

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

*Caroline Weber\**

*April 2014*

## **Abstract**

The elasticity of taxable income (ETI) is a central parameter for tax policy debates. This paper shows that mean reversion prevents most estimators employed in the literature from obtaining consistent estimates of the ETI. A new method is proposed that will resolve inconsistency due to mean reversion under testable assumptions regarding the degree of serial correlation in the error term. Using this procedure, I estimate an ETI of 0.858, which is about twice as large as the estimates found in the most frequently cited paper on this subject (Gruber and Saez, 2002). The corresponding elasticity of broad income is 0.475.

---

Many thanks to David Agrawal, Gerald Auten, Dora Benedek, Wojciech Kopczuk, Alexander Gelber, Jim Hines, Laura Kawano, Emmanuel Saez, Håkan Selin, Joel Slemrod, Jeff Smith, anonymous referees, University of Michigan seminar participants, and International Institute of Public Finance 2010 conference 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). All remaining errors are my own.

\*Assistant Professor, Department of Economics, University of Oregon: cweber5@uoregon.edu.

JEL Codes: H21, H24

# 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.<sup>1</sup> 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 (Feldstein, 1999; Chetty, 2009). 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—that the instrument is uncorrelated with transitory shocks in the error term—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<sup>2</sup>—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 a large 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 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. The first main contribution of this paper is to define an

---

<sup>1</sup>The marginal net-of-tax rate is one minus the marginal tax rate.

<sup>2</sup>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.

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<sup>3</sup>), 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. I then use the Michigan IRS Tax Panel data set for the years 1979-1990 to empirically estimate the ETI using this instrument. My preferred baseline ETI estimate is 0.858. The corresponding estimate for the elasticity of broad income is 0.475, likely highlighting the degree to which there were significant changes in itemizations in response to the tax reform. To the extent that adjustments in itemized deductions do not incur significant costs (Chetty, 2009) or itemized deductions, such as charitable giving, create positive externalities, it is the elasticity of broad income, not taxable income, that is relevant for welfare analysis.

I conduct a variety of sensitivity analyses and estimate local polynomial regressions to provide a graphical examination of the identifying variation exploited in this regression. A few key findings from this analysis are: (i) the estimates do not appear to be biased by heterogeneous income trends or an anticipation response, although this possibility cannot be completely ruled out, (ii) the estimates are not biased by shifting from C-corporation to S-corporation status in response to the tax reform, and (iii) there is a homogeneous response to tax rate changes, on average, across the income distribution with one exception: above the top one percent of the income distribution, the response is more elastic. The higher elasticity at the top has a minimal effect on the estimates because, while the elasticity

---

<sup>3</sup>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.

visually increases, the change is not statistically significant; this is mostly due to the fact that this data set does not oversample high-income individuals.

## 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 (Saez et al., 2009). This paper exclusively discusses panel-based difference-in-differences. Identification usually comes from differences in tax rate changes across individuals brought about by a tax reform. Because 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 large tax rate changes to those lower in the income distribution who experience low or no tax rate changes. Estimating the ETI 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 (base-year 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.<sup>4</sup> Because the instrument is still a function of

---

<sup>4</sup>At least this has been traditionally true. More recently, data sets from other countries which have better demographic data have been employed. And, in the U.S., Singleton (2011) has linked two different

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 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 between 0.4 to 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 federal reforms were targeted tax increases (the Omnibus Budget Reconciliation Acts of 1990 and 1993). These estimates ranged from 0.19 to 0.38 (Carroll, 1998; Giertz, 2010). There has since been a large literature using the same methods to estimate the ETI in other countries (Saez et al., 2009).

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)

---

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 studies to make use of demographic instruments to estimate the ETI were Moffitt and Wilhelm (2000) who used the Survey of Consumer Finances (SCF) and a very recent paper by Burns and Ziliak (2013) who make use of the Current Population Survey (CPS), not actual tax return data sets.

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 all the proposed estimators to determine whether they provide consistent estimates of the ETI and proposes new methods when necessary. I begin in Section 3 by setting up a simple model of income and within this, characterizing mean reversion and income trends. In Section 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 provides sensitivity analyses. Section 5 concludes.

### 3 Model Setup

To formalize the main issues that make estimating the ETI difficult, I set up a simple model of the taxable income process in this section. This process is consistent with the theoretical model that drives estimation strategies in this literature and 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. For example, the literature has explored the role of income effects (Gruber and Saez, 2002) and tax-base effects (Kopczuk, 2005). If such additional complexities matter, they will need to be addressed in order to obtain a consistent estimate of the ETI. 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.

Let individuals' log income  $\ln(Y_{it})$  be governed by the following equation:<sup>5</sup>

$$\ln(Y_{it}) = \varepsilon \ln(1 - \tau_{it}) + \ln(\mu_{it}) + \ln(\nu_{it}), \quad (1)$$

---

<sup>5</sup>I assume the response takes one period; this is relaxed in the empirical application.

or in differences as:

$$\Delta \ln(Y_{it}) = \varepsilon \Delta \ln(1 - \tau_{it}) + \Delta \ln(\mu_{it}) + \Delta \ln(\nu_{it}), \quad (2)$$

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.<sup>6</sup> Neither  $\mu_{it}$  nor  $\nu_{it}$  include adjustments to the current tax rate; that is, they measure permanent and transitory income levels when the tax rate is zero. Additionally,  $\tau_{it}$  is the individual's marginal tax rate, which is a function of log income  $\ln(Y_{it})$  and the tax code  $c_t$ :

$$\tau_{it} = f(\ln(Y_{it}), c_t). \quad (3)$$

Assume income grows at the same rate on average, regardless of the individual's income level. Then,  $g_t = \Delta \ln(\mu_t)$  is the period-specific homogeneous income growth rate. Suppressed in  $g_t$  is everything that affects the income-growth profile of an individual, which is likely not homogeneous across income levels. I relax this assumption in Section 4.3; for now it simplifies notation and the analysis conducted prior to Section 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:

$$\ln(\nu_{it}) = \sum_{k=1}^K \phi_k \ln(\nu_{it-k}) + \ln(\xi_{it}), \quad (4)$$

where  $K$  is the order of autocorrelation and  $|\phi_k| < 1$  for all  $k$ . Let  $\ln(\xi_{it}) \sim iid(0, \sigma_\xi^2)$  for all time periods. I assume the same serial correlation process for all individuals for notational convenience, which may not be accurate. For example, Kopczuk (2012) shows the process is different for business owners and wage earners. As long as the instrument chosen is independent of each earnings process separately, the estimates will not be biased.

Mean reversion is a term that describes certain behaviors of the transitory component of

---

<sup>6</sup>I assume that  $\varepsilon$  is the same for all individuals; this is relaxed in Section 4.2.3.

income.<sup>7</sup> Because  $\mathbb{E}[\ln(\xi_t)] = 0$ , individuals receive a mean zero shock each period.<sup>8</sup> When  $\phi_k = 0$  for all  $k$ , mean reversion at the individual level is strong, because current income is no longer a function of transitory income in previous periods. Hence, if individuals have high or low incomes relative to their permanent income level this period, on average, their incomes will return to their mean level in the next period. As  $\phi_k \rightarrow 1$ , mean reversion weakens. Thus, when examining data with  $0 \leq \phi_k < 1$ , 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  $\sigma_\xi^2$  also affects the severity of mean reversion. If there is no income volatility (i.e.  $\sigma_\xi^2=0$ ), then there is no mean reversion, because all transitory shocks are zero. As  $\sigma_\xi^2$  increases, the magnitude of mean reversion also increases. As noted in Section 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 Section 4.2.3.

## 4 Data and Estimation

In this section, I derive conditions under which a consistent estimate of the ETI is identified. I also implement the results empirically, which quantifies the bias induced by incorrect methods in this context, and ultimately provides consistent estimates of the ETI under certain assumptions.

This section proceeds as follows. Section 4.1 describes the data that will be used in the empirical analysis. Section 4.2 conducts baseline theoretical and empirical analysis. Sections 4.3 through 4.7 provide evidence of the validity of the instrument and robustness of the estimates using both parametric and semi-parametric (graphical) methods.

---

<sup>7</sup>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 Kopczuk (2003, 2005), Moffitt and Wilhelm (2000), and Saez et al. (2012), among others.

<sup>8</sup>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  $t$ .



## 4.1 Data

This section describes the data used in the regression analysis in Sections 4.2 through 4.7. I use 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. Because I will use instruments that are a function of lagged income, I restrict the estimation to the years 1983-1990. 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. For an extensive discussion of this data set and TRA86, see Gruber and Saez (2002).

The definition of taxable income used to construct the dependent variable and the income splines used in some specifications 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—changes 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. The estimates will also be biased if the tax base changes fall disproportionately on individuals in a particular income range. 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.<sup>9</sup> Addressing this issue more completely is beyond the scope of this paper.<sup>10</sup> As is common in this literature, I exclude capital gains entirely.<sup>11</sup> Taxable

---

<sup>9</sup>Heim (2006) shows that using a constant-law measure of taxable income will generally lead to biased estimates unless the cross-price elasticities between sources of income that are taxable under one of the alternative definitions of taxable income and the portion of income that is taxable under all regimes is zero.

<sup>10</sup>Kopczuk (2005) addresses this issue more directly by controlling for changes in the tax base directly in the estimating equation.

<sup>11</sup>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.

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.<sup>12</sup> There are three marital status categories in total: single, married, and head of household/widowed with a dependent child. Marital status indicators are the only covariates used by Gruber and Saez (2002). I add several additional covariates that are available in the tax return data set, although I will show in Section 4.7 that these additional covariates have little effect on the estimates. I include a set of indicators for the number of dependent children in the household: zero, one, two, three, and four or more dependent children. I also include a set of indicators for the number of individuals in the taxpaying unit that were over age 65 or blind and a full set of state indicators.

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.<sup>13</sup> Using an income cutoff is common practice in the literature.<sup>14</sup> Much of the justification for including an income cutoff—too much mean reversion at the low end of the income distribution—will likely be resolved by the instruments ultimately used in this section. However, the low end of the income distribution may remain a poor comparison group for other reasons and they will be censored at some income level (which may change over time) because not all individuals will file who are below the filing threshold. This is particularly true in this data set before the large EITC expansion in the 1990's, which encouraged many of these individuals to file. The proposed \$10,000 cutoff is not endogenous

---

<sup>12</sup>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.

<sup>13</sup>Less than twenty individuals, whose 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.

<sup>14</sup>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. Excluding a certain portion of the population can improve the mean reversion problem, but it only eliminates the problem if everyone left in the sample experiences the same transitory income shock.

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 Section 4.2.3 and show that the estimates are robust to alternative income cutoffs in Section 4.7. Descriptive statistics for the preferred baseline estimates are provided in Table 1.

**Table 1: Descriptive Statistics**

	Mean	Std.Dev
Taxable Income	\$37,033	\$40,668
Federal Tax Rate	23.811	7.502
State Tax Rate	4.575	3.211
Single Indicator	0.243	0.429
Married Indicator	0.705	0.456
Under Age 65 and Not Blind Indicator	0.901	0.299
One Person Over Age 65 or Blind Indicator	0.060	0.238
No Dependents Indicator	0.541	0.498
One Dependent Indicator	0.178	0.383
Two Dependents Indicator	0.196	0.397
Three Dependents Indicator	0.060	0.238
Observations	24,576	-

Taxable income is in 1992 dollars. These summary statistics are for years 1983-1990 and match the restrictions imposed by the baseline estimates in Table 2.

## 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 (2) in an estimable form, which yields:

$$\Delta \ln(Y_{it}) = \varepsilon \Delta \ln(1 - \tau_{it}) + \alpha_{t-1} + \Delta \ln(\nu_{it}), \quad (5)$$

where  $\alpha_{t-1}$  are year fixed effects. 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  $g_t$ . If all individuals do not share the same income trend, including year fixed effects is not enough to produce consistent estimates. I address this case in Section 4.3.

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.  $\mathbb{E}[\Delta \ln(Y_t)' \Delta \ln(\nu_t)] > 0$ ), 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. Section 4.2.1 uses the income process laid out in Section 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. Section 4.2.2 uses the same method to examine whether there are alternative instruments which are exogenous under certain assumptions. Section 4.2.3 empirically tests whether each instrument discussed in the previous two subsections is orthogonal to the error term. It then provides baseline ETI estimates.

#### 4.2.1 Using the Predicted Net-of-Tax Rate as an Instrument

How to address the endogeneity of the net-of-tax rate variable 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  $t - 1$ . I will refer to this instrument as  $1 - \tau_{it}^p$ . In this subsection, I focus exclusively on the ability of this instrument to solve the endogeneity problem. I discuss alternative instruments in Section 4.2.2.

There are two necessary conditions for an instrument to be valid: (1) the instrument cannot be weak, i.e.:

$$|\text{cov}[\Delta \ln(1 - \tau_t), \Delta \ln(1 - \tau_t^p) | X_t]| \quad (6)$$

is large, and (2) the instrument is orthogonal to the error term:

$$\text{cov}[\Delta \ln(1 - \tau_t^p), \Delta \ln(\nu_t)] = \mathbb{E}[\Delta \ln(1 - \tau_t^p)' \Delta \ln(\nu_t) | X_t] = 0, \quad (7)$$

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):

$$plim(\hat{\varepsilon}_{IV}) = \varepsilon + \frac{cov[\Delta \ln(1 - \tau_t^p), \Delta \ln(\nu_t) | X_t]}{cov[\Delta \ln(1 - \tau_t), \Delta \ln(1 - \tau_t^p) | X_t]}, \quad (8)$$

and the estimates will be inconsistent.

The predicted net-of-tax rate instrument likely does not suffer from the weak instrument problem in practice. For example, the F-statistics for the first-stage results estimated by Gruber and Saez (2002) are all between 20 and 100. But, it is not clear that the instrument is orthogonal to the error term. In particular, to obtain identification in equation (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. I will address how this discussion is altered if a different form of identification is used in a paragraph later in this section. In this case, higher levels of  $Y_{it-1}$  will generate higher values of  $\Delta \ln(1 - \tau_{it}^p)$  in the case of a tax decrease and lower values in the case of a tax increase. Therefore,  $cov[\Delta \ln(1 - \tau_t^p), \ln(\nu_{t-1})] > 0 \Rightarrow cov[\Delta \ln(1 - \tau_t^p), \Delta \ln(\nu_t)] < 0$  in the case of a tax decrease, and  $cov[\Delta \ln(1 - \tau_t^p), \ln(\nu_{t-1})] < 0 \Rightarrow cov[\Delta \ln(1 - \tau_t^p), \Delta \ln(\nu_t)] > 0$  in the case of a tax increase. Since the denominator of the ratio in equation (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,  $\mathbb{E}[\Delta \ln(1 - \tau_t^p)' \Delta \ln(\nu_t)] = 0$  only if  $\mathbb{E}[\ln(Y_{t-1})' \Delta \ln(\nu_t)] = 0$ .

The severity of the endogeneity problem when  $\mathbb{E}[\ln(Y_{t-1})' \Delta \ln(\nu_t)] \neq 0$  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.<sup>15</sup> When

---

<sup>15</sup>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.

there is no serial correlation, that is,  $K = 0$ ,

$$\begin{aligned}
\mathbb{E}[\ln(Y_{t-1})' \Delta \ln(\nu_t)] &= \mathbb{E}[\ln(Y_{t-1})'( \ln(\nu_t) - \ln(\nu_{t-1}))] \\
&= \mathbb{E}[\ln(Y_{t-1})'( \ln(\xi_t) - \ln(\xi_{t-1}))] \\
&= -\mathbb{E}[\ln(Y_{t-1})' \ln(\xi_{t-1})] = -\sigma_\xi^2 < 0.
\end{aligned} \tag{9}$$

The last line of (9) relies on the assumption that  $\ln(\xi_{t-1})$  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,  $cov(\ln(\mu_{t-1}), \ln(\xi_{t-1}))$ . But, the expression is guaranteed to remain negative unless  $cov(\ln(\mu_{t-1}), \ln(\xi_{t-1}))$  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,

$$\begin{aligned}
\mathbb{E}[\ln(Y_{t-1})' \Delta \ln(\nu_t)] &= \mathbb{E}[\ln(Y_{t-1})'( \ln(\nu_t) - \ln(\nu_{t-1}))] \\
&= \mathbb{E}[\ln(Y_{t-1})'(\phi_1 \ln(\nu_{t-1}) + \ln(\xi_t) - \ln(\nu_{t-1}))] \\
&= \mathbb{E}[\ln(Y_{t-1})'(\phi_1 - 1)\ln(\nu_{t-1})] = (\phi_1 - 1)\sigma_\nu^2 \\
&= \frac{(\phi_1 - 1)\sigma_\xi^2}{1 - \phi_1^2} = -\frac{\sigma_\xi^2}{1 + \phi_1} < 0.
\end{aligned} \tag{10}$$

The last line of (10) takes advantage of the fact that  $\sigma_\nu^2 = \frac{\sigma_\xi^2}{1 - \phi_1^2}$  when the process is AR(1) and covariance stationary. Note that equation (10) equals zero if  $\phi_1 = 1$ , but this implies that income follows a unit-root process, which generates an alternative set of issues to address. Unless the error term follows a unit root process,  $\ln(Y_{it-1})$  is correlated with the error term. A non-zero covariance is also found when higher orders of serial correlation are considered. The general notion that this instrument remains endogenous has been well-acknowledged in the literature.

If there is a tax reform for which the tax change is the same for all income levels, the endogeneity problem discussed above is eliminated, but it also eliminates identification because everyone experiences the same treatment, and there is no variation to exploit in order

to estimate the elasticity. Alternatively, if there is a tax reform that affected some individuals, but not others within a given income class, the endogeneity problem is mitigated.<sup>16</sup>

Given the remaining endogeneity of the instrument, researchers, beginning with Auten and Carroll (1999), have tried to solve the problem by including controls for  $\ln(Y_{it-1})$ . As Saez (2003, p.1250) observes, “...if  $\epsilon [\Delta \ln(\nu_{it})]$  depends on  $z_1 [Y_{it-1}]$ , the instrument, which is a function of  $\text{taxinc}_1 [Y_{it-1}]$ , is likely to be correlated with the error term  $\epsilon [\Delta \ln(\nu_{it})]$ . However by controlling for any smooth function of  $\text{taxinc}_1 [Y_{it-1}]$  in the regression setup in both stages, it is possible to get rid of the correlation between  $\epsilon [\Delta \ln(\nu_{it})]$  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),  $\ln(1 - \tau_t^p)$  is some constant value; so from that perspective, the endogeneity problem is solved—conditional on  $\ln(Y_{t-1})$ ,  $\Delta \ln(1 - \tau_t^p)$  no longer covaries with  $\Delta \ln(\nu_t)$ .

If  $\ln(Y_{t-1})$  were a valid proxy for the components of the error term that are correlated with  $\ln(Y_{t-1})$ , then it is a valid control. But this is never true. For example, when  $K = 1$ , one wants a perfect proxy for  $\ln(\nu_{t-1})$ , but  $\ln(Y_{t-1})$  includes two variables:  $\ln(\nu_t)$ —the variable we want to proxy for—and another variable,  $\ln(\mu_t)$ . It is well known that taking a variable that is a perfect proxy for  $\ln(\nu_t)$  and adding a non-constant variable that is not controlled for elsewhere in the regression will not fully absorb  $\ln(\nu_t)$  from the error term. Put formally, if  $\ln(Y_{t-1})$  is employed as a proxy,  $\ln(\mu_{t-1})$  will be contained in the error term and  $\text{cov}(\ln(Y_{t-1}), \ln(\mu_{t-1})) = \sigma_\mu^2 > 0$ .<sup>17</sup> 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 then 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 Section 4.3.

---

<sup>16</sup>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.

<sup>17</sup>The only exception to this statement occurs if high levels of permanent income were correlated with low shocks, but as I noted before, this is not likely the case.

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

$$\Delta \ln(Y_{it-1}) = \ln(\nu_{it-1}) + g_{t-1} - \ln(\nu_{it-2}), \quad (11)$$

which will remain endogenous because the latter two terms will be relegated to the error term and  $\text{cov}(\Delta \ln(Y_{t-1}), -\ln(\nu_{t-2})) = (1 - \phi_1)\sigma_\nu^2 > 0$  when  $K = 1$ . This covariance weakens as  $\phi_1 \rightarrow 1$ .<sup>18</sup> A similar result holds for all  $K \neq 1$ . As a last alternative, consider:

$$\begin{aligned} \Delta \ln(Y_{it-1}) - \Delta \ln(Y_{it-2}) &= (\ln(\nu_{it-1}) + g_{t-1} - \ln(\nu_{it-2})) \\ &\quad - (\ln(\nu_{it-2}) + g_{t-2} - \ln(\nu_{it-3})) \\ &= \ln(\nu_{it-1}) - \ln(\nu_{it-3}), \end{aligned} \quad (12)$$

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  $\text{cov}(\Delta \ln(Y_{t-1}) - \Delta \ln(Y_{t-2}), -\ln(\nu_{t-3})) = (1 - \phi_1^2)\sigma_\nu^2 > 0$ . A similar result holds for all  $K \neq 1$ . 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 orthogonal to the error term. 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.<sup>19</sup>

---

<sup>18</sup>Kopczuk (2003, 2005) suggests this proxy could be improved by including it as a 10-piece spline instead. But, this does not resolve the problem.

<sup>19</sup>This is not to say that these controls are not useful; it is just that they are useful in controlling for heterogeneous income trends, which will be addressed in Section 4.3, not as a proxy for transitory income shocks.



### 4.2.2 Alternative Instruments

Section 4.2.1 demonstrates that  $\Delta \ln(1 - \tau_{it}^p)$  is not exogenous as an instrument, regardless of the additional controls used. This section considers other possible instruments. Instruments that incorporate post-reform taxable income, such as those proposed in Carroll (1998) and Blomquist and Selin (2010), have also been considered in the literature; however, neither have had a significant impact because of the strong assumptions that are required for these instruments to be valid. Among other issues, over 15 percent of individuals in the data set used in this paper have taxable income changes in 1986 that were large enough that they were forced to change tax brackets between 1986 and 1987, regardless of where in the income distribution they were located (unless they were above the top tax bracket in both periods). Heterogeneous responses certainly play a role in determining the magnitude of the taxable income change, and thus, on the margin, whether individuals change tax brackets or not. It is well-known that when instruments are correlated with heterogeneous responses, the resulting estimates are biased (Murray, 2006). Thus, these instruments will not be further considered here.

The rest of this section considers a potential instrument that has not been previously considered in the literature.<sup>20</sup> 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  $\ln(Y_{it-1})$ . When the instrument is constructed from the appropriate lag, the instrument will be orthogonal to the error term. 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 (e.g., Anderson and Hsiao, 1982).

To understand when this approach resolves the endogeneity problem, suppose  $\ln(Y_{it-2})$  is used to instrument for the tax rate. Now, the relevant condition for the instrument to be orthogonal to the error term is  $\mathbb{E}[\ln(Y_{t-2})' \Delta \ln(\nu_t)] = 0$ . When  $K = 0$ , this condition can be

---

<sup>20</sup>Although this particular instrument has not previously been employed in the static panel literature that estimates the ETL, it has been employed by Holmlund and Soderstrom (2008) in a dynamic panel setting.

rewritten as:

$$\begin{aligned}\mathbb{E}[\ln(Y_{t-2})'\Delta\ln(\nu_t)] &= \mathbb{E}[\ln(Y_{t-2})'(ln(\xi_t) - ln(\xi_{t-1}))] \\ &= \mathbb{E}[\ln(Y_{t-2})'\ln(\xi_t)] - \mathbb{E}[\ln(Y_{t-2})'\ln(\xi_{t-1})] = 0.\end{aligned}\tag{13}$$

So, this instrument is clearly valid when  $K = 0$ . Now consider when  $K = 1$ :

$$\begin{aligned}\mathbb{E}[\ln(Y_{t-2})'\Delta\ln(\nu_t)] &= \mathbb{E}[\ln(Y_{t-2})'(ln(\nu_t) - ln(\nu_{t-1}))] \\ &= \mathbb{E}[\ln(Y_{t-2})'(ln(\xi_t) + \phi_1 ln(\nu_{t-1}) - ln(\nu_{t-1}))] \\ &= \mathbb{E}[\ln(Y_{t-2})'(ln(\xi_t) + (\phi_1 - 1)ln(\xi_{t-1}) \\ &\quad + \phi_1(\phi_1 - 1)ln(\nu_{t-2}))] \\ &= \phi_1(\phi_1 - 1)\mathbb{E}[\ln(Y_{t-2})'\ln(\nu_{t-2})] = \phi_1(\phi_1 - 1)\sigma_\nu^2,\end{aligned}\tag{14}$$

which does not equal zero unless there is a unit-root. However, because  $\mathbb{E}[\ln(Y_{t-l})\ln(\nu_{t-l})] = \sigma_\nu^2$  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  $|(\phi_1 - 1)|\sigma_\nu^2 > |\phi_1(\phi_1 - 1)|\sigma_\nu^2$ . Hence, the instrument is strictly more orthogonal to the error term 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 (14) 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  $\mathbb{E}[\ln(Y_{t-l})'\Delta\ln(\nu_t)] = 0$  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),  $\mathbb{E}[\ln(Y_{t-3})\Delta\ln(\nu_t)]$  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.<sup>21</sup> Hence, in Section 4.2.3, I will consider a test that will tell us

---

<sup>21</sup>Moffitt and Gottschalk (2002) find that the earnings structure of income in the U.S. is best approximated by an ARMA(1,1) process.

whether or not I can reject the hypothesis that  $\mathbb{E}[\ln(Y_{t-l})\Delta\ln(\nu_t)] = 0$ , and I will abstract away from whether the true underlying process is moving-average or serially correlated.

### 4.2.3 Empirical Results for Instrument Selection

Sections 4.2.1 and 4.2.2 have examined the choice of instrument theoretically. In this section, I empirically examine when I can reject the null hypothesis that the instrument is exogenous. I highlight the magnitude of the biases induced by incorrect methods, and then implement a method from Section 4.2.2, for which we can no longer reject the null hypothesis that the instrument is exogenous.

Table 2 provides empirical estimates of equation (5) for each of the proposed instrument and income control combinations considered above. Equation (5) can be interpreted as a continuous treatment difference-in-differences equation. This table estimates two-year differences. Other difference lengths will be considered in Table 4. A full set of marital status, dependent children, over age 65 or blind, state, and year indicator variables are also included in each regression. When income controls are included, I use a five-piece spline, rather than including the income control directly. This is standard in the literature and is used to make the control more flexible. Using higher-order splines has a very minimal effect on the results that follow.

I use a Difference-in-Sargan test to examine whether a given instrument is exogenous. This test is an over-identification test that compares the covariance of the instrument and residual for the instrument(s) we want to test relative to the same covariance for the instrument(s) that we have hypothesized are exogenous. This test has been found to have more power than the Sargan test, often also called a J-test, which computes the covariance of the instrument and residual for all instruments included (Arellano and Bond, 1991).<sup>22</sup>

---

<sup>22</sup>These tests will introduce a pretest bias (Guggenberger and Kumar, 2011), although this is mitigated by the fact that I will use these tests as evidence of which instruments are valid, but use my originally hypothesized choice of lags—two, three, and four—as my preferred estimates, even though the null hypothesis is marginally not rejected for one lag. Also, the inference reported is robust to inference based on the Anderson-Rubin test, which Guggenberger and Kumar (2011) show is less subject to these size distortions.

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. At the very least, the new estimates are substantially closer to obtaining consistency than any previously reported.<sup>23</sup> 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.<sup>24</sup> 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.<sup>25</sup> Column (5) will test whether income lagged two periods is, in fact, exogenous. Columns (1)-(5) are restricted so that the same individuals appear in each to enhance comparability across the columns.

Table 2 Columns (1)-(3) consider specifications in which the instrument is shown to be endogenous in Section 4.2.1, except under extreme assumptions. Column (1) uses the predicted tax rate instrument  $\Delta \ln(1 - \tau_t^p)$ . The estimate is 0.000 and the Difference-in-Sargan test p-value is 0.003; therefore, I can strongly reject that the instrument is orthogonal to the error term, as expected. Column (2) adds splines in log base-year income  $\ln(Y_{t-2})$ . The spline break points occur at \$25,000, \$50,000, \$75,000, and \$100,000.<sup>26</sup> While the estimate changes substantially—it increases by an order of magnitude to 0.279 and becomes statistically significant at the ten percent level<sup>27</sup>—the Difference-in-Sargan test p-value is even smaller (0.000). Column (3) again repeats Column (1), but adds the lagged value of the dependent variable as a spline  $\Delta \ln(Y_{t-1})$ . The estimate is 0.058 and is insignificant.

---

<sup>23</sup>The size of these tests is very good according to Arellano and Bond (1991).

<sup>24</sup>Each predicted tax rate instrument is constructed by running taxable income in the prediction year through TAXSIM for each outcome year of interest. For example, suppose I am constructing a two-year difference predicted tax rate instrument as a function of income lagged two periods. I take income in period  $t - 4$  and run it through TAXSIM for the years  $t - 2$  and  $t$ . I then use these tax rates to construct the difference.

<sup>25</sup>The Difference-in-Sargan and Sargan tests are equivalent unless there are at least at least two instruments that are assumed exogenous.

<sup>26</sup>A 17-piece spline with breakpoints every \$10,000 up to the top 1 percent of the income distribution in my sample yields an almost identical point estimate and the Difference-in-Sargan p-value remains 0.000.

<sup>27</sup>If I used the same covariates as Gruber and Saez (2002), the estimate in this column is 0.352 with a p-value of 0.020.

**Table 2: Instrument Selection<sup>1</sup>**

	(1)	(2)	(3)	(4)	(5)
$\Delta \ln(1 - \tau_t)$	0.000 (0.125)	0.279* (0.149)	0.038 (0.125)	0.681*** (0.252)	0.858*** (0.311)
1 <sup>st</sup> Spline <sup>2</sup>		-0.128*** (0.022)	-0.350*** (0.047)		
2 <sup>nd</sup> Spline		-0.074*** (0.023)	0.508*** (0.153)		
3 <sup>rd</sup> Spline		-0.103* (0.062)	-0.276 (0.239)		
4 <sup>th</sup> Spline		0.082 (0.135)	-0.164 (0.205)		
5 <sup>th</sup> Spline		-0.104 (0.064)	0.274*** (0.094)		
Instruments <sup>3</sup>	0,2,3 lags	0,2,3 lags	0,2,3 lags	1,2,3 lags	2,3,4 lags
Observations	24,576	24,576	24,576	24,576	24,576
Individuals	6,044	6,044	6,044	6,044	6,044
Diff-in-Sargan p-value	0.003	0.000	0.048	0.229	0.868
First Stage F-statistic	487.9	369.4	484.7	145.0	108.2

<sup>1</sup>Each column is estimated using 2SLS for two-year differences. Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. Indicator variables for marital status, number of dependents, whether over 65 or blind, and state and year fixed effects are also included in estimation.

<sup>2</sup>In Column (2), the splines are a function of log base-year income. The spline break points are \$25,000, \$50,000, \$75,000, and \$100,000. In Column (3), the splines are a function of  $\Delta \ln(Y_{t-1})$ . The spline break points are quintiles of  $\Delta \ln(Y_{t-1})$ .

<sup>3</sup>This row lists 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 the 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 is exogenous using the Difference-in-Sargan test.

The null hypothesis that the instrument is exogenous is rejected at the five percent level.<sup>28</sup>

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 not true in this context. The results also highlight that mean reversion is substantial in the U.S., because it would otherwise be difficult to reject the null hypothesis that these instruments are exogenous.

Table 2 Columns (4)-(5) consider instruments that are exogenous under certain, more reasonable assumptions that were proposed in Section 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 0.229, and therefore cannot be rejected at the 10 percent level. Column (5) assumes that an instrument lagged three and four periods is exogenous and tests whether we can reject

<sup>28</sup>I have also combined the controls proposed in Columns (2) and (3). The p-value on the test of for the Difference-in-Sargan test in this case is 0.007.

the null that two lags are exogenous. The p-value on this test is 0.868. The ETI estimate in Column (5) is 0.858 and is statistically significant at the one percent level. This is my preferred baseline estimate.<sup>29</sup> This estimate is more than twice as large as the Column (2) estimate that is based on the most frequently cited paper on this subject (Gruber and Saez, 2002).

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 accordingly. This is consistent with the theoretical analysis in Section 4.2.1, which showed that the estimates are biased downwards when endogenous instruments were used (for a tax decrease). As we move rightwards across these columns, the lags used as instruments increase; however the standard errors do not increase linearly with this change. Replacing the original instrument with an instrument lagged one period doubles the standard errors, while increasing the instrument by another lag only increases the standard errors by 20 percent. This suggests that much of the increase in standard errors from Column (1) to (4) reflects the fact that the original instrument is quite endogenous. All the first-stage F-statistics in Table 2 are above 100, indicating that weak instruments are not of concern here. Moreover, each tax rate instrument is individually significant at the one percent level, except in Column (2) as shown in appendix Table A.1.

### 4.3 Heterogeneous Growth Rates

The theoretical and empirical results have so far assumed that  $g_t$  is the same, on average, for individuals at all points in the income distribution. 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). However, a homogeneous growth rate may or may not be a legitimate assumption in practice depending on the context, and this fact has

---

<sup>29</sup>I prefer these estimates over Column (4) given that I am not far from rejecting the null hypothesis in Column (4) and the power of this test may be slightly weak as discussed above. Additionally, the fact that the estimates rise by about a standard deviation from Column (4) to (5) suggests that the instruments are less exogenous in Column (4).

been recognized by researchers in this area.<sup>30</sup>

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.<sup>31</sup> If it is ignored, the income-class-varying portion of the growth rate will end up in the error term, and is likely highly persistent. As a result, it will likely be correlated with lags of base-year income used to construct the tax rate instrument that are otherwise 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 rich demographic data are not usually available with tax return data.<sup>32</sup> Moreover, when estimating the ETI, perhaps even these are not enough to fully absorb heterogeneous income trends. The literature has addressed this challenge by including base-year income controls. While they are endogenous when included directly, base-year income controls can be instrumented using the same lags as are used to instrument for the tax rate variable or the lags can be included directly, replacing the endogenous control.<sup>33</sup> When such suitable lags are used, the income variables effectively control for permanent income plus an uncorrelated measurement error.<sup>34</sup> This measurement error is uncorrelated with permanent income by definition. Therefore, when instrumented, log income proxies for permanent income, which is what is relevant for determining the heterogeneity in growth

---

<sup>30</sup>Auten and Carroll (1999) were the first to relax this assumption.

<sup>31</sup>Another alternative considered in some papers, such as Singleton (2011), is to examine a tax change that affects only a narrow portion of the income distribution. If  $g_t$  is the same for both the comparison and control groups this will indeed resolve this problem.

<sup>32</sup>A few studies have had access to additional demographic controls (e.g., Carroll (1998), Auten and Carroll (1999), and Singleton (2011)) and report that controlling for these in addition to base-year income does not significantly change the ETI estimates.

<sup>33</sup>Another alternative is proposed in Gelber (2010). He constructs predicted spline segments from a regression of the change in taxable income on the spline segments in a year in which there were no tax changes. Then, he constructs a new dependent variable—the log change in taxable income minus the predicted spline segments. This method is valid as long as the spline segments are instrumented with lags for years in which the base-year is not sufficiently exogenous, so that the spline coefficients will be consistent.

<sup>34</sup>Its possible that some highly persistent components of transitory income will be treated as equivalent to permanent income in these regressions.

rates, and is a best-case scenario in terms of being a proxy for things such as age and education level. That said, it remains an imperfect proxy because it still includes a measure of both permanent and (lagged) transitory income.

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.<sup>35</sup> 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.<sup>36</sup> For this reason, including additional pairs of years aids in identifying these growth rates separately from the tax changes.

Table 3 examines the importance of heterogeneous income controls empirically. Column (1) repeats Column (5) of Table 2 as a benchmark, excluding those with negative constant-law taxable income lagged two, three, and four periods, so that all columns in this table will include the same observations. 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 in a previous year than others.<sup>37</sup> Empirically, this increases the estimate by about half a standard deviation.

Table 3 Column (2) adds a spline in log base-year taxable income, which, as noted in the previous paragraphs, is expected to be endogenous. I use a five-piece spline with break points at \$25,000, \$50,000, \$75,000, and \$100,000.<sup>38</sup> These estimates more than double relative to

---

<sup>35</sup>Gruber and Saez (2002) were the first to use income splines when estimating the ETI.

<sup>36</sup>Gruber and Saez (2002) test this assumption and find that it does not matter much in the 1979-1990 Michigan IRS Tax Panel data set.

<sup>37</sup>However, there is no reason to believe that this additional cutoff alters the endogeneity of the tax rate or income spline instruments, and it does not in practice.

<sup>38</sup>These cut points are more appropriate than quintiles of the distribution because the highest quintile in this data set is at \$48,917.96, which does not allow for effective control of heterogeneous growth trends at the top of the income distribution; that said, the results are robust to including a quintile spline instead.



**Table 3: Heterogeneous Income Trends<sup>1</sup>**

	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \ln(1 - \tau_t)$	1.070*** (0.326)	2.399*** (0.548)	1.128*** (0.438)	1.268*** (0.449)	1.244*** (0.449)	1.140*** (0.439)
1 <sup>st</sup> Spline <sup>2</sup>		-0.175*** (0.031)	-0.147*** (0.028)	-0.020 (0.015)		-0.070 (0.071)
2 <sup>nd</sup> Spline		-0.160*** (0.035)	-0.060** (0.031)	-0.063** (0.028)		-0.056 (0.062)
3 <sup>rd</sup> Spline		-0.267*** (0.092)	-0.100 (0.103)	-0.023 (0.092)	17- Piece Spline	-0.037 (0.193)
4 <sup>th</sup> Spline		-0.122 (0.165)	0.234 (0.202)	0.193 (0.182)		0.289 (0.388)
5 <sup>th</sup> Spline		-0.126* (0.070)	-0.030 (0.069)	-0.009 (0.064)		-0.035 (0.097)
Spline lags included <sup>3</sup>	-	0	0,2,3,4	2,3,4	2,3,4	0(2,3,4)
Observations	22,729	22,729	22,729	22,729	22,729	22,729
Individuals	5,520	5,520	5,520	5,520	5,520	5,520
Diff-in-Sargan p-value <sup>4</sup>	0.826	0.850	0.894	0.904	0.895	0.802
First Stage F-statistic	101.5	53.06	64.54	58.74	59.56	10.99

<sup>1</sup>Each column is estimated using 2SLS for two-year differences. The predicted net-of-tax rate instruments are always constructed based on income lagged two-, three-, and four periods. Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. Indicator variables for marital status, number of dependents, whether over 65 or blind, and state and year fixed effects are also included in estimation.

<sup>2</sup>The spline break points are \$25,000, \$50,000, \$75,000, and \$100,000. In Columns (2)-(4), the spline coefficients are the sum of the coefficients of all spline lags included.

<sup>3</sup>This row lists the spline lags included in each column. For example, Column (3) lists the splines as no lag, two lags, three lags, and four lags. The splines are all included directly, except when they are listed inside parentheses. The splines inside parentheses are used as instruments for the spline outside parentheses.

<sup>4</sup>These are p-values from testing whether all instruments lagged two periods are exogenous.

the estimates in Column (1).

To highlight that the increase in the Column (2) estimate is due to the use of an inappropriate control, rather than a relevant (albeit slightly endogenous) control that is subsequently dropped, I keep this spline in Column (3) and add splines of income lagged two, three, and four periods. The spline coefficients in this column are the sum of the coefficients for a given spline piece across all splines included. The estimates in Column (3) remain significant at the one percent level, but are now indistinguishable from the estimates in Column (1), which included no income splines. Column (4) drops the base-year income spline, which has a minimal effect on the estimates.

The results suggest that controlling for heterogeneous income trends is not important in this data set. One could be concerned that this is because I did not include enough spline segments. In Column (5) of Table 3, I replace the five-piece spline with a 17-piece spline.

The 17-piece spline has break points every \$10,000 up to the top one percent of the income distribution observed in this data set (about \$170,000). The results show that the estimates are highly robust to the number of splines included—the coefficient in Column (4) is 1.268 and the coefficient in Column (5) is 1.244 and both are significant at the one percent level.<sup>39</sup> Section 4.5 will explore the finding that heterogeneous income trends are not important for this data set in more detail.

An alternative to including lagged income splines directly is to use the lagged income splines as an instrument for the base-year income spline. Both methods should generally yield similar estimates, although including the spline pieces directly is a more flexible specification. The advantage of considering a specification that uses lagged splines as instruments is that it allows me to test the relevance and exogeneity of the income spline lags. Column (6) reports this specification. The ETI estimate remains unchanged. The first stage F-statistic is 10.99, which is well above the weak instrument threshold for a regression that includes six variables that require an instrument. The tax rate variable and spline pieces also individually pass an Angrist-Pischke first stage F-test. In this regression, the Difference-in-Sargan test examines whether all instruments (predicted tax rate and income spline pieces) lagged two periods are exogenous. The p-value on this test is 0.802, which is far from being able to reject the null hypothesis that the instruments are exogenous.

## 4.4 Different Difference Lengths

In this section, I estimate and compare ETI estimates based on one-, two-, and three-year differences. These estimates are reported in Table 4. Columns (1) and (2) estimate the ETI based on one-year differences without and with heterogeneous income trends, respectively. Columns (3) and (4) estimate the same specifications for two-year differences (which are repeated from the previous table). Columns (5) and (6) do the same for three-year differences. Column (7) estimates the three-year difference ETI with a 17-piece instead of a five-piece

---

<sup>39</sup>This is consistent with what others have found. For example, Gruber and Saez (2002) use a 10-piece spline, but note that their results are similar when they use a 20-piece spline.

income spline. The estimates with heterogeneous income trends are all significant and slightly increase as the difference window increases (0.973 to 1.268 to 1.361), although the estimates are far from being significantly different. Comparing the ETI estimates in Columns (6) and (7) highlights the increased value of switching to a higher-order spline in three-year differences. However, while this decrease of 0.1 (from 1.361 to 1.265) is perhaps economically significant, it is far from statistically significant.

Up to this point, I have used Difference-in-Sargan tests to determine whether the instrument is exogenous or endogenous. A disadvantage of this test is that it is possible to reject the null because two very different local average treatment effects (LATE) are identified with the exogenous versus potentially endogenous instruments, rather than a true endogeneity problem. For one-year differences, there is another test that can be used which is not an over identification test—the Arellano-Bond test. This test was also used in Arellano and Bond (1991) and it tests whether there is serial correlation in the error term; that is, whether there is serial correlation in first differences. First-order serial correlation will naturally exist in first differences unless there is a unit root. The standard instrument used in Columns (1), (2), and (3) of Table 2 are only valid if control variables can eliminate this serial correlation.<sup>40</sup> In order to perform this test, I use the baseline one-year difference specification found in Table 4, Column (1). I find there is strong first-order serial correlation (the p-value is 0.0000). When I add one or both of the controls that have been proposed in the literature (and were previously considered in Table 2 Columns (2) and (3)), the z-statistics do decrease, but the p-values remain 0.0000. Whenever the lagged dependent variable is included as a control, I can reject the null at the five (and often one) percent level that there is no second- and third-order serial correlation as well, which is highly consistent with the results in Section 4.2.1.

The literature has interpreted the overall similarity of estimates across different difference

---

<sup>40</sup>One caveat is that this test is designed for a much longer panel than the panel used in this paper. However, given the strength of these results (and their consistency with what was found using the Difference-in-Sargan test), it seems unlikely they are qualitatively different in a longer panel.

**Table 4: Different Difference Lengths<sup>1</sup>**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \ln(1 - \tau_t)$	0.996*** (0.380)	0.973** (0.459)	1.070*** (0.326)	1.268*** (0.449)	0.880*** (0.330)	1.361** (0.624)	1.265** (0.613)	1.593*** (0.504)
Difference Length	1 year	1 year	2 years	2 years	3 years	3 years	3 years	2 years
Spline Included <sup>2</sup>	-	5-piece	-	5-piece	-	5-piece	17-piece	5-piece
Base-Years Excluded	-	-	-	-	-	-	-	85, 87
Observations	28,877	28,877	22,729	22,729	18,160	18,160	18,160	15,129
Individuals	6,250	6,250	5,520	5,520	5,172	5,172	5,172	5,367
Diff-in-Sargan p-value <sup>3</sup>	0.334	0.331	0.826	0.904	0.538	0.691	0.707	0.648
First Stage F-statistic	80.82	58.29	101.5	58.74	100.8	33.12	33.79	40.10

<sup>1</sup>Each column is estimated using 2SLS. The predicted net-of-tax rate instruments and spline lags are always constructed based on income lagged two-, three-, and four periods. Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. Indicator variables for marital status, number of dependents, whether over 65 or blind, and state and year fixed effects.

<sup>2</sup>Splines, when included, are lagged two, three, and four periods.

<sup>3</sup>These are p-values from testing whether all instruments lagged two periods are exogenous.

lengths as evidence that the short-run and long-run responses are similar (e.g. Gruber and Saez, 2002). This is probably partially true; however, this explanation overlooks the fact that none of these estimates are identifying the parameter they are ostensibly measuring. The problem lies in the nature of overlapping differences. Suppose, as for TRA86, the tax rate changes are phased-in over two years—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 1986 tax changes, and two years of the 1987 tax change response. The 1984-1987 difference only captures a one-year response to the 1987 tax changes. The estimate is thus 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 difference 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 though there is no tax change this period. For the TRA86 phase-in, most individuals experienced larger tax rate changes in 1987 than 1988. To the extent that the responses in these two years 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.

Estimating a two-year response excluding the 1985-1987 and 1987-1989 differences will potentially identify a two-year response to TRA86. This coefficient is reported in Table 4 Column (8); it increases by about two-thirds of a standard deviation, which may be reasonable given that this parameter is now no longer partly estimating a one-year effect, which is slightly smaller. However, there is an important concern. When this method is used, there is no control for the taxable income changes in the years just on either side of the tax reform, which matters if they are substantially different from those in other years, on average. For this reason, it is an interesting robustness check, but not my preferred method.

## 4.5 Understanding the Identifying Variation

Section 4.2.3 established an instrument set for which we could not reject the null hypothesis that the instruments are valid, Section 4.3 showed that heterogeneous income trends are not important in this context, and Section 4.4 highlighted that these results are similar across one-, two-, and three-year differences. This section explores the new instruments used in these sections—lagged predicted tax rate changes—in much more depth, using both estimation and graphical analysis to highlight the identifying variation and provide some sensitivity analysis.

Table 5 Column (1) reproduces the baseline specification found in Table 2 Column (5).<sup>41</sup> Columns (2), (3), and (4) repeat this specification including only the instrument lagged two, three, or four periods, respectively. The estimates for the two and three lag instruments are almost identical and are significant at the one percent level. The ETI estimate for the four

---

<sup>41</sup>I do not use heterogeneous income trends in this section because they were shown not to matter previously, and I want the relationship between these estimates and graphical evidence that I will present shortly to be as tight as possible. All estimates presented in this section are robust to the inclusion of heterogeneous income trends.

**Table 5: Understanding the Identifying Variation<sup>1</sup>**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\Delta \ln(1 - \tau_t)$	0.858*** (0.311)	0.909** (0.361)	0.907** (0.450)	0.735* (0.402)	0.971** (0.391)	0.917* (0.493)	1.104* (0.609)
Instrument lags <sup>2</sup>	2,3,4	2	3	4	2	2	2
Two-Lag Log Income Cutoff	-	-	-	-	<12.5	<11.5	$\geq 11.5$
Observations	24,576	24,576	24,576	24,576	23,978	23,397	1,179
Individuals	6,044	6,044	6,044	6,044	5,868	5,825	756
First Stage F-statistic	108.2	229.1	161.0	161.0	212.2	139.7	86.25

<sup>1</sup>Each column is estimated using 2SLS for two-year differences. Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. Indicator variables for marital status, number of dependents, whether over 65 or blind, and state and year fixed effects are also included in estimation.  
<sup>2</sup>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.

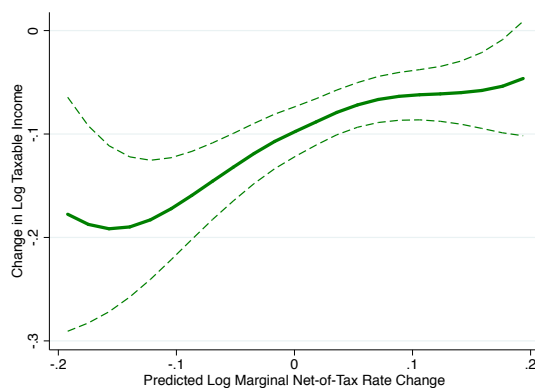
lag instrument falls by about half a standard deviation, but remains statistically significant. The fact that the four lag predicted tax rate instrument still identifies a significant ETI that is almost the same as for fewer lags is rather impressive and reflects the large permanent component of income over time. The consistency of the estimates across columns suggests that (i) each of the instruments are identifying roughly the same LATE, and (ii) all three of these instruments are equally exogenous.

This section proceeds to graphically analyze the predicted tax rate instrument lagged two periods. The estimate in Column (2) increases by about 0.1 (one-third of a standard deviation) when no covariates besides year fixed effects are included, suggesting that the graphical analysis will not be affected by ignoring covariates. Figures 1 through 4 plot fourth-order local polynomial regressions. This polynomial most accurately captures the shape of the line, which could only be captured with lower-order polynomials if a small (and thus very noisy) bandwidth were used.

Figure 1 provides a graphical, semi-parametric representation of the relationship between the outcome and instrument for the specification estimated in Table 5 Column (2). This figure plots a local polynomial regression of the change in log taxable income ( $\Delta \ln(Y_t)$ ) on the predicted log marginal net-of-tax rate change lagged two periods before the base-year ( $\Delta \ln(1 - \tau_t^{p,2lags})$ ). The thin dashed-lines represent a 95 percent confidence interval. Only the two-year difference from 1986 to 1988, which captures the response to TRA86, is included

in this figure. Most tax changes in the sample take place during this two-year window, and focusing on a single difference allows me to make valid comparisons across the distribution of  $\Delta \ln(1 - \tau_t^{p,2lags})$ .<sup>42</sup> For the same reason, Figures 2, 3, and 4 are also restricted to this two-year difference. Figure 1 depicts the local polynomial regression for  $\Delta \ln(1 - \tau_t^{p,2lags}) \in [-.2, .2]$ , which includes 97 percent of the sample.<sup>43</sup> There is a steady increase in the average taxable income change as the predicted marginal net-of-tax rate change increases.

**Figure 1: Mean Taxable Income Change by Predicted Tax Change Level (1986)**



This is a fourth-order local polynomial regression plot of the mean 2-year change (from 1986 to 1988) in log taxable income as the predicted 2-year marginal net-of-tax rate lagged two periods rises over the interval  $[-.2, .2]$ . 97 percent of the data lies within this interval. The thin dashed lines are 95 percent confidence intervals. The bandwidth is 0.16.

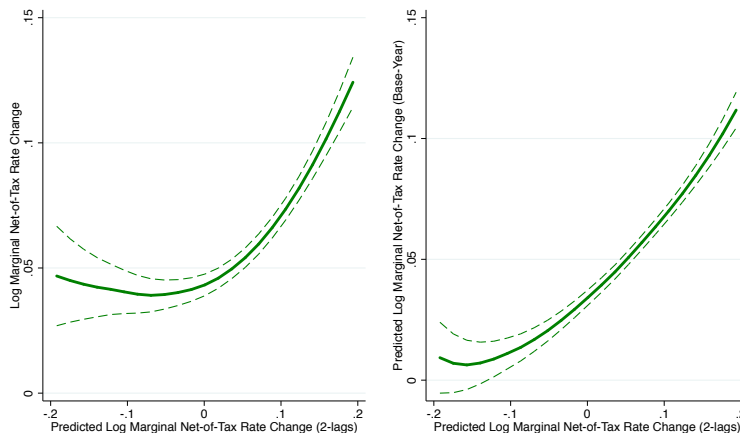
The left panel of Figure 2 provides a graphical representation of the first-stage of the estimates in Table 5 Column (2). For positive values of  $\Delta \ln(1 - \tau_t^{p,2lags})$  (predicted tax decreases), the instrument appears to be performing quite well; however, the figure raises concerns that the instrument is weak for negative values. Moreover, the instrument seems to be predicting outcomes well in this region (see Figure 1), which suggests the exclusion

<sup>42</sup>For example, if I include the difference 1984-1986, there would be a mass point at  $\Delta \ln(1 - \tau_t^{p,2lags}) = 0$ . We could now no longer compare the response of those at zero to those that faced significant tax changes because the change in log taxable income does vary some, on average, across years.

<sup>43</sup>Extreme observations are excluded because the standard errors on these observations are large and make the rest of the graph difficult to read. Restricting the specification in Table 5 to include only individuals  $\Delta \ln(1 - \tau_t^{p,2lags}) \in [-.2, .2]$  decreases the ETI estimate by about 2/5 of a standard deviation; further restricting the specification to include only  $\Delta \ln(1 - \tau_t^{p,2lags}) \in [-.1, .1]$  increases the ETI estimate by about 2/5 of a standard deviation. The estimates remain statistically significant.

restriction is violated. To shed light on this issue, the right panel provides an alternative local polynomial regression, replacing the y-axis with the tax rate change individuals face if their income remained constant at their base-year income level ( $\Delta \ln(1 - \tau_t^p)$ ). This measure of treatment is highly correlated with the instrument across the entire distribution—both positive and negative—which provides a reason that the instrument and actual tax rate changes are weakly correlated for tax rate increases. On average, these individuals must have crossed tax brackets after the tax reform (either on purpose in response to their tax rate change or because of transitory income shocks or income growth) more often than individuals with positive instrument values, and this makes the actual tax rate change less informative for them. For an extensive discussion of these issues, see Weber (2013). As a sensitivity check, I estimate the specification in Table 5, Column (2) replacing the actual tax rate change with  $\Delta \ln(1 - \tau_t^p)$ . The ETI estimate from this regression is 0.77 and the standard errors decrease by 23 percent relative to the estimate in Table 5, Column (2).

**Figure 2: Mean Tax Rate Change by Predicted Tax Rate Change Level (1986)**



The left panel is a fourth-order local polynomial regression plot of the mean 2-year change (from 1986 to 1988) in the marginal net-of-tax rate as the predicted marginal net-of-tax rate lagged two periods increases over the interval  $[-.2, .2]$ . The thin dashed lines are 95 percent confidence intervals. The right panel replaces the x-axis with an alternative measure of treatment—the base-year predicted tax rate change. The bandwidth is 0.16.

Figure 3 examines heterogeneous income growth rates across the income distribution for the two-year difference 1986 to 1988 ( $\ln(Y_{88}) - \ln(Y_{86})$ ) when there was a tax reform



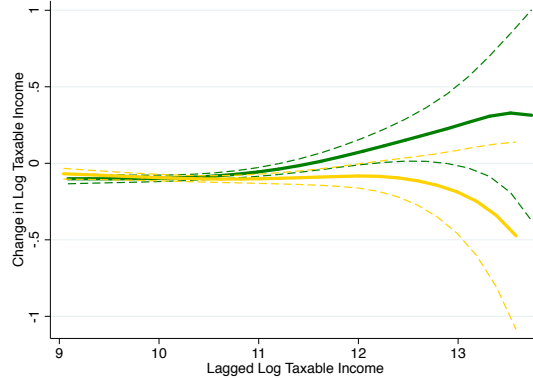
and from 1984 to 1986 ( $\ln(Y_{86}) - \ln(Y_{84})$ ) when there was no tax reform. The figure plots a local polynomial regression of  $\Delta \ln(Y_t)$  on taxable income lagged two periods before the base-year.<sup>44</sup> It is clear from the 1986 to 1988 plot that income growth increases as taxable income rises. This pattern could be due either to the fact that the largest tax rate decreases occurred at the top of the income distribution or because individuals at the top of the income distribution were experiencing high growth rates in all years. The 1984 to 1986 plot provides no evidence of the second explanation; the changes in taxable income in this plot are extremely flat across the taxable income distribution, and actually decline slightly at the top, although this decrease is not statistically significant. The data suggests that the decline is not due to anticipatory responses of high income individuals waiting to realize income until 1987 because, looking at the two years separately, the response is flat for both years, and the decline at the top is stronger in 1984-1985 than 1985-1986. Although not pictured here, the change from 1988-1990 is similarly flat, except there is an insignificant uptick in growth above \$450,000. This picture reinforces the results in Section 4.3—there appears to be no significant heterogeneous income trend across taxable income levels in the data. While highly suggestive, this analysis cannot rule out other potential stories, such as the 1988-1990 figure is flat because there is a positive heterogeneous income trend in these years, but this is offset by individuals with large tax rate decreases (i.e. high income individuals) shifting income into 1988 and shifting deductions out of 1988 into earlier years because 1988 is the first year the tax changes of TRA86 are fully phased in. Moreover, both heterogeneous income trends and anticipatory responses could be an important factor in other contexts, and should be carefully examined before being dismissed.

Figure 4 plots the outcome— $\Delta \ln(Y_t)$ —and the treatment as measured by the instrument— $\Delta \ln(1 - \tau_t^{p,2lags})$ —across the distribution of two-year lagged taxable income. Together, they present my main regression specification graphically across two-year lagged taxable income

---

<sup>44</sup>One could replace the x-axis with base-year log taxable income. The results are very similar to those found here, except at the ends which remain fairly endogenous because of substantial mean reversion on average at the top and bottom of the income distribution.

**Figure 3: Mean Taxable Income Change by Taxable Income Level**

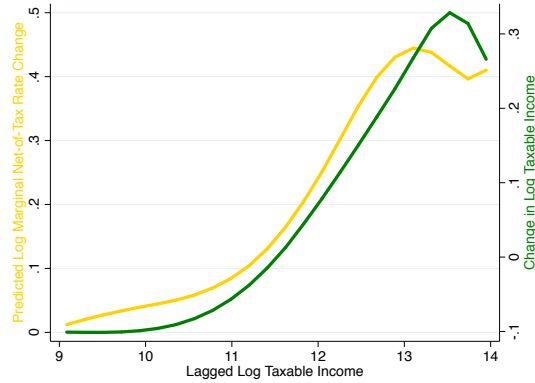


This figure is a fourth-order local polynomial regression plot of the mean two-year change in taxable income after TRA86 (green line, 1986 to 1988) relative to before the tax reform (yellow line, 1984 to 1986) as two-year lagged taxable income increases. The thin dashed lines are 95 percent confidence intervals. The bandwidths are 2.71 (green line) and 3.06 (gold line). The figure is trimmed to exclude lagged log taxable income below \$10,000.

levels for the two-year change from 1986 to 1988 (capturing TRA86). If there is a homogeneous treatment effect across the income distribution, these two lines should move in lock-step. The two lines move together until the very top of the income distribution around when log taxable income is 12.5, which is equivalent to \$268,337. Above this point, it appears that the ETI is more elastic than elsewhere in the distribution because the change in taxable income is now increasing more rapidly than the increase in the net-of-tax rate change. This is consistent with evidence in Saez (2004), which found that the top one percent are most responsive to tax changes.

Saez et al. (2012) show that when those with the highest tax rate change are more elastic than the rest of the distribution, this will bias estimates upward. To consider the degree to which the higher elasticity of individuals at the top of the income distribution biases the estimates, Table 5 Column (5) estimates the same specification as Column (2), but excludes individuals with two-year lagged taxable income above \$268,337. This decrease the sample size about about 2.4 percent and leaves the estimates almost unchanged. This is not surprising for three reasons: (i) this is a small portion of the data and the change in the elasticity is not statistically significant, (ii) the rest of the income distribution responds homogeneously, on average, and (iii) it occurs at a point when the change in the net-of-tax

**Figure 4: Mean Taxable Income Change and Mean Tax Rate Change by Taxable Income Level (1986)**



This figure presents two fourth-order local polynomial regression plots. The green line represents the average outcome—a two-year log change in taxable income—across taxable income levels (repeated from Figure 3) and the gold line represents the average instrument—a two-year change in the predicted log marginal net-of-tax rate—across taxable income levels. The confidence intervals are suppressed because the confidence intervals on the green line are quite wide at the top. The two lines are never significantly different. The bandwidths are 1.74 (gold line) and 2.71 (green line).

rate is leveling out, not increasing dramatically. Regarding (i) note that this data set does not oversample high-income individuals. This issue is much more relevant in a data set where the top one percent is oversampled. Regarding (ii), one can essentially read the elasticity that will be estimated, ignoring individuals above \$268,337, directly from the Figure 4. From the bottom to the top of the income distribution, the change in log taxable income increases by about 0.25 and the net-of-tax rate change increases by about 0.35, yielding an elasticity of 0.71.

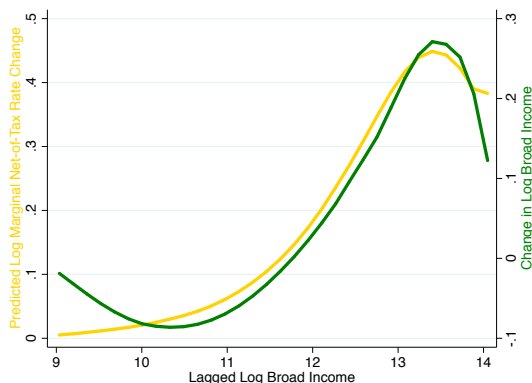
To further reinforce the homogeneity of response across the income distribution numerically, Columns (6) and (7) of Table 5 estimates the ETI separately for individuals with two-year lagged log taxable income below 11.5 (\$98,716) and those above. These two estimates are very similar—0.917 for individuals below \$98,716 and 1.104 for individuals above—and are both statistically significant at the ten percent level.

## 4.6 Broad Income and Weighted Estimates

This section covers two additional important topics for welfare analysis—the elasticity of broad income and weighted elasticity estimates. Broad income is constant-law gross income, excluding capital gains, minus constant-law above-the line deductions; in practice, there are very few above-the-line deductions included because most changed over these years. The most important difference between broad and taxable income is that itemized deductions are not included in broad income. To the extent that changing the amount itemized is not costly, broad income, rather than taxable income is the relevant parameter for welfare analysis (Chetty, 2009). Moreover, we think some itemized deductions, such as charitable giving, create positive externalities. If charitable giving is currently under or optimally supplied in response to the current tax rate, a response along this margin to a tax rate decrease increases deadweight loss, rather than decreasing it. Lastly, there could be a larger response to anticipated future changes among itemized deductions; although, at least for charitable giving, there is only such a response amongst individuals making over \$200,000 per year (Bakija and Heim, 2011), which comprise a very small fraction of my data. Column (1) of Table 6 estimates the elasticity of broad income using the same restrictions as were used in estimating the elasticity of taxable income. The estimate is small—0.277—and insignificant. However, plotting a local polynomial regression of this specification across lagged broad income levels (see Figure 5) shows that there is a different income trend for individuals with lagged log broad income levels below 10.5 (\$36,316). I address this issue in two alternative ways. In Column (2), I restrict the sample to include all individuals with lagged broad income greater than \$36,316. In Column (3), I include a five-piece spline of lagged log broad income. The approaches yield similar estimates and both estimates are significant at the ten percent level. The elasticity of broad income estimate in Column (2) is 0.475 and the p-value is 0.076. I cannot reject the null hypothesis that the tax rate instrument lagged two periods is exogenous—the p-value on the Difference-in-Sargan test is 0.855. This estimate is comparable to the baseline ETI estimate found in Table 2 Column (5). The ETI estimate is

almost twice the size of the elasticity of broad income estimate, suggesting that a substantial portion of the response to TRA86 occurred via changes in itemizations.

**Figure 5: Mean Broad Income Change and Mean Tax Rate Change by Broad Income Level (1986)**



This figure presents two fourth-order local polynomial regression plots. The green line represents the average outcome—a two-year log change in broad income—across broad income levels, and the gold line represents the average instrument—a two-year change in the predicted log marginal net-of-tax rate—across taxable income levels. The confidence intervals are suppressed because the confidence intervals on the green line are quite wide at the top. The two lines are never significantly different. The bandwidths are 1.74 (gold line) and 2.71 (green line).

Up to this point, I have not weighted the estimates by income. Because the weighted ETI and weighted elasticity of broad income are the relevant parameters for welfare analysis, it is valuable to consider the sensitivity of the estimates to income weighting. However, weighting estimates using an endogenous variable—base-year income—is not valid. For a given income two periods ago (and thus instrument value), this weighting scheme places more (less) weight on individuals that face positive (negative) shocks today and likely a fall (rise) back to the mean tomorrow. This biases the ETI estimate towards zero, which is evidenced by the results from the specification that uses this method to estimate the ETI in Column (5). The estimate is 0.856 and the standard error is extremely large—0.582—which is a 25 percent increase over the standard errors for the same unweighted estimate. Column (6) weights the estimates by income lagged two periods, which is still a very good measure of today’s income—the coefficient on a regression of base-year income on income lagged two periods is 0.898 (p-value: 0.000)—but is not endogenous. The estimated ETI in Column (6) is almost identical to the same unweighted estimate (Table 3 Column (6)). This is not surprising given

**Table 6: Broad Income and Weighted Estimates<sup>1</sup>**

	Broad Income (BI)				Weighted Taxable Income (TI)		Weighted BI
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\Delta \ln(1 - \tau_t)$	0.277 (0.217)	0.475* (0.268)	0.510* (0.295)	0.677** (0.314)	0.856 (0.582)	1.195*** (0.438)	0.699** (0.321)
Spline lags included <sup>2</sup>	-	-	BI: 2,3,4	TI: 2,3,4	TI: 2,3,4	TI: 2,3,4	BI: 2,3,4
Two-Lag Log BI Cutoff	-	>10.5	-	-	-	-	-
Income Weights?	No	No	No	No	Base-Year	2-lags	2-lags
Observations	25,097	14,911	24,765	23,175	22,729	22,729	24,341
Individuals	6,157	3,741	6,065	5,628	5,520	5,520	5,949
Diff-in-Sargan p-value	0.510	0.855	0.630	0.608	0.505	0.530	0.830
First Stage F-statistic	94.41	61.97	57.42	50.39	39.31	36.91	34.42

<sup>1</sup>Each column is estimated using 2SLS for two-year differences. The predicted net-of-tax rate instruments are always constructed based on income lagged two-, three-, and four periods. Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. Indicator variables for marital status, number of dependents, whether over 65 or blind, and state and year fixed effects are also included in estimation.

<sup>2</sup>This is a list of the spline lags included in each column. For example, Column (3) includes income splines lagged two, three, and four periods before the base-year. In Columns (3) and (7), the splines are constructed using broad income. In Columns (4), (5), and (6), the splines are constructed using taxable income. The spline break points are \$25,000, \$50,000, \$75,000, and \$100,000. The spline coefficients are the sum of the coefficients of all spline lags included.

the graphical analysis in Section 4.5. Column (7) estimates a weighted elasticity of broad income, by repeating the specification in Column (3), but adding taxable income weights lagged two periods. The estimates are about half a standard deviation larger and significant at the five percent level.

## 4.7 Robustness Checks

This section provides additional sensitivity analysis. Table 7 presents these results. Column (1) repeats the baseline two-year ETI estimates in Table 2 Column (5) as a reference. All the other columns impose a restriction on this specification. Columns (2) and (3) consider alternative base-year taxable income cutoffs at the bottom of the income distribution—\$20,000 and \$30,000. The estimates increase by about 0.1 each time the income cutoff is increased, but the estimates are otherwise unchanged.

Individuals whose income is over \$200,000 in all years they are observed in the data do not have a state code, and are thus assigned no state taxes. Up to this point, they have been included, along with an indicator variable for missing state income tax rates.

**Table 7: Sensitivity Analysis<sup>1</sup>**

	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \ln(1 - \tau_t)$	0.858*** (0.311)	0.963*** (0.359)	1.082*** (0.394)	0.872*** (0.310)	0.804* (0.432)	0.919*** (0.311)
Income Cutoff	\$10,000	\$20,000	\$30,000	\$10,000	\$10,000	\$10,000
Include if Missing State	Yes	Yes	Yes	No	Yes	Yes
Include if Ever Schedule E	Yes	Yes	Yes	Yes	No	Yes
Include all Covariates	Yes	Yes	Yes	Yes	Yes	No
Observations	24,576	17,619	11,772	24,548	16,525	24,576
Individuals	6,044	4,493	3,217	6,037	4,161	6,044
Diff-in-Sargan p-value	0.868	0.545	0.993	0.822	0.925	0.772
First Stage F-statistic	108.2	74.66	63.23	110.0	55.91	108.6

<sup>1</sup>Each column is estimated using 2SLS for two-year differences. The predicted net-of-tax rate instruments are always constructed based on income lagged two-, three-, and four periods. Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. Indicator variables for marital status, number of dependents, whether over 65 or blind, and state and year fixed effects are also included in estimation, except in Column (6) which includes only marital status indicators and year fixed effects.

Table 7 Column (4) drops these 28 individuals from the sample; the ETI estimate is almost unchanged. Another concern that has been highlighted in other research on TRA86 is that there was substantial income shifting (Slemrod, 1996). In particular, individuals changed from C-corporation status to S-corporation status, which caused a large increase in individual income, without any real change in economic behavior taking place.<sup>45</sup> It is concerning if this were contributing to the results. Column (5) addresses this concern by dropping all individuals who had any Schedule E income in any year that they are observed in the data. This includes both S-corporation and partnership income. The standard errors increase due to the decrease in sample size, but the ETI estimate is almost unchanged—0.804 compared to the baseline estimate of 0.858—and remains significant at the 10 percent level. In Gruber and Saez (2002), the only covariates included in the analysis were marital status indicators and year fixed-effects. Column (6) repeats Column (1) including only these covariates. The estimates increase by 0.1, revealing that the specification is relatively insensitive to the addition of these covariates.

<sup>45</sup>Capital gains realizations were another major form of income shifting discussed in Slemrod (1996), but capital gains are excluded from constant-law taxable income, so this form of income shifting is not a concern here.

## 5 Conclusion

As aptly summarized by Saez et al. (2012), in their recent *Journal of Economic Literature* 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. The baseline ETI estimate obtained in this paper is 0.858 and the baseline elasticity of broad income is 0.475. The difference between these two estimates suggests that about half of the ETI is due to changes in itemizations in response to the tax reform.

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 (Saez et al., 2009). The theoretical results in this paper are equally applicable to these other countries. Given the extreme nature of the assumptions needed 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 public availability. 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 are likely not completely overcome in this paper. This paper is also



silent about extensive margin responses and those that enter or exit mid-year are potentially in the sample. Several important extensions that have been previously considered in the literature, but are beyond the scope of this paper include estimating the response to the tax base separately (Kopczuk, 2005) and estimating the compensated elasticity (Gruber and Saez, 2002). Identifying the response to past and future tax changes is an important avenue for future research. More should also be learned about the long-run response to tax rate changes in the form of investment decisions and human capital accumulation.

## References

- Anderson, T. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18:47–82.
- Arellano, M. and Bond, S. R. (1991). Some specification tests for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, 58:277–298.
- Auten, G. and Carroll, R. (1999). The effect of income taxes on household behavior. *Review of Economics and Statistics*, 81(4):681–693.
- Bakija, J. and Heim, B. (2011). How does charitable giving respond to incentives and income? New estimates from panel data. *National Tax Journal*, 64(2, part 2):615–650.
- Blomquist, S. and Selin, H. (2010). Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates. *Journal of Public Economics*, 94:878–889.
- Burns, S. K. and Ziliak, J. P. (2013). Identifying the elasticity of taxable income. University of Kentucky working paper.
- Carroll, R. (1998). Do taxpayers really respond to changes in tax rates? Evidence from the 1993 act. Office of Tax Analysis Working Paper 78, U.S. Department of Treasury, Washington D.C.
- Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics*, 1:451–488.
- Feenberg, D. R. and Coutts, E. (1993). An introduction to the TAXSIM model. *Journal of Policy Analysis and Management*, 12(1):189–194.
- Feldstein, M. S. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics*, 81(4):674–680.

- Giertz, S. (2008). Panel data techniques and the elasticity of taxable income. CBO Working Paper 2008-20, Congressional Budget Office, Washington D.C.
- Giertz, S. (2010). The elasticity of taxable income during the 1990s: New estimates and sensitivity analyses. *Southern Economic Journal*, 77(2):406–433.
- Gruber, J. and Saez, E. (2002). The elasticity of taxable income: Evidence and implications. *Journal of Public Economics*, 84(1):1–32.
- Guggenberger, P. and Kumar, G. (2011). On the size distortions of tests after an over identifying restrictions pretest. *Journal of Applied Econometrics*, forthcoming.
- Heim, B. T. (2006). The elasticity of taxable income: Evidence from a new panel of tax returns. Working Paper 2008:28, Institute for Labour Market Policy Evaluation (IFAU).
- Holmlund, B. and Soderstrom, M. (2008). Estimating dynamic income responses to tax reforms: Swedish evidence. Working Paper 2008:28, Institute for Labour Market Policy Evaluation (IFAU).
- Kopczuk, W. (2003). Tax bases, tax rates and the elasticity of reported income. NBER Working Papers 10044.
- Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics*, 89(11-12):2093–2119.
- Kopczuk, W. (2012). The polish business “flat” tax and its effect on reported incomes: A pareto improving tax reform? mimeo.
- Long, J. (1999). The impact of marginal tax rates on taxable income: Evidence from state income tax differentials. *Southern Economic Journal*, 65(4):855–869.
- MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics*, 18:83–114.
- Moffitt, R. A. and Gottschalk, P. (2002). Trends in transitory variance of earnings in the united states. *The Economic Journal*, 112(478):C68–C73.
- Moffitt, R. A. and Wilhelm, M. (2000). Taxation and the labor supply decisions of the affluent. In Slemrod, J., editor, *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*. Harvard University Press and Russell Sage Foundation, New York.
- Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, 20(4):111–132.
- Saez, E. (2003). The effect of marginal tax rates on income: A panel study of ‘bracket creep’. *Journal of Public Economics*, 87:1231–1258.
- Saez, E. (2004). Reported incomes and marginal tax rates, 1960-2000: Evidence and policy implications. In Poterba, J., editor, *Tax Policy and the Economy*, volume 18, pages 117–173. The MIT Press, Cambridge.

- Saez, E., Slemrod, J., and Giertz, S. (2009). The elasticity of taxable income with respect to marginal tax rates: A critical review. NBER Working Paper No. 15012.
- Saez, E., Slemrod, J., and Giertz, S. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, 50(1):3–50.
- Singleton, P. (2011). The effect of taxes on taxable earnings: Evidence from the 2001 and related U.S. federal tax acts. *National Tax Journal*, 64(2, Part 1):323–352.
- Slemrod, J. (1996). High-income families and the tax changes of the 1980s: The anatomy of behavioral response. In Feldstein, M. and Poterba, J. M., editors, *Empirical Foundations of Household Taxation*, pages 169 – 192. Chicago: University of Chicago Press and National Bureau of Economic Research.
- Weber, C. (2013). The fixed bracket average treatment effect: A constructive alternative to late analysis for tax policy. University of Oregon Working Paper.

**Table A.1: First-Stage Estimates for Table 2<sup>1</sup>**

	(1)	(2)	(3)	(4)	(5)
$\Delta \ln(1 - \tau_t^p)$	0.572*** (0.017)	0.520*** (0.017)	0.565*** (0.017)		
$\Delta \ln(1 - \tau_t^{p,1lag})$				0.229*** (0.016)	
$\Delta \ln(1 - \tau_t^{p,2lags})$	0.056*** (0.013)	0.031** (0.012)	0.062*** (0.013)	0.105*** (0.014)	0.159*** (0.014)
$\Delta \ln(1 - \tau_t^{p,3lags})$	0.049*** (0.014)	0.019 (0.013)	0.051*** (0.014)	0.094*** (0.015)	0.096*** (0.015)
$\Delta \ln(1 - \tau_t^{p,4lags})$					0.122*** (0.015)
Observations	24,576	24,576	24,576	24,576	24,576
Individuals	6,044	6,044	6,044	6,044	6,044
First Stage F-statistic	487.9	369.4	484.7	145.0	108.2

<sup>1</sup>Heteroskedasticity-robust standard errors clustered by the individual are in parentheses. All covariates are suppressed. The covariates in each column are the same covariates that are included in the 2SLS estimates reported in Table 2.