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

Caroline Weber*

October 2011

Abstract

The elasticity of taxable income (ETI) is a central parameter for tax policy debates. This paper shows that most estimators employed in the literature fail to obtain consistent estimates of the ETI. A new method is proposed that will provide consistent estimates under testable assumptions regarding the degree of serial correlation in the error term. Using this procedure, I estimate an ETI of 1.046, which is more than twice as large as the estimates found in the most frequently cited paper on this subject (Gruber and Saez, 2002). I also consider an alternative definition of treatment, which was proposed in Weber (2011). It has a minimal effect on the estimates, but decreases standard errors by 16 percent.

*Ph.D. Candidate, Department of Economics, University of Michigan: ceweber@umich.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.\(^1\) An individual’s response to a tax change could take a number of forms including a labor supply response, a change in tax avoidance strategies (e.g. changing the amount of itemized deductions accrued), or a change in the extent of tax evasion. The ETI captures all of these. It is informative on its own and has also been shown, under certain assumptions, to be a sufficient statistic for marginal deadweight loss.\(^2\) Therefore, it has been a popular parameter for public finance economists to estimate and obtaining a consistent estimate is valuable for policy debates.

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

---

\(^1\)The marginal net-of-tax rate is one minus the marginal tax rate.
\(^2\)For example, see Feldstein (1999) and Chetty (2009).
\(^3\)I use instrument exogeneity throughout this paper as it is used in the ETI literature; that is, an instrument is exogenous if the instrument is uncorrelated with transitory shocks in the error term.
\(^4\)The standard estimating equation regresses the log change in taxable income on the log change in the marginal net-of-tax rate and other covariates.
poses new methods when necessary, where the conditions required for consistency when estimating the response to a marginal tax rate change are laid out in Weber (2011). I use the Michigan IRS Tax Panel data set for the years 1979-1990 for empirical applications.

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

The second key contribution of this paper is to propose a new instrument that is exogenous under testable assumptions regarding the degree of serial correlation in the error term. Given the additional requirements for a consistent estimate derived in Weber (2011), this is the only instrument that has the potential to provide a consistent estimate under reasonable assumptions. Using this instrument, my preferred baseline estimate is 1.046. There are likely two main reasons why my preferred baseline estimate is more than twice as large as the estimates found in the most frequently cited paper on this subject (Gruber and Saez, 2002). First, I show that the instrument used in (Gruber and Saez, 2002) biases the estimates downwards when the tax reform decreases marginal tax rates at the top of the income distribution. Second, with an additional auxiliary assumption, this parameter can be interpreted as a Fixed-Bracket Average Treatment Effect (FBATE); that is, the parameter is identified from the subpopulation of taxpayers with no incentive to cross a tax bracket line due to a tax reform or transitory income shock (Weber, 2011). This subpopulation may be more responsive during the time period in which the estimation occurs; however, if

\(^5\)The predicted net-of tax rate change is the change in the net-of-tax rate if an individual had their base-year income in both years.
those excluded will respond the same, on average, in the long-run, FBATE is the relevant parameter for welfare analysis.

The paper also addresses the use of income splines to control for heterogeneous income trends at different ranges in the income distribution. While theoretically a reasonable idea, the implementation is often such that the splines are endogenous; in fact, eliminating endogeneity of the splines changes the sign of the majority of spline coefficients in my preferred baseline estimates. Overall, controlling for heterogeneous income trends plays a relatively minor role when the difference length is short (i.e. one year), but as expected, increases as the difference length increases.

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.\(^6\) The latter has been used most often and will be the focus of this paper.\(^7\) Identification usually comes from differences in tax rate changes across individuals brought about by a tax reform.\(^8\) Given that tax reforms frequently change tax rates more for high income individuals, identification is often obtained by comparing the taxable income response of individuals at the top of the income distribution who experience a large tax rate change to those lower in the income distribution who experience a low or no tax rate change. Estimating the ETI accurately requires data that will provide a precise measure of individuals’ taxable income, which makes administrative tax return data attractive.

The ETI is obtained by regressing the log change in taxable income on the log change in the net-of-tax rate (as well as other covariates). Without using an instrument for the log change in the net-of-tax rate, it is clearly endogenous because it is a function of taxable income.

---

\(^6\) For a comparison of these methods, see Saez et al. (2010).

\(^7\) Throughout the paper, difference-in-differences is synonymous with panel-based difference-in-differences.

\(^8\) Papers that obtain identification by other means are beyond the scope of this paper.
income—the dependent variable in the regression. As a result, all regression-based studies of the ETI use an instrument for the log change in the net-of-tax rate. The most common instrument is the value for the change in the net-of-tax rate given the tax reform if individuals earned their base-year income (income is income in the first year of the difference) in both years. Instruments that are only a function of taxable income are employed because tax return data sets are normally used and usually do not have rich demographic data, which could provide alternative instruments.\footnote{At least this has been traditionally true. But, more recently, data sets from other countries which have much better demographic data have been employed. And, in the U.S., Singleton (2011) has linked two different data sets, one of which also provides much better demographic data. However, alternative instruments based on demographics have not been used in any of these studies. The only study to make use demographic instruments to estimate the ETI was Moffitt and Wilhelm (2000) who used the SCF, not an actual tax return data set.} Since the instrument is still a function of the dependent variable, there is no guarantee that the instrument employed is exogenous. The literature has identified two problems that can cause remaining endogeneity of the instrument: mean reversion and heterogenous 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 within the range of 0.4-0.62 depending on the functional form of base-year income used as a regressor (Auten and Carroll, 1999; Gruber and Saez, 2002). The ETI was also estimated using 1990’s data, in which the predominant federal reforms were targeted tax increases (the Omnibus Budget Reconciliation Acts of 1990 and 1993). The estimates ranged from 0.17-0.38 (Carroll, 1998; Giertz, 2005). There has since been a large literature using the same methods to estimate the ETI in other countries (Saez et al., 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) agrees that this variation is due to varying degrees of model mis-specification, but does not derive the exact nature of the biases in each. Several authors have tried to get around this problem by proposing alternative instruments (e.g., Blomquist and Selin, 2010). Still, the estimation methods commonly employed remain those laid out in Auten and Carroll (1999) or Gruber and Saez (2002).

This paper examines each of the proposed estimators to determine whether any of them provide consistent estimates of the ETI and proposes new methods when necessary. I begin in Section 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 provide an interpretation of these estimates. Section 5 concludes.

3 Model Setup

To facilitate the analysis of the issues inherent in estimating the ETI, I set up a simple model of the taxable income process in this section. This process is consistent with the theoretical model that drives estimation strategies in this literature that is formally laid out in Gruber and Saez (2002). Of course, the actual process may be more complex than that laid out in

---

10 Sometimes in the ETI literature, the income concept considered is broad income, rather than taxable income. Broad income generally refers to income before deductions, credits and itemization. Assume income means taxable income throughout the paper unless stated otherwise.
this section. If such additional complexities matter, they will need to be addressed in order to obtain a consistent estimate of the ETI. Hence, to the extent that the case considered here is a special case in more complex estimation strategies, one can think about the analysis based on this model as providing necessary conditions for obtaining a consistent estimate of the ETI, but perhaps not sufficient.

Let individuals' log income \( \ln(Y_{it}) \) be governed by the following equation:

\[
\ln(Y_{it}) = \varepsilon \ln(1 - \tau_{it}^r) + \ln(\mu_{it}) + \ln(\nu_{it}),
\]

or in differences as:

\[
\Delta \ln(Y_{it}) = \varepsilon \Delta \ln(1 - \tau_{it}^r) + \Delta \ln(\mu_{it}) + \Delta \ln(\nu_{it}),
\]

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. Additionally, \( \tau_{it}^r \) is the marginal tax rate to which an individual responds, which is a function of pre-response income \( Y_{it}^r \) and the tax code \( c_t \):

\[
\tau_{it}^r = f(Y_{it}^r, c_t),
\]

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

---

11 For example, the literature has explored the role of income effects (Gruber and Saez, 2002) and tax-base effects (Kopczuk, 2005).

12 For expositional ease, I choose to examine a simple framework in which all the important issues arise. It is relatively clear how to apply the methods described in Subsection 4.2.2 to obtain consistent estimates for many extensions examined in the literature.

13 In this specification, I am assuming that the entire response takes one period. I relax this assumption in the empirical application.

14 For now, I assume that \( \varepsilon \) is the same for all individuals. I relax this assumption in Subsection 4.2.4.
individual, which is not likely homogeneous across individuals. In Subsection 4.3, I will relax this assumption, but for now, assuming homogeneous trends will simplify notation and the analysis conducted prior to Subsection 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:

\[
\ln(\nu_t) = \sum_{k=1}^{K} \phi_k \ln(\nu_{t-k}) + \ln(\xi_{it}),
\]

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

To characterize the determinants of mean reversion, first note that mean reversion has to do only with the transitory component of income. Since \( \mathbb{E}[\ln(\xi_t)] = 0 \), individuals receive a mean zero shock each period. When \( \phi_k = 0 \) for all \( k \), mean reversion at the individual level is very strong, because current income is no longer a function of transitory income in previous periods. Hence, even if individuals have very high or very low incomes relative to their permanent income level this period, on average, their incomes will return to their mean level in the following period. As \( \phi_k \to 1 \), mean reversion weakens. Therefore, when examining data with \( 0 \leq \phi_k < 1 \) in a given year, if one looks at high income individuals,

---

15 I choose to write the income process in this simplified way, because, given the limited demographics in the tax return data sets commonly employed, separate causes of heterogeneous growth rates cannot be identified empirically.

16 Allowing \( \ln(\xi_{it}) \) to have a nonzero mean is equivalent to changing permanent income. Any trends in \( \ln(\xi_{it}) \) are equivalent to trends in permanent income. Hence, without loss of generality, I can assume that \( \mathbb{E}[\ln(\xi_t)] = 0 \).

17 Covariance stationarity along with the assumption that \( \ln(\xi_{it}) \sim iid(0, \sigma^2_\xi) \) implies: \( \mathbb{E}[\ln(\nu_t)] = \mathbb{E}[\ln(\nu_{t-1})] = 0 \) for all \( t \), \( \mathbb{V}ar[\ln(\nu_t)] = \mathbb{V}ar[\ln(\nu_{t-1})] = \sigma^2_\nu \) for all \( t \). And \( \text{Cov}[\ln(\nu_t), \ln(\nu_s)] = \phi_1 \sigma^2_\nu \) for all \( s \neq t \) when \( K = 1 \).

18 Thinking about mean reversion as being caused by a serially correlated error term or, more generally, by transitory income shocks is not unique to this paper. For example, see Kopczuk (2003, 2005), Moffitt and Wilhelm (2000), and Saez et al. (2010), among others.

19 Throughout this paper, for any variable \( w, w_{it} \) is the value this variable takes on for a single individual, and \( w_t \) is the corresponding vector of all individuals at time \( t \) for this variable. All statements made about this vector hold for all \( t \).
it will seem as though, on average, their income falls in the following year, and the reverse is true for low income individuals, even though individuals experience a mean zero shock every period. The actual volatility of transitory income $\sigma^2_\xi$ also affects the severity of mean reversion each period. If there was no income volatility in an economy (i.e. $\sigma^2_\xi=0$), then there would be no mean reversion, because all transitory shocks would be equal to zero. As $\sigma^2_\xi$ increases, the magnitude of mean reversion also increases. As noted in Section 2, mean reversion is believed to be substantial in the U.S. context, and I will provide additional empirical evidence that this is, in fact, the case in Subsection 4.2.4.

4 Data and Estimation

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

This section proceeds as follows. Subsection 4.1 describes the data that will be used in the empirical analysis. Subsection 4.2 conducts baseline theoretical and empirical analysis. Subsection 4.3 conducts theoretical and empirical analysis allowing for heterogeneous growth rates in income at different ranges in the income distribution.

4.1 Data

This subsection describes the data that are used in the regression analysis Subsections 4.2 and 4.3. This section uses the Michigan IRS Tax Panel data set for the years 1979-1990. This is the only publicly available panel tax return data set in the U.S. Given that I will use instruments that are a function of lagged income, I restrict the estimation to the years
The major tax change that takes place during this period is TRA86. This was a substantial reform that changed both the tax rate and the tax base. It also substantially reduced the number of tax brackets in the U.S. tax system. It decreased tax rates for most individuals, particularly at the top of the income distribution and the reform was phased-in; that is, the tax rates were adjusted to their new level over a period of several years. For an extensive discussion of this data set and TRA86, see Gruber and Saez (2002).

The definition of taxable income used in the construction of the dependent variable and income splines in the regressions in Subsections 4.2 and 4.3 is defined in each year so that the tax base is constant across reforms. This is common in the literature; without this adjustment, the dependent variable—the log change in taxable income—would change mechanically as the definition of the tax base changes. To the extent that the tax base alters the tax rate faced by a taxpayer, not making this adjustment could substantially bias the estimates. It is widely recognized in the literature that this mitigates, but does not necessarily resolve the problem, because tax base changes often induce taxpayers to shift from one form of taxable income to another. Addressing this issue more completely is beyond the scope of this paper. As is common in this literature, I exclude capital gains entirely.

Taxable income for each year is in 1992 dollars.

The tax rate variables used in the regressions include both state and federal tax rate changes. All tax rate variables are generated using TAXSIM.

---

20 In Table 2, I do consider one specification that includes all the years, and show that the results are robust to this inclusion.

21 The estimates could also be biased if the tax base changes fall disproportionately on individuals in a particular income range.

22 Kopczuk (2005) tried to address this issue more directly by controlling for changes in the tax base directly in the estimating equation.

23 In general, my income measures are defined as in Gruber and Saez (2002), but a few improvements are made to 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.

24 An overview of TAXSIM can be found in Feenberg and Coutts (1993). I use the full version of TAXSIM, which is available exclusively on the NBER server.
verage elasticity. However, for most of this section, I refrain from weighting the estimates because this data set does not oversample high-income individuals; therefore, income-weighted estimates place substantial weight on individuals whose response I can measure relatively imprecisely in this data. I do provide income-weighted estimates in Subsection 4.2.4; the estimates decrease slightly and there is a sizable increase in the standard errors. The only additional covariates included in the regressions in this section are marital status indicator variables. There are three marital status categories in total: single, married, and head of household/widowed with a dependent child.

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

---

25 This is common in U.S. studies, because the publicly available tax return data lacks additional covariates. For example, see Gruber and Saez (2002). I discuss the extent to which this lack of covariates may or may not affect the estimates later in the paper.

26 Less than twenty individuals in each specification, whose actual log tax change plus their predicted log tax change is greater than one, are excluded. The absolute value of $\Delta \ln(Y_{it})$ is censored at 7. These restrictions are the same as those used in Gruber and Saez (2002) and they have very minor effects on the estimates.

27 The first paper to exclude low-income individuals was Auten and Carroll (1999). They excluded everyone whose taxable income fell below the 22 percent marginal tax rate bracket in 1985, which corresponded to $21,020 in 1985 dollars. Gruber and Saez (2002) exclude everyone with broad income under $10,000. Most subsequent papers that use an income cutoff follow the Gruber and Saez selection rule. Note that excluding a certain portion of the population can improve the mean reversion problem, but it only resolves it completely in the extreme case that everyone left in the sample experiences the same transitory income shock.
reason to exclude those at the very bottom of the income distribution in practice. The use of constant-law taxable income in the dependent variable usually creates a particular type of sample selection. Near zero, constant-law taxable income can take on negative values, which are turned to missing when this variable is converted to logs. This excludes all individuals who earned negative constant-law taxable income in one year of the difference. Since income splines are also a function of constant-law taxable income, this excludes all individuals with negative constant-law taxable income in the base year. This means that just above zero in base-year income, individuals are only included in the sample if they received a positive shock (or a very small negative shock) between periods, which is more likely to generate a substantial bias in the estimates than a clean income cutoff. The proposed $10,000 cutoff is not endogenous as long as the instruments used are not significantly correlated with transitory income shocks in base-year income. I will discuss whether or not this condition holds for the instruments chosen in Subsection 4.2.4. Descriptive statistics for the preferred baseline estimates are provided in Table 1.

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:

\[ \Delta \ln(Y_{it}) = \varepsilon \Delta \ln(1 - \tau_{it}) + \alpha_{t-1} + \eta_{it}, \] (5)

where \(\alpha_{t-1}\) are year fixed effects and \(\eta_{it} = \Delta \ln(\nu_{it})\). The year fixed effects control for any omitted variables in differences that are the same, on average, for all individuals at a given time \(t\), including the homogeneous growth rate \(g_t\). Of course, if all individuals do not share the same income trend \(g_t\), the year fixed effects are no longer enough to produce consistent estimates. I will address this particular case in Subsection 4.3. I assume that \(1 - \tau_{it}' = 1 - \tau_{it}\)
and address what happens when this is not the case at the end of Subsection 4.2.4.

Additionally, assume, as is the case for TRA86 and most other tax reforms, that the tax rate schedule is graduated and the regression includes individuals at all income levels. Then, higher values of $\Delta \ln(\nu_{it})$ lead to higher values of $\Delta \ln(Y_{it})$ (i.e. $\mathbb{E}[\Delta \ln(Y_{it})' \Delta \ln(\nu_{it})] > 0$), all else equal, which in turn lead to lower values of $\Delta \ln(1 - \tau_{it})$. Hence, as has been widely recognized in the literature, $\Delta \ln(1 - \tau_{it})$ is endogenous. Subsection 4.2.1 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. Subsection 4.2.2 uses the same method to examine whether there are alternative instruments which are exogenous under certain assumptions. Subsection 4.2.3 proposes a test, which will allow me to test empirically whether each instrument discussed in the previous two subsections is exogenous. Subsection 4.2.4 conducts this instrument exogeneity test for each of the proposed instrument and control combinations discussed in the previous two subsections and provides baseline ETI estimates.

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

How to address the endogeneity of the net-of-tax rate term is very important, and it has duly received a large discussion in the literature. By far the most frequently used instrument for $1 - \tau_{it}$ is the value for $1 - \tau_{it}$ if an individual’s income was $Y_{it-1}$ in year $t$ and the tax code was that of year $t$, that is, the predicted net-of-tax rate based on income in year $t - 1$. In this subsection, I will focus exclusively on the ability of this instrument to solve the endogeneity problem. I will refer to this instrument as $1 - \tau_{it}^p$. I discuss alternative instruments in Subsection 4.2.2.

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

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

(6)
is large, and that the instrument is exogenous:

\[ \text{cov}[\Delta \ln(1 - \tau_{it}^p), \eta_t] = \mathbb{E}[\Delta \ln(1 - \tau_{it}^p)|\eta_t, X_t] = 0, \quad (7) \]

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):

\[ \text{plim}(\hat{\varepsilon}_{IV}) = \varepsilon + \frac{\text{cov}[\Delta \ln(1 - \tau_{it}^p), \eta_t|X_t]}{\text{cov}[\Delta \ln(1 - \tau_{it}), \Delta \ln(1 - \tau_{it}^p)|X_t]}, \quad (8) \]

and the estimates will be inconsistent.

The predicted net-of-tax rate instrument likely does not suffer from the weak instrument problem in practice. But, it is not clear that the instrument exogeneity condition is satisfied. 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. 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, \( \text{cov}[\Delta \ln(1 - \tau_{it}^p), \ln(\nu_{it - 1})] > 0 \Rightarrow \text{cov}[\Delta \ln(1 - \tau_{it}^p), \eta_t] < 0 \) in the case of a tax decrease, and \( \text{cov}[\Delta \ln(1 - \tau_{it}^p), \ln(\nu_{it - 1})] < 0 \Rightarrow \text{cov}[\Delta \ln(1 - \tau_{it}^p), \eta_t] > 0 \) 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.

---

28 Whether or not this condition holds is determined empirically using an F-test.
29 Technically, to obtain consistent IV estimates, there are two additional conditions that must be satisfied, namely that the instruments are linearly independent and that there are as at least as many instruments as there are endogenous variables. Neither of these conditions are ever violated here, and so it will be assumed that these conditions hold throughout.
30 For example, the F-statistics for the first-stage results estimated by Gruber and Saez (2002) are all between 20 and 100.
31 This is by far the most common form of identification because tax reforms typically take this form. I will address how this discussion is altered if a different form of identification is used in a paragraph later in this section.
decrease and biased upwards in the case of a tax increase. Therefore, $\mathbb{E}[\Delta \ln(1 - \tau_p^t)'\eta_t] = 0$ only if $\mathbb{E}[\ln(Y_{t-1})'\eta_t] = 0$.

The severity of the endogeneity problem when $\mathbb{E}[\ln(Y_{t-1})'\eta_t] \neq 0$ is clearly a function of both the degree of serial correlation and volatility of transitory income. For notational ease, I am going to assume that the true ETI is zero in the analysis that follows.\(^{32}\) When there is no serial correlation, that is, $K = 0$, \(^{33}\)

$$
\mathbb{E}[\ln(Y_{t-1})'\eta_t] = \mathbb{E}[\ln(Y_{t-1})'(\ln(\nu_t) - \ln(\nu_{t-1}))]
= \mathbb{E} [\ln(Y_{t-1})'(\ln(\xi_t) - \ln(\xi_{t-1}))]
= -\mathbb{E} [\ln(Y_{t-1})'\ln(\xi_{t-1})]
= -\sigma_\xi^2 < 0.
$$

(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, $\text{cov}(\ln(\mu_{t-1}),\ln(\xi_{t-1}))$. But, the expression is guaranteed to remain negative unless $\text{cov}(\ln(\mu_{t-1}),\ln(\xi_{t-1}))$ is negative, so that high values of $\ln(\xi_{t-1})$ are offset by low values of $\ln(\mu_{t-1})$. It is highly unlikely that this is the case.

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

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

$$
\mathbb{E}[\ln(\nu_{t-3})'\eta_t] = \mathbb{E}[\ln(\nu_{t-3})'(\ln(\nu_t) - \ln(\nu_{t-3}))]
= \mathbb{E} [\ln(\nu_{t-3})'(\ln(\xi_t) - \ln(\xi_{t-3}))]
= -\mathbb{E} [\ln(\nu_{t-3})'\ln(\xi_{t-3})]
= -\sigma_\xi^2 < 0.
$$
Now, suppose there is first-order serial correlation, that is $K = 1$. Then,

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

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

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

$$
= (\phi_1 - 1)^2 \sigma_\nu^2
$$

$$
= \frac{\sigma_\xi^2}{1 - \phi_1^2} = -\frac{\sigma_\xi^2}{1 + \phi_1} < 0.
$$

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. Compared to the case when there is no serial correlation as given by (9), (10) is larger in absolute value. Also, note that (10) would equal zero if $\phi_1 = 1$, but this would imply that income follows a unit-root process, which would generate an alternative set of issues to be addressed. Unless the error term follows a unit root process, $\ln(Y_{it-1})$ is correlated with the error term.\(^{35}\) The general notion that this instrument remains endogenous has been well-acknowledged in the literature.

Of course, if there was a tax reform for which the tax change was the same for all income levels, the endogeneity problem discussed in the previous paragraphs would be eliminated, but it would also eliminate identification because everyone would experience the same treatment, and there would be no variation to exploit in order to estimate the elasticity. Al-

\(^{34}\) Again, this derivation is for one-year differences. For three-year differences, the covariance is strictly larger assuming the covariance between transitory shocks and income stays constant over time because this covariance is given by:

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

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

$$
= \mathbb{E}[\ln(Y_{t-3})' (\phi_1^3 - 1) \ln(\nu_{t-3})]
$$

$$
= (\phi_1^3 - 1)^2 \sigma_\nu^2
$$

$$
= \frac{(\phi_1^3 - 1)^2 \sigma_\xi^2}{1 - \phi_1^2} < 0,
$$

and $|\phi_1^3 - 1| > |\phi_1 - 1|$.

\(^{35}\) A non-zero covariance is also found when higher orders of serial correlation are considered.
ternatively, if there was a tax reform that affected some individuals, but not others within a given income class, the endogeneity problem would be mitigated. Different tax changes at different points in the income distribution in different years has been suggested as an alternative way to resolve the endogeneity problem. If there are no year effects, then this will indeed help mitigate the problem, but if year fixed-effects are included in the regression, such variation is absorbed in the year fixed-effects and does not aid in identifying the ETI.

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

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

$$ln(Y_{it-1}) = ln(\nu_{it-1}) + ln(\mu_{it-1}).$$ (11)

---

36For example, Long (1999) just uses state tax rate variation and most ETI studies in the U.S. combine federal and state tax rate variation.

37The first paper to include this control is Auten and Carroll (1999).
And, $\text{cov}(ln(Y_{t-1}), ln(\mu_{t-1})) = \sigma^2_{\nu} > 0$. Hence, $ln(Y_{t-1})$ will produce a biased estimate of $ln(\nu_{t-1})$ and therefore remain endogenous. This control is thus valid only when the original instrument is valid. But, in that case, it is not needed (at least not to solve an endogeneity problem). I will return to a discussion of whether it is relevant for heterogeneous income trends in Subsection 4.3).

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

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

(12)

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

$$\Delta ln(Y_{it-1}) - \Delta ln(Y_{it-2}) = (ln(\nu_{it-1}) + g_{t-1} - ln(\nu_{it-2})) - (ln(\nu_{it-2}) + g_{t-2} - ln(\nu_{it-3}))$$

(13)

$$= ln(\nu_{it-1}) - ln(\nu_{it-3}),$$

where the last line follows if $g_{t-1} = g_{t-2}$. This control solves both the issues raised with $\Delta ln(Y_{it-1})$, but generates a new, similar problem, namely $\text{cov}(\Delta ln(Y_{t-1}) - \Delta ln(Y_{t-2}), -ln(\nu_{t-3})) = (1 - \phi_1^2)\sigma^2_{\nu} > 0$. Hence, this control is not valid, either. Therefore, there are no income-based controls that have been proposed that are expected to make this instrument exogenous.

Note, also, that some demographics are always used, and occasionally more extensive demographic covariates are used (e.g. Carroll, 1998; Auten and Carroll, 1999; Singleton, 2011). However, in general these covariates are variables such as occupation and education level,

---

38 The only exception to this statement would be if high levels of permanent income were correlated with low shocks, but as I noted before, this is not likely the case.
39 A similar result holds for all $K \neq 1$.
40 A similar result holds for all $K \neq 1$.
41 Kopczuk (2003, 2005) suggests this proxy could be improved by including it as a 10-piece spline instead. But, this does not resolve the problem.
42 A similar result holds for all $K \neq 1$. 

18
which likely proxy for permanent, not transitory income. As a result, these are also not expected to resolve the endogeneity problem.\footnote{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 Subsection 4.3, not as a proxy for transitory income shocks.}

### 4.2.2 Alternative Instruments

Subsection 4.2.1 demonstrates that $\Delta \ln(1 - \tau_p^t)$ is not exogenous as an instrument, regardless of the additional controls used. There are two possible instrument types that I will consider in this subsection.\footnote{There is also a third possibility that has been proposed in the literature. Carroll (1998) suggests constructing the tax rate instrument based on average income over the sample period. Blomquist and Selin (2010) show that this instrument is exogenous under the assumptions of covariance stationarity and normality of the error term. However, this instrument will ultimately suffer from the same problem as that proposed by Blomquist and Selin (2010), so I do not discuss it separately here.} First, I consider a potential instrument that has not been previously considered in the literature.\footnote{Although this particular instrument has not previously been employed in the static panel literature that estimates the ETI, it has been employed by Holmhund and Soderstrom (2008) in a dynamic panel setting.} In particular, suppose that instead of making the predicted tax rate instrument a function of $\ln(Y_{it-1})$, it was instead a function of some lag of $\ln(Y_{it-1})$. If the relevant instrument exogeneity condition holds, then this will indeed resolve the endogeneity problem brought about by mean reversion. This approach is standard for resolving endogeneity problems in the dynamic panel literature, which always puts lags of the left-hand side variable on the right-hand side.\footnote{For example, see Anderson and Hsiao (1982).}

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

\[
E[\ln(Y_{t-2})' \eta_t] = E[\ln(Y_{t-2})'(\ln(\xi_t) - \ln(\xi_{t-1}))] = E[\ln(Y_{t-2})' \ln(\xi_t)] - E[\ln(Y_{t-2})' \ln(\xi_{t-1})] = 0.\tag{14}
\]
So, this instrument would clearly be valid when \( K = 0 \). Now consider the case where \( K = 1 \):

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

\[
= E[\ln(Y_{t-2})'\left(\ln(\xi_t) + \phi_1\ln(\nu_{t-1}) - \ln(\nu_{t-1})\right)]
\]

\[
= E[\ln(Y_{t-2})'\left(\ln(\xi_t) + (\phi_1 - 1)\ln(\xi_{t-1}) + \phi_1(\phi_1 - 1)\ln(\nu_{t-2})\right)]
\]

\[
= \phi_1(\phi_1 - 1)E[\ln(Y_{t-2})'\ln(\nu_{t-2})],
\]

\[
= \phi_1(\phi_1 - 1)\sigma^2_{\nu},
\]

which does not equal zero when there is no unit-root. However, since \( E[\ln(Y_{t-1})'\ln(\nu_{t-1})] = \sigma^2_{\nu} \) is the same for all \( l \), the covariance is strictly less than the covariance when the instrument was based on \( \ln(Y_{t-1}) \) because \( |(\phi_1 - 1)|\sigma^2_{\nu} > |\phi_1(\phi_1 - 1)|\sigma^2_{\nu} \). Hence, the endogeneity problem is strictly better than it was before. If the error process is truly serially correlated, the recursive structure of the error term will cause conditions like that found in (15) to always be violated, regardless of the number of lags chosen. However, the value of the covariance will get arbitrarily small as the number of lags increase. Alternatively, one could rephrase this statement in terms of a testable hypothesis. If enough lags are used, eventually the null hypothesis that \( E[\ln(Y_{t-1})'\ln(\eta_t)] = 0 \) will not be rejected. It is also possible that the true underlying process is not serially correlated, but rather a moving-average process. In this case, the same basic idea holds, but the recursive structure is gone. For example, if the error process is MA(1), \( E[\ln(Y_{t-3})'\ln(\eta_t)] \) will equal zero exactly. In practice, it is usually not possible to distinguish between serial correlation that dies out quickly and a moving average process.\(^{47}\) Hence, in Subsection 4.2.3, I will consider a test that will tell us whether or not I can reject the hypothesis that \( E[\ln(Y_{t-1})'\ln(\eta_t)] = 0 \), and I will abstract away from whether the true underlying process is moving-average or serially correlated.

There is also another possible type of instrument that has already been proposed in the literature that has been proven to be valid only for even differences (e.g. it is valid for two-

\(^{47}\)MaCurdy (1982), in his study on the properties of the error component of earnings, makes this observation. In his study he finds that he cannot distinguish between an earnings process that is AR(1) or MA(2).
I repeat their proof here, using a two-year difference with first-order autocorrelation as an example. Suppose, I constructed the tax rate instrument as a function of $\ln(Y_{t-1})$. Now, the relevant instrument exogeneity condition is $E[\ln(Y_{t-1})'(\ln(\nu_t) - \ln(\nu_{t-2}))] = 0$, which holds exactly:

$$E[\ln(Y_{t-1})'(\ln(\nu_t) - \ln(\nu_{t-2}))] = E[\ln(\nu_{t-1})'(\phi_1\ln(\nu_{t-1}) + \ln(\xi_t))]$$

$$= \phi_1\sigma^2_{\nu} - \phi_1\sigma^2_{\nu} = 0.$$

Blomquist and Selin (2010) also assume that $\ln(\nu_t)$ is distributed normally, which allows them to conclude that $\ln(Y_{t-1})$ and $(\ln(\nu_t) - \ln(\nu_{t-2}))$ are independent. This in turn implies that any function of $\ln(Y_{t-1})$ is independent of $(\ln(\nu_t) - \ln(\nu_{t-2}))$, including the tax rate instrument. Therefore, under these assumptions, constructing $\Delta\ln(1 - \tau^p_{it})$ as a function of $\ln(Y_{it-1})$ will be a valid instrument. Now, consider a three-year difference. There is no longer a period that lies directly in the middle of the difference. Suppose an instrument one

---

48Auten and Carroll (1999) and others have raised concerns about instruments being constructed as a function of post-response income. This only matters to the extent that individuals move across tax brackets in response to the tax change. If cross-bracket movement is due to transitory income shocks only, then it will be resolved as long as the instrument is uncorrelated with these transitory shocks. A more fundamental problem occurs if somehow the behavioral model underlying individuals’ responses suggests that there are cases when their response will cause them to move from one tax bracket to the next. However, in this case it can be shown that neither pre- nor post-response income instruments are guaranteed to resolve the problem. A more formal discussion of this issue is left for future work.

49In general, zero covariance does not imply independence, which, in turn, implies that this instrument may not be valid if the error terms are not distributed normally, at least for short differences. Also, the same results hold if a higher order serially correlated process or an MA process is assumed instead.
period after the base-year is being considered as a potential instrument.\footnote{If the instrument used is two periods after the base-year, it can be shown that the bias is the same in magnitude, but opposite in sign.}

$$
\begin{align*}
\mathbb{E}[\ln(Y_{t-1})' (\ln(\nu_t) - \ln(\nu_{t-3}))] &= \mathbb{E}[\ln(\nu_{t-1})' (\ln(\nu_t) - \ln(\nu_{t-3}))] \\
&= \mathbb{E}[\ln(\nu_{t-1})'(\phi_1 \ln(\nu_{t-1}) + \ln(\xi_t))] \\
&\quad - \mathbb{E}[(\phi_1^2 \ln(\nu_{t-3}) + \phi_1 \ln(\xi_{t-2}) + \ln(\xi_{t-1}))(\ln(\nu_{t-3}))] \\
&= \phi_1 \sigma^2_\nu - \phi_1^2 \sigma^2_\nu > 0.
\end{align*}
$$

(17)

So, the instrument is not exogenous. However, in looking at the first line, note that if a test showed in the method I discussed previously that an instrument one period away from the end years was approximately valid, this instrument would also be approximately valid. Therefore, even though exact exogeneity no longer holds, if the difference is long enough, the instrument will still be approximately exogenous in practice.

### 4.2.3 Testing Instrument Validity

In this subsection, I consider tests of instrument validity that will determine which, if any, lags of $\ln(Y_{ut-1})$ are valid for constructing the predicted net-of-tax rate instrument discussed in Subsection 4.2.2.\footnote{These tests will introduce a pretest bias (Guggenberger and Kumar, 2011). However, it is mitigated by the fact that I will use these tests in the next subsections as evidence of which instruments are valid, but use my originally hypothesized choice of lags—two, three, and four—even though the null hypothesis is marginally not rejected for one lag. In general, researchers should be cautious about agnostically implementing this method without considering the size distortions this may induce. Also, the inference reported in the next subsections is robust to using inference based on the Anderson-Rubin test, which Guggenberger and Kumar (2011) show is less subject to size distortions from non-exogeneity, which would arise if the over-identification test failed to reject an instrument that is, in fact, endogenous.} If serial correlation did not cause the estimator to be inconsistent, then I could estimate the model, obtain a consistent estimate of $\hat{\eta}_{it}$, and use this estimate to estimate the serial correlation properties of $\eta_{it}$. But, this is not the case. The dynamic panel literature faces a very similar problem and employs a variety of alternative tests to assess whether or not the instruments used are valid. One common method is tests based on over-identifying restrictions (Arellano and Bond, 1991). One such test is the Sargan test.\footnote{This test also goes by the name J-test, which might be more familiar to readers.}
The test statistic is given by:

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

(18)

where \( Z \) is a matrix of all instruments used stacked over all time periods, and \( p \) is the number of instruments. Note that the middle piece of this sandwich estimator is an estimate of the heteroskedasticity-robust covariance matrix. And, recall from the previous subsection that the condition which must hold in order for the instruments to be valid is given by \( \mathbb{E}[Z_t' \hat{\eta}_t] = 0 \) for all \( t \). Therefore, when this statistic is close to zero, the instruments are valid (null hypothesis), and when it is far away, they are not (alternative hypothesis). How effective this method is at detecting all instruments that are, in fact, endogenous depends on the power of the test; that is, how often the test fails to reject the null hypothesis when it is in fact false. Arellano and Bond (1991) provide some suggestive evidence about this issue in a dynamic panel setting.\(^{53}\) In their setup, the null hypothesis is rejected 78 percent of the time at the 10 percent level when it is in fact false if the underlying serial correlation is 0.3, and only 47 percent of the time when the underlying correlation is 0.2. They find that a similar test, the Difference-in Sargan test, has higher power:

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

(19)

where \( Z_I \) is a matrix of all instruments believed to be valid under the null and alternative hypothesis and \( p_I \) is the number of these instruments. The null hypothesis of this test is that all instruments are valid, and the alternative hypothesis is that instruments that are being tested (those that are valid under null and not valid under the alternative hypothesis) are not valid. Note that the second component should be small. If it is large, the entire instrument set should be rejected based on the Sargan test. Using the Difference-in-Sargan

\(^{53}\)Their setting is not quite the same as the one considered in this paper, so these results are only suggestive. Running a simulation study of these tests in this particular context is an area for future research.
test in the Arellano-Bond setup, the null hypothesis is rejected 91 percent of the time at the 10 percent level when it is in fact false if the underlying serial correlation is 0.3, and 60 percent of the time when the underlying correlation is 0.2.

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

\subsection{Empirical Results}

Subsections 4.2.1 and 4.2.2 have examined the choice of instrument theoretically. Subsection 4.2.3 provided a way in which the degree of endogeneity can be quantified empirically, which is used in this subsection to provide a quantitative analog to Subsections 4.2.1 and 4.2.2, highlighting the biases induced by incorrect methods, and ultimately obtaining a consistent estimate of the ETI. This subsection concludes by interpreting this parameter in light of Weber (2011), and considers an alternative definition of treatment which provides a substantial decrease in standard errors.

Table 2 provides empirical estimates for each of the proposed instrument and control combinations considered above. This table estimates two-year differences so that all the different possible instruments can be considered using the same difference length.\textsuperscript{55} The estimating equation is given by (5) (with the addition of marital status indicators) using instruments and sometimes additional income controls as proposed in the previous subsections.\textsuperscript{56} Note that for each of the income controls, I use a 5-piece spline, rather than just

\textsuperscript{54}The size of these tests is very good according to Arellano and Bond (1991); that is, the tests do not incorrectly reject the null hypothesis when it is true more than they should. Therefore, in interpreting these results, one should not worry more than usual that a consistent estimator has been rejected.

\textsuperscript{55}Recall that the instrument proposed by Blomquist and Selin (2010) is only applicable for even differences.

\textsuperscript{56}The instruments are tested in the same order as they were discussed in previous subsections.
including the income control directly. This is standard in the literature, since it makes the control more flexible. This equation can be interpreted as a continuous treatment difference-in-differences equation. In Columns (1)-(4), I assume that a predicted tax rate instrument constructed from income lagged two and three periods prior the base-year are exogenous. This means that both of these instruments are included in the instrument set in each column and are used to test the exogeneity of another, potentially endogenous, instrument using the Difference-in-Sargan test. Column (5) will test whether income lagged two periods is, in fact, exogenous. Columns (1)-(6) are restricted so that the same individuals appear in each to enhance comparability across the columns.

Table 2 Columns (1)-(3) consider specifications where the instrument is shown to be endogenous in Subsection 4.2.1, except under extreme assumptions. Column (1) uses the predicted tax rate instrument $\Delta ln(1 - \tau^p_t)$. The estimate is 0.144 and the p-value from the Difference-in-Sargan test is 0.006; therefore, I can strongly reject instrument exogeneity, as predicted by the theoretical results. Column (2) adds splines in log base-year income $ln(Y_{t-2})$. While the estimate changes substantially—it more than doubles to 0.453 and becomes statistically significant at the one percent level—the p-value of the Difference-in-Sargan test is even smaller (0.000). Column (3) again repeats Column (1), but adds the lagged value of the dependent variable as a spline $\Delta ln(Y_{t-1})$. The estimate is 0.171 and is marginally insignificant. The null hypothesis that the instrument is exogenous is rejected at the 10 percent level. The results in Columns (1)-(3) are strongly consistent with the theoretical analysis, and highlight that while the literature has believed that these splines can do a lot to resolve the endogeneity of the instrument created by mean reversion, it is simply

---

57 I have also tried higher-order splines and this has a minimal effect on the results that follow.
58 Each predicted tax rate instrument is constructed by running actual taxable income in the prediction year through TAXSIM for each outcome year of interest. For example, suppose I am constructing a one-year difference predicted tax rate instrument as a function of income lagged two periods. I take income in period $t - 3$ and run it through TAXSIM for the years $t - 1$ and $t$. I then use these tax rates to construct the difference.
59 Note that the Difference-in-Sargan test and the Sargan are equivalent unless there are at least two instruments that are assumed exogenous.
60 I have also combined the controls proposed in Columns (2) and (3). The p-value on the test of instrument exogeneity in this case is 0.004.
not true in this context. The results also highlight that mean reversion is a substantial issue in the U.S., because it would otherwise be difficult to reject the null hypothesis that these instruments are not exogenous.

Table 2 Columns (4)-(6) consider instruments that are exogenous under certain, more reasonable assumptions that were proposed in Subsection 4.2.2. Column (4) still assumes that two and three lags are exogenous and tests whether one lag is also exogenous. The p-value is 0.163, and therefore can barely not be rejected at the 10 percent level. Column (5) assumes that an instrument lagged three and four periods is exogenous and tests whether can reject the null that two lags are exogenous. The p-value on this test is 0.490. The ETI estimate in Column (5) is 1.046 and is statistically significant at the 1 percent level. The estimates in Column (5) are my preferred baseline estimates. Looking at Columns (1), (4), and (5), the instruments become more exogenous as the lags of income used to construct the instruments increase, and the estimates increase as the instruments move towards being exogenous. This is consistent with the theoretical analysis in Subsection 4.2.1, which showed that the estimates would be biased downwards when the endogenous instruments were used (for a tax decrease). The estimate in Column (5) is more than twice as large as the Column (2) estimates, where the latter is a commonly used method in the literature, originally proposed in Gruber and Saez (2002).

One disadvantage of using lagged income in the construction of the instrument is that the first years of the data set are used exclusively for constructing lags; it is for this reason, that only the years 1983-1990 are used in my preferred baseline estimate. Furthermore, using lags increases standard errors and decreases F-statistics. Regarding the latter concern, F-statistics do decline, but they are still far from a critical level, since they all remain over 100. The standard errors do increase, but they do this even when all columns are restricted

---

61I prefer these estimates over Column (4) given that I barely fail to reject the null hypothesis in Column (4) and the power of this test may be slightly weak as discussed in Subsection 4.2.3.

62Note that I could drop the fourth lag instrument after the tests are over; however, the use of the fourth lag decreases standard errors, even though an additional year of data can no longer be used in the second-stage estimating equation.
to the years 1983-1990. This suggests much of the increase is reflective of the fact that the original treatment measure is quite endogenous. Later in this section, I will consider an alternative definition of treatment that reduces the standard errors.

Table 2 Column (6) considers the instrument proposed by Blomquist and Selin (2010). This estimate is 1.145 and statistically significant at the 1 percent level.\(^{63}\) The estimates in Columns (5) and (6) are similar, as would be expected if the instruments in both columns are exogenous. Recall, in the data subsection, I noted that all individuals with base-year income below $10,000 were excluded from the estimation. It is reasonable theoretically to assume that the instruments employed are not correlated with base-year income, and if they were, I would be able to reject the null hypothesis that the instruments are exogenous.\(^{64}\)

Table 2 Column (7) estimates the same specification as Column (6), but includes all years, 1979-1990; it is the one ETI estimate in this paper that exploits variation from both ERTA81 and TRA86. ERTA81 also provided marginal tax rate decreases to most taxpayers. Although the reforms share a lot in common, the fact that the estimate remains stable when ERTA81 is included, provides some evidence that this method is robust. It also suggests that the other estimates are losing out on an opportunity to decrease standard errors, but otherwise little is lost by restricting the data set to 1983-1990.

Table 2 Columns (8) and (9) repeat Column (5) for one- and three-year differences, respectively. The estimates for one- two- and three-year differences are similar overall.\(^{65}\) The literature has interpreted this similarity as evidence that the short-run and long-run responses are quite similar (e.g. Gruber and Saez, 2002). This is certainly part of the story; however, accepting this explanation as the whole story potentially overlooks the fact that none of these estimates are identifying the parameter they are ostensibly measuring. The

---

\(^{63}\) I do not include any additional instruments in this specification to test for instrument exogeneity, because the estimates are assumed exogenous. However, if I do include instruments lagged three and four periods the p-value on the Difference-in-Sargan test is 0.401.

\(^{64}\) Note that imposing a cutoff as a function of lagged income is not necessarily better or worse when lags of income are used as instruments.

\(^{65}\) In a set of estimates not reported here, which restrict Columns (5), (8), and (9) so that they include the same taxpayers, the one-year estimates are slightly smaller than the two- and three-year estimates, which are almost identical.
fundamental problem lies in the nature of overlapping differences. To see the problem, suppose (as was the case for TRA86) that the tax rate changes take place in two years (i.e. it is phased-in)—1987 and 1988—and it takes three years for individuals to respond fully to the tax reform. Consider three-year differences. The difference that spans 1986-1989 captures the long-run response to the tax changes that took place in 1986, and two years of the response to the tax changes that took place in 1987. In contrast, the 1984-1987 difference only captures a one-year response to the tax changes implemented in 1987. Therefore, the estimate is a combination of short-run, medium-run, and long-run responses.

One-year differences suffer from a variant of the same problem. The 1986-1987 estimates a one-year response to the 1987 tax changes. The 1987-1988 difference estimates a one-year response to the 1988 tax changes, but also picks up the second-year response to the 1987 tax changes to the extent that the tax changes in the second year are correlated with those in the first. The 1988-1989 difference will suggest a response (the second-year response to the tax change in the previous year and the third-year response to the tax changes two years before), even when there is not a tax change this period. This will be absorbed in the year fixed-effect. The phase-in was such that most individuals experienced larger tax rate changes in 1986-1987 relative to 1987-1988. Therefore, the response to the tax reform in 1986-1987 was necessarily larger overall than the response in 1987-1988; otherwise, significant estimates would not be obtained (the data would suggest individuals responding the same, or perhaps even less to tax changes that were larger in magnitude). Further, to the extent that the responses in these two years were more similar than the gap in tax rate changes suggests, the estimates are biased downwards. Note that if the reform were not phased-in, one-year differences would effectively capture the short-run response. For this reason, tax changes with no phase-in period (and more generally, with tax reforms relatively far apart) are preferred.

This paper emphasizes proper interpretation of estimates using varying difference lengths. However, this issue along with any anticipatory response or short-term income shifting in-
duced by TRA86, are not resolved in this paper. The obvious solution to these problems is to control directly for leads and lags of the tax change. However, Weber (2011) effectively rules this option out in the U.S. context, because if individuals respond to leads and lags of tax rates, it is almost guaranteed that it will not be possible to identify a causal parameter for the lead and lag terms. A non-causal parameter would both be uninteresting on its own and ineffective at resolving the issues with the static estimates.

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

Weber (2011) shows that the instrument in Table 2 Column (6), which was proposed by Blomquist and Selin (2010), clearly violates this condition as long as there is some heterogeneity among individuals that are considering crossing the tax bracket line because the

---

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

67For example, suppose the frictions manifest themselves by generating imperfect bunching around the tax kink. If there is a tax decrease above a given tax kink, there will be individuals above this kink before the tax reform who attempt to bunch at that kink after the reform by choosing a taxable income slightly below the tax kink. The actual treatment for each of these individuals is the tax rate change above the tax kink brought about by the reform, but the researcher would assign them an even larger tax decrease: the decline from the original marginal tax rate above the tax kink to the marginal tax rate below the kink.
instrument is a function of post-response income; that is, the instrument strongly predicts who did, in fact, cross the tax bracket line. Column (5) may also violate this condition, but to a much lesser degree; it only violates it if individuals that choose to cross a tax bracket line today for a particular reason were clumped in the same tax bracket several years before. Therefore, the estimates in Column (5) are strictly preferred to those in Column (6), and I will consider an alternative lower bound estimate that provides a consistent lower bound under weaker conditions, if Column (5) does, in fact, violate this assumption, too. If the assumption holds, Weber (2011) shows that the parameter obtained is a Fixed-Bracket Average Treatment Effect (FBATE).

FBATE identifies the average treatment effect for individuals with no incentive to cross a tax kink in response to a tax reform or transitory shock in taxable income. Interpreting the estimates in Table 2 Column (5) in this light provides a new interpretation regarding their size. Certainly, part of the increase relative to estimates such as those found in Gruber and Saez (2002) is due to the decrease in endogeneity of the instruments. However, part of the explanation also likely lies in the fact that FBATE identifies the ETI for a particular subpopulation. Both instrument exogeneity and the additional assumption discussed in the last paragraph contribute to the fact that individuals who face large transitory income shocks do not contribute to the FBATE estimate (i.e. they are included in the estimation but, given the instrument, their response cancels out). To the extent that the estimates in Columns (1)-(3) measure this response (albeit endogenously), those estimates will be lower if individuals do not respond as strongly to tax changes brought about by changes in transitory income, which seems plausible.

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

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

70 For a more technical and extensive discussion of each point made in this paragraph as well as a derivation of FBATE, see Weber (2011).
The fact that it is only possible identify a causal parameter for a subpopulation also has important implications regarding the degree to which these parameters are relevant for welfare analysis. If individuals who receive transitory income shocks would otherwise (and will eventually) respond just as those who do not experience these shocks, then an FBATE parameter is actually more representative of the long-run effect of a tax reform on the population than one that accurately measures each individual’s short-run response.

As highlighted by Chetty (2011), the bounds on the structural parameter relevant for welfare analysis are tighter when marginal tax rate changes are high, because it induces more individuals to overcome their adjustment costs. But, these very same reforms (TRA86 is a prime example) induce a lot of individuals to cross tax bracket lines when they overcome these adjustment costs, and therefore, obtaining a causal parameter (i.e. FBATE) may be more challenging in these contexts; ultimately, the causal estimate will need to be independent of those who face high adjustment costs, such that if they respond, they will cross tax bracket lines. It is overcoming adjustment costs that present the biggest threat to the ability of the estimates in Table 2 Column (5) to obtain FBATE, and if FBATE is obtained, these estimates may be too low in the sense that they exclude some highly responsive individuals. Also, if FBATE fails for this reason, the estimates are likely biased downwards since TRA86 provided a marginal tax rate decrease for most individuals.71

Given that FBATE identifies the parameter for a subpopulation that does not change tax brackets between years, redefining the treatment as the predicted net-of-tax rate as a function of base-year income, should not change the estimate obtained with the original definition of treatment if the assumptions necessary to obtain FBATE are satisfied, but it will decrease standard errors. Furthermore, as long as the instruments used satisfy instrument exogeneity, Weber (2011) shows that this estimate identifies a lower bound on the true average treatment effect.

---

71 This is because individuals who choose to overcome their adjustment costs in response to TRA86 are responding to the tax rate decrease in a higher tax bracket, but the treatment they get assigned is a tax rate increase because researchers observe them in the lower tax bracket in period $t - 1$ and in the higher tax bracket in period $t$.  

31
These estimates are presented in Table 3. Column (1) replicates the baseline specification for two-year differences found in Column (5) of Table 2. Column (2) is the same as Column (1), except treatment is defined as the predicted net-of-tax rate using base-year income. The estimate in Column (2) is 0.975, which is quite similar to the estimate in Column (1)—this suggests the estimates in Column (1) do satisfy the conditions necessary to identify FBATE, or at least the bias due to any violations is not large. The standard errors decrease by 16 percent, suggesting that there is a substantial efficiency gain from employing this alternative treatment definition. Columns (3)-(6) do the same thing for one- and three-year differences and the results are similar to those found for one-year differences (the standard errors decrease by 17-19 percent).

4.3 Heterogeneous Growth Rates and Time Trends

Up to now, the theoretical and empirical results have assumed that $g_t$ is the same, on average, for individuals at all points in the income distribution. As noted above, if $g_t$ is the same for all individuals, it will simply be absorbed in the constant term (or if there are more than one pair of years, it will be absorbed by the year fixed-effects). Moreover, including more pairs of years in the regression does not aid in identifying $g_t$ because it is different for each year. However, a homogeneous growth rate may not a legitimate assumption in practice (at least not in the United States), and this fact has been recognized by researchers in this area.\footnote{Auten and Carroll (1999) were the first to relax this assumption.}

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.\footnote{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.} If it is not controlled for, the income-class-varying portion of the growth rate will end up in the error term, and is likely highly persistent. Therefore, it will likely be correlated with lags of base-year income used to construct the tax rate instrument that would otherwise be exogenous. In most other literatures, this issue is
dealt with by controlling for known factors that influence $g$. For example, MaCurdy (1982) included “...family background variables, education, age, interactions between education and age, and dummy variables for each year of the sample” in his study on the properties of the error structure of earnings. But, such data are not usually available in tax return data. Additionally, in ETI research, perhaps even these would not be enough to fully eliminate the endogeneity problem. Hence, the literature has turned to a second-best alternative, namely including base-year income controls. While they are endogenous when included directly, they can be instrumented using the same lag as is used to instrument for the tax rate variable. When such a suitable lag is used, the income controls effectively control for permanent income plus an uncorrelated measurement error.\footnote{This measurement error is uncorrelated with permanent income by definition. Therefore, when instrumented, log income is a valid control for permanent income, which is what is relevant for determining the heterogeneity in the growth rates, and is a best-case scenario in terms of being a proxy for things such as age and education level.} This measurement error is uncorrelated with permanent income by definition. Therefore, when instrumented, log income is a valid control for permanent income, which is what is relevant for determining the heterogeneity in the growth rates, and is a best-case scenario in terms of being a proxy for things such as age and education level.\footnote{If the income class-specific growth trend increases linearly with income, then including \( \ln(Y_{i,t-1}) \), 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. 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 \( g_{j,t} = g_j \), that is, the heterogenous time trends do not vary over time. Then, including additional pairs of years}

If the income class-specific growth trend increases linearly with income, then including \( \ln(Y_{i,t-1}) \), 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.\footnote{Gruber and Saez (2002) were the first to use splines of income when estimating the ETI. Gruber and Saez (2002) test this assumption and find that it does not matter much.} 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 \( g_{j,t} = g_j \), that is, the heterogenous time trends do not vary over time.\footnote{Gruber and Saez (2002) were the first to use splines of income when estimating the ETI. 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.} Then, including additional pairs of years
aids in identifying these growth rates separately from the tax changes.\footnote{Another alternative is proposed in Gelber (2010). Here the spline coefficients are obtained by regressing the change in taxable income on the spline segments in a year in which there were no tax changes. Then, a new dependent variable is constructed; it is the log change in taxable income minus the spline coefficients times the spline segments in the base year. This method imposes the same assumption that the spline coefficients do not change across years. Just as when the spline segments are included in the main regression directly, this method is valid as long as the spline segments are instrumented with lags, so that the spline coefficients will be consistent.}

Table 4 examines the importance of heterogeneous income controls empirically. I begin with two-year differences. I use a 5-piece spline in the log of constant-law taxable income. In choosing the number of spline segments used, the researcher must balance preferences for flexibility with a desire to keep identification. Empirically, this issue has resolved itself, since a wide range in the number of spline segments employed yield similar estimates.\footnote{For example, Gruber and Saez (2002) use a 10-piece spline, but note that their results change very little when a 20-piece spline is employed. Although not reported here, I find the same robustness to the order of spline used.}

Since the splines are in logs of constant-law taxable income, negative values are excluded from the estimation, both for the base-year (which were already excluded due to the income cutoff at $10,000) and in each year used as an instrument. The results will use two, three, and four lags of income as instruments, so Column (1) repeats Column (5) of Table 2 as a benchmark, excluding those with negative constant-law taxable income lagged two, three, and four periods. Note that this may change the composition of the estimates slightly because individuals with some types of income are more likely to have negative constant-law taxable income than others (and this is likely persistent over time).\footnote{However, there is no reason to believe that this additional cutoff alters the endogeneity of the tax rate or income spline instruments, although it could have a slight effect in practice.}

Empirically, this cutoff has a minimal effect on the estimates.

Table 4 Column (2) adds the base-year income spline directly, which as noted in the previous paragraphs is expected to be endogenous. Column (3) assumes that income splines as a function of three and four lags of income are exogenous and tests whether the income splines as a function two lags are also exogenous.\footnote{The results assuming two and three lags are exogenous and testing whether one lag is exogenous are not reported here, but I find that I can reject the null hypothesis that one lag is exogenous at the 1 percent level.} Including the endogenous income splines
in Column (2) more than doubles the elasticity estimates; however, when instrumented with the appropriate lag, the estimate is 1.461, which is only slightly higher than the estimate in Column (1) without controls for heterogeneous income growth. This suggests that using a lag of income is important; in fact, the sign of the majority of segments of the income spline flip between Column (2) and Column (3) and none of the coefficients on the splines are significant in Column (3). The p-value of the Difference-in-Sargan test for instrument exogeneity of the spline coefficients in Column (3) is 0.227. Column (3) also suggests that controlling for heterogeneous growth rates has a minimal effect on the estimates. This finding is consistent with the fact that researchers who have been able to include more demographic controls have found that they have little effect on the estimated coefficients (Carroll, 1998; Auten and Carroll, 1999).

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

Note that the estimating equation would remain valid if I included the lags of income directly as proxy variables, rather than instrumenting for the income spline. If included as a proxy variable, the researcher is left to guess whether or not the proxy variable is lagged.

---

82I keep the restriction that constant-law taxable income must be positive in each lag so that the composition of each column is the same. When I relax this restriction, I can reject the null that two lags of broad income are an exogenous instrument. When I include three, four, and five lags as instruments instead, I cannot reject that three lags are endogenous. This alternative approach has a minimal effect on the estimated ETI coefficient, so these estimates are not reported here.
enough periods to be exogenous, and whether or not the proxy variable is still close enough to the base-year to be a relevant control. By using an instrument, I was able to test the former using the Difference-in-Sargan test and the latter by testing for weak instruments.

Up to this point, I have not weighted the estimates by income, because, while this is the appropriate parameter estimate for welfare analysis, this data set does not oversample high-income individuals. As a result, the weighted estimates put substantial weight on individuals whose responses can be estimated with relatively little precision. Column (5) of Table 4 repeats Column (3), but weights the estimates by income. Weighting has a minimal effect on the estimates, but it does increase the standard errors by more than 30 percent. Columns (6) and (7) take advantage of the increase in efficiency of the alternative definition of treatment proposed in Subsection 4.2.4 to attempt to obtain weighted estimates with lower standard errors. Column (6) repeats Column (4) of Table 3, but includes heterogeneous income controls. Column (7) adds income weights to Column (6). As expected, the estimates in Column (5) and (7) are quite similar, but the standard errors are almost 20 percent lower. Still, the standard errors on the estimates in Column (7) are too large to be of much practical use.

Table 4 Columns (8) and (9) repeat Column (3) for one- and three-year differences, respectively. The estimates are 0.749 and 1.704. These results highlight that controlling for heterogeneous income trends is more important as the difference length increases, as expected.

---

83 In practice when a proxy variable has been employed, the proxy is constructed several years before the first year examined in the estimation for all years. Sometimes these proxy variables are expected to be equally informative more than ten years later. For example, see Blomquist and Selin (2010), or Holmlund and Soderstrom (2008). Individual’s permanent incomes can change a lot over a decade; so, while the proxy may be exogenous, it may not be controlling very meaningfully for the appropriate growth rate towards the end of the sample period.
5 Conclusion

As aptly summarized by Saez et al. (2010), in their recent *Journal of Economic Perspectives* article on the ETI, a longstanding problem in the ETI panel data estimation literature “is that the identification assumptions lack transparency because they mix assumptions regarding mean reversion and assumptions regarding changes in income inequality.” This paper has carefully disentangled these two issues both theoretically and empirically. The modal approach in the literature—to try to simultaneously rectify mean reversion and heterogeneous income trends with the use of some type of base-year income splines—resolves neither problem. They are ineffective both theoretically and empirically, and the magnitude of the estimates change substantially when alternative methods that resolve the issues properly are employed.

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

The U.S.-centered nature of this paper is partially due to the fact that mean reversion is believed to be particularly strong in the U.S., and partially because I am using a U.S. data set. But, there is a large literature that estimates the ETI for other countries. 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

---

84See Saez et al. (2010) for a review of this literature.
that are based on the methods discussed in this paper are also inconsistent.

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

\textsuperscript{85}I do not attempt to obtain a compensated elasticity in this paper because to do so would require separately identifying the income effect, as in Gruber and Saez (2002). Given the substantial changes in the tax base as part of TRA86, it is unlikely that I could obtain a convincing estimate of the income effect in this context. Researchers estimating these two effects separately with another data set would likely be well-served to use the alternative definition of treatment proposed in Subsection 4.2.4 and its income effect analog in order to maximize efficiency of these estimates.
References


**Table 1: Descriptive Statistics**

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

Taxable income is in 1992 dollars. These summary statistics are for years 1983-1990 and match the restrictions imposed by the baseline estimates given in Table 2, Column (5).
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \ln(1-\tau_t)$</td>
<td>0.144</td>
<td>0.453***</td>
<td>0.171</td>
<td>0.693***</td>
<td>1.046***</td>
<td>1.145***</td>
<td>0.906***</td>
<td>0.780**</td>
<td>0.808***</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.138)</td>
<td>(0.117)</td>
<td>(0.237)</td>
<td>(0.299)</td>
<td>(0.315)</td>
<td>(0.234)</td>
<td>(0.332)</td>
<td>(0.309)</td>
</tr>
<tr>
<td>$1^{st}$ Quintile Spline$^2$</td>
<td></td>
<td>$-0.174^{***}$</td>
<td>$-0.284^{***}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.048)</td>
<td>(0.046)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2^{nd}$ Quintile Spline</td>
<td></td>
<td>0.137</td>
<td>0.335**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.084)</td>
<td>(0.147)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$3^{rd}$ Quintile Spline</td>
<td></td>
<td>$-0.036$</td>
<td>0.132</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.086)</td>
<td>(0.225)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$4^{th}$ Quintile Spline</td>
<td></td>
<td>$-0.046$</td>
<td>$-0.471^{**}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.083)</td>
<td>(0.190)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$5^{th}$ Quintile Spline</td>
<td></td>
<td>0.035</td>
<td>0.270***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.057)</td>
<td>(0.086)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Instruments$^3$: 0,2,3 lags 0,2,3 lags 0,2,3 lags 1,2,3 lags 2,3,4 lags 1 lead 1 lead 2,3,4 lags 2,3,4 lags

Years Included: 82-90 82-90 82-90 82-90 83-90 83-90 79-90 83-90 83-90

Difference Length: 2 years 2 years 2 years 2 years 2 years 2 years 2 years 1 year 3 years

Observations: 29,547 29,547 29,547 29,547 25,087 25,087 80,419 31,646 19,723

Individuals: 6,602 6,602 6,602 6,602 6,224 6,224 31,971 6,899 5,693

Diff-in-Sargan p-value: 0.006 0.000 0.063 0.163 0.427 0.895 0.112

F-statistic: 601.4 440.7 597.6 157.0 569.6 226.5 548.6 88.47 114.9

---

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

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

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

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

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

2 This is a list of the predicted net-of-tax rate instruments used in each column. For example, Column (1) lists the instruments as two lags, three lags, and four lags. This means that predicted net-of-tax rate instruments are constructed for this column as a function of income two, three, and four periods before the base-year. The second two instruments in the list in each column are used to test whether the first instrument listed is exogenous using the Difference-in-Sargan test.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta \ln (1 - \tau_t))</td>
<td>1.180***</td>
<td>2.467***</td>
<td>1.461***</td>
<td>1.240***</td>
<td>1.033*</td>
<td>0.749*</td>
<td>1.704***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.311)</td>
<td>(0.496)</td>
<td>(0.414)</td>
<td>(0.395)</td>
<td>(0.544)</td>
<td>(0.384)</td>
<td>(0.598)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\Delta \ln (1 - \tau_t^p))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.232***</td>
<td>0.954**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.281)</td>
<td>(0.454)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Quintile Spline(^2)</td>
<td>-0.177***</td>
<td>-0.179</td>
<td>-0.483***</td>
<td>-0.135</td>
<td>0.072</td>
<td>0.142</td>
<td>0.317</td>
<td>-0.860</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.366)</td>
<td>(0.162)</td>
<td>(0.574)</td>
<td>(0.325)</td>
<td>(0.520)</td>
<td>(0.314)</td>
<td>(0.543)</td>
<td></td>
</tr>
<tr>
<td>2nd Quintile Spline</td>
<td>0.040</td>
<td>0.375</td>
<td>0.990***</td>
<td>0.297</td>
<td>-0.073</td>
<td>-0.155</td>
<td>-0.482</td>
<td>1.518</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.719)</td>
<td>(0.334)</td>
<td>(0.996)</td>
<td>(0.641)</td>
<td>(0.909)</td>
<td>(0.598)</td>
<td>(1.042)</td>
<td></td>
</tr>
<tr>
<td>3rd Quintile Spline</td>
<td>0.020</td>
<td>-0.666</td>
<td>-1.102***</td>
<td>-0.506</td>
<td>-0.238</td>
<td>-0.150</td>
<td>0.146</td>
<td>-1.361</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.664)</td>
<td>(0.399)</td>
<td>(0.749)</td>
<td>(0.591)</td>
<td>(0.687)</td>
<td>(0.501)</td>
<td>(0.944)</td>
<td></td>
</tr>
<tr>
<td>4th Quintile Spline</td>
<td>-0.125</td>
<td>0.607</td>
<td>0.903**</td>
<td>0.396</td>
<td>0.351</td>
<td>0.205</td>
<td>-0.001</td>
<td>0.609</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.468)</td>
<td>(0.354)</td>
<td>(0.481)</td>
<td>(0.415)</td>
<td>(0.446)</td>
<td>(0.306)</td>
<td>(0.663)</td>
<td></td>
</tr>
<tr>
<td>5th Quintile Spline</td>
<td>0.061</td>
<td>-0.177</td>
<td>-0.347**</td>
<td>-0.145</td>
<td>-0.141</td>
<td>-0.132</td>
<td>0.032</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.191)</td>
<td>(0.171)</td>
<td>(0.192)</td>
<td>(0.168)</td>
<td>(0.177)</td>
<td>(0.109)</td>
<td>(0.273)</td>
<td></td>
</tr>
<tr>
<td>(\Delta \ln (1 - \tau_t)) Instrument lags(^3)</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
</tr>
<tr>
<td>Spline Instrument lags(^3)</td>
<td>None</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
<td>2,3,4</td>
</tr>
<tr>
<td>Years Included</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
<td>83-90</td>
</tr>
<tr>
<td>Difference Length</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>2 years</td>
<td>1 year</td>
<td>3 years</td>
</tr>
<tr>
<td>Income Weights?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>22,991</td>
<td>22,991</td>
<td>22,991</td>
<td>22,991</td>
<td>22,991</td>
<td>22,991</td>
<td>28,735</td>
<td>18,159</td>
<td></td>
</tr>
<tr>
<td>Individuals</td>
<td>5,650</td>
<td>5,650</td>
<td>5,650</td>
<td>5,650</td>
<td>5,650</td>
<td>5,650</td>
<td>6,186</td>
<td>5,172</td>
<td></td>
</tr>
<tr>
<td>Diff-in-Sargan p-value(^4)</td>
<td>0.227</td>
<td>0.203</td>
<td>0.163</td>
<td>0.293</td>
<td>0.245</td>
<td>0.672</td>
<td>0.550</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

\(^2\) In Column (4), the splines are a function of log base-year broad income. In all other columns, the splines are a function of log base-year taxable income. The spline coefficients give the marginal change from the previous spline coefficient.

\(^3\) This is a list of the predicted net-of-tax rate instruments used in each column. For example, Column (1) lists the instruments as two lags, three lags, and four lags. This means that predicted net-of-tax rate instruments are constructed for this column as a function of income two, three, and four periods before the base-year. The second two instruments in the list in each column are used to test whether the first instrument listed is exogenous using the Difference-in-Sargan test.

\(^4\) These are the p-values from testing whether a given income spline is exogenous.