Estimating Intensive and Extensive Tax Responsiveness*

Abby Alpert†
David Powell‡

February 2020

Abstract

Influential literatures have exploited tax policy changes to estimate the effects of income taxes on either intensive or extensive margin decisions. We extend this quasi-experimental approach to jointly estimate intensive and extensive margin tax elasticities to address selection issues that have hindered consistent estimation of labor supply effects. The extensive margin equation provides a way to control for selection in the intensive margin equation while consistent estimation of the intensive margin provides estimates of after-tax returns to working for non-workers, a necessary input to study extensive behavioral responses. We apply this approach to study the tax responsiveness of older workers.

Keywords: Labor Supply Estimation, Income Taxes, Sample Selection, Nonlinear Budget Sets, Retirement
JEL Classification: C33, H24, J20, J26

*We would like to thank Michael Dworsky, Jon Gruber, John Laitner, Nicole Maestas, Kathleen Mullen, Jim Poterba, Susann Rohwedder, Hui Shan, David Stapleton and conference and seminar participants at the Association for Public Policy Analysis and Management Fall Research Conference, Michigan Retirement Research Center Conference, National Tax Association Conference, RAND, and the Retirement Research Center Consortium for their helpful comments. We received especially insightful comments from our conference discussants Robert Carroll, David John, and David Richardson. We also thank Dan Feenberg for assistance with the NBER TAXSIM program. The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Retirement Research Consortium. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA or any agency of the Federal Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States Government or any agency thereof.

†The Wharton School and NBER, University of Pennsylvania, alpertab@wharton.upenn.edu
‡RAND, dpowell@rand.org, (703) 413-1100

1 Introduction

An influential tax literature has estimated how households respond to income tax rates on the intensive margin using tax schedule changes and nonlinearities in the tax code to identify this relationship.[1] A separate literature, often focused on female labor supply, has exploited similar policy changes to study extensive margin behavioral responses to income taxes (e.g., Eissa and Hoynes (2004); Eissa et al. (2008)). Selection issues are problematic in both contexts and, more broadly, whenever intensive and extensive labor supply decisions are studied in isolation. In this paper, we show that the legislative tax changes exploited in these tax literatures permit joint estimation of the intensive and extensive margin equations in a manner that directly addresses the threats posed by selection in each context.

We build on insights in the labor supply and tax literatures recognizing the variation in incentives induced by the nonlinear budget sets created by the tax schedule (e.g., Hausman (1985b); Blomquist and Newey (2002); Keane and Moffitt (1998)) and introduce an approach to jointly identify the intensive and extensive labor supply equations from tax schedule changes. The proposed method should be useful broadly in the tax and labor supply literatures.

We contribute two key innovations to the approaches often used in these literatures. First, we extend the Gruber and Saez (2002) empirical framework to isolate behavioral responses to taxes on the extensive margin. Gruber and Saez (2002) show that nonlinearities in the tax schedule provide separate variation in marginal tax rates and after-tax income, which identifies the intensive margin substitution and income effects, respectively. We add to this approach by recognizing that nonlinearities in the tax schedule can also be used to separately identify yet another important dimension: the tax-based incentives to participate in the labor force. We are then able to separately estimate substitution and income effects along both the intensive and extensive margins using quasi-experimental variation.

Second, we account for issues of selection and unobserved earnings for non-workers by estimating the intensive and extensive margin equations together. The two main issues we address are: 1) in the intensive margin equation, the estimated earnings effect is conditional on working, yet the decision to work may be endogenous to tax incentives, thus leading to biased estimates; and 2) in the extensive margin equation, the main explanatory

[1] See Auten and Carroll (1999); Gruber and Saez (2002); Feldstein (1995); Burns and Ziliak (2017); Giertz (2007); Auten et al. (2008); Heim (2009); Singleton (2011); Saez et al. (2012) for a few examples.

variable of interest (after-tax income if the individual works) is not observed for non-workers. Our approach shows how the selection and imputation issues that have plagued estimation of either equation independently can be resolved if these equations are not estimated in isolation.

This method uses policy-driven changes in the after-tax incentives to work as a selection instrument for the decision to work. After-tax non-labor income affects labor force participation, but does not – conditional on the marginal tax rate and after-tax income – independently affect intensive labor supply outcomes, making it an ideal selection instrument since it plausibly satisfies the exclusion restriction required to implement a sample selection model. This provides a method for obtaining consistent estimates of the intensive margin equation. Once we have consistent estimates of the intensive labor supply parameters, we then can use these parameters to predict individual labor earnings for everyone in the sample, even those who do not work.

The elasticity of taxable income and male labor supply literatures often ignore selection concerns. In contrast, the female labor supply literature typically focuses on the extensive margin. Models of the extensive margin of labor supply model the decision to work as a function of the pecuniary return to working. In this literature, the main explanatory variable of interest (i.e., wages or total earnings) is missing for those who do not work. The literature has addressed this issue in one of two ways.

First, for individuals who are not working, this literature imputes wages or total earnings as if they had worked, assuming that workers and non-workers are the same conditional on covariates (see Meyer and Rosenbaum (2001) and Blau and Kahn (2007)). Second, other studies use selection models to impute earnings for non-workers. The excluded variable identifying the selection equation is the number of children or presence of young children (Eissa and Hoynes (2004); Eissa et al. (2008)) and/or non-labor income (e.g., Heim (2007); Kumar and Liang (2016)).[2] We build on this framework, but improve on the identification of the selection equation using quasi-experimental variation derived from nonlinearities in the tax schedule and changes to the tax code over time. This strategy offers credibly exogenous variation in the selection mechanism and has broader applicability than previous instruments. Given the known challenges in finding appropriate selection instruments for labor force participation, our approach should be useful more broadly in the labor literature.

[2] Huber and Mellace (2014) test the appropriateness of these instruments. Other papers rely on distributional assumptions or other individual-level characteristics (e.g., Kimmel and Kniesner (1998)) to identify the participation equation.

We illustrate the potential of this approach by studying tax schedule changes (detailed below) during the period 1999-2007. A rich literature on the elasticity of taxable income (ETI) has used tax schedule changes to identify behavioral responses to taxes (e.g., Auten and Carroll (1999); Gruber and Saez (2002)) and a number of studies have found economically meaningful aggregate behavioral responses to tax policies during our study period (see Giertz (2007); Auten et al. (2008); Heim (2009); Singleton (2011); Saez et al. (2012)) using this approach. We study the effects of income tax changes on the labor supply behavior of older individuals in the United States. Economists and policymakers have long been interested in understanding the effects of economic incentives on the retirement decisions of older workers. Due to the potential benefits of systematic delays in retirement, there are large literatures investigating the labor supply responses to Social Security benefits (see Feldstein and Liebman (2002) for a review), pensions (e.g., Samwick (1998); French and Jones (2012)), and Medicare (e.g., Blau and Gilleskie (2006); French and Jones (2011)). The tax code is a potentially useful, but often overlooked, policy lever to encourage individuals to earn more and remain in the labor force longer.

While there exists a large literature which estimates the effects of taxes on labor supply (summarized in Keane (2011)) and on taxable income (summarized in Saez et al. (2012)), these studies often explicitly exclude older individuals from the analysis or estimate aggregate effects combining all age groups. Auten and Carroll (1999) limit their sample to ages 25-55, Feldstein (1995) excludes individuals over age 65, and the majority of the other studies estimate effects for an aggregate or younger population. The labor supply literature, summarized in Hausman (1985a), Blundell and MaCurdy (1999), and Keane (2011), also does not, generally, study older workers and frequently even eliminates them from the analysis. Many influential studies have selected on individuals by age, usually using a cutoff of 50, 55, or 60.[3] The exclusion of the older population is likely due to selection concerns given the low rates of labor force participation.

We focus on older ages because the extensive margin is especially important for this demographic and, consequently, accounting for selection concerns is potentially crucial. We study the population ages 62 to 74 in the 2000 Census and 2001-2008 American Community Surveys (ACS) which provide detailed information on earnings and employment. Because tax rates and earnings are mechanically linked, we exploit exogenous variation in federal

[3]See, for example, Hausman (1985a); Blomquist and Hansson-Brusewitz (1990); Triest (1990); Eissa (1995); Blundell et al. (1998); Ziliak and Kniesner (2005); Blomquist and Selin (2010) which are just a small subsample of these studies.

income tax rates originating from two major legislative tax schedule changes that occurred during this time period: the Economic Growth and Tax Relief Reconciliation Act of 2001 and the Jobs and Growth Tax Relief Reconciliation Act of 2003.

Most relevant to our study, Laitner and Silverman (2012) simulate the effects of eliminating the payroll tax for older ages and conclude that this policy would delay retirement by, on average, one year. Gustman and Steinmeier (2015) also estimate a structural life-cycle model and find increases in full-time work at ages 65+ if the employee portion of the payroll tax were eliminated.[4] Our paper takes a different approach than this small literature on tax policy for older workers by using tax policy changes as a source of identification, intersecting more directly with the elasticity of taxable income literature, and thereby providing the first "quasi-experimental" evidence of the impact of income taxes on the labor supply decisions of older individuals. There are tradeoffs to the two approaches. Our results may speak more to the short-term effects of tax policy *changes*, but may not capture the full long-term effects that arise when individuals are able to anticipate age-specific tax reductions.[5] On the other hand, while a structural life-cycle model explicitly considers the dynamic aspects of labor supply decisions in response to income taxes, that approach requires making assumptions regarding a household utility function, modeling disability trajectories, and imposing further restrictions. This structural literature also tends to ignore intensive margin concerns for older workers, modeling individuals as selecting only between full-time work and not working.

We follow the approach of a long-standing and influential tax literature which studies the observed behavioral changes resulting from legislative tax changes, extending it to address behavior at older ages, and introducing a method to use natural experiments to jointly estimate intensive and extensive margin responsiveness. We believe that this contribution is a critical step – both generally and in the context of older workers – and can later be merged with life-cycle models to relax structural assumptions in those models which are necessary to isolate intensive margin and extensive margin incentives.

One literature which specifically addresses labor supply effects of taxes for older workers is the literature studying the effects of the Social Security Annual Earnings Test

[4] A small literature has estimated calibrated life-cycle models and found welfare gains when taxes are age-dependent (e.g., Weinzierl (2011), Karabarbounis (2016)).

[5] Some have suggested scope for more age-targeted tax policy (e.g., Kremer (2002)). Banks and Diamond (2010) in the Mirrlees Review recommend increasing the age dependence of taxes, calling the idea "a case of theory being ahead of policy, with research on tax design needed." Meanwhile, some economists have recommended eliminating the payroll tax after certain ages or after Social Security receipt (Biggs (2012); Laitner and Silverman (2012)).

(Friedberg (2000); Gruber and Orszag (2003); Song and Manchester (2007); Haider and Loughran (2008); Gelber et al. (2020b,a, 2018)). Notably, this literature has primarily relied on a static framework as well. Findings in this literature are mixed, though recent evidence suggests that Social Security recipients are very responsive to the earnings test. Given that the earnings test is not a pure tax since it returns the benefits in an actuarially fair manner, it is possible that individuals' responsiveness to income tax changes may be even larger. A related literature has also studied behavioral responses to the taxing of Social Security benefits (Page and Conway, 2015).

Our results suggest that taxes have a statistically significant and economically large impact on labor force participation and retirement decisions for older workers. On the extensive margin, we estimate large compensated participation elasticities, defined as responsiveness to the additional after-tax income earned from working: 3.9 for women and 0.7 for men.[6] These estimates are smaller than or similar to those produced by the structural models referenced earlier. Using these estimates, we predict that the elimination of the employee portion of Social Security payroll taxes for older workers would increase the percentage of women working by 4.3 percentage points and men by 1.4 percentage points, corresponding to increases of 16% and 4%, respectively, from baseline rates. On the intensive margin, we estimate that individuals respond to the marginal net-of-tax rate, the amount that a worker keeps for an additional $1 in earnings. We estimate large compensated elasticities for both women and men. Overall, our estimates suggest substantial scope for impacting labor force decisions of older workers through the tax code.

We present evidence that the results are not driven by underlying trends correlated with our tax instruments. In fact, we find that our results are generally unaffected by including next year's tax measures. The main estimates are also robust to testing for confounding effects related to decisions to work part-time. We also consider whether policies related to the Social Security Earnings Test during this time period could be driving the results. Our estimates are robust to accounting for these policy changes.

In the next section, we describe the data. Section 3 includes the model and empirical strategy. We present our results in Section 4, including policy simulation estimates. Section 5 concludes.

[6]The elasticities for non-participation are often reported in the literature. These are 1.5 for women and 0.5 for men in this context.

2 Data

We use the 2000 Census and the 2001-2008 American Community Surveys (ACS) for adults ages 62-74. These data sets contain a rich set of information including detailed demographics, income, and labor supply information. The sample size is also large which is important given that we are focusing on a narrow age-group and aim to independently identify three separate, but correlated, tax variables.[7] The Census and the ACS provide equivalent variables (Ruggles et al. (2015)) and are often linked together (e.g., Coile and Levine (2010)). Since the income variables refer to the previous year, our sample spans 1999-2007. These years bookend the major tax policies that we study and have the advantage of preceding the Great Recession.

The detailed income variables of the Census and ACS are beneficial for generating tax variables. We use NBER's TAXSIM program (Feenberg and Coutts (1993)) to derive tax rates, tax liability, and labor taxes for each individual based on their household income and family characteristics. We use federal and states taxes plus one-half of FICA taxes in our calculations of tax rates and tax liability.[8] Some sources of non-labor income (e.g., Social Security benefits) are taxed differently than labor income and thus should not be included in the determination of the marginal tax rate with respect to labor income. Since we focus on labor supply incentives, we define the marginal tax rate as the additional taxes for the next dollar of a person's *labor* income. Our tax liability measures include tax liability from all income sources and account for the unique treatment of each type of income by the tax code, as captured by TAXSIM.[9]

The advantages and disadvantages of studying tax responsiveness in secondary data sources such as the Census relative to administrative data are discussed thoroughly in Burns and Ziliak (2017). For example, information on many deductions is not collected in the Census and American Community Surveys. On the other hand, we have access to a richer set of demographic characteristics, which will be beneficial for our identification strategy.

[7]We also find similar results using the Health and Retirement Study (HRS) in a previous version of this paper. However, the Census and ACS combined have a much larger sample size, providing more precision.

[8]TAXSIM includes the age 65 deduction when applicable but, otherwise, does not use age information when calculating household tax information. Consequently, TAXSIM will assign the EITC to individuals ages 65+. We obtain the TAXSIM calculations of the EITC and subtract these values for individuals age 65 or older.

[9]The Census and American Community Surveys do not include information on capital gains, but "most previous studies have also excluded capital gains from their analysis" (Gruber and Saez (2002)). Capital gains are also excluded in more recent tax research (e.g., Burns and Ziliak (2017)).

We can also proxy for retirement using the employment status variable. We define "retired" in our data by two criteria: (1) no individual labor earnings and (2) self-declared as not in the labor force.[10] Consequently, we can study two dimensions of the extensive margin: "Not Working" if no labor earnings in the previous year; and "Retirement" if not in the labor force and no labor earnings in the previous year.

We present summary statistics for our data in Table C.1. We observe low employment rates for this population: 27% for women and 40% for men. These low rates highlight the need to account for systematic selection into working when modeling intensive labor supply decisions.

3 Model and Empirical Strategy

In this section, we build on the tax literature to develop a basic theoretical framework for modeling intensive and extensive labor supply responses to income taxes. We use this model to derive our empirical specifications.

3.1 Theoretical Framework

We consider a static framework where an individual maximizes utility that is a function of consumption and labor. The budget constraint includes labor income, non-labor income (assumed exogenous here but not in the empirical analysis) and tax liability which is a function of both labor earnings and non-labor income. The utility function also includes a parameter related to the cost of working and is similar in spirit to the model found in Eissa et al. (2008). The individual solves the following maximization problem:

$$\max_{c,L} U(c, L) - \mathbf{1}(L > 0)q \quad \text{s.t.} \quad c = L + y^o - T[L + y^o]$$

where c represents consumption, L is labor earnings ($U_L < 0$), y^o is non-labor income, and $y = L + y^o$ is total income. $T[y]$ is total tax liability given total income y and is non-linear in y. q represents a fixed cost of working. The fixed cost of working is equal to zero for those who do not work and we assume $q > 0$.

[10] One caveat is that the survey asks respondents about their current self-reported labor force status, while the labor earning and tax variables all relate to the previous year. We do not view this as a limitation for our analysis, however, given that we are using the "retirement" variable as a more permanent indicator of leaving the labor force.

A. Intensive Margin

If we assume an interior solution, then the first-order conditions imply:

$$\frac{U_L}{U_c} = -(1-\tau), \qquad c = L + y^o - T[L + y^o].$$

where τ represents the marginal tax rate ($T' = \tau$). The insight from these equations is that changes in labor earnings (conditional on working) are a function of changes in $1-\tau$ (the marginal net-of-tax rate) and changes in $L+y^o-T[L+y^o]$ (after-tax income). Labeling after-tax income as ATI, the model states that labor earnings can be written as $L = L(1-\tau, ATI)$. Consequently, the tax schedule alters intensive labor supply in the following manner:

$$dL = -\frac{\partial L}{\partial(1-\tau)}d\tau + \frac{\partial L}{\partial ATI}dATI. \tag{1}$$

Define $\zeta^I = \frac{1-\tau}{L}\frac{\partial L}{\partial(1-\tau)}$ and $\eta^I = \frac{ATI}{L}\frac{\partial L}{\partial ATI}$. The I superscript is used to denote that these are intensive margin elasticities. ζ^I is the intensive margin substitution effect, which we interpret as a compensated elasticity, given that we are holding after-tax income constant. η^I is the intensive margin income effect.[11] Substituting these terms into equation (1), we get

$$\frac{dL}{L} = -\zeta^I\frac{d\tau}{1-\tau} + \eta^I\frac{dATI}{ATI}.$$

The corresponding regression specification is

$$\ln L = \alpha^I + \zeta^I \ln(1-\tau) + \eta^I \ln\left[y - T(y)\right] + \epsilon^I.$$

Gruber and Saez (2002) note that the effect of the marginal net-of-tax rate (substitution effect) and the effect of after-tax income (income effect) can be separately identified empirically due to the non-linearities in the budget constraint. The budget constraint is non-linear because the tax schedule sets different marginal tax rates for distinct segments of total income. Changes in the marginal tax rate are the same for everyone on the same segment (i.e., tax bracket) of total income, but changes in after-tax income vary depending on a person's distance from the kink in the budget constraint.

[11] The model in Gruber and Saez (2002) begins with taxable income as a function of $1-\tau$ and virtual income but concludes with the same specification as equation (3.1) and interprets the parameters in the same manner that we do.

Figure 1 plots an illustrative case. For simplicity, we graph the nonlinear budget set created by a tax schedule with two tax brackets. After-tax income is an increasing, nonlinear function of taxable income. We consider the case where the marginal tax rate in the top bracket is reduced between periods $t = 0$ and $t = 1$, while the tax rate in the lower tax bracket remains constant. Person A is located in the lower tax bracket, while persons B and C are in the top tax bracket. Comparing A and B, it is clear that the tax schedule change reduces the marginal tax rate for person B, while leaving the marginal tax rate for person A unaffected. Comparing B and C, we observe that while both individuals experience the same change in the marginal tax rate, they experience different changes in after-tax income (labeled ΔATI). Thus, the marginal tax rate and after-tax income are separately identified due to the nonlinearities in the budget constraint.

B. Extensive Margin

Individuals may decide not to work and this decision is also related to the tax schedule. In the above equations, we can solve for interior solutions c^* and L^*. Then, we can compare the utility from working to the utility from not working. Consider an individual that is indifferent between working and not working:

$$U(L^* + y^o - T[L^* + y^o], L^*) - q = U(y^o - T[y^o], 0).$$

For the extensive margin, the model tells us that the decision to work (W) is dependent on after-tax income, after-tax non-labor income ($ATNI$), and the individual's labor earnings if they work (represented by L): $W = W(ATI, ATNI, L)$. Changes in working status due to tax schedule changes can be decomposed into three variables:

$$dW = \frac{\partial W}{\partial ATI} dATI + \frac{\partial W}{\partial ATNI} dATNI + \frac{\partial W}{\partial L} dL.$$

Define $\zeta^E = \frac{ATI}{W} \frac{\partial W}{\partial ATI}$ and $\eta^E = \frac{ATNI}{W} \frac{\partial W}{\partial ATNI}$. These are the extensive margin substitution effect and income effect elasticities, respectively. We interpret ζ^E as a compensated elasticity given that we hold non-labor after-tax income constant. This term tells us how individuals react to additional after-tax income if they work, holding the after-tax income they would receive if they did not work constant (i.e., this term is the additional amount of after-tax income due to working). Finally, $\omega^E = \frac{L}{W} \frac{\partial W}{\partial L}$, which relates the disutility of additional labor earnings to the probability of working. Following from the model, holding after-tax (labor and non-labor) income constant, additional pre-tax labor earnings reduces utility. Individuals who must earn more in pre-tax labor earnings to make the same amount in after-tax income are less likely to work. Substituting these parameters into the above equation, we arrive at the following relationship:

$$\frac{dW}{W} = \zeta^E \frac{dATI}{ATI} + \eta^E \frac{dATNI}{ATNI} + \omega^E \frac{dL}{L}.$$

From this equation, we derive our regression specification

$$P(\text{Work} = 1) = F\left\{\phi^E + \beta^E \ln\left[L + y^o - T(L + y^o)\right] + \theta^E \ln\left[y^o - T(y^o)\right] + \rho^E \ln L + \epsilon^E\right\}.$$

This specification differs slightly from some papers in the tax literature that have studied the decision to work (e.g., Eissa and Hoynes (2004); Gelber and Mitchell (2011)), though it follows naturally from our model. Other studies have started with a specification which assumes the decision to work is a function of $\frac{L-(T(L+y^0)-T(y^o))}{L}$, which is the share of labor earnings that an individual keeps if they work. This variable is related to our specification, except that it combines all three of our variables into one term. Taking the log of this term produces a similar expression as found in the equation above: $\ln[L + y^0 - T(L + y^0) - (y^0 - T(y^0))] - \ln L$. The main difference is that this measure takes the log of the *difference* in after-tax earnings from working versus not working, whereas we include the log of after-tax income and the log of after-tax non-labor income separately. This allows us to estimate the response to the *proportional* increase in after-tax income from working (i.e., allowing for different effects depending on whether the person has a small or large amount of unearned income), rather than the level increase and is consistent with the intensive margin equation which assumes behavioral responses to the log of after-tax income. Furthermore, our specification includes the log of pre-tax labor earnings as a separate variable. Converting $\ln\left[L + y^o - T(L + y^o)\right]$ to a rate variable would require the assumption that $\omega^E = -\zeta^E$, a restriction that is not implied by our derivation and one that we do not enforce.

Nonlinearities in the tax schedule can also be used to separately identify the effects of the additional after-tax income earned due to working – after-tax labor income (ATLI)[12] – from after-tax income and the marginal net-of-tax rate, as illustrated in Figure 2. Unlike

[12] We define ATLI as the additional amount of after-tax income received due to working. Holding after-tax income fixed, changes in ATLI are perfectly and inversely related to changes in ATNI. Thus, when we discuss our selection instrument, it is equivalent to discuss identification of ATNI and identification of ATLI. An increase in ATLI is an additional incentive to work since it implies additional monetary gains to working in after-tax dollars. A decrease in ATNI, holding ATI constant, implies the equivalent incentive to work.

prior work which uses cross-sectional heterogeneity in non-labor income alone as identifying variation, we show that tax code changes (conditional on after-tax non-labor income) identify independent variation in after-tax labor income.

Consider two different people who initially have identical pre-tax total income (marked C, as in Figure 1) but different levels of non-labor (NL) income (e.g., spousal earnings). Neither this identification discussion nor the empirical strategy assumes that non-labor income is exogenous. Non-labor income for person 1 and 2 are represented by C_1^{NL} and C_2^{NL}, respectively. The additional after-tax income earned by person 1 due to working is represented by the vertical distance between C_1^{NL} and C. Suppose the tax rate decreases for the top tax bracket between periods $t = 0$ and $t = 1$, as before. After-tax labor income (labeled $\Delta ATLI_{C1}$ and $\Delta ATLI_{C2}$), the pecuniary incentive to work, has increased more for person 1 than for person 2 following the tax cut. Holding everything else constant, person 1 benefits from the tax cut only if she works. However, person 2 benefits from the tax cut regardless of whether or not she works, so the additional amount she earns due to the tax cut if she works relative to if she does not work is smaller. The benefits of working have increased for person 2 but have increased even more for person 1. Equivalently, the tax change increases after-tax non-labor income (i.e., ATNI) for person 2, but does not change after-tax non-labor income for person 1. Thus, differential policy-driven changes in ATLI (or, equivalently, ATNI), holding the tax rate and ATI constant, result from variation in distance from the kink using *only* non-labor income.

We can find two people who experience the same change in the marginal tax rate and after-tax income, but experience different *changes* in after-tax labor income (equivalently, different changes in ATNI) due to a legislative tax change while holding all components of pre-tax income fixed. Consequently, all exogenous variation originates from legislative changes interacted with initial differences in income (and types of income). As explained below, we control for fixed effects which account for the independent effects of initial differences in earnings, non-labor income, etc. Thus, we are *not* using non-labor income variation alone to identify the participation equation but, instead, leveraging these differences to exploit quasi-experiment variation. We take advantage of the separate identification of the marginal tax rate, after-tax income, and after-tax non-labor income to estimate the intensive and extensive margin effects of income taxes.

3.2 Empirical Strategy

Our empirical strategy models and estimates the impact of taxes on both the intensive and extensive margins of labor supply for older workers, using the insights of the above theoretical framework. We condition on a set of individual characteristics (X_{it}) and generate simulated instruments using baseline income information based on these exact same characteristics.[13] While we do not have individual-level longitudinal data, we control for the direct effects of X_{it} such that there is no identifying variation due to differences between two people under the same tax code. We discuss the intensive margin first, followed by the extensive margin, and then introduce the simulated instrument strategy.

3.2.1 Intensive Margin Effect

We begin by modeling intensive labor supply decisions, measured as labor earnings.[14] Given that the labor supply literature has consistently found that men and women respond to labor market incentives in different ways, we perform all analyses separately by gender. Our specification models changes in labor earnings as a function of changes in the marginal net-of-tax rate (substitution effect) and changes in after-tax income (income effect). This equation is similar to the main specification used in the elasticity of taxable income literature:

$$\ln L_{it} = \alpha_t + X'_{it}\delta + \beta^I \ln(1 - \tau_{it}) + \theta^I \ln\left(y_{it} - T_t(y_{it})\right) + \epsilon_{it}, \tag{2}$$

where L is own-labor income (pre-tax), τ is the marginal tax rate, such that $\ln(1 - \tau_{it})$ represents the log of the marginal net-of-tax rate for person i in period t. y is total household income (including non-labor income) and $T(y)$ is total tax liability for income y. $\ln\left(y_{it} - T_t(y_{it})\right)$ is the log of after-tax income. The specification includes year effects, and X is a vector of covariates. These covariates are important because our identification strat-

[13] With individual-level panel data, we could predict tax information in each year for the person based on initial income while conditioning on individual fixed effects. With repeated cross-sections, the X_{it} variables account for unobserved heterogeneity (in the same way individual fixed effects would) and the instruments vary based only on the *interaction* of the tax code with X_{it}.

[14] We use labor earnings as our primary outcome of interest because it is the product of a host of choices that may respond to tax incentives such as hours worked, amenity preferences, and effort. Labor earnings is thus a useful summary metric that combines all of these components. We are also specifically interested in the potential ramifications of policies that alter older individuals' incentives to work and the subsequent impact on earnings as a means of supplementing or replacing Social Security benefits. Moreover, our model suggests that individuals respond to the additional income earned by participating in the labor force so we need to estimate labor earnings for each individual in our sample to construct this measure. The tax code can and does tax labor income in a different way than it taxes other income. For drawing policy implications, it is important to understand how labor income responds to taxes independent of other sources of income.

egy compares outcomes of individuals with the same covariates over time. Our covariates include indicators based on age group,[15] race,[16] education,[17] and marital status. Because spouses may coordinate their labor force decisions, we also include indicators for spousal age groups,[18] spousal race, and spousal education.[19]

We restrict estimation of equation (2) to individuals with positive labor earnings, which motivates our concerns about systematic selection (discussed in Section 3.3.2). The substitution and income effects are separately identified using legislative tax schedule changes so that β^I can be interpreted as a compensated elasticity (Gruber and Saez (2002)). We expect this parameter to be positive.

3.2.2 Extensive Margin Effect

We also estimate extensive margin effects. According to our theoretical model, an individual's decision to work is a function of the amount of after-tax income that she would make if she worked, the amount of after-tax income she would make if she did not work, and the pre-tax labor earnings if she worked. We model the extensive margin following Section 3.1.B:

$$P\left(\text{Work}_{it}=1\right)=F\left(\phi_t+X_{it}'\gamma+\beta^E\ln\left[L_{it}+y_{it}^o-T_t(L_{it}+y_{it}^o)\right]+\theta^E\ln\left(y_{it}^o-T_t(y_{it}^o)\right)+\rho^E\ln L_{it}+\nu_{it}\right),\quad (3)$$

where y_{it}^o is non-labor income. $T(L+y^o)$ represents the total tax liability if the individual works and earns labor income L. $T(y^o)$ is the individual's tax liability if they do not work. This specification models the probability of working in period t as a function of the income in after-tax dollars that the individual receives if she works, conditional on the income in after-tax dollars that the individual receives if she does not work. We interpret β^E as a compensated elasticity and expect it to be positive for the probability of working. We expect θ^E to be negative as increasing after-tax non-labor income provides a disincentive to work. The specification also includes the term $\ln L_{it}$, which is derived from our model. Additional after-tax income for working should induce more people to work, but it is also

[15] We include indicators for age groups 62-63, 64-67, 68-71, and 72-74.

[16] Using the Census and ACS categories, we include indicators for white, black, and other.

[17] We categorize educational attainment into 4 groups: less than high school, high school graduate, some college, and college graduate.

[18] We use the same age groups as before, except that we also include indicators for spouses outside of the 62-74 range.

[19] The spousal variables are all equal to 0 for single individuals.

important to account for the amount of pre-tax labor earnings that are needed to achieve a given level of after-tax income. For a fixed $\ln[L_{it} + y^o_{it} - T_t(L_{it} + y^o_{it})]$, we would expect that a larger L would deter work. Thus, we expect ρ^E to be negative and interpret it as a measure of the disutility of additional work, since higher L represents more hours and effort for a given level of after-tax income. L_{it} and, consequently, $L_{it} + y^o_{it} - T_t(L_{it} + y^o_{it})$ are not observed for non-workers. We discuss how we address this missing data issue in the next section. We also estimate equation (3) with "retired" as the outcome variable (defined in Section 2).

3.3 Identification Challenges

Equations (2) and (3) pose a few identification challenges. First, changes in labor earnings mechanically increase tax rates and tax liabilities such that OLS will not provide consistent estimates of equation (2). Similarly, in equation (3), individuals with higher L (and consequently, higher tax liabilities) may be more likely to work for reasons unrelated to after-tax earnings. Second, we do not observe L for individuals who do not work, and L is important for constructing two of the variables in the extensive margin equation. Third, we can only estimate the intensive margin equation (2) for a selected sample of individuals who participate in the labor force. We discuss how we address these endogeneity and selection issues here.

3.3.1 Instruments

To address the mechanical relationship between earnings and taxes, we create a set of instruments to isolate plausibly exogenous variation in the tax variables. Our two structural equations (2) and (3) include three tax-related variables: the marginal net-of-tax rate, after-tax income, and after-tax non-labor income. We implement an instrumental variable strategy that exploits independent variation in these tax-related variables derived solely from legislative tax schedule changes.

We take advantage of changes in federal tax policy during our study period that changed tax-based incentives for reasons unrelated to individual changes in labor supply. During our sample period, there were two key tax reforms: 1) the Economic Growth and Tax Relief Reconciliation Act (EGTRRA) of 2001, which reduced tax rates for nearly every tax bracket with especially large changes for those with low income; 2) the Jobs and Growth Tax Relief Reconciliation Act (JGTRRA) of 2003, which also reduced tax rates, primarily focusing on relatively higher income households. Figure A1 shows changes in the federal marginal tax rate across our study period. For married couples filing jointly, reductions in the marginal tax rate over this time period ranged from 0 to 46 percent, depending on the household's adjusted gross income.

We employ a simulated instrumental variables strategy which exploits the differential effect that these tax policy changes had on households based on their observed characteristics. This strategy is motivated by Gruber and Saez (2002) but shares similarities with other simulated instrumental variable strategies (Currie and Gruber (1996a,b)). Our implementation is most similar to the approach used in Burns and Ziliak (2017) and Powell (2019).

Our instrumental variable strategy takes the 1999 sample and runs these observations through TAXSIM to calculate the tax variables for each observation for each year from 1999 through 2007. We use the 1999 sample even when constructing instruments for 2000-2007. We then estimate the relationship between the simulated tax variables and the covariates defined by X in equations (2) and (3) in each year and predict values for each tax variable based only on covariates. Finally, each observation is assigned the predicted values based on their covariates and the year – these are the "simulated instruments." Consequently, two individuals with the same covariates X (which we control for in our specifications) will be assigned different values of the instrument only because they face different tax schedules due to legislative changes.

Our method for constructing the instruments can be summarized by the following steps. We discuss construction of the simulated log of the marginal net-of-tax rate. The instruments for the other tax variables are generated using the same procedure.

1. Holding real income and household characteristics constant, we simulate the log of the marginal net-of-tax rate for every observation in the 1999 sample assuming that they were subject to the year s tax code. We represent this variable by $\ln(1-\tau_{i,1999}^{s})$, where $\tau_{i,1999}^{s}$ is the tax rate that person i in year 1999 would have faced under the year s tax schedule given the same real income and household characteristics. We implement this step for $s = 1999, \ldots, 2007$.
2. For each s, we regress $\ln(1-\tau_{i,1999}^{s})$ on X, where X is the same vector of covariates as in the intensive and extensive margin equations. Let δ_s represent the coefficients on X for the regression for year s. These coefficients parameterize the relationship between

covariates and the predicted log of the marginal net-of-tax rate.

3. Using these coefficients, we predict the log of marginal net-of-tax rate for each observation in the sample in each year: i.e., $\widehat{\ln(1-\tau^s_{it})} = X'_{it}\widehat{\delta}_s$.

4. Observations in year t are assigned $\widehat{\ln(1-\tau^t_{it})}$.

The sample and covariates are held constant across regressions in Step 2. Moreover, the inputs used to generate $\ln(1-\tau^s_{it})$ are identical across years with the slight caveat that we adjust all income measures for inflation to hold real income constant.[20] Consequently, the *only* reason that $\ln(1-\tau^s_{i,1999})$ varies for $s \neq s'$ is because the tax code has changed. As a result, our instruments only vary from one year to another due to tax code changes.

Restricting the instruments to vary based only on covariates allows us to account for the *exact* function that generated the instruments in our specifications. We control for X_{it}, accounting for the independent effects of household characteristics on labor supply changes, and we control for the tax policy changes (i.e., year fixed effects) in our extensive and intensive margin equations. Thus, identification originates solely from the interaction of X_{it} and the tax policy changes. The X_{it} variables in the main specifications also control for the independent effects of differences in baseline earnings and non-labor income (i.e., our method does not require exogenous non-labor income).[21] In the end, we construct three instruments using this method: predicted log of the marginal net-of-tax rate ($\widehat{\text{MTR}_{it}}$), predicted log of after-tax income ($\widehat{\text{ATI}_{it}}$), and predicted log of after-tax non-labor income ($\widehat{\text{ATNI}_{it}}$). We generate all instruments separately by gender. Note that the tax variables in equations (2) and (3) include state taxes, but we do not use state policy variation for identification.[22]

The first and second instruments will be used for identification of equation (2). The third instrument will be included as a determinant of selection into labor force participation in our selection equation, which we discuss below. All three instruments are used to identify

[20] We use the inflation rate since the tax code is typically indexed to inflation.

[21] The tax literature following Gruber and Saez (2002), predicts tax changes based on initial income and characteristics and changes in the tax code. One concern with this approach is that the instruments, which are functions of initial income, may be correlated with mean reversion and income trends (Weber (2014); Burns and Ziliak (2017)). Predicting the instruments based on covariates alleviates these concerns and our approach should be less susceptible to biases resulting from mean reversion and secular trends.

[22] To the extent that certain covariates predict differential exposure to various state tax schedule changes, then that variation is exploited. However, we are not including state fixed effects or directly predicting tax changes using state information.

the extensive margin equation.[23]

In addition, we also construct a measure of the predicted (based solely on individual and spousal characteristics) probability of working in 1999 using the exact same approach to derive the instruments. We interact this predicted measure with year indicators (except 1999) and include these variables in all specifications to account for potential differential trends in labor force behavior, similar in spirit to controls included in Burns and Ziliak (2017) to account for secular trends. We test more explicitly for confounding trends in Section 4.5.

3.3.2 Selection

Our second identification challenge is that we do not observe labor earnings for individuals who do not work. The concerns that arise from this are two-fold. First, the intensive margin labor supply equation is estimated for a selected sample of individuals who work. Second, the extensive margin labor supply equation contains two variables which are unobserved for non-workers.

We address these issues by estimating the intensive and extensive margin labor supply equations together. Combining the extensive and intensive margin equations is helpful for two reasons. First, the extensive margin equation provides a useful exclusion restriction to identify the selection mechanism in the intensive margin labor supply equation. This point has been recognized frequently in the labor supply literature which typically relies on distributional assumptions, number of children, and/or non-labor income to separately identify the participation equation. To control for selection in the intensive margin equation, we need an instrument that affects labor force participation, but does not independently affect labor earnings conditional on participation. Fortunately, the extensive margin equation (equation (3)) includes a variable that is excluded from the intensive margin equation (equation (2)), after-tax non-labor income, which is identified with quasi-experimental tax variation. Thus, we can use predicted after-tax non-labor income as an exogenous shock to employment. This is an ideal instrument for selection in the intensive margin equation since after-tax non-labor income affects labor force participation, but does not – conditional on the marginal net-of-tax rate and after-tax income – independently affect labor income. We find

[23] The extensive margin equation includes three endogenous variables. The use of $\widehat{\text{ATI}}_{it}$ and $\widehat{\text{ATNI}}_{it}$ as instruments to identify the extensive margin equation is straightforward. $\widehat{\text{MTR}}_{it}$ is also an appropriate instrument because it independently shocks $\ln L_{it}$. This independent variation is necessary to identify the equation.

that this selection instrument has a strong relationship with labor force participation.

We use this selection instrument to estimate a probit model of labor force participation (Heckman (1979)) and condition on a flexible function of the estimated index in the intensive margin equation to adjust for selection. In our preferred specifications, we use a semi-parametric approach to correct for selection which does not assume normality of the error term and allows us to identify solely from independent variation in the excluded instrument, not distributional assumptions. In this semi-parametric approach, we assume:

$$E\left[\epsilon_{it}|\widehat{\text{MTR}_{it}},\widehat{\text{ATI}_{it}},\widehat{\text{ATNI}_{it}},X_{it},\alpha_t,\text{Work}_{it}=1\right]=\lambda(\mathbf{W}'_{it}\zeta). \tag{4}$$

where $\mathbf{W}$ includes our instruments for the intensive labor supply equation, the selection instrument, and all exogenous variables in equation (2). We do not assume any functional form for $\lambda(\cdot)$ and instead use a series approximation, as suggested in Newey (2009). We estimate the selection equation using the monotone rank estimator introduced in Cavanagh and Sherman (1998), which requires no distributional assumptions to obtain consistent estimates (up to scale).

Second, estimating the intensive and extensive equations together is useful because the intensive labor supply equation provides consistent predictions of labor earnings for non-workers and we can use these predictions to estimate the extensive margin labor supply equation. After we have estimated the intensive labor earnings equation, we predict earnings and calculate tax variables for each person in the sample, including those who do not have any labor earnings. We use these estimates to construct the otherwise unobserved labor earnings (and related tax variables).

3.4 Implementation

Our method for estimating the intensive and extensive margin labor supply equations proceeds in four steps. We describe the technical details in Appendix Section A. First, we estimate the selection equation and predict the selection adjustment term. Second, we estimate the intensive labor supply equation using 2SLS, conditioning on a flexible function of the selection adjustment term. Because the selection adjustment term is estimated, we bootstrap for inference which accounts for the inclusion of an estimated term in the intensive margin equation. Third, we use the parameter estimates from this equation to predict labor earnings for the entire sample including those who do not work. We also estimate tax liabilities and after-tax income given these labor earnings estimates. Fourth, we estimate the extensive margin equation using the estimated labor earnings and after-tax income variables derived from the intensive margin equation.

4 Results

Before discussing the regression results, we provide suggestive graphical evidence relating the tax incentives to work to the probability of working. We use our instruments for after-tax income and after-tax non-labor income to generate the additional log income that an individual receives from working: $\widehat{\text{ATI}_{it}} - \widehat{\text{ATNI}_{it}}$. We create "cells" based on the covariates included in our regression: age group, education, race, spousal age group, spousal education, spousal race, and gender. We calculate the change in the predicted tax incentives to work for each cell from 1999 to 2007. We then divide the cells into terciles based on the change in the predicted tax-based incentives and graph this against the change in the fraction of individuals working in those cells over the same time period. Figure A2 shows this relationship. This approach mimics our empirical strategy, linking predicted changes in the tax-based incentives to work to changes in the probability of working. We find a positive monotonic relationship between our instruments, representing the predicted change in the incentives to work, and the fraction of individuals working.

Next, we present our regression results in the order that the equations are estimated: selection equation, intensive margin equation, and extensive margin equation. For the latter two equations, we include estimates with (1) no selection adjustment; (2) a selection adjustment method generated by a probit regression; (3) a semi-parametric selection adjustment method. In the extensive margin estimation, the type of selection adjustment refers to the method used to impute earnings (and the corresponding tax variables). We provide the three sets of results for comparison but favor the semi-parametric estimates since they account for selection relying only on variation in our excluded selection instrument. Whether the selection adjustments matter in practice is an empirical question, and we present multiple approaches to assess their importance.

4.1 Selection Adjustment

In Table 1, we present results for the selection equation separately for women and men. Column (1) shows the results from a probit regression estimating the probability of working as a function of the three tax instruments and covariates. The predicted log of after-tax non-labor income is excluded from the intensive margin equation, separately identifying the selection adjustment term. Column (2) presents semi-parametric estimates of the same selection equation using the monotone rank estimator. The monotone rank estimator estimates the index without any distributional assumptions. We normalize coefficients in Table 1 so that the sum of the square of all coefficients in one model is equal to 1. Columns (3) and (4) show the analogous results for men. We estimate a negative coefficient on the excluded instrument. As we would expect, additional after-tax non-labor income decreases the probability of working. Conditional on after-tax non-labor income, after-tax income (if the individual works) increases the probability of working. The estimates for the excluded selection instrument are statistically significant regardless of the estimation procedure used.

For both women and men, we have identified a variable which predicts labor force participation and is theoretically excluded from the intensive labor supply equation. We use the predictions from these estimates in our intensive margin estimation to account for selection.

4.2 Intensive Margin

Table 2 presents the results from 2SLS estimation of the intensive labor supply equation. We interpret the coefficients on the marginal net-of-tax rate as compensated elasticities since we separately account for income effects (Gruber and Saez, 2002). When no selection adjustment is made (Column (1)), we estimate an elasticity of 1.830 for women. This estimate is large and statistically significant from zero. We also estimate a positive income effect, which we also would not expect. When we account for selection in Column (2), the coefficient on the marginal net-of-tax rate variable decreases in magnitude to 1.526. The Column (2) estimate imposes distributional assumptions on the selection equation. When we relax this assumption in Column (3), we estimate a similar effect of 1.716. This estimate is statistically significant at the 1% level. Our income effect estimate is close to zero, though imprecisely estimated.

For men, we estimate elasticities of 3.269 without the selection adjustment and 3.718 and 3.035 with the probit and semi-parametric selection adjustments, respectively. The estimates are statistically significant at the 1% level. We cannot reject that the income effect is equal to zero and the confidence intervals are quite large.

We have little existing empirical data on the intensive margin responsiveness of older workers to taxes to make an appropriate comparison. The finding that men are more responsive than women is different than what we would normally expect to find for younger populations. However, it may simply be the case the older men are extremely responsive to intensive margin incentives while older women who work are also very responsive but not to the same extent.[24] Overall, we find that both women and men are very responsive to marginal tax rates at older ages, which is consistent with findings of intensive margin responsiveness to the earnings test given that we would expect even more responsiveness to a pure tax.

Interestingly, the intensive margin substitution effect estimate does not appear to be very sensitive to adjustments for selection. However, accounting for possible selection bias has advantages in our empirical approach beyond estimating consistent coefficients for the marginal net-of-tax rate and after-tax income variables. The selection adjustment allows for consistent estimation of *all* the parameters in the intensive margin equation. This is critical because we use the parameters from estimating the intensive margin equation to generate predicted labor earnings for the extensive margin equation. These predicted earnings are a function of all covariates included in the specification, and it is possible that selection may bias the coefficients estimated for other included variables, and consequently, the labor earnings predictions that are generated from this equation. Our selection corrections address this concern.

4.3 Extensive Margin

After estimating the intensive margin equation, it is possible to predict the log of after-tax income if working and the log of pre-tax labor earnings for everyone in the sample, including individuals who do not work. The extensive margin equation also includes the log of after-tax non-labor income. These three variables are endogenous and we use all three tax instruments to identify the extensive margin equation.

Table 3 presents the instrumental variable estimates for the extensive margin equation using 2SLS.[25] In Column (1), using earnings predicted from the intensive margin equation without a selection adjustment, we estimate that a 1% increase in after-tax income –

[24]We will find that women are much more responsive on the extensive margin. These results suggest that men may decide to work or not in a manner less dependent on taxes than women, but then they decide how much to work based on tax incentives. Women, however, decide whether to work based on tax incentives but then these is less (but still substantial) variation in how much they decide to work.

[25]IV-Probit generates similar estimates.

holding after-tax non-labor income constant – *decreases* the probability of working by 0.099 percentage points. We estimate a negative relationship between after-tax non-labor income and the probability of working, as we would expect. Finally, holding after-tax income constant, we estimate that higher pre-tax labor earnings increase the probability of working, which is the opposite relationship that we would expect.

When we adjust for selection using the parametric adjustment, the estimates have the same sign. Column (3) shows the results when the semi-parametric adjustment is made, our preferred approach. The estimates are the expected sign for all three variables. We estimate that a 1% increase in after-tax income increases the probability of working by 1.0 percentage points for women. We also estimate that an increase in after-tax non-labor income decreases the probability of working. Finally, a 1% increase in pre-tax labor income decreases the probability of working by 0.055 percentage points. We also find that, conditional on the after-tax monetary return to work, higher pre-tax earnings (i.e., more effort to make the same amount in post-tax dollars) reduces the probability of working.

For men, the results are more consistent across the different models. The Column (6) estimates imply that a 1% increase in after-tax income increases the probability of working by 0.284 percentage points for men. An increase in after-tax non-labor income decreases the probability of working. Finally, a 1% increase in pre-tax labor income decreases the probability of working by 0.187 percentage points.

Our implied elasticities with respect to after-tax income are large – 3.9 for women and 0.7 for men – due to the low labor force participation rates of these populations.[26] These estimates are not directly comparable to the previous literature since we are estimating the labor force participation responsiveness with respect to after-tax income (holding after-tax non-labor income constant), whereas previous studies typically estimate participation elasticities with respect to after-tax *labor* income, defined as labor earnings minus the additional taxes paid because of those labor earnings.[27] A 10% increase in after-tax income (holding after-tax non-labor income constant) is much larger than a 10% increase in after-tax labor income and should generate larger behavioral responses. In our data, a 10% increase in after-tax income is about 5 times as large as a 10% increase in after-tax labor income.

[26]Among individuals ages 62 to 74, the probability of working is 0.272 for women and 0.397 for men. We use these probabilities to construct elasticities from the estimates from Table 3. We calculate these elasticities using $\widehat{\beta^E} \times \frac{1}{P(work)}$.

[27]Our paper departs from the literature in several other ways as well.

Eissa et al. (2008) summarizes the literature on the effects of taxes on non-elderly female household heads as finding a central value of 0.7 while Hotz and Scholz (2003) reports participation elasticities as large as 1.69.[28] Overall, our results suggest meaningful scope for impacting labor force participation of older individuals through the tax code. The large elasticities are consistent with recent research that this age group is very responsive to the Earnings Test on the extensive margin (Gelber et al., 2020b), even though this policy is not a tax. We also note below (in Section 4.4) that our results imply much smaller employment effects (Laitner and Silverman, 2012) or similar effects (Gustman and Steinmeier, 2015) than those found in the literature for this age group.

In Table C.2, we study retirement. Focusing on the semi-parametric results (Columns (3) and (6)), we find that the estimates have similar magnitudes as the estimates for working, suggesting that the large increases in working due to after-tax income are almost entirely driven by reductions in retirement.[29] These results suggest that tax changes may have longer-term employment effects for the 62-74 population.

4.4 Policy Simulations

In this section, we simulate the labor force participation ramification of eliminating the employee portion of the payroll tax at age 65.[30] Since this policy increases the generosity of the tax code at specific ages, implementation of these polices may also have dynamic effects if we believe that individuals will shift their labor supply to periods in the life-cycle where they would earn more. We cannot study this possibility without imposing more restrictive assumptions, but these estimates may be most relevant for understanding the short-term consequences of the introduction of these policies. The following estimates are likely lower bounds of the permanent effects if individuals would delay some of their labor supply until the period when they face lower taxes. We also do not quantify any corresponding labor supply reductions at younger ages.

We assume that an equivalent lump sum tax is levied on each person such that we can ignore income effects.[31] For each person, we predict (1) the probability of working under

[28]Hotz and Scholz (2003) summarize participation elasticities for women in the literature: 0.85 in Dickert et al. (1995), 1.16 in Eissa and Liebman (1996), 0.96 in Keane and Moffitt (1998), 0.70 in Meyer and Rosenbaum (2001), 0.29 in Eissa and Hoynes (2004), and between 0.97 and 1.69 in Hotz et al. (2010).

[29]This pattern of results is not surprising given that, for this age group, almost the entire population that is not working identifies as being out of the labor force.

[30]Laitner and Silverman (2012) eliminates both the employee and employer portions.

[31]We also hold pre-tax labor income constant.

the current tax rules in period t and (2) the probability of working under the "counterfactual" tax rules (i.e., eliminating the employee portion of the payroll tax), substituting in $\ln\left[L_{it} + y^o_{it} - T^c_t(L_{it} + y^o_{it})\right]$, where T^c_t represents the counterfactual tax burden in period t given the elimination of the employee portion of FICA taxes. The difference in these probabilities gives us the effect of this policy on labor supply behavior. Table 4 shows the results of this simulation.

In our baseline sample, 27.2% of women and 39.7% of men earn positive wages. Elimination of the employee FICA taxes would increase this percentage by 4.3 percentage points for women and 1.4 percentage points for men. We estimate nearly equivalent reductions in retirement. The results imply a 16% increase in the fraction of women working and a 4% increase in the fraction of men working. Laitner and Silverman (2012) find that the elimination of the full payroll tax would, on average, extend working lives by one year. Our estimates imply smaller changes, but these effects are still large and economically significant. Our results are also similar in magnitude – though slightly larger – than those found in Gustman and Steinmeier (2015).

4.5 Sensitivity Analysis

In this section, we study the robustness of our results. For each test, we repeat our entire procedure and use the semi-parametric selection adjustment method as our preferred specification. The robustness results are reported in Table C.3. The first row shows the intensive margin results for the coefficient on the marginal net-of-tax rate variable only (corresponding to Table 2). The second row presents the extensive margin estimates (corresponding to Table 3). The third row shows the results from the elimination of the employee portion of the payroll tax simulation.

First, in columns 1 and 2 of Panel A, we repeat our main findings. Second, we test for the possibility that part-time work transitions are affecting our estimates. It is possible that equation (2) does not fully encapsulate discrete intensive margin decisions such as the decision between full-time and part-time work. In principle, we could model this behavior explicitly.[32] Note that the cited literature on older workers and taxes models labor supply behavior as a decision between full-time work and not working, ignoring all intensive margin decisions. Thus, our approach represents a contribution by including an intensive margin

[32]This extension would require an additional equation and an equivalent instrument for the pecuniary incentives to work part-time (versus full-time).

decision. Instead of adding yet another dimension, we simply evaluate the possible effects of assuming that individuals respond to the marginal tax rate for intensive labor supply decisions. For working individuals that report usually working less than 35 hours per week, we calculate their hourly wages and then recalculate their labor earnings, assuming that they worked 35 hours per week. The intensive margin results are similar to the main estimates. The extensive margin estimates are smaller for women but larger for men.

Third, we previously noted that the literature has found that the Social Security Earnings Test affects labor supply decisions for the older population. The incentives resulting from the Earnings Test changed during our time period and one concern is that these changes might also have different effects over the income distribution and correlate with our tax instruments. The most significant policy change over our sample period occurred in 2000 when the earnings test was removed for individuals at the full retirement age or older as part of the Senior Citizens Freedom to Work Act of 2000. Before 2000, the test applied to individuals claiming Social Security up to age 69. For columns (5) and (6), we constructed the implied Social Security earnings test tax rate and liability and generated predicted values in the same manner as our tax instruments. We included these variables in our main specifications. The results are similar to our main estimates, suggesting that changing incentives implicit in the earnings test are not confounding our analysis.

Finally, we study whether the results are driven by underlying earnings and employment trends over this time period. We include one-year leads of the main policy variables to test whether the behavioral effects load on the year t or the year $t+1$ measures. Because we do not observe next year's individual-level tax variables, we include leads of the predicted values (i.e., the instruments). Alternatively, we can replace the year t variables with predicted variables as well (to place all variables on equal footing) and the results are similar. The results are presented in Table C.4. We find that the main estimates are generally unaffected by including leads. Moreover, the estimates associated with the lead variables tend to be much smaller in magnitude and are never statistically different from zero at the 5% level. Underlying trends are unlikely to be associated with year t tax variables so precisely.

5 Conclusion

This paper models both the intensive and extensive margins of labor supply, using each margin to enable consistent estimation of the other. Both of these equations pose challenges for estimation even with appropriate instrumental variables due to possible selection bias and unobserved earnings. The extensive labor supply equation, however, provides a natural exclusion restriction to account for selection in the intensive labor supply equation. The use of quasi-experimental tax variation to identify the participation equation is, to our knowledge, new to the labor supply literature, which has often noted that selection instruments meeting the required conditions are difficult to find. Moreover, the intensive labor supply equation provides a means of imputing a crucial variable in the extensive margin equation (earnings for individuals who do not work), allowing us to generate consistent estimates for that equation as well. This marks an improvement over the existing literature which has frequently imputed earnings without adjusting for selection, adopted selection instruments that are likely independently related to earnings, and used methods requiring strong distributional assumptions.

We show that the different policy tax measures are separately identified by tax schedule changes and nonlinearities in the tax code. We apply this approach to understand the tax responsiveness of older workers' labor supply. We find statistically significant and economically meaningful effects of taxes on labor force participation for older workers. These findings suggest scope for extending the working lives of older workers through the tax code. Since the prior labor supply and tax literatures rarely study the older segment of the population, this paper fills a large gap in these literatures and provides important estimates about the potential incentives in the tax code. We predict that tax reductions would cause this population to remain in the labor force longer and delay retirement. These findings are not driven by trends and the large estimated intensive margin elasticities are not due to systematic decisions to work part-time. Instead, we simply find that this population is especially responsive to intensive and extensive incentives to work. The results are generally consistent with extensive margin estimates generated from structural models in the literature as well as the responsiveness of older workers to the Earnings Test, which is not a pure tax and should therefore produce muted behavioral responses in comparison.

We find that accounting for selection is important. Interestingly, the intensive margin substitution effect is generally insensitive to whether and how we account for selection. However, the intensive margin equation helps generate variables necessary for the extensive margin equation and, in that case, we see that accounting for selection makes a substantial difference. The methods introduced in this paper should be useful more broadly in the tax and labor supply literatures for estimating intensive and extensive labor supply responses.

References

Auten, Gerald and Robert Carroll, "The Effect Of Income Taxes On Household Income," *The Review of Economics and Statistics*, November 1999, *81* (4), 681–693.

_ , _ , and Geoffrey Gee, "The 2001 and 2003 Tax Rebate Reductions: An Overview and Estimate of the Taxable Income Response," *National Tax Journal*, September 2008, *61* (3), 345–364.

Banks, James and Peter A. Diamond, "The base for direct taxation," in James A. Mirrlees et al., ed., *The Mirrlees Review*, Oxford: Oxford University Press, 2010, pp. 548–648.

Biggs, Andrew, "A Payroll Tax Cut Could Help Social Security," *Wall Street Journal*, April 2012. April 24.

Blau, David M and Donna B Gilleskie, "Health insurance and retirement of married couples," *Journal of Applied Econometrics*, 2006, *21* (7), 935–953.

Blau, Francine D. and Lawrence M. Kahn, "Changes in the Labor Supply Behavior of Married Women: 1980 – 2000," *Journal of Labor Economics*, 2007, *25*, 393–438.

Blomquist, N Sören and Urban Hansson-Brusewitz, "The effect of taxes on male and female labor supply in Sweden," *Journal of Human Resources*, 1990, pp. 317–357.

Blomquist, Sören and Håkan Selin, "Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates," *Journal of Public Economics*, 2010, *94* (11), 878–889.

_ and Whitney Newey, "Nonparametric estimation with nonlinear budget sets," *Econometrica*, 2002, *70* (6), 2455–2480.

Blundell, Richard, Alan Duncan, and Costas Meghir, "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 1998, *66* (4), 827–861.

_ and Thomas MaCurdy, "Labor supply: A review of alternative approaches," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3, Elsevier, 1999, chapter 27, pp. 1559–1695.

Burns, Sarah K and James P Ziliak, "Identifying the elasticity of taxable income," *The Economic Journal*, 2017, *127* (600), 297–329.

Cavanagh, Christopher and Robert P. Sherman, "Rank estimators for monotonic index models," *Journal of Econometrics*, 1998, *84* (2), 351 – 381.

Chernozhukov, Victor and Han Hong, "An MCMC approach to classical estimation," *Journal of Econometrics*, 2003, *115* (2), 293–346.

Coile, Courtney C and Phillip B Levine, "Recessions, reeling markets, and retiree well-being," Technical Report, National Bureau of Economic Research 2010.

Currie, Janet and Jonathan Gruber, "Health Insurance Eligibility, Utilization of Medical Care, and Child Health," *Quarterly Journal of Economics*, 1996, *111*, 431–466.

_ and _ , "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women," *The Journal of Political Economy*, 1996, *104*, 1263–1296.

Dickert, Stacy, Scott Houser, and John Karl Scholz, "The Earned Income Tax Credit and transfer programs: a study of labor market and program participation," in "Tax Policy and the Economy, Volume 9," MIT Press, 1995, pp. 1–50.

Eissa, Nada, "Taxation and labor supply of married women: the Tax Reform Act of 1986 as a natural experiment," Technical Report, National Bureau of Economic Research 1995.

_ **and Hilary Williamson Hoynes**, "Taxes and the labor market participation of married couples: the earned income tax credit," *Journal of Public Economics*, August 2004, *88* (9-10), 1931–1958.

_ **and Jeffrey B Liebman**, "Labor Supply Response to the Earned Income Tax Credit," *The Quarterly Journal of Economics*, 1996, *111* (2), 605–637.

_ **, Henrik Jacobsen Kleven, and Claus Thustrup Kreiner**, "Evaluation of four tax reforms in the United States: Labor supply and welfare effects for single mothers," *Journal of Public Economics*, April 2008, *92* (3-4), 795–816.

Feenberg, Daniel and Elisabeth Coutts, "An introduction to the TAXSIM model," *Journal of Policy Analysis and Management*, 1993, *12* (1), 189–194.

Feldstein, Martin, "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy*, 1995, *103* (3).

_ **and Jeffrey B Liebman**, "Social Security," *Handbook of Public Economics*, 2002, *4*, 2245–2324.

French, Eric and John Bailey Jones, "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 2011, *79* (3), 693–732.

_ **and John Jones**, "Public pensions and labor supply over the life cycle," *International Tax and Public Finance*, 2012, *19* (2), 268–287.

Friedberg, Leora, "The Labor Supply Effects of the Social Security Earnings Test," *The Review of Economics and Statistics*, February 2000, *82* (1), 48–63.

Gelber, Alexander, Adam Isen, and Jae Song, "The Role of Social Security Benefits in the Initial Increase of Older Womens Employment," *Women Working Longer: Increased Employment at Older Ages*, 2018, pp. 239–268.

Gelber, Alexander M. and Joshua W. Mitchell, "Taxes and Time Allocation: Evidence from Single Women and Men," *The Review of Economic Studies*, 2011.

Gelber, Alexander M, Damon Jones, and Daniel W Sacks, "Earnings adjustment frictions: Evidence from the Social Security earnings test," *American Economic Journal: Applied Economics*, 2020, *12* (1), 1–31.

_ **, _ , _ , and Jae Song**, "The employment effects of the Social Security earnings test," *Journal of Human Resources*, 2020.

Giertz, Seth, "The Elasticity of Taxable Income over the 1980s and 1990s," *National Tax Journal*, 2007, *60* (4).

Gruber, Jon and Emmanuel Saez, "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, April 2002, *84* (1), 1–32.

Gruber, Jonathan and Peter Orszag, "Does The Social Security Earnings Test Affect Labor Supply And Benefits Receipt?," *National Tax Journal*, December 2003, *56*, 755–773.

Gustman, Alan L and Thomas L Steinmeier, "Effects of Social Security Policies on Benefit Claiming, Retirement and Saving," *Journal of Public Economics*, 2015, *129*, 51–62.

Haider, Steven J. and David S. Loughran, "The Effect of the Social Security Earnings Test on Male Labor Supply: New Evidence from Survey and Administrative Data," *Journal of Human Resources*, 2008, *43* (1).

Hausman, Jerry, "Taxes and Labor Supply," in Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, New York: Elsevier, 1985, pp. 213–263.

Hausman, Jerry A, "The econometrics of nonlinear budget sets," *Econometrica: Journal of the Econometric Society*, 1985, pp. 1255–1282.

Heckman, James J, "Sample Selection Bias as a Specification Error," *Econometrica*, January 1979, *47* (1), 153–61.

_ , "Varieties of Selection Bias," *American Economic Review*, May 1990, *80* (2), 313–18.

Heim, Bradley T, "The incredible shrinking elasticities married female labor supply, 1978–2002," *Journal of Human Resources*, 2007, *42* (4), 881–918.

_ , "The effect of recent tax changes on taxable income: Evidence from a new panel of tax returns," *Journal of Policy Analysis and Management*, 2009, *28* (1), 147–163.

Hotz, V. Joseph and John Karl Scholz, "The Earned Income Tax Credit," in "Means-Tested Transfer Programs in the United States," University of Chicago Press, 2003, pp. 141–197.

Hotz, V Joseph, Charles H Mullin, and John Karl Scholz, "The Earned Income Tax Credit and labor market participation of families on welfare," Technical Report June 2010.

Huber, Martin and Giovanni Mellace, "Testing exclusion restrictions and additive separability in sample selection models," *Empirical Economics*, 2014, *47* (1), 75–92.

Karabarbounis, Marios, "A Roadmap for Efficiently Taxing Heterogeneous Agents," *American Economic Journal: Macroeconomics*, 2016.

Keane, Michael and Robert Moffitt, "A structural model of multiple welfare program participation and labor supply," *International Economic Review*, 1998, pp. 553–589.

Keane, Michael P, "Labor supply and taxes: A survey," *Journal of Economic Literature*, 2011, *49* (4), 961–1075.

Kimmel, Jean and Thomas J Kniesner, "New evidence on labor supply:: Employment versus hours elasticities by sex and marital status," *Journal of Monetary Economics*, 1998, *42* (2), 289–301.

Kremer, Michael, "Should taxes be independent of age?," *Unpublished paper Harvard University*, 2002.

Kumar, Anil and Che-Yuan Liang, "Declining female labor supply elasticities in the United States and implications for tax policy: Evidence from panel data," *National Tax Journal*, 2016, *69* (3), 481.

Laitner, John and Dan Silverman, "Consumption, retirement and social security: Evaluating the efficiency of reform that encourages longer careers," *Journal of Public Economics*, 2012, *96* (7–8), 615 – 634.

Meyer, Bruce D. and Dan T. Rosenbaum, "Welfare, The Earned Income Tax Credit, And The Labor Supply Of Single Mothers," *The Quarterly Journal of Economics*, August 2001, *116* (3), 1063–1114.

Newey, Whitney K., "Two-step series estimation of sample selection models," *The Econometrics Journal*, January 2009, *12* (Supplement s1), S217–S229.

Page, Timothy F and Karen Smith Conway, "The labor supply effects of taxing Social Security benefits," *Public Finance Review*, 2015, *43* (3), 291–323.

Powell, David, "The Distortionary Effects of the Health Insurance Tax Exclusion," *American Journal of Health Economics*, 2019, *5* (4), 428–464.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek, *Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]*, Minneapolis: University of Minnesota, 2015.

Saez, Emmanuel, Joel Slemrod, and Seth Giertz, "The elasticity of taxable income with respect to marginal tax rates: a critical review," *Journal of Economic Literature*, March 2012, *50* (1), 3–50.

Samwick, Andrew A, "New evidence on pensions, Social Security, and the timing of retirement," *Journal of Public Economics*, 1998, *70* (2), 207–236.

Schafgans, Marcia M.A. and Victoria Zinde-Walsh, "On Intercept Estimation In The Sample Selection Model," *Econometric Theory*, February 2002, *18* (01), 40–50.

Singleton, Perry, "The Effect of Taxes on Taxable Earnings: Evidence from the 2001 and Related U.S. Federal Taxes," *National Tax Journal*, June 2011, *64*, 323–352.

Song, Jae G. and Joyce Manchester, "New evidence on earnings and benefit claims following changes in the retirement earnings test in 2000," *Journal of Public Economics*, April 2007, *91* (3-4), 669–700.

Subbotin, Viktor, "Asymptotic and bootstrap properties of rank regressions," MPRA Paper 9030, University Library of Munich, Germany November 2007.

Triest, Robert K, "The effect of income taxation on labor supply in the United States," *Journal of Human Resources*, 1990, pp. 491–516.

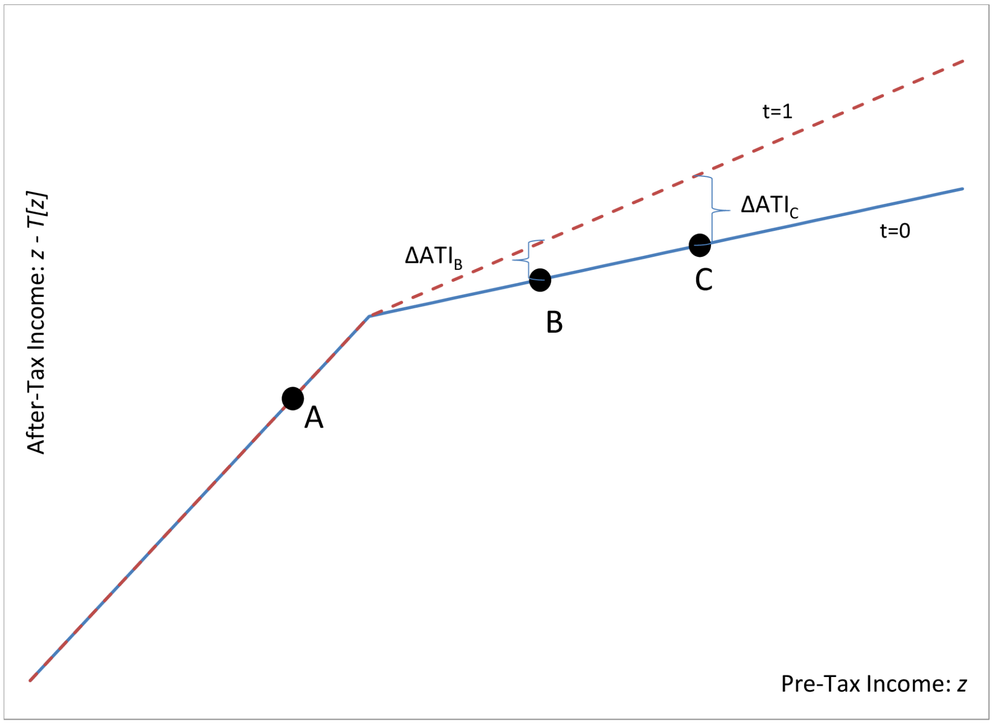
Weber, Caroline E, "Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences," *Journal of Public Economics*, 2014, *117*, 90–103.

Weinzierl, Matthew, "The surprising power of age-dependent taxes," *The Review of Economic Studies*, 2011, *78* (4), 1490–1518.

Ziliak, James P and Thomas J Kniesner, "The effect of income taxation on consumption and labor supply," *Journal of Labor Economics*, 2005, *23* (4), 769–796.

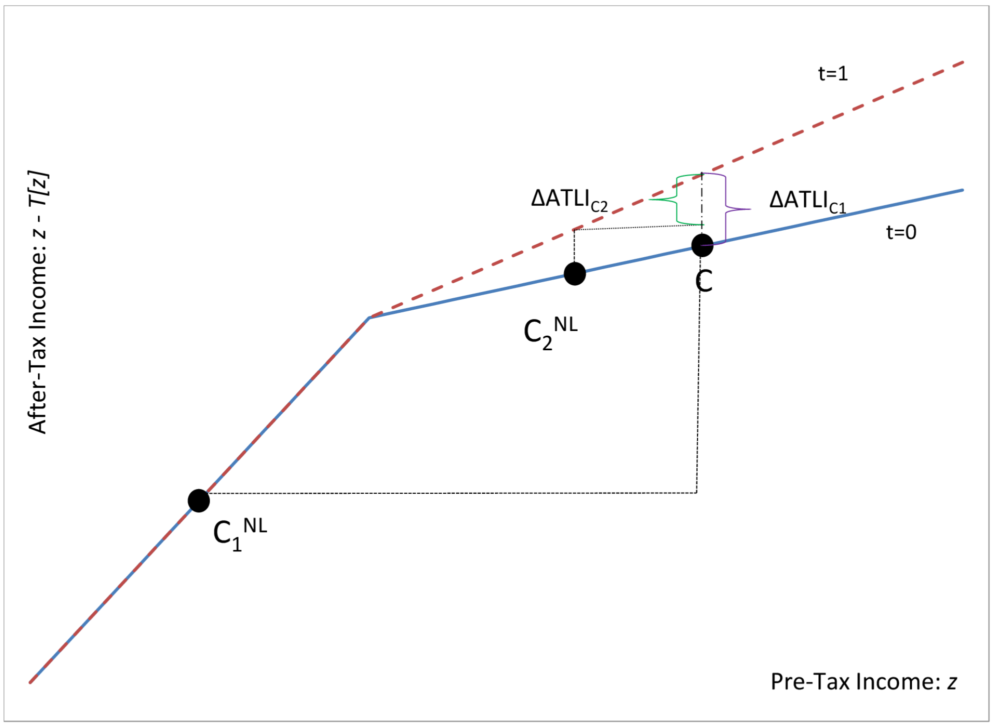
Figures

Figure 1: Marginal Tax Rates and After-Tax Income



Notes: This figure graphs after-tax income as a function of pre-tax income for a tax schedule with two brackets. The marginal tax rate for the top bracket decreases in period $t = 1$. Holding pre-tax income constant, persons A and B experience different changes in their marginal tax rates. Persons B and C experience the same change in marginal tax rates but different changes in after-tax income (ΔATI).

Figure 2: After-Tax Labor Income



Notes: This figure is the same as Figure 1 but with a focus on Person C. Person C with very little non-labor income (C_1^{NL}) experiences a large increase in the after-tax incentive to work ($\Delta ATLI$, after-tax labor income). Person C with more non-labor income (C_2^{NL}) experiences a smaller increase in the after-tax incentive to work. Both experience the same change in the marginal tax rate and after-tax income but different changes in after-tax labor income.

Tables

Table 1: Selection Equation

Dependent Variable:	I(Work)			
	Women		Men	
	(1)	(2)	(3)	(4)
Predicted ln(1 − Marginal Tax Rate)	0.297	0.088	-0.526*	-0.144
	[-0.211, 0.805]	[-0.144, 0.319]	[-1.095, 0.042]	[-0.423, 0.134]
Predicted ln(After-Tax Income)	0.066	0.255**	0.475	0.439***
	[-0.970, 1.101]	[0.008, 0.442]	[-0.719, 1.669]	[0.152, 0.726]
Predicted ln(After-Tax Non-Labor Income)	-0.020***	-0.030**	-0.026***	-0.030***
	[-0.032, -0.008]	[-0.054, -0.006]	[-0.039,-0.016]	[-0.051, -0.009]
Probit	X		X	
Monotone Rank		X		X
Observations	1,658,099	1,658,099	1,422,223	1,422,223

Notes: Significance Levels: *10%, **5%, ***1%. 95% confidence intervals in parentheses estimated using bootstrap. Coefficients are scaled so that sum of the square of the coefficients is equal to 1. Other variables included: year dummies; age group fixed effects; education fixed effects; race fixed effects; spousal age group fixed effects; spousal education fixed effects; and spousal race fixed effects. We also control for the predicted probability of working (based on covariates) in 1999 interacted with year dummies (except 1999) to account for potential differential labor force participation trends.

Table 2: Intensive Labor Supply Equation, 2SLS

Dependent Variable:	ln(Labor Income)					
	Women			**Men**		
Selection Adjustment:	None	Probit	Semi-Parametric	None	Probit	Semi-Parametric
	(1)	(2)	(3)	(4)	(5)	(6)
ln(1-MTR)	1.830***	1.526***	1.716***	3.269***	3.718***	3.035***
	[0.213, 2.695]	[0.806, 2.246]	[0.994, 2.439]	[1.674, 4.863]	[1.594, 5.842]	[1.514, 4.557]
ln(After-Tax Income)	0.958***	0.940	0.004	-0.473	-1.086	-0.407
	[0.260, 1.657]	[-0.853, 2.732]	[-7.852, 7.860]	[-9.805, 8.859]	[-7.047, 4.877]	[-10.257, 9.445]
Observations	451,765	451,765	451,765	565,162	565,162	565,162

Notes: Significance Levels: *10%, **5%, ***1%. 95% confidence intervals in parentheses estimated using bootstrap. Other variables included: year dummies; age group fixed effects; education fixed effects; race fixed effects; spousal age group fixed effects; spousal education fixed effects; and spousal race fixed effects. We also control for the predicted probability of working (based on covariates) in 1999 interacted with year dummies (except 1999) to account for potential differential labor force participation trends.

Table 3: Extensive Labor Supply Equation (Working)

Dependent Variable:	I(Work)					
	Women			**Men**		
Selection Adjustment:	None	Probit	Semi-Parametric	None	Probit	Semi-Parametric
	(1)	(2)	(3)	(4)	(5)	(6)
ln(After-Tax Income)	-0.099***	-0.197***	1.055***	0.213***	0.172	0.284***
	[-0.162, -0.037]	[-0.292, -0.103]	[0.914, 1.195]	[0.078, 0.348]	[-0.225, 0.569]	[0.202, 0.367]
ln(After-Tax Non-Labor Income)	-0.036	-0.031	-0.055***	-0.047***	-0.052***	-0.027***
	[-0.089, 0.018]	[-0.070, 0.007]	[-0.071, -0.041]	[-0.028, -0.066]	[-0.078, -0.027]	[-0.047, -0.008]
ln(Pre-Tax Labor Income)	0.147*	0.243***	-0.169***	-0.133***	-0.084	-0.187***
	[-0.080, 0.304]	[0.084, 0.403]	[-0.245, -0.091]	[-0.192, -0.074]	[-0.231, 0.063]	[-0.244, -0.130]
Observations	1,658,099	1,658,099	1,658,099	1,422,223	1,422,223	1,422,223

Notes: Significance Levels: *10%, **5%, ***1%. 95% confidence intervals in parentheses. Other variables included: year dummies; age group fixed effects; education fixed effects; race fixed effects; spousal age group fixed effects; spousal education fixed effects; and spousal race fixed effects. We also control for the predicted probability of working (based on covariates) in 1999 interacted with year dummies (except 1999) to account for potential differential labor force participation trends.

Table 4: Policy Simulations

Eliminating Employee Portion of Payroll Tax				
Outcome:	Working	Working	Retired	Retired
Effect of Age-Specific Payroll Tax	0.043***	0.014***	-0.038***	-0.014***
	[0.038, 0.049]	[0.010, 0.018]	[-0.056, -0.019]	[-0.018, -0.010]
Baseline Rate	0.272	0.397	0.719	0.591
Sample	Women	Men	Women	Men

Notes: Significance Levels: *10%, **5%, ***1%. Estimates use results from Tables 3 and C.2 to simulate effects of eliminating the employee portion of the payroll tax. We calculate after-tax labor income with and without the payroll tax, comparing the probabilities of not working and retiring. 95% Confidence Intervals in brackets.

Appendix: Online Publication Only

A Implementation Details

We explain the more technical details of the empirical strategy here. We describe each step in the order that it is estimated.

A.1 Step 1:

In the first step, we model the selection mechanism. When we report estimates that do not account for selection, this step is skipped. We must include all of the instruments used in the intensive labor supply equation. In the end, we estimate

$$P\left(\text{Work}_{it}=1\right) \quad = \quad F\Big(\phi_t \ + \ X_{it}'\gamma \ + \ \beta_1\widehat{\text{MTR}_{it}} \ + \ \beta_2\widehat{\text{ATI}_{it}} \ + \ \beta_3\widehat{\text{ATNI}_{it}}), \eta\Big) \quad (5)$$

The predictions provided by equation (5) are used as selection adjustments for the intensive equation. We do this in two different ways. First, we assume that $F(\cdot)=\Phi(\cdot)$ and estimate equation (5) using a probit regression. This method is frequently used in the literature. However, instead of including an inverse Mills ratio (Heckman (1979)), we condition on a flexible function of the estimated index.

Second, we use a monotone rank estimator introduced in Cavanagh and Sherman (1998). This estimator does not estimate $F(\cdot)$ but provides $\sqrt{n}$-consistent estimates up to scale of the coefficients in the argument of the function. We then predict the index function, which we denote as $W_{it}'\hat{\zeta}$. The selection correction term is a function of this index and we follow the method of Newey (2009) by approximating this term with a spline using $W_{it}'\hat{\zeta}$.[33] The advantage of this approach is that the maximum rank estimator requires no distributional assumptions to obtain consistent estimates.

To implement the monotone rank estimator, we generate initial values through a probit regression and maximize the objective function using an MCMC optimization algorithm (see Chernozhukov and Hong (2003)). Confidence intervals are generated using a bootstrap.[34]

[33] Newey (2009) recommends the use of a spline over a power series.
[34] Subbotin (2007) discusses properties of bootstrapping rank regression estimates.

A.2 Step 2:

The second step estimates the intensive labor supply equation. Because of selection, we estimate the following:

$$\ln L_{it} = \alpha_t + X_{it}'\delta + \beta^I \ln(1-\tau_{it}) + \theta^I \ln\left(y_{it} - T_t(y_{it})\right) + \epsilon_{it} \tag{6}$$

where $\epsilon_{it} = \lambda(W_{it}'\zeta) + \mu_{it}$. We use $\widehat{\text{MTR}_{it}}$ and $\widehat{\text{ATI}_{it}}$ as instruments. We bootstrap Steps 1 and 2 jointly to account for the inclusion of an estimated term in equation (6).

We should highlight that 2SLS includes the selection adjustment terms in the first stage as well. This has practical importance in our strategy. Notice that for individuals not working, we do not observe their marginal net-of-tax rate if they had actually worked. We predict this variable from the first-stage regression in the same way that we will predict labor earnings.

We use 2SLS to obtain consistent estimates. Once we have consistent estimates for equation (6), we can predict $\ln L_{it}$ for our entire sample: $\widehat{\ln L_{it}}$. This includes people who did not work. When using the Newey (2009) method, the constant term is not separately identified from the selection correction term. A method to estimate the constant term was introduced in Heckman (1990). Schafgans and Zinde-Walsh (2002) discuss consistency and asymptotically normality of this estimator. However, we are interested in imputing the variable $\ln\left[L_{it} + y_{it}^o - T_t(L_{it} + y_{it}^o)\right]$ (as well as $\ln L_{it}$). It is well-known that the expected value of a variable is *not* the exponential of the expected value of the log of that variable (i.e., $E[y] \neq \exp(E[\ln y])$). Instead, we use the Heckman (1990) approach to estimate the distribution of after-tax income (if working).

$$\text{Let} \quad \mathcal{U} = \{\widehat{\epsilon}_{it} \quad | \quad \mathbf{W}_{it}'\widehat{\zeta} > \gamma_N\},$$

where $\mathbf{W}_{it}'\widehat{\zeta}$ was defined in equation (4). γ_N is the bandwidth or smoothing parameter: $\gamma_N \to \infty$ as $N \to \infty$.[35] On average, these residuals are the estimate of the constant using Heckman (1990). We use the full set of residuals in this set and integrate over the distribution. Our estimate of $E\left[\ln\left[L_{it} + y_{it}^o - T_t(L_{it} + y_{it}^o)\right]\right]$ is

$$\int_{\epsilon \in \mathcal{U}} \ln\left[\exp(\widehat{\ln L_{it}} + \epsilon) + y_{it}^o - T_t(\exp(\widehat{\ln L_{it}} + \epsilon) + y_{it}^o\right] d\epsilon. \tag{7}$$

[35]In practice, we set γ_N at the 99^{th} percentile of the distribution of $\mathbf{W}_{it}'\widehat{\zeta}$.

In practice, we estimate our after-tax income (for workers) in the following manner. We first create $\mathcal{U}$. Then, for each person, we randomly select an element of $\mathcal{U}$. The exponential of the predicted value of $\ln L_{it}$ plus this term provides one possible value of L_{it} for that individual. Next, using NBER's Taxsim program, we calculate the individual's tax liability given their non-labor income and this predicted value of labor earnings. Given this information, we can create one possible value of the log of after-tax income given that the individual works. We perform this imputation ten times in total and take the average.[36] This is our after-tax income variable for the extensive margin equation. We also predict the $\widehat{\ln L_{it}}$ term in the same manner.[37]

A.3 Step 3:

Next, we estimate

$$P\left(\text{Work}_{it}=1\right)=F\left(\phi_t+X_{it}'\gamma+\beta^E\ln\left[L_{it}+y_{it}^o\widehat{-T_t}(L_{it}+y_{it}^o)\right]+\theta^E\ln\left(y_{it}^o-T_t(y_{it}^o)\right)+\rho^E\widehat{\ln L_{it}}+\nu_{it}\right)$$

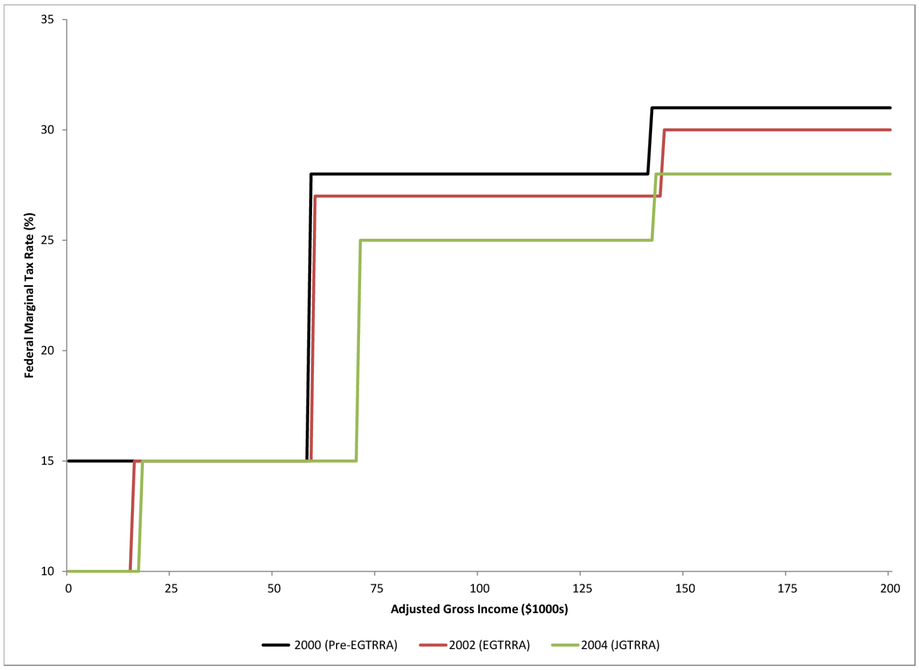
We estimate this equation using 2SLS. IV-Probit estimates are similar.

[36]We stop at ten instead of integrating over the entire set for computational reasons. This may lead to measurement error in our measure, but should not affect the consistency of the estimates (since the IV method will address measurement error concerns). Averaging in additional imputations would likely decrease noise.

[37]For this variable, there is no need to integrate over the set of residuals in $\mathcal{U}$, but for the sake of consistency, we construct the two measures using the same method.

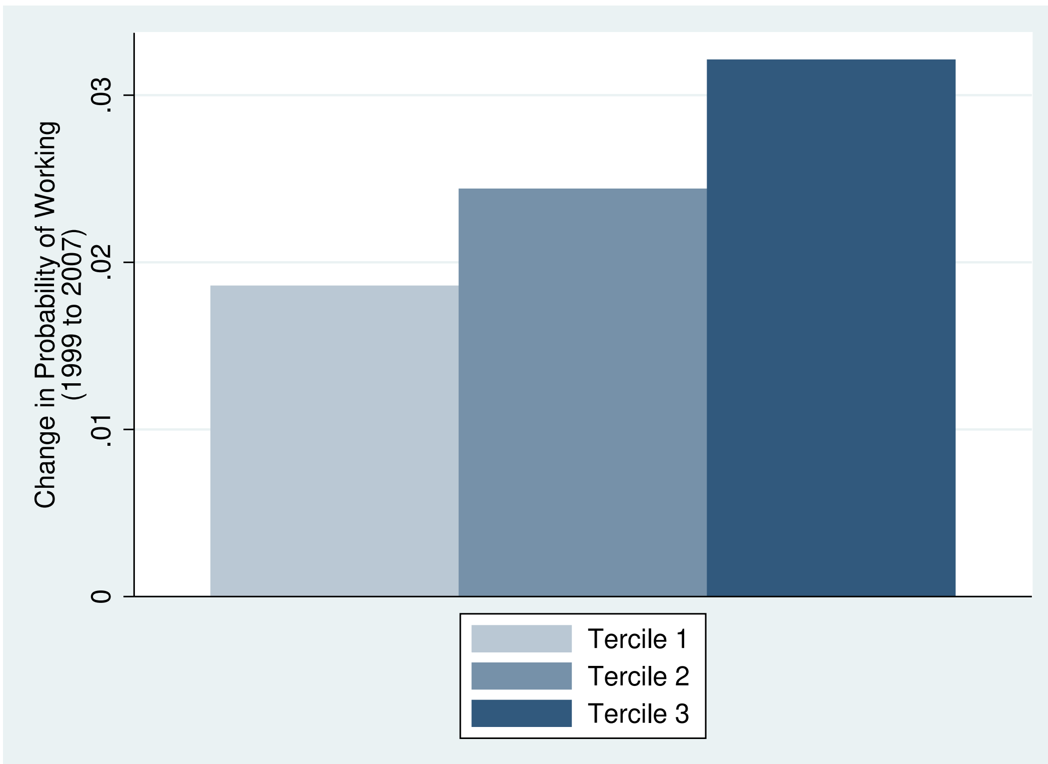
B Additional Figures

Figure A1: Federal Marginal Tax Rates, Pre- and Post-Tax Reforms



Notes: Marginal tax rates are for married couples filing jointly. Income is in constant 2013 dollars. Source: "U.S. Federal Individual Income Tax Rates History, 1862-2013 (Nominal and Inflation Adjusted Brackets)," Tax Foundation.

Figure A2: Change in the Probability of Working by Percentage Change in Predicted After-Tax Labor Income: 1999 to 2007



Notes: The y-axis is the change in the probability of working between 1999 and 2007 for each bin (i.e., the fraction working in 2007 minus the fraction working in 1999). The bins are defined by the change in the predicted log of after-tax income minus the predicted log of after-tax non-labor income (where these predictions are the instruments discussed in the text and entirely dependent on covariates). This is a measure of changes in tax-based incentives to work due to legislative policy changes only. We divide the sample into terciles based on the magnitude of this change: Tercile 1 experiences the smallest increase and Tercile 3 experiences the largest increase. This approach mimics our empirical strategy. The y-axis is the change in the probability of working.

C Additional Tables

Table C.1: Descriptive Statistics

	Women	Men
Demographics		
Age	67.7	67.5
Less than HS	24.4%	24.0%
HS Grad	37.4%	29.7%
Some College	22.3%	21.5%
College Grad	15.8%	24.8%
Labor Outcomes		
Personal Labor Earnings	$5,960.94	$16,667.96
Works	27.2%	39.7%
Retired	71.9%	59.1%
Total Household Income	$43,479.99	$56,371.47
Tax Variables		
Marginal Tax Rate	19.3	22.5
N	1,658,099	1,422,223

Notes: All dollar values expressed in nominal dollars.

Table C.2: Extensive Labor Supply Equation (Retirement)

Dependent Variable:	I(Work)					
	Women			**Men**		
Selection Adjustment:	None	Probit	Semi-Parametric	None	Probit	Semi-Parametric
	(1)	(2)	(3)	(4)	(5)	(6)
ln(After-Tax Income)	0.143***	0.281***	-0.914***	-0.054	0.002	-0.289***
	[0.081, 0.205]	[0.180, 0.381]	[-1.373, -0.454]	[-0.453, 0.344]	[-3.311, 3.315]	[-0.379, -0.199]
ln(After-Tax Non-Labor Income)	0.024	0.021	0.053***	0.051***	0.049	0.029***
	[-0.028, 0.075]	[-0.018, 0.060]	[0.033, 0.075]	[0.033, 0.069]	[-0.082, 0.179]	[0.009, 0.049]
ln(Pre-Tax Labor Income)	-0.169**	-0.310***	0.080	0.057	-0.002	0.165***
	[-0.320, -0.019]	[-0.477, -0.144	[-0.049, 0.209]	[-0.102, 0.216]	[-1.228,1.223]	[0.106, 0.224]
Observations	1,658,099	1,658,099	1,658,099	1,422,223	1,422,223	1,422,223

Notes: Significance Levels: *10%, **5%, ***1%. 95% confidence intervals in parentheses. Other variables included: year dummies; age group fixed effects; education fixed effects; race fixed effects; spousal age group fixed effects; spousal education fixed effects; and spousal race fixed effects. We also control for the predicted probability of working (based on covariates) in 1999 interacted with year dummies to account for potential differential labor force participation trends.

Table C.3: Robustness Tests

	Main Results		Part-Time Adjustment		Account for Earnings Test Changes	
	(1) Women	(2) Men	(3) Women	(4) Men	(5) Women	(6) Men
	1. Intensive Margin Equation - Outcome is ln(Labor Income)					
ln(1-MTR)	1.716***	3.035***	1.602**	3.437***	1.642***	3.175**
	[0.994, 2.439]	[1.514, 4.557]	[0.265, 2.939]	[2.011, 4.863]	[0.200, 3.084]	[0.161, 6.189]
	2. Extensive Margin Equation - Outcome is I(Work)					
ln(After-Tax Labor Income)	1.055***	0.284***	0.692***	0.454*	0.922***	0.513***
	[0.914, 1.195]	[0.202, 0.367]	[0.363, 1.021]	[-0.022, 0.930]	[0.581, 1.263]	[0.428, 0.598]
	3. Simulations - Outcome is P(Working)					
Payroll Tax Simulation	0.043***	0.014**	0.032***	0.019*	0.039***	0.024***
	[0.038, 0.049]	[0.010, 0.018]	[0.017, 0.047]	[-0.001, 0.039]	[0.032, 0.046]	[0.021, 0.029]

Notes: Significance Levels: *10%, **5%, ***1%. 95% Confidence intervals in brackets estimated in same manner as previous results. The first row presents estimates from the intensive margin equation. The second row includes estimates from the extensive margin equation. The third row presents estimates from the payroll tax simulation. All estimates refer to semi-parametric adjustment results.
Columns (1) and (2) repeat the main results presented in earlier tables.
Columns (3) and (4) reassign the labor earnings of part-time workers to earnings if they made the same hourly wage but worked 35 hours per week.
Columns (5) and (6) include predicted policy variables related to the Social Security Earnings Test (generated in the same manner as the instruments).

Table C.4: Timing of Effects

	Women	Men
	Intensive Margin	
ln(1-MTR)	2.714**	3.742***
	[0.509, 4.919	[0.779, 6.705]
ln(1-MTR) in $t+1$	-0.136	-0.243
	[-0.578, 0.304]	[-0.715, 0.229]
	Extensive Margin	
ln(After-Tax Labor Income)	0.751***	0.504***
	[0.425, 1.078]	[0.265, 0.743]
ln(After-Tax Labor Income) in $t+1$	-0.187	-0.326*
	[-0.601, 0.227]	[-0.683, 0.030]

Notes: Significance Levels: *10%, **5%, ***1%. 95% Confidence intervals in brackets estimated in same manner as previous results. All estimates refer to semi-parametric adjustment results with same covariates as before. The $t+1$ values refer to the predicted values (i.e., instrument). Results are similar if we replace the year t variables with the predicted values as well. The first two rows present estimates from the intensive margin equation. The bottom two rows include estimates from the extensive margin equation.