseem, 2011) or assumed that the density is biased downwards because the bunching individuals came from the region above the tax kink (Chetty et al., 2011b). The latter assumption is incorrect as long as individuals further above the tax kink are responding in the same way as those near the tax kink and the counterfactual region is not near the next tax kink (in the region near the maximum income taxed at a given rate, there is a decline because these individuals are no longer being replaced by individuals further up in the distribution).

23 This is the most common way of constructing the counterfactual density in the literature (Saez, 2010; Kleven and Waseem, 2011). The other approach that has been used estimates a global polynomial to construct the counterfactual (Chetty et al., 2011b). This method is ideal when the distribution has a shape that is well approximated by a low-order polynomial (even if ultimately a higher-order polynomial is used). This does not apply for this distribution, because
total mass in the bunching region: $\hat{B} = \hat{H}^* - \frac{\delta_b}{\delta_c}(\hat{H}^* + \hat{H}^*).$

The other important consideration in constructing these estimates is the choice of the size of the imperfect bunching region, $2\delta_b$, and the size of the counterfactual region, $\delta_c$. I set $\delta_b = \delta_c = \$1,000$ for the baseline estimates, which are the same as those used by Saez (2010), to analyze the second tax kink. The choices for $\delta_b$ and $\delta_c$ are based on the following considerations. First, the gap between the first and second kink for two-child families since 1996 is $\$3,470$. Therefore, it must be that $2\delta_b + \delta_c \leq 3,470$, so that the counterfactual region for each kink does not include part of the bunching region for the other tax kink. If $\delta_b$ is smaller than the imperfect bunching region, the estimates will be biased downwards because part of the bunching will be excluded from $\hat{H}^*$ and will instead be included in $\hat{H}^*$. Assuming the counterfactual density is flat, choosing a $\delta_b$ that is larger than the imperfect bunching region does not bias the estimates, but will otherwise. In practice, choosing a large $\delta_b$ induces a bias in the estimates, because the extent to which $\hat{h}^*$ and $\hat{h}^*$ provide an accurate estimate of $h_0(z^*)$ usually declines further away from $z^*$. There is not much that can be done in the way of sensitivity analysis over the whole sample period given the constraint that the first kink imposes. I pursue this issue below, when I look exclusively at the years 1988-1993, when the first kink was about twice as far away as it was in 1994-2006.

I examine self-employed individuals and wage-earners separately because people in these two employment states arguably differ in their capacity to respond to tax kinks on several dimensions. An individual is defined as self-employed if they have any self-employment income. One way in which individuals can bunch at the tax kink is to simply not report income above the kink—tax evasion. This is quite difficult to do with wage and salary income and not get caught, as this income is generally subject to withholding and information reporting. Self-employment income faces neither withholding nor information reporting.

Tax avoidance is another way in which individuals can respond. There are few tax avoidance possibilities on earnings for wages because all wages are reported on a W-2 and thus face withholding and information reporting requirements as discussed above. Still, there is some flexibility if workers there are two kinks that are relatively close together and are not defined by the same type of income. In the early years of the EITC when the kinks were further apart, it is possible, although perhaps not completely persuasive, to get estimates using this method. The results obtained are similar. In later years, it becomes impossible, particularly for those with two children, to use a global polynomial to control for one kink, and accurately estimate the counterfactual for the other. The global polynomial does a good job of estimating the counterfactual when it is approximately linear, which is the same assumption imposed when the method in this paper is used.

24 The gap increases for married families by $\$1,000 in the years 2002-2004 and $\$2,000 in the years 2005-2006.

25 Note that there is a minimum requirement on withholding for wage-earning individuals. It is $\$221 for single individuals and $\$667 for married individuals in 2006 dollars. This should, in general, not have an effect on the estimates, unless individuals hold two jobs, one of which is below the threshold and they claim to only have one job on the Form W-4 for their primary job (Form W-4 is completed by individuals so that their employer knows how much to withhold). If individuals decide to report the income from their second job if they face the lower tax rate, but not if they face the higher rate, a small part of the response by these individuals could be tax evasion.

26 It is possible that such individuals do receive cash payments for work done on the side for an employer.
are able to substitute towards non-monetary forms of compensation. Self-employed individuals have a greater opportunity for tax avoidance. These individuals are allowed to deduct their expenses from their gross income, which allows them to have an extra expense, say, which would place them at the tax kink without engaging in anything illegal.

A labor supply response could also move these individuals towards the tax kink. Self-employed individuals face lower adjustment costs, on average, associated with altering their labor supply decisions, so it is more feasible for them to make minor adjustments.

Each of these three possibilities—evasion, avoidance, and a labor supply response—all suggest that the income adjustments will be larger and more precise for the self-employed compared to wage-earners. The top panels of Figures 3.5 and 3.6 display the bunching at the second EITC kink for self-employed and wage-earning individuals for all years, respectively.\textsuperscript{27} The densities are normalized so that they integrate to one. The variables \( \hat{z}^*, \hat{\tau}_l, \) and \( \hat{\tau}_h \) refer to the population weighted averages for each sample. The solid yellow line below the density is the estimated counterfactual density of individuals in the bunching region. The bunching depicted in the figures is more sharply defined for the self-employed. The corresponding elasticities are given in Table 3.3 Columns (1) and (2). Standard errors for the elasticities were obtained by nonparametric bootstrap (500 replications). The elasticity estimate of adjusted gross income with respect to the marginal net-of-tax rate for self-employed individuals is 0.063 and is significant at the one percent level, while the estimate for wage-earners is 0.010 and insignificant.

Table 3.3 Columns (3)-(5) and the second panel of Figures 3.5 and 3.6 examine years 1988-1993 for self-employed and wage-earners, respectively. The key advantage of examining the early years of the tax credit is that the first kink was about twice as far away for filers with two or more dependents, so a broader range of \( \delta_b \) and \( \delta_c \) are feasible.\textsuperscript{28} I fix \( \delta_c = $1,000 \). For the self-employed, I report estimates for \( \delta_b = $2,000 \) and \( \delta_b = $1,000 \) in Columns (3) and (4); the counterfactual density is drawn in Figure 3.5 for the case of \( \delta_b = $2,000 \). The estimates are similar for both and visually \( \delta_b = $2,000 \) appears correct, but there are efficiency gains when using \( \delta_b = $1,000 \). The elasticity estimate for \( \delta_b = $1,000 \) is 0.095 and is statistically significant at the 10 percent level. One could even make the case that \( \delta_b \) should be $3,000. In this case, the estimate roughly doubles and this rise

\textsuperscript{27}The estimates are population weighted.

\textsuperscript{28}Also, in the later years, the second kink interacted with several other tax schedule features. First, the filing threshold for married individuals was within $1,000 of the second kink in 1994-2006. It is unlikely that this had a substantial effect on filing, as it was clearly in the best interest of individuals on both sides of the tax kink to file to claim the EITC, but it is possible. Second, in 2002, the Saver’s Credit was introduced and the first notch for the single filers is near the second EITC kink. Ramnath (2010) documents bunching at these kinks, but the number of individuals applying for the Saver’s Credit was quite small, so it is unlikely that this would generate a substantial bias in these estimates. If either of these other programs were playing a significant role in the size of the overall estimates, those estimates would be biased upwards.
is driven by the fall in density between -$4,000 and -$3,000, which is now part of the counterfactual density. For wage-earners, the plotted density strongly supports $\delta_b = $3,000.

The fact that the bunching is more spread out for wage-earning individuals is consistent with the fact that they have less flexibility in adjusting their earned income. The elasticity estimate for wage-earners is given in Column (5). It is 0.104 and is statistically significant at the five percent level. The wage-earner and self-employed elasticities are about the same (once the difference in the precision of bunching is taken into account) for these years. There are several possible explanations for this finding. First, when self-employed individuals engage in tax evasion, they may move all the way to the first tax kink (in the region between the first and second kink, they receive no additional EITC benefit, but do have to pay additional payroll tax on their self-employment earnings). Alternatively, while it may be easier for the self-employed to alter their earned income, wage-earners make up the difference by adjusting their unearned income more. Given the discussion in Subsection 3.3.3, these parameters can be interpreted as a lower bound relative to what would be found further in the phase-out region, where the amount of awareness needed to respond precisely is lower.

3.4.2 Changes in Average Income

So far, I have presented evidence that individuals respond to the second EITC kink in a small, but statistically significant, way. However, this analysis cannot be informative regarding which types of income are most responsive, nor does the literature have any precedent for such an examination beyond splitting the sample by those with and without a particular type of income (e.g. self-employed income) and examining whether the elasticity is different across the two groups. When the response is on the intensive margin, this approach will uncover the relevant response heterogeneity. However, when the response is on the extensive margin, the response of these individuals will be captured in the elasticity of those categorized as not having this type of income, because the researchers will only observe these individuals after they have given up all of this type of income. This is especially problematic for types of unearned income where the likelihood of an extensive margin response is higher. The rest of this section proposes a new alternative method of examining behavior around the tax kink for different types of income to provide some evidence regarding which types of income are the most responsive and are thus contributing to the elasticities found above.

To explain this method, I use the example of investment income, and assume that a tax kink is introduced as part of a tax reform, such that the reform discretely increases the marginal tax rate above the tax kink. Suppose individuals are able to bunch perfectly at the tax kink, there is no growth in investment income, and investment income is the only type of income that responds to the
increase in the marginal tax rate above the kink. Then, investment income will decline above the tax kink as depicted in the top panel of Figure 3.7. Note that investment income is a part of AGI, and thus a change in investment income moves an individual’s location on both the horizontal and vertical axes. Below the tax kink, investment income does not respond. At the tax kink, average income is comprised of all those that were at the kink before the reform, plus all those who move to the kink in response to the tax reform (these are all the individuals whose desired percent decrease in investment income at the new tax rate is at least as large as the gap between their pre-reform AGI and the tax kink). This point in Figure 3.7 is located half way between the old and new investment income levels, which occurs when there is a uniform distribution of investment income, everyone holds the same amount of investment income pre-reform, and everyone responds in the same way. In general, this point can be anywhere within the old and new levels, or even above the old level.\textsuperscript{29} Above the kink, investment income is at a new lower level. Under these assumptions, one could drop the point at the tax kink and then implement a regression discontinuity design to estimate the effect of the higher tax rate above the kink. Dropping the point at the tax kink effectively eliminates all manipulation of the running variable, which would otherwise bias the estimates.

With imperfect bunching and investment income growth, the picture changes to look like that in the bottom panel of Figure 3.7; investment income growth, even without imperfect bunching, mitigates the discrete drop in investment income above the kink if individuals are responding along the intensive margin. This occurs because individuals above the kink respond by decreasing their investment income by a certain percentage, but also started with more. This figure allows other types of income to be growing as well; their rate of growth is a determinant of the slope of investment income. But, it assumes investment income is the only type of income that responds to the tax rate. I discuss what happens when this assumption is relaxed below. Imperfect bunching causes investment income to decline once individuals start responding to the tax kink at $-\delta_b$; that is, for a given level of AGI, investment income will be lower (it can also increase for the same reason as described in the perfect bunching case when investment income increases at the point of the tax kink). By the time AGI has exceeded the imperfect bunching region, investment income will be at a new, lower growth rate.\textsuperscript{30}

By calculating the percent change in average investment income in a region above $\delta_b$ relative to

\textsuperscript{29}The latter is more likely to occur as the response by those with investment income decreases and the number of individuals with no investment income increases.

\textsuperscript{30}Because imperfect bunching exists and the outcome (investment income) and assignment variable (AGI) are simultaneously realized, calculating the change in slope by taking the limit as AGI approaches the tax kink (using a method similar to that in Card et al. (2009)), would necessarily yield an estimate of approximately zero (i.e. with imperfect bunching, the individuals are approximately the same just on either side of the tax kink).
the counterfactual investment income that would have existed absent the tax change, backing out the
everosity of investment income with respect to the marginal net-of-tax rate is straightforward. The
counterfactual is empirically constructed from the region \([z^* - \delta_b - \delta_c, z^* - \delta_b]\); for the counterfactual
to be valid, investment income must continue to grow at the same rate in the region in which the
everosity is calculated. However, the standard errors on the counterfactual increase as AGI increases
above \(-\delta_b\), because the accuracy with which I can predict the counterfactual using investment income
in the region \([-\delta_b - \delta_c, -\delta_b]\) decreases as AGI increases. For this reason, I instead calculate the
everosity in the region of the tax kink \([-\delta_b, \delta_b]\), and acknowledge that this is a lower bound on the
true everosity.\(^{31}\)

I will now define some notation, which is depicted in Figure 3.7.

Let \(X_{i+}^*\) be average investment income for bin \(i\) and let \(X_{i-}^*\) be the counterfactual investment
income in the same bin. The everosity can then be calculated using the following formula:

\[
\hat{\epsilon} = \frac{\sum_{i=1}^{n_b} \ln \left( \frac{X_{i-}^*}{X_{i+}^*} \right) / n_b}{\sum_{j=1}^{n_c} \ln \left( \frac{1 - \tau_{1j}}{1 - \tau_{1k}} \right) / n_c},
\]

where \(n_b\) is the number of bins in the region \([-\delta_b, \delta_b]\) and \(n_c\) is the number of individuals in the
region \([-\delta_b, \delta_b]\).

Note that this everosity estimate is also a lower bound on the underlying average everosity
(assuming investment income is increasing, on average, in AGI) because investment income is en-
dogenous with respect to AGI; that is, when investment income declines, so does AGI, pushing down
the estimated response at a given AGI level. However, a positive response will be found as long as
investment income is not the sole source of growth in AGI. Moreover, like the density method used
above, this method implicitly assumes that there are no other factors that need to be held fixed in
order to obtain a consistent estimate, which makes it possible to obtain a non-causal relationship.
For example, suppose earned income decreases in response to the tax rate, but only for those with
low investment income. This would lead to a spurious response in investment income. This will
be partly ruled out in the discussion that follows; still, these are informative results, meant to pro-
vide evidence of the apparent response before proceeding to the next section, which analyzes the
investment income response in a more precise, causal way.

The everosity estimates from equation (3.2) are given in Table 3.3 Columns (6)-(9). The standard
errors are obtained by a nonparametric bootstrap method (500 replications). I exclude those with
unemployment insurance income from the analysis because a decline in unemployment insurance

\(^{31}\)When the former method is used, the results are qualitatively similar, but the standard errors are substantially
larger.
should be accompanied by an increase in earned income; this method is not designed to address this type of mechanical correlation. The first three columns examine the wage-earning sample. Columns (6) and (7) report the results for the two components of AGI—earned and unearned income—respectively. Figure 3.8 is the corresponding figure. Both the figure and the elasticity estimate for earned income (-0.015) point to the following fact: earned income grew at a faster rate (although not statistically significant) above the tax kink relative to below. This occurs mechanically if other types of income included in AGI are growing slower in response to the higher tax rate and earned income is not responding, or at least responding in a very small way, relative to the other types of income that do respond (the mechanical response is small because earned income is a much larger share of AGI than unearned income). The magnitude of the earned income response is consistent with it being entirely mechanical. In contrast, Figure 3.8 shows a decline in unearned income above the tax kink. The corresponding elasticity is 0.516, but is statistically insignificant, likely reflecting the overall volatility of unearned income and also that some types of unearned income are expected to be responsive, while others are not expected to be responsive.

Table 3.3 Column (8) and the top panel of Figure 3.9 examine one type of unearned income—investment income. Although capital gains or losses are a lumpy realization of savings, I include these in this analysis of investment income, because most individuals that have capital gains also have dividend or interest income. If capital gains are excluded, and individuals respond to the higher marginal tax rate above the tax kink by decreasing capital gains, this will likely bias the estimates downwards. Both the figure and the elasticity provide evidence that investment income declined substantially above the tax kink—the elasticity estimate is 1.879 and is statistically significant at the 5 percent level.

In Table 3.3 Column (9) and the second panel of Figure 3.9, I consider whether self-employment income for self-employed individuals declined significantly in the region around the kink. Both the figure and the elasticity estimate suggest that it did; however, the estimates are not significant.

This section has found a statistically significant response to the phase-out region tax rate among both wage-earners and self-employed individuals, and for the former, it is concentrated among non-labor income, particularly investment income. The elasticity of investment income with respect to the marginal net-of-tax rate is 1.88 and can be interpreted as a lower bound on the true elasticity.

32 The sample is restricted to those for whom the sum of dividend, interest, positive capital gains, and tax exempt interest income is under $2,300 in all years, which is $500 below the investment income asset limit in 1996-2006; this corresponds to my sample restriction in the next section, and is imposed here for comparability. The results are similar when it is not imposed.
3.5 Difference-in-Differences Estimates

This section focuses specifically on the responsiveness of investment income to the EITC phase-out marginal tax rate, exploiting differential variation in the phase-out tax rate between one- and two-dependent families. The evidence is consistent with the evidence from the last section, pointing to a significant decline in investment income in response to changes in the EITC schedule. There is no evidence that bias-inducing selection or tax evasion are driving the results, although both are considered.

The results from the last section inform the method used here. In particular, I do not rely on a comparison of individuals above and below the tax kink because bunching occurs at the tax kink, which generates selection into a particular observed tax rate and thus prevents a causal parameter from being obtained. Weber (2012b) shows that picking an instrument to resolve this type of selection is difficult, and no obvious instrument exists in this context. To the extent that imperfect bunching is due to income volatility, selection into a particular tax bracket may be mitigated, but this creates a new problem—the parameters of the treatment are not known with certainty in advance, making it difficult for individuals to respond.

I proceed by defining a region based on earned income within the phase-out region and identify the parameters of interest by relying on differential variation in the marginal tax rates between families with one and two children. Defining the region in this way will not necessarily yield consistent estimates if earned income responds to the tax rate. The evidence in the previous section is consistent with the assumption that earned income does not respond to the tax rate for wage-earners, but not for self-employed individuals. Therefore, I restrict the sample to wage-earning individuals. Given the results in Section 3.4, a response in earned income would necessarily be small, but I cannot completely rule it out; therefore, later in this section, I will discuss the conditions under which such a response would induce a bias in the estimates, and provide evidence that this form of selection does not appear to be having a statistically significant impact on the estimates.

The identifying variation in this section comes from the differential variation in the phase-out EITC rate between one- and two-child families over time. This variation is plotted in Figure 3.10. There, I plot the marginal net-of-tax rate, which is one minus the marginal tax rate; this is the variable that will be used in the main estimating equation. Given the form of identifying variation, the specifications in this section assume that individuals do not make endogenous fertility decisions or decisions to claim an extra dependent in a way that is correlated with their investment income response. These types of exogeneity condition are assumed in much of the EITC literature, as most
of it relies on differential variation in tax rates across the number of children for identification. I consider this story later in this section and find no evidence that the estimates are biased by individuals selecting the number of children they have in a way that is correlated with their outcomes.

Before proceeding to the main estimation, I provide graphical evidence of the response in investment income over time by number of children, and estimate a binary treatment (i.e. did an individual have any interest or dividend income) difference-in-differences equation. For this analysis, I split the years into two groups, pre and post. Pre includes all years through the end of the EITC expansion (1988-1995), and post includes all the years after the expansion (1996-2006). I estimate the following difference-in-differences intent-to-treat (ITT) specification for wage-earning individuals with earned income in the range $18,000-$30,000:

$$INV_{it} = \alpha + \gamma_{post} \times twokids_{it} + \beta_1 twokids_{it} + \beta_2 X_{it} + \eta_t + \nu_{it},$$

(3.3)

where $\gamma$ is the parameter of interest, $INV_{it}$ is an indicator for whether an individual has positive interest or dividend income, $twokids$ is an indicator for whether the individual has more than one child, $post$ is an indicator for the observation being in 1996-2006, $X_{it}$ are additional covariates, and $\eta_t$ are year fixed effects.\(^{33}\) The vector $X_{it}$ includes earnings, earnings squared, married, earnings x $twokids$, earnings squared x $twokids$, and married x $twokids$. I pool data over all years in the sample. All income values are in 2006 dollars.\(^{34}\) I choose $18,000 as the lower cutoff because it is at least $1,000 above the second kink point in all years; thus, the estimates will likely not capture bunching decisions around the kink point. It is above the filing threshold for each type of marital status and above the positive tax rate threshold for all single and head of household taxpayers, but not for married individuals. For them, the positive tax rate is implemented part way through the window examined in this section. However, given that families with one dependent face this positive tax rate first, this will bias the estimates downward, because it will make one-dependent families appear more responsive relative to two-child households. Earned income of $30,000 was chosen as the baseline upper bound, because it is at least $1,000 below the end of the phase-out region for all years. The baseline estimates are robust to the choice of upper and lower bound; I will analyze the sensitivity of the main estimates to the bounds in Table 3.6.

I drop individuals with more than $2,300 of investment income (the EITC definition of investment income is dividend, interest, tax exempt interest and positive capital gain income) in any year. This

\(^{33}\) A separate indicator for $post$ is not included. It is not needed because year-fixed effects are included.

\(^{34}\) I use the deflator used by the IRS, which is the average CPI-U from September in year $t - 2$ through August in year $t - 1$. 52
decision is driven by the fact that there was an asset limit imposed in 1996, excluding those with more than $2,800 dollars of investment income in 2006 dollars from claiming the EITC. Before 1996, 1.1 percent of two-child families held investment income over this limit and were 10 percent more likely to have investment income over the limit than one-child families. Two-child families are also more apt to respond to the limit by decreasing their investment income below the threshold because they will lose more in benefits than an equivalent one-child household. Consequently, estimates that include individuals with investment income over the limit or impose the limit only after 1996 will be biased upward. I impose a limit of $500 less than the asset limit to avoid differential rates of selection by number of children into having investment income just under the cutoff.

Table 3.4 provides descriptive statistics for the baseline specification. Average interest and dividend income reported, including all zeros, is $48.60. About 20 percent of individuals have positive interest and dividend income. Almost one-third of the sample is married, and the average number of children is 1.4. Fifty-nine percent use a paid preparer.

The estimated $\gamma$ from equation (3.3) is -0.036 and the p-value is 0.000. The main estimation below defines treatment as $\log(1 - \tau)$. Rescaling the estimated $\gamma$ using two-stage least squares so that it is directly comparable to the main estimates, yields an estimate of 0.789 with a p-value of 0.000. This estimate implies that, for a one percent decrease in the marginal net-of-tax rate, the likelihood an individual opens an interest- or dividend-bearing account declines by 0.008. For example, if the saving rate was 0.4 before a 10 percent decline in the marginal net-of-tax rate, it would be 0.32 afterwards.

Figure 3.11 provides graphical evidence of the behavioral response to changes in the EITC phase-out tax rate of the choice to hold income in interest- or dividend-bearing accounts over time using the same sample restrictions. The estimated line comes from a cubic in year interacted with post and number of children. In the early years, when there was no (or a small) difference between the tax rates for one- and two-child households, their investment income was not statistically distinguishable, suggesting that one-child families provide a good comparison group for two-child families. It takes a year or two for individuals to adjust to the new tax rates, which is consistent with other EITC literature (Saez, 2010). In 1996, there is a statistically significant drop in the likelihood that two-child families have investment income and this difference remains statistically significant through

---

35 There are a few special circumstances when this limit does not hold.
36 Note that another possible type of selection could occur if a substantial number of new families filed for the EITC after the phase-out rate increased (which coincided with an increase in EITC generosity) and these new filers were much less likely to have investment income. I cannot completely rule this out, but it is worth noting that while EITC claims were rising over this period, there was no discrete jump at 1996, nor is the pattern consistent with the pattern in Figure 3.11. Additionally, adding a control for the total number of EITC claimants by year in equation (3.3) has a minimal effect on the estimates.
2006. This graphical evidence is consistent with the estimates obtained from the simple difference-in-differences estimates; the gap between the likelihood that taxpayers with one and two children have investment income is about 0.03 in the post period, and this increase in the gap is stable over time. This suggests that the additional covariates included in equation (3.3) do not have a substantial effect on the estimates.

The main specification in this section takes advantage of all the variation in $\log(1 - \tau)$ by including it directly and estimating the following intent-to-treat (ITT) specification:

\[
INV_{it} = \alpha + \gamma \log(1 - \tau_{it}) + \beta_1 twokids_{it} + \beta_2 X_{it} + \eta_t + \nu_{it},
\]  

(3.4)

where the covariates and sample restrictions are the same as for equation (3.3). Note that one may wish to estimate the response of investment income to the after-tax rate of return $\log((1 - \tau_{it})r_t)$, where $r_t$ is the average annual real interest rate. Because $\log((1 - \tau_{it})r_t) = \log((1 - \tau_{it})) + \log(r_t)$ and there are year fixed-effects in this model, the estimates obtained from this specification are equivalent. The same argument applies for including the state EITC rates when state-year fixed effects are included (except for Wisconsin).\textsuperscript{37}

The estimates given by equation (3.4) produce ITT estimates because all individuals are included as long as they have earned income in the specified range, regardless of whether they actually decide to claim the EITC, or are even eligible based on their overall AGI level. It also ignores other tax rates and credits faced. I choose not to report treatment on the treated (TOT) estimates to address these issues, because the value of such estimates is limited. The estimates are quite similar in practice (only four percent of my baseline sample was not EITC-eligible) and, in this context, given the nature of the selection and its correlation with the ITT measure, Weber (2012b) shows that the estimate will be biased.

In Tables 3.5, 3.6, and 3.7, the top rows of each column estimate the extensive margin effect—the dependent variable is an indicator for whether the individual receives any interest or dividend income. These estimates are the mean marginal effects from a probit specification. The bottom rows estimate the elasticity of dividend and interest income with respect to the marginal net-of-tax rate. These estimates come from a two-tiered model proposed by Cragg (1971).\textsuperscript{38} The first tier estimates the extensive margin decision and is estimated using the same probit model as is used in the top rows. The second tier estimates the intensive margin effect (i.e. conditional on choosing to

\textsuperscript{37} It does not apply to Wisconsin because a larger percentage of the federal credit is given to families with two children.

\textsuperscript{38} This type of model is also referred to as a hurdle model.
save a positive amount, how much does the individual save) using a truncated normal regression, where the dependent variable is the log of interest and dividend income. Taking the log of interest and dividend income is attractive because the specification directly produces an elasticity, and it is necessary because the unscaled dependent variable has a large left skew that, if left untransformed, would violate the assumptions of the estimation method. The reported coefficients are mean marginal effects from this two-tier model. Standard errors were calculated using the delta method. I do not report the second tier; however it is worth noting that the second tier coefficient on $\log(1 - \tau_{it})$ is never statistically significant. This suggests that the behavioral response occurs along the extensive margin and much of what the elasticities are capturing is the average amount by which dividend and interest income changes when individuals choose to open, or close, an interest- or dividend-bearing account.

Table 3.5 presents some results. Baseline estimates are given in Column (1). The extensive margin estimate is 0.612, which means that a one percent increase in the marginal net-of-tax rate increases the probability of having interest or dividend income by 0.006. The phase-out marginal net-of-tax rate decreased by 9.6 percent over the last two decades, which implies a 0.058 decrease in the likelihood that individuals have interest or dividend income. The overall decrease in the likelihood that individuals hold interest or dividend income for EITC recipients in this population is 0.134; hence, among those examined in this section, 43.5 percent of the decline in the fraction of EITC recipients with savings in income-bearing accounts over the last two decades in the can be explained by changing EITC incentives. Overall, the extensive margin estimate implies that 39.4 percent of the decline in the fraction of EITC recipients with savings in income-bearing accounts over the last two decades can be explained by changing EITC incentives.39

The elasticity estimate is 3.047, which implies that a one percent increase in the marginal net-of-tax rate increases interest and dividend income by 3.047 percent. The elasticity estimate implies that for each one percent increase in the marginal net-of-tax rate, dividend and interest income increases by $1.48 at the mean. Suppose that the interest rate is 5 percent. Then, if individuals spend their income (as opposed to saving it, but not placing it in an interest- or dividend-bearing account), this implies an increase of $29.62 at the mean in the annual amount in savings. Among those that have positive savings in 1988-1990, this implies a $185.82 increase in savings. Because the marginal net-of-tax rate decreased by 9.6 percent over this period, these estimates imply that the annual amount in savings, among those with positive saving in 1988-1990, would have been

---

39These estimates include all EITC eligible individuals with one or two dependents, who were wage-earners with investment income less than $2,300.
$1,775.82 higher in 2006 if the disincentives to save had remained at their 1988 level.

Table 3.5 Column (2) examines a broader measure of investment income, including capital gains or losses, partnership income, other gains or losses, and IRA contributions;\(^{40}\) this has a minimal effect on the estimates. I exclude capital gains in the baseline estimates because it is a lumpy realization of savings, rather than the annual amount. Column (3) includes the controls \(X\) in a more flexible way, by interacting each control with an indicator variable for the observation occurring before 1992 and another indicator variable for the observation occurring after 1998. The year 1992 is two years before the largest expansions in the gap between one- and two-child families, and 1998 is two years after the last change in the tax rate. This specification allows for different trends in investment income in the years before 1992 and after 1998. The estimates increase slightly and the additional controls have a minimal effect on the standard errors. Column (4) adds state fixed-effects to control for differences in state EITC generosity.

Table 3.5 Columns (5)-(7) examine heterogeneity in the baseline estimates by taxpayer type. Column (5) estimates the response separately for unmarried (single and head of household) individuals and married individuals; this estimate is perhaps not too interesting on its own, but the lack of difference between these two types lends credibility to the identification used in Subsection 3.5.1. Married individuals are slightly less responsive than unmarried individuals and the difference in responsiveness is highly insignificant.

One might expect that those who temporarily face the EITC schedule, may be less responsive, since savings is often a longer-term decision, and awareness of detailed features of the schedule may be low. On the other hand, if an individual is a temporary EITC recipient because of an unemployment spell, their response could be more elastic, as it may be an appealing time to draw down their savings. I use an indicator for having unemployment insurance as a proxy for temporarily facing the EITC schedule due to a negative shock in earned income; it is of course not a perfect proxy as an individual could be in the EITC range both with and without becoming unemployed for part of the year. The results are reported in Table 3.5 Column (6).\(^{41}\) The results suggest that, on average, those with unemployment insurance receipts do not respond differently; the estimates are not statistically different.

Table 3.5 Column (7) considers how the response differs among those that do and do not use

\(^{40}\)Note that all negative values are truncated at zero and estimated using the same two-tier model above to take this truncation into account.

\(^{41}\)Unemployment insurance indicators and a complete set of interactions with the other covariates are included in the estimating equation to control for underlying differences in investment income among those with unemployment insurance.
a paid preparer. Those who choose to have a paid preparer may be more responsive to begin with, and the paid preparer may educate them about the costs and benefits associated with having investment income when in the phase-out region. With regards to the latter, using a paid preparer in the current year can be interpreted as a proxy for having a paid preparer in the past, because learning in the current year is not helpful unless the tax preparer encourages individuals to engage in tax evasion. Both those with and without a paid preparer respond, but the response of those who use a paid preparer is twice as large. This supports the hypothesis that paid preparers increase individuals’ understanding of the EITC schedule.

Table 3.6 considers two possible forms of selection discussed above, which could be biasing the estimates. Column (1) repeats Column (4) in Table 3.5. The first type of selection occurs when individuals select into or out of the sample by adjusting their earned income in response to the EITC phase-out tax rate. If individuals are moving in and out of the sample and this selection is correlated with the likelihood they respond along the investment income margin (or have investment income), then the estimates may be biased. The estimates remain unbiased only if this selection is occurring at a constant rate—that is, entry and exit are equal. Assuming a positive correlation among earned and investment income responses, the estimates will be biased downwards (upwards) if exit (entry) is larger. This occurs because those who are entering or exiting will adjust their investment income more to a given change in the marginal tax rate than individuals whose earned income does not respond to the tax change. When exit (entry) is larger, my estimates will exclude (include) more individuals that are highly responsive. The last section found no evidence of a response in earned income; however, a small response among certain taxpayers could exist. If responsiveness changes monotonically with earned income, the entry and exit rates will converge as the estimation window narrows, making the bias decrease. Moreover, if I divide the sample in half and estimate the response over $18,000-$24,000 and $24,000-$30,000, both estimates should experience a reduction in bias, thus moving the estimates in the same direction. Table 3.6 Columns (2) and (3) split the data in this way and find that the estimates are not statistically different from one another and that the coefficients move in opposite directions relative to the baseline, providing no support for this type of bias-inducing selection.

Another potential obstacle to obtaining consistent estimates in this section is that individuals can choose the number of children they wish to have (or to claim on their tax return). Empirically, the literature suggests that these responses are relatively small (Baughman and Dickert-Conlin, 2003; 42 Paid preparer usage and a complete set of interactions with paid preparer are included in the estimating equation to control for underlying differences in investment income among those with a paid preparer.)
McCubbin, 2000). This will bias the estimates upwards if the correlation between choosing to have a second child to obtain a higher EITC amount and choosing to have lower dividend and interest income is greater than the same correlation for choosing to have a first child. The incentives for having a first child are higher than those for having a second for all years in the sample, but the incentives for claiming a second child have increased over time more than those for having a first child, and these changes are correlated with changes in the phase-out tax rate, which would bias the estimates upwards. While there is a strong incentive to claim a second child, there is no incentive to claim a third child, except in the state of Wisconsin.\(^{43}\) Table 3.6 Column (4) drops two-child families and instead includes three-child families, effectively removing the selection problem for two children, but leaving the selection for one child; now, any selection for one-child families that is correlated with the phase-out tax rate will bias the estimates downwards. There are about four times fewer three-child than two-child families, which increases the standard errors, all else equal. While the estimates are no longer statistically significant, the point estimates are close to the baseline estimates. These results provide no evidence that selection into having a second child plays a significant role in the baseline estimates.

While the existence of tax evasion could not bias the estimates, it would substantially change the interpretation and welfare implications of the results. In general, it is appropriate to assume that the elasticity of the tax evasion with respect to the marginal net-of-tax rate is a function of the costs associated with engaging in tax evasion.\(^{44}\) In this context, this implies that there is more evasion below the information reporting threshold at $10 and evasion will increase as the marginal tax rate increases. Figure 3.12 provides evidence on the amount of interest income and dividend income evasion below the information reporting line. The estimates are based on the density test proposed by McCrary (2008), which relies on the same bin size and bandwidth selection as described in Section 3.4, except that the bandwidth is chosen as the average of the bandwidth chosen below the information reporting notch and above the notch. In order to make a fair comparison, I choose the bandwidth when the data is restricted to $10 on either side of the tax notch (otherwise, the bandwidth would be far too large to obtain a reasonable estimate below the information reporting notch). The analysis is done in nominal dollars.\(^{45}\)

The top panel of Figure 3.12 examines interest income over years 1988-2006. There is a 120

\(^{43}\)Wisconsin provides more EITC to those with three children.

\(^{44}\)This point comes up frequently in the elasticity of taxable income literature (e.g., Saez et al., 2012).

\(^{45}\)The reason is that interest and dividend income is rounded to the nearest dollar on the tax form, so if I were to renormalize the variable to 0 at the information reporting notch and then convert the variable to real dollars, there are a large number at 0, which is just above the notch, and then a gap until plus or minus one. Moreover, the number now in minus 2 to minus 1 is lower because the inflation rate scales many of them out, but does not have the same effect at 0. Therefore, the estimates when using real dollars substantially overstate the true effect.
percent increase in the density just above the information reporting threshold and this estimate is statistically significant at the 1 percent level. This suggests that there is substantial evasion in interest income below $10, or alternatively individuals often forget to report their interest income unless they receive a tax form (1099-INT) at the end of the year with the amount earned, which only occurs if there was information reporting. If the elasticity with respect to the tax rate is positive and increasing with a decrease in the cost, the jump should increase after 1996, relative to the small EITC years (1988-1993); but neither of these separate estimates is statistically different from the estimate for all years, and the estimate for the years after 1996 is lower, not higher. This suggests that evasion occurs, but that it does not vary systematically with the tax rate. The bottom panel considers dividend income; however, the amount of dividend income in this range is about five times less, so the results are not statistically significant; in fact, there does not appear to be any evidence of evasion on this margin.

As an additional check on evasion below $10 in particular, I replace all reports less than or equal to $10 with zeros for all years. This forces the response of all with income less than $10 to be zero, which eliminates any evasion response of these individuals as well as any other response. These estimates are displayed in Column (5) of Table 3.6. They are slightly lower, but not statistically different from the baseline estimates. The decrease could be due to eliminating this form of tax evasion, but may also be due to the fact that I have eliminated the possibility of response for all those with dividend or interest income under $10.

3.5.1 Separating the Income and Substitution Effects

The estimation conducted up to this point has estimated a single parameter capturing the response to the marginal tax rate. This parameter encompasses both a substitution and an income effect. In particular, an increase in the phase-out tax rate has two effects: 1) the marginal cost of holding a dollar in an interest- or dividend-bearing account increases (substitution effect), and 2) the individuals’ wealth changes because the EITC amount changes (income effect). Wealth increases if the location of the end of the phase-out region is held fixed. Wealth decreases if the location of the beginning of the phase-out region is held fixed. This section adds a control for the income effect, so that the substitution effect—the main parameter of interest in this paper—is not biased. I find that the estimates in previous section were biased downwards, although the difference is not statistically significant.

The sign of the income effect on saving is ambiguous. As the EITC increases, individuals have more income, which may lead them to save more. As discussed in Subsection 3.2.2, a key mechanism
considered in the context of other need-based programs is that these programs decrease the need for precautionary savings. This is true in the context of the EITC as well because, for individuals in the phase-out region, a negative shock in income is met with an increase in the amount of EITC received. This suggests that the sign should be negative.

To estimate the two effects separately, I add a log function of after-tax earned income \((\text{earnings} + \tau(\text{end} - \text{earnings}))\), where \text{end} is the end of the phase-out region and \(\tau\) is the EITC phase-out marginal tax rate.\(^{46}\) Note that actual after-tax income is defined as \((\text{agi} + \tau(\text{end} - \text{agi}))\) (assuming \(\text{agi} > \text{earnings}\)) and \(\tau\) will change if individuals are not located in the phase-out region given their AGI. In this specification, individuals that decrease their investment income by a larger percentage in response to the tax rate change would receive a smaller income treatment by definition because AGI includes investment income. This would create a nontrivial bias in the estimates. Note that the existing literature that separately estimates the substitution and income effects (e.g. Gruber and Saez, 2002) estimates the income effect in this way, and therefore the results from this literature are biased. Even using earnings in the definition of the proxy for the EITC benefit amount requires a stronger assumption on earnings than before—earnings cannot respond to the marginal tax rate. If earnings respond and the response is positively correlated with the investment income response, this will create a downwards bias in the income effect estimates, but I find no evidence that earnings responds in this paper. Even if earnings responds, this bias is, by definition, smaller than the bias induced by using AGI, because earnings are one component of AGI. The proxy precisely captures the change in the EITC benefit received for a given level of earnings, as long as AGI is not outside the phase-out region. The income effect term can therefore be interpreted as an ITT estimate.

Empirically, the identification of the income effect separately from the substitution effect comes from two sources. One source is variation in \text{end}, which is the end of the phase-out region. The variable \text{end} increased for two-child taxpayers in 1996 and married taxpayers in 2002 and 2005. For this analysis to be valid, I must assume that marital status does not change endogenously with the changes in \text{end}. If more responsive individuals are more likely to get married once \text{end} increases, this will bias the income effect estimates upwards. The personal income tax schedule also changed in 2002 and the change affected married, head of household, and single individuals differently. Prior to 2002, all of the individuals who faced a positive personal income tax rate in the EITC sample examined here faced a 15 percent income tax. In 2002, a new 10 percent bracket was created. To ensure that all individuals experienced the same tax rate change in 2002 (so there is no differential effect of this

---

\(^{46}\)Gruber and Saez (2002) show that the income effect measure should take this form.
tax change which might be attributed to a change in end, I drop all single individuals. The income effect will also be partly identified from variation in $\tau$; changes in $\tau$ matter for both the substitution and income effect terms, but they enter each in a different way. Note that without including flexible controls in earnings, identification of the income effect would also come from variation in earnings. However, investment income likely increases with earnings, regardless of the tax rate, making this variation invalid.

The results are presented in Table 3.7. Columns (1) and (2) replicate the baseline specification given in Columns (1) and (4) in Table 3.5 with the restricted sample used in this subsection. Columns (3) and (4) add the income effect term to this specification. There is not enough variation to separately identify the income effect at a statistically significant level. The substitution effect gets slightly larger when controlling for the income effect, but the difference is not statistically significant. When the income effect is positive and it is ignored in the estimation, the estimates should be biased downwards because an increase in the phase-out tax rate usually coincided with an increase in EITC benefits in the period examined.

### 3.6 Conclusion

When designing the EITC, non-labor income was likely included as a determinant of the amount of EITC received to ensure the credit was going to those that were low-wealth. This paper finds that an unintended consequence of this provision is to substantially distort non-labor income, particularly investment income; in fact, along the intensive margin, the non-labor income distortions are far larger than the earned income distortions for wage-earning individuals. Such behavior induces deadweight loss because engaging in such behavior is not costless for these individuals regardless of whether a decline in investment income translates into a decline in savings. To the extent that the behavioral response estimates in this paper reflect an actual decline in savings, this creates additional concerns, particularly since the government would like to encourage low-income households to save more, as evidenced by programs like the Saver’s Credit and Individual Development Accounts.

This paper provides evidence that the response in investment income is substantial. Indeed, nearly 40 percent of the decline in saving in income-bearing accounts by EITC recipients over the last two decades can be explained by the changing incentives for saving caused by the EITC schedule. The response is twice as large among those that use a paid preparer, consistent with an increased

---

47 Most unmarried individuals with children file as a head of household, so this does not drop many individuals from the sample.

48 A quadratic in earnings is used in the paper; however the results are not sensitive to including a cubic or quartic polynomial in earnings instead.
awareness of the relevant incentives among this group.

The policy implications for EITC design depend on how policymakers weight the deadweight loss induced by this provision, relative to its original intent—to exclude individuals that were not actually poor from receiving the EITC. Making the EITC amount exclusively a function of earned income would eliminate the distortion found in this paper, but would increase the number of claimants by 4 percent; these additional claimants would receive an average of $1,014 in benefits, even though their average AGI is $52,179.
This figure displays the EITC schedule in earned income for taxpayers with two dependents in 1996-2006. 75 percent of these taxpayers have unearned income less than or equal to zero and thus face the EITC schedule given by the solid line. 25 percent have positive unearned income and face an EITC schedule, such that the phase-out line is shifted in towards the dashed line. For about a quarter of these individuals, the line is shifted far enough that the plateau region is eliminated; they face a spike in their EITC schedule.
The top figure plots the entire 3-dimensional surface in earned income, unearned income, and the EITC amount for taxpayers with two dependents in 1996-2006. The bottom figure provides contour lines of the top figure. When earned income is low, EITC benefits remain low, regardless of unearned income levels. This is because, when earned plus unearned income (AGI) is large enough to reach the second kink, individuals are still forced to take the benefit based on their earned income because it is lower than their AGI benefit. As earned income rises, the EITC first rises and then falls, either with or without a plateau region depending on the amount of unearned income, as was depicted in Figure 3.1. As unearned income rises, it first has no effect on the EITC and then causes it to decline at a certain threshold, where the threshold varies depending on the individual’s earned income amount.
This figure displays the main EITC calculation form for 2006. On line 1, individuals are asked to write down their earned income. On line 2, they calculate their credit using a tax table, and are asked to stop if they are ineligible. On line 3, they enter their AGI. On line 4, they are asked whether their earned income and AGI are equal (i.e., they have no unearned income). If they are, they can skip the rest of the form and enter the credit amount from line 2 on line 6, where line 6 is the amount of EITC claimed. If earned income and AGI are not equal, they proceed to the next question before line 5, which asks them whether their AGI places them in the phase-out region. If it does not, they again skip the rest of the form and enter the credit amount from line 2 on line 6. If their AGI does place them in the phase-out region, line 5 asks them to calculate their EITC based on their AGI. They are then required to take the minimum of lines 2 and 5 and put this amount on line 6—this is the amount of EITC they will receive.
Figure 3.4: Estimating Elasticity from Bunching Around a Tax Kink

This figure illustrates how bunching at a tax kink can be used to identify the response to the change in marginal tax rate at the tax kink. $z^*$ is the tax kink, $2\delta_b$ is the width of the imperfect bunching region, $\delta_c$ is the width of the counterfactual region. $h^*_-$ and $h^*_+$ are the densities in the counterfactual regions, $h^*$ is the density in the bunching region, each of which is depicted by the blue line in its respective region. $H^* = 2\delta_b h^*$ is the cumulative density in the imperfect bunching region. $H^*_- = \delta_c h^*_-$ and $H^*_+ = \delta_c h^*_+$ are the cumulative densities in the counterfactual region. $B$ is the amount of imperfect bunching in the bunching region, which is given by the area between the solid blue and dashed yellow lines.
All adjusted gross income (AGI) values are in 2006 dollars, renormalized so that the kink ($z^*$) is at zero for all years. The estimates are population weighted. The density estimates are normalized to integrate to one. 95 percent confidence intervals are given by the dashed lines. The scatter-plot provides the unsmoothed density estimates. The region inside the long-dash vertical lines is used to calculate the bunching at the tax kink, and the region between the short-dash and long-dash lines is used to calculate the counterfactual density. The implied counterfactual is given by the solid yellow line. The bin size and bandwidth are 27 and 797 for the top panel and are 83 and 1,469 for the bottom panel.
All adjusted gross income (AGI) values are in 2006 dollars, renormalized so that the kink ($z^*$) is at zero for all years. The estimates are population weighted. The density estimates are normalized to integrate to one. 95 percent confidence intervals are given by the dashed lines. The scatter-plot provides the unsmoothed density estimates. The region inside the long-dash vertical lines is used to calculate the bunching at the tax kink, and the region between the short-dash and long-dash lines is used to calculate the counterfactual density. The implied counterfactual is given by the solid yellow line. The bin size and bandwidth are 17 and 745 for the top panel and are 55 and 2,310 for the bottom panel.
This figure shows how a lower bound on the elasticity of response of a given income type $X$ can be calculated in the region around the tax kink. $X^*_+$ (the solid blue line) gives the realized amount of investment income. $X^*_-$ (the yellow dashed line) gives the counterfactual amount of investment income. The shaded grey region is the implied decline in investment in the region around the tax kink that is due to the increase in marginal tax rate above the kink.
Wage-earners for years 1988-2006 with no unemployment insurance income are included. All adjusted gross income (AGI) values are in 2006 dollars, renormalized so that the kink ($z^*$) is at zero for all years. The estimates are population weighted. The region inside the dashed vertical lines is used to calculate the elasticity. Realized average income is given by the solid blue line. The implied counterfactual is given by the dashed yellow line and is constructed using investment income in the region between the short-dash and long-dash vertical lines (-4,000 to -3,000). The bin size and bandwidth for earned income are 31 and 2,518 and are 31 and 3,787 for unearned income.
Panel A includes wage-earners for years 1988-2006 with no unemployment insurance income whose EITC investment income is less than $2,300. Panel B includes all individuals with positive self-employment income and no unemployment insurance income for years 1988-1993. The estimates are population weighted. All adjusted gross income (AGI) values are in 2006 dollars, renormalized so that the kink ($z^*$) is at zero for all years. The region inside the long-dash vertical lines is used to calculate the elasticity. Realized average income is given by the solid blue line. The implied counterfactual is given by the dashed yellow line and is constructed using investment income in the region between the short-dash and long-dash vertical lines (-4,000 to -3,000). The bin size and bandwidth are 31 and 2,417 for Panel A and are 88 and 1,384 for Panel B.
Figure 3.10: Variation in the Marginal Net-of-Tax Rate by Number of Dependents

This figure plots the variation in the marginal net-of-tax phase-out rate by year and number of dependents. The marginal net-of-tax rate for one dependent is depicted by the blue line with circles and the marginal net-of-tax rate for two or more dependents is depicted as the yellow line with diamonds. The marginal net-of-tax rate is one minus the marginal tax rate.

Figure 3.11: Likelihood Individuals have Investment Income by Year and Number of Dependents

This figure plots the likelihood an individual holds investment income by year for one (blue line) and two dependents (yellow line) based on the restrictions used in the baseline estimates, with a break at 1996 (marked by the vertical black line). The thin dashed lines are 95 percent confidence intervals. The scatter plot gives the value for each dependent-year pair. The lines are from a cubic polynomial regression in year interacted with a pre/post indicator and number of dependents.
These estimates include any individual with positive interest or dividend income, subject to the baseline sample restrictions in Section 3.5. The implied log difference in density at the information reporting threshold for interest income is 1.21 (0.21) and for dividend income is 0.18 (0.33). Years 1996-2006 are included. Interest and dividend income is in nominal dollars. The estimates are population weighted. The density estimates are normalized to integrate to one. The smoothed density estimates do not start at zero because they cannot be estimated accurately (or could, but not with the same bandwidth) until interest or dividend income is greater than or equal to the bandwidth.
Table 3.1: EITC Schedule Parameters

<table>
<thead>
<tr>
<th>EITC Range</th>
<th>Bracket</th>
<th>Marginal Tax Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1979-1984 (in 2006 $), Qualifying Children ≥ 1 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$9,779</td>
<td>-10%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$9,779-$11,735</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$11,735-$19,559</td>
<td>12.5%</td>
</tr>
<tr>
<td><strong>1985-1986 (in 2006 $), Qualifying Children ≥ 1 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$9,061</td>
<td>-11%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$9,061-$11,780</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$11,780-$19,935</td>
<td>12.22%</td>
</tr>
<tr>
<td><strong>1987 (in 2006 $), Qualifying Children ≥ 1 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$10,743</td>
<td>-14%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$10,743-$12,227</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$12,227-$27,267</td>
<td>10%</td>
</tr>
<tr>
<td><strong>1988-1990 (in 2006 $), Qualifying Children ≥ 1 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$10,747</td>
<td>-14%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$10,747-$16,933</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$16,933-$31,979</td>
<td>10%</td>
</tr>
<tr>
<td><strong>1991 (in 2006 $), Qualifying Children = 1 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$10,748</td>
<td>-16.7%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$10,748-$16,935</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$16,935-$31,988</td>
<td>11.93%</td>
</tr>
<tr>
<td><strong>1991 (in 2006 $), Qualifying Children ≥ 2 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$10,608</td>
<td>-17.3%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$10,608-$16,935</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$16,935-$31,988</td>
<td>12.36%</td>
</tr>
<tr>
<td><strong>1992 (in 2006 $), Qualifying Children = 1 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$10,752</td>
<td>-17.6%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$10,752-$16,929</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$16,929-$31,986</td>
<td>12.57%</td>
</tr>
<tr>
<td><strong>1992 (in 2006 $), Qualifying Children ≥ 2 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$10,752</td>
<td>-18.4%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$10,752-$16,929</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$16,929-$31,986</td>
<td>13.14%</td>
</tr>
<tr>
<td><strong>1993 (in 2006 $), Qualifying Children = 1 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$10,754</td>
<td>-18.5%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$10,754-$16,928</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$16,928-$31,983</td>
<td>13.21%</td>
</tr>
<tr>
<td><strong>1993 (in 2006 $), Qualifying Children ≥ 2 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$10,754</td>
<td>-19.5%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$10,754-$16,928</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$16,928-$31,983</td>
<td>13.93%</td>
</tr>
<tr>
<td><strong>1994 (in 2006 $), Qualifying Children = 1 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$10,434</td>
<td>-26.3%</td>
</tr>
<tr>
<td>Plateau</td>
<td>$10,434-$14,810</td>
<td>0%</td>
</tr>
<tr>
<td>Phase-out</td>
<td>$14,810-$31,983</td>
<td>15.98%</td>
</tr>
<tr>
<td><strong>1994 (in 2006 $), Qualifying Children ≥ 2 :</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase-in</td>
<td>0-$11,343</td>
<td>-30%</td>
</tr>
</tbody>
</table>
Plateau $11,343-$14,810 0%
Phase-out $14,810-$34,058 17.68%

1995 (in 2006 $), Qualifying Children = 1 :³
Phase-in 0-$8,083 -34%
Plateau $8,083-$14,815 0%
Phase-out $14,815-$32,013 15.98%

1995 (in 2006 $), Qualifying Children ≥ 2 :³
Phase-in 0-$11,338 -36%
Plateau $11,338-$14,815 0%
Phase-out $14,815-$35,001 20.22%

1996-2006 (in 2006 $), Qualifying Children = 1 :³,4
Phase-in 0-$8,080 -34%
Plateau $8,080-$14,810 0%
Phase-out $14,810-$32,001 15.98%

1996-2006 (in 2006 $), Qualifying Children ≥ 2 :³,4
Phase-in 0-$11,340 -34%
Plateau $11,340-$14,810 0%
Phase-out $14,810-$36,348 21.06%

¹The EITC schedule was fixed in nominal dollars in these years. Therefore, in real terms, the bracket cutoffs declined substantially over this time. In each case, the real values in the last year of the range are listed.
²The beginning of the phase-out range is underlined in each year because it marks the location of the second tax kink.
³The parameters are inflation-indexed, but are then rounded. In each case, the real values based on the actual values in the last year in the range are listed.
⁴For married taxpayers, the phase-out region was shifted right by $1,000 in 2002-2004 and $2,000 in 2005-2006.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Fraction ≠ 0</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wage Income</td>
<td>105,848</td>
<td>16,017.45</td>
<td>9,672.61</td>
<td>0.984</td>
<td>0.124</td>
</tr>
<tr>
<td>Self-Employment Income</td>
<td>105,848</td>
<td>1,042.46</td>
<td>4,517.18</td>
<td>0.157</td>
<td>0.364</td>
</tr>
<tr>
<td>Unearned Income</td>
<td>105,848</td>
<td>571.66</td>
<td>10,649.28</td>
<td>0.336</td>
<td>0.472</td>
</tr>
<tr>
<td>Dividend &amp; Interest Income</td>
<td>105,848</td>
<td>104.69</td>
<td>1,677.13</td>
<td>0.186</td>
<td>0.389</td>
</tr>
<tr>
<td>Capital Gains or Losses</td>
<td>105,848</td>
<td>8.57</td>
<td>1,828.51</td>
<td>0.022</td>
<td>0.147</td>
</tr>
<tr>
<td>Partnership Income</td>
<td>105,848</td>
<td>-66.17</td>
<td>10,386.36</td>
<td>0.031</td>
<td>0.174</td>
</tr>
<tr>
<td>Other Gains or Losses</td>
<td>105,848</td>
<td>-14.42</td>
<td>3,178.40</td>
<td>0.005</td>
<td>0.067</td>
</tr>
<tr>
<td>Taxable Refunds</td>
<td>105,848</td>
<td>16.46</td>
<td>296.59</td>
<td>0.032</td>
<td>0.176</td>
</tr>
<tr>
<td>Taxable Pension, Annuity &amp; IRA Distributions</td>
<td>105,848</td>
<td>16.38</td>
<td>296.67</td>
<td>0.051</td>
<td>0.219</td>
</tr>
<tr>
<td>Social Security Income</td>
<td>105,848</td>
<td>5.11</td>
<td>118.08</td>
<td>0.003</td>
<td>0.055</td>
</tr>
<tr>
<td>Unemployment Insurance</td>
<td>105,848</td>
<td>335.56</td>
<td>1,372.48</td>
<td>0.107</td>
<td>0.309</td>
</tr>
<tr>
<td>Deductions</td>
<td>105,848</td>
<td>103.94</td>
<td>855.12</td>
<td>0.050</td>
<td>0.218</td>
</tr>
<tr>
<td>Married?</td>
<td>105,848</td>
<td>0.28</td>
<td>0.45</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Number of Dependents</td>
<td>105,848</td>
<td>1.41</td>
<td>0.49</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tax Preparer?</td>
<td>102,161</td>
<td>0.62</td>
<td>0.49</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Claim EITC?</td>
<td>105,848</td>
<td>0.92</td>
<td>0.28</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

This table reports summary statistics for all EITC eligible individuals with one or two dependents. Tax preparation variables are missing for 1990 and are occasionally missing in other years. All income values are in 2006 dollars.
Table 3.3: Elasticities at Second EITC Kink

<table>
<thead>
<tr>
<th></th>
<th>Bunching Elasticity Estimates</th>
<th>Income Elasticity Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity</td>
<td>(1)</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>0.063***</td>
<td>−0.015</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Type</td>
<td>Self-employed</td>
<td>Wage-earner</td>
</tr>
<tr>
<td></td>
<td>Wage-Earner</td>
<td>Wage-earner</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>(7)</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>0.516</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.451)</td>
</tr>
<tr>
<td>Income Type</td>
<td>AGI</td>
<td>Earned</td>
</tr>
<tr>
<td>Years</td>
<td>1988-2006</td>
<td>1988-2006</td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>(8)</td>
</tr>
<tr>
<td></td>
<td>1.04</td>
<td>1.879**</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.832)</td>
</tr>
<tr>
<td></td>
<td>(4)</td>
<td>(9)</td>
</tr>
<tr>
<td></td>
<td>0.095*</td>
<td>1.125</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(1.060)</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.104**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td></td>
</tr>
<tr>
<td>Years</td>
<td>1988-1993</td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ_b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>3,000</td>
</tr>
<tr>
<td></td>
<td>δ_c</td>
<td>1,000</td>
</tr>
</tbody>
</table>

The bunching estimates are calculated by first calculating the amount of imperfect bunching around the tax kink using a local linear regression as described in the text and the corresponding figures. Then, equation (3.1) is used to construct elasticities based on this estimate. The income elasticity estimates are calculated by first calculating average income of the given income type in the region around the tax kink using a local linear regression as described in the text and the corresponding figures. Then, equation (3.2) is used to calculate elasticities based on this estimate. The standard errors are calculated via non-parametric bootstrap (500 replications). These estimates exclude individuals that did not file for the EITC, had no dependents, or had more than two dependents. The wage-earner estimates include only wage-earners and the self-employed include only self-employed individuals. $\delta_b$ is the width of the imperfect bunching region and $\delta_c$ is the width of the counterfactual region.
Table 3.4: Repeated-Cross-Section Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Fraction ≠ 0</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earned Income</td>
<td>31,480</td>
<td>23,679.15</td>
<td>3,444.38</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unearned Income</td>
<td>31,480</td>
<td>1,160.53</td>
<td>7,695.71</td>
<td>0.351</td>
<td>0.477</td>
</tr>
<tr>
<td>Dividend &amp; Interest Income</td>
<td>31,480</td>
<td>48.60</td>
<td>205.94</td>
<td>0.206</td>
<td>0.404</td>
</tr>
<tr>
<td>Capital Gains or Losses</td>
<td>31,480</td>
<td>-10.35</td>
<td>227.29</td>
<td>0.016</td>
<td>0.125</td>
</tr>
<tr>
<td>Partnership Income</td>
<td>31,480</td>
<td>245.71</td>
<td>5,971.41</td>
<td>0.028</td>
<td>0.164</td>
</tr>
<tr>
<td>Other Gains or Losses</td>
<td>31,480</td>
<td>-7.32</td>
<td>1,777.94</td>
<td>0.001</td>
<td>0.029</td>
</tr>
<tr>
<td>Unemployment Insurance</td>
<td>31,480</td>
<td>316.48</td>
<td>1,389.29</td>
<td>0.100</td>
<td>0.300</td>
</tr>
<tr>
<td>Married?</td>
<td>31,480</td>
<td>0.28</td>
<td>0.45</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Number of Dependents</td>
<td>31,480</td>
<td>1.40</td>
<td>0.49</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tax Preparer?</td>
<td>30,217</td>
<td>0.59</td>
<td>0.49</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EITC Eligible?</td>
<td>31,480</td>
<td>0.96</td>
<td>0.19</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

This table reports summary statistics for individuals in the baseline estimates. Tax preparation variables are missing for 1990 and are occasionally missing in other years. All income values are in 2006 dollars.
Table 3.5: Baseline Estimates

<table>
<thead>
<tr>
<th>Extensive Margin:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log(1 - \tau) )</td>
<td>0.612***</td>
<td>0.613***</td>
<td>0.868***</td>
<td>0.816***</td>
<td>0.867***</td>
<td>0.824***</td>
<td>0.353*</td>
</tr>
<tr>
<td></td>
<td>(0.174)</td>
<td>(0.156)</td>
<td>(0.213)</td>
<td>(0.221)</td>
<td>(0.267)</td>
<td>(0.217)</td>
<td>(0.188)</td>
</tr>
<tr>
<td>( \log(1 - \tau) \times \text{married} )</td>
<td>-0.098</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.306)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \log(1 - \tau) \times \mathbb{1}[ui &gt; 0] )</td>
<td>0.114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.240)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \log(1 - \tau) \times \text{pdprep} )</td>
<td>0.7366***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value of sum</td>
<td>0.005</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Elasticity:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log(1 - \tau) )</td>
<td>3.047***</td>
<td>3.000***</td>
<td>3.682***</td>
<td>3.435***</td>
<td>3.897***</td>
<td>3.505***</td>
<td>1.292*</td>
</tr>
<tr>
<td></td>
<td>(0.810)</td>
<td>(0.724)</td>
<td>(0.835)</td>
<td>(0.936)</td>
<td>(1.061)</td>
<td>(0.917)</td>
<td>(0.748)</td>
</tr>
<tr>
<td>( \log(1 - \tau) \times \text{married} )</td>
<td>-0.806</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.605)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \log(1 - \tau) \times \mathbb{1}[ui &gt; 0] )</td>
<td>0.171</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.306)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \log(1 - \tau) \times \text{pdprep} )</td>
<td>3.3626***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value of sum</td>
<td>0.021</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Broad Measure of Investment Income?</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>No</th>
<th>No</th>
<th>No</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls x 1992, 1998 included?</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State FE included?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The extensive margin estimates are mean marginal effects from a probit specification, where the dependent variable is an indicator variable for whether the taxpayer had positive interest or dividend income. The elasticity estimates are mean marginal elasticities from a two-tier model, where the dependent variable in the first tier is an indicator variable for whether the taxpayer had positive interest or dividend income and the dependent variable in the second tier is the log of interest and dividend income. The dependent variables in the second column are constructed from the sum of dividend, interest, partnership, capital gains and losses, other gains and losses, and IRA contributions. All columns include covariates \( X \) (as specified in the paper) and year fixed-effects. All income values are in 2006 dollars. Estimates are population weighted. Estimates are for years 1988-2006. Standard errors clustered by year are in parentheses.
Table 3.6: Robustness Checks

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extensive Margin:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\log(1 - \tau)$</td>
<td>0.816***</td>
<td>0.912***</td>
<td>0.711***</td>
<td>0.596</td>
<td>0.538**</td>
</tr>
<tr>
<td></td>
<td>(0.221)</td>
<td>(0.305)</td>
<td>(0.197)</td>
<td>(0.579)</td>
<td>(0.253)</td>
</tr>
<tr>
<td>Elasticity:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\log(1 - \tau)$</td>
<td>3.435***</td>
<td>2.539**</td>
<td>4.554***</td>
<td>3.063</td>
<td>2.981***</td>
</tr>
<tr>
<td></td>
<td>(0.936)</td>
<td>(1.141)</td>
<td>(0.920)</td>
<td>(2.496)</td>
<td>(0.966)</td>
</tr>
<tr>
<td>Controls x 1992, 1998 included?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State FE included?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Two dependent taxpayers included?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Three dependent taxpayers included?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>31,480</td>
<td>16,821</td>
<td>14,652</td>
<td>22,996</td>
<td>31,480</td>
</tr>
</tbody>
</table>

The extensive margin estimates are mean marginal effects from a probit specification, where the dependent variable is an indicator variable for whether the taxpayer had positive interest or dividend income. The elasticity estimates are mean marginal elasticities from a two-tier model, where the dependent variable in the first tier is an indicator variable for whether the taxpayer had positive interest or dividend income and the dependent variable in the second tier is the log of interest and dividend income. The last column replaces all values for interest and dividend income with zero whenever these income types are less than $10 before constructing the dependent variables. All columns include covariates $X$ (as specified in the paper) and year fixed-effects. All income values are in 2006 dollars. Estimates are population weighted. Estimates are for years 1988-2006. Standard errors clustered by year are in parentheses.
Table 3.7: Substitution and Income Effects

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extensive Margin:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\log(1 - \tau)$</td>
<td>0.638***</td>
<td>0.781***</td>
<td>0.810*</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>(0.190)</td>
<td>(0.268)</td>
<td>(0.445)</td>
<td>(0.617)</td>
</tr>
<tr>
<td>$\log(earnings + \tau(\text{end-earnings}))$</td>
<td>0.202</td>
<td>0.202</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.389)</td>
<td>(0.611)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elasticity:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\log(1 - \tau)$</td>
<td>3.144***</td>
<td>3.347***</td>
<td>4.456**</td>
<td>4.501</td>
</tr>
<tr>
<td></td>
<td>(0.889)</td>
<td>(1.135)</td>
<td>(2.066)</td>
<td>(3.250)</td>
</tr>
<tr>
<td>$\log(earnings + \tau(\text{end-earnings}))$</td>
<td>1.543</td>
<td>1.377</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.813)</td>
<td>(3.204)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls x 1992, 1998 included?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>State FE included?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Single Filers Included</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>30,266</td>
<td>30,266</td>
<td>30,266</td>
<td>30,266</td>
</tr>
</tbody>
</table>

The extensive margin estimates are mean marginal effects from a probit specification, where the dependent variable is an indicator variable for whether the taxpayer had positive interest or dividend income. The elasticity estimates are mean marginal elasticities from a two-tier model, where the dependent variable in the first tier is an indicator variable for whether the taxpayer had positive interest or dividend income and the dependent variable in the second tier is the log of interest and dividend income. All columns include covariates $X$ (as specified in the paper) and year fixed-effects. All income values are in 2006 dollars. Estimates are population weighted. Estimates are for years 1988-2006. Standard errors clustered by year are in parentheses. Note that single filers is referring to the taxpayer category; unmarried head of household filers are still included.
CHAPTER IV

Obtaining a Consistent Estimate of the Elasticity of Taxable Income Using Difference-in-Differences

4.1 Introduction

Shortly after significant income tax rate reductions due to the Tax Reform Act of 1986 (TRA86) in the U.S., tax researchers began to estimate individuals’ responses to taxation, as measured by their reported taxable income. The elasticity of taxable income (ETI) is the percent change in individuals’ reported taxable income in response to a one percent change in their marginal net-of-tax rate.\(^1\) An individual’s response to a tax change could take a number of forms including a labor supply response, a change in tax avoidance strategies (e.g. changing the amount of itemized deductions accrued), or a change in the extent of tax evasion. The ETI captures all of these. It is informative on its own and has also been shown, under certain assumptions, to be a sufficient statistic for marginal deadweight loss.\(^2\) Therefore, it has been a popular parameter for public finance economists to estimate and obtaining a consistent estimate is valuable for policy debates.

One necessary condition for consistency—instrument exogeneity\(^3\)—remains a topic of substantial discussion in the literature. Concerns about endogeneity of proposed instruments for the independent variable of interest—the log change in the marginal net-of-tax rate\(^4\)—have given rise to proposals of numerous ways to adjust the standard difference-in-differences estimating model to address this endogeneity. Kopczuk (2005) and Giertz (2008) have examined many of these proposals simultaneously and shown that there is an alarming degree of variation in the ETI estimates based on U.S. data depending on the exact model chosen (both find estimates ranging from -1 to 1). Kopczuk (2003, 2005) agrees that this variation is due to varying degrees of estimating model mis-specification, but

---

\(^{1}\) The marginal net-of-tax rate is one minus the marginal tax rate.

\(^{2}\) For example, see Feldstein (1999) and Chetty (2009b).

\(^{3}\) I use instrument exogeneity throughout this paper as it is used in the ETI literature; that is, an instrument is exogenous if the instrument is uncorrelated with transitory shocks in the error term.

\(^{4}\) The standard estimating equation regresses the log change in taxable income on the log change in the marginal net-of-tax rate and other covariates.
does not formally prove whether any of the variants he examines provide a consistent estimate.

This paper examines which methods provide a consistent estimate of the ETI and proposes new methods when necessary, where the conditions required for consistency when estimating the response to a marginal tax rate change are laid out in Weber (2012b). I use the Michigan IRS Tax Panel data set for the years 1979-1990 for empirical applications.

The first main contribution of this paper is to define an income process that fits within the standard estimation strategy and derive the conditions necessary for a potential instrument to be exogenous in this context. This paper focuses exclusively on instruments that remain a function of taxable income and formally shows that, under reasonable assumptions, most of these existing instruments are not exogenous. Empirical tests for exogeneity support the theoretical results and both types of results show that the addition of various forms of income-based controls, while popularly believed to help eliminate the endogeneity of the most commonly used instrument (the change in the predicted net-of-tax rate\(^5\)), are not effective in the U.S. context.

The second key contribution of this paper is to propose a new instrument that is exogenous under testable assumptions regarding the degree of serial correlation in the error term. Given the additional requirements for a consistent estimate derived in Weber (2012b), this is the only instrument that has the potential to provide a consistent estimate under reasonable assumptions. Using this instrument, my preferred baseline estimate is 1.046. There are likely two main reasons why my preferred baseline estimate is more than twice as large as the estimates found in the most frequently cited paper on this subject (Gruber and Saez, 2002). First, I show that the instrument used in (Gruber and Saez, 2002) biases the estimates downwards when the tax reform decreases marginal tax rates at the top of the income distribution. Second, with an additional auxiliary assumption, this parameter can be interpreted as a Fixed-Bracket Average Treatment Effect (FBATE); that is, the parameter is identified from the subpopulation of taxpayers with no incentive to cross a tax bracket line due to a tax reform or transitory income shock (Weber, 2012b). This subpopulation may be more responsive during the time period in which the estimation occurs; however, if those excluded will respond the same, on average, in the long-run, FBATE is the relevant parameter for welfare analysis.

The paper also addresses the use of income splines to control for heterogeneous income trends at different ranges in the income distribution. While theoretically a reasonable idea, the implementation is often such that the splines are endogenous; in fact, eliminating endogeneity of the splines changes the sign of the majority of spline coefficients in my preferred baseline estimates. Overall, controlling

\(^5\)The predicted net-of tax rate change is the change in the net-of-tax rate if an individual had their base-year income in both years.
for heterogeneous income trends plays a relatively minor role when the difference length is short (i.e. one year), but as expected, increases as the difference length increases.

4.2 Background

This section briefly reviews the evolution of the ETI estimation literature. A variety of estimation methods have been employed to estimate the ETI, including difference-in-differences based on repeated cross-sections, share analysis, and panel-based difference-in-differences.\textsuperscript{6} The latter has been used most often and will be the focus of this paper.\textsuperscript{7} Identification usually comes from differences in tax rate changes across individuals brought about by a tax reform.\textsuperscript{8} Given that tax reforms frequently change tax rates more for high income individuals, identification is often obtained by comparing the taxable income response of individuals at the top of the income distribution who experience a large tax rate change to those lower in the income distribution who experience a low or no tax rate change. Estimating the ETI accurately requires data that will provide a precise measure of individuals’ taxable income, which makes administrative tax return data attractive.

The ETI is obtained by regressing the log change in taxable income on the log change in the net-of-tax rate (as well as other covariates). Without using an instrument for the log change in the net-of-tax rate, it is clearly endogenous because it is a function of taxable income—the dependent variable in the regression. As a result, all regression-based studies of the ETI use an instrument for the log change in the net-of-tax rate. The most common instrument is the value for the change in the net-of-tax rate given the tax reform if individuals earned their base-year income (income is income in the first year of the difference) in both years. Instruments that are only a function of taxable income are employed because tax return data sets are normally used and usually do not have rich demographic data, which could provide alternative instruments.\textsuperscript{9} Since the instrument is still a function of the dependent variable, there is no guarantee that the instrument employed is exogenous. The literature has identified two problems that can cause remaining endogeneity of the instrument: mean reversion and heterogeneous income trends. Both of these problems will be discussed extensively below, so I will hold off on providing a formal definition until then. All researchers that employ this instrument have included some function of base-year income in hope of

\textsuperscript{6}For a comparison of these methods, see Saez et al. (2012).
\textsuperscript{7}Throughout the paper, difference-in-differences is synonymous with panel-based difference-in-differences.
\textsuperscript{8}Papers that obtain identification by other means are beyond the scope of this paper.
\textsuperscript{9}At least this has been traditionally true. But, more recently, data sets from other countries which have much better demographic data have been employed. And, in the U.S., Singleton (2011) has linked two different data sets, one of which also provides much better demographic data. However, alternative instruments based on demographics have not been used in any of these studies. The only study to make use demographic instruments to estimate the ETI was Moffitt and Wilhelm (2000) who used the SCF, not an actual tax return data set.
resolving these two problems.

Early estimates of the ETI were based on U.S. data from the 1980’s, where the major federal tax reforms were the Economic Recovery Tax Act of 1981 (ERTA81) and TRA86. These were predominantly tax decreases, and produced estimates within the range of 0.4-0.62 depending on the functional form of base-year income used as a regressor (Auten and Carroll, 1999; Gruber and Saez, 2002). The ETI was also estimated using 1990’s data, in which the predominant federal reforms were targeted tax increases (the Omnibus Budget Reconciliation Acts of 1990 and 1993). The estimates ranged from 0.17-0.38 (Carroll, 1998; Giertz, 2005). There has since been a large literature using the same methods to estimate the ETI in other countries (Saez et al., 2012).

A more recent literature has suggested that there is no guarantee that the base-year income controls selected in these early papers will resolve the endogeneity of the instrument. Additionally, concerns have been raised about what is the appropriate comparison group (i.e. should it be all individuals that do not receive a tax change in the tax reform considered or some subset that are nearest in income level to the treated individuals). Kopczuk (2005) and Giertz (2008) conduct sensitivity analyses to document the instability of the estimates to the choice of base-year income covariates (they consider a much wider range of functional forms for the income controls than those used by previous authors) and comparison group. They find estimates that range from less than -1 to greater than 1. Kopczuk (2003, 2005) agrees that this variation is due to varying degrees of model mis-specification, but does not derive the exact nature of the biases in each. Several authors have tried to get around this problem by proposing alternative instruments (e.g., Blomquist and Selin, 2010). Still, the estimation methods commonly employed remain those laid out in Auten and Carroll (1999) or Gruber and Saez (2002).

This paper examines each of the proposed estimators to determine whether any of them provide consistent estimates of the ETI and proposes new methods when necessary. I begin in Section 4.3 by setting up a simple model of income and within this, characterizing mean reversion and income trends. In Section 4.4, I show theoretically and empirically which instruments and base-year income controls are appropriate to obtain a consistent estimate of the ETI, obtain ETI estimates, and provide an interpretation of these estimates. Section 4.5 concludes.
4.3 Model Setup

To facilitate the analysis of the issues inherent in estimating the ETI, I set up a simple model of the taxable income process in this section.\(^{10}\) This process is consistent with the theoretical model that drives estimation strategies in this literature that is formally laid out in Gruber and Saez (2002). Of course, the actual process may be more complex than that laid out in this section.\(^{11}\) If such additional complexities matter, they will need to be addressed in order to obtain a consistent estimate of the ETI. Hence, to the extent that the case considered here is a special case in more complex estimation strategies, one can think about the analysis based on this model as providing necessary conditions for obtaining a consistent estimate of the ETI, but perhaps not sufficient.\(^ {12}\)

Let individuals’ log income $\ln(Y_{it})$ be governed by the following equation:\(^ {13}\)

\begin{equation}
\begin{aligned}
\ln(Y_{it}) &= \varepsilon \ln(1 - \tau_{it}^\tau) + \ln(\mu_{it}) + \ln(\nu_{it}), \\
\text{or in differences as:} \\
\Delta \ln(Y_{it}) &= \varepsilon \Delta \ln(1 - \tau_{it}^\tau) + \Delta \ln(\mu_{it}) + \Delta \ln(\nu_{it}),
\end{aligned}
\end{equation}

where $\Delta \ln(Y_{it}) = \ln \left( \frac{Y_{it}}{Y_{it-1}} \right)$, $\mu_{it}$ is permanent income, $\nu_{it}$ is transitory income, and $\varepsilon$ is the ETI.\(^ {14}\) Additionally, $\tau_{it}^\tau$ is the marginal tax rate to which an individual responds, which is a function of pre-response income $Y_{it}^\tau$ and the tax code $c_t$:

\begin{equation}
\tau_{it}^\tau = f(Y_{it}^\tau, c_t),
\end{equation}

where $\ln(Y_{it}^\tau) = \ln(\mu_{it}) + \ln(\nu_{it})$. For now, assume that all individuals’ income grows at the same rate on average, regardless of their income level (that is, assume homogeneous income trends). Then I can define $g_t = \Delta \ln(\mu_t)$ as the period-specific homogeneous income growth rate. Note that suppressed in $g_t$ is everything that affects the income-growth profile of an individual, which is

\footnotesize
\(^{10}\)Sometimes in the ETI literature, the income concept considered is broad income, rather than taxable income. Broad income generally refers to income before deductions, credits and itemization. Assume income means taxable income throughout the paper unless stated otherwise.\(^ {11}\)For example, the literature has explored the role of income effects (Gruber and Saez, 2002) and tax-base effects (Kopczuk, 2005).\(^ {12}\)For expositional ease, I choose to examine a simple framework in which all the important issues arise. It is relatively clear how to apply the methods described in Subsection 4.4.2.2 to obtain consistent estimates for many extensions examined in the literature.\(^ {13}\)In this specification, I am assuming that the entire response takes one period. I relax this assumption in the empirical application.\(^ {14}\)For now, I assume that $\varepsilon$ is the same for all individuals. I relax this assumption in Subsection 4.4.2.4.
not likely homogeneous across individuals. In Subsection 4.4.3, I will relax this assumption, but for now, assuming homogeneous trends will simplify notation and the analysis conducted prior to Subsection 4.4.3 is orthogonal to this issue.

Suppose that the transitory income component $\ln(\nu_{it})$ is serially correlated, and is generated by the following process:

\begin{equation}
\ln(\nu_{it}) = \sum_{k=1}^{K} \phi_k \ln(\nu_{it-k}) + \ln(\xi_{it}),
\end{equation}

where $K$ is the order of autocorrelation and $|\phi_k| < 1$ for all $k$. Note that I have assumed that serial correlation is the same for all individuals and all time periods. Let $\ln(\xi_{it}) \sim iid(0, \sigma^2_\xi)$ for all time periods. And, let $\ln(\nu_{it})$ be covariance stationary.

To characterize the determinants of mean reversion, first note that mean reversion has to do only with the transitory component of income. Since $E[\ln(\xi_t)] = 0$, individuals receive a mean zero shock each period. When $\phi_k = 0$ for all $k$, mean reversion at the individual level is very strong, because current income is no longer a function of transitory income in previous periods. Hence, even if individuals have very high or very low incomes relative to their permanent income level this period, on average, their incomes will return to their mean level in the following period. As $\phi_k \to 1$, mean reversion weakens. Therefore, when examining data with $0 \leq \phi_k < 1$ in a given year, if one looks at high income individuals, it will seem as though, on average, their income falls in the following year, and the reverse is true for low income individuals, even though individuals experience a mean zero shock every period. The actual volatility of transitory income $\sigma^2_\xi$ also affects the severity of mean reversion each period. If there was no income volatility in an economy (i.e. $\sigma^2_\xi = 0$), then there would be no mean reversion, because all transitory shocks would be equal to zero. As $\sigma^2_\xi$ increases, the magnitude of mean reversion also increases. As noted in Section 4.2, mean reversion is believed to be substantial in the U.S. context, and I will provide additional empirical evidence that this is, in fact, the case in Subsection 4.4.2.4.

---

15 I choose to write the income process in this simplified way, because, given the limited demographics in the tax return data sets commonly employed, separate causes of heterogeneous growth rates cannot be identified empirically. 
16 Allowing $\ln(\xi_{it})$ to have a nonzero mean is equivalent to changing permanent income. Any trends in $\ln(\xi_{it})$ are equivalent to trends in permanent income. Hence, without loss of generality, I can assume that $E[\ln(\xi_t)] = 0$. 
17 Covariance stationarity along with the assumption that $\ln(\xi_{it}) \sim iid(0, \sigma^2_\xi)$ implies: $E[\ln(\nu_t)] = E[\ln(\nu_{t-1})] = 0$ for all $t$, $\text{Var}[\ln(\nu_t)] = \text{Var}[\ln(\nu_{t-1})] = \sigma^2_\nu$ for all $t$. And $\text{Cov}[\ln(\nu_t), \ln(\nu_{t-1})] = \phi_s \sigma^2_\nu$ for all $s \neq t$ when $K = 1$. 
18 Thinking about mean reversion as being caused by a serially correlated error term or, more generally, by transitory income shocks is not unique to this paper. For example, see Kopeckz (2003, 2005), Moffitt and Wilhelm (2000), and Saez et al. (2012), among others. 
19 Throughout this paper, for any variable $w$, $w_{it}$ is the value this variable takes on for a single individual, and $w_t$ is the corresponding vector of all individuals at time $t$ for this variable. All statements made about this vector hold for all $t$. 

---

87
4.4 Data and Estimation

In this section, I theoretically derive conditions under which a consistent estimate of the ETI is identified. I also implement the results empirically, which provides an illustration of the bias induced if incorrect methods are used, and ultimately provides consistent estimates of the ETI under certain assumptions. When these assumptions do not hold, I estimate an alternative specification that is expected to provide a lower bound on the ETI under weaker assumptions.

This section proceeds as follows. Subsection 4.4.1 describes the data that will be used in the empirical analysis. Subsection 4.4.2 conducts baseline theoretical and empirical analysis. Subsection 4.4.3 conducts theoretical and empirical analysis allowing for heterogeneous growth rates in income at different ranges in the income distribution.

4.4.1 Data

This subsection describes the data that are used in the regression analysis Subsections 4.4.2 and 4.4.3. This section uses the Michigan IRS Tax Panel data set for the years 1979-1990. This is the only publicly available panel tax return data set in the U.S. Given that I will use instruments that are a function of lagged income, I restrict the estimation to the years 1982-1990.\(^{20}\) The major tax change that takes place during this period is TRA86. This was a substantial reform that changed both the tax rate and the tax base. It also substantially reduced the number of tax brackets in the U.S. tax system. It decreased tax rates for most individuals, particularly at the top of the income distribution and the reform was phased-in; that is, the tax rates were adjusted to their new level over a period of several years. For an extensive discussion of this data set and TRA86, see Gruber and Saez (2002).

The definition of taxable income used in the construction of the dependent variable and income splines in the regressions in Subsections 4.4.2 and 4.4.3 is defined in each year so that the tax base is constant across reforms. This is common in the literature; without this adjustment, the dependent variable—the log change in taxable income—would change mechanically as the definition of the tax base changes. To the extent that the tax base alters the tax rate faced by a taxpayer, not making this adjustment could substantially bias the estimates.\(^{21}\) It is widely recognized in the literature that this mitigates, but does not necessarily resolve the problem, because tax base changes often induce taxpayers to shift from one form of taxable income to another. Addressing this issue more

\(^{20}\)In Table 4.2, I do consider one specification that includes all the years, and show that the results are robust to this inclusion.

\(^{21}\)The estimates could also be biased if the tax base changes fall disproportionately on individuals in a particular income range.
completely is beyond the scope of this paper.\textsuperscript{22} As is common in this literature, I exclude capital gains entirely.\textsuperscript{23} Taxable income for each year is in 1992 dollars.

The tax rate variables used in the regressions include both state and federal tax rate changes. All tax rate variables are generated using TAXSIM.\textsuperscript{24} Weighting estimates by income is common, since the estimate desired for welfare analysis is the income-weighted average elasticity. However, for most of this section, I refrain from weighting the estimates because this data set does not oversample high-income individuals; therefore, income-weighted estimates place substantial weight on individuals whose response I can measure relatively imprecisely in this data. I do provide income-weighted estimates in Subsection 4.4.2.4; the estimates decrease slightly and there is a sizable increase in the standard errors. The only additional covariates included in the regressions in this section are marital status indicator variables.\textsuperscript{25} There are three marital status categories in total: single, married, and head of household/widowed with a dependent child.

Most individuals with constant-law taxable income greater than $10,000 in the base-year whose marital status did not change between the two years in the differences are included in the estimation.\textsuperscript{26} Using an income cutoff is common practice in the literature.\textsuperscript{27} Much of the justification for including an income cutoff—namely that there is too much mean reversion at the low end of the income distribution—will likely be resolved by the instruments ultimately used in this section. If the low end of the income distribution provides a poor comparison group, this would remain a reasonable justification for excluding them. Ultimately, I will control for heterogeneous income trends in this section using income splines. With this control, individuals at the low end of the income distribution would make a poor comparison group if individuals’ income grows at a non-constant rate over time and this trend changes differentially for those at the low end of the income distribution. Even if individuals at the lower end were a reasonable comparison group in theory, there often remains an

\begin{itemize}
\item \textsuperscript{22}Kopczuk (2005) tried to address this issue more directly by controlling for changes in the tax base directly in the estimating equation.
\item \textsuperscript{23}In general, my income measures are defined as in Gruber and Saez (2002), but a few improvements are made to make the income definitions more consistent across years. These changes have very minor effects on Gruber and Saez’s original estimates. More generally, despite the large tax base changes in TRA86, adjusting taxable income definitions to make them consistent across years does not have much affect on the estimates.
\item \textsuperscript{24}An overview of TAXSIM can be found in Feenberg and Coutts (1993). I use the full version of TAXSIM, which is available exclusively on the NBER server.
\item \textsuperscript{25}This is common in U.S. studies, because the publicly available tax return data lacks additional covariates. For example, see Gruber and Saez (2002). I discuss the extent to which this lack of covariates may or may not affect the estimates later in the paper.
\item \textsuperscript{26}Less than twenty individuals in each specification, whose actual log tax change plus their predicted log tax change is greater than one, are excluded. The absolute value of $\Delta \ln(Y_{it})$ is censored at 7. These restrictions are the same as those used in Gruber and Saez (2002) and they have very minor effects on the estimates.
\item \textsuperscript{27}The first paper to exclude low-income individuals was Auten and Carroll (1999). They excluded everyone whose taxable income fell below the 22 percent marginal tax rate bracket in 1985, which corresponded to $21,020 in 1985 dollars. Gruber and Saez (2002) exclude everyone with broad income under $10,000. Most subsequent papers that use an income cutoff follow the Gruber and Saez selection rule. Note that excluding a certain portion of the population can improve the mean reversion problem, but it only resolves it completely in the extreme case that everyone left in the sample experiences the same transitory income shock.
\end{itemize}
important reason to exclude those at the very bottom of the income distribution in practice. The use of constant-law taxable income in the dependent variable usually creates a particular type of sample selection. Near zero, constant-law taxable income can take on negative values, which are turned to missing when this variable is converted to logs. This excludes all individuals who earned negative constant-law taxable income in one year of the difference. Since income splines are also a function of constant-law taxable income, this excludes all individuals with negative constant-law taxable income in the base year. This means that just above zero in base-year income, individuals are only included in the sample if they received a positive shock (or a very small negative shock) between periods, which is more likely to generate a substantial bias in the estimates than a clean income cutoff. The proposed $10,000 cutoff is not endogenous as long as the instruments used are not significantly correlated with transitory income shocks in base-year income. I will discuss whether or not this condition holds for the instruments chosen in Subsection 4.4.2. Descriptive statistics for the preferred baseline estimates are provided in Table 4.1.

4.4.2 Instrument Selection

Before using the data described in the last subsection to estimate the ETI empirically, this subsection theoretically examines which instruments will be exogenous using the model of income set up in Section 3. To do this, I first rewrite equation (4.2) in an estimable form, which yields:

\[ \Delta \ln(Y_{it}) = \varepsilon \Delta \ln(1 - \tau_{it}) + \alpha_{t-1} + \eta_{it}, \]

where \( \alpha_{t-1} \) are year fixed effects and \( \eta_{it} = \Delta \ln(\nu_{it}) \). The year fixed effects control for any omitted variables in differences that are the same, on average, for all individuals at a given time \( t \), including the homogeneous growth rate \( g_t \). Of course, if all individuals do not share the same income trend \( g_t \), the year fixed effects are no longer enough to produce consistent estimates. I will address this particular case in Subsection 4.4.3. I assume that \( 1 - \tau_{it} = 1 - \tau_{it} \) and address what happens when this is not the case at the end of Subsection 4.4.2.4.

Additionally, assume, as is the case for TRA86 and most other tax reforms, that the tax rate schedule is graduated and the regression includes individuals at all income levels. Then, higher values of \( \Delta \ln(\nu_{it}) \) lead to higher values of \( \Delta \ln(Y_{it}) \) (i.e. \( E[\Delta \ln(Y_{it})' \Delta \ln(\nu_{it})] > 0 \)), all else equal, which in turn lead to lower values of \( \Delta \ln(1 - \tau_{it}) \). Hence, as has been widely recognized in the literature, \( \Delta \ln(1 - \tau_{it}) \) is endogenous. Subsection 4.4.2.1 uses the income process laid out in Section 4.3 to theoretically examine whether this endogeneity can be addressed using the predicted net-of-tax rate
as an instrument and additional income-based control variables. Subsection 4.4.2.2 uses the same method to examine whether there are alternative instruments which are exogenous under certain assumptions. Subsection 4.4.2.3 proposes a test, which will allow me to test empirically whether each instrument discussed in the previous two subsections is exogenous. Subsection 4.4.2.4 conducts this instrument exogeneity test for each of the proposed instrument and control combinations discussed in the previous two subsections and provides baseline ETI estimates.

4.4.2.1 Using the Predicted Net-of-Tax Rate as an Instrument

How to address the endogeneity of the net-of-tax rate term is very important, and it has duly received a large discussion in the literature. By far the most frequently used instrument for $1 - \tau_{it}$ is the value for $1 - \tau_{it}$ if an individual’s income was $Y_{it-1}$ in year $t$ and the tax code was that of year $t$, that is, the predicted net-of-tax rate based on income in year $t - 1$. In this subsection, I will focus exclusively on the ability of this instrument to solve the endogeneity problem. I will refer to this instrument as $1 - \tau_{it}^p$. I discuss alternative instruments in Subsection 4.4.2.2.

There are two conditions relevant for assessing instrument validity: that the instrument is not weak, i.e. that

\[(4.6) \quad |\text{cov} [\Delta \ln(1 - \tau_t), \Delta \ln(1 - \tau_{it}^p)|X_t]|\]

is large,\(^{28}\) and that the instrument is exogenous:

\[(4.7) \quad \text{cov}[\Delta \ln(1 - \tau_{it}^p), \eta_t] = \mathbb{E}[\Delta \ln(1 - \tau_{it}^p)' \eta_t|X_t] = 0,\]

where $X_t$ are any other covariates included in the regression. When these do not hold, the asymptotic bias can be given by the following equation (when there are no time-varying conditioning variables):

\[(4.8) \quad \text{plim}(\hat{\varepsilon}_{IV}) = \varepsilon + \frac{\text{cov}[\Delta \ln(1 - \tau_{it}^p), \eta_t|X_t]}{\text{cov}[\Delta \ln(1 - \tau_t), \Delta \ln(1 - \tau_{it}^p)|X_t]},\]

and the estimates will be inconsistent.\(^{29}\)

The predicted net-of-tax rate instrument likely does not suffer from the weak instrument problem.

---

\(^{28}\)Whether or not this condition holds is determined empirically using an F-test.

\(^{29}\)Technically, to obtain consistent IV estimates, there are two additional conditions that must be satisfied, namely that the instruments are linearly independent and that there are as at least as many instruments as there are endogenous variables. Neither of these conditions are ever violated here, and so it will be assumed that these conditions hold throughout.
in practice.\textsuperscript{30} But, it is not clear that the instrument exogeneity condition is satisfied. In particular, to obtain identification in equation (4.5), variation in the tax rate change across individuals within a given year is necessary. For now, assume this variation is due to larger tax rate changes for higher income levels and smaller tax rate changes for lower income levels.\textsuperscript{31} In this case, higher levels of \( Y_{it-1} \) will generate higher values of \( \Delta \ln(1 - \tau_t^{p}) \) in the case of a tax decrease and lower values in the case of a tax increase. Therefore, \( \text{cov}[\Delta \ln(1 - \tau_t^{p}), \ln(\nu_{i-1})] > 0 \Rightarrow \text{cov}[\Delta \ln(1 - \tau_t^{p}), \eta_t] < 0 \) in the case of a tax decrease, and \( \text{cov}[\Delta \ln(1 - \tau_t^{p}), \ln(\nu_{i-1})] < 0 \Rightarrow \text{cov}[\Delta \ln(1 - \tau_t^{p}), \eta_t] > 0 \) in the case of a tax increase. Since the denominator of the ratio in equation (4.8) is always positive, this implies that, in the absence of additional controls, the IV estimate will be biased downwards in the case of a tax decrease and biased upwards in the case of a tax increase. Therefore, \( \mathbb{E}[\Delta \ln(1 - \tau_t^{p})'\eta_t] = 0 \) only if \( \mathbb{E}[\ln(Y_{i-1})'\eta_t] = 0 \).

The severity of the endogeneity problem when \( \mathbb{E}[\ln(Y_{i-1})'\eta_t] \neq 0 \) is clearly a function of both the degree of serial correlation and volatility of transitory income. For notational ease, I am going to assume that the true ETI is zero in the analysis that follows.\textsuperscript{32} When there is no serial correlation, that is, \( K = 0 \), \textsuperscript{33}

\[
\mathbb{E}[\ln(Y_{i-1})'\eta_t] = \mathbb{E}[\ln(Y_{i-1})' (\ln(\nu_i) - \ln(\nu_{i-1}))]
\]

\[
= \mathbb{E}[\ln(Y_{i-1})' (\ln(\xi_i) - \ln(\xi_{i-1}))]
\]

\[
= -\mathbb{E}[\ln(Y_{i-1})' \ln(\xi_{i-1})]
\]

\[
= -\sigma^2_{\xi} < 0.
\]

The last line of (4.9) relies on the assumption that \( \ln(\xi_{i-1}) \) is i.i.d. If the transitory income shocks are not actually independent of permanent income, the form will be slightly different, because there will be an additional piece, \( \text{cov}(\ln(\mu_{i-1}), \ln(\xi_{i-1})) \). But, the expression is guaranteed to remain

\textsuperscript{30}For example, the F-statistics for the first-stage results estimated by Gruber and Saez (2002) are all between 20 and 100.

\textsuperscript{31}This is by far the most common form of identification because tax reforms typically take this form. I will address how this discussion is altered if a different form of identification is used in a paragraph later in this section.

\textsuperscript{32}When the ETI is positive, there will be an additional term in all the covariance derivations below, which accounts for the fact that, under a progressive tax schedule, individuals with higher transitory income shocks will face higher marginal tax rates, on average, and will respond to this by lowering their taxable income levels. This will mitigate the results below slightly, but do not change the overall conclusions.

\textsuperscript{33}This derivation is for one-year differences. For three-year differences, the covariance is equivalent, assuming the covariance between transitory shocks and income stays constant over time, because this covariance is given by:

\[
\mathbb{E}[\ln(Y_{i-3})'\eta_t] = \mathbb{E}[\ln(Y_{i-3})' (\ln(\nu_i) - \ln(\nu_{i-3}))]
\]

\[
= \mathbb{E}[\ln(Y_{i-3})' (\ln(\xi_i) - \ln(\xi_{i-3}))]
\]

\[
= -\mathbb{E}[\ln(Y_{i-3})' \ln(\xi_{i-3})]
\]

\[
= -\sigma^2_{\xi} < 0.
\]
negative unless $\text{cov}(\ln(\mu_{t-1}), \ln(\xi_{t-1}))$ is negative, so that high values of $\ln(\xi_{t-1})$ are offset by low values of $\ln(\mu_{t-1})$. It is highly unlikely that this is the case.

Now, suppose there is first-order serial correlation, that is $K = 1$. Then,

$$
E[\ln(Y_{t-1})' \eta_t] = E[\ln(Y_{t-1})' (\ln(\nu_t) - \ln(\nu_{t-1}))]
= E[\ln(Y_{t-1})' (\phi_1 \ln(\nu_{t-1}) + \ln(\xi_t) - \ln(\nu_{t-1}))]
= E[\ln(Y_{t-1})' (\phi_1 - 1)\ln(\nu_{t-1})]
= (\phi_1 - 1) \sigma^2_v
= \frac{(\phi_1 - 1) \sigma^2_v}{1 - \phi_1^2} = -\frac{\sigma^2_\xi}{1 + \phi_1} < 0.
$$

The last line of (4.10) takes advantage of the fact that $\sigma^2_v = \frac{\sigma^2_\nu}{1 - \phi_1^2}$ when the process is AR(1) and covariance stationary. Compared to the case when there is no serial correlation as given by (4.9), (4.10) is larger in absolute value. Also, note that (4.10) would equal zero if $\phi_1 = 1$, but this would imply that income follows a unit-root process, which would generate an alternative set of issues to be addressed. Unless the error term follows a unit root process, $\ln(Y_{it-1})$ is correlated with the error term. The general notion that this instrument remains endogenous has been well-acknowledged in the literature.

Of course, if there was a tax reform for which the tax change was the same for all income levels, the endogeneity problem discussed in the previous paragraphs would be eliminated, but it would also eliminate identification because everyone would experience the same treatment, and there would be no variation to exploit in order to estimate the elasticity. Alternatively, if there was a tax reform that affected some individuals, but not others within a given income class, the endogeneity problem would be mitigated. Different tax changes at different points in the income distribution in different years has been suggested as an alternative way to resolve the endogeneity problem. If there are no

\[34\text{Again, this derivation is for one-year differences. For three-year differences, the covariance is strictly larger assuming the covariance between transitory shocks and income stays constant over time because this covariance is given by:}
E[\ln(Y_{t-3})' \eta_t] = E[\ln(Y_{t-3})' (\ln(\nu_t) - \ln(\nu_{t-3}))]
= E[\ln(Y_{t-3})' (\phi_3^2 \ln(\nu_{t-3}) + \phi_3^2 \ln(\xi_{t-2}) + \phi_1 \ln(\xi_{t-1}) + \ln(\xi_t) - \ln(\nu_{t-3}))]
= E[\ln(Y_{t-3})' (\phi_3^2 - 1)\ln(\nu_{t-3})]
= (\phi_3^2 - 1) \sigma^2_v
= \frac{(\phi_3^2 - 1) \sigma^2_v}{1 - \phi_3^2} < 0,
\]

and $|\phi_3^2 - 1| > |\phi_1 - 1|$.

\[35\text{A non-zero covariance is also found when higher orders of serial correlation are considered.}
\]

\[36\text{For example, Long (1999) just uses state tax rate variation and most ETI studies in the U.S. combine federal and state tax rate variation.}
\]
year effects, then this will indeed help mitigate the problem, but if year fixed-effects are included in the regression, such variation is absorbed in the year fixed-effects and does not aid in identifying the ETI.

Given the remaining endogeneity of the instrument, researchers have tried to solve the problem by including controls for $\ln(Y_{it-1})$.

As Saez (2003, p.1250) observes, “...if $\epsilon[\eta_{it}]$ depends on $z_1[Y_{it-1}]$, the instrument, which is a function of $taxinc_1[Y_{it-1}]$, is likely to be correlated with the error term $\epsilon[\eta_{it}]$. However by controlling for any smooth function of $taxinc_1[Y_{it-1}]$ in the regression setup in both stages, it is possible to get rid of the correlation between $\epsilon[\eta_{it}]$ and the instruments.” Saez is correct that, conditional on a given value of base-year income (and any other controls included in the model such as marital status), $\ln(1-\tau^p_t)$ is some constant value; so from that perspective, the endogeneity problem is solved—conditional on $\ln(Y_{t-1})$, $\Delta \ln(1-\tau^p_t)$ no longer covaries with $\eta_t$.

But, another problem arises. The value of $\ln(Y_{t-1})$ itself is a valid control only if it does not covary with $\eta_t$, that is $E[\ln(Y_{t-1})'\eta_t] = 0$, which, as I showed above, holds only in the case of a unit-root. An alternative way of thinking about the issue is that if $\ln(Y_{t-1})$ was a valid proxy for the components of the error term that are correlated with $\ln(Y_{t-1})$, then it would be fine as a control. But this is never the case. For example, when $K=1$, one would like a perfect proxy for $\ln(\nu_{t-1})$. But, if $\ln(Y_{t-1})$ is employed as this proxy, $\ln(\mu_{t-1})$ will be contained in the error term because:

$$ln(Y_{it-1}) = ln(\nu_{it-1}) + ln(\mu_{it-1}).$$

And, $cov(ln(Y_{t-1}), ln(\mu_{t-1})) = \sigma^2_\mu > 0$. Hence, $ln(Y_{t-1})$ will produce a biased estimate of $ln(\nu_{t-1})$ and therefore remain endogenous. This control is thus valid only when the original instrument is valid. But, in that case, it is not needed (at least not to solve an endogeneity problem). I will return to a discussion of whether it is relevant for heterogeneous income trends in Subsection 4.4.3).

Kopczuk (2003, 2005) suggested the following alternative control:

$$\Delta ln(Y_{it-1}) = ln(\nu_{it-1}) + g_{t-1} - ln(\nu_{it-2}),$$

which will remain endogenous because the latter two terms will be relegated to the error term and $cov(\Delta ln(Y_{it-1}), -ln(\nu_{it-2})) = (1-\phi_1)\sigma^2_\nu > 0$ when $K = 1$. This covariance weakens as $\phi_1 \to 1$.

---

37 The first paper to include this control is Auten and Carroll (1999).
38 The only exception to this statement would be if high levels of permanent income were correlated with low shocks, but as I noted before, this is not likely the case.
39 A similar result holds for all $K \neq 1$.
40 A similar result holds for all $K \neq 1$.
41 Kopczuk (2003, 2005) suggests this proxy could be improved by including it as a 10-piece spline instead. But,
As a last alternative, consider
\[
\Delta \ln(Y_{it-1}) - \Delta \ln(Y_{it-2}) = (\ln(\nu_{it-1}) + g_{t-1} - \ln(\nu_{it-2})) - (\ln(\nu_{it-2}) + g_{t-2} - \ln(\nu_{it-3})) \\
= \ln(\nu_{it-1}) - \ln(\nu_{it-3}),
\]
where the last line follows if $g_{t-1} = g_{t-2}$. This control solves both the issues raised with $\Delta \ln(Y_{it-1})$, but generates a new, similar problem, namely $\text{cov}(\Delta \ln(Y_{it-1}) - \Delta \ln(Y_{it-2}), -\ln(\nu_{it-3})) = (1 - \phi_1^2)\sigma_\nu^2 > 0$. Hence, this control is not valid, either. Therefore, there are no income-based controls that have been proposed that are expected to make this instrument exogenous. Note, also, that some demographics are always used, and occasionally more extensive demographic covariates are used (e.g. Carroll, 1998; Auten and Carroll, 1999; Singleton, 2011). However, in general these covariates are variables such as occupation and education level, which likely proxy for permanent, not transitory income. As a result, these are also not expected to resolve the endogeneity problem.

### 4.4.2.2 Alternative Instruments

Subsection 4.4.2.1 demonstrates that $\Delta \ln(1 - \tau_{it}^p)$ is not exogenous as an instrument, regardless of the additional controls used. There are two possible instrument types that I will consider in this subsection. First, I consider a potential instrument that has not been previously considered in the literature. In particular, suppose that instead of making the predicted tax rate instrument a function of $\ln(Y_{it-1})$, it was instead a function of some lag of $\ln(Y_{it-1})$. If the relevant instrument exogeneity condition holds, then this will indeed resolve the endogeneity problem brought about by mean reversion. This approach is standard for resolving endogeneity problems in the dynamic panel literature, which always puts lags of the left-hand side variable on the right-hand side.

To understand when this approach would resolve the endogeneity problem, suppose that $\ln(Y_{it-2})$ is used to instrument for the tax rate. Now, the relevant instrument exogeneity condition is
\[ E[\ln(Y_{t-2})'] \eta_t] = 0. \] Suppose that \( K = 0 \). Then, this condition can be rewritten as:

\[
(4.14) \quad E[\ln(Y_{t-2})' \eta_t] = E[\ln(Y_{t-2})' (\ln(\xi_t) - \ln(\xi_{t-1}))] \\
= E[\ln(Y_{t-2})' \ln(\xi_t)] - E[\ln(Y_{t-2})' \ln(\xi_{t-1})] = 0.
\]

So, this instrument would clearly be valid when \( K = 0 \). Now consider the case where \( K = 1 \):

\[
(4.15) \quad E[\ln(Y_{t-2})' \eta_t] = E[\ln(Y_{t-2})' (\ln(\nu_t) - \ln(\nu_{t-1}))] \\
= E[\ln(Y_{t-2})' (\ln(\xi_t) + \phi_1 \ln(\nu_{t-1}) - \ln(\nu_{t-1}))] \\
= E[\ln(Y_{t-2})' (\ln(\xi_t) + (\phi_1 - 1) \ln(\xi_{t-1}) + \phi_1 (\phi_1 - 1) \ln(\nu_{t-2}))] \\
= \phi_1 (\phi_1 - 1) E[\ln(Y_{t-2})' \ln(\nu_{t-2})], \\
= \phi_1 (\phi_1 - 1) \sigma^2_{\nu},
\]

which does not equal zero when there is no unit-root. However, since \( E[\ln(Y_{t-1}) \ln(\nu_{t-1})] = \sigma^2_{\nu} \) is the same for all \( l \), the covariance is strictly less than the covariance when the instrument was based on \( \ln(Y_{t-1}) \) because \(|(\phi_1 - 1)\sigma^2_{\nu} > |\phi_1 (\phi_1 - 1)\sigma^2_{\nu}|\). Hence, the endogeneity problem is strictly better than it was before. If the error process is truly serially correlated, the recursive structure of the error term will cause conditions like that found in (4.15) to always be violated, regardless of the number of lags chosen. However, the value of the covariance will get arbitrarily small as the number of lags increase. Alternatively, one could rephrase this statement in terms of a testable hypothesis. If enough lags are used, eventually the null hypothesis that \( E[\ln(Y_{t-1})' \ln(\eta_t)] = 0 \) will not be rejected. It is also possible that the true underlying process is not serially correlated, but rather a moving-average process. In this case, the same basic idea holds, but the recursive structure is gone. For example, if the error process is \( MA(1) \), \( E[\ln(Y_{t-1})' \ln(\eta_t)] \) will equal zero exactly. In practice, it is usually not possible to distinguish between serial correlation that dies out quickly and a moving average process.\(^{47}\) Hence, in Subsection 4.4.2.3, I will consider a test that will tell us whether or not I can reject the hypothesis that \( E[\ln(Y_{t-1})' \eta_t] = 0 \), and I will abstract away from whether the true underlying process is moving-average or serially correlated.

There is also another possible type of instrument that has already been proposed in the literature that has been proven to be valid only for even differences (e.g. it is valid for two-year differences, but not for one- or three-year differences). Blomquist and Selin (2010) use Swedish data to estimate a single decade-long difference from 1981 to 1991, and use 1986 income, that is income in the middle

---

\(^{47}\)MaCurdy (1982), in his study on the properties of the error component of earnings, makes this observation. In his study he finds that he cannot distinguish between an earnings process that is \( AR(1) \) or \( MA(2) \).
year, as an instrument. \footnote{Auten and Carroll (1999) and others have raised concerns about instruments being constructed as a function of post-response income. This only matters to the extent that individuals move across tax brackets in response to the tax change. If cross-bracket movement is due to transitory income shocks only, then it will be resolved as long as the instrument is uncorrelated with these transitory shocks. A more fundamental problem occurs if somehow the behavioral model underlying individuals’ responses suggests that there are cases when their response will cause them to move from one tax bracket to the next. However, in this case it can be shown that neither pre- nor post-response income instruments are guaranteed to resolve the problem. A more formal discussion of this issue is left for future work.}

I repeat their proof here, using a two-year difference with first-order autocorrelation as an example. Suppose, I constructed the tax rate instrument as a function of $\ln(Y_{t-1})$. Now, the relevant instrument exogeneity condition is $E[\ln(Y_{t-1})'(\ln(\nu_t) - \ln(\nu_{t-2}))] = 0$, which holds exactly:

$$E[\ln(Y_{t-1})'(\ln(\nu_t) - \ln(\nu_{t-2}))] = E[\ln(\nu_{t-1})'(\ln(\nu_t) - \ln(\nu_{t-2}))]$$

$$= E[\ln(\nu_{t-1})'(\phi_1\ln(\nu_{t-1}) + \ln(\xi_t))]$$

$$- E[(\phi_1\ln(\nu_{t-2}) + \ln(\xi_{t-1}))'\ln(\nu_{t-2})]$$

$$= \phi_1\sigma^2_\nu - \phi_1\sigma^2_\nu = 0.\footnote{In general, zero covariance does not imply independence, which, in turn, implies that this instrument may not be valid if the error terms are not distributed normally, at least for short differences. Also, the same results hold if a higher order serially correlated process or an MA process is assumed instead.}$$

Blomquist and Selin (2010) also assume that $\ln(\nu_t)$ is distributed normally, which allows them to conclude that $\ln(Y_{t-1})$ and $(\ln(\nu_t) - \ln(\nu_{t-2}))$ are independent. This in turn implies that any function of $\ln(Y_{t-1})$ is independent of $(\ln(\nu_t) - \ln(\nu_{t-2}))$, including the tax rate instrument. \footnote{If the instrument used is two periods after the base-year, it can be shown that the bias is the same in magnitude, but opposite in sign.}

Therefore, under these assumptions, constructing $\Delta \ln(1 - r^p_{it})$ as a function of $\ln(Y_{it-1})$ will be a valid instrument. Now, consider a three-year difference. There is no longer a period that lies directly in the middle of the difference. Suppose an instrument one period after the base-year is being considered as a potential instrument:\footnote{If the instrument used is two periods after the base-year, it can be shown that the bias is the same in magnitude, but opposite in sign.}

$$E[\ln(Y_{t-1})'(\ln(\nu_t) - \ln(\nu_{t-3}))] = E[\ln(\nu_{t-1})'(\ln(\nu_t) - \ln(\nu_{t-3}))]$$

$$= E[\ln(\nu_{t-1})'(\phi_1\ln(\nu_{t-1}) + \ln(\xi_t))]$$

$$- E[(\phi_1^2\ln(\nu_{t-3}) + \phi_1\ln(\xi_{t-2}) + \ln(\xi_{t-1}))'\ln(\nu_{t-3})]$$

$$= \phi_1\sigma^2_\nu - \phi_1^2\sigma^2_\nu > 0.$$
approximately exogenous in practice.

4.4.2.3 Testing Instrument Validity

In this subsection, I consider tests of instrument validity that will determine which, if any, lags of \( \ln(Y_{it-1}) \) are valid for constructing the predicted net-of-tax rate instrument discussed in Subsection 4.4.2.2.\(^{51}\) If serial correlation did not cause the estimator to be inconsistent, then I could estimate the model, obtain a consistent estimate of \( \hat{\eta}_{it} \), and use this estimate to estimate the serial correlation properties of \( \eta_{it} \). But, this is not the case. The dynamic panel literature faces a very similar problem and employs a variety of alternative tests to assess whether or not the instruments used are valid. One common method is tests based on over-identifying restrictions (Arellano and Bond, 1991). One such test is the Sargan test.\(^{52}\) The test statistic is given by:

\[
(4.18) \quad s = \hat{\eta}'Z \left( \sum_{i=1}^{N} Z_i'\hat{\eta}_i\hat{\eta}_i'Z_i \right)^{-1} Z_i'\hat{\eta}_i \sim \chi^2_p,
\]

where \( Z \) is a matrix of all instruments used stacked over all time periods, and \( p \) is the number of instruments. Note that the middle piece of this sandwich estimator is an estimate of the heteroskedasticity-robust covariance matrix. And, recall from the previous subsection that the condition which must hold in order for the instruments to be valid is given by \( \mathbb{E}[Z_i'\eta_i] = 0 \) for all \( t \).

Therefore, when this statistic is close to zero, the instruments are valid (null hypothesis), and when it is far away, they are not (alternative hypothesis). How effective this method is at detecting all instruments that are, in fact, endogenous depends on the power of the test; that is, how often the test fails to reject the null hypothesis when it is in fact false. Arellano and Bond (1991) provide some suggestive evidence about this issue in a dynamic panel setting.\(^{53}\) In their setup, the null hypothesis is rejected 78 percent of the time at the 10 percent level when it is in fact false if the underlying serial correlation is 0.3, and only 47 percent of the time when the underlying correlation

---

\(^{51}\) These tests will introduce a pretest bias (Guggenberger and Kumar, 2011). However, it is mitigated by the fact that I will use these tests in the next subsections as evidence of which instruments are valid, but use my originally hypothesized choice of lags–two, three, and four—even though the null hypothesis is marginally not rejected for one lag. In general, researchers should be cautious about agnostically implementing this method without considering the size distortions this may induce. Also, the inference reported in the next subsections is robust to using inference based on the Anderson-Rubin test, which Guggenberger and Kumar (2011) show is less subject to size distortions from non-exogeneity, which would arise if the over-identification test failed to reject an instrument that is, in fact, endogenous.

\(^{52}\) This test also goes by the name J-test, which might be more familiar to readers.

\(^{53}\) Their setting is not quite the same as the one considered in this paper, so these results are only suggestive. Running a simulation study of these tests in this particular context is an area for future research.
is 0.2. They find that a similar test, the Difference-in Sargan test, has higher power:

\[
(4.19) \quad ds = \hat{\eta}' Z \left( \sum_{i=1}^{N} Z_i \hat{\eta}_i \hat{\eta}_i' Z_i \right)^{-1} Z' \hat{\eta} - \hat{\eta}' \right) Z \left( \sum_{i}^{N} Z_i \hat{\eta}_i \hat{\eta}_i' Z_i \right)^{-1} Z' \hat{\eta} \sim \chi^2_{p-I},
\]

where \( Z_I \) is a matrix of all instruments believed to be valid under the null and alternative hypothesis and \( p_I \) is the number of these instruments. The null hypothesis of this test is that all instruments are valid, and the alternative hypothesis is that instruments that are being tested (those that are valid under null and not valid under the alternative hypothesis) are not valid. Note that the second component should be small. If it is large, the entire instrument set should be rejected based on the Sargan test. Using the Difference-in-Sargan test in the Arellano-Bond setup, the null hypothesis is rejected 91 percent of the time at the 10 percent level when it is in fact false if the underlying serial correlation is 0.3, and 60 percent of the time when the underlying correlation is 0.2.

Therefore, in the estimation section, I will determine the number of valid lags based on the Difference-in-Sargan test, because it has higher power. Given that the power of these tests is not perfect for all levels of serial correlation, this test may fail to reject some instruments that are in fact not valid. I have shown that the bias must decrease as the lag used increases when the error term is serially correlated. Therefore, at the very least, the estimator is substantially closer to obtaining consistency than any one previously used.\(^{54}\)

### 4.4.2.4 Empirical Results

Subsections 4.4.2.1 and 4.4.2.2 have examined the choice of instrument theoretically. Subsection 4.4.2.3 provided a way in which the degree of endogeneity can be quantified empirically, which is used in this subsection to provide a quantitative analog to Subsections 4.4.2.1 and 4.4.2.2, highlighting the biases induced by incorrect methods, and ultimately obtaining a consistent estimate of the ETI. This subsection concludes by interpreting this parameter in light of Weber (2012b), and considers an alternative definition of treatment which provides a substantial decrease in standard errors.

Table 4.2 provides empirical estimates for each of the proposed instrument and control combinations considered above. This table estimates two-year differences so that all the different possible instruments can be considered using the same difference length.\(^{55}\) The estimating equation is given by (4.5) (with the addition of marital status indicators) using instruments and sometimes additional

---

\(^{54}\)The size of these tests is very good according to Arellano and Bond (1991); that is, the tests do not incorrectly reject the null hypothesis when it is true more than they should. Therefore, in interpreting these results, one should not worry more than usual that a consistent estimator has been rejected.\(^{55}\)Recall that the instrument proposed by Blomquist and Selin (2010) is only applicable for even differences.
income controls as proposed in the previous subsections. Note that for each of the income controls, I use a 5-piece spline, rather than just including the income control directly. This is standard in the literature, since it makes the control more flexible. This equation can be interpreted as a continuous treatment difference-in-differences equation. In Columns (1)-(4), I assume that a predicted tax rate instrument constructed from income lagged two and three periods prior the base-year are exogenous. This means that both of these instruments are included in the instrument set in each column and are used to test the exogeneity of another, potentially endogenous, instrument using the Difference-in-Sargan test. Column (5) will test whether income lagged two periods is, in fact, exogenous. Columns (1)-(6) are restricted so that the same individuals appear in each to enhance comparability across the columns.

Table 4.2 Columns (1)-(3) consider specifications where the instrument is shown to be endogenous in Subsection 4.4.2.1, except under extreme assumptions. Column (1) uses the predicted tax rate instrument $Δln(1 − τ^p_t)$. The estimate is 0.144 and the p-value from the Difference-in-Sargan test is 0.006; therefore, I can strongly reject instrument exogeneity, as predicted by the theoretical results. Column (2) adds splines in log base-year income $ln(Y_{t−2})$. While the estimate changes substantially—it more than doubles to 0.453 and becomes statistically significant at the one percent level—the p-value of the Difference-in-Sargan test is even smaller (0.000). Column (3) again repeats Column (1), but adds the lagged value of the dependent variable as a spline $Δln(Y_{t−1})$. The estimate is 0.171 and is marginally insignificant. The null hypothesis that the instrument is exogenous is rejected at the 10 percent level. The results in Columns (1)-(3) are strongly consistent with the theoretical analysis, and highlight that while the literature has believed that these splines can do a lot to resolve the endogeneity of the instrument created by mean reversion, it is simply not true in this context. The results also highlight that mean reversion is a substantial issue in the U.S., because it would otherwise be difficult to reject the null hypothesis that these instruments are not exogenous.

Table 4.2 Columns (4)-(6) consider instruments that are exogenous under certain, more reasonable assumptions that were proposed in Subsection 4.4.2.2. Column (4) still assumes that two and...
three lags are exogenous and tests whether one lag is also exogenous. The p-value is 0.163, and therefore can barely not be rejected at the 10 percent level. Column (5) assumes that an instrument lagged three and four periods is exogenous and tests whether can reject the null that two lags are exogenous. The p-value on this test is 0.490. The ETI estimate in Column (5) is 1.046 and is statistically significant at the 1 percent level. The estimates in Column (5) are my preferred baseline estimates. Looking at Columns (1), (4), and (5), the instruments become more exogenous as the lags of income used to construct the instruments increase, and the estimates increase as the instruments move towards being exogenous. This is consistent with the theoretical analysis in Subsection 4.4.2.1, which showed that the estimates would be biased downwards when the endogenous instruments were used (for a tax decrease). The estimate in Column (5) is more than twice as large as the Column (2) estimates, where the latter is a commonly used method in the literature, originally proposed in Gruber and Saez (2002).

One disadvantage of using lagged income in the construction of the instrument is that the first years of the data set are used exclusively for constructing lags; it is for this reason, that only the years 1983-1990 are used in my preferred baseline estimate. Furthermore, using lags increases standard errors and decreases F-statistics. Regarding the latter concern, F-statistics do decline, but they are still far from a critical level, since they all remain over 100. The standard errors do increase, but they do this even when all columns are restricted to the years 1983-1990. This suggests much of the increase is reflective of the fact that the original treatment measure is quite endogenous. Later in this section, I will consider an alternative definition of treatment that reduces the standard errors.

Table 4.2 Column (6) considers the instrument proposed by Blomquist and Selin (2010). This estimate is 1.145 and statistically significant at the 1 percent level. The estimates in Columns (5) and (6) are similar, as would be expected if the instruments in both columns are exogenous. Recall, in the data subsection, I noted that all individuals with base-year income below $10,000 were excluded from the estimation. It is reasonable theoretically to assume that the instruments employed are not correlated with base-year income, and if they were, I would be able to reject the null hypothesis that the instruments are exogenous.

---

61I prefer these estimates over Column (4) given that I barely fail to reject the null hypothesis in Column (4) and the power of this test may be slightly weak as discussed in Subsection 4.4.2.3.
62Note that I could drop the fourth lag instrument after the tests are over; however, the use of the fourth lag decreases standard errors, even though an additional year of data can no longer be used in the second-stage estimating equation.
63I do not include any additional instruments in this specification to test for instrument exogeneity, because the estimates are assumed exogenous. However, if I do include instruments lagged three and four periods the p-value on the Difference-in-Sargan test is 0.401.
64Note that imposing a cutoff as a function of lagged income is not necessarily better or worse when lags of income are used as instruments.
Table 4.2 Column (7) estimates the same specification as Column (6), but includes all years, 1979-1990; it is the one ETI estimate in this paper that exploits variation from both ERTA81 and TRA86. ERTA81 also provided marginal tax rate decreases to most taxpayers. Although the reforms share a lot in common, the fact that the estimate remains stable when ERTA81 is included, provides some evidence that this method is robust. It also suggests that the other estimates are losing out on an opportunity to decrease standard errors, but otherwise little is lost by restricting the data set to 1983-1990.

Table 4.2 Columns (8) and (9) repeat Column (5) for one- and three-year differences, respectively. The estimates for one- two- and three-year differences are similar overall.\textsuperscript{65} The literature has interpreted this similarity as evidence that the short-run and long-run responses are quite similar (e.g. Gruber and Saez, 2002). This is certainly part of the story; however, accepting this explanation as the whole story potentially overlooks the fact that none of these estimates are identifying the parameter they are ostensibly measuring. The fundamental problem lies in the nature of overlapping differences. To see the problem, suppose (as was the case for TRA86) that the tax rate changes take place in two years (i.e. it is phased-in)—1987 and 1988—and it takes three years for individuals to respond fully to the tax reform. Consider three-year differences. The difference that spans 1986-1989 captures the long-run response to the tax changes that took place in 1986, and two years of the response to the tax changes that took place in 1987. In contrast, the 1984-1987 difference only captures a one-year response to the tax changes implemented in 1987. Therefore, the estimate is a combination of short-run, medium-run, and long-run responses. One-year differences suffer from a variant of the same problem. The 1986-1987 estimates a one-year response to the 1987 tax changes. The 1987-1988 difference estimates a one-year response to the 1988 tax changes, but also picks up the second-year response to the 1987 tax changes to the extent that the tax changes in the second year are correlated with those in the first. The 1988-1989 difference will suggest a response (the second-year response to the tax change in the previous year and the third-year response to the tax changes two years before), even when there is not a tax change this period. This will be absorbed in the year fixed-effect. The phase-in was such that most individuals experienced larger tax rate changes in 1986-1987 relative to 1987-1988. Therefore, the response to the tax reform in 1986-1987 was necessarily larger overall than the response in 1987-1988; otherwise, significant estimates would not be obtained (the data would suggest individuals responding the same, or perhaps even less to tax changes that were larger in magnitude). Further, to the extent that the responses in these two years

\textsuperscript{65}In a set of estimates not reported here, which restrict Columns (5), (8), and (9) so that they include the same taxpayers, the one-year estimates are slightly smaller than the two- and three-year estimates, which are almost identical.
were more similar than the gap in tax rate changes suggests, the estimates are biased downwards. Note that if the reform were not phased-in, one-year differences would effectively capture the short-run response. For this reason, tax changes with no phase-in period (and more generally, with tax reforms relatively far apart) are preferred.

This paper emphasizes proper interpretation of estimates using varying difference lengths. However, this issue along with any anticipatory response or short-term income shifting induced by TRA86, are not resolved in this paper. The obvious solution to these problems is to control directly for leads and lags of the tax change. However, Weber (2012b) effectively rules this option out in the U.S. context, because if individuals respond to leads and lags of tax rates, it is almost guaranteed that it will not be possible to identify a causal parameter for the lead and lag terms. A non-causal parameter would both be uninteresting on its own and ineffective at resolving the issues with the static estimates.

Up to this point, I have assumed that $\tau_{it} = \tau_{rt}$ for all individuals; that is, the tax rate that researchers observe, $\tau_{it}$, which is a function of post-response income, is equivalent to the tax rate individuals face when making their decisions about how much to respond, $\tau_{rt}$. However, Weber (2012b) highlights that this assumption likely does not hold in practice. In a classical analysis of individuals’ responses to a marginal tax rate change, $\tau_{it} = \tau_{rt}$ always because individuals never cross tax bracket lines in response to a tax rate change, but the model assumes away dynamic issues such as responding to transitory income shocks and overcoming adjustment costs and other frictions, the latter of which is a particularly well-accepted feature of individuals’ responsiveness (e.g. Powell and Shan, 2012; Saez, 2010; Chetty, 2011). When $\tau_{it} \neq \tau_{rt}$, the treatment is mismeasured, which, if not addressed properly, will bias the estimates. A bias will exist in this case because the first-stage estimate does not accurately capture the average marginal tax rate faced when individuals decide how much to respond. Therefore, Weber (2012b) imposes an additional assumption: in each case, the decision to cross a tax bracket line is independent of the instrument.

Weber (2012b) shows that the instrument in Table 4.2 Column (6), which was proposed by Blomquist and Selin (2010), clearly violates this condition as long as there is some heterogeneity among individuals that are considering crossing the tax bracket line because the instrument is a

---

66 This has been done in Bakija and Heim (2011), Gertz (2008), and Holmlund and Soderstrom (2008), among others.

67 For example, suppose the frictions manifest themselves by generating imperfect bunching around the tax kink. If there is a tax decrease above a given tax kink, there will be individuals above this kink before the tax reform who attempt to bunch at that kink after the reform by choosing a taxable income slightly below the tax kink. The actual treatment for each of these individuals is the tax rate change above the tax kink brought about by the reform, but the researcher would assign them an even larger tax decrease: the decline from the original marginal tax rate above the tax kink to the marginal tax rate below the kink.
function of post-response income; that is, the instrument strongly predicts who did, in fact, cross
the tax bracket line. Column (5) may also violate this condition, but to a much lesser degree; it
only violates it if individuals that choose to cross a tax bracket line today for a particular reason
were clumped in the same tax bracket several years before. Therefore, the estimates in Column (5)
are strictly preferred to those in Column (6),\textsuperscript{68} and I will consider an alternative estimate that is
expected to provide a consistent lower bound under weaker conditions, if Column (5) does, in fact,
vioLate this assumption, too. If the assumption holds,\textsuperscript{69} Weber (2012b) shows that the parameter
obtained is a Fixed-Bracket Average Treatment Effect (FBATE).\textsuperscript{70}

FBATE identifies the average treatment effect for individuals with no incentive to cross a tax
kink in response to a tax reform or transitory shock in taxable income. Interpreting the estimates
in Table 4.2 Column (5) in this light provides a new interpretation regarding their size. Certainly,
part of the increase relative to estimates such as those found in Gruber and Saez (2002) is due to
the decrease in endogeneity of the instruments. However, part of the explanation also likely lies in
the fact that FBATE identifies the ETI for a particular subpopulation. Both instrument exogeneity
and the additional assumption discussed in the last paragraph contribute to the fact that individuals
who face large transitory income shocks do not contribute to the FBATE estimate (i.e. they are
included in the estimation but, given the instrument, their response cancels out). To the extent
that the estimates in Columns (1)-(3) measure this response (albeit endogenously), those estimates
will be lower if individuals do not respond as strongly to tax changes brought about by changes in
transitory income, which seems plausible.

The fact that it is only possible identify a causal parameter for a subpopulation also has important
implications regarding the degree to which these parameters are relevant for welfare analysis. If
individuals who receive transitory income shocks would otherwise (and will eventually) respond
just as those who do not experience these shocks, then an FBATE parameter is actually more
representative of the long-run effect of a tax reform on the population than one that accurately
measures each individual’s short-run response.

As highlighted by Chetty (2011), the bounds on the structural parameter relevant for welfare
analysis are tighter when marginal tax rate changes are high, because it induces more individuals to

\textsuperscript{68}Depending on the exact circumstances, some of the biases induced in Column (6) may cancel out, but unless
one knew that they cancelled out exactly, it remains difficult to interpret this parameter; it is for this reason that the
estimates in Column (5), and ultimately the lower bound estimates discussed below, are strictly preferred.

\textsuperscript{69}There is one more condition needed in order for me to interpret the resulting parameter as an average treatment
effect: for those that do not face an incentive to cross tax bracket lines, there cannot be heterogeneity in their
responsiveness that is correlated with the instrument.

\textsuperscript{70}For a more technical and extensive discussion of each point made in this paragraph as well as a derivation of
FBATE, see Weber (2012b).
overcome their adjustment costs. But, these very same reforms (TRA86 is a prime example) induce a lot of individuals to cross tax bracket lines when they overcome these adjustment costs, and therefore, obtaining a causal parameter (i.e. FBATE) may be more challenging in these contexts; ultimately, the causal estimate will need to be independent of those who face high adjustment costs, such that if they respond, they will cross tax bracket lines. It is overcoming adjustment costs that present the biggest threat to the ability of the estimates in Table 4.2 Column (5) to obtain FBATE, and if FBATE is obtained, these estimates may be too low in the sense that they exclude some highly responsive individuals. Also, if FBATE fails for this reason, the estimates are likely biased downwards since TRA86 provided a marginal tax rate decrease for most individuals.\footnote{This is because individuals who choose to overcome their adjustment costs in response to TRA86 are responding to the tax rate decrease in a higher tax bracket, but the treatment they get assigned is a tax rate increase because researchers observe them in the lower tax bracket in period $t-1$ and in the higher tax bracket in period $t$.}

Given that FBATE identifies the parameter for a subpopulation that does not change tax brackets between years, redefining the treatment as the predicted net-of-tax rate as a function of base-year income, should not change the estimate obtained with the original definition of treatment if the assumptions necessary to obtain FBATE are satisfied, but it will decrease standard errors. Furthermore, as long as the instruments used satisfy instrument exogeneity, Weber (2012b) shows that this estimate is expected to identify a lower bound on the true average treatment effect.

These estimates are presented in Table 4.3. Column (1) replicates the baseline specification for two-year differences found in Column (5) of Table 4.2. Column (2) is the same as Column (1), except treatment is defined as the predicted net-of-tax rate using base-year income. The estimate in Column (2) is 0.975, which is quite similar to the estimate in Column (1)—this suggests the estimates in Column (1) do satisfy the conditions necessary to identify FBATE, or at least the bias due to any violations is not large. The standard errors decrease by 16 percent, suggesting that there is a substantial efficiency gain from employing this alternative treatment definition. Columns (3)-(6) do the same thing for one- and three-year differences and the results are similar to those found for one-year differences (the standard errors decrease by 17-19 percent).

4.4.3 Heterogeneous Growth Rates and Time Trends

Up to now, the theoretical and empirical results have assumed that $g_t$ is the same, on average, for individuals at all points in the income distribution. As noted above, if $g_t$ is the same for all individuals, it will simply be absorbed in the constant term (or if there are more than one pair of years, it will be absorbed by the year fixed-effects). Moreover, including more pairs of years in the regression does not aid in identifying $g_t$ because it is different for each year. However, a homogeneous
growth rate may not a legitimate assumption in practice (at least not in the United States), and this fact has been recognized by researchers in this area.\textsuperscript{72}

Once $g_t$ varies by income level, such that $g_{jt}$ is constant for all individuals within income class $j$, but varies across income classes, controlling for variation in $g_t$ across income classes is crucial for obtaining consistent estimates.\textsuperscript{73} If it is not controlled for, the income-class-varying portion of the growth rate will end up in the error term, and is likely highly persistent. Therefore, it will likely be correlated with lags of base-year income used to construct the tax rate instrument that would otherwise be exogenous. In most other literatures, this issue is dealt with by controlling for known factors that influence $g$. For example, MaCurdy (1982) included “...family background variables, education, age, interactions between education and age, and dummy variables for each year of the sample” in his study on the properties of the error structure of earnings. But, such data are not usually available in tax return data. Additionally, in ETI research, perhaps even these would not be enough to fully eliminate the endogeneity problem. Hence, the literature has turned to a second-best alternative, namely including base-year income controls. While they are endogenous when included directly, they can be instrumented using the same lag as is used to instrument for the tax rate variable. When such a suitable lag is used, the income controls effectively control for permanent income plus an uncorrelated measurement error.\textsuperscript{74} This measurement error is uncorrelated with permanent income by definition. Therefore, when instrumented, log income is a valid control for permanent income, which is what is relevant for determining the heterogeneity in the growth rates, and is a best-case scenario in terms of being a proxy for things such as age and education level.\textsuperscript{75}

If the income class-specific growth trend increases linearly with income, then including $ln(Y_{it-1})$, instrumented with the appropriate lag, will be enough to obtain consistent estimates. However, if the relationship is believed to be non-linear, splines should be employed, an observation that has been widely acknowledged in the literature.\textsuperscript{76} When splines are used, identification becomes more challenging because the income class-specific growth trend must be identified separately from the behavioral response to the tax rate, where the tax rate change also varies with income levels (this is, after all, how identification is obtained in the first place). Usually, when splines are employed,
it is assumed that \( g_{jt} = g_j \), that is, the heterogeneous time trends do not vary over time.\(^77\) Then, including additional pairs of years aids in identifying these growth rates separately from the tax changes.\(^78\)

Table 4.4 examines the importance of heterogeneous income controls empirically. I begin with two-year differences. I use a 5-piece spline in the log of constant-law taxable income. In choosing the number of spline segments used, the researcher must balance preferences for flexibility with a desire to keep identification. Empirically, this issue has resolved itself, since a wide range in the number of spline segments employed yield similar estimates.\(^79\) Since the splines are in logs of constant-law taxable income, negative values are excluded from the estimation, both for the base-year (which were already excluded due to the income cutoff at $10,000) and in each year used as an instrument. The results will use two, three, and four lags of income as instruments, so Column (1) repeats Column (5) of Table 4.2 as a benchmark, excluding those with negative constant-law taxable income lagged two, three, and four periods. Note that this may change the composition of the estimates slightly because individuals with some types of income are more likely to have negative constant-law taxable income than others (and this is likely persistent over time).\(^80\) Empirically, this cutoff has a minimal effect on the estimates.

Table 4.4 Column (2) adds the base-year income spline directly, which as noted in the previous paragraphs is expected to be endogenous. Column (3) assumes that income splines as a function of three and four lags of income are exogenous and tests whether the income splines as a function two lags are also exogenous.\(^81\) Including the endogenous income splines in Column (2) more than doubles the elasticity estimates; however, when instrumented with the appropriate lag, the estimate is 1.461, which is only slightly higher than the estimate in Column (1) without controls for heterogeneous income growth. This suggests that using a lag of income is important; in fact, the sign of the majority of segments of the income spline flip between Column (2) and Column (3) and none of the coefficients on the splines are significant in Column (3). The p-value of the Difference-in-Sargan test for instrument exogeneity of the spline coefficients in Column (3) is 0.227. Column (3) also

\(^77\)Gruber and Saez (2002) test this assumption and find that it does not matter much.

\(^78\)Another alternative is proposed in Gelber (2010). Here the spline coefficients are obtained by regressing the change in taxable income on the spline segments in a year in which there were no tax changes. Then, a new dependent variable is constructed; it is the log change in taxable income minus the spline coefficients times the spline segments in the base year. This method imposes the same assumption that the spline coefficients do not change across years. Just as when the spline segments are included in the main regression directly, this method is valid as long as the spline segments are instrumented with lags, so that the spline coefficients will be consistent.

\(^79\)For example, Gruber and Saez (2002) use a 10-piece spline, but note that their results change very little when a 20-piece spline is employed. Although not reported here, I find the same robustness to the order of spline used.

\(^80\)However, there is no reason to believe that this additional cutoff alters the endogeneity of the tax rate or income spline instruments, although it could have a slight effect in practice.

\(^81\)The results assuming two and three lags are exogenous and testing whether one lag is exogenous are not reported here, but I find that I can reject the null hypothesis that one lag is exogenous at the 1 percent level.
suggests that controlling for heterogeneous growth rates has a minimal effect on the estimates. This finding is consistent with the fact that researchers who have been able to include more demographic controls have found that they have little effect on the estimated coefficients (Carroll, 1998; Auten and Carroll, 1999).

The results up to this point assume that heterogeneous growth rates can be proxied by constant-law taxable income. Another reasonable candidate for this proxy is constant-law broad income. Constant-law taxable income begins with constant-law broad income and then adjusts for constant-law exemptions, and either constant-law itemizations or a constant-law standard deduction. I consider how this changes the results in Table 4.4 Column (4), which repeats Column (3), except the spline is in broad, rather than taxable income. The type of income used in this context to control for heterogeneous income trends does not have a substantial effect on the estimated ETI coefficient. The estimated ETI (1.240) is quite similar to the ETI in Column (3), although all the coefficients on the spline in Column (4) are statistically significant at the 5 percent level, which was not true in Column (3).

Note that the estimating equation would remain valid if I included the lags of income directly as proxy variables, rather than instrumenting for the income spline. If included as a proxy variable, the researcher is left to guess whether or not the proxy variable is lagged enough periods to be exogenous, and whether or not the proxy variable is still close enough to the base-year to be a relevant control. By using an instrument, I was able to test the former using the Difference-in-Sargan test and the latter by testing for weak instruments.

Up to this point, I have not weighted the estimates by income, because, while this is the appropriate parameter estimate for welfare analysis, this data set does not oversample high-income individuals. As a result, the weighted estimates put substantial weight on individuals whose responses can be estimated with relatively little precision. Column (5) of Table 4.4 repeats Column (3), but weights the estimates by income. Weighting has a minimal effect on the estimates, but it does increase the standard errors by more than 30 percent. Columns (6) and (7) take advantage of the increase in efficiency of the alternative definition of treatment proposed in Subsection 4.4.2.4

---

82I keep the restriction that constant-law taxable income must be positive in each lag so that the composition of each column is the same. When I relax this restriction, I can reject the null that two lags of broad income are an exogenous instrument. When I include three, four, and five lags as instruments instead, I cannot reject that three lags are endogenous. This alternative approach has a minimal effect on the estimated ETI coefficient, so these estimates are not reported here.

83In practice when a proxy variable has been employed, the proxy is constructed several years before the first year examined in the estimation for all years. Sometimes these proxy variables are expected to be equally informative more than ten years later. For example, see Blomquist and Selin (2010), or Holmlund and Soderstrom (2008). Individual’s permanent incomes can change a lot over a decade; so, while the proxy may be exogenous, it may not be controlling very meaningfully for the appropriate growth rate towards the end of the sample period.
to attempt to obtain weighted estimates with lower standard errors. Column (6) repeats Column (4) of Table 4.3, but includes heterogeneous income controls. Column (7) adds income weights to Column (6). As expected, the estimates in Column (5) and (7) are quite similar, but the standard errors are almost 20 percent lower. Still, the standard errors on the estimates in Column (7) are too large to be of much practical use.

Table 4.4 Columns (8) and (9) repeat Column (3) for one- and three-year differences, respectively. The estimates are 0.749 and 1.704. These results highlight that controlling for heterogeneous income trends is more important as the difference length increases, as expected.

4.5 Conclusion

As aptly summarized by Saez et al. (2012), in their recent *Journal of Economic Perspectives* article on the ETI, a longstanding problem in the ETI panel data estimation literature “is that the identification assumptions lack transparency because they mix assumptions regarding mean reversion and assumptions regarding changes in income inequality.” This paper has carefully disentangled these two issues both theoretically and empirically. The modal approach in the literature—to try to simultaneously rectify mean reversion and heterogeneous income trends with the use of some type of base-year income splines—resolves neither problem. They are ineffective both theoretically and empirically, and the magnitude of the estimates change substantially when alternative methods that resolve the issues properly are employed.

Using these methods, I obtain a baseline ETI estimate of 1.046. Under the appropriate assumptions, this estimate can be interpreted as an FBATE estimate. The increase in magnitude of the estimate relative to frequently cited estimates, such as (Gruber and Saez, 2002), are likely due to: 1) the substantial violations of the conditions needed to satisfy FBATE induced a downward bias in the estimates in most cases, and 2) individuals in the subpopulation that identifies FBATE are often expected (although there are a few notable exceptions) to be more responsive during the time in which estimation occurs. Assuming those effectively excluded from FBATE will respond the same, on average, in the long-run, this estimate is the relevant parameter for welfare analysis.

The U.S.-centered nature of this paper is partially due to the fact that mean reversion is believed to be particularly strong in the U.S., and partially because I am using a U.S. data set. But, there is a large literature that estimates the ETI for other countries.\(^{84}\) The theoretical results in this paper are equally applicable to these other countries. Given the extreme nature of the assumptions needed

\(^{84}\)See Saez et al. (2012) for a review of this literature.
to produce a consistent estimate using the methodologies most commonly employed, it is likely that many of the estimates obtained for other countries that are based on the methods discussed in this paper are also inconsistent.

While much has been addressed in this paper, there are still several important avenues for future research. I chose the Michigan panel data set for years 1979-1990 because of its widespread use and ease of access (i.e. it is publicly available). However, it is not ideal in several dimensions. It does not oversample high-income taxpayers, substantial tax base changes accompany the tax rate changes, and the tax reform was both anticipated and phased-in. Each of these issues presents challenges that were discussed in the paper, but are likely not completely overcome. Additionally, with a data set improved in these dimensions, one could confidently estimate an income-weighted, compensated elasticity.\(^{85}\)

\(^{85}\)I do not attempt to obtain a compensated elasticity in this paper because to do so would require separately identifying the income effect, as in Gruber and Saez (2002). Given the substantial changes in the tax base as part of TRA86, it is unlikely that I could obtain a convincing estimate of the income effect in this context. Researchers estimating these two effects separately with another data set would likely be well-served to use the alternative definition of treatment proposed in Subsection 4.4.2.4 and its income effect analog in order to maximize efficiency of these estimates.
Table 4.1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxable Income</td>
<td>$36,889.00</td>
<td>$40,599.69</td>
</tr>
<tr>
<td>Federal Tax Rate</td>
<td>23.71</td>
<td>7.63</td>
</tr>
<tr>
<td>State Tax Rate</td>
<td>4.46</td>
<td>3.28</td>
</tr>
<tr>
<td>Single Dummy</td>
<td>0.245</td>
<td>-</td>
</tr>
<tr>
<td>Married Dummy</td>
<td>0.703</td>
<td>-</td>
</tr>
<tr>
<td>Observations</td>
<td>25,087</td>
<td>-</td>
</tr>
</tbody>
</table>

Taxable income is in 1992 dollars. These summary statistics are for years 1983-1990 and match the restrictions imposed by the baseline estimates given in Table 4.2, Column (5).
Table 4.2: Baseline Results

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta \ln(1 - \tau_t) )</td>
<td>0.144</td>
<td>0.453***</td>
<td>0.171</td>
<td>0.693***</td>
<td>1.046***</td>
<td>1.145***</td>
<td>0.906***</td>
<td>0.780**</td>
<td>0.808***</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.138)</td>
<td>(0.117)</td>
<td>(0.237)</td>
<td>(0.299)</td>
<td>(0.315)</td>
<td>(0.234)</td>
<td>(0.332)</td>
<td>(0.309)</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Quintile Spline&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>-0.174***</td>
<td>-0.284***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.048)</td>
<td>(0.046)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Quintile Spline</td>
<td>0.137</td>
<td>0.335**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.147)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Quintile Spline</td>
<td>-0.036</td>
<td>0.132</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(0.225)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; Quintile Spline</td>
<td>-0.046</td>
<td>-0.471**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.190)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt; Quintile Spline</td>
<td>0.085</td>
<td>0.270***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.086)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruments&lt;sup&gt;3&lt;/sup&gt;</td>
<td>0.23 lags</td>
<td>0.23 lags</td>
<td>0.23 lags</td>
<td>1.23 lags</td>
<td>2.34 lags</td>
<td>1 lead</td>
<td>1 lead</td>
<td>2.34 lags</td>
<td>2.34 lags</td>
</tr>
<tr>
<td>Years Included</td>
<td>82-90</td>
<td>82-90</td>
<td>82-90</td>
<td>82-90</td>
<td>83-90</td>
<td>83-90</td>
<td>79-90</td>
<td>83-90</td>
<td>83-90</td>
</tr>
<tr>
<td>Difference Length</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>1 year</td>
<td>3 years</td>
</tr>
<tr>
<td>Observations</td>
<td>29,547</td>
<td>29,547</td>
<td>29,547</td>
<td>29,547</td>
<td>25,087</td>
<td>25,087</td>
<td>80,419</td>
<td>31,646</td>
<td>19,723</td>
</tr>
<tr>
<td>Individuals</td>
<td>6,602</td>
<td>6,602</td>
<td>6,602</td>
<td>6,602</td>
<td>6,224</td>
<td>6,224</td>
<td>31,971</td>
<td>6,899</td>
<td>5,693</td>
</tr>
<tr>
<td>Diff-in-Sargan p-value</td>
<td>0.006</td>
<td>0.000</td>
<td>0.063</td>
<td>0.163</td>
<td>0.427</td>
<td></td>
<td>0.895</td>
<td></td>
<td>0.112</td>
</tr>
<tr>
<td>F-statistic</td>
<td>601.4</td>
<td>440.7</td>
<td>597.6</td>
<td>157.0</td>
<td>116.4</td>
<td>226.5</td>
<td>548.6</td>
<td>88.47</td>
<td>114.9</td>
</tr>
</tbody>
</table>

1 Each column is estimated using 2SLS. Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. Indicator variables for marital status and base years are also included in estimation.

2 In Column (2), the splines are a function of log base-year income. In Column (3), the splines are a function of \( \Delta \ln(Y_{t-1}) \). The spline coefficients give the marginal change from the previous spline coefficient.

3 This is a list of the predicted net-of-tax rate instruments used in each column. For example, Column (1) lists the instruments as no lag, two lags, and three lags. This means that predicted net-of-tax rate instruments are constructed for this column as a function of base-year income, income two periods before the base-year, and income three periods before the base-year. The second two instruments in the list in each column are used to test whether the first instrument listed is exogenous using the Difference-in-Sargan test.
Table 4.3: Alternative Definition of Treatment

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \ln(1 - \tau_i)$</td>
<td>1.046***</td>
<td>0.780**</td>
<td>0.808***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.299)</td>
<td>(0.332)</td>
<td>(0.309)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta \ln(1 - \tau_i^p)$</td>
<td></td>
<td>0.975***</td>
<td>0.677***</td>
<td>0.770***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.251)</td>
<td>(0.268)</td>
<td>(0.265)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruments$^2$</td>
<td>2.3, 4 lags</td>
<td>2.3, 4 lags</td>
<td>2.3, 4 lags</td>
<td>2.3, 4 lags</td>
<td>2.3, 4 lags</td>
<td>2.3, 4 lags</td>
</tr>
<tr>
<td>Years Included</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
</tr>
<tr>
<td>Difference Length</td>
<td>2 years</td>
<td>2 years</td>
<td>1 year</td>
<td>1 year</td>
<td>3 years</td>
<td>3 years</td>
</tr>
<tr>
<td>Observations</td>
<td>24,036</td>
<td>24,036</td>
<td>30,315</td>
<td>30,315</td>
<td>18,922</td>
<td>18,922</td>
</tr>
<tr>
<td>Individuals</td>
<td>5,875</td>
<td>5,875</td>
<td>6,582</td>
<td>6,582</td>
<td>5,385</td>
<td>5,385</td>
</tr>
<tr>
<td>Diff-in-Sargan p-value</td>
<td>0.490</td>
<td>0.728</td>
<td>0.815</td>
<td>0.993</td>
<td>0.147</td>
<td>0.236</td>
</tr>
<tr>
<td>F-statistic</td>
<td>114.4</td>
<td>264.0</td>
<td>90.15</td>
<td>195.3</td>
<td>111.3</td>
<td>244.2</td>
</tr>
</tbody>
</table>

$^1$Each column is estimated using 2SLS. Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. Indicator variables for marital status and base years are also included in estimation.

$^2$This is a list of the predicted net-of-tax rate instruments used in each column. For example, Column (1) lists the instruments as two lags, three lags, and four lags. This means that predicted net-of-tax rate instruments are constructed for this column as a function of income two, three, and four periods before the base-year. The second two instruments in the list in each column are used to test whether the first instrument listed is exogenous using the Difference-in-Sargan test.
Table 4.4: Heterogeneous Income Trends

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta \ln(1 - \tau_t) )</td>
<td>1.180***</td>
<td>2.467****</td>
<td>1.461***</td>
<td>1.240***</td>
<td>1.033*</td>
<td>0.749*</td>
<td>1.704***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.311)</td>
<td>(0.496)</td>
<td>(0.414)</td>
<td>(0.395)</td>
<td>(0.544)</td>
<td>(0.384)</td>
<td>(0.598)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta \ln(1 - \tau_t^p) )</td>
<td></td>
<td></td>
<td></td>
<td>1.232***</td>
<td>0.954**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.281)</td>
<td>(0.454)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Quintile Spline(^2)</td>
<td>-0.177****</td>
<td>-0.179</td>
<td>-0.483***</td>
<td>-0.135</td>
<td>0.072</td>
<td>0.142</td>
<td>0.317</td>
<td>-0.860</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.366)</td>
<td>(0.162)</td>
<td>(0.574)</td>
<td>(0.325)</td>
<td>(0.520)</td>
<td>(0.314)</td>
<td>(0.543)</td>
<td></td>
</tr>
<tr>
<td>2nd Quintile Spline</td>
<td>0.040</td>
<td>0.375</td>
<td>0.990***</td>
<td>0.297</td>
<td>-0.073</td>
<td>-0.155</td>
<td>-0.482</td>
<td>1.518</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.719)</td>
<td>(0.334)</td>
<td>(0.996)</td>
<td>(0.641)</td>
<td>(0.909)</td>
<td>(0.598)</td>
<td>(1.042)</td>
<td></td>
</tr>
<tr>
<td>3rd Quintile Spline</td>
<td>0.020</td>
<td>-0.666</td>
<td>-1.102***</td>
<td>-0.506</td>
<td>-0.238</td>
<td>-0.150</td>
<td>0.146</td>
<td>-1.361</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.664)</td>
<td>(0.399)</td>
<td>(0.749)</td>
<td>(0.591)</td>
<td>(0.687)</td>
<td>(0.501)</td>
<td>(0.944)</td>
<td></td>
</tr>
<tr>
<td>4th Quintile Spline</td>
<td>-0.125</td>
<td>0.607</td>
<td>0.903**</td>
<td>0.396</td>
<td>0.351</td>
<td>0.205</td>
<td>-0.001</td>
<td>0.609</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.468)</td>
<td>(0.354)</td>
<td>(0.481)</td>
<td>(0.415)</td>
<td>(0.446)</td>
<td>(0.306)</td>
<td>(0.663)</td>
<td></td>
</tr>
<tr>
<td>5th Quintile Spline</td>
<td>0.061</td>
<td>-0.177</td>
<td>-0.347***</td>
<td>-0.145</td>
<td>-0.141</td>
<td>-0.132</td>
<td>0.032</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.191)</td>
<td>(0.171)</td>
<td>(0.192)</td>
<td>(0.168)</td>
<td>(0.177)</td>
<td>(0.109)</td>
<td>(0.273)</td>
<td></td>
</tr>
<tr>
<td>( \Delta \ln(1 - \tau_t) ) Instrument lags(^3)</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
</tr>
<tr>
<td>Spline Instrument lags(^3)</td>
<td>None</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
<td>2.3,4</td>
</tr>
<tr>
<td>Years Included</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
</tr>
<tr>
<td>Difference Length</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>1 year</td>
<td>3 years</td>
<td></td>
</tr>
<tr>
<td>Income Weights?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>22,991</td>
<td>22,991</td>
<td>22,991</td>
<td>22,991</td>
<td>22,991</td>
<td>22,991</td>
<td>28,735</td>
<td>18,159</td>
<td></td>
</tr>
<tr>
<td>Individuals</td>
<td>5,650</td>
<td>5,650</td>
<td>5,650</td>
<td>5,650</td>
<td>5,650</td>
<td>5,650</td>
<td>6,186</td>
<td>5,172</td>
<td></td>
</tr>
<tr>
<td>Diff-in-Sargan p-value(^4)</td>
<td>0.227</td>
<td>0.203</td>
<td>0.163</td>
<td>0.293</td>
<td>0.245</td>
<td>0.672</td>
<td>0.550</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1Each column is estimated using 2SLS. Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. Indicator variables for marital status and base years are also included in estimation.
2In Column (4), the splines are a function of log base-year broad income. In all other columns, the splines are a function of log base-year taxable income. The spline coefficients give the marginal change from the previous spline coefficient.
3This is a list of the predicted net-of-tax rate instruments used in each column. For example, Column (1) lists the instruments as two lags, three lags, and four lags. This means that predicted net-of-tax rate instruments are constructed for this column as a function of income two, three, and four periods before the base-year. The second two instruments in the list in each column are used to test whether the first instrument listed is exogenous using the Difference-in-Sargan test.
4These are the p-values from testing whether a given income spline is exogenous.
CHAPTER V

Conclusion

Tax researchers often want to estimate the behavioral response to tax changes using pre-reform characteristics as instruments for the observed tax rate change, but Chapter 2 provides a cautionary tale. The Wald estimates obtained from this exercise are only useful to the extent that they provide a causal average treatment effect that is relevant for deadweight loss. This paper brings these basic ideas back to the forefront in this literature, showing that many papers do not obtain a causal average treatment effect. Even for those that do obtain a causal average treatment effect, which I show can be interpreted as a Fixed-Bracket Average Treatment Effect (FBATE), it is not guaranteed that this parameter will be relevant for welfare analysis. Therefore, in each context, researchers should carefully consider whether the methodology considered in this chapter is appropriate to use in their context.

Chapter 3, which analyzes the effect of the EITC on saving, estimates ITT parameters rather than Wald estimates precisely because of the concerns raised about the latter in Chapter 2. I also implement a new method of analyzing the effect of tax changes around tax kinks. These methods provide evidence that the response in investment income to the tax rate increases over the period are substantial. Indeed, nearly 40 percent of the decline in saving in income-bearing accounts by EITC recipients over the last two decades can be explained by the changing EITC incentives. The response is twice as large among those that use a paid preparer, consistent with an increased awareness of the relevant incentives among this group. The policy implications for EITC design depend on how policymakers weight the deadweight loss induced by this provision, relative to its original intent—to exclude individuals that were not actually poor from receiving the EITC. Making the EITC amount exclusively a function of earned income would eliminate the distortion found in this paper, but would increase the number of claimants by 4 percent; these additional claimants would receive an average of $1,014 in benefits, even though their average AGI is $52,179.
Chapter 4 takes Chapter 2 very seriously and carefully chooses an instrument that will satisfy the conditions needed to obtain FBATE, in a context where most previous estimators did not satisfy basic instrument exogeneity conditions and no estimates satisfied the conditions needed to obtain FBATE. Using these methods, I obtain a baseline ETI estimate of 1.046. Under the appropriate assumptions, this estimate can be interpreted as an FBATE estimate. The increase in magnitude of the estimate relative to frequently cited estimates, such as (Gruber and Saez, 2002), are likely due to: 1) the substantial violations of the conditions needed to satisfy FBATE induced a downward bias in the estimates in most cases, and 2) individuals in the subpopulation that identifies FBATE are often expected (although there are a few notable exceptions) to be more responsive during the time in which estimation occurs. Assuming those effectively excluded from FBATE will respond the same, on average, in the long-run, this estimate is the relevant parameter for welfare analysis.

This dissertation has made progress designing and implementing estimation methods that are appropriate given the particular issues that need to be considered when the treatment of interest is a tax rate change. More research on this front is needed.


