THE ECONOMIC JOURNAL

The Economic Journal, Doi: 10.1111/ecoj.12299 © 2015 Royal Economic Society. Published by John Wiley & Sons, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

IDENTIFYING THE ELASTICITY OF TAXABLE INCOME*

Sarah K. Burns and James P. Ziliak

We use matched panels from the Current Population Survey along with a grouping instrumental variables estimator to provide new estimates of the elasticity of taxable income. Our identification strategy exploits the fact that federal and state tax reforms over the past three decades have differentially affected cohorts across states and over time. We find that the elasticity is in the range of 0.4–0.55. The implication of our new estimates for tax policy is that the revenue-maximising tax rate is nearly 30 percentage points lower than that obtained when we use the typical identification strategy in the literature.

Understanding how taxable income responds to marginal tax rates, that is, the elasticity of taxable income (ETI), has become a focal outcome for research and policy on the optimal design of the income tax and transfer system (Feldstein, 1995; Auten and Carroll, 1999; Saez, 2001; Gruber and Saez, 2002; Kopczuk, 2005, 2012; Heim, 2009; Blomquist and Selin, 2010; Brewer *et al.*, 2010; Giertz, 2010; Saez *et al.*, 2012; Kleven and Schultz, 2014; Weber, 2014). The ETI provides a summative measure of the income response to tax policy, whether due to behavioural labour supply changes, income composition changes, or changes in timing of income receipt. Moreover, under certain assumptions, the ETI can serve as a sufficient statistic for optimal tax analysis. For example, Saez *et al.* (2012) show that conditional on the shape of the income distribution, how income changes with changes in the marginal tax rate determines the revenue-maximising rate of taxation for high earners. Chetty (2009) notes further that even though the sufficiency of the ETI for marginal deadweight loss calculations breaks down if the marginal resource cost of sheltering income does not equal the tax rate, a weighted average of the ETI along with the total earned income elasticity is sufficient for marginal excess burden. Thus, pinning down the empirical magnitude of the ETI has important ramifications for the design of the tax system.

In this article, we present new evidence on the ETI, with particular emphasis on how the ETI varies with key economic and statistical decisions. We begin with the canonical ETI specification and identification scheme that regresses the change in log annual income on the change in the log net-of-tax share, defined as one minus the marginal tax rate, conditional on initial log income to control for possible regression-to-the-mean effects. Because the change in log net-of-tax rate is likely endogenous, most authors have adopted some variant of the identification scheme proposed by Gruber and Saez (2002) whereby the change in log net-of-tax share is instrumented with the

* Corresponding author: James P. Ziliak, Department of Economics and Center for Poverty Research, University of Kentucky, Lexington, KY 40506-0034, USA. Email: jziliak@uky.edu.

We thank two anonymous referees, the editor, Seth Giertz, Richard Blundell, Caroline Weber and seminar participants at the 2011 National Tax Association Meetings, the 2012 IZA Conference on Recent Advances in Labour Supply Modeling, Indiana University, and University of Wisconsin for helpful comments on earlier versions. All errors are our own.

predicted change in log net-of-tax share that would obtain if incomes grew from 1 year to the next solely due to inflation.

Our first point of departure with the prior literature is to propose a new approach to identifying the ETI based on a Wald-type grouping instrumental variables estimator (Heckman and Robb, 1985; Angrist, 1991; Blundell *et al.*, 1998). Several authors have examined the exogeneity of the individual-level synthetic tax instrument used in much of the literature (Moffitt and Wilhelm, 2000; Kopczuk, 2005; Blomquist and Selin, 2010; Giertz, 2010; Weber, 2014). Because the synthetic instrument is based on the individual's income at time $(t-1)$, it is likely to be correlated, at least weakly, with the model error term except under fairly stringent conditions. Our approach is to extend the Blundell *et al.* (1998) estimator to admit fixed growth-rate heterogeneity in incomes across cohorts defined by date-of-birth and education level and US states, and then to invoke the identifying assumption that tax reforms over the past three decades have differentially affected cohorts across states and over time. This assumption then permits us to construct a new instrument, which is the cohort-state-year mean of the synthetic tax rate. That is, in lieu of using the person-specific synthetic instrument, we average it up to the cohort-state-year level, thus attenuating the correlation between the original instrument and the idiosyncratic model error term.

We also distinguish our work from the prior literature with the use of two-year matched panels from the Current Population Survey (CPS).[1] With few exceptions (Moffitt and Wilhelm, 2000), the ETI literature uses some variant of taxpayer panel data. The advantage of tax panels over the CPS is the quality of data for measuring deductions and exemptions necessary to move from gross income to taxable income, coupled with the fact that most tax panels follow the same person for several years whereas the maximum panel length in the CPS is two years. However, this is weighed against limitations of tax data such as the fact that it is often not publically available, it generally has limited demographic information and it does not necessarily capture the low end of the distribution because many poor families have non-filing episodes. For example, in ETI papers it is standard to truncate low-income families from the sample, for example, with incomes below $20,000 in Auten and Carroll (1999), or $10,000 in Gruber and Saez (2002), under the assumption that this truncation is likely to impart little bias in the ETI. While some authors have examined how the elasticity changes when the truncation point changes, truncation has not been modelled formally. We utilise the extensive demographic information in the CPS to examine how the ETI is affected by the inclusion of controls for human capital and by formally testing for truncation bias (Barnow *et al.*, 1980; Heckman and Robb, 1985; Heckman *et al.*, 2006).

The estimates of the ETI using our new identification strategy range from 0.25 to 0.55 depending on measure of income and narrow down to a tighter range of 0.4 to 0.55 when using more exogenous controls for heterogeneous income trends defined at the cohort level. These latter estimates are 2–3 times larger than those obtained using the standard person-specific synthetic instrument, and are double the modal estimate

[1] To our knowledge we are the first to use matched CPS panels to estimate the ETI. Singleton (2011) uses a cross section of the CPS linked to administrative wage data from the IRS's Detailed Earnings. He attempts to identify the ETI through the marriage penalty relief provision contained in the 2001 Economic Growth and Tax Relief Reconciliation Act.

of 0.25 in the survey by Saez *et al.* (2012). We also find that controlling for additional demographics reduces the ETI by about 20%, and that this is more important than formal statistical controls for truncation bias caused by dropping low-income families from the analysis.

1. Estimation and Identification of the ETI

The canonical approach to estimating the effect of taxation on labour supply is to assume that a taxpayer maximises a utility function over a composite consumption good c and hours of work h, $U(c, h)$, subject to a budget constraint of $c = wh + V + N - T(wh + N)$, where V *is* non-taxable non-labour income, N is taxable non-labour income, w is the pre-tax hourly wage rate, $T(.)$ is the tax function and the price of consumption has been normalised to 1. Solving the optimisation problem results in an optimal hours of work function of $h[w(1 - \tau), N^v]$, where τ is the marginal tax rate and N^v is virtual non-labour income $N + V + \tau wh - T(.)$, which is that level of compensation needed to make the worker behave as if they faced a constant marginal tax rate on all taxable income. In this framework, both the after-tax wage and virtual non-labour income are treated as endogenous in estimation since the tax rate an individual faces is an implicit function of hours of work. Feldstein (1995) argued that this approach missed other behavioural responses to tax law changes such as shifting compensation from taxable to non-taxable income, or changes in the timing of compensation. Instead, he posited workers preferences over consumption and an income supply function, y, $U(c, y)$, which coupled with the budget constraint, results in an income supply function of $y(1 - \tau, N^v)$ that depends on the net-of-tax share $(1 - \tau)$ and virtual non-labour income. Like the labour supply predecessor, both the net-of-tax share and virtual incomes are treated as endogenous in estimation.

Gruber and Saez (2002) extended the Feldstein approach by motivating the income supply model within the context of the Slutsky equation in elasticity form, which relates how income supply responds to infinitesimal changes in net-of-tax shares and captures both substitution and income effects of tax law changes. For the empirical counterpart of their model they replaced the continuous time derivative from the Slutsky equation with a discrete time change from period $t - 1$ to t:

$$\Delta \ln y_{it} = \beta \Delta \ln(1 - \tau_{it}) + \gamma \Delta \ln N_{it}^{v} + \epsilon_{it}, \tag{1}$$

where $\Delta \ln y_{it} = \ln y_{it} - \ln y_{it-1}$, $\Delta \ln(1 - \tau_{it}) = \ln(1 - \tau_{it}) - \ln(1 - \tau_{it-1})$ and $\Delta \ln N_{it}^{v} = \ln N_{it}^{v} - \ln N_{it-1}^{v}$.[2] In log first difference form β is the compensated ETI. As Gruber and Saez found that γ was near zero, or that income effects were small, most of the subsequent literature has ignored income effects in their empirical applications and thus remain silent on distinguishing whether the ETI reflects compensated or uncompensated effects, though it is generally assumed it is the former such that the null hypothesis is $\beta \geq 0$. We follow the recent work and ignore income effects for the ETI model.

[2] The theoretical model of Gruber and Saez (2002) refers to virtual non-labour income, but for the empirical counterpart they use after-tax income.

The actual empirical model estimated in the literature is more akin to:

$$\Delta \ln y_{it} = \beta \Delta \ln(1 - \tau_{it}) + \delta f(y_{it-1}) + \boldsymbol{x}_{it}\boldsymbol{\theta} + \mu_t + \epsilon_{it}, \tag{2}$$

where $f(y_{it-1})$ is some function of base-year income such as the log of income or a spline in income to control for mean reversion in income growth as well as trends in inequality, $\boldsymbol{x}_{it}$ is a vector of demographics, μ_t is a control for aggregate time effects such as a linear trend or time dummies, and $\epsilon_{it} \sim \mathrm{iid}(0, \sigma_\epsilon^2)$. Because the marginal tax rate is a function of the person's income, the standard OLS assumption that $\mathrm{E}[\Delta \ln\ (1 - \tau_{it})\epsilon_{it}] = 0$ is likely to be violated, and thus it is necessary to instrument for the endogenous regressor. Gruber and Saez (2002) propose a just-identified model based on the instrument $\Delta \ln(\widehat{1-\tau_{it}}) = \ln(1 - \hat{\tau}_{it}) - \ln(1 - \tau_{it-1})$, where $\hat{\tau}_{it}$ is the marginal tax rate that the individual would face in year t if income in year t differed from its $t - 1$ value only by an inflation adjustment. This synthetic marginal tax rate instrument is valid provided that it only reflects changes in tax law.

1.1. *Cohort-based Identification*

Several authors have raised concerns over the exogeneity of the synthetic tax instrument and have examined the sensitivity of the estimated ETI to alternative identification strategies (Moffitt and Wilhelm, 2000; Kopczuk, 2005; Blomquist and Selin, 2010; Giertz, 2010; Weber, 2014). The overarching concern is whether or not controlling for $f(y_{it-1})$ is sufficient to purge the model of any correlation between the synthetic instrument and the error term ϵ_{it}. For example, Blomquist and Selin (2010) and Weber (2014) propose more robust alternatives that lead to larger estimates of the ETI than with standard synthetic instruments. Blomquist and Selin instrument the 10-year change in the log net-of-tax share from 1981 to 1991 with a predicted change in net-of-tax share. Specifically, they regress taxable income in 1981 and 1991 on taxable income in 1986, along with a host of demographic variables, and construct the instrument as $\Delta \ln(\widehat{1-\tau_{it}}) = \ln(1 - \hat{\tau}_{i1991}(\hat{y}_{i1991})) - \ln(1 - \hat{\tau}_{i1981}(\hat{y}_{i1981}))$. Using reduced-form predictions based on demographics has its roots in the static labour supply literature (Hausman, 1981; Mroz, 1987; MaCurdy *et al.*, 1990; Blundell and MaCurdy, 1999; Keane, 2011), and, as shown. In Griliches and Hausman (1986), a wide difference like a 10-year change reduces the role of measurement error. With standard iid assumptions on the error term, using information between years (such as the midpoint of 1986) is a valid instrument. Weber (2014), on the other hand, identifies the change in the net of tax share using further lags in taxable income to construct the instrument. The idea here is that provided that the idiosyncratic error term ϵ_{it} has limited memory, for example, MA(1) or MA(2), then lags of taxable income dated y_{t-k-1}, where k is the moving average order, are valid to construct instruments. This is akin to the panel-based labour supply literature that uses lagged after-tax wages to estimate the wage elasticity of labour supply (Ziliak and Kniesner, 1999, 2005).

We propose a new approach to identifying the ETI that builds on the grouping instrumental variables estimator that also traces its roots to the labour supply literature (Heckman and Robb, 1985; Angrist, 1991; Blundell *et al.*, 1998). The idea of the grouping estimator is to purge correlation between the idiosyncratic error and the endogenous regressor by assigning individuals to groups and to use group-specific means as instruments instead of person-specific information. Specifically, consider the following restatement of (2):

$$\Delta \ln y_{icjt} = \beta \Delta \ln(1 - \tau_{icjt}) + \delta f(y_{icjt-1}) + x_{icjt}\theta + \gamma_c + \tau_j + \mu_t + \epsilon_{icjt}, \tag{3}$$

where $c = 1, \ldots, C$ is the cohort that individual i belongs, $j = 1, \ldots, J$ is the state of residence, γ_c is a fixed cohort effect, and π_j is a fixed state effect. With this model specification we invoke the following assumptions:

$$\mathrm{E}[\epsilon_{icjt}|c, j, t] = \gamma_c + \tau_j + \mu_t, \tag{A.1}$$

$$\{\mathrm{E}[\Delta \ln(1-\tau_{icjt})|c,j,t] - \mathrm{E}[\Delta \ln(1-\tau_{icjt})|c] - \mathrm{E}[\Delta \ln(1-\tau_{icjt})|j] - \mathrm{E}[\Delta \ln(1-\tau_{icjt})|t]\}^2 \neq 0. \tag{A.2}$$

Assumption A.1 implies the exclusion restrictions for identification are that unobservable differences in changes in average taxable income across groups can be summarised by a permanent cohort effect (γ_c), a permanent state effect (π_j) and an additive time effect (μ_t). Assumption A.2 states that changes in the net-of-tax share grow differentially across the groups and is equivalent to a rank condition for identification. It requires that variation remains in net-of-tax share changes after controlling for fixed cohort, state and time effects. We follow Blundell *et al.* (1998) and define a cohort based on date of birth and education attainment. Because our data span several decades and multiple tax reforms, various date-of-birth cohorts face different tax systems, and since tax reforms differentially affected those at different points of the income distribution, we further separate cohorts by education attainment, which is a proxy for permanent income (Blundell and MaCurdy, 1999; Attanasio and Weber, 2010).

Assumptions A.1 and A.2 offer a set of exclusion restrictions for identification; namely, a full interaction of cohort-state-time effects. That is, in the structural (3) the fixed cohort, state, and time effects enter linearly, which means in the reduced-form prediction of the change in net-of-tax share we can exploit non-linearities by interacting all of the cohort, state, and time effects. This results in a heavily over-identified model because of our data (described below) on persons in one of 39 cohorts, residing in the 50 states plus the District of Columbia, and 27 years.[3] These interactions reflect any changes that affect cohorts differentially across states and time, not just tax policy, and while this 'black box' form of identification may be powerful, it deviates from the literature's focus on using specific tax policy reforms to estimate the ETI. Consequently, instead of the full interactions we construct a single instrument to just identify the model as is typical in the literature. We take advantage of the fact that tax rates vary across cohorts, states, and time and thus construct a synthetic instrument based on cohort-state-year mean change in the log net-of-tax share. Specifically, we take the individual-level Gruber–Saez instrument above, $\Delta \ln(\widehat{1-\tau_{it}})$, and compute the

[3] This implies a potential of 53,702 cohort-state-year overidentifying restrictions if all cohorts are observed in all 27 years and in all 50 states and DC. Our sample is unbalanced, but we do have 27,979 cohort-state-year groups.

mean value for each cohort-state-year cell, $\overline{\ln\{(1-\hat{\tau}_{it})/(1-\tau_{it-1})\}}_{cjt}$. Under assumption A.1 this instrument is valid because the correlation between the group mean tax rate and the idiosyncratic error term is likely to be negligible, and provided that the federal and state tax policy changes over the past three decades affected cohorts differently then assumption A.2 is satisfied. For completeness, in Section 4 below we note estimates from the over-identified model.

Our model extends the Blundell *et al.* (1998) grouping IV estimator in several ways. First, since their empirical application was to married women's labour supply in the United Kingdom, they only controlled for cohort and time effects. Because there are important differences in tax policy across US states over and above federal tax reforms that can aid in identification, our estimator incorporates this additional state-level heterogeneity. Second, Blundell *et al.* (1998) specified their model in levels and not changes in part because they derived their estimator for the explicit purpose of repeated cross-section data since they did not have individual-level panel data. With panel data on a large number of cross-sections we can combine individual panel heterogeneity along with cohort heterogeneity that is not possible in repeated cross-sections without strong assumptions on the dynamics of cohorts (Moffitt, 1993; Verbeek and Vella, 2005). Our estimator in fact captures a greater extent of unobserved heterogeneity in levels that has been a concern in the ETI literature.

To see this consider the levels analogue to (3) (ignoring controls for base-year income trends and demographics for simplicity):

$$\ln y_{icjt} = \beta \ln(1-\tau_{icjt}) + \gamma_{c0} + \gamma_c t + \pi_{j0} + \pi_j t + \widetilde{\mu}_t + u_{icjt}, \tag{4}$$

where (γ_{c0}, π_{j0}) are fixed cohort and state effects affecting taxable income levels, $\gamma_c t$ and $\pi_j t$ are time-varying cohort and state effects that capture heterogeneous growth rates in incomes, and $\tilde{\mu}_t$ is a renormalised time effect. The error term $u_{icjt} = \rho_i + e_{icjt}$ contains person-specific and time-invariant unobserved heterogeneity, ρ_i, along with an error e_{icjt} assumed to be MA(1) such that $\Delta e_{icjt} = \epsilon_{icjt}$.[4] Under these assumptions, if we take first differences of (4), the time invariant factors ρ_i, γ_{c0} and π_{j0} each drop out. However, we are still left with the unobserved heterogeneous trend factor loadings γ_c and π_j because $\pi_c = \pi_c t - \pi_c(t-1)$ and $\pi_j = \pi_j t - \pi_t(t-1)$. It is the unobserved heterogeneity in cohorts and states in income growth that is missing in the labour supply setting of Blundell *et al.* (1998) but which is likely important in the ETI context. That is, the model in (3) is robust to person-specific unobserved heterogeneity found in panel studies of the ETI, as well as cohort- and state-level growth-rate heterogeneity that generally has not been found in either panel or repeated cross-section studies.

We note that there is nothing about our grouping estimator that requires the use of the CPS and indeed it can be readily implemented in taxpayer panel data as well. As described in Section 3 we define our cohorts based on year of birth and education attainment but, because education is not available in most tax panels, at least in the US, we also show how the estimates vary if cohort is redefined to only include birth year.

[4] π_{j0} enters the model as a fixed effect under the assumption that individuals do not change states across survey years, which is a requirement to be followed in the CPS. We discuss this further in Section 3.

1.2. *Robustness to Demographics and Truncation*

Taxpayer panels generally are limited by a parsimonious set of demographic controls such as the number and ages of adults and children, along with marital status. However, if taxable income, and the attendant response to tax changes, is also determined by other demographics such as race, gender and education attainment, which are standard in Mincer-type wage equations (Heckman *et al.*, 2006), then the conditional mean assumption underlying (3), $E[\epsilon_{icjt} \mid \Delta\ln(1-\tau_{icjt}), f(y_{icjt-1}), x_{icjt}, \gamma_c, \pi_j, \mu_t] = 0$, will be violated and the estimated ETI $\hat{\beta}$ will suffer from omitted variables bias. With an expanded set of demographics available in the CPS we test whether the ETI is robust to inclusion of these potential confounders. In our baseline models, we take a parsimonious approach and enter them linearly via x_{icjt}, though we also explore the robustness to more complicated non-linear control functions $g(x_{icjt})$, specifically a full interaction of the demographics as well as propensity scores (Barnow *et al.*, 1980).

Another threat to consistent identification of the ETI in (3) is possible truncation bias. The typical paper in the literature truncates the year-one income below some threshold – \$20,000 in Auten and Carroll (1999), \$10,000 in Gruber and Saez (2002) – and assumes that the data below the threshold are missing at random (conditional on marital status and base-year income). Under the assumption of normally distributed errors we can write the truncated conditional mean of changes in log taxable income as:

$$E(\Delta \ln y_{icjt}|y_{icjt-1} > \varphi) = \beta\Delta\ln(1-\tau_{icjt}) + \delta f(y_{icjt-1}) + x_{icjt}\theta + \gamma_c + \pi_j + \mu_t + \sigma\lambda_{icjt}, \quad (5)$$

where

$$\lambda_{icjt} = \frac{\phi\left[\frac{(\varphi - z'_{icjt}\boldsymbol{\Gamma})}{\sigma}\right]}{1-\Phi\left[\frac{(\varphi - z'_{icjt}\boldsymbol{\Gamma})}{\sigma}\right]}$$

is the inverse Mills ratio for the normal distribution, φ is the truncation point (e.g. \$10,000), z'_{icjt} is the vector of right hand side variables in the regression, $\boldsymbol{\Gamma}$ is the vector of parameters, and ϕ, Φ are the pdf and cdf of the normal distribution respectively (Wooldridge, 2002). In general the omission of the last term in (5) will lead to biased estimates of $\hat{\beta}$, and the proposed grouping instrument described in the last Section may not be sufficient to eliminate the bias. This assumption of no truncation bias implicitly assumes that changes in labour force composition in response to tax reforms will not affect the ETI, nor will changes in composition owing to years of non-filing among low-income families whose incomes tend to be highly volatile and increasingly so over the past three decades (Hardy and Ziliak, 2014). However, Meyer and Rosenbaum (2001) attributed upwards of 60% of the increase in labour force participation of single mothers in the 1990s to expansions in the EITC. Many of these women do not work full time, and yet are quite responsive to tax policy, and thus could affect estimates of the ETI. To our knowledge this assumption has not been tested formally in the literature (though some test the robustness of results to alternative thresholds).

The conditional mean of the truncated normal model in (5) can be relaxed so that the mechanism that determines whether income is above the threshold is different from the dollar amount, which is sometimes known as a generalised tobit in labour supply models and is functionally similar to the Heckman two-step selection model (Heckman, 1979; Zabel, 1993; Wooldridge, 2002). This means that we do not need to use the same set of covariates to determine the truncation point and taxable income changes and thus can use additional exclusion restrictions to identify the inverse Mills ratio and not just non-linearity in the functional form. In implementing (5) we estimate a reduced-form probit of whether or not income exceeds \$10,000 so that it is a function of the exogenous demographics, fixed state, cohort and time effects, along with state macroeconomic and policy variables. With the control for truncation, assumptions A.1 and A.2 are modified as:

$$\mathrm{E}(\epsilon_{icjt}|c, j, t, \phi) = \gamma_c + \pi_j + \mu_t + \sigma\lambda_{cjt}(\varphi), \tag{A.1$'$}$$

$$\begin{aligned}&\{\mathrm{E}[\Delta\ln(1-\tau_{icjt})|c, j, t, \varphi] - \mathrm{E}[\Delta\ln(1-\tau_{icjt})|c, \varphi] - \mathrm{E}[\Delta\ln(1-\tau_{icjt})|j, \varphi] \\ &- \mathrm{E}[\Delta\ln(1-\tau_{icjt})|t, \varphi] - \sigma_\tau\lambda_{cjt}\}^2 \neq 0,\end{aligned} \tag{A.2$'$}$$

where identification of the ETI from (A.2′) now requires that net-of-tax shares change differentially across cohorts, states and time, as well as over changes in sample composition above the threshold φ. Our approach below is to first compare the baseline estimates to a model that includes all individuals with positive incomes, regardless of level and then to a model that formally controls for the truncation process under assumptions A.1′ and A.2′.

2. Data

The primary economic and demographic information comes from the Annual Social and Economic Supplement of the CPS for calendar years 1979–2008 (interview years 1980–2009), which is fielded in March of each year. The March CPS contains rich data on labour and non-labour income as well as detailed family demographics – including those relevant for tax purposes (marital status, dependents, etc.). Our sample consists of married and unmarried heads of households ages 25–60 (with and without children or other dependents) who are assigned into one of thirteen 5-year birth cohorts and three education levels (less than high school, high school only and more than high school) for a total of 39 five year birth by education cohorts. There is a trade-off between heterogeneity and measurement error in constructing cohorts (Deaton, 1997) – too few cohorts increases the odds that unmeasured heterogeneity will plague estimates, while too few observations per cohort increases the odds that measurement error may affect estimates because of outliers. To address the former concern Blundell *et al.* (1998) use 10-year birth cohorts and two education groups but because the CPS is a larger survey we are able to allow greater heterogeneity with shorter birth windows and an additional education category. For the latter concern we follow Blundell *et al.* (1998) and drop cohort-year cells with fewer than 50 observations. Table A1 shows that this results in a loss of 20 cohort-years out of a possible 444, or 4.5%, but this amounts to just 0.3% of total observations. We test this assumption below and find our results are robust.

2.1. *Panel Data in the CPS*

The CPS employs a rotating survey design so that a respondent is in sample for 4 months, out 8 months, and in another 4 months. This makes it possible to match approximately one-half of the sample from one March interview to the next, and thus to create a series of two-year panels over our sample period.[5] Following the recommended Census procedure we perform an initial match of individuals on the basis of five variables: month in sample (months 1–4 for year 1, months 5–8 for year 2); gender; line number (unique person identifier); household identifier; and household number. We then cross check the initial match on three additional criteria: race, state of residence and age of the individual. If the race or state of residence of the person changed we delete that observation and, if the age of the person falls or increases by more than two years (owing to the staggered timing of the initial and final interviews), then we delete those observations on the assumption that they were bad matches. These additional criteria were important prior to the 1986 survey year but thereafter the five base criteria match most observations. In concordance with the literature, we exclude individuals whose marital status changes from one year to the next as large changes in income unrelated to tax policy are expected for this group. There were major survey redesigns in the 1980s and 1990s so it is not possible to match across the 1985–6 waves and the 1995–6 waves. This yields a matched time series across 29 years with gaps in calendar years 1984–5 and 1994–5.

A possible concern with matched CPS is with sample attrition affecting our income series. The CPS sample domain is household addresses and not individuals, so that if a person moves between March surveys then the Census Bureau interviews the new occupant at the address and does not follow the original respondent. Under the assumption that the probability of attrition is unobserved and time invariant (i.e. a fixed effect), then the first-differenced model in (3) will remove the latent probability of attrition and our estimates will be purged of possible attrition bias. Ziliak and Kniesner (1998) found evidence in favour of this assumption in their study of the effects of income taxation on life cycle labour supply. If there is time-variation in the attrition process but it is only trending linearly across cohorts and states, then our controls for cohort and state effects in (3) will capture this influence. However, if there is time-variation in the factor loading on the unobserved individual-level heterogeneity then differencing will not eliminate potential attrition bias unless the factor loading is randomly distributed across the population. A conservative interpretation then is that data from matched CPS provides estimates of the ETI among the population of non-movers.

2.2. *Income and Tax Data*

We use two variants of income for the dependent variable akin to those used in much of the ETI literature. The first, broad (gross) income is defined as total family income less social security income. Total family income includes most components of income

[5] Some studies use three-year differences rather than one-year differences used here. The use of one year differences may result in elasticities reflecting more income-shifting behaviour but, given the structure of the CPS design, it is not possible to examine three-year differences.

reported on Form 1,040 such as earnings of the head (and spouse if present) as well as unemployment compensation, worker compensation, social security, public assistance, retirement benefits, survivor benefits, interest income, dividends, rents, child support, alimony, financial assistance and other income. Gruber and Saez (2002) exclude social security income and capital gains owing to their differential tax treatment over the 1980s, and we do so as well.

Our second, and narrower, income definition is taxable income defined as broad income less estimated exemptions and deductions which are obtained from NBER's *TAXSIM* program. A disadvantage of CPS data relative to tax panel data is that information on many deductions (e.g. home mortgage interest expense, moving expenses, charitable contributions and medical expenses) used to arrive at adjusted gross income (AGI) and taxable income are not collected.[6] While many of these are typically omitted in the literature in order to achieve a consistent definition over the years, our measure of taxable income will be broader than those calculable from tax panel data.[7] We therefore expect our estimates of taxable income elasticities to be on the low end of the literature.[8] Our broad income measures, however, should be quite comparable to those used in the past literature. All income data are deflated by the personal consumption expenditure deflator (PCE) with 2008 base year. Following Gruber and Saez (2002), in our baseline sample we drop observations with year-one real broad income less than $10,000. In addition, because taxable income may be zero or negative, the log of which is undefined, we restrict our sample to broad income greater than $10,000 and taxable income greater than zero in order to have identical sample sizes across income definitions. We then test how the ETI is affected by this truncation as described in subsection 1.2.

Prior to our matching across waves, we delete those observations with imputed income as Bollinger and Hirsch (2006) show that such imputed data may impart bias in regression coefficients. That is, if a respondent does not know their income or refuses to report their income, then the Census uses what is called a 'sequential hot deck' procedure to assign income to the individual. That is, they match the non-respondent to a respondent on the basis of 12 demographic characteristics and assign the income of the respondent to the non-respondent (Welniak, 1990). Bollinger and Hirsch (2006) demonstrate that such a process can bias not only the coefficients on matched

[6] There is evidence to suggest that the self employed have an incentive to under-report income to tax authorities and, while such incentives should not hold for reporting to household surveys, Hurst *et al.* (2014) provide some evidence that income under-reporting in the Consumer Expenditure Survey and Panel Study of Income Dynamics among the self employed is comparable to under-reporting rates in tax return data. Although there is no evidence on this issue in the CPS that we are aware of, we assume that income reports among the self employed are no better or worse in the CPS than in tax return data.

[7] For instance, we calculate the ratio of taxable income to broad income in our sample to be roughly 85% while studies based on tax panels tend to have lower ratios ranging from 60% to 70% (Gruber and Saez, 2002; Heim, 2009; and Giertz, 2010).

[8] TAXSIM estimates whether the tax unit itemises based on estimated state tax liability (which is deductible for itemisers in the US) and other deductions such as home mortgage interest payments, child care expenses and charitable contributions. Because the CPS does not collect information on these other deductions, whether or not a family is estimated to itemise largely depends on state tax liability. Across our sample period, we estimate that about one quarter of the families itemise, which is 6–10 percentage points lower than reported by the IRS in a typical year (See http://www.taxpolicycenter.org/taxfacts/displayafact.cfm?Docid=173).

variables but also other non-match variables and recommend that imputed observations be deleted.

In a procedure that is distinct from imputation, the Census Bureau also top codes the incomes of high-income earners in a bid to ensure confidentiality. The method of top-coding has varied over the years, complicating analyses of income inequality and potentially this article as well. From 1976–87 Census collected income from 11 sources for the CPS and, starting in 1988 onward, they collect income from 24 sources. Each income source has its own top code value and thus broad income may include some income sources that are top coded and some that are not. The top-code value was a fixed dollar threshold until 1996 when Census started using the mean income of top-coded individuals within cells (determined by up to 12 demographic variables). For example, if in 1995 a person reported \$500,000 in earnings, then the Census recorded the earnings of that person as \$150,000. In 1996, that same person earning \$500,000 would have recorded the mean earnings of all persons within their demographic cell. This clearly creates the possibility of a jump discontinuity that could affect research with the CPS. To mitigate this problem, Larrimore *et al.* (2008) were granted access to the internal CPS data at Census to construct a cell-mean top code series for 1976–95 to be used in conjunction with the post-1995 Census procedure and this series is made publically available.[9] Consequently, before matching we incorporate this series of consistently derived income top codes. Note that these cell means change over time and thus for individuals who are top coded in both years we will register a change in their earnings based on the change in the mean value within the cell that they belong. Burkhauser *et al.* (2012) find that using the consistent top code method results in CPS measures of income inequality tracking closely those from proprietary tax return data and much better than unadjusted public-use CPS data. Top coding affects 2.7% of our sample, so as a check on the sensitivity of the ETI to top coding, in the robustness section we drop the entire time-series of an observation if they are in the top 5% of the income distribution in any year.

Tax rates at the federal and state level are estimated for each family in each year using the NBER *TAXSIM* program in conjunction with basic information on labour income, taxable non-labour income and dependents. The federal and state taxes include the respective EITC code for each tax year and state, thus allowing for the possibility of negative tax payments. As discussed in Section 2, because the change in net-of-tax shares is endogenous to the change in income, we instrument the actual change in tax rates with a predicted tax change, $\Delta\ln(\widehat{1-\tau_{it}})$. To obtain $\hat{\tau}_{it}$, we inflate each individual's year one income by the increase in the PCE and run it through *TAXSIM* as year two synthetic income. For the instrument in the grouping IV model, we compute the mean value of the synthetic tax rate for each cohort-state-year.

Our sample selection process results in 198,285 longitudinally linked observations over the 1979–2008 calendar years. Summary statistics for the matched CPS data are shown in Appendix Table A2.

[9] Since Census did not release the cell-mean top code values for 1996–2002, Larrimore *et al.* (2008) also construct these cell means for the series. See also https://cps.ipums.org/cps/inctaxcodes.shtml for information.

Figure 1 shows the life-cycle net of tax rates plotted for the 13 different birth cohorts by education level. It is clear that cohorts in the lowest education level have the highest net of tax shares (i.e. face the lowest marginal tax rates), while cohorts in the highest

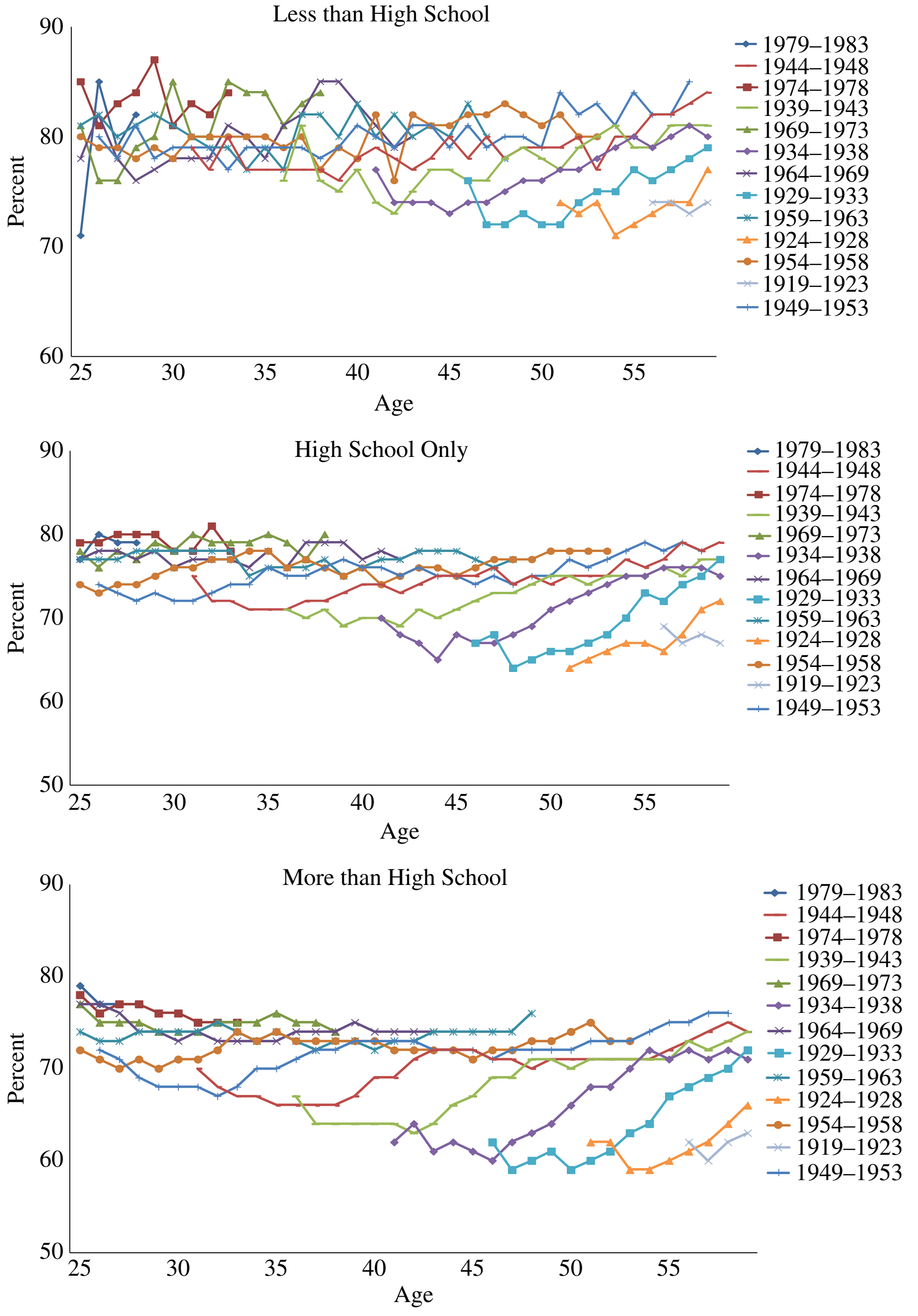

Fig. 1. *Life-cycle Net of Tax Rates for the 5-year Birth-year Cohorts By Level of Education*

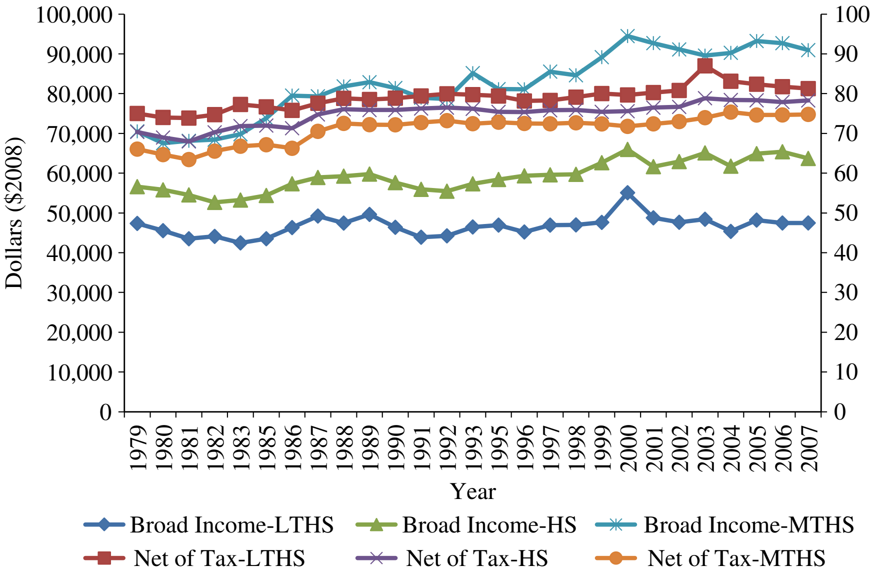

Fig. 2. *Trends in Real Broad Income and Net of Tax Shares by Education Group*

education group face the lowest after-tax shares. This is consistent with progressive taxation assuming income is rising with education attainment. Evidence of variation within the education groups is also present, which is necessary for identification of the grouping estimator (see also Table A3 documenting the year-to-year variation in our cohort-state-year instrument). For example, among birth cohorts with more than high school (3rd panel in Figure 1), at age 40 the more recent cohorts face a higher net-of-tax share, which is consistent with tax reforms reducing marginal tax rates. Moreover, there appears to be a life-cycle trend of rising net of tax shares, which in this case appears more pronounced for the older cohorts.

Figure 2 depicts trends in the levels of real broad income (left axis) and the net of tax shares (right axis) aggregated across the state and year-of-birth cohorts but separate by education attainment and year. The Figure makes clear that broad income was relatively stable for the high school and less-than-high school samples but shows secular growth among those with more than high school, a fact widely documented in the inequality literature. The Figure also shows that net of tax shares increased the most in the 1980s after the 1986 tax reform and that broad income appears to increase in periods where the net of tax share increases. That said, there are also changes in income that do not coincide with changes in net of tax shares, especially in the late 1980s and again in the late 1990s among the more than high school group. In the former period it appears in a couple of years that broad income changes before the net of tax shares, while in the latter period it appears there might be time variation in income growth. It is important to note, however, that no causal relationship can be inferred from these simple time-series figures for the reasons detailed in the methods Section above; namely, the levels in Figure 2 ignore latent heterogeneity at the person, cohort- and state-level, likely regression-to-the mean effects and the endogeneity of the net-of-tax share. This presages the need in our empirical model as presented in (3) to transform into first differences, to instrument for the net-of-tax share and to control for base-year income and cohort and state heterogeneity in income growth. In Section 4 we also test the robustness of the estimated ETI to a variety of possible confounding influences, including time-variation in cohort growth-rate heterogeneity.

Before proceeding, however, we present some graphical evidence on the causal relationship between our proposed cohort-state-year instrument and broad-income changes. Figure 3 contains the time-series of changes in log broad income along with changes in the cohort-state-year instrument for net of tax shares by education group. This differs from Figure 2 in several ways. First, Figure 3 is in log first differences and not levels. Second, we partial out the effect of base-year income trends by regressing the change in log broad income on the 10-piece spline in base year log income and a series of 27 year dummies and then plot the estimated coefficients on the year dummies in Figure 3. Note that since we do not include a constant term in the regression, the scale on the left-axis also reflects the income spline. Third, because year-to-year changes are noisy, we plot a lowess smoothing line of each time series.[10] The Figure reveals a clear positive relationship between income changes and net of tax share changes, with the latter leading the former. As noted in Saez *et al.* (2012, p. 25), in the context of a difference-in-difference identification scheme, if each of the groups faces tax reforms then consistency of the estimator hinges on the elasticities being the same across groups. A simple bivariate time series regression of change in broad income on change in net of tax shares for each panel in Figure 3 yields estimated ETIs of 0.24, 0.20, 0.27, which are qualitatively very similar and statistically the same (i.e. we cannot reject the null hypothesis that they are equal), and thus lending confidence in the consistency of our identification strategy. Moreover, we improve on this graphical strategy with our empirical model in (3) by pooling these groups together and controlling for other observed and unobserved heterogeneity, which combines identification from both the panel ETI literature and difference-in-difference literature.

A casual glance at the time–series relationship between the instrument and outcome in Figure 3 appears that it is stronger in the 1980s than in more recent decades. We explore this in more detail in Figure 4 where we use an approach similar in spirit to Weber (2014) and present local polynomial regressions of the residual change in log broad income on the residual change in the cohort-state-year net-of-tax share instrument.[11] The residuals are obtained by regressing the change in log of broad income (or the change in the log cohort-state-year instrument) on the 10-piece spline in log base-year income, thus netting out heterogeneous income trends. The Figure depicts four residual plots along with 95% confidence bands – one pooling all changes across all years, one using the post-1986 tax reform from 1986 to 1987, one for the post-1993 tax change from 1995–6[12] and one

[10] The lowess smoother is implemented in Stata as a running-line locally-weighted least squares regression. The bandwidth is set at 0.5, and is robust to other bandwidth choices in the same neighbourhood, noting that the smoothness of the curve increases in the bandwidth parameter.

[11] Figures 4 and 5 are produced using the lpolyci graph command in Stata. The third-order local polynomial regression is estimated with an epanechnikov kernel and bandwidth parameter of 0.16 as in Weber (2014). The samples in the Figures consist of 98% of the observations on average, as the extreme outliers exert undue influence on the polynomial estimator. As such, the ranges of the axes vary with each panel in the Figures since they draw from different time periods.

[12] The 1993–4 change is not possible owing to the missing year of 1994 in the matched CPS.

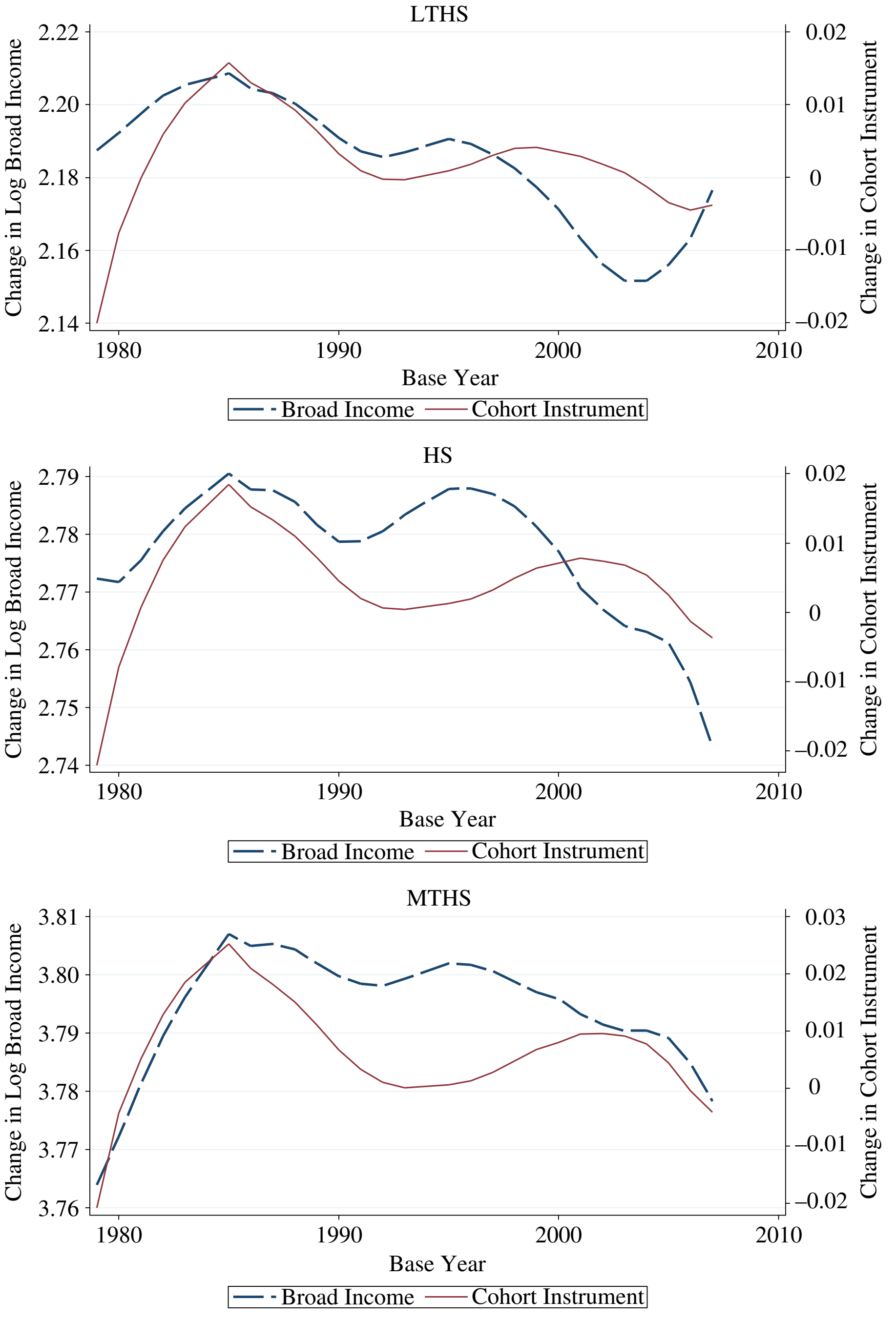

Fig. 3. *Trends in Change in Broad Income and Cohort-tax Instrument*

for the post-2003 tax change from 2003 to 2004. The first panel in Figure 4 shows that across all years and tax reforms the average change in log broad income is increasing in the change in the cohort-state-year instrument, as we require for identification of the

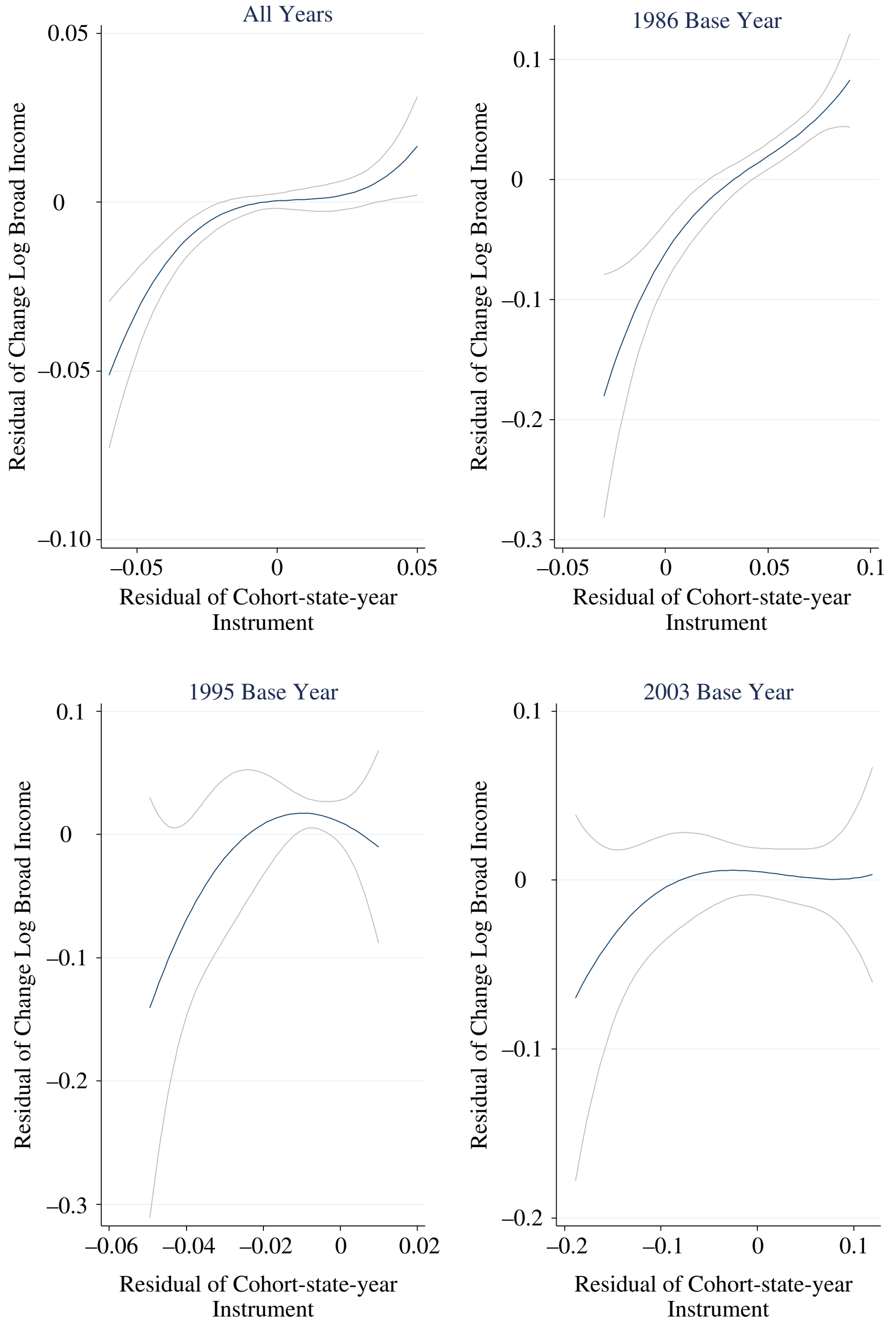

Fig. 4. *Average Change in Log Broad Income by Change in Cohort Instrument*

ETI. This is most evident in the second panel for the change immediately after the 1986 tax reform. The flattening out of the relationship in the first panel around zero change in the instrument seems to emanate from later tax changes in the 1990s and 2000s, as seen in panels three and four of the Figure. This is not necessarily because the relationship between the net-of-tax share regressor and instrument gets weaker over time. Indeed, Figure 5 presents parallel plots to those in Figure 4 but focuses

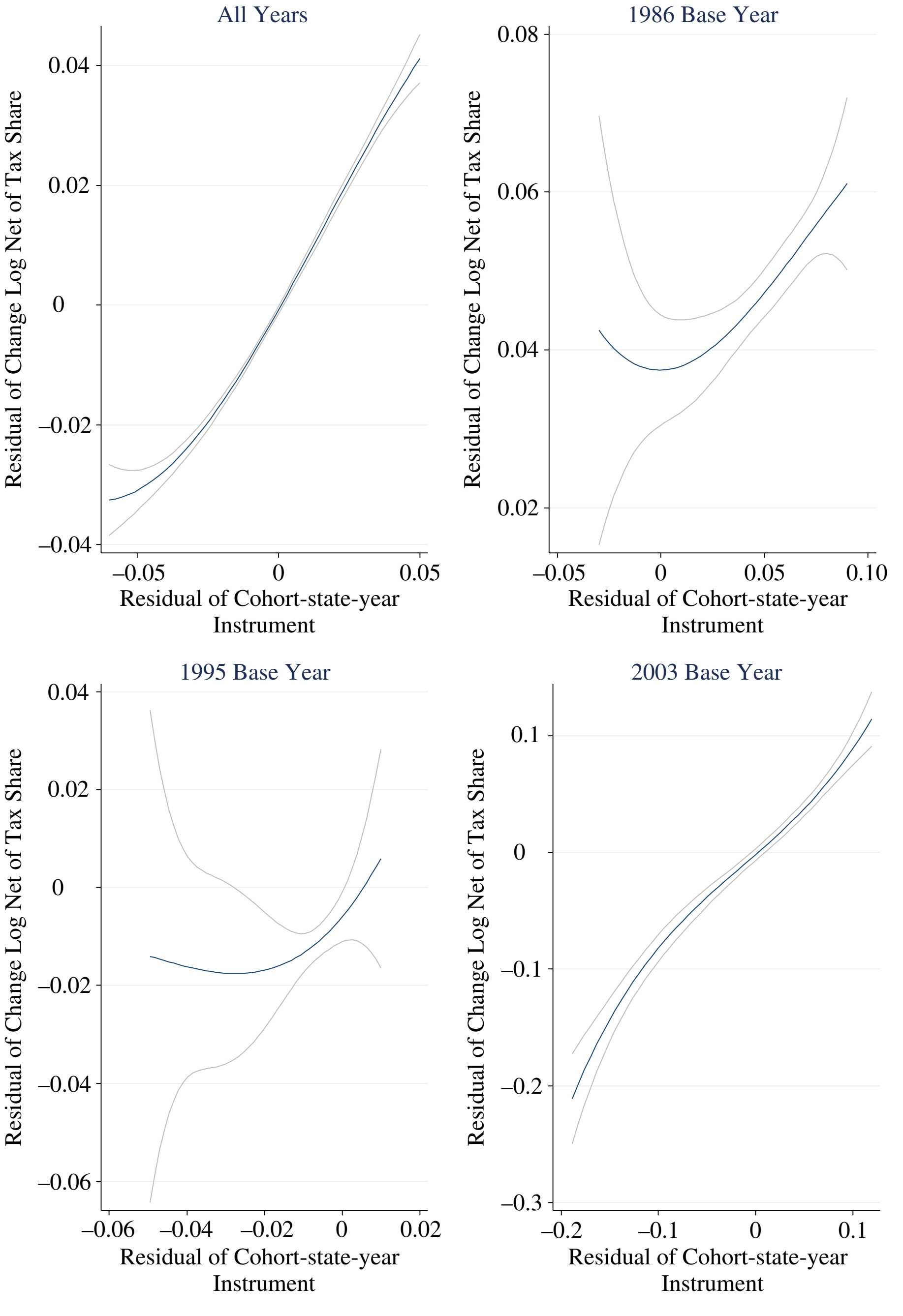

Fig. 5. *Average Change in Log Net-of-tax Share by Change in Cohort Instrument*

instead on the residual changes in the actual log net-of-tax share (i.e. the regressor in the model) and the cohort-state-year synthetic net-of-tax share (i.e. the instrument in the model). As the first panel depicts, the relationship between the endogenous regressor and instrument is strong across the entire distribution of the instrument across all years and tax reforms, which is crucial for identification of our empirical model. While there is some evidence that it weakens in the mid 1990s, it is again quite strong in the 2000s. Because of this possible change over time, in the robustness section below, we more formally examine how the estimates of the ETI change by decade.

3. Results

Our first objective is to replicate the baseline results from Gruber and Saez (2002) using matched two-year samples from the CPS along with the canonical ETI specification and person-specific synthetic instrument identification strategy in (2). This will inform us whether and how estimates of the ETI using CPS data differ from tax panel estimates. We then compare this estimate with our grouping instrumental variables estimator from (3) where we aggregate the person-specific synthetic instrument to the cohort-state-year level. All instrumental variables regressions control for marital status, state fixed effects and time fixed effects for initial year and are weighted by year one broad income.[13] The grouping IV models also include cohort fixed effects as regressors. The standard errors are robust to heteroscedasticity and are clustered at the state- and year-level as suggested by Cameron *et al.* (2011).[14]

The baseline estimates are reported in Table 1. The Table has two panels corresponding to the control for mean reversion: one with a 10-piece spline in year-one log income and the other with the level of year-one log income. The former is more flexible and as such is preferred over the latter. For each panel we report the ETI, along with the corresponding first-stage F-test of the null that the instrument is uncorrelated with the endogenous regressor. In the first column of Table 1, the broad-income estimates of 0.119 for the model with the 10-piece spline and 0.138 for the model with the level of log income are remarkably similar to the corresponding estimates of 0.120 and 0.170 in Gruber and Saez (2002, Table 4) using their person-level synthetic instrument strategy. This suggests that for measures of broad income the CPS compares favourably with tax panel data.

In column (2) we show the corresponding estimate based on our new cohort-state-year synthetic instrument, where in this case the estimated ETI increases by 150% to 0.291 and 0.348, respectively, depending on income control.[15] The use of person-

[13] Following Gruber and Saez (2002) we censor the level of broad income at \$1 million in constructing weights.

[14] The Cameron *et al.* (2011) multi-way clustering is for non-nested clusters. In our sample, both individuals and cohorts are nested within states, and thus they recommend using the highest level of aggregation to create non-nested clusters, which in our case is at the state level.

[15] There is a large loss of efficiency when clustering at the state and year level. This is from the year dimension, which has 27 clusters, and not the state dimension with 51 clusters. Indeed, clustering only at the state-level yields standard errors 50% lower than those reported in the Tables. However, because theory predicts that the ETI is positive, hypothesis tests based on one-tail still suggest that estimates are statistically significant at the 10% level or better in column (2) and at the 5% level or better in column (4).

Table 1

Estimates of the Elasticity of Taxable Income Under Alternative Identification Strategies

	Broad income		Taxable income	
	Gruber–Saez synthetic instrument (1)	Cohort-state-year synthetic instrument (2)	Gruber–Saez synthetic instrument (3)	Cohort-state-year synthetic instrument (4)
Income control				
Spline of ln(income)	0.119	0.291	0.149	0.431
	(0.091)	(0.194)	(0.109)	(0.245)
First-stage F-test of instrument relevance	116	147	112	159
Level of ln(income)	0.138	0.348	0.218	0.547
	(0.075)	(0.191)	(0.091)	(0.237)
First-stage F-test of instrument relevance	122	129	131	144
Observations	198,285	198,285	198,285	198,285

Notes. Standard errors in parentheses are robust to heteroscedasticity and clustered at the state and year level. Gruber–Saez models use the person-specific synthetic instrument, and the Cohort-state-year models use the synthetic instrument aggregated to the Cohort-state-year level. All regressions are weighted by income and include controls for marital status, state effects and time effects for initial year. Cohort-state-year models also include cohort fixed effects. Income range is real year-one broad income greater than $10,000 and positive taxable income. The first-stage F-test is the first-stage test of the null that the instrument is uncorrelated with the endogenous regressor and is cluster robust.

specific income at $(t-1)$ to construct the synthetic instrument leads to significant attenuation of the ETI, which is in accord with recent results of Blomquist and Selin (2010) and Weber (2014). This attenuation is not due to a weak instrument problem as the first-stage F-tests are very large (the benchmark of weak instruments is an F-test < 10) and as depicted in Figures 4 and 5, and instead must be due to correlation between the instrument and the error term. In the last two columns of Table 1 we report the corresponding estimates for taxable income. The Gruber–Saez identification yields ETI estimates of 0.149 and 0.218. These are considerably lower than those reported in Gruber and Saez (0.4 and 0.6, respectively), which is not surprising given the more limited controls for deductions and exemptions in the CPS. However, in column (4) the grouping IV estimates jump by 150–200% to 0.431 and 0.547, again highlighting the concerns for tax policy with attenuation from using person-specific $(t-1)$ income in the Gruber–Saez instrument.

3.1. *The Effects of Demographics and Truncation on the ETI*

We first subject our new baseline estimates of 0.291 and 0.431 based on the grouping IV estimator with 10-piece spline in base-year income to possible bias from omission of potentially important confounding demographics and truncation. In Table 2 we append to (3) a quadratic in age, education attainment (dummies for high school, some college, college, and graduate degree, with less than high school the omitted group), indicators for children under the age of 6 and under 18, race (indicators for

Table 2

Estimates of the Elasticity of Taxable Income with Controls for Demographics and Truncation

	Additional linear demographics	Full interaction of demographics	All observations, baseline demographics	Truncation, baseline demographics	Truncation and additional linear demographics
	(1)	(2)	(3)	(4)	(5)
Broad income					
Elasticity	0.234	0.238	0.395	0.263	0.234
	(0.153)	(0.166)	(0.145)	(0.178)	(0.153)
First-stage F-test of instrument relevance	162	156	152	151	162
Taxable income					
Elasticity	0.358	0.362	0.477	0.407	0.357
	(0.201)	(0.214)	(0.241)	(0.227)	(0.198)
First-stage F-test of instrument relevance	175	170	136	163	175
Observations	198,285	198,285	214,088/203,085	198,285	198,285

Notes. Standard errors in parentheses are robust to heteroscedasticity and clustered at the state and year level. All models use the Cohort-state-year synthetic instrument, are weighted by income and include controls for a 10-piece spline in initial income, marital status, cohort effects, state effects, and time effects for initial year. Additional demographics in column (1) include gender (female), race (controls for African American and other with white as the omitted group), education (high school, some college, college and graduate degree with less that high school as the omitted group), indicators for children under age 6 and 18, and age and age squared. Column (2) includes a full interaction of the demographics in column (1). Column (3) includes all positive income observations, along with baseline demographics. Column (4) includes a control function for truncation bias with baseline demographics, while column (5) includes the demographics in column (1). Except for column (3), income range is real year-one broad income greater than $10,000 and positive taxable income. The first-stage F-test is the first-stage test of the null that the instrument is uncorrelated with the endogenous regressor, and is robust to state-year clustering in the variance.

African American, Other, with white as the omitted group), and gender. In column (1) we enter these demographics linearly, while in column (2) we present results from a saturated model with a full interaction of these dummy variables that appends an additional 70 demographic interactions (aka the ‘kitchen sink’ regression, Wooldridge, 2002, p. 619). All models continue to include controls for fixed cohort, state and time effects.

In column (1) of Table 2 we see that the ETI falls by about 20% to 0.234 for broad income and 0.358 for taxable income, suggesting that canonical estimates of the ETI are overstated due to the omission of demographics. There is no evidence of important non-linearities in the demographic effects as shown in column (2) and thus hereafter we focus on the more parsimonious specification.[16] In Table 3 we demonstrate which demographics seem to matter most for this change in the ETI by first adding in controls for gender (columns 1 and 5), age composition of children (columns 2 and

[16] In lieu of including the full interactions of demographics in the main structural equation, we instead estimate a reduced-form probit model of the probability that broad income exceeds $10,000 and taxable income is positive on both the linear and the full interaction of demographics. We then append the predicted probability (propensity score) to the model. The estimated ETI is 0.277 (0.177), which suggests that it does not control as flexibly for the demographics as the kitchen sink model reported in Table 2.

Table 3

The Effect of Demographics on the Elasticity of Taxable Income

	Broad income				Taxable income			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Elasticity	0.287 (0.192)	0.283 (0.189)	0.239 (0.167)	0.234 (0.153)	0.425 (0.243)	0.417 (0.243)	0.364 (0.216)	0.358 (0.201)
Controls include								
Marital status	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort and state effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Children	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Age	No	No	Yes	Yes	No	No	Yes	Yes
Race and education	No	No	No	Yes	No	No	No	Yes
First-stage F-test of instrument relevance	148	149	157	162	160	160	168	175
Observations	198,285	198,285	198,285	198,285	198,285	198,285	198,285	198,285

Notes. Standard errors in parentheses are robust to heteroscedasticity and clustered at the state and year level in the Cohort-state-year models. All models use the Cohort-state-year synthetic instrument, are weighted by income, and include controls for a 10-piece spline in initial income, cohort effects, state effects, and time effects for initial year. Additional demographics include gender (female), indicators for children under age 6 and between 6 and 18, age and age squared, race (controls for African American and other with white as the omitted group), education (high school, some college, college, and graduate degree with less that high school as the omitted group). Income range is real year-one broad income greater than $10,000 and positive taxable income. The first-stage F-test is the first-stage test of the null that the instrument is uncorrelated with the endogenous regressor and is robust to state-year clustering in the variance.

6), quadratic in the household head's age (columns 3 and 7) and education and race (columns 4 and 8). As gender, age and age of children are available in most tax panel data sets but education and race are generally not available, the sensitivity of the estimates to the inclusion of these variables is useful for applying our estimator to standard tax panel data sets. In Table 3 we see that the estimates change most with the inclusion of age of the household head and only trivially with the addition of education and race. This is good news for applying our estimator in tax panels as a significant extent of potential bias from omitted confounders can be captured with already measured demographics. As we discuss below, however, this good news is tempered somewhat by the fact that we define cohorts by 5-year birth and three education groups and thus the cohort fixed effects already absorb most of the effect of education.

In columns (3)–(5) of Table 2 we examine the role of truncation in the estimated ETI. We first apply the grouping IV estimator to a sample that includes all positive observations using the baseline demographics of marital status as in Table 1; that is, we remove the criteria that real broad income be at least \$10,000. In column (3) we see that the ETI increases from 0.291 to 0.395 for broad income (the corresponding estimate is 0.306 with additional demographic controls), suggesting that truncating the sample at \$10,000 does attenuate the ETI by about 25% and highlighting the importance of an extensive margin decision at the low end of the income distribution.[17]

In the last two columns of Table 2 we apply the grouping IV estimator to the conditional truncation model of (5). We construct the inverse Mills ratio based on a probit model of the probability that year-one real broad income is over \$10,000 and taxable income is positive and report the probit estimates in Appendix Table A4. In the reduced form probit model we include the observable demographics, along with fixed cohort, state and year effects. While in theory the non-linear functional form of the inverse Mills ratio assures identification, in practice this identification is often tenuous and leads to collinearity problems. Thus, we add some exclusion restrictions motivated from the literature on the labour supply of low-income families (Meyer and Rosenbaum, 2001). Specifically, we use state-level variables that change over time and that are likely to reflect the labour-market and macroeconomic conditions facing the poor, including employment per capita, the unemployment rate, the poverty rate, the inflation-adjusted state minimum wage, real personal income per capita and the combined real values of welfare (TANF) and food stamp (SNAP) benefits for a family of 3.[18] In column (4) we see that the benchmark estimates of Table 1 are fairly robust to truncation bias, falling by 6–10% once we control for the inverse Mills ratio. In column (5) we combine the demographic and truncation bias controls of columns (1) and (4) and see that the estimates are unchanged relative to column (1), meaning that controlling for demographics is sufficient for estimates of the ETI, conditional on truncation.[19] That

[17] Note that to properly model the tax incentives affecting the extensive margin decision we need to modify the equation to use the average tax rate and not the marginal tax rate (except in the restrictive tobit-type framework where the same rates affect both margins).

[18] The state-level variables were merged into the CPS using data from the University of Kentucky Center for Poverty Research Welfare Database http://www.ukcpr.org/EconomicData/UKCPR_National_Data_Set_12_14_11.xlsx

[19] Indeed, the inverse Mills ratio is statistically significant with baseline demographics in column (4) with a sizable negative coefficient, but is not economically or statistically significant in column (5).

said, the results in column (3) suggest that the general practice of dropping low-income individuals in ETI studies likely misses important behavioural responses to tax policy such as the EITC (Meyer and Rosenbaum, 2001).

3.2. *State-cohort Income Trends*

Several authors have raised concerns not only about the tax rate but about whether controlling for lagged person-specific income may cause bias in the ETI because of a correlation with the net-of-tax change and the model error term (Kopczuk, 2005; Weber, 2014). In Table 4, we apply the ideas of the grouping estimator and instead of using the 10-piece spline of the individual income, $f(y_{it-1})$, we use a 10-piece spline of cohort-state mean income, $f(y_{jct-1})$. Since cohort by state incomes change over time, the group-income spline is identified separately from the cohort, state and time fixed effects in the regression model. The first and third columns of Table 4 show that when using a group-year income spline the estimated ETI is 45% higher with broad income and 33% higher with taxable income. The second and fourth columns show that the ETI is more robust to inclusion of additional demographics with the group-income spline than with the individual income spline in Table 2, yielding ETIs of about 0.41 and 0.55, or only about 5% smaller instead of 20% in Table 2.

In Table 5 we return to the issue of implementing our estimator using information available in standard tax panels. Specifically, we redefine cohorts to be a function only of 5-year date-of-birth, and not birth-by-education, and then sequentially add observed demographics as in Table 3. We conduct this exercise both for models using the person-specific year-one income controls and for cohort-state-year income controls. The grouping IV estimates of 0.346 for broad income and 0.516 for taxable income in

Table 4

Sensitivity of the Elasticity of Taxable Income to Cohort-state-year Income Controls

	Broad income		Taxable income	
	Cohort-state-year income	With additional linear demographics	Cohort-state-year income	With additional linear demographics
	(1)	(2)	(3)	(4)
Elasticity	0.426	0.410	0.576	0.545
	(0.152)	(0.133)	(0.194)	(0.174)
First-stage F-test of instrument relevance	139	145	151	157
Observations	198,285	198,285	198,285	198,285

Notes. Standard errors in parentheses are robust to heteroscedasticity and clustered at the state and year level. All models use the Cohort-state-year synthetic instrument, are weighted by income, and include controls for a 10-piece spline in initial cohort-level income, marital status, cohort effects, state effects, and time effects for initial year. Additional demographics include gender (female), race (controls for African American and other with white as the omitted group), education (high school, some college, college, and graduate degree with less that high school as the omitted group), indicators for children under age 6 and 18, and age and age squared. Income range is real year-one broad income greater than $10,000 and positive taxable income. The first-stage F-test is the first-stage test of the null that the instrument is uncorrelated with the endogenous regressor, and is robust to state-year clustering in the variance.

Table 5

Estimates of the Elasticity of Taxable Income Defining Cohort by Birth-year Alone

	Broad income				Taxable income			
	Marital status	Add gender	Add children	Add age, age^2	Marital status	Add gender	Add children	Add age, age^2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Elasticity	0.346	0.342	0.330	0.254	0.516	0.510	0.507	0.415
	(0.291)	(0.289)	(0.284)	(0.303)	(0.343)	(0.342)	(0.345)	(0.294)
First-stage F-test for instrument relevance	192	192	191	191	210	210	207	206
Cohort-state-year income controls								
Elasticity	0.475	0.473	0.466	0.462	0.654	0.651	0.633	0.600
	(0.187)	(0.187)	(0.184)	(0.180)	(0.206)	(0.207)	(0.201)	(0.194)
First-stage F-test for instrument relevance	179	180	180	176	200	200	197	195
Observations	198,285	198,285	198,285	198,285	198,285	198,285	198,285	198,285

Notes. Standard errors in parentheses are robust to heteroscedasticity and clustered at the state and year level. All models use the Cohort-state-year synthetic instrument, are weighted by income, and include controls for a 10-piece spline in initial income, marital status, cohort effects, state effects, and time effects for initial year. Cohort is defined only by 5-year date of birth, and not interacted with education. The cohort-state-year income controls models use a 10-piece spline in base-year income aggregated to the cohort (birth-year only) level. The first-stage F-test is the first-stage test of the null that the instrument is uncorrelated with the endogenous regressor and is robust to state-year clustering in the variance.

the top panel are almost 20% higher than the baseline cohort IV estimates and are less precisely estimated. This suggests that some of the demographic effect is captured in the base-case cohort grouping that includes education and this additional heterogeneity adds important precision to the ETI.[20] The bottom panel of Table 5 using the cohort-level year-one controls, as in Table 4, suggests that this model is more robust to demographic factors and offers a promising alternative identification strategy for tax panels.[21]

3.3. *Additional Robustness Checks*

In Table 6 we consider several further variants of the baseline group IV model from Table 1 using person-specific year-one income controls (top panel) and from Table 4 using aggregated cohort-state-year income controls (bottom panel). First, we address the issue of whether there is time-variation in cohort-income growth by adding interactions of cohort-by-decade to the model of (3). It is not possible to include a full interaction of cohort and year as this nearly absorbs all the variation underlying the instrument (only state difference remain) but we can include a more parametric version by allowing cohort trends to vary by decade. In both the top and bottom panels of columns (1) and (5) the addition of cohort decadal trends results in attenuation of the ETI by about 15% from the baseline case but, given that these additional controls impose a greater challenge to identification, the results remain remarkably robust.

We next drop the top 5% of family income earners in columns (2) and (6) to examine the potential influence of top coding. Recall that the top code varies by year based on the cell-mean income of top-coded persons across detailed 12-point demographic cells defined by the Census Bureau. Thus, if they are top coded in any year then they are dropped from the sample. Here, we find that the ETI is attenuated by about 45% in the top panel and 75% in the bottom panel, suggesting that a substantial contribution to identification comes from the top of the income distribution, a point made by many others in the literature. In the third and seventh columns we no longer exclude cohorts with fewer than 50 observations per cohort-year. This has little effect on the ETI. In columns (4) and (8) we estimate the model in (3)

[20] Recent research by Chetty *et al.* (2014) links 1098-T tax forms that indicate attendance in a higher education institution with other administrative files. In results not tabulated we redefined our cohorts to be 5-year birth interacted with whether the individual had more than a high school education. This is a test of whether proxying human capital with post-secondary education, much like what one would get if they used US tax data linked to 1098-Ts, is sufficient compared to our cohorts of less than high school, high school and more than high school. In this case the ETI with respect to broad income was 0.31 (SE = 0.25) and taxable income was 0.40 (SE = 0.28) in models that control for baseline marital status, which are quite similar to the estimates in Table 1.

[21] In another set of results, we also ran the cohort-state-year model with only the federal tax rate included in the net-of-tax share with negligible change in ETI estimates in Table 2 (0.25 instead of 0.23 for broad income). We then re-estimated the grouping IV model where we ignore state variation both in constructing the tax rate and in the grouping instrument. Specifically we measure the tax rate by the federal rate alone, drop state fixed effects and aggregate the group IV only by cohort and year. In this case the baseline-estimated ETI increases to 0.507 (SE = 0.498) for broad income and 0.628 (SE = 0.691) for taxable income but then falls considerably with the additional demographics. This suggests that grouping by state is an important source of variation for the ETI and this additional level of heterogeneity yields estimates considerably more robust to inclusion of demographics.

Table 6

Robustness of the Elasticity of Taxable Income to Alternative Samples and Identification

	Broad income				Taxable income			
	Cohort-decade interactions	Drop top 5%	Keep cohort-year cells with <50 observations	BDM estimator	Cohort-decade interactions	Drop top 5%	Keep cohort-year cells with <50 observations	BDM estimator
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Elasticity	0.236	0.156	0.307	−2.805	0.377	0.289	0.445	−3.231
	(0.178)	(0.114)	(0.198)	(0.303)	(0.226)	(0.157)	(0.248)	(0.400)
First-stage F-test for instrument relevance	143	83	133	14	154	88	143	15
Cohort-state-year income controls								
Elasticity	0.360	0.101	0.454	−3.139	0.509	0.264	0.605	−3.585
	(0.130)	(0.095)	(0.163)	(0.334)	(0.171)	(0.136)	(0.205)	(0.425)
First-stage F-test for instrument relevance	138	78	124	16	213	81	135	17
Observations	198,285	188,371	198,781	198,285	198,285	188,371	198,781	198,285

Notes. Standard errors in parentheses are robust to heteroscedasticity and clustered at the state and year level. All models use the Cohort-state-year synthetic instrument, are weighted by income, and include controls for a 10-piece spline in initial income, marital status, cohort effects, state effects, and time effects for initial year. The cohort-state-year income controls models use a 10-piece spline in base-year income aggregated to the cohort (birth-year by education) level. Income range is real year-one broad income greater than \$10,000 and positive taxable income. The first-stage F-test is the first-stage test of the null that the instrument is uncorrelated with the endogenous regressor and is robust to state-year clustering in the variance. The BDM estimator in columns (4) and (8) implements the Blundell *et al.* (1998) two-step grouping estimator with overidentifying restrictions based on interactions of state and cohort, state and year, and cohort and year.

Table 7

Robustness of the Elasticity of Taxable Income to Alternative Time Periods

	Broad income			Taxable income		
	1980s	1990s	2000s	1980s	1990s	2000s
	(1)	(2)	(3)	(4)	(5)	(6)
Elasticity	0.458	0.320	−0.051	0.671	0.411	0.051
	(0.140)	(0.402)	(0.151)	(0.179)	(0.408)	(0.219)
First-stage F-test of instrument relevance	87	32	92	96	34	100
Observations	77,379	64,190	56,716	77,379	64,190	56,716

Notes. Standard errors in parentheses are robust to heteroscedasticity and clustered at the state and year level. All models use the Cohort-state-year synthetic instrument, are weighted by income, and include controls for a 10-piece spline in initial income, cohort effects, state effects, and time effects for initial year. Additional demographics include gender (female), race (controls for African American and other with white as the omitted group), education (high school, some college, college, and graduate degree with less that high school as the omitted group), indicators for children under age 6 and 18, and age and age squared. Income range is real year-one broad income greater than $10,000 and positive taxable income. The first-stage F-test is the first-stage test of the null that the instrument is uncorrelated with the endogenous regressor and is robust to state-year clustering in the variance.

using the overidentified reduced-form strategy of Blundell *et al.* (1998). That is, in the first stage we predict the change in net-of-tax share using interactions of state and cohort, state and year and cohort and year as over-identifying restrictions. We then include the residual from the first stage to control for the endogeneity of the change in the net-of-tax share in (3). The estimated ETIs are negative across all specifications, suggesting that this identification strategy in effect ‘overfits’ the first stage and absorbs the across cohort-state-year variation needed to identify the ETI relative to our just-identified model.[22]

In a final robustness check, we estimate the baseline model in (3) by decade, including the additional demographics as controls, to examine whether or not the tax reforms in particular decades are offering differing weight to identification, as was hinted in Figures 3–5. The estimates in Table 7 make concrete that the 1980s tax reforms are most important, followed by the 1990s, which is consistent with the tax-panel estimates in Giertz (2007). The reforms in the 2000s, which were less comprehensive compared to the 1980s, contribute little to the overall variation.

4. Conclusion

We present new estimates of the ETI using matched panels from the CPS along with a grouping instrumental variables estimator. With few exceptions the literature has

[22] In an earlier version of this article we presented estimates of wage elasticities of labour supply for men and women using the Blundell *et al.* (1998) estimator. There we found positive compensated wage effects, consistent with theory, suggesting that pre-tax wages in the labour supply model provide the variation needed to implement the overidentified BDM estimator not found in the net-of-tax share alone.

relied upon taxpayer panel data and this is the first use of matched panels of the CPS to the ETI literature, which enabled us to explore the importance of demographics such as education and race not typically available in tax panels. The grouping instrumental variables estimator utilised variation in tax policy across birth and education cohorts, states and time to identify the ETI. This group-level synthetic instrument was strongly correlated with individual-level actual changes in the net-of-tax share and, while it is not possible to test the correlation of the instrument with the error term in the just-identified framework, we argued that an instrument at the cohort-state-year level is less likely to be correlated with the structural error than the individual-level synthetic instrument utilised in much of the literature.

Our preferred estimates suggest that the elasticity with respect to broad income is about 0.4 and with respect to taxable income is about 0.55. These are based on our grouping estimator that also controls for person-level demographics such as age, education, race, gender and family structure, along with lagged cohort-state-year income to control for heterogeneous trends. Even though our taxable-income elasticity is probably understated because of the lack of detailed data on deductions and exemptions in the CPS, it is more than double the modal estimate of 0.25 reported in the survey by Saez *et al.* (2012). Our results are consistent with larger estimates in the recent work from Blomquist and Selin (2010) and Weber (2014), suggesting an emerging consensus that the ETI may be larger than previously thought after adopting more robust identification strategies. The implications of our estimates are significant for tax policy. Using the formula for the revenue-maximising tax rate in Saez (2001) of $\tau^* = 1/(1 + a \times e)$, where a is the Pareto parameter characterising the shape of the income distribution in the highest tax bracket and e is the ETI, the revenue-maximising rate falls from 82% based on a Pareto parameter of 1.5 and ETI of 0.15 from standard models to 55% with an ETI of 0.55 from our grouping IV estimator.

Because it is possible to construct a similar cohort-state-year grouping instrumental variables estimator in large taxpayer panels, our results suggest that this could be a fruitful alternative for identification in future research on the ETI in tax panels. Most tax panels have information on age of the individual, gender, marital status, state or municipality and the number and age composition of dependents. The variables often missing, at least in the US context, are race and education. European tax panels are generally much richer in this domain (Blomquist and Selin, 2010; Kleven and Schultz, 2014). However, there are ongoing efforts in the US that link Internal Revenue Service tax data to the CPS (Jones, 2014) and other survey and/or administrative data (Chetty *et al.*, 2014) with measures of human capital that could ultimately prove very promising for tax-based research.

Appendix A. Summary Statistics and First-Stage Estimates

Table A1

Cohort-year Cells Dropped due to Observations Less than 50

Cohort	Year(s) with <50 observations
1	2004–6
2	2005
4	1999–2001
5	1999
7	1995–6
10	1989–90
16	1979
22	2006–7
25	2002
26	2002
28	1997
31	1992
39	1982
Total cohort-years dropped	20 out of 444
Total observations dropped	655 out of 214,818

Table A2

Summary Statistics, 1979–2008

	Mean	SD
Demographics		
Married	0.685	0.464
Single	0.189	0.391
Female	0.306	0.461
White	0.874	0.332
African American	0.085	0.279
Other	0.041	0.198
Age	41.527	9.416
Child under 6	0.213	0.409
Child under 18	0.405	0.491
Less than high school	0.119	0.324
High school graduate	0.345	0.475
More than high school	0.536	0.499
Income and tax rates		
Broad income	70,447	56,475
Taxable income	61,324	54,189
Net-of-tax share (Federal + State)	73.370	11.434

Note. Sample size is 198,285.

Table A3

Variation in After-tax Share Cohort Based Instrument $\overline{\{\ln[(1-\hat{\tau}_{it})/(1-\tau_{it-1})]_{jct}\}}$ *By Education Level*

Year	Less than high school			High school only			More than high school		
	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*N*
1979–80	−0.022	0.020	1,632	−0.026	0.017	2,814	−0.028	0.018	3,026
1980–1	−0.015	0.019	1,448	−0.017	0.017	2,653	−0.019	0.017	2,760
1981–2	0.016	0.016	1,448	0.022	0.016	2,806	0.033	0.020	3,166
1982–3	0.017	0.015	1,414	0.022	0.014	2,764	0.030	0.015	3,012
1984–85	0.006	0.014	1,149	0.010	0.010	2,609	0.014	0.010	3,023
1985–86	0.003	0.010	1,126	0.004	0.007	2,752	0.004	0.007	3,237
1986–87	0.040	0.029	1,160	0.052	0.023	3,017	0.063	0.021	3,592
1987–88	0.013	0.025	1,176	0.016	0.020	3,364	0.031	0.021	4,007
1988–87	0.000	0.006	1,109	−0.001	0.005	3,430	−0.001	0.005	4,273
1989–90	−0.001	0.007	1,134	−0.001	0.004	3,578	−0.001	0.004	4,700
1990–91	0.000	0.015	1,075	0.001	0.010	3,594	0.002	0.010	4,650
1991–92	0.001	0.012	952	0.002	0.013	3,233	0.001	0.009	5,074
1992–93	−0.001	0.010	624	−0.002	0.006	2,255	−0.004	0.007	3,774
1993–94	0.003	0.034	518	−0.001	0.018	1,878	0.000	0.011	3,492
1995–96	−0.001	0.020	686	0.000	0.010	2,350	0.001	0.007	4,071
1996–97	0.002	0.014	645	0.002	0.010	2,329	0.002	0.009	4,012
1997–98	0.013	0.041	637	0.010	0.027	2,136	0.005	0.015	3,886
1998–99	0.004	0.019	609	0.002	0.011	2,023	0.001	0.007	3,707
1999–00	0.000	0.011	595	−0.001	0.011	1,953	−0.001	0.006	3,432
2000–01	0.017	0.029	626	0.009	0.019	2,321	0.009	0.012	4,287
2001–02	0.000	0.026	589	0.004	0.019	2,250	0.006	0.010	4,358
2002–03	0.022	0.036	625	0.026	0.023	2,207	0.028	0.016	4,578
2003–04	−0.033	0.101	474	0.003	0.068	1,880	0.022	0.044	3,991
2004–05	−0.002	0.020	582	−0.001	0.011	1,981	−0.002	0.006	4,207
2005–06	0.001	0.019	584	0.001	0.011	2,052	0.002	0.006	4,487
2006–07	0.002	0.017	525	0.000	0.010	2,088	0.001	0.007	4,622
2007–08	−0.010	0.047	541	−0.003	0.017	2,042	−0.003	0.010	4,819

Note. The sample is real year-one broad income greater than $10,000 and positive taxable income.

Table A4

First Stage Probit Estimates for Truncation Model

	Baseline model	Expanded demographic model
Married	0.836***	0.919***
	(0.012)	(0.015)
Single	0.001	0.164***
	(0.013)	(0.014)
Female		−0.275***
		(0.011)
Other race		−0.253***
		(0.023)
Black		−0.329***
		(0.014)
Children under 6		−0.421***
		(0.013)
Children aged 6–18		−0.254***
		(0.011)

Table A4
(*Continued*)

	Baseline model	Expanded demographic model
Age of head		0.073***
		(0.006)
Age of head squared		−0.001***
		(0.000)
State employment *per capita*	1.501***	1.434***
	(0.534)	(0.546)
State poverty rate	−0.026***	−0.025***
	(0.003)	(0.003)
Real state minimum wage	−0.020	−0.017
	(0.014)	(0.014)
State unemployment rate	−0.011*	−0.012**
	(0.006)	(0.006)
Real state *per capita* income	−0.000	−0.001
	(0.003)	(0.003)
Real state combined TANF-SNAP	0.001	0.001
	(0.001)	(0.001)
Head education high school		0.457***
		(0.094)
Head education some college		0.612***
		(0.133)
Head education college		0.918***
		(0.133)
Head education graduate degree		1.065***
		(0.134)
Constant	0.470	−1.346**
	(0.483)	(0.565)
Observations	214087	214087

Notes. Standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

Institute for Defense Analyses
University of Kentucky

Submitted: 2 December 2012
Accepted: 12 April 2015

Additional Supporting Information may be found in the online version of this article:

Data S1.

References

Angrist, J. (1991). 'Grouped data estimation and testing in simple labor supply models', *Journal of Econometrics*, vol. 47(2–3), pp. 243–66.

Attanasio, O. and Weber, G. (2010). 'Consumption and saving: models of intertemporal allocation and their implications for public policy', *Journal of Economic Literature*, vol. 48(3), pp. 693–751.

Auten, G. and Carroll, R. (1999). 'The effect of income taxes on household income', *Review of Economics and Statistics*, vol. 81(4), pp. 681–93.

Barnow, B, Cain, G. and Goldberger, A. (1980). 'Issues in the analysis of selectivity bias', in (E. Stromsdorfer and G. Farkas, eds.), *Evaluation Studies Review Annual, vol. 5*, pp. 43–59, Beverly Hills, CA: Sage Publications.

Blomquist, S. and Selin, H. (2010). 'Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates', *Journal of Public Economics*, vol. 94(11–12), pp. 878–89.

Blundell, R. and MaCurdy, T. (1999). 'Labor supply: a review of alternative approaches', in (O. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, vol. 3A, pp. 1559–695, Amsterdam: North-Holland.

Blundell, R., Duncan, A. and Meghir, C. (1998). 'Estimating labor supply responses using tax reforms', *Econometrica*, vol. 66(4), pp. 827–61.

Bollinger, C. and Hirsch, B. (2006). 'Match bias from earnings imputation in the current population survey: the case of imperfect matching', *Journal of Labor Economics*, vol. 24(3), pp. 483–519.

Brewer, M., Saez, E. and Shepherd, A. (2010). 'Means testing and tax rates on earnings', in (S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba, eds.), *Dimensions of Tax Design: The Mirrlees Review*, pp. 90–201, Oxford: Oxford University Press.

Burkhauser, R., Feng, S., Jenkins, S. and Larrimore, J. (2012). 'Recent trends in top income shares in the USA: reconciling estimates from March CPS and IRS tax return data', *Review of Economics and Statistics*, vol. 94(2), pp. 371–88.

Cameron, A.C., Gelbach, J. and Miller, D. (2011). 'Robust inference with multiway clustering', *Journal of Business and Economic Statistics*, vol. 29(2), pp. 238–49.

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

Chetty, R., Friedman, J. and Rockoff, J. (2014). 'Measuring the impacts of teachers II: teacher value-added and student outcomes in adulthood', *American Economic Review*, vol. 104(9), pp. 2633–79.

Deaton, A. (1997). *Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Baltimore, MD: Johns Hopkins University Press.

Feldstein, M. (1995). 'The effect of marginal tax rates on taxable income: a panel study of the 1986 tax reform act', *Journal of Political Economy*, vol. 103(3), pp. 551–72.

Giertz, S. (2007). 'The elasticity of taxable income over the 1980s and 1990s', *National Tax Journal*, vol. 60(4), pp. 743–68.

Giertz, S. (2010). 'The elasticity of taxable income in the 1990s: new estimates and sensitivity analyses', *Southern Economic Journal*, vol. 77(2), pp. 406–33.

Griliches, Z. and Hausman, J. (1986). 'Errors in variables in panel data', *Journal of Econometrics*, vol. 31(1), pp. 93–118.

Gruber, J. and Saez, E. (2002). 'The elasticity of taxable income: evidence and implications', *Journal of Public Economics*, vol. 84(1), pp. 1–32.

Hardy, B. and Ziliak, J.P. (2014). 'Decomposing trends in income volatility: the wild ride at the top and the bottom', *Economic Inquiry*, vol. 52(1), pp. 459–76.

Hausman, J. (1981). 'Labor supply', in (H. Aaron and J. Pechman, eds.), *How Taxes Affect Economic Behavior*, pp. 27–84, Washington, DC: The Brookings Institution.

Heckman, J. (1979). 'Sample selection bias as specification error', *Econometrica*, vol. 47(1), pp. 153–61.

Heckman, J. and Robb, R. (1985). 'Alternative methods for evaluating the impact of interventions', in (J. Heckman and B. Singer, eds.), *Longitudinal Analysis of Labor Market Data*, pp. 156–246, Cambridge: Cambridge University Press.

Heckman, J., Lochner, L. and Todd, P. (2006). 'Earnings functions, rates of return, and treatment effects: the Mincer equation and beyond', in (E. Hanushek and F. Welch, eds.), *Handbook of the Economics of Education*, vol. 1, pp. 307–458, Amsterdam: North Holland.

Heim, B. (2009). 'The effect of recent tax changes on taxable income: evidence from a new panel of tax returns', *Journal of Policy Analysis and Management*, vol. 28(1), pp. 147–63.

Hurst, E., Li, G. and Pugsley, B. (2014). 'Are household surveys like tax forms: evidence from income underreporting of the self-employed', *Review of Economics and Statistics*, vol. 96(1), pp. 19–33.

Jones, M. (2014). 'Changes in EITC eligibility and participation, 2005–2009', CARRA Working Paper Series No.2014-04, US Census Bureau.

Keane, M. (2011). 'Labor supply and taxes: a survey', *Journal of Economic Literature*, vol. 49(4), pp. 961–1075.

Kleven, H. and Schultz, E. (2014). 'Estimating taxable income responses using Danish tax reforms', *American Economic Journal: Economic Policy*, vol. 6(4), pp. 271–301.

Kopczuk, W. (2005). 'Tax bases, tax rates, and the elasticity of reported income', *Journal of Public Economics*, vol. 89(11–12), pp. 2093–119.

Kopczuk, W. (2012). 'The Polish business "flat" tax and its effect on reported incomes: a Pareto improving tax reform?', Working Paper, Columbia University.

Larrimore, J., Burkhauser, R., Feng, S. and Zayatz, L. (2008). 'Consistent cell means for topcoded incomes in the public use March CPS (1976–2007)', *Journal of Economic and Social Measurement*, vol. 33(2–3), pp. 89–128.

MaCurdy, T., Green, D. and Paarsch, H. (1990). 'Assessing empirical approaches for analyzing taxes and labor supply', *Journal of Human Resources*, vol. 25(3), pp. 415–89.

Meyer, B. and Rosenbaum, D. (2001). 'Welfare, the earned income tax credit, and the labor supply of single mothers', *Quarterly Journal of Economics*, vol. 116(3), pp. 1063–114.

Moffitt, R. (1993). 'Identification and estimation of dynamic models with a time series of repeated cross sections', *Journal of Econometrics*, vol. 59(1–2), pp. 99–123.

Moffitt, R. and Wilhelm, M. (2000). 'Taxation and the labor supply decisions of the affluent', in (J. Slemrod, ed.), *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, pp. 193–234, New York: Russell Sage Foundation.

Mroz, T. (1987). 'The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions', *Econometrica*, vol. 55(4), pp. 765–99.

Saez, E. (2001). 'Using elasticities to derive optimal income tax rates', *Review of Economic Studies*, vol. 68(1), pp. 205–29.

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

Singleton, P. (2011). 'The effect of taxes on taxable earnings: evidence from the 2001 and related US federal tax acts', *National Tax Journal*, vol. 64(2), pp. 323–52.

Verbeek, M. and Vella, F. (2005). 'Estimating dynamic models from repeated cross sections', *Journal of Econometrics*, vol. 127(1), pp. 83–102.

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

Welniak, E. (1990). 'Effects of the March Current Population Survey's new processing system on estimates of income and poverty', Working Paper, US Census Bureau.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

Zabel, J. (1993). 'The relationship between hours of work and labor force participation in four models of labor supply behavior', *Journal of Labor Economics*, vol. 11(2), pp. 387–416.

Ziliak, J.P. and Kniesner, T. (1998). 'The importance of sample attrition in life-cycle labor supply estimation', *Journal of Human Resources*, vol. 33(2), pp. 507–30.

Ziliak, J.P. and Kniesner, T. (1999). 'Estimating life-cycle labor-supply tax effects', *Journal of Political Economy*, vol. 107(2), pp. 326–59.

Ziliak, J.P. and Kniesner, T. (2005). 'The effect of income taxation on consumption and labor supply', *Journal of Labor Economics*, vol. 23(4), pp. 769–96.