# Stanford | Institute for Economic Policy Research (SIEPR)

# USING NON-LINEAR BUDGET SETS TO ESTIMATE EXTENSIVE MARGIN RESPONSES: METHOD AND EVIDENCE FROM THE SOCIAL SECURITY EARNINGS TEST

### Alexander M. Gelber
UC Berkeley Goldman School of Public Policy and NBER

### Damon Jones
University of Chicago,
Harris School of Public Policy and NBER

### Daniel W. Sacks
Indiana University, Kelley School of Business

### Jae Song
Social Security Administration

# Using Non-Linear Budget Sets to Estimate Extensive Margin Responses: Method and Evidence from the Social Security Earnings Test

Alexander M. Gelber, Damon Jones, Daniel W. Sacks, and Jae Song[*]

June 2018

## Abstract

We develop a method for estimating the effect of a kinked or notched budget set on workers' employment decisions, and we use it to estimate the impact of the Social Security Old-Age and Survivors Insurance (OASI) Annual Earnings Test (AET). The AET reduces OASI claimants' current OASI benefits in proportion to their earnings in excess of an exempt amount. Using a Regression Kink Design and Social Security Administration data, we document that the discontinuous change in the benefit reduction rate at the exempt amount causes a corresponding change in the slope of the employment rate. We develop conditions in a general setting under which we can use such patterns to estimate the elasticity of the employment rate with respect to the effective average net-of-tax rate. Our resulting point estimate for the AET indicates an elasticity of at least 0.49, suggesting that the AET reduces employment by more than one percentage point in the group we study.

# 1    Introduction

Non-linear price schedules are common across a wide variety of economic applications, including labor supply (*e.g.* Hausman, 1981), electricity demand (*e.g.* Ito, 2014), and retirement savings (*e.g.* Bernheim *et al.*, 2015). Taxes and government transfers in particular often create kink points, which are discontinuous changes in the marginal incentive to earn or work more, or notches, which are discrete changes in the level of net income as a function of earnings. For example, many means-tested government transfer programs reduce or eliminate transfer benefits as income rises above threshold levels, and the individual income tax in many countries generates a piecewise linear budget set with kinks at each point at which the marginal tax rate (MTR) jumps. Previous literature has developed methods for explicitly using kink points or notches in the effective tax schedule to estimate responses at the intensive margin, the choice of earnings or hours worked conditional on working (*e.g.* Burtless and Hausman, 1978; Hausman, 1981; Saez, 2010; Chetty *et al.*, 2011; Kleven and Waseem, 2013; see the review in Kleven, 2016). However, tax and transfer policies affect not only intensive margin responses, but also responses at the extensive margin, the choice of whether or not to be employed (Cogan, 1981; Saez, 2002; Eissa *et al.*, 2008). More generally, the effects of incentives on earnings and the employment rate are both key questions across several fields in economics.

In this paper, we develop a method for estimating the effects of budget sets with kinks or notches on workers' extensive margin decisions. Our method relies on clear patterns in data: under the conditions we specify, a kink in a budget set leads to a kink in the employment rate, and a notch leads to a discontinuous change in the level of the employment rate. Specifically, suppose individuals face fixed costs of employment such as commuting costs (Cogan, 1981), and consider the probability of employment as a function of the desired earnings individuals would choose if employed. If the individual faces a linear budget set, this function should presumably be smooth (continuously differentiable), as illustrated in the dashed line in Figure 1A. Now consider a budget set with a convex kink at which the effective marginal tax rate discontinuously rises. Figure 2A shows that in this case the slope of the average net-of-tax rate (ANTR) decreases discontinuously at the kink point. The ANTR is defined as the fraction of an individual's income that the individual keeps net of taxes and benefit reduction when earning a positive amount rather than earning zero, reflecting the individual's incentive to be employed.[1] In this environment, we show that there should robustly be a discontinuous decrease at the kink point in the slope of the probability of employment among individuals who are constrained from making small, intensive margin adjustments. This is illustrated in the solid line in Figure 1A. Intuitively, these intensive margin frictions prevent individuals from locating at the kink even if they might otherwise like to do so under the kinked budget constraint, driving them to non-employment instead and fueling a discontinuous decrease in the slope of employment.[2]

Thus, observing such a change in slope represents a novel type of evidence that a kinked budget set has an extensive margin effect. Our assumptions are therefore testable, as this change in slope should arise if

---

[1] Formally, for a tax schedule $T(z)$, we define the ANTR as $ANTR \equiv 1 - [T(z) - T(0)]/z$, where $z$ is pre-tax earnings.

[2] Our method also applies at non-convex kinks: the slope of the employment rate discontinuously increases for those constrained from adjusting away from non-convex kinks.

individuals face fixed costs of employment as well as intensive margin frictions. This prediction does not rely on parametric assumptions on the utility function, nor on assuming a representative agent; indeed the distribution of employment responses in the population, as well as other distributional assumptions, are left unrestricted. Using a Regression Kink Design (RKD), we can estimate the change in slope of the employment rate (Nielsen *et al.*, 2010; Card *et al.*, 2015), which reflects average extensive margin effects for those near the budget set kink.

The elasticity of participation with respect to the ANTR can be expressed as a function of the change in the slope of the employment rate at the kink, as well as other observable moments in the data. All else equal, the larger the change in the slope of employment at the kink point, the larger the elasticity of employment with respect to the ANTR. If some individuals do not face such frictions and some do, then we estimate a lower bound on the average elasticity among both of these two groups without making parametric or distributional assumptions. We suggest that this method for estimating (a lower bound on) extensive margin elasticities is valid without such assumptions not only in the case of employment (*i.e.* a corner solution in leisure demand), but also in the case of demand for other commodities in the presence of fixed costs of extensive margin decisions as well as intensive margin frictions.

Parallel logic implies that in the presence of such a friction, a notch in the budget set leads to a discontinuous change in the level of employment as a function of the desired earnings individuals would choose if employed. This is illustrated by the solid line in Figure 1B. Again, this prediction is non-parametrically valid within the model we specify and does not rely on assuming a representative agent. The magnitude of the discontinuity can be estimated using a Regression Discontinuity Design (RDD). The elasticity of participation can again be expressed as a function of the magnitude of the discontinuity as well as other estimable parameters, and we suggest this method is also generalizable to demand for other commodities.

One collection of prior literature explicitly uses kinked or notched budget sets to estimate intensive margin responses, while abstracting from the extensive margin that we study (*e.g.* Burtless and Hausman, 1978; Hausman, 1981; Saez, 2010; Chetty *et al.*, 2011; Kleven and Waseem, 2013; Kleven, 2016). Previous literature on the effect of the marginal incentives to be employed on extensive margin employment decisions, including the literature based on non-linear budget sets, has generally imposed parametric restrictions to generate labor supply estimates (Manski, 2014), or has not explicitly related causal responses to the budget constraint (Blundell and MaCurdy, 1999). For example, prior literature contains estimates of extensive margin responses to employment incentives under the distributional assumptions in Tobit or discrete choice models (McDonald and Moffitt, 1980; Moffitt, 1983; Hoynes, 1996; Keane and Moffitt, 1998; Blundell and Shephard, 2012; see the surveys in Keane, 2011, or Saez *et al.*, 2012). Similarly, Alpert and Powell (2014) develop a method that can be used when capital and labor income are both observed under certain restrictions on functional forms. Other papers have performed program evaluations of such effects using randomized experiments (*e.g.* Bitler *et al.*, 2006) or other techniques such as difference-in-differences (*e.g.* Eissa and Liebman, 1996), but with empirical models that do not explicitly take into account how the data relate

to the budget constraint (Blundell and MaCurdy, 1999). For example, such papers do not model intensive margin responses to kinks or notches jointly with extensive margin responses. We specify the conditions under which our method can be used, and our model allows us to be explicit about the about the population and parameter that our elasticity refers to.

Two recent papers have investigated the effects of a notch on extensive margin decisions in very different contexts than ours. Kleven *et al.* (2014) explore the effect of a kink on migration using differences-in-differences methods, and Kopczuk and Munroe (2015) show that a housing tax notch creates "unraveling" of productive matches within a bargaining model for the particular housing context. Our work develops methods applicable in the context of a general budget set kink or notch, with an application to the employment decision.[3]

None of these existing methods applies the RKD or RDD design in the novel context of explicitly using kinks or notches in budget sets to estimate extensive margin labor supply or earnings responses.[4] Thus, a key feature of our method is that its central predictions can be transparently confirmed in a simple graph of the data showing whether the slope or level of the employment rate changes discontinuously at the kink. Our method explicitly focuses on discontinuities in the data to drive the results. As our empirical method is valid under a different set of assumptions than alternative methods for estimating extensive margin responses, we see our method as complementary to other approaches.

Substantively, we apply our method for kinks to show that extensive margin responses are important in the context of the Social Security Old-Age and Survivors Insurance (OASI) Annual Earnings Test (AET). Although we focus on the AET, we also explain how our method can be used in other contexts, and we show empirically that in a broader sample of ages—not just older Americans subject to the AET—conditions hold that allow our method to be used. The AET reduces OASI claimants' current OASI benefits as a proportion of earnings in excess of an exempt amount. For example, for OASI claimants aged 62 to the year of attaining the Normal Retirement Age (NRA) of 66 in 2018, current OASI benefits are reduced by one dollar for every additional two dollars earned above $17,040. This could reduce OASI claimants' incentives for additional work. Reductions in current benefits due to the AET sometimes lead to increases in later benefits; nonetheless, as we discuss, several factors may explain why individuals' earnings still respond to the AET.

The AET is a natural context for studying the extensive margin effects of non-linear budget sets. The AET is a significant factor affecting the income of many older Americans. The AET's large benefit reduction

---

[3] A small literature has developed non-parametric methods for estimating a different parameter, or in a substantially different context. Blomquist *et al.* (2015) non-parametrically estimate the effect of a non-linear budget set on expected labor income, without separating the intensive and extensive margin responses that are necessary for welfare evaluation. In a welfare program creating kinks and notches, as well as under-reporting of income, welfare stigma, and hassle costs of welfare work requirements, Kline and Tartari (2016) develop non-parametric bounds on labor supply responses in the context of the particular features of their setting, whereas our paper focuses on behavior around a general kink or notch.

[4] In Austria, Manoli and Weber (2016) use an RKD to study retirement decisions relying on a Saez (2010)-type model of intensive margin bunching in retirement at the retirement age, while Manoli and Weber (forthcoming) study bunching in retirement at notches created by severance payments using intensive margin methods. Similarly, Card *et al.* (2015) and Card *et al.* (2016) use an RKD to estimate the effect of unemployment benefits on unemployment, but in their context the unemployment benefit does not depend on contemporaneous earnings and therefore does not create the kinked budget constraint studied in this paper or in other literature on kinked budget constraints such as Hausman (1981) or Saez (2010).

rate (BRR) can lead to a large budget set kink, suggesting that the AET could have important impacts on employment that policy-makers wish to understand. Particularly as the economy-wide employment-to-population ratio has fallen by over four percentage points since its peak in 2000 (Aaronson *et al.*, 2014), many policy makers are interested in ways to affect employment rates (Abraham and Kearney, 2017), including potential changes to the AET (Tergesen, 2016). Indeed, the AET is a leading candidate in helping to explain the upward spike in the hazard of retirement at age 62 (Gruber and Wise 1999). The importance of the AET is now increasing as the NRA gradually rises from 65 for those born in 1937 and earlier, to 67 for those born in 1960 and later, exposing more OASI claimants to the AET. Even in 2003—the last year for which we have benefits data—when the NRA was only 65.5, 538,000 beneficiaries lost fully \$4.30 billion of current total benefits to the AET.[5]

To estimate the employment effects of the AET, we use Social Security Administration (SSA) data on a 25 percent random sample of the U.S. population in birth cohorts 1918 to 1923 over the years 1978 to 1987.[6] These data cover 11,397,336 observations of calendar year earnings and other information on 1,424,667 individuals in our full sample. Using an RKD, we uncover a novel fact: as a function of past earnings—which captures desired earnings as we verify—the slope of the employment rate among 63 to 64 year-olds discontinuously decreases around the budget set kink created by the AET at the exempt amount.[7] We conduct several placebo and other robustness checks to verify that we have found a true effect on earnings, as opposed to an underlying nonlinearity in the employment rate.

A baseline specification shows that the point estimate of the elasticity of employment with respect to the ANTR is at least 0.49.[8] Simulations relying on these estimates show that eliminating the AET would cause an increase in the annual employment rate at ages 63 to 64 of 1.4 percentage points in the group we study near the kink, reflecting a 2.5 percent increase in employment. Using our extensive margin elasticity from this paper, as well as the AET intensive margin elasticities from Gelber *et al.* (2013), we calculate that earnings in the group we study decrease by 9.8 percent due to the AET. Of the total earnings reductions due to the AET, 27.6 percent are associated with extensive margin exit decisions. Our model and results therefore highlight both the potential importance of the AET to older workers' earnings and employment decisions, as well as the importance of the extensive margin responses in addition to the intensive margin responses around the kink that are the focus of much recent literature following Saez (2010).

Our new estimates of the effect of the AET on employment relate to long lines of previous research on the effect of Social Security and other types of pensions on retirement or employment decisions (see

---

[5] If earnings and/or employment are reduced by the AET, then this calculated reduction in benefits reflects a lower bound on the reduction in current benefits we would hypothetically observe if earnings and employment were inert in response to the AET.

[6] For our calculation of the \$4.3 billion aggregate benefit reduction alone, we use the Benefits and Earnings Public Use File one percent population sample, because this file has information on all birth cohorts until the more recent year 2003 that illustrates the aggregate benefit reduction better after the NRA had risen above 65. We use our 25 percent random sample of administrative data from 1978 to 1987 everywhere else because these data allow much more statistical power to perform the other estimates in the paper.

[7] We focus on 63 to 64 year-olds for several reasons: this group is relevant to policy, as the AET currently applies to this group but not to those over NRA; this is the youngest group subject to the AET for which measured employment would be affected by the AET, as we explain in detail later; and thus this is the group for which age 60 earnings represents the best proxy for desired earnings.

[8] For consistency with the previous literature on kink points that has focused on the effect of taxation, we sometimes use "tax" as shorthand for "tax-and-transfer," as in our use of the term "ANTR" to apply to the AET, while recognizing that the AET reduces Social Security benefits and is not administered through the tax system.

the literature review in Blundell *et al.*, 2016). With respect to the AET in particular, much of the previous literature has focused on the policy's intensive margin effect (*e.g.* Burtless and Moffitt, 1985; Friedberg, 1998; Friedberg, 2000; Song and Manchester, 2007; Gelber *et al.*, 2013), typically finding moderate substitution elasticities at the intensive margin. Given the clear but moderate responses at the intensive margin, it is arguably surprising that nearly all prior studies find little evidence of extensive margin responses. Indeed, much of the earlier empirical literature on the AET concludes that the policy has little meaningful effect on the labor supply of older men.[9] More recent work has examined the effect of the AET on employment decisions using a difference-in-differences framework. Several of these papers find little evidence for an effect on the employment rate (Gruber and Orszag, 2003; Song, 2004; Song and Manchester, 2007; Haider and Loughran, 2008), while Friedberg and Webb (2009) find a significant effect in some specifications in the Current Population Survey.[10]

We interpret our evidence as showing a clear and robust impact of the AET on the annual employment rate, in contrast to the bulk of previous literature. The combination of a large administrative dataset with individual-level microdata (also used in various forms in Song, 2004, Song and Manchester, 2007, and Haider and Loughran, 2008) and our novel identification strategy based on discontinuities leads to precise estimates of sizeable elasticities. Our paper is the first that has estimated a significant impact of the AET at the extensive margin through explicitly modeling individuals' budget sets. However, we estimate the effect of the AET on those locating near the exempt amount in the policy-relevant 63 to 64 year-old group to which the AET currently applies—a younger group than those studied in the difference-in-differences literature cited above. Thus, our results are not directly comparable to this previous literature, but provide relatively novel estimates for a younger group.

The paper proceeds as follows. Section 2 presents our model. Section 3 describes the policy environment. Section 4 explains our empirical strategy. Section 5 discusses the data. Section 6 presents empirical evidence on the response to the AET and performs counterfactual simulations. Section 7 concludes.

## 2   Model

We model how a kink in the budget set impacts an individual's decision of whether or not to have a positive amount of earnings, which we refer to as the "employment" decision. In particular, we explore under what conditions the introduction of a kink in the effective tax schedule should cause a corresponding kink in the employment rate.[11] We also explore the case of a budget set notch. Throughout, we make use of a potential outcomes framework (Rubin, 1974).

---

[9] See Viscusi (1979), Burtless and Moffit (1985), Gustman and Steinmeier (1985, 1991), Vroman (1985), Honig and Reimers (1989), and Leonesio (1990).

[10] Examining the effects of earnings tests in other countries using difference-in-differences designs, Baker and Benjamin (1999) find a significant effect of the Canadian earnings test on weeks worked per year but no significant effect on employment at some point during the year, and Disney and Smith (2002) find inconclusive evidence on the impact of the U.K. earnings test on the employment rate. French (2005) uses method of simulated moments in a lifecycle model to estimate the effects of health, wealth (including all forms of pension wealth, not just incentives created by the the AET), and wages on labor supply and retirement; simulations based on these estimates imply that eliminating the AET would cause individuals to retire later on average.

[11] To clarify, we use "kink" in two senses—both to describe a discontinuity in the effective marginal tax rate in the budget set, and to describe a discontinuity in the slope of an outcome variable (in our case the employment rate).

We index two potential states of the world by $j \in \{0,1\}$. In state 0 individuals face a linear tax $\tau_0$, *i.e.* $T_0(z) = \tau_0 z$, where $T_j(\cdot)$ denotes their tax liability in state $j$ and $z$ denotes their pre-tax earnings. Alternatively, in the case of a kink, in state 1 the tax schedule exhibits a change in slope at $z^*$:

$$T_1(z) = \begin{cases} \tau_0 z, & \text{if } z < z^* \\ \tau_0 z^* + \tau_1(z - z^*) & \text{if } z \geq z^* \end{cases} \tag{1}$$

In the case of a notch, the tax schedule in state 0 is the same, but now, in state 1, the tax schedule exhibits a "downward" notch at $z^*$, where "downward" indicates an increase in the tax liability or a decrease in transfers. As in Kleven and Waseem (2013), the notched schedule takes the form: $T_1(z) = \tau_0 \cdot z + [dT + d\tau \cdot z] \cdot \mathbf{1}\{z > z^*\}$, where $dT$ is a "pure notch," *i.e.* an intercept shift in the budget set, and $d\tau$ captures a "proportional notch," *i.e.* a change in level and slope. In our application we interpret $z$ as earnings and $T$ as a tax schedule, but an analogous model should apply to consumption of any good (or "bad") under a price schedule with a convex kink or downward notch.

Following much previous literature on employment responses to kinks or notches (*e.g.* Hausman, 1981; Saez, 2010; Kleven and Waseem, 2013), we begin with a model in which individuals' utility depends on consumption and earnings, $u(c, z; n)$, where the partial effect of an increase in $z$ on utility is negative as it requires effort to increase earnings, and the partial effect of an increase in $c$ on utility is positive. We assume $u(\cdot)$ is a function of class $C^2$. We do not make assumptions constraining the nature of income effects, *e.g.* we do not assume that utility is quasi-linear, that leisure is a normal good, or other such assumptions. As in Saez (2010) and much subsequent literature, we model the determination of earnings, rather than hours worked, as earnings (but not hours worked) are observed in many administrative datasets. However, our method can easily be adapted to apply to the determination of hours worked. Our index of "ability" is $n$; the marginal rate of substitution of $c$ for $z$ is decreasing in $n$ at all levels of $c$ and $z$.[12] To simplify the model for the time being, we begin by assuming that individuals maximize utility subject to a static budget constraint:

$$c_{nj} = z_{nj} - T_j(z_{nj}) \tag{2}$$

where $z_{nj}$ is realized earnings for an individual with ability $n$ in state $j$. At an interior solution, the first-order condition, $(1 - T_j'(z))u_c + u_z \equiv 0$, implicitly defines the earnings supply function (we suppress subscripts here as shorthand).

## 2.1 Intensive Margin Responses

Given this setup, we briefly review the intensive margin effect of a kink or notch. As shown in Saez (2010), a kinked budget set leads to a discontinuity in the earnings density at the kink due to intensive margin responses. Assuming a smooth distribution of ability $n$, a range of individuals who would earn between $z^*$ and $z^* + \Delta z^*$ in state 0 will respond in state 1 by reducing earnings to the kink at $z^*$. This is referred to as

---

[12] This implies a standard single-crossing property assumed in these models, which generates rank preservation in earnings, conditional on earning a positive amount.

"bunching" at the kink. The reduction in earnings $\Delta z^*$ of the "marginal buncher"—*i.e.* the buncher who earns the most, $z^* + \Delta z^*$, in state 0—can be related to the size of the change in the marginal tax rate at the kink, $d\tau \equiv \tau_1 - \tau_0$, in order to estimate an intensive margin elasticity (Saez, 2010).

In the case of a notched budget set, intensive margin responses are analyzed in Kleven and Waseem (2013). Those who would earn between $z^*$and $z^* + \triangle z^N$ in state 0 bunch at $z^*$ in state 1. In state 1, in the absence of adjustment frictions no one will locate in the strictly dominated region above the notch. However, in the presence of adjustment frictions, there may be "inert" individuals who remain in the dominated region rather than bunching. In this case, the amount of bunching in combination with the share of earners remaining in the dominated region can be used to estimate a structural intensive margin elasticity.

## 2.2 Extensive Margin Responses

In addition to the intensive margin response that much recent literature has focused on, individuals may also respond at the extensive margin. In the model of a kink briefly described in Section 2.1 and exposited in greater detail elsewhere such as Saez (2010), preferences and budget sets are convex, which restricts a small tax change only to affect the choice between zero and infinitesimally small earnings supply (*e.g.* Kleven and Kreiner, 2005). To capture the realistic pattern of potential entry to or exit from non-trivial levels of earnings, we introduce a fixed cost of employment (Cogan, 1981; Eissa *et al.*, 2008). Utility conditional on working is now given by:

$$u\left(c_{nj}, z_{nj}; n\right) = v\left(c_{nj}, z_{nj}; n\right) - q_{nj} \cdot 1\left\{z_{nj} > 0\right\} \tag{3}$$

where $j \in \{0,1\}$ indexes the state of the world, and the state-specific, additively separable fixed cost of employment, $q_{nj}$, is drawn from a distribution with CDF $G\left(q|n\right)$ and pdf $g\left(q|n\right)$. If an agent does not work, she receives a reservation level of utility of $u\left(c^0, 0; n\right) = v^0$ in either state of the world.[13],[14] We can interpret our framework as accommodating extensive margin frictions in the form of a fixed cost, as such frictions could be considered part of the fixed cost of working. We often refer to the fixed cost of employment as the fixed cost of working.

We pay special attention to whether or not the individual locates at a corner, *i.e.* $z_{nj} = 0$. Let $\tilde{z}_{nj}$ denote the optimal level of earnings in state $j$ conditional on working. This is chosen by maximizing $u\left(c, z; n\right)$ subject to (2). The individual works in state $j$ if:

$$v\left(\tilde{z}_{nj} - T_j\left(\tilde{z}_{nj}\right), \tilde{z}_{nj}; n\right) - q_{nj} > v^0 \tag{4}$$

Our key behavioral response of interest is the extensive margin response to the presence of a kink or notch. Here we define an individual's type by their optimal interior earnings conditional on working in state 0, *i.e.* $\tilde{z}_{n0}$. (We will often refer to $\tilde{z}_{n0}$ as "counterfactual" or "desired" earnings.) An isomorphism exists

---

[13] Without loss of generality, the outside option, $v^0$, does not vary with $n$ or $j$. This is because cross-sectional and state-specific variation in the outside option is not separately identified from the fixed cost of entry $q_{nj}$. We therefore collapse all such variation into the fixed cost of entry.

[14] Writing the fixed cost $q$ as separable from $v$, as in (3) above, simplifies the exposition. Without loss of generality, this model is equivalent to a model in which these are not separable *per se*, and instead we express utility simply as $v\left(c, z; n\right)$. Letting $c_n^0$ be consumption when not working, we can then posit a discontinuity in $v\left(c, z; n\right)$ at the boundary of the support of $z$ that reflects the fixed cost. Thus, we can define a fixed cost $q_n$ as: $q_n \equiv \lim_{z->0+}\left[v\left(c_n^0, 0; n\right) - v\left(c_n^0, z, n\right)\right]$.

between this earnings amount and ability $n$, and for empirical purposes using an earnings amount is natural to implement. The probability of working in state $j$ conditional on type $\tilde{z}_{n0}$ is:

$$\Pr\left(z_{nj} > 0 \mid \tilde{z}_{n0}\right) = \Pr\left(q_{nj} \leq v\left(\tilde{z}_{nj} - T_j\left(\tilde{z}_{nj}\right), \tilde{z}_{nj}; n\right) - v^0 \mid \tilde{z}_{n0}\right) = G\left(\bar{q}_{nj} \mid n\right) \tag{5}$$

where:

$$\bar{q}_{nj} \equiv v\left(\tilde{z}_{nj} - T_j\left(\tilde{z}_{nj}\right), \tilde{z}_{nj}; n\right) - v^0 \tag{6}$$

is the critical value for the fixed cost of employment that leaves the agent indifferent between working and not working. We allow the $G(\cdot)$ function to vary across individuals so that we have two sources of heterogeneity: (1) preferences captured by the $v(\cdot)$ function pin down intensive margin heterogeneity but also affect the extensive margin through $\bar{q}_{n1}$, and (2) the unrestricted heterogeneity in the $G(\cdot)$ function allows for differences in extensive margin responses independent of the $v(\cdot)$ function.[15] We make a number of assumptions regarding smoothness in heterogeneity. First, we assume that $G\left(q_{nj} \mid n\right)$ is continuous. Second, we assume that the partial derivative of $G\left(q_{nj} \mid n\right)$ with respect to $q_{nj}$, $g\left(q_{nj} \mid n\right)$, is continuous in $q_{nj}$ and $n$. Third, we similarly assume that the partial derivative of $G\left(q_{nj} \mid n\right)$ with respect to $n$, $\partial G\left(q_{nj} \mid n\right) / \partial n$, is continuous in $q_{nj}$ and $n$. Finally, we assume that the CDF of $n$ is continuously differentiable.

## 2.3   Modeling Incentives with a Kink or Notch

To demonstrate the impact of a kink or notch on the decision to work, we illustrate the extensive margin incentives created by a kink or notch in Figure 2. Here we plot the $ANTR \equiv 1 - [T(z) - T(0)]/z$, as a function of desired earnings. The ANTR measures the share of pre-tax income that is kept after taxes when working and earning $z$. With a linear tax schedule, the ANTR is constant at $1 - \tau_0$. With a kinked tax schedule, the ANTR decreases above $z^*$, and the slope of the ANTR decreases discontinuously at $z^*$. With a notched schedule, the level of the ANTR discontinuously decreases at $z^*$ (and its slope changes as well).

In Figure 3 we illustrate the indifference curves governing the extensive margin decision under the alternative tax schedules. We model the fixed cost of working visually by allowing agents to choose a level of earnings along the prevailing tax schedule, or to earn zero and receive a level of consumption of $c^0$.[16]

Panels A and B show a kinked tax schedule, while Panels C and D show a notch. In Panels A and C the agent's optimal level of earnings conditional on working is $z^*$. In this case she prefers earning $z^*$ to earning zero. Her response to the kink or notch is simply a reduction in earnings. In Panels B and D the agent similarly has optimal earnings of $z^*$ conditional on having positive earnings. In this case the individual's preferences lead her to earn zero rather than earning at the kink. Next we formally explore such responses.

---

[15] If extensive margin responses were instead only driven by the value function $v(\cdot)$, we might generate unexpected predictions or restrictions on the employment function. For example, if $G(\cdot)$ were homogeneous in our model, we would necessarily require that labor force participation be upward sloping as a function of $\tilde{z}_{n0}$.

[16] For the purposes of this figure, we can redefine $c^0$ using the following identity $u\left(c^0, 0; n\right) \equiv v^0 + q_{nj}$.

## 2.4 Employment Probability with a Kink and Unconstrained Intensive Margin Responses

In modeling the employment response to non-linear budget sets, we focus on the case of a kink because our empirical application features a kink. We therefore begin with the kink setting, initially considering the standard context in which individuals are free to adjust their earnings anywhere on the intensive margin. In other words, individuals' earnings, conditional on having positive earnings $\tilde{z}_{nj}$, may differ across the two tax schedules, and earnings choices are subject to no constraints other than the budget constraint $c = z - T(z)$. Let the employment function in state 1, conditional on counterfactual, interior earnings in state 0, be $\Pr(z_{n1} > 0 | \tilde{z}_{n0})$. This is the probability of having zero earnings in state 1 as a function of the level of earnings in state 0. We have shown that $\Pr(z_{n1} > 0 | \tilde{z}_{n0}) = G(\bar{q}_{n1} | n)$. We now explore how this function changes as $\tilde{z}_{n0}$ changes. In Appendix A, we prove the following result:

**Proposition 1** *If the individuals can freely adjust their earnings on the intensive margin, then the employment probability, as a function of earnings in state 0, will exhibit no first-order change in slope at $z^*$:*

$$\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d\Pr(z_{n1} > 0 | \tilde{z}_{n0})}{d\tilde{z}_{n0}} = \lim_{\tilde{z}_{n0} \to z^{*-}} \frac{d\Pr(z_{n1} > 0 | \tilde{z}_{n0})}{d\tilde{z}_{n0}}. \tag{7}$$

Figure 1A illustrates this result. The $x$-axis measures counterfactual earnings in state 0 conditional on having positive earnings, *i.e.* $\tilde{z}_{n0}$. The $y$-axis plots an illustrative employment rate. The dashed line represents a presumed smooth relationship between the employment rate in state 0 under a linear tax schedule, *i.e.* $\Pr(z_{n0} > 0 | \tilde{z}_{n0})$, and earnings conditional on having positive earnings, *i.e.* $\tilde{z}_{n0}$. The dotted line plots the employment rate in state 1 under a kinked tax schedule, $\Pr(z_{n1} > 0 | \tilde{z}_{n0})$, while the $x$-axis continues to plot $\tilde{z}_{n0}$. In this case, we assume that individuals are unrestricted in their earnings choices. We see that the employment function is unchanged at counterfactual earnings levels below $z^*$, as the tax schedule remains the same in the two states. Above $z^*$ we see a gradual decrease in the probability of positive earnings in state 1 relative to state 0, due to the decrease in the ANTR (recall Figure 2A). Nonetheless, the kink in the ANTR does not translate into a kink in the employment rate. Intuitively, the ability to adjust on the intensive margin smoothes the first-order changes in the slope of the ANTR at $z^*$.

## 2.5 Employment Probability with a Kink and Constrained Intensive Margin Responses

Although the kink in the budget set does not lead to a kink in employment when individuals are free to adjust to any earnings level, a kink in the employment rate arises when frictions impede intensive margin adjustment. To illustrate the ideas as simply and transparently as possible, we begin with the case that individuals are completely restricted from earning other amounts at the intensive margin, so that $\tilde{z}_{n1} \equiv \tilde{z}_{n0}$. Individuals are still allowed to vary their extensive margin choices across the two states. Numerous papers have found evidence for such restrictions on labor supply or earnings, for example due to constraints on

hours or earnings choices, or fixed costs of adjustment that would prevent adjustment to the kink for those in this region (*e.g.* Altonji and Paxson, 1988; Dickens and Lundberg, 1993; Chetty *et al.*, 2007; Chetty *et al.*, 2011; Gelber *et al.*, 2013). Modeling and estimating frictions that could give rise to such restrictions is the focus of other work (Gelber *et al.* 2013); for the purpose of this paper it is not necessary to take a stand on what specific process gives rise to such restrictions, as the existence of such restrictions is sufficient to generate our results. In Appendix A, we prove the following result:

**Proposition 2** *If individuals are not able to adjust earnings on the intensive margin, i.e.* $\tilde{z}_{n1} \equiv \tilde{z}_{n0}$, *then the employment probability, as a function of desired earnings conditional on employment in state 0, will exhibit a first-order change in slope (i.e. a kink) at* $z^*$. *This kink is given by:*

$$\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d \Pr(z_{n1} > 0 | \tilde{z}_{n0})}{d\tilde{z}_{n0}} - \lim_{\tilde{z}_{n0} \to z^{*-}} \frac{d \Pr(z_{n1} > 0 | \tilde{z}_{n0})}{d\tilde{z}_{n0}} = -d\tau \cdot \lambda_{n^*} \cdot g(\overline{q}_{n^*1} | n^*) \tag{8}$$

*where* $\lambda_n \equiv v_c$ *is the marginal utility of consumption, and* $\overline{q}_{n^*1}$, $n^*$, *and* $\lambda_{n^*}$ *all refer to the individual for whom* $\tilde{z}_{n0} = z^*$.

Returning to Figure 1A, the solid line depicts the relationship between counterfactual earnings in state 0 and the probability of positive earnings when a kink is present in state 1 and $\tilde{z}_{n1} \equiv \tilde{z}_{n0}$. The slope of the employment rate now discontinuously changes at $z^*$, where the ANTR also changes slope. We have closed one of the channels through which individuals respond to the increase in tax liability, and thus the slope of the employment rate decreases discontinuously at $z^*$. Equation (8) has an intuitive interpretation: the kink in the employment rate is proportional to $d\tau$, the size of the kink in the tax schedule; $\lambda_{n^*}$, the marginal utility of after-tax income; and $g(\overline{q}_{n^*1} | n^*)$, the density of workers who are on the margin of entering employment in state 0. These parameters apply to the individual earning $z^*$ in state 0. Because the kink in the employment rate we model is only detectable in the presence of frictions in intensive margin adjustment (whether due to fixed costs or constraints on earnings), our method can therefore also provide an incidental test of whether intensive margin frictions exist.

In Appendix B.1, we extend the model to allow for an arbitrary set of discrete earnings choices (other than $z^*$ for the constrained types), rather than the parsimonious and transparent—but more restrictive—assumption of this section that $\tilde{z}_{n1} \equiv \tilde{z}_{n0}$.[17] We show that there is a discontinuity in the slope of the employment rate as long as individuals who earn near $z^*$ in state 0 are constrained from making small adjustments exactly to $z^*$ in state 1. We also demonstrate in this case that the extensive margin behavior for those near $z^*$ is not affected by the ability to make discrete adjustments to other earnings levels, *e.g.* a part-time job.

---

[17] In fact, one can allow for some continuity in the choice set as long as, loosely speaking, individuals cannot make adjustments exactly to $z^*$. See Appendix B.1 for details.

## 2.6 Estimating an Extensive Margin Elasticity with a Kink

When a kink is created in the employment rate due to a kink in the budget set, we may use this behavioral response to estimate an extensive margin elasticity $\eta$ for a given average net-of-tax rate $1 - a \equiv 1 - [T(z) - T(0)]/z$. Using the fact that $\Pr(z > 0) = G(\bar{q})$, we have:

$$\eta \equiv \frac{d\Pr(z > 0)}{d(1 - a)} \frac{1 - a}{\Pr(z > 0)} = g(\bar{q}) \frac{\partial \bar{q}}{\partial(1 - a)} \frac{1 - a}{\Pr(z > 0)}$$

For now we maintain the assumption of $\tilde{z}_{n1} \equiv \tilde{z}_{n0}$ as in Section 2.5 and denote $\beta$ the estimated discontinuity at $z^*$ in the slope of the employment probability. Given (8) and the fact that $\partial \bar{q}/\partial(1 - a) = \lambda z$, we have:

$$\eta^* = g(\bar{q}_{n^*1}|n^*) \frac{\partial \bar{q}_{n^*1}}{\partial(1 - a)} \frac{1 - a}{\Pr(z_{n1} > 0| \tilde{z}_{n0} = z^*)} = \frac{\beta}{\alpha} \cdot \frac{1 - a}{\Pr(z_{n1} > 0| \tilde{z}_{n0} = z^*)} \tag{9}$$

where $\eta^*$ is the extensive margin elasticity for the individual earning $z^*$ in state 0, and $\alpha = -d\tau/z^*$ is the magnitude of the kink at $z^*$ in the slope of the ANTR, $1 - a$.[18] In other words, our model naturally suggests an RKD approach to estimating the elasticity of extensive margin participation with respect to the ANTR. Of course, although this approach yields a single elasticity, if the function $g(\cdot)$ is heterogeneous across different sub-groups then the elasticity may also vary across these sub-groups. Note also that although our approach does not allow us to estimate the extent to which the marginal utility of income $\lambda_{n^*}$ and the fixed cost distribution $g(\bar{q}_{n^*1}|n^*)$ separately contribute to $\eta^*$, the extensive margin elasticity $\eta^*$ on its own can allow us to perform counterfactual simulations of the implications of different tax schedules (*e.g.* Eissa *et al.*, 2008).

## 2.7 Interpreting the Observed Elasticity with a Kink

Thus far we have derived an expression for the extensive margin elasticity under a kink assuming that individuals cannot adjust on the intensive margin across tax schedules (or in Appendix B.1, assuming that individuals can make adjustments anywhere, except that those near $z^*$ in state 0 cannot make adjustments in state 1 to $z^*$). In fact, some individuals may be able to adjust on the intensive margin to $z^*$. Suppose that under a kink we observe both intensive margin bunching (described in Section 2.1) and a discontinuity at counterfactual earnings $z^*$ in the slope of the employment rate. In this case we discuss how to interpret the "observed" extensive margin elasticity, $\hat{\eta} = (\hat{\beta}/\alpha) \cdot (1 - a)/\Pr(z_{n1} > 0| \tilde{z}_{n0} = z^*)$, where $\hat{\beta}$ is the estimated kink at $z^*$ in the employment rate. Below we explore frameworks that allow for both bunching among some individuals and a kink in the employment rate among others. We show that the observed elasticity, $\hat{\eta}$, can be interpreted as a weak lower bound on the "structural" elasticity $\eta^*$, *i.e.* $\hat{\eta} \leq \eta^*$, where "observed" and "structural" are used in a sense analogous to Chetty (2012). Here the "observed" elasticity refers to what we observe due to an increase in marginal tax rates above the kink among those earning near $z^*$, which is in part affected by the lack of an extensive margin response among those unconstrained at the intensive margin. The "structural" elasticity refers to the elasticity we would hypothetically observe in response to a

---

[18] To see this, note that below $z^*$, $(1 - a) = 1 - \tau_0$ and $\partial(1 - a)/z = 0$, while above $z^*$, $(1 - a) = 1 - (\tau_0 z^* + \tau_1(z - z^*))/z = 1 - \tau_1 + d\tau z^*/z$ and $\partial(1 - a)/z = -d\tau z^*/z^2$. When $z = z^*$, this reduces to $-d\tau/z^*$.

change in average tax rates everywhere, among those earning near $z^*$ and including both those constrained and those unconstrained on the intensive margin.[19]

### 2.7.1   Model with Mixture of Types

One approach to capturing both bunching and a kink in the employment probability is to posit a model with two types of individuals: Type $A$ that can adjust on the intensive margin, and Type $B$ that cannot (*e.g.* Kleven and Waseem, 2013). We have shown that among Type $A$ agents, the employment function has a continuous slope. Among Type $B$ agents, the slope is discontinuous at $z^*$. Let $\pi_B^* = \Pr\left(B \mid \tilde{z}_{n0} = z^*\right)$ be the probability of being Type $B$ conditional on having earnings at $z^*$ in state 0. It follows that:

$$\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d \Pr\left(z_{n1} > 0 \mid \tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} - \lim_{\tilde{z}_{n0} \to z^{*-}} \frac{d \Pr\left(z_{n1} > 0 \mid \tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} = -\pi_B^* \cdot d\tau \cdot g_B\left(\bar{q}_{n*1} \mid n^*\right) \tag{10}$$

where $g_B\left(\cdot\right)$ is the pdf of fixed costs among Type $B$ agents. In this sense, our estimate of the extensive margin elasticity is attenuated by a factor $\pi_B^*$ and can therefore be considered a weak lower bound on the elasticity among Type $B$ agents with state 0 earnings $z^*$. Among Type $A$ agents who earn above $z^*$ in state 0, there may also be a response to the kink that increases gradually above $z^*$ as in Figure 1A, but our method only picks up responses among Type $B$ agents. Nonetheless, the observed elasticity is a lower bound on the elasticity among all of those earning $z^*$ in state 1:

$$\hat{\eta} = \frac{\hat{\beta}}{\alpha} \cdot \frac{1 - a}{\Pr\left(z_{n1} > 0 \mid \tilde{z}_{n0} = z^*\right)} = \pi_B^* \cdot \eta_B^* \leq \pi_A^* \cdot \eta_A^* + \pi_B^* \cdot \eta_B^* = \eta^* \tag{11}$$

In principle it would be possible to use the observed elasticity $\hat{\eta}$, together with an estimate of the fraction constrained $\pi_B^*$, to estimate the structural elasticity among constrained agents, $\eta_B^* = \hat{\eta}/\pi_B^*$. However, estimating $\pi_B^*$ requires more restrictive assumptions, including assuming that types $A$ and $B$ have the same distribution of $q$ and $n$, as we explain in Appendix B.2. Our observed elasticity remains of interest regardless of the underlying proportion of agents displaying different types of behavior, both in the sense that policy-makers are interested in the raw employment effects of changing policy parameters like the average tax rate, and in the sense that it reflects a lower bound on the structural elasticity.

### 2.7.2   Model with a Fixed Cost of Intensive Margin Adjustment

In an alternative model of intensive margin frictions, individuals face a fixed cost of adjusting earnings on the intensive margin in response to variation in the tax schedule (see Gelber *et al.*, 2013, for a detailed exposition of this model). Such frictions could reflect a variety of factors, including lack of knowledge of a tax regime, the cost of negotiating a new contract with an employer, or the time and financial cost of job search. With a fixed intensive margin adjustment cost individuals will only adjust if the utility gain of intensive margin adjustment exceeds the fixed cost. Recall that for individuals earning $\tilde{z}_{n0} < z^*$ there is no change in the tax schedule from state 0 to state 1, and therefore $\tilde{z}_{n1} = \tilde{z}_{n0}$. Gelber *et al.* (2013) show that due to the fixed cost of intensive margin adjustment individuals with $\tilde{z}_{n0} > z^*$ for whom $\tilde{z}_{n0}$ is sufficiently close to $z^*$ will also prefer to keep earnings fixed across the two tax schedules, *i.e.* $\tilde{z}_{n1} = \tilde{z}_{n0}$. The reason is that the utility

---

[19] Although we use the term "structural elasticity," the extensive margin elasticity depends on the number of individuals who are just indifferent between working and not working—*i.e.* $g\left(\bar{q} \mid n^*\right)$—which varies depending on the employment level.

gain from adjusting on the intensive margin converges to zero as $\tilde{z}_{n0}$ approaches $z^*$: the optimal level of earnings is $z^*$ in state 1 for this group. In this case, in a close enough neighborhood around $z^*$, individuals behave as in Section 2.5, and our results from Section 2.5 follow. In other words, a fixed cost of intensive margin adjustment can rationalize the assumption that some individuals do not adjust to $z^*$ in state 1, and it follows that the observed elasticity reflects the structural elasticity.[20]

## 2.8 Extensive Margin Response to a Notch

Our method can be extended to the case of a notch. As in the case of a kink, we first explore the setting in which individuals are free to adjust earnings on the intensive margin. We then contrast this with a setting with frictions on the intensive margin, as documented in Kleven and Waseem (2013). In Appendix A we prove the following:

**Proposition 3** *If individuals can freely adjust their earnings on the intensive margin, then the employment probability, as a function of earnings in state 0, will be continuous at $z^*$:*

$$\lim_{\tilde{z}_{n0}\to z^{*+}} \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right) = \lim_{\tilde{z}_{n0}\to z^{*-}} \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right). \tag{12}$$

*Furthermore, the employment probability, as a function of earnings in state 0, will exhibit no first-order change in slope at $z^*$:*

$$\lim_{\tilde{z}_{n0}\to z^{*+}} \frac{d\Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} = \lim_{\tilde{z}_{n0}\to z^{*-}} \frac{d\Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right)}{d\tilde{z}_{n0}}. \tag{13}$$

This is illustrated in the dotted line in Figure 1B. Its slope is continuous at $z^*$.

**Proposition 4** *If individuals are not able to adjust earnings on the intensive margin, i.e. $\tilde{z}_{n,1} = \tilde{z}_{n,0}$, then the employment probability, as a function of desired earnings in state 0, will exhibit a first-order change in levels (i.e. a discontinuity) at $z^*$. This discontinuity is given by:*

$$\lim_{\tilde{z}_{n0}\to z^{*+}} \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right) - \lim_{\tilde{z}_{n0}\to z^{*-}} \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right) = G\left(\bar{q}_{n^*1}^+ \middle| n^*\right) - G\left(\bar{q}_{n^*0} \middle| n^*\right) \tag{14}$$

*where the critical earnings levels are defined as $\bar{q}_{n^*0} \equiv v\left(z^* - T_0\left(z^*\right), z^*; n^*\right) - v^0$ and $\bar{q}_{n^*1}^+ \equiv v\left(z^* - T_1^+\left(z^*\right), z^*; n^*\right) = v\left(z^* - T_0\left(z^*\right) - dT - d\tau \cdot z^*, z^*; n^*\right)$ and $T_1^+\left(z^*\right) \equiv \lim_{z\to z^{*+}} T_1\left(z\right)$. Moreover, using a first-order approximation for the $G\left(\cdot\right)$ and $v\left(\cdot\right)$, we have:*

$$\lim_{\tilde{z}_{n0}\to z^{*+}} \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right) - \lim_{\tilde{z}_{n0}\to z^{*-}} \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right) \approx -\left[dT + d\tau \cdot z^*\right] \cdot \lambda_{n^*} \cdot g\left(\bar{q}_{n^*0} \middle| n^*\right) \tag{15}$$

*where $\lambda_{n^*} \equiv v_c$ is the marginal utility of consumption, and $\bar{q}_{n^*1}^+$, $\bar{q}_{n^*0}$, $n^*$, and $\lambda_{n^*}$ all refer to the individual for whom $\tilde{z}_{n0} = z^*$.*

This is illustrated in the solid line in Figure 1B. This line shows a discontinuity in the level of employment at $z^*$.

---

[20] If there is heterogeneity in the fixed cost of intensive margin adjustment, then as long as the fixed cost is strictly positive — *i.e.* $\phi > 0$ for all $\phi$ — then the above still holds in a neighborhood near $z^*$.

In the case of intensive margin frictions, the employment probability also features a discontinuity in slope, due to the discontinuity in the ANTR shown in Figure 2b. The magnitude of this change in slope depends on the functional form of the utility function and the distribution of fixed costs. As an example, in Appendix A we provide a result after imposing some restrictions on both functions.

Compared to the case of a kink in the tax schedule, we continue to predict continuity in the level and slope of the employment probability as a function of counterfactual earnings when intensive margin adjustment is possible. However, once we introduce frictions on the intensive margin, the employment probability exhibits a discontinuity in both levels and slope. This suggests using a Regression Discontinuity Design (RDD) to recover our parameter of interest non-parametrically using the discontinuity in levels. If we estimate a discontinuity $D$ in the employment probability at $z^*$, we can use this to recover a discrete version of the extensive margin elasticity:

$$
\eta^N = \frac{\triangle \Pr\left(z > 0\right)}{\triangle\left(1-a\right)} \frac{1-a}{\Pr\left(z>0\right)} = \frac{G\left(\bar{q}_{n^*0}|\,n^*\right) - G\left(\bar{q}_{n^*1}^+|\,n^*\right)}{\triangle\left(1-a\right)} \frac{1-a}{\Pr\left(z>0\right)} = \frac{D}{-[dT/z^* + d\tau]} \cdot \frac{1-a}{\Pr\left(z_{n1} > 0|\,\tilde{z}_{n0} = z^*\right)}.
$$
(16)

As in the case of a kink, with a notch and both constrained and unconstrained types, the observed elasticity will reflect an attenuated version of the structural elasticity. This is governed by an analogous set of formulas to (10) and (11) in the case of a kink.

# 3 Policy Environment

## 3.1 Annual Earnings Test (AET) Rules

We apply the kink framework from the prior section in the context of estimating responses to the kink created by the AET. Social Security Old-Age and Survivors Insurance (OASI) provides annuity income to older Americans and to survivors of deceased workers. Individuals with sufficient years of eligible earnings can claim OASI benefits through their own work history as early as age 62, the Early Entitlement Age (EEA). They can claim full benefits once they reach the Normal Retirement Age (NRA), which is 65 for individuals in our sample. The AET reduces current OASI benefits in proportion to earnings above an exempt amount. For example, consider a 63-year-old earning $23,040 in 2018, receiving $1,000 in monthly benefits, and facing a $17,040 exempt amount. Her annual benefits would be reduced by $3,000 = ($23,040 − $17,040) × 50%, equal to 3 months of benefits. For those whose age is at the NRA and over, the real exempt amount is substantially higher than for those below the NRA. Over the years we study, 1978 to 1987, the exempt amount for those under NRA ranged from $9,787 to $11,517, while the exempt amount for those NRA and over was on average around $3,600 higher than the amount for those under NRA. The exempt amount by year and age range is shown in Appendix Figure 1. For those under the NRA but above the EEA—the main group that our empirical work studies—the benefit reduction rate (BRR) was 50 percent throughout the period we study, 1978 to 1987. During our period, the AET applied to earnings from ages 62 to 71 during

1978 to 1982, and from ages 62 to 69 during 1983 to 1987.

During the period we study, the AET rules for married couples imply that the couple's total benefit is reduced at the rate of one dollar for every two dollars that each spouse's individual current earnings exceed the exempt amount applying to each spouse. How this occurs depends on how benefits are claimed. If each spouse is a primary OASI beneficiary, then the AET reduces each spouse's separate benefit by the BRR multiplied by that spouse's individual current earnings in excess of the exempt amount. If one spouse is a primary beneficiary and the other is a secondary or dual-entitled beneficiary, the couple's total benefits are reduced by the BRR multiplied by the primary beneficiary's current individual earnings in excess of the exempt amount, and further reduced by the BRR multiplied by the secondary or dual beneficiary's current individual earnings in excess of the exempt amount. In either case, the relevant amount for applying the AET is each individual's current annual earnings, which we observe in our data.

When current OASI benefits are lost to the AET, future scheduled benefits may be increased in some circumstances. This is sometimes referred to as "benefit enhancement." For beneficiaries below the NRA in particular, the benefit enhancement, known as the "actuarial adjustment," raises future benefits whenever a claimant earns over the AET exempt amount.[21] Future benefits are raised by 0.55 percent per month of benefits withheld for the first three years of AET assessment. Returning to the example above, consider the 63-year-old receiving \$1,000 in monthly benefits due to the AET. Upon reaching the NRA, her monthly benefits would increase by around \$16.50 = 0.0055 x 3 x \$1,000. On average, this adjustment is roughly actuarially fair when considering the timing of claiming OASI (Diamond and Gruber, 1999).

## 3.2 Applying the Model in our Policy Setting

Our theoretical model for kinks in Section 2 generally applies to any context with a kinked budget set, and it is specifically applicable in the case of the kink the AET creates at the exempt amount. Following previous literature (*e.g.* Friedberg, 1998; Friedberg, 2000), we model the AET as creating a positive implicit MTR for some individuals above the exempt amount, consistent with the empirical finding in this previous literature that some individuals bunch at the exempt amount. The MTR we assume is the BRR associated with the reduction in the family's total benefits (as described above), as is appropriate in a unitary model of family labor supply (Becker, 1976). Thus, we interpret our results within this model as the effect of the AET applying to a given spouse's earnings (but potentially reducing both spouses' combined current OASI benefits at the BRR) on that spouse's employment.

Even after taking benefit enhancement into account, there are a number of reasons that the AET could create a positive implicit MTR, and therefore affect earnings and employment decisions. Those whose expected lifespan is shorter than average should expect to collect OASI benefits for less time than average, implying that the AET is financially punitive. Liquidity-constrained individuals or those who discount faster than average may also reduce work in response to the AET. Finally, many individuals also may not

---

[21]Social Security Administration (2012); Gruber and Orzag (2003). Gelber *et al.* (2013) show that there is no evidence of a bunching response to the upward notch created by this aspect of the policy.

understand the AET benefit enhancement or other aspects of OASI (Honig and Reimers, 1989; Liebman and Luttmer, 2015; Brown *et al.*, 2013). Indeed, the earnings test is widely viewed as a pure tax. Most popular guides highlight the Earnings Test but do not note the subsequent adjustment in benefits (Gruber and Orszag, 2003). During the period that we study, the popular guide *Your Income Tax* (J.K. Lasser Institute, 1987), for example, warned readers that if "you are under age 70, Social Security benefits are reduced by earned income," but did not mention benefit adjustment. This is consistent with survey evidence showing that most older adults understand that the AET reduces current benefits but do not know that later benefits may be affected (Brown and Perron, 2011). We follow most previous work and do not distinguish among the potential reasons for a response to the AET, though Gelber *et al.* (2013) explore some of these potential mechanisms.[22]

Our policy setting has a number of specific features that can be incorporated into the method for estimating elasticities using kinks in Section 2. First, individuals are subject to the AET only when they choose to claim OASI; this mirrors other transfer programs that must be claimed to receive benefits. In Appendix B.4 we extend our model to show under mild assumptions that when claiming can respond jointly with working, the kink in the probability of earning a positive amount will be less negative than in the case in which claiming is exogenous. Nonetheless, we can recover the elasticity among those claiming OASI by dividing the observed elasticity by the probability of claiming at $z^*$. In the most general case in which we relax these mild assumptions, we can still obtain a lower bound on this employment elasticity by subtracting the kink in claiming from the kink in the employment probability. This is one of several reasons discussed over the course of the paper that we may estimate a lower bound on the structural elasticity.

Second, in Section 2 we examined a static setting in which individuals only consider the current period's budget set and incentives. In a dynamic setting we may observe individuals serially facing a linear tax (corresponding to state 0), followed temporally by a tax that creates a convex kink (corresponding to state 1); indeed this is the case in our empirical application, in which individuals are not subject to the AET initially and become subject later. In such a dynamic setting with two or more periods, our static framework will hold only in a model with borrowing constraints, myopia, or no income effects. In Appendix B.3 we present a fully dynamic, multi-period model with a joint decision over saving and earnings, and we show that under a set of empirically relevant assumptions our results for kinks still hold: if intensive margin frictions exist, a change in the slope of the employment rate occurs at the budget set kink, and we can still recover the elasticity using (9). This appendix also discusses the interpretation of the parameter our elasticity represents, depending on whether the tax change is anticipated or unanticipated. If agents act as if they do not anticipate the change in an initial period—consistent with our empirical findings in this sample as well as the results in Gelber *et al.* (2013) and Gelber *et al.* (2016)—then we can interpret the results as

---

[22]Our framework provides certain secondary contributions relative to existing theoretical treatments of the AET (Blinder *et al.*, 1980; Burkhauser and Turner, 1978; Vroman, 1985; Burtless and Moffitt, 1985; Gustman and Steinmeier, 1985; Honig and Reimers, 1989; Gustman and Steinmeier, 1991; Friedberg, 1998). First, we focus on the AET's extensive margin effects; in many cases, existing models focus on the AET's intensive margin incentives. Note, however, that Vroman (1979) and Friedberg and Webb (2009) describe informally how constraints on labor supply could lead the AET to have extensive margin effects. Second, and more important, we introduce a new methodology that exploits the sharp changes in incentives around the AET exempt amount to estimate extensive margin effects. We do not model shifts from the formal labor market to "off-the-books" employment as discussed in Christensen (1990).

reflecting the impact of an unanticipated change in policy. Third, after presenting a basic dynamic model in Appendix B.3, we extend this model to tailor it to our particular policy setting by showing that our method still applies when reductions in current benefits due to the AET can lead to increases in later benefits, as under benefit enhancement.

# 4    Empirical Strategy

Our empirical strategy is tailored to our empirical context, involving the kink in the budget set created by the AET. Our model indicates that to estimate the extensive margin effect of the kink, we should seek to estimate the change in the slope of employment at the kink, which can be accomplished with an RKD. We therefore focus on the empirical implementation of our method for kinks using an RKD. In the context of a budget set notch, it would be possible to estimate the discontinuity in the level of the employment probability in (14) using an RDD (see the survey in Lee and Lemieux, 2010), in which case parallel implementation considerations would arise to those discussed below.

## 4.1    Regression Kink Design

Recent work on RKDs shows that under certain conditions, a change in the slope of the treatment intensity can be used to identify local treatment effects by comparing the relative magnitudes of the kink in the treatment variable and the induced kink in the outcome variable (Nielsen *et al.*, 2010). Under smoothness assumptions that parallel the smoothness assumptions of our model, the estimates can be interpreted as a treatment-on-the-treated or "local average response" parameter (Card *et al.*, 2015).

In our context, the treatment intensity is the effective ANTR for individual $i$ at time $t$, $ANTR_{it}$, and the primary outcome variable is the probability of positive earnings after becoming subject to the AET, $\Pr(L_{it}=1)$, where $L_{it}$ is a dummy for positive earnings (and where $\Pr(L_{it}=1)$ in the empirical model corresponds to $\Pr(z_{n1}>0)$ in the theoretical model). The $ANTR_{it}$ is a function of earnings conditional on working, which we take as desired earnings on a linear budget set $\tilde{z}_{i0t}$ (corresponding to $\tilde{z}_{n0}$ in the model in Section 2 above). Consistent with Figure 2, the AET creates a discontinuity at the exempt amount in the slope of the ANTR as a function of $\tilde{z}_{i0t}$. In this case, we can estimate the causal effect of the AET on the employment rate by estimating the change at the exempt amount in the slope of the employment rate as a function of $\tilde{z}_{i0t}$, in comparison to the change at the exempt amount in the slope of the ANTR. Specifically, using the RKD we can estimate the marginal effect of the ANTR on the employment probability as:

$$\frac{\partial \Pr(L_{it}=1)}{\partial ANTR_{it}}\bigg|_{\tilde{z}_{i0t}=z^*} = \frac{\lim\limits_{\tilde{z}_{i0t}\to z^{*+}}\frac{\partial \Pr(L_{it}=1|\tilde{z}_{i0t})}{\partial \tilde{z}_{i0t}} - \lim\limits_{\tilde{z}_{i0t}\to z^{*-}}\frac{\partial \Pr(L_{it}=1|\tilde{z}_{i0t})}{\partial \tilde{z}_{i0t}}}{\lim\limits_{\tilde{z}_{i0t}\to z^{*+}}\frac{\partial ANTR_{it}(\tilde{z}_{i0t})}{\partial \tilde{z}_{i0t}} - \lim\limits_{\tilde{z}_{i0t}\to z^{*-}} - \frac{\partial ANTR_{it}(\tilde{z}_{i0t})}{\partial \tilde{z}_{i0t}}} \tag{17}$$

That is, the marginal effect we can estimate is the change at the exempt amount in the slope of the employment probability as a function of desired earnings, divided by the change at the exempt amount in the slope of the ANTR as a function of desired earnings.

The numerator of (17) can be estimated by relating $\Pr(L_{it}=1)$ to desired earnings:

$$\Pr\left(L_{it}=1\right)=\sum_{j=0}^{J}\delta_j\left(\tilde{z}_{i0t}-z_t^*\right)^j + D\cdot\sum_{j=0}^{J}\beta_j\left(\tilde{z}_{i0t}-z_t^*\right)^j + \mu_t + \varepsilon_{it} \tag{18}$$

where $D=1\left\{\tilde{z}_{i0t}\geq z_t^*\right\}$ is an indicator for being above the exempt amount, the $\mu_t$ are time fixed effects, $\varepsilon_{it}$ is an error term, and $\beta_1$ is the change in the slope of $\Pr\left(L_{it}=1\right)$ at the exempt amount. We calculate $\Pr\left(L_{it}=1\right)$ at the individual level by averaging an indicator for employment over a range of ages, specifically ages 63 and 64 in our main specification. We retain the $t$ subscript to allow for the fact that in different specifications we may investigate employment over different time periods. This RKD yields a non-parametric estimate of the effect, because asymptotically the estimator uses only data arbitrarily close to the exempt amount and is unbiased without any functional form restriction on the underlying function to be estimated (Card *et al.*, 2015). Our model in Section 2 makes no specific functional form assumptions, and expression (9) for the elasticity only relies on expectations and partial derivatives, all of which can be estimated non-parametrically. Similarly, in the case of a budget set notch, an RDD would yield estimates that are non-parametrically valid (Lee and Lemieux 2010).

Identification of the effect of the AET on employment relies on two key assumptions (Card *et al.*, 2015). First, in the neighborhood of the exempt amount there is no discontinuity in the slope of age 63 to 64 employment that occurs for reasons unrelated to the AET. For example, beneficiaries' earnings could in principle be affected by other public programs, or by their human capital or work experience as manifested in their hourly wage. Married beneficiaries' earnings could also be affected by their spouses' effective MTRs. We follow Saez (2010) and subsequent literature studying the effects of public programs in assuming that these factors jointly would have a smooth effect on earnings in the neighborhood of $z^*$.[23] Second, conditional on unobservables, the distribution of the assignment variable, $\tilde{z}_{i0t}$, is smooth (*i.e.* the p.d.f. is continuously differentiable) in this neighborhood. These assumptions may not hold if we observe sorting in relation to the exempt amount, as indicated by a discontinuous change at the exempt amount in the slope or level of the density of the assignment variable, or in the distribution of predetermined covariates.

## 4.2 Measuring the ANTR

We would like to measure the ANTR in period $t$ (*i.e.* ages 63 to 64) using the rules of the AET and the earnings level $\tilde{z}_{i0t}$, the counterfactual level of earnings the individual would choose if there were a linear tax schedule in period $t$. We do not observe this counterfactual level of earnings, because earnings at ages 63 to 64 (analogous to $\tilde{z}_{n1}$ from the model above) are endogenous to the AET, and furthermore earnings are not observed for those who are not employed at those ages. Because of this endogeneity, actual observed

---

[23] The 1978 and 1986 amendments to the Age Discrimination in Employment Act (ADEA) extended the ages at which age discrimination in employment was prohibited, but this did not have a discontinuous effect on elderly work incentives at the exempt amount. The 1977 Social Security Amendments increased the Delayed Retirement Credit for those 65 to 69 beginning in 1982, and the maximum age to which the AET applied decreased from 71 to 69 in 1983, but again neither of these should confound our RKD strategy. Other programs, such as Medicaid, Supplemental Security Income, Disability Insurance (DI), or taxes such as unemployment insurance payroll taxes distort earnings incentives at low earnings levels. We eliminate essentially all DI recipients from our sample through the restriction to those who claimed at 62 or later. The kinks created by other programs are typically at least several thousand dollars away from the AET convex kink, and we verify that the kink in employment at the AET exempt amount in particular is statistically significant relative to placebo kink locations.

earnings at ages 62, 63, or 64 cannot serve as adequate proxies for this counterfactual earnings level.

Many other papers have grappled with the issue of how to proxy for earnings or wages if individuals choose to work, and thus how to proxy for the incentive to work. Given that the econometrician does not directly observe counterfactual earnings, there is no alternative to making some assumption. One solution is a selection correction in the context of the effect of wages on labor supply, which generally requires functional form assumptions (Heckman, 1979) or very powerful instruments (Powell, 1994). Another solution is to use demographics to impute earnings if an individual works (*e.g.* Meyer and Rosenbaum, 2001), which is more difficult in settings such as ours with a limited number of demographic variables in our administrative data.

Our solution to this issue is to estimate a bound on the elasticity by assuming that the ANTR an individual would face at age 63 or 64 at her desired earnings level is the same as the ANTR she would face given her age 60 earnings level. Thus, we proxy for desired earnings conditional on working at ages 63 and 64 in the absence of the AET, $\tilde{z}_{i0t}$, using actual earnings at age 60, $z_i^{60}$. Individuals do not face the AET kink at age 60, and therefore at age 60 their budget set is on average approximately linear in the region of the exempt amount. We use earnings at age 60, rather than 61, to avoid any anticipatory manipulation of earnings, and we show that there is no evidence for anticipatory manipulation in age 60 earnings in our sample. Age 60 is the closest age to 63-64 that our data show is not subject to concerns about manipulation, and therefore serves as the best proxy for counterfactual earnings. In our baseline RKD specification, our running variable is the distance between age 60 earnings in a given year and the exempt amount applying to those aged 62 to 64 in that year.

The use of a proxy variable and the inability to observe $ANTR_{it}$ raise challenges in the estimation of the first stage of a "fuzzy RKD" (see Battistin and Chesher, 2014, on measurement error in the general average treatment effect framework). In our case, we rely on an analytic expression for the denominator of equation (17). Specifically, we calculate the kink in $ANTR_{it}$ using the AET rules, *i.e.* $\alpha \equiv d\tau/z^*$, where $d\tau$ is the BRR. Next, we must address the use of a proxy for desired earnings, $z_i^{60}$, which may suffer from measurement error. Following Card *et al.* (2015), we model the measurement error in our proxy as:

$$z_i^{60} = \tilde{z}_{i0t} + p_{it} \cdot v_{it}, \tag{19}$$

where $v_{it}$ is a continuously distributed random variable, and $p_{it}$ is an indicator variable that equals zero with probability $\pi(\tilde{z}_{i0t}, v_{it}) > 0$. Thus, the error in our observed proxy is a composite of a continuously distributed variable and a mass point at zero. This implies that with some positive probability, $\pi$, earnings at age 60 equals desired earnings conditional on working in period $t$ under a linear tax schedule. Note that we need not assume that the measurement error is mean zero, only that the distribution of $v_{it}$, conditional on $z_i^{60}$, is a smooth function of $z^{60}$ and that $\pi(\tilde{z}_{i0t}, v_{it}) > 0$.

Importantly, if $\pi(\tilde{z}_{i0t}, v_{it}) < 1$, so that desired earnings are variable from age 60 to ages 63 to 64, then our procedure should overestimate the absolute value of the change in the ANTR at the exempt amount in the first stage, and therefore we should estimate a lower bound on the elasticity. We show this formally in Appendix A; this is an implication of Card *et al.* (2015). Since the lower bound we estimate will be large,

our conclusion will be that the elasticity is large. If there is no such persistence in desired earnings, then we would not see a kink in the employment rate; thus, we can test whether $\pi\left(\tilde{z}_{i0t}, v_{it}\right) > 0$. In principle, it would be possible to put assumptions on the structure of the measurement error, but these assumptions would be unverifiable.

Although we use lagged earnings as a proxy for desired earnings in our specific context, this need not be a feature of our method in general. As a general matter in the potential outcomes framework, we need a proxy for counterfactual desired earnings. This can be achieved in a number of ways; for example, in the context of a randomized experiment we may be able to generate this counterfactual by comparing to a control group that is unaffected by a policy, thus allowing a researcher to estimate the elasticity itself more directly (see Kline and Tartari, 2016, for such an approach using a randomized experiment to generate counterfactuals under alternative budget constraints).

## 4.3    Comparison to Other Identification Strategies

Our method involves a different set of assumptions than existing approaches for estimating extensive margin elasticities. As noted, relative to discrete choice methods or a selection correction, our method is valid non-parametrically in the context of the economic model we specify. Our method also allows transparent identification through the RKD (or RDD) approach, relying on clear graphical depiction of the results.

Relative to the difference-in-differences approach that has often been used (*e.g.* Eissa and Liebman, 1996; Meyer and Rosenbaum, 2001; Eissa and Hoynes, 2004; Gelber and Mitchell, 2012), our formal model allows us to be explicit about the population and parameter that our elasticity refers to, as in the distinction between the "observed" and "structural" elasticities discussed in Section 2.7. Our paper estimates an observed elasticity driven by individuals who face intensive margin frictions, much as previous papers have implicitly estimated observed extensive margin elasticities that could be driven in part by responses among constrained individuals. Our model highlights that extensive margin elasticities, in the presence of kinked budget sets such as the ones created by the AET or Earned Income Tax Credit, can reflect "observed" elasticities that can be influenced by the non-response on the extensive margin of those who bunch at the kink.

Using the RKD approach, one avoids the assumptions of difference-in-differences (DD). Suppose the employment probability is a function of the ANTR, which varies across groups and time due to policy changes. We can use this variation to perform a DD estimate of the effect of the ANTR (or a transformation thereof) on the employment rate (or a transformation thereof), for example using the policy variation as an instrument for the ANTR under a DD instrumental variables approach. However, to do so we must postulate some parametric relationship—whether linear, log, semi-log, *etc.*—between employment and the ANTR. Our RKD approach does not require any such parametric assumption. A DD estimator also makes other assumptions, including the "parallel trends" assumption, that are not required for our RKD. In Gelber *et al.* (2017), we are investigating the effects of the AET using a DD approach comparing those below and above the exempt amount before and after age 62, allowing estimates applying to a broader sample. This

requires the assumptions above, as well as panel data over more than two time periods: in the DD context counterfactual (*i.e.* lagged) earnings must be observed prior to both the pre-62 employment observation, as well as the post-62 employment observation. Thus, our RKD setting requires fewer time periods of data — two rather than at least three.

Our method also has potential disadvantages relative to a DD design. For example, in many cases estimating a precise RKD may require large sample sizes of the sort we have obtained in our SSA data. Furthermore, like other empirical work that exploits sharp discontinuities in incentives, our estimate is local to those near the kink. In general, it may be possible to address this to some extent—for example, in principle one could investigate whether estimates are heterogeneous when the kink is placed in different locations. In comparison, the DD approach also may identify local effects—those in the region of policy variation—that may not apply more broadly. Our RKD approach also makes the assumptions above, including the continuity of the first derivative. As Card *et al.* (2015) note, it is also possible to probe the underlying smoothness assumption by testing for kinks in predetermined variables. We are also able to address these assumptions by assessing how our results compare to those in a placebo sample that is not subject to the AET (of course, placebos may also be available in a DD context).

Ultimately, our approach offers a new and different method, with both potential advantages and disadvantages relative to other methods. At the least, our new approach both allows an explicit understanding of the interpretation of the extensive margin parameter we estimate, and enlarges the set of existing strategies for estimating extensive margin responses, thus complementing existing strategies. We view our approach as particularly well suited to settings in which it is possible to obtain precise kink estimates and convincingly verify the smoothness assumptions.

## 4.4    Additional Specification Considerations

Implementation of an RKD requires a number of choices regarding specification. For our main results, we implement the Calonico *et al.* (2014) data-driven method for bandwidth selection. We report confidence intervals corrected for bias following Calonico *et al.* (2014), and we use a triangular kernel to weight the data near the exempt amount. Card *et al.* (2015) use both linear and quadratic specifications in their analysis. Calonico *et al.* (2014) propose an RKD estimator in which the quadratic specification can be used to correct for the bias in the linear estimator, while Ganong and Jäger (2015) advocate a cubic specification. We implement linear, quadratic, and cubic versions of (18), to investigate the robustness of our results. We use the linear specification without additional controls as our baseline because it minimizes the corrected Akaike Information Criterion (AICc) and the Bayesian Information Criterion (BIC) for our main outcome (the employment rate at ages 63 to 64), as well as for nearly every other outcome.

We must also decide whether to allow a discontinuity at the exempt amount in the level of the outcome variable and whether to control for covariates (Ando, 2013). We allow for a discontinuity, although the results are virtually unchanged if we do not. We present results with and without controlling for covariates.

Thus, for each sample or outcome we can produce estimates using six regressions: the linear, quadratic, and cubic regressions; and a version of each further adding predetermined covariates using Calonico *et al.* (2016).

Finally, we focus primarily on employment outcomes at ages 63 and 64, even though the AET begins to apply as early as age 62 and remains in effect at ages 65 and older in the period we study. We make this choice due to the following considerations. The AET first applies to claimants when they reach OASI eligibility at age 62, but it does not make sense to examine the effect of the AET on whether an individual has positive earnings in the calendar year she turns age 62. The reason is that we observe calendar year earnings, and we measure "age" in a given calendar year coarsely as the highest age a person attains during that calendar year. If an individual claims OASI during the calendar year she turns age 62, the AET only applies to earnings in the calendar months after the individual claims. If the claimant earns at all during this calendar year—even during months prior to claiming OASI—then she will have positive earnings in this calendar year. Thus, a person who is induced by the AET to stop earning after claiming would appear in the data with positive earnings during this calendar year, and therefore would appear to have no measured response to the AET. As a result, it appears highly unlikely that we would observe a measurable employment response to the AET at age 62 as we measure it in the data, even if the AET has a substantial impact on employment. We expect measurable effects only to appear as early as age 63. We choose 64 as the oldest age at which to examine employment effects because age 60 earnings are a better proxy for desired earnings at ages 63 to 64 than for older ages. Moreover, at age 65 individuals with earnings near the under-NRA exempt amount are only exposed to this exempt amount—as opposed to the much higher exempt amount applying to those at NRA and above—for part of the year.

## 4.5    Mapping the Estimates to the Model

The RKD statistical model can be related directly to our theoretical model. First, the smoothness conditions imposed on preferences and heterogeneity in Section 2 are analogs to the assumptions in Card *et al.* (2015) that allow us to interpret the RKD estimate as a treatment-on-the-treated parameter. Second, the RKD coefficients map directly to the parameters of our model. The parameter $\beta$ in equation (9) corresponds to the $\beta_1$ from (18), and we calculate $\alpha$ analytically, as described above. We can then calculate the remaining elements of equation (9), $1 - a$ and $\Pr(z_{n1} > 0 \mid \tilde{z}_{n0} = z^*)$, using data on individuals who have age 60 earnings near $z^*$. Moreover, just as the RKD returns an estimate that is local to agents located at the kink, our theoretical model identifies parameters that apply to agents located at the kink.

As described in Section 4.2, we will use earnings at age 60 to proxy for desired earnings at ages 63 and 64 in the absence of the AET, positing equation (19). Thus, in terms of our model in Section 2, we require that the ability distribution remain relatively stable between ages 60 and 63 to 64. We still allow $\pi(\tilde{z}_{i0t}, v_{it}) < 1$ in Section 4.2, *i.e.* we do not require this distribution to be absolutely static across this time span. Furthermore, this assumption places no restrictions on the evolution of the fixed cost distribution, $G(q \mid n)$, over time. We observe a declining employment rate between ages 60 and 63 to 64, which can be

accommodated by a rightward shift in the distribution $G(q|n)$ at older ages. While we use age 60 earnings to proxy for desired earnings at ages 63 to 64 as above, our RKD effectively uses the employment rate at ages 63 to 64 for those with age 60 earnings just under the exempt amount to reflect state 0 employment rates, while the employment rate at ages 63 to 64 for those with age 60 earnings just over the exempt amount reflects state 1 employment rates. In our model, we require that the distribution $G(q|n)$ is the same across states 0 and 1. The empirical analogue is thus a requirement that the distribution of fixed costs $G(q|n)$ is same for those with age 60 earnings just below and just above the exempt amount.

# 5    Data

We implement our RKD estimation strategy using the restricted-access Social Security Administration Master Earnings File (MEF) linked to the Master Beneficiary Record (MBR). The data contain Federal Insurance Contributions Act (FICA) pre-tax earnings for all Social Security Numbers (SSNs) in the U.S. in each calendar year.[24] Separate information is available on self-employment and non-self-employment earnings. The data are from W-2s, mandatory forms filed with the Internal Revenue Service for each employee for whom the firm withholds taxes and/or to whom remuneration exceeds a modest threshold. Thus, we have data on earnings regardless of whether an employee files taxes. The data longitudinally follow individuals over time.

The MBR contains information on exact date of birth, exact date of death, month and calendar year of claiming OASI, race, and sex. In the calendar year after an individual dies, earnings and employment appear in the dataset as zeroes. Thus, some of an effect on employment could in principle be mediated through an effect on mortality, which would affect the interpretation—but not the validity—of our results. The effects we estimate are relevant to policy, in the sense that they reflect the overall effect on employment.

The AET is applied to pre-tax OASI benefits; this affects the results negligibly relative to measuring after-tax benefits, because OASI benefits only became taxable in 1984, and after 1984 benefits were only taxable above an income threshold much higher than the AET exempt amount. By examining pre-tax benefits, we answer a policy-relevant question: how a given change in the AET BRR would affect employment. After-tax benefits are slightly smaller than pre-tax benefits—and the benefit reduction rate associated with after-tax benefits should change by slightly less at the exempt amount—suggesting that the effect of the pre-tax BRR should if anything reflect a lower bound on the effect of the after-tax BRR, and again strengthening our case that we will estimate a lower bound on the elasticity.

Due to institutional considerations and computational constraints, we focus on a specific set of cohorts observed over certain calendar years. We focus on a sample that is subject to the AET in 1978 and after. The reason is that we observe only calendar year earnings, and the AET has effectively applied to calendar year earnings beginning in 1978 (Gelber *et al.*, 2013, Appendix A). Before 1978, the AET was applied to

---

[24] Data are also available on total W-2 earnings, as opposed to FICA earnings that are capped at the maximum taxable earnings level. We use FICA earnings because in 1978 to 1980 the data on total W-2 earnings are subject to measurement error that is not an issue in the FICA earnings data (Utendorf, 2001/2002; Kopczuk, Saez, and Song, 2010). The cap on FICA earnings at the maximum taxable earnings level does not affect our results: the AET exempt amount was far below the maximum taxable earnings level, so the FICA earnings data are capped at a level far above the top of the range included in our bandwidth.

quarterly earnings, which we do not observe in the data. This limits our focus to individuals born in 1918 or later, who turned age 60 in 1978 or later. Due to computational constraints, we were able to obtain data on individuals born from 1918 until 1923, and due to these constraints it was also necessary to draw a 25 percent random sample of this group. This leaves 3,629,183 sampled individuals; we included data on these individuals for all calendar years in the panel. We observe these individuals until they turn 64, which occurs as late as 1987 (in the 1923 birth cohort).

Next, from this initial sample, we drop individuals who claimed OASI or Disability Insurance prior to age 62—usually those claiming on a spouse's record—so that our measure of earnings prior to 62 is not affected by potential contemporaneous impacts of the AET (19.5 percent of the sample); those with missing values for gender (0.049 percent of the remaining sample); those with positive self-employment earnings at age 60 (7.82 percent of the remaining sample), as these earnings are often subject to manipulation (*e.g.* Chetty *et al.*, 2013); and those with negative measured earnings at some point between ages 50 and 70 (0.035 percent of the remaining sample). Since the key running variable is earnings at age 60, we restrict attention to individuals with positive earnings at age 60 (47.05 percent of the remaining sample). We include all remaining individuals, including those who collect benefits as retired workers, auxiliary beneficiaries, or survivors. This leaves us with a sample of 1,424,667 individuals. We use data on individuals from ages 57 to 64, or eight calendar years per person, so that the total number of individual-year observations is 11,397,336. We do not limit the sample in any given year to individuals claiming in that year, since claiming OASI is potentially endogenous to the AET. Our data extract does not link spouses to one another.[25]

Several features of the data merit discussion. First, these administrative data are subject to little measurement error. Second, earnings as measured in the dataset are not subject to manipulation through tax deductions, credits, or exemptions, and they are subject to third-party reporting (among the non-self-employed). Third, like most other administrative datasets, the data do not contain information on hours worked, hourly wage rates, amenities at individuals' jobs, underground earnings, assets, savings, or consumption. They also do not contain data on unearned income or marital status.

Table 1 shows summary statistics for our sample. The main sample shown is restricted to the 95,960 individuals who have positive age 60 earnings within $2,797 of the exempt amount and who satisfy the other sample restrictions. This bandwidth, $2,797, is the one we choose in our regressions for our main outcome using the procedure of Calonico *et al.* (2014). Our main outcome in our regressions is the mean annual employment rate among 63 to 64 year-olds, *i.e.* the percent of the corresponding calendar years when the individual has positive earnings; the mean level of this variable in the main sample is 56.53 percent.

The exempts amounts during our period, ranging from $9,787 to $11,517, are close to mean annual

---

[25] A non-convex kink in the budget set occurs where the AET has completely phased out the OASI benefit, and therefore the BRR jumps from positive to zero. We are not able to investigate the effects of this non-convex kink in a reliable way. First, we do not directly observe an individual's OASI benefits in each year in our data extract. Second, the effect of the non-convex kink should be spread throughout the earnings distribution, because the non-convex kink occurs at a different earnings level for each individual. Thus, we would not expect any sharp change in aggregate behavior at any particular average earnings level due to this non-convex kink. It is also not possible use individual-level data, because we are not able to link spouses. If a spouse is claiming on a primary beneficiary's record, then the non-convex kink occurs at the earnings level at which the family's entire benefit is completely phased out; thus, we would need information on whether a spouse is claiming on the primary beneficiary's record to observe the non-convex kink accurately.

earnings, $10,977, at ages 63 to 64 in the full sample of those with positive or zero age 60 earnings (not restricted to being close to the exempt amount); and close to median earnings, $9,166, at ages 63 to 64 among those with positive age 60 earnings (not restricted to being close to the exempt amount). Thus, our estimates apply to a group near the exempt amount that is close to this mean and median in these broader populations; this is relevant for interpreting the regression results. Mean earnings (including zeroes) at ages 63 to 64 is $5,814 in our main sample. The mean age of claiming OASI is 63.13. This sample is 67.99 percent female; the large female percentage is a consequence of focusing on individuals near the exempt amount at age 60. For comparison we also show the full sample of those with positive earnings at age 60, not restricted to those around the exempt amount but satisfying the other sample restrictions. Throughout the paper, dollar figures are expressed in real 2010 terms.

# 6    Empirical Results

## 6.1    Preliminary Analysis

We begin our analysis with initial validity checks of our empirical method, demonstrating that individuals do not appear to sort around the exempt amount when choosing their earnings at age 60. First, Figure 4 shows that the density of earnings at age 60 (solid circles) appears continuous around the exempt amount. The regressions in Table 2 confirm that the density of observations is smooth in the region of the exempt amount. Following Landais (2015), in these regressions we use the procedure of McCrary (2008) to estimate discontinuities in the level and slope of the density function at the exempt amount, allowing for linear, quadratic, or cubic functions of the running variable. To test further for sorting, we also test for bunching at the exempt amount. Figure 4 shows that the density of earnings at age 60 appears smooth near the exempt amount, and that the amount of bunching, calculated using a method similar to Chetty *et al.* (2011) — and described in detail in Appendix C — is statistically insignificant.

For comparison, Figure 4 also shows the earnings distribution at age 62 (hollow circles), when claimants are subject to the AET. At age 62 we see a markedly different pattern than at age 60, with a large, statistically significant spike in the age 62 earnings density near the exempt amount. Recalling our model in Section 2, this is evidence that some individuals do not face frictions in adjusting at the intensive margin to the exempt amount.

Similarly, Figure 5 shows that there is no clear visual discontinuity in the slope of predetermined covariates (sex, birth year, and probability of being white). In this figure and subsequent figures, the $x$-axis measures our running variable, earnings at age 60. The range of the $x$-axis is [-$3,000, $3,000], which corresponds roughly to the Calonico *et al.* (2014) bandwidth of $2,797. On the $y$-axis is the outcome in question, taking means within $500 bins. The data are pooled across all years of the sample. Table 2 confirms that in the baseline linear specification, there is no significant discontinuity in the slope of the fraction white or female, although there is a small, significant discontinuity in the slope of the year of birth. Nonetheless, there is no significant discontinuity in year of birth in the cubic specification, in which we will still find a significant

discontinuity in our main outcome, the age 63 to 64 annual employment rate. Moreover, the significance of the discontinuity in the slope year of birth does not survive a Bonferroni correction that is relevant here since we have no theoretical presumption of a change in slope. Finally, controlling for these covariates, including year of birth, will have no material effect on the results. Note that although we show our data aggregated by bin means, our regressions here and elsewhere are run using individual-level data as in (18). All of this evidence of covariate smoothness shows that agents act as if they do not anticipate the imposition of the AET at age 60, consistent with this version of our dynamic model in Appendix B.3 (and consistent with the evidence in Gelber $et\ al.$, 2013, and Gelber $et\ al.$, 2016, of no reaction to OASI, and the AET in particular, until closer to when individuals are affected by it).

To support the validity of using age 60 earnings as a proxy for age 63 to 64 desired earnings, we show that desired earnings remain stable across a "placebo" set of ages. Specifically, we show that the distribution of real earnings growth from one period to a subsequent period exhibits a spike at zero, as postulated in (19). Figure 6 shows that from age 59 to age 60—a placebo set of ages near the EEA during which our sample is not subject to the AET—a noticeable spike in real earnings growth does occur near zero percent growth, as workers' wages grow at exactly the rate of CPI in a substantial fraction of cases.

Although we focus on the case of the AET in this paper, it would also be possible to use our method to estimate extensive margin responses in other contexts. It seems natural that the assumption that $\pi\left(\tilde{z}_{i0t}, v_{it}\right) > 0$—an assumption that allows our method to be used as explained above—would apply when using lagged earnings to predict desired earnings for workers in other contexts as well. This is bolstered by Appendix Figure D.2, which shows a large spike at zero nominal earnings growth from one year to the next in the SSA data to which we have access, covering the 1918 to 1923 cohorts over calendar years 1951 to 1984 and encompassing ages 28 to 59. Thus, it should be possible to use our method, implemented in this case through using lagged earnings to proxy for desired earnings, when studying extensive margin responses to other policies applying in other age ranges.

## 6.2 Main Results Documenting a Kink in Employment

Having passed these validity tests, we proceed to our main results. In Figure 7 we see a sharp decrease at the exempt amount in the slope of the annual employment rate at ages 63 to 64: under the exempt amount the slope appears positive and steep, whereas above the exempt amount the slope appears to change immediately to nearly flat. Appendix Figure D.3 shows that this visual pattern—a sharp decrease in slope at the exempt amount—is also clear with a wider bandwidth of $6,000. Table 3, Column 1, reports the estimated coefficient $\beta_1$ from (18), confirming that the change in the slope is statistically significant and substantial in the baseline linear specification that uses a Calonico $et\ al.$ (2014) bandwidth of $2,797.

Even without estimating an elasticity, these results constitute a novel source of evidence that the AET causes an extensive margin effect.[26] In the context of our model in Section 2, evidence of a kink in the

---

[26]Similarly, it would be possible to investigate whether the AET causes movement to other specific earnings levels. For example, for those constrained at the intensive margin the AET could cause movement to part-time work with earnings below the exempt amount. However, the level of earnings associated with this part-time work could be different for different individuals and is likely

employment rate also implies that some individuals face frictions in adjusting earnings on the intensive margin. Thus, these results support the assumptions necessary to estimate the elasticity.

Our estimates represent the effects of the AET while holding other factors constant. We do not interpret the discontinuity in the slope of earnings at the exempt amount as reflecting changes in demand by firms, since such demand changes should have been materially similar on either side of the exempt amount as should any general equilibrium effects of the policy change more broadly. Like other papers based on local variation, including others in the AET literature, our identification strategy does not attempt to address general equilibrium impacts of the AET. We interpret our measured responses as potentially including the effects of misperception of substitution incentives or other frictions that could affect adjustment.

## 6.3    Robustness and Validity Checks

These basic results pass several robustness and validity checks. The estimates remain significant in Table 3 under the quadratic and cubic specifications, with moderately larger estimates. The results are also robust to controls for demographics in Column 2 (dummies for sex, year of birth, and race groups). In Figure 8 we demonstrate that the size of the estimated kinks remains relatively stable across different bandwidths.

As a "placebo" test in the period prior to being subject to the AET, Figure 9, Panel A shows no clear visible change in the slope of employment at ages 56 and 57, as a function of age 60 earnings. Like ages 63 and 64, ages 57 and 56 are respectively three and four years away from age 60, the age at which the running variable is measured. In Table 3, Column 4 we present the estimates from a corresponding regression in which the dependent variable is the annual employment rate at ages 56 to 57. We maintain the same running variable as earlier, *i.e.* age 60 earnings, since we would like to test for a spurious relationship between earnings at age 60 and lagged employment. Across all three polynomial specifications, mean annual employment at ages 56 to 57 shows an insignificant change in slope.[27]

Furthermore, in Figure 9, Panel B and Table 4 we show that the kink arises precisely at ages 63 and 64—not earlier—which we consider our single most convincing evidence that we have uncovered a true effect of the AET. Figure 9, Panel B shows the employment probability by single year of age from 61 to 64. There is no clear kink at ages 61 and 62, and a visible kink only arises at ages 63 and 64. Recall that no kink should appear at age 62 because the AET only applies for part of the year. Table 4 confirms that the kink estimates are insignificant at ages 61 and 62, but they are substantial and highly significant at ages 63 and 64. We are particularly reassured by the small and insignificant kink at ages 61 and 62. This placebo check provides further evidence that nonlinearities in the probability of employment (conditional on earnings at

---

spread relatively smoothly throughout the population, making it difficult to perform a formal test of this hypothesis given that SSA only has data on earnings, not hours worked. A dataset with information on hours worked could allow a better test of this hypothesis, if part-time work tends to require common numbers of hours (*e.g.* 20 hours per week).

[27] As a "placebo" exercise, with ideal data it would be possible to examine whether individuals show a kink in the employment rate as a function of age 60 earnings even after the AET has been eliminated for ages NRA and above after 2000. However, this is difficult to implement for three reasons. First, by the year 2000 when the AET was eliminated for those NRA and above, the birth cohorts in our sample are very old (age 77 or older) and have very low employment rates. Second, even with the necessary data, age 60 earnings may be a poor proxy for desired earnings at ages NRA and above. Third, both before and after 2000, individuals face the AET when they claim before NRA. If extensive margin responses have persistence, then we might expect a persistent kink even after NRA. At ages 65 and over in our sample, individuals face a higher exempt amount, and age 60 earnings represents a weaker proxy for desired earnings, so it is less straightforward to interpret the results. However, we do see some evidence of a persistent response (available upon request).

age 60) do not drive our kink estimates. The absolute value of the kink grows substantially from age 63 to age 64, from 1.60 to 1.98, only a modest fraction of which can be explained by the higher fraction claiming at 64 (79.5 percent) than at 63 (74.5 percent). This is consistent with gradual adjustment to the AET, which Gelber *et al.* (2013) document at the intensive margin.

As another assessment of the validity of our approach, we conduct a permutation test in the spirit of Ganong and Jäger (2015). In particular, we estimate a set of placebo changes in slope in the age 63 to 64 employment rate, using the same specification as our main estimates except that we examine the change in slope at placebo locations of the exempt amount away from the true exempt amount. In Figure 10, we show that the point estimate recovered from the actual exempt amount is located well below the distribution of placebo estimates, reinforcing the view that we are detecting a true effect of the AET. This permutation test shows a significant kink ($p = 0.025$).[28] Following Landais (2015), we show in Figure 11 that the R-squared of the regression in the baseline linear specification is maximized at the actual location of the kink, rather than at placebo kink locations, again consistent with our finding a true effect of the AET.[29]

## 6.4 Estimating an Extensive Margin Elasticity

Our model allows us to go beyond documenting the effects of the AET by estimating an elasticity using equation (9). In this equation we require an estimate of the average marginal tax rate; we use a marginal tax rate incorporating the AET BRR as well as the average federal and state income and payroll marginal tax rates, calculated using the TAXSIM calculator of the National Bureau of Economic Research and information on individuals within the bandwidth distance from the kink in the Statistics of Income data in the years we examine. Table 5 shows that the elasticity estimated using the full sample in the baseline linear specification is 0.49. As discussed above, we interpret this as a lower bound on the elasticity. Since the lower bound is large, our main conclusion is that the elasticity is large.

Separating the regressions by demographic group, the elasticity for women is 0.49, and the elasticity for men is 0.25, consistent with the typical stylized fact that women have larger employment elasticities than men.[30] Whites and non-whites show similar employment elasticities, although the estimate is much more precise in the larger white population. It is also of interest how the elasticity may vary depending on an individual's mean prior earnings, which we measure from ages 40 to 59. The elasticity point estimates are similar among those with mean prior earnings that are above the median *vs.* below the median, although the elasticity estimate is more precise in the below-median group.

A variety of considerations indicate that these estimated elasticities are robust; that if anything they

---

[28] Relative to our baseline specification, we also find similar, statistically significant results when we run our regressions using data at the $50 bin mean level (or within bins of other sizes). This is an alternative way of demonstrating that the results are significant and robust to filtering out within-bin correlations; however, these results are superseded by the permutation test results as the latter demonstrate that the kink is statistically unusual relative to the distribution of placebo estimates.

[29] Recall that our running variable measures earnings relative to the exempt amount applying to OASI claimants when an individual is age 60. Since the exempt amount is rising modestly over time (Figure D.1), the exempt amount individuals face rises slightly from age 60 to ages 63 and 64 when we measure employment. However, it rises by only $75.77 on average, far smaller than the bin size of $500 that we use in our figures. Thus, our statistical and graphical results are similar to the baseline when we instead measure our running variable as age 60 earnings relative to the exempt amount that later applies to the same individual at ages 63 to 64.

[30] Our study of women as well as men differs from the pre-2003 empirical literature on the employment effects of the AET cited above, as well as Baker and Benjamin (1999) and Haider and Loughran (2008), all of which focus on studying only men. Gruber and Orszag (2003) find suggestive evidence of an impact among women but not among men.

reflect lower bounds; and that these large elasticities make sense given that we are studying older workers, who have marginal attachment to employment in many cases. Appendix Table D.1 shows that the elasticity is modestly larger in the quadratic or cubic specifications, reaching 0.66 in the quadratic and 0.82 in the cubic. The results are similar with or without controls. In the baseline first stage, we calculated the exact change in slope of the ANTR, under the assumption that desired earnings are fixed at their age 60 level; to show that the results are robust to instead running a first stage regression analogous to the reduced form regression, in Appendix Table D.2 we show that the elasticities are similar when we assume that desired earnings are fixed at their age 60 level but estimate the change in the slope of the ANTR using a regression whose bandwidth is chosen using Calonico *et al.* (2014).

The elasticity calculations are mechanically affected by how we specify the first stage. Given that the stringency of the AET, and therefore the effective ANTR, is potentially affected by benefit enhancement, we consider our elasticity estimates to be illustrative, as the scaling of the first stage could be affected by accounting for benefit enhancement. However, it is also important to note that if we took benefit enhancement into account, the first stage absolute value of the change in the ANTR at the exempt amount would be smaller, and thus the estimated elasticity would be larger. As a result, if anything we would estimate larger elasticities, strengthening our conclusion that the lower bound on the elasticity is large. Moreover, the kink estimates in Table 3 and Table 4 are unaffected by the scaling, and our counterfactual estimates of the effect of policy changes on employment would be unaffected by the scaling of the first stage. Finally, if individuals (mis-)perceive the AET as a pure tax (without taking benefit enhancement into account), then we have specified the first stage in a way that accurately captures their perceptions.

We noted earlier that the first stage also could be affected by error in using lagged earnings to measure desired earnings. As an illustration of the magnitude of this factor, the signal-to-noise ratio is 0.61 based on the coefficient from a regression of an individual's earnings at age 61 on the individual's earnings at age 58 and a constant term, for those with age 58 earnings within \$3,000 of the exempt amount. From ages 50 to 58, the signal-to-noise ratio from age $t$ to $t+3$ varies in a similar range, from 0.57 ($t$=56) to 0.63 ($t$=51). This suggests that a moderate degree of upward adjustment to our elasticity estimates would be entailed based on the signal-to-noise ratio, again strengthening our case that we have estimated a large lower bound on the elasticity.

The relatively large elasticities we estimate are within the range of elasticities estimated using microdata for groups with low employment attachment, though they fall outside the range of some previous surveys of extensive margin elasticities for the full population estimated using microdata. Chetty *et al.* (2012) report Hicksian and Frisch (*i.e.* constant-marginal-utility-of-wealth) extensive margin elasticities of 0.25, in a meta-analytic synthesis of microdata estimates not specific to particular population groups.[31] Saez (2002) discusses that the Negative Income Tax (NIT) experiments find small extensive margin elasticities for males of about 0.2, but find participation elasticities for those groups less attached to the labor force—*e.g.* second

---

[31] In a survey of macroeconomic literature, Chetty *et al.* (2012) find average Hicksian elasticities of 0.17 and average Frisch elasticities of 2.77.

earners, single household heads, and the young—between 0.5 and 1. Our setting studying older workers at or near retirement age is most comparable to estimates for groups with low employment attachment, and therefore our estimates are comparable to previous results for such groups. Indeed, a small group of theoretical and empirical studies has found relatively large labor supply or earnings responses among the elderly in particular (French, 2005; Laitner and Silverman, 2012; Alpert and Powell, 2014), but none using our non-parametric method based on discontinuities (and applied to administrative data).

## 6.5 Accounting for Claiming Behavior

Among those with age 60 earnings within our baseline bandwidth relative to the exempt amount, 75 and 80 percent of the sample has claimed by ages 63 and 64, respectively. The fact that some have not claimed OASI affects the interpretation of the elasticity we estimate. In particular, the elasticity we have estimated so far should be interpreted as the elasticity *inclusive* of the impact of the fact that some individuals have not yet claimed. We estimate this elasticity in the baseline because this is policy-relevant, in the sense that it informs us how a given change in policy would affect employment in this group, given the effects of the percentage of our sample that claims.

At the same time, an alternative elasticity concept is also of interest: the "conditional" elasticity among those in the sample claiming OASI (*i.e.* we "condition" on claiming OASI). In estimating this conditional elasticity, the smaller the fraction that claims by 63 or 64, the smaller the implied first stage will be, and thus the larger the conditional elasticity estimate will be. Since we are estimating a lower bound on the unconditional elasticity above, perhaps the simplest way to address estimation of the conditional elasticity—given endogenous claiming behavior—is to note that if anything the estimated conditional elasticity would be still larger than the unconditional elasticity, if the first stage were attenuated due to the fact that some individuals have not claimed by 63 or 64.

Beyond noting this, we can divide the unconditional elasticity by the share claiming, as shown in Appendix B.4, to calculate a conditional elasticity. Here we can use the claiming behavior of those with $z_i^{60}$ just below the kink to impute the percentage claiming among those at the kink, *i.e.* those to whom the elasticity estimates apply. Using the percent claiming at ages 63 and 64, this method implies that to estimate the conditional elasticity, we would inflate the elasticity estimates by 29.9 percent relative to those shown in Table 5.[32] Thus, for example, the baseline elasticity would be 0.63 rather than 0.49. Appendix Table D.3 shows the elasticities and standard errors calculated in this way.

In principle it would also be possible to adjust the conditional elasticity estimates further to account for the fact that the fraction that claims may be directly affected by the AET itself, as in Appendix B.4. However, we show in the Appendix that if the AET has no effect on the fraction claiming, then no further adjustment is necessary to estimate a lower bound on the conditional elasticity. Appendix Figure D.4 shows that the probability of claiming shows no clear change in slope at the exempt amount, and corresponding regressions show no robust effect of the AET on claiming at these ages: a placebo test parallel to our previous

---

[32] We calculate 29.9 percent as $100 * (1/[(0.795 + 0.745)/2] - 1)$.

test in the spirit of Ganong and Jäger (2015) shows $p = 0.15$ for the two-sided test of equality of the coefficient with zero.[33] Because there is no evidence for an effect on claiming, there is little case for implementing our most flexible adjustment for endogenous claiming in Appendix B.4.3, though this bounding technique is in principle more applicable in other OASI settings that may feature a clear effect on the claiming probability.

## 6.6   Counterfactual Simulations

We can use our elasticity estimates to understand the effect of changing the parameters of the AET, under the assumption that our extensive margin elasticity estimates apply throughout the earnings range shown in our graphs, from age 60 earnings $3,000 under the exempt amount to $3,000 over. One key issue for policy-makers is the impact of eliminating the AET (Tergesen, 2016). We simulate the increase in the employment rate among 63 and 64 year-olds that would result from eliminating the AET entirely, using our baseline elasticity estimate of 0.49 in Table 5. As before we assume that at ages 63 and 64, desired earnings conditional on employment in the absence of the AET is given by earnings at age 60. To calculate the baseline probability of participation at ages 63 to 64 at different earnings levels, we bin the earnings distribution into $500 bins as in our figures and calculate this probability, $\Pr(PE_{0i})$. For each age 60 earnings bin, we calculate the mean percent change in the average net-of-tax rate associated with moving from the AET's current 50 percent BRR to the ANTR associated no AET (but keeping other taxes in place), $(ANTR_{1i} - ANTR_{0i})/ANTR_{0i}$. We then use our estimated elasticity $\hat{\eta}$ to calculate the predicted change in the employment rate if the AET were eliminated, $\hat{\eta}[(ANTR_{1i} - ANTR_{0i})/ANTR_{0i}]\Pr(PE_{0i})$, and we aggregate across bins (weighting by the fraction of the population in each bin).

This calculation shows that eliminating the AET would increase the employment rate by 1.4 percentage points in the group with age 60 earnings within $3,000 of the exempt amount, or a 2.5 percent increase. We consider this a lower bound for the several reasons described earlier. This counterfactual exercise illustrates that the observed elasticity can be useful for counterfactual predictions about employment levels under alternative AET parameters.

We can also use our elasticity estimates to calculate the change in earnings due to the AET, as well as the fraction of the change that is due to extensive margin responses. We use our baseline extensive margin elasticity estimate of 0.49. To calculate the intensive margin response, we assume quasilinearity as in much recent literature on earnings responses to kinked budget sets (*e.g.* Saez, 2010; Chetty *et al.*, 2011; Kleven and Waseem, 2013; Kleven *et al.* 2014). For simplicity we assume that the intensive margin elasticity with respect to the marginal net-of-tax share is a constant 0.35, and that the intensive margin adjustment cost is $278, corresponding to the Gelber *et al.* (2013) baseline point estimates. We calculate the increase in earnings associated with the elimination of the AET through each channel, intensive and extensive. For an

---

[33] Our evidence is compatible with Gruber and Orszag's (2003) evidence on claiming: our results pertain to 63-64 year olds, whereas Gruber and Orszag (2003) find an effect of the AET on claiming among older groups.

individual $i$, the estimated total earnings response $\Delta E_i$ is:

$$\Delta E_i = \Pr(PE_{no,i})z_{no,i} - \Pr(PE_{AET,i})z_{AET,i}$$
$$= \Pr(PE_{AET,i})(z_{no,i} - z_{AET,i}) + [\Pr(PE_{no,i}) - \Pr(PE_{AET,i})]z_{no,i} \qquad (20)$$

where $\Pr(PE_{AET})$ and $z_{AET}$ are the probability of positive earnings under the AET and earnings under the AET, respectively, and $\Pr(PE_{no})$ and $z_{no}$ refer to the levels of these variables once the AET has been eliminated. The second line shows that the total response can be decomposed into an intensive margin response (the first term) and an intensive margin response (the second term).

Our results show that in the group within \$3,000 of the exempt amount, mean annual earnings at ages 63-64 (including zeroes) increase by 9.8 percent (\$579) due to eliminating the AET, and 27.6 percent of the increase in earnings is due to the extensive margin effect on the employment rate. Focusing only on the intensive margin impact would overlook over one quarter of the total earnings response.

Our specific predictions are influenced by multiple factors: the elasticity is sizeable; the change in the BRR due to eliminating the AET is large; but for those near the AET exempt amount that we study, the change in the ANTR is relatively small. The elasticity could also be different at other points in the earnings distribution, and moreover some individuals may tend to be unresponsive (*e.g.* those with zero earnings at age 60) while others might be more responsive (*e.g.* those with age 60 earnings farther above the exempt amount who experience a larger change in the ANTR, but not so far that the effect of the AET on the ANTR greatly diminishes). Ultimately, our calculations do demonstrate that eliminating the AET can cause large increases in earnings and employment, even using our lower bound point estimates.

## 7 Conclusion

We develop a method for estimating the extensive margin effect of budget sets kinks or notches, relying on empirically estimable quantities and a transparent empirical design based on predictions from an economic model that does not require parametric or distributional assumptions. Using this method applied in the case of a kink, we show that the AET has a large impact on extensive margin earnings decisions. Under our preferred specification, the point estimate shows that the elasticity of the employment rate with respect to the ANTR is at least 0.49. We interpret this as a lower bound on the elasticity for several reasons. Estimating this large lower bound is useful because it clearly distinguishes our study from previous literature that largely finds the AET has little effect on employment.

This large response suggests that the AET can have important effects on the employment of older workers. Our point estimates would imply that eliminating the AET would increase the employment rate among affected older workers by 1.4 percentage points (2.5 percent) in the group we study, and would increase earnings by 9.8 percent. 27.6 percent of the aggregate earnings increase due to eliminating the AET is associated with extensive margin decisions, demonstrating that focusing on intensive margin responses, as in Saez (2010) and much other literature, can abstract from an important component of the earnings impacts

of tax and transfer programs. These estimates also suggest that the impact of eliminating the AET on the budget is more positive than the budget-neutral estimates of the Congressional Budget Office (2000), which abstracted from extensive margin effects anywhere (including in the group around the kink we study).

Estimates of the full response to taxation must also capture responses that could move an individual from above to below a convex kink, including moving to zero earnings. Indeed, our empirical approach could be easily extended to estimate the effect of a kink on the probability of locating at any location or set of locations below the kink (not only at zero earnings but also at positive earnings levels below the kink), perhaps due to frictions like a limited earnings menu. If anything this would strengthen our conclusion that the earnings losses due to the AET are large.

Our results are consistent with the hypothesis that some individuals are constrained from adjusting on the intensive margin. The presence of frictions is consistent with recent literature including Chetty (2012), Kleven and Waseem (2013), or Gelber *et al.* (2013). It is also consistent with Abraham and Houseman's (2005) findings from a self-reported survey that suggest frictions among older workers in particular, and Baker and Benjamin's (1999) findings consistent with frictions in Canada. Our model points out that as a result of these frictions, the observed extensive margin elasticity in our setting need not reflect the structural elasticity, but the observed elasticity generally provides a lower bound.

This work could usefully be extended in a number of directions. First, in this paper we apply our method in the case of the AET kink. Future work could fruitfully find an empirical application of our method in the case of a budget set notch.

Second, in the context of the earnings decision, the estimation method requires: (1) a kinked or notched budget set; (2) information on earnings or labor supply; and (3) a proxy for desired earnings on a linear budget set. Thus, it appears that our method is applicable in many other contexts in which individuals make extensive margin employment decisions, even outside of the AET, particularly because lagged earnings can proxy for current desired earnings in a broad range of ages as we show. Although in our context we find large effects of the marginal incentive to earn, our results do not imply that such effects are as large in all settings as the literature suggests older individuals' labor supply may be particularly elastic. It would be fruitful to use our method to estimate employment elasticities in other contexts.

Third, Eissa *et al.* (2008) show that knowledge of both extensive and intensive margin elasticities are necessary to calculate the deadweight cost of tax wedges. Thus, the extensive margin estimates in this paper could be used to inform such calculations. However, to apply our results in this setting more directly, it would be necessary to add frictions of the sort we have modeled to the Eissa *et al.* (2008) framework. As this paper focuses on methodology and estimation, we leave this worthy exercise for future work.

Fourth, future work could also attempt to investigate how to adjust elasticity estimates for error in measuring desired earnings, whether through a context with a randomized experiment or through assumptions that put additional structure on the measurement error. Fifth, future work could also further examine responses to kinks when individuals mis-perceive their incentives. Sixth, we are also currently investigating

the mortality effect of the AET using a similar strategy.

Finally, although we have modeled employment as a key first context, our method could be used to estimate the elasticity of other extensive margin consumption decisions in the presence of non-linear pricing: the slope of the extensive margin consumption decision should track the slope of the average price, inducing an estimable kink or discontinuity in demand. Since non-linear pricing is found in many economic applications, our method should have considerable scope to be adapted in other economic contexts.

# References

**Aaronson, S., Cajner, T., Fallick, B., Galbis-Reig, F., Smith, C., and Wascher, W.** "Labor Force Participation: Recent Developments and Future Prospects," *Brookings Papers on Economic Activity*, 2014(2), 197-275.

**Abraham, K., and Houseman, S.** (2005). "Work and Retirement Plans among Older Americans." In *Reinventing the Retirement Paradigm*, R. Clark and O. Mitchell, eds. Oxford, UK; New York: Oxford University Press, 70-91.

**Abraham, K., and Kearney, M.** (2017). "The Factors Behind the Decline in U.S. Employment Rates: A Review of the Evidence." University of Maryland Working Paper.

**Alpert, A., and Powell, D.** (2014). "Estimating Intensive and Extensive Tax Responsiveness: Do Older Workers Respond to Income Taxes." RAND Working Paper.

**Altonji, J. G. and Paxson, C. H.** (1988). "Labor Supply Preferences, Hours Constraints, and Hours-Wage Trade-Offs," *Journal of Labor Economics*, 6, 254-276.

**Brown, S.K, and Perron, R**. (2011). "Assessing Current and Future Beneficiaries' Knowledge of Social Security Benefits," American Association of Retired Persons research report.

**Ando, M.** (Forthcoming). "How Much Should We Trust Regression-Kink-Design Estimates?" *Empirical Economics*.

**Baker, M., and Benjamin, D.** (1999). "How do retirement tests affect the labor supply of older men?" *Journal of Public Economics*, 71, 27-51.

**Battistin, E., and Chesher, A.** (2014). "Treatment Effect Estimation with Covariate Measurement Error." *Journal of Econometrics*, 178(2), 707-715.

**Becker, G.** (1976). "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology." *Journal of Economic Literature*, 14(3), 817-826.

**Bernheim, B. D., Fradkin, A., and Popov, I.** (2015). "The Welfare Economics of Default Options in 401(k) Plans," *American Economic Review,* 105(9), 2798-2837.

**Bitler, M. P.**, **Gelbach, J. B., and Hoynes, H. W.** (2006), "What Mean Impacts Miss: Distributional Effects Of Welfare Reform Experiments," *American Economic Review*, 96(4), 988-1012.

**Blinder, A. S., Gordon, R. H., and Wise, D. E.** (1980). "Reconsidering the Work Disincentive Effects of Social Security." *National Tax Journal*, 33(4), 431-442.

**Blomquist, S., Kumar, A., Liang, C.-Y., and Newey, N.** (2015). "Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income." Uppsala University Working Paper.

**Blundell, R., French, E. and Tetlow, G.** (2016). "Retirement Incentives and Labor Supply." In *Handbook of the Economics of Population Aging*, John Piggott and Alan Woodland, eds., 457-566. Amsterdam: Elsevier.

**Blundell, R., and MaCurdy, T.** (1999). "Labor Supply: A Review of Alternative Approaches." In Ashenfelter, O. and Card, D., eds., *Handbook of Labor Economics* 3A, 1560-1695. Amsterdam: Elsevier.

**Blundell, R. and Shephard, A.** (2012). "Employment, Hours of Work and the Optimal Taxation of Low-Income Families." *Review of Economic Studies*, 79, 481-510.

**Brown, J., Kapteyn, A., Mitchell, O., and Mattox, T.** (2013). "Framing the Social Security Earnings Test." Wharton Pension Research Council Working Paper.

**Burkhauser, R. V., and Turner, J. A.** (1978). "A Time Series Analysis on Social Security and the Market Work of Men at Younger Ages." *Journal of Political Economy*, 86(4), 701-715.

**Burtless, G., and Hausman, J. A.** (1978). "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment."*Journal of Political Economy*, 86(6), 1103-1130.

**Burtless, G., and Moffitt, R. A.** (1985). "The Joint Choice of Retirement Age and Postretirement Hours of Work." *Journal of Labor Economics*, 3, 209-236.

**Calonico, S., Cattaneo, M., and Titiunik, R.** (2014). "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82(6), 2295-2326.

**Calonico, S., Cattaneo, M., Farrell, M. and Titiunik, R.** (2016). "Regression Discontinuity Designs Using Covariates." University of Michigan Working Paper.

**Card, D., Lee, D., Pei, Z., and Weber, A.** (2015). "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica*, 83(6), 2463-2483.

**Card, D., Johnston, A., Leung, P., Mas, A., and Pei, Z.** (2016). "The Effect of Unemployment Benefits on the Duration of Unemployment Insurance Receipt: New Evidence from a Regression Kink Design in Missouri, 2003-2013." *American Economic Review Papers and Proceedings*, 105(5), 126-130.

**Chetty, R.** (2012). "Bounds on Elasticities with Optimization Frictions." *Econometrica*, 80(3), 969-1018.

**Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L.** (2011). "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics*, 126, 749-804.

**Chetty, R., Friedman, J. N., and Saez, E.** (2013). "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review*, 103(7), 2683-2721.

**Chetty, R., Guren, A., Manoli, D., and Weber, A.** (2012). "Does Indivisible Labor Explain the Difference Between Mciro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities." In Acemoglu, D., Parker, J., and Woodford, M., eds., *NBER Macroeconomics Annual 2012* (Vol. 27, pp. 1-56). Chicago: University of Chicago Press.

**Chetty, R., Looney, A. and Kroft, K.** (2007), "Salience and Taxation: Theory and Evidence." NBER Working Paper No. 13330.

**Christensen, K.** (1990). "Bridges Over Troubled Water: How Older Workers View the Labor Market," in P. Doeringer, ed., *Bridges to Retirement: Trends in the Labor Market for Older Workers* (pp. 175-207). Ithaca, NY: ILR Press, Cornell University.

**Cogan, J.** (1981). "Fixed Costs and Labor Supply." *Econometrica*, 49, 945-963.

**Congressional Budget Office**. (2000). *Pay-As-You-Go Estimate: H.R. 5, Senior Citizens' Freedom to Work Act of 2000.* Washington, D.C.: U.S. Government Printing Office.

**Diamond, P.** (1980). "Income Taxation with Fixed Hours of Work." *Journal of Public Economics*, 13(1), 101-110.

**Diamond, P., and Gruber, J.** (1999). "Social Security and Retirement in the United States." In Gruber, J. and Wise, D., eds., *Social Security and Retirement around the World* (pp. 437-473). Chicago: University of Chicago Press.

**Dickens, W., and Lundberg, S.** (1993). "Labor Supply with Hours Restrictions." *International Economic Review*, 34(1), 169-192.

**Disney, R., and Smith, S.** (2002). "The labour supply effect of the abolition of the earnings rule for older workers in the United Kingdom." *Economic Journal*, 112, C136-C152.

**Eissa, N., and Hoynes, H. W.** (2004). "Taxes and the Labor Market Participation of Married Couples: the Earned Income Tax Credit." *Journal of Public Economics*, 88(9-10), 1931-1958.

**Eissa, N., Kleven, H., and Kreiner, C.** (2008). "Evaluation of four tax reforms in the United States: Labor supply and welfare effects for single mothers." *Journal of Public Economics*, 92(3-4), 795-816.

**Eissa, N., and Liebman, J.** (1996). "Labor Supply Response to the Earned Income Tax Credit." *Quarterly Journal of Economics*, 111(2), 605-637.

**Engelhardt, G., and Kumar, A.** (2014). "Taxes and the Labor Suplly of Older Americans: Recent Evidence from the Social Security Earnings Test." *National Tax Journal*, 67(2), 443-458.

**Figinski, T., and Neumark, D.** (2015). "Does Eliminating the Earnings Test Increase the Incidence of Low Income Among Older Women?" NBER Working Paper No. 21601.

**French, E.** (2005). "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies*, 72, 395-427.

**Friedberg, L.** (1998). "The Social Security Earnings Test and Labor Supply of Older Men." In Poterba, J. M., ed., *Tax Policy and the Economy* (pp. 121-150). Chicago: University of Chicago Press.

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

**Friedberg, L., and Webb, A.** (2009). "New Evidence on the Labor Supply Effects of the Social Security Earnings Test." In Brown, J.R., and Poterba, J. M., eds., *Tax Policy and the Economy* (pp. 1-35). Chicago: University of Chicago Press.

**Ganong, P., and Jäger, S.** (2015). "A Permutation Test for the Regression Kink Design." Harvard University Working Paper.

**Gelber, A., and Mitchell, J.** (2012). "Taxes and Time Allocation: Evidence from Single Women and Men." *Review of Economic Studies*, 79(3), 863-897.

**Gelber, A., Isen, A., and Song, J.** (2016). "The Effect of Pension Income on Elderly Earnings: Evidence from Social Security and Full Population Data." UC Berkeley Working Paper.

**Gelber, A., Jones, D., and Sacks, D.** (2013). "Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test." NBER Working Paper No. 19491.

**Gelber, A., Jones, D., Sacks, D., and Song, J.** (2017). "The Labor Force Participation Effects of the Social Security Earnings Test." Indiana University Working Paper.

**Gelber, A., Moore, T., and Strand, A.** (Forthcoming). "The Effect of Disability Insurance Payments on Beneficiaries' Earnings." *American Economic Journal: Economic Policy*.

**Gruber, J., and Orzag, P.** (2003). "Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt." *National Tax Journal*, 56, 755-773.

**Gruber, J., and Wise, D.** (1999). "Introduction and Summary." In Gruber, J., and Wise, D., eds., *Social Security and Retirement Around the World* (pp. 1-35). Chicago: University of Chicago Press.

**Gustman, A., and Steinmeier, T.** (1985). "The 1983 Social Security Reforms and Labor Supply Adjustment of Older Individuals in the Long Run." *Journal of Labor Economics*, 3(2), 237-253.

**Gustman, A., and Steinmeier, T.** (1991). "Changing the Social Security Rules for Work after 65." *Industrial Labor and Relations Review*, 44(4), 733-745.

**Haider, S., and Loughran, D.** (2008). "The Effect of the Social Security Earnings Test on Male Labor Supply: New Evidence from Survey and Administrative Data." *Journal of Human Resources*, 43(1), 57-87.

**Hausman, J.** (1981). "Labor Supply." In Aaron, H. J. and Pechman, J. A., eds., *How Taxes Affect Economic Behavior* (pp. 27-71). Washington, DC: Brookings Institution.

**Heckman, J.** (1979). "Sample Selection Bias as a Specification Error." *Econometrica*, 47(1), 153-161.

**Honig, M., and Reimers, C.** (1989). "Is it Worth Eliminating the Retirement Test?" *American Economic Review*, 79(2), 103-107.

**Hoynes, H. W.** (1996). "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation Under AFDC-UP." *Econometrica*, 64, 295-332.

**Ito, K.** (2014). "Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing." *American Economic Review*, 104(2), 537-563.

**J. K. Lasser Institute** (1998). *Your Income Tax: 1998*. New York: Macmillan General Reference.

**Keane, M. P.** (2011). "Labor Supply and Taxes: A Survey." *Journal of Economic Literature*, 49(4), 961-1075.

**Keane, M. P. and Moffitt, R.** (1998). "A Structural Model of Multiple Welfare Program Participation and Labor Supply." *International Economic Review*, 39, 553-589.

**Kleven, H.** (2016). "Bunching." *Annual Review of Economics*, 8, 435-464.

**Kleven, H., and Kreiner, C.** (2005). "Labor Supply Behavior and the Design of Tax and Transfer Policy." *Danish Journal of Economics*, 143, 321-358.

**Kleven, H., Landais, C., Saez, E., and Schultz, E.** (2014). "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark." *Quarterly Journal of Economics*, 129(1), 333-378.

**Kleven, H. J., and Waseem, M.** (2013). "Using Notches to Uncover Optimization Frcitions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics*, 128(2), 669-723.

**Kline, P., and Tartari, M.** (2016). "Bounding Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach." *American Economic Review*, 106, 972-1014.

**Kopczuk, W., and Munroe, D.** (2015). "Mansion Tax: The effect of Transfer Taxes on the Residential Real Estate Market." *American Economic Journal: Economic Policy*, 7(2), 214-257.

**Kopczuk, W., Saez, E., and Song, J.** (2010). "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937." *Quarterly Journal of Economics*, 125(1), 91-128.

**Laitner, J., and Silverman, D.** (2012). "Consumption, Retirement, and Social Security: Evaluating the Efficiency of Reform that Encourages Longer Careers." *Journal of Public Economics*, 96, 615-634.

**Landais, C.** (2015). "Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design." *American Economic Journal: Economic Policy*, 7(4), 243-278.

**Lee, D. S., and T. Lemieux** (2010). "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48, 281-355.

**Leonesio, M. V.** (1990). "The Effects of the Social Security Earnings Test on the Labor-Market Activity of Older Americans: A Review of Empirical Evidence." *Social Security Bulletin*, 53(5), 2-21.

**Liebman, J., and Luttmer, E.** (2015). "Would People Behave Differently If They Better Understood Social Security? Evidence From a Field Experiment." *American Economic Journal: Economic Policy*, 7(1), 275-99.

**Manoli, D., and Weber, A.** (2016). "The Effects of the Early Retirement Age on Retirement Decisions." NBER Working Paper 22561.

**Manoli, D., and Weber, A.** (Forthcoming). "Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions." *American Economic Journal: Economic Policy.*

**Manski, C.** (2014). "Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy," *Quantitative Economics*, 5, 145-174.

**McCrary, J.** (2008). "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142(2), 698-714.

**McDonald, J. F., and Moffitt, R. A.** (1980). "The Uses of Tobit Analysis." *Review of Economics and Statistics*, 62(2), 318-321.

**Meyer, B. D., and Rosenbaum, D. T.** (2001). "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics*, 116(3), 1063-1114.

**Moffitt, R.** (1983). "An Economic Model of Welfare Stigma." *American Economic Review*, 73(5), 1023-1035.

**Nielsen, H. S., Sorensen, T., and Taber, C.** (2010). "Estimating the Effect of Student Aid on College Enrollment from a Government Grant Policy Reform." *American Economic Journal: Economic Policy*, 2(2), 185-215.

**Figinski, T. and Neumark, D.** (2018). "Does eliminating the Earnings Test increase the incidence of low income among older women?" *Research on aging*, 40(1), 27-53.

**Powell, J. L.** (1994). "Estimation of Semiparametric Models." In Engle, R. F., and McFadden, D., eds., *Handbook of Econometrics* (pp. 2443-2521). Amsterdam: Elsevier.

**Rubin, D. B.** (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66.5, 688-701.

**Saez, E.** (2002). "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses." *Quarterly Journal of Economics*, 117(3), 1039-1073.

**Saez, E.** (2010). "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2, 180-212.

**Saez, E., Slemrod, J. and Giertz, S.** (2012). "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature*, 50(1), 3-50.

**Social Security Administration.** (2012). *Social Security Handbook.* Washington, D.C.: SSA.

**Song, J. G.** (2004). "Evaluating the Initial Impact of Eliminating the Retirement Earnings Test." *Social Security Bulletin*, 65(1), 1-15.

**Song, J. G., and Manchester, J.** (2007). "New Evidence on Earnings and Benefit Claims Following Changes in the Retirement Earnings Test." *Journal of Public Economics*, 91, 669-700.

**Tergesen, A.** (2016). "How to Get People to Delay Retirement." *Wall St. Journal*, March 21.

**Turner, L.** (2014). "The Road to Pell is Paved with Good Intentions: The Economic Incidence of Federal Student Grant Aid." University of Maryland Working Paper.

**Utendorf, K.** (2001/2002). "The Upper Part of the Earnings Distribution in the United States: How Has It Changed?" *Social Security Bulletin*, 64 (3), 1-11.

**Viscusi, K.** (1979). *Welfare of the Elderly: An Economic Analysis and Policy Prescription.* New York, NY: John Wiley and Sons.

**Vroman, W.** (1985). "Some Economic Effects of the Retirement Test." In Ehrenberg, R., ed., *Research in Labor Economics* (Vol. 7, pp. 31-89). Greenwich, CT: Jai Press, Inc.

Figure 1: Extensive Margin Response by State 0 Counterfactual Earnings for a Kink (Panel A) and a Notch (Panel B)

Panel A: Kinked Tax Schedule
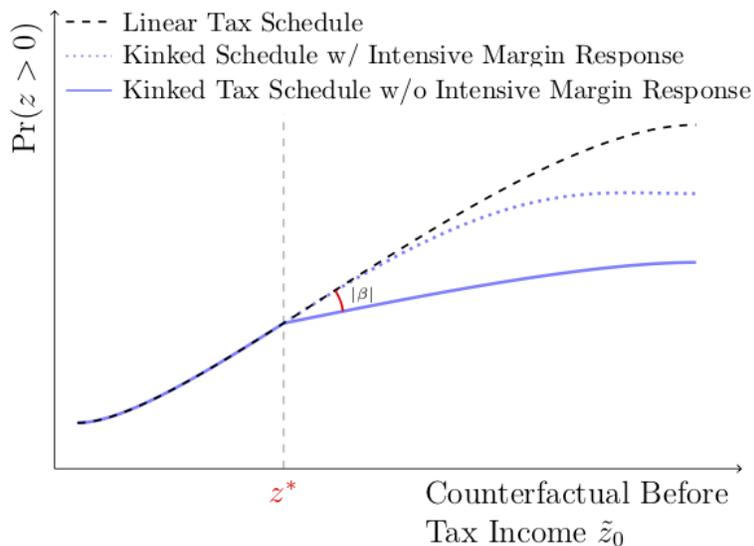


Panel B: Notched Tax Schedule
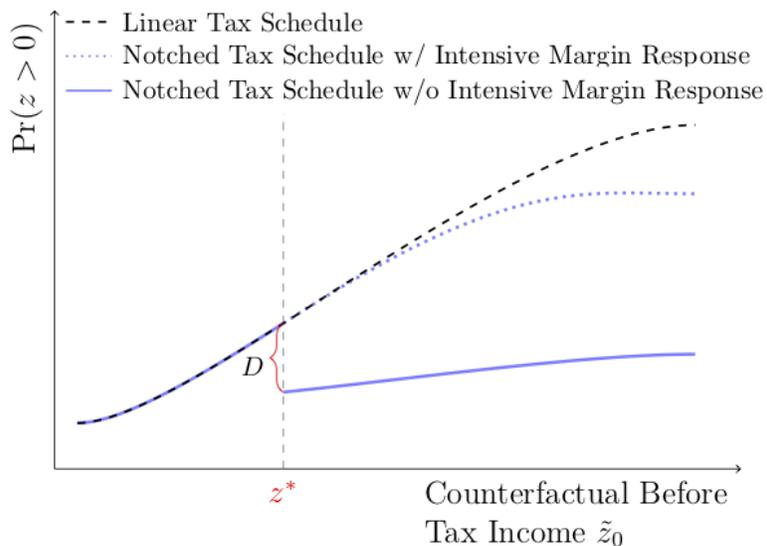


Notes: The figures illustrate features of the model described in Section 2. The $x$-axes show desired income if employed on a linear budget set in state 0, $i.e.$ $\tilde{z}_0$. The $y$-axes show a hypothetical probability of employment, under three scenarios: a linear schedule in state 0 (dashed line); a kinked or notched tax schedule in state 1 when individuals can make intensive margin adjustments (dotted line); and a kinked (Panel A) or notched (Panel B) tax schedule in state 1 when individuals cannot make intensive margin adjustments (solid line). $\beta$ refers to the change in slope at the exempt amount of the employment probability as a function of desired earnings when a kink is present. $D$ refers to the discontinuity at the exempt amount under a notch.

Figure 2: Extensive Margin Incentives for a Kink (Panel A) and a Notch (Panel B)

Panel A: Kinked Tax Schedule



Panel B: Notched Tax Schedule
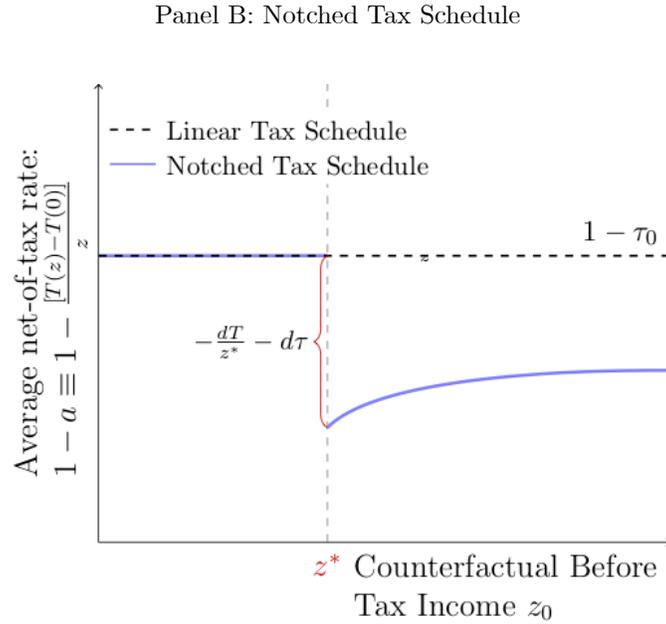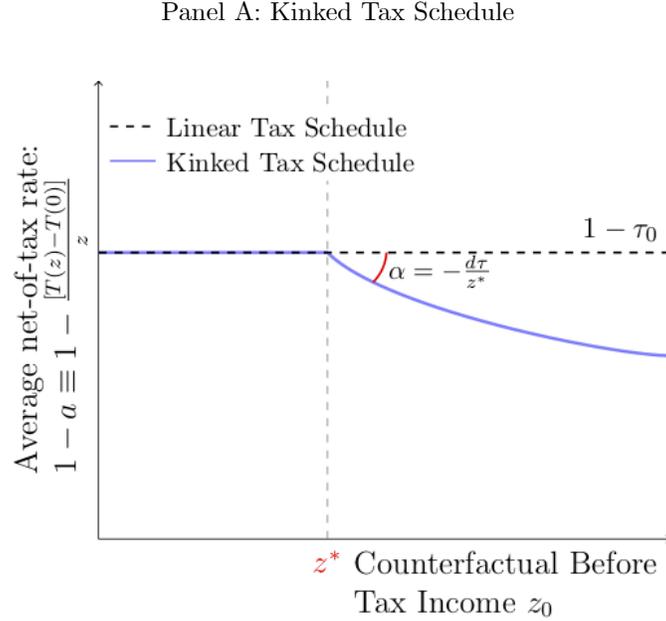


Notes: The figures show the ANTR ($y$-axes) as a function of before-tax income ($x$-axes). In both panels, incentives under a linear tax schedule in which the ANTR is equal to $1 - \tau_0$ everywhere are represented by a dashed line. In Panel A, incentives under a kinked tax schedule in which the ANTR is equal to $1 - \tau_0$ below the kink point $z^*$, and $1 - \tau_1 + d\tau z^*/z$ above $z^*$, are represented by a solid line. In Panel B, incentives under a notched tax schedule in which the ANTR is $1 - \tau_0$ below $z^*$ and $1 - dT/z - \tau_1$ above $z^*$ are likewise represented by a solid line. Panel A shows that under a kink, the slope of this graph discontinuously decreases at the kink point $z^*$, due to the imposition of the marginal tax on earnings above $z^*$. Panel B shows that under an notch, the ANTR discontinuosly drops at $z^*$ (and its slope discontinuously rises). $1 - a$ is defined as the average net-of-tax rate, i.e. $1 - a = 1 - [T(z) - T(0)]/z$.

Figure 3: Intensive and Extensive Margin Responses to a Kink (Panels a and b) or Notch (Panels c and d)
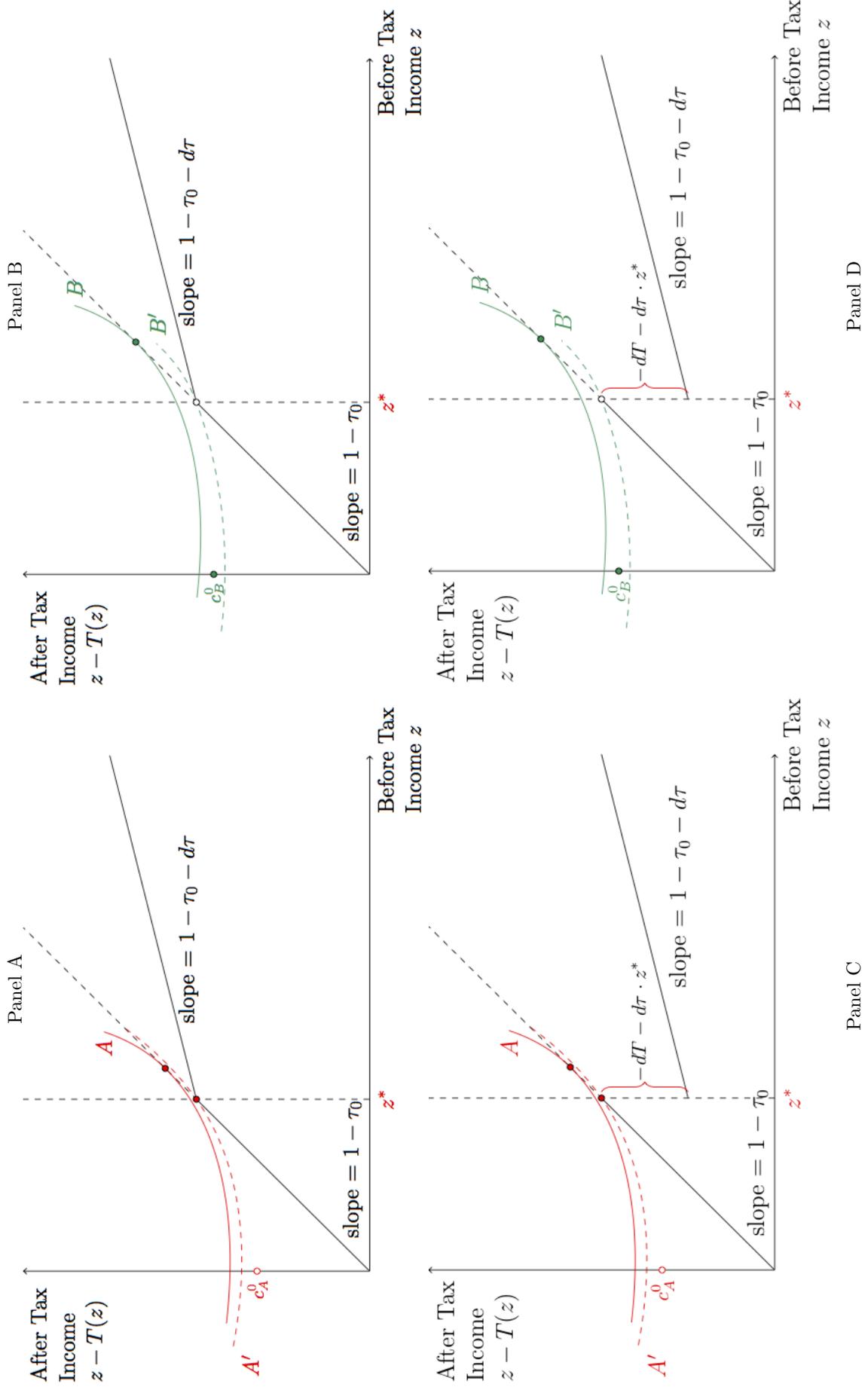


Notes: The figure depicts potential responses to a kinked or notched budget set. In Panel A the agent reduces earnings to the kink at $z^*$, preferring this to the outside level of consumption. In Panel B the agent prefers the outside option of earning zero to the optimal level of earnings $z^*$ conditional on being employed. Panels C and D show analogous patterns in the case of a notch.

Figure 4: Pre-AET Earnings Distribution and Intensive Margin Response to the AET
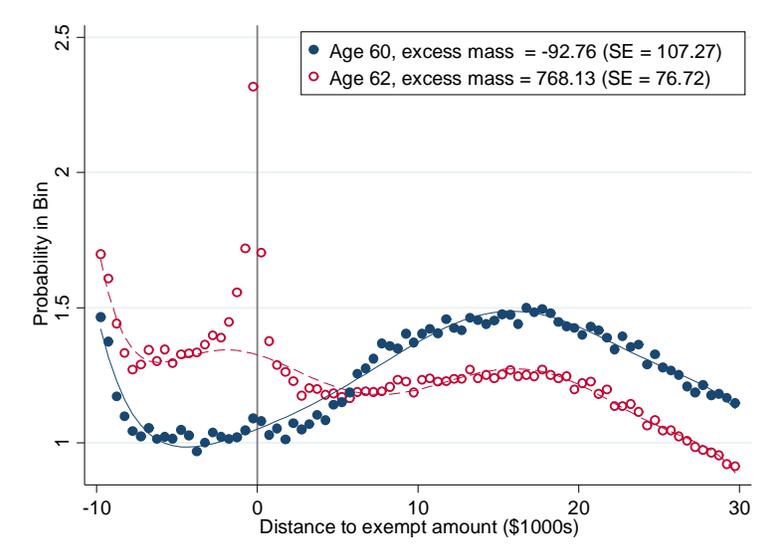


Notes: The figure shows the actual earnings density (dotted plot) and smooth fit (solid or dashed line). The data shown are means within bins of width $500. The exempt amount is normalized to zero and shown with the vertical line at zero. Following Chetty *et al.* (2011), the smooth density is estimated using a seventh-order polynomial, excluding the region within a $3,000 bandwidth of the exempt amount. The source is a 25 percent random sample of the Social Security Administration Master Earnings File linked to the Master Beneficiary Record. See the Appendix for a description of how we estimate excess mass. The sample consists of individuals with age 60 earnings that are positive and within $40,000 of the exempt amount, born 1918 to 1923, with no self-employment income at age 60, and excluding individuals who ever have negative earnings at ages 50-57 or 63-70.

Figure 5: Predetermined Covariates
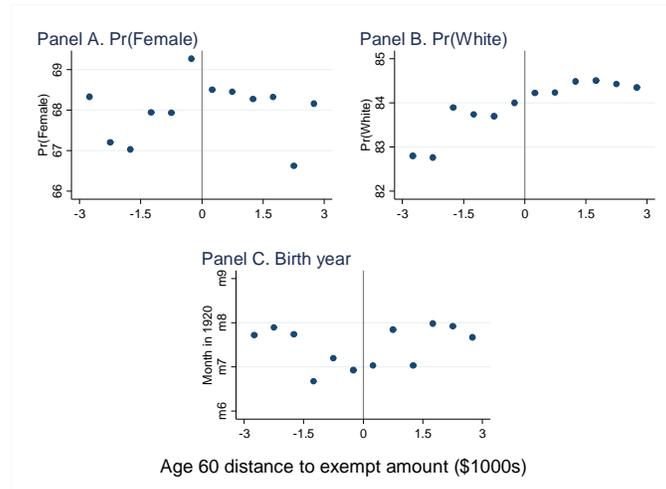


Note: The figure shows the bin means of predetermined covariates as a function of the distance to the age 60 exempt amount. The figure demonstrates that there are no clear visual changes in slope in any of these covariates at the age 60 exempt amount, consistent with the assumptions necessary for the validity of the regression kink design.

Figure 6: Histogram of Percent Real Earnings Growth, Ages 59 to 60
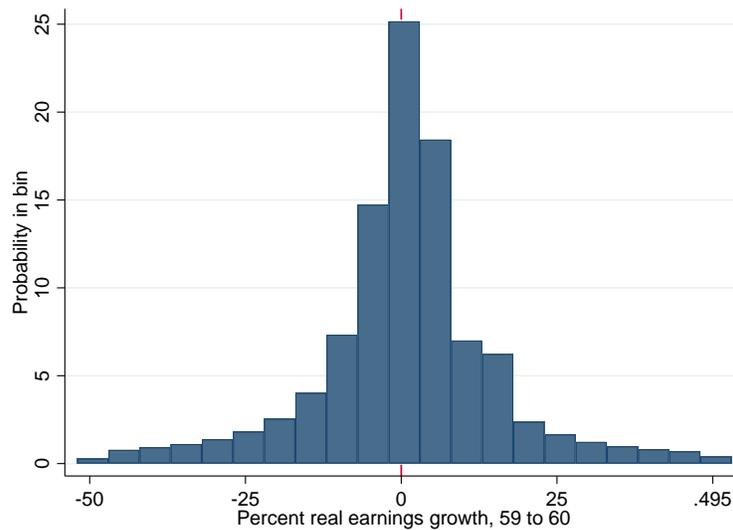


Notes: This histogram shows that there is a large mass near zero percent real earnings growth across a "placebo" set of ages, 59 and 60, when individuals do not face the AET. This indicates that a substantial mass of individuals have no growth in desired real earnings, consistent with the assumptions necessary for our RKD to estimate a lower bound on the elasticity as described in the main text. Real earnings in each year are calculated using the CPI-U. See other notes to Figure 4.

Figure 7: Employment Rate Near the Exempt Amount, Ages 63 to 64
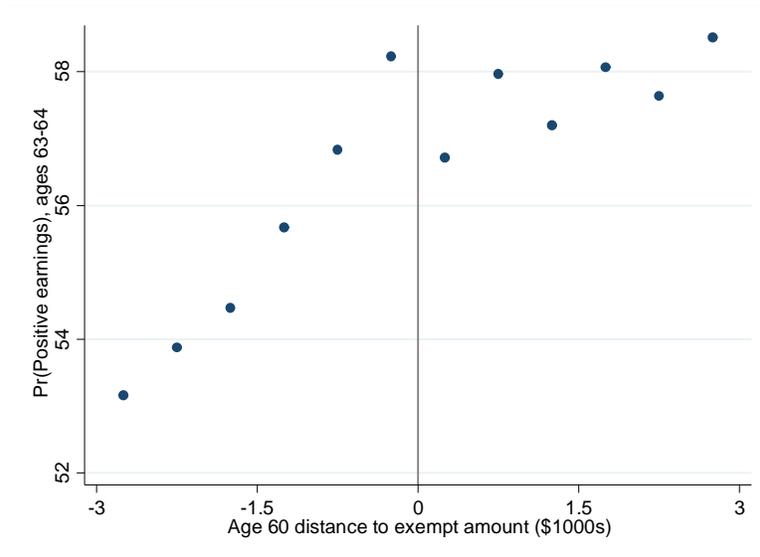


Notes: The figure plots the mean annual employment rate, *i.e.* the probability of positive earnings, at ages 63 to 64 averaged, as a function of the distance of age 60 earnings from the exempt amount. The sample is individuals with positive age 60 earnings and no age 60 self-employment income, born 1918 to 1923. See other notes to Figure 4.

Figure 8: Kink Estimate by Bandwidth
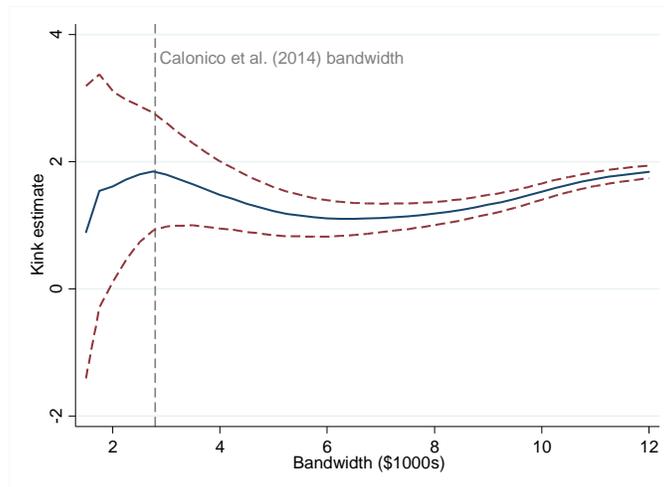


Notes: The figure shows that as a function of the bandwidth, the estimated change in slope at the exempt amount of the mean annual probability of age 63 to 64 employment is relatively stable. The solid line shows the point estimates, and the dashed lines show the 95 percent confidence intervals. See other notes to Figure 4.

44

Figure 9: Placebo Tests

Panel A: Probability of Positive Earnings, Ages 56 to 57
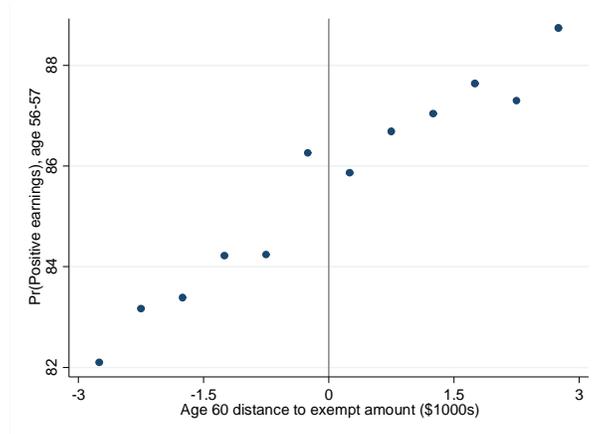
Panel B: Probability of Positive Earnings by Single Year of Age, Ages 61 to 64
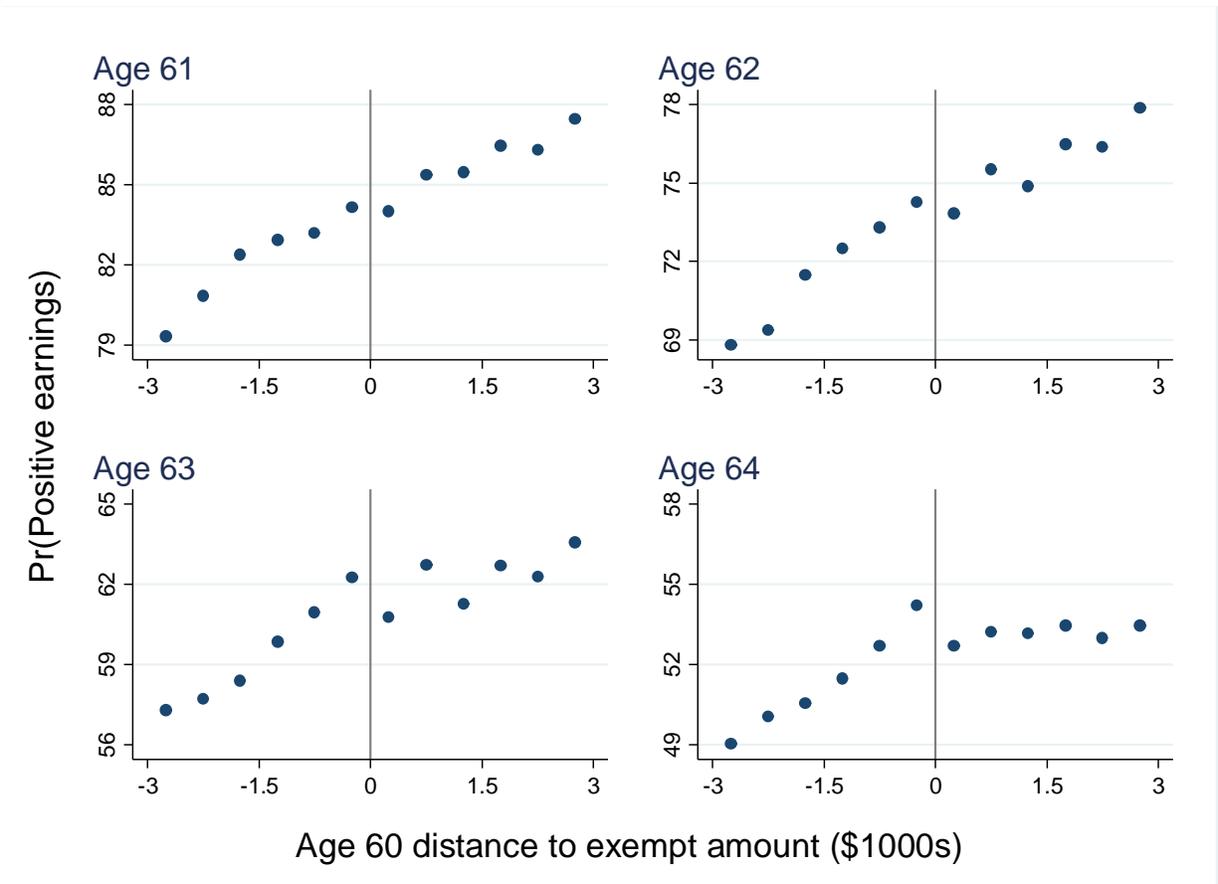
Notes: Panel A plots the annual probability of positive earnings at ages 56 to 57, as a function of the distance of age 60 earnings from the exempt amount. In Panel B, each figure plots the probability of positive earnings for each single year of age from 61 to 64. See other notes to Figure 4.

Figure 10: Permutation Test
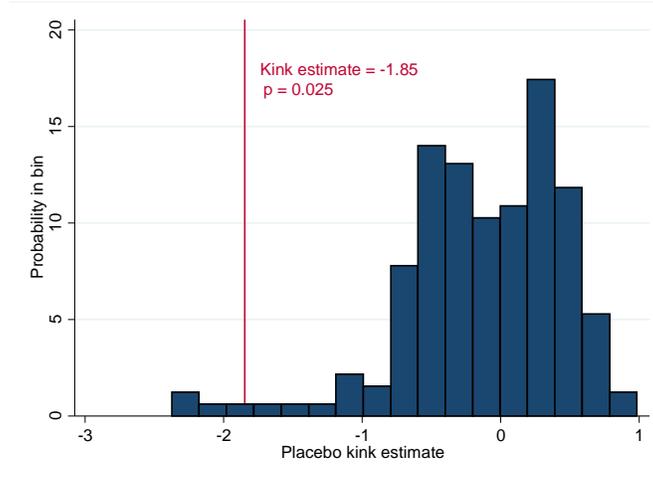


Notes: The figure plots the histogram of placebo kink estimates from a permutation test in the spirit of Ganong and Jäger (2015). We estimate a set of placebo changes in slope in the mean annual age 63 to 64 employment rate, using the same specification as our main estimates except that we examine the change in slope at placebo locations of the exempt amount away from the true exempt amount. The figure shows that the point estimate recovered from the actual exempt amount, shown by the vertical line in the figure, is located well below the significant majority of the distribution of placebo estimates, reinforcing the view that we are detecting a true effect of the AET. This permutation test shows a significant kink (p=0.025). See other notes to Figure 4.

Figure 11: R-Squared by Placebo Kink Location
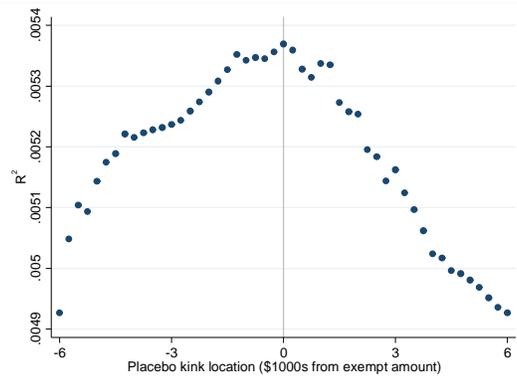


Notes: The figure plots the R-squared of our baseline specification against the "placebo" kink location relative to the exempt amount, following Landais (2015). The vertical line denotes the actual location of the exempt amount. As described in the text, we estimate a set of placebo changes in slope in the mean annual age 63 to 64 employment rate, using the same specification as our main estimates except that we examine the change in slope at placebo locations of the exempt amount away from the true exempt amount. The figure shows that the R-squared is maximized at the true location of the placebo kink, supporting our hypothesis that we have found a true kink in the data rather than a spurious underlying nonlinearity in the relationship between the yearly employment rate at ages 63 to 64 and age 60 earnings. See other notes to Figure 4.

Table 1: Summary Statistics

| | (1) Within $2,797 of Exempt Amount | (2) Full Earnings Range |
|---|---|---|
| Percent of years with positive earnings, ages 63 to 64 | 56.53 (45.92) | 63.27 (44.05) |
| Annual earnings, ages 63 to 64 (including zeroes) | $5,813 ($8,134) | $20,115 ($24,599) |
| Claim age (if claimed) | 63.13 (2.61) | 63.68 (2.14) |
| Percent female | 68.00 (46.65) | 41.85 (49.33) |
| Percent white | 83.96 (36.70) | 89.30 (30.91) |
| Year of birth | 1920.62 (1.68) | 1920.52 (1.69) |
| $N$ | 95,960 | 1,425,600 |

Notes: The table shows the means of each variable shown in the row headings, with the standard deviation in parentheses, from a 25 percent random sample of the Social Security Administration Master Earnings File and Master Beneficiary Record. The sample is all individuals born 1918-1923 with positive earnings at age 60, and satisfying the other sample restrictions described in the main text. In Column 1, the bandwidth relative to the AET exempt amount, $2,797, corresponds to the bandwidth used in our main regressions. For comparison, Column 2 shows individuals from the full range of earnings. Earnings are in thousands of real 2010 dollars. N's refer to the number of individuals.

Table 2: Initial Tests of Smoothness

| | (1) Density | (2) White (Indicator) | (3) Female (Indicator) | (4) Year of Birth |
|---|---|---|---|---|
| | | Estimated Kink | | |
| Linear | -0.057 (0.067) | -0.0005 (0.36) | -0.035 (0.34) | 0.052 (0.018)*** |
| $N$ | 141,110 | 172,383 | 330,556 | 102,640 |
| Quadratic | -0.26 (0.36) | -0.074 (1.48) | 0.0069 (0.28) | 0.074 (0.017)*** |
| $N$ | 141,110 | 696,139 | 348,920 | 304,306 |
| Cubic | -0.26 (0.36) | -0.98 (0.69) | -2.25 (0.72)*** | 0.038 (0.028) |
| $N$ | 141,110 | 282,460 | 362,174 | 308,690 |

Notes: The table presents tests for the smoothness of the earnings density, and of three pre-determined covariates (indicators for being white or female, and year of birth). The table reports coefficients corresponding to $\beta_1$ in equation (18), reflecting the change in slope in the outcome at the age 60 exempt amount, as a function of age 60 earnings relative to the exempt amount. The column headings indicate the outcome variable, while the row headings indicate the specification of the running variable (linear, quadratic, or cubic). The number of individuals N in each regression is shown below the standard error. In Columns 2 to 4, robust standard errors following Calonico *et al.* (2014) are reported in parentheses, and the bandwidth is also chosen using Calonico *et al.* (2014). The resulting difference in bandwidths is the reason for the discrepancy across specifications in the number of individuals included in these regressions. The N's are the same across the specifications for the density outcome in Column 1 because following previous literature, we use the McCrary (2008) procedure to calculate the bandwidth, bin size, and standard error for the density outcome. Here and in the following tables, *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10, all from two-tailed tests of equality with zero. See other notes to Table 1.

Table 3: Main Results Documenting Kink in Employment

|  | (1) | (2) | (3) |
|---|---|---|---|
|  | Main | With | Age 56-57 |
|  | RKD | Controls | (Placebo) |
| Linear | -1.85 | -1.68 | -0.59 |
|  | (0.72)*** | (0.66)*** | (0.37) |
| N | 95,960 | 104,665 | 106,241 |
| Quadratic | -2.47 | -2.27 | 0.11 |
|  | (1.02)*** | (0.89)*** | (0.53) |
| N | 160,785 | 172,979 | 277,187 |
| Cubic | -3.11 | -2.57 | -0.27 |
|  | (1.07)*** | (1.04)*** | (0.59) |
| N | 273,241 | 326,782 | 407,773 |

Notes: The table shows regression kink estimates showing the change at the age 60 exempt amount in the slope of the mean annual probability of positive earnings at ages 63 and 64 as a function of age 60 distance to the exempt amount, corresponding to $\beta_1$ in equation (18). Column 1 is our main specification described in the main text. In Column 1, the Calonico *et al.* (2014) bandwidth is \$2,797 for the linear specification, \$3,047 for the quadratic, and \$3,093 for the cubic. Column 2 adds indicators for demographic categories (sex, race groups, and year of birth). Column 3 shows a placebo test in which the mean probability of positive earnings at ages 56 and 57 is the dependent variable. Robust standard errors, following Calonico *et al.* (2014), are in parentheses. The number of individuals in each regression is shown below the standard error. See other notes to Table 2.

Table 4: RKD Estimates by Single Year of Age

|  | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
|  | Age 61 | Age 62 | Age 63 | Age 64 |
| Estimate of slope | -0.33 | -0.90 | -1.60 | -1.98 |
| discontinuity | (0.81) | (0.91) | (0.72)** | (0.74)*** |
| N | 80,148 | 244,465 | 105,871 | 98,502 |

Notes: The table presents regression kink estimates showing the change at the age 60 exempt amount in the slope of the probability of positive earnings at each age separately from 61 to 64, as a function of the age 60 distance to the exempt amount, corresponding to $\beta_1$ in equation (18). We our use our baseline linear specification. The age in question is shown in each column heading. See other notes to Table 2.

Table 5: Elasticity Estimates

|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
|  | Full |  |  |  | Non- | High Prior | Low Prior |
|  | Sample | Men | Women | White | White | Earnings | Earnings |
| Elasticity | 0.49 | 0.25 | 0.49 | 0.47 | 0.41 | 0.48 | 0.40 |
|  | (0.19)*** | (0.19) | (0.21)** | (0.18)*** | (0.43) | (0.48) | (0.16)*** |
| N | 95,960 | 68,971 | 66,251 | 93,722 | 39,271 | 19,574 | 101,709 |

Notes: The table presents the estimates for the full sample and various subgroups, from the baseline linear specification. The "high prior earnings" and "low prior earnings" subgroups refer to individuals for whom mean earnings from ages 40 to 59 were above the median and below the median, respectively. The elasticity is calculated as described in the text, using the AET rules to calculate analytically the first stage change in slope of the ANTR at the exempt amount; averaging across calendar years, the mean first stage change in the slope of the ANTR at the exempt amount is -4.88. N's refer to the number of individuals included in each "reduced form" regression (18). The N's differ across specifications both because the samples differ, and because the the Calonico *et al.* (2014) bandwidths differ across samples. See other notes to Table 2.

# Appendix: For Online Publication

## A    Proofs of Propositions and Other Claims

**Proposition 1:** In general the slope of $\frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}}$ will be given by:

$$\frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} = g(\bar{q}_{n1}|n)\frac{d\bar{q}_{n,1}}{d\tilde{z}_{n0}} + \frac{\partial G(\bar{q}_{n1}|n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} \tag{A.1}$$

Focusing on the first term in the expression for $d\Pr(z_{n1}>0|\tilde{z}_{n0})/d\tilde{z}_{n0}$ in (A.1), we have:

$$\begin{aligned}
\frac{d\bar{q}_{n,1}}{d\tilde{z}_{n0}} &= \frac{\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)}{\partial\tilde{z}_{n1}}\frac{d\tilde{z}_{n1}}{d\tilde{z}_{n0}} + \frac{\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} \\
&= \frac{\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}}
\end{aligned} \tag{A.2}$$

When agents are unrestricted in their intensive margin earnings choice, we can set the first term on the right side of (A.2) to zero. For those with $\tilde{z}_{n0}<z^*$ or $\tilde{z}_{n0}>z^*+\Delta z^*$, $\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)/\partial\tilde{z}_{n1}=0$ due to the envelope theorem.[34] For those with $z^*\leq\tilde{z}_{n0}\leq z^*+\Delta z^*$, we have $d\tilde{z}_{n1}/d\tilde{z}_{n0}=0$, since $\tilde{z}_{n1}=z^*$ for everyone in this set—*i.e.* these agents bunch at $z^*$. Substituting for $d\bar{q}_{n1}/dz_{n0}$ in (A.1) using (A.2), we have:

$$\frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} = g(\bar{q}_{n1}|n)\frac{\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} + \frac{\partial G(\bar{q}_{n1}|n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} \tag{A.3}$$

when individuals are able to adjust on both the intensive and extensive margins.

Our smoothness assumptions imply that this slope is continuous, and in particular it is continuous at $z^*$ since $n$, $\bar{q}_{n1}$, $\tilde{z}_{n1}$, $T_1(\cdot)$ and $\partial G(\bar{q}_{nj}|n)/\partial n$ are all continuous in $\tilde{z}_{n0}$ at $z^*$. Furthermore, $g(\cdot)$ and $\partial v/\partial n$ are likewise continuous in their arguments. Thus, we have:

$$\lim_{\tilde{z}_{n0}\to z^{*+}}\frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} = \lim_{\tilde{z}_{n0}\to z^{*-}}\frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} \tag{A.4}$$

That is, the employment probability does not exhibit any first-order change in slope at $z^*$, even though the ANTR does feature such a discontinuity.

**Proposition 2:** If $\tilde{z}_{n1}\equiv\tilde{z}_{n0}$, the general expression for $d\Pr(z_{n1}>0|\tilde{z}_{n0})/d\tilde{z}_{n0}$ from (A.1) still holds. However, we now have a slightly different expression for the critical level of fixed costs, which is now evaluated at $\tilde{z}_{n0}$, implying $\bar{q}_{n1}\equiv v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)-v^0$. Accordingly, we have a different expression for $d\bar{q}_{n1}/d\tilde{z}_{n0}$ relative to (A.2). Since $\tilde{z}_{n1}=\tilde{z}_{n0}$ for everyone, we have:

$$\frac{d\bar{q}_{n1}}{d\tilde{z}_{n0}} = \frac{\partial v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)}{\partial\tilde{z}_{n0}} + \frac{\partial v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} \tag{A.5}$$

where the key difference is that $\partial v/\partial z$ and $\partial v/\partial n$ are evaluated at $\tilde{z}_{n0}$ instead of $\tilde{z}_{n1}$. For those with $\tilde{z}_{n0}<z^*$, since $T_1(\cdot)=T_0(\cdot)$ and $\tilde{z}_{n1}=\tilde{z}_{n0}$, it is still the case that $\partial v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)/\partial\tilde{z}_{n0}=0$ due to the envelope theorem. However, the first term in (A.5) for those with $\tilde{z}_{n0}>z^*$ is now:

$$\begin{aligned}
\frac{\partial v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)}{\partial\tilde{z}_{n0}} &= (1-\tau_1)v_c+v_z \\
&= \lambda_n\left[(1-\tau_1)+\frac{v_z}{v_c}\right]
\end{aligned} \tag{A.6}$$

where $\lambda_n\equiv v_c$, and $v_c$ and $v_z$ are the partial derivatives of $v(\cdot)$ with respect to $c$ and $z$, respectively, evaluated at $(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)$.

---

[34] In this and other similar expressions elsewhere we evaluate the partial derivative of $v$ with respect to $z$ allowing both earnings and consumption to change via the budget constraint, but holding $n$ constant.

Thus, we now have:

$$\frac{d\Pr\left(z_{n1}>0\middle|\tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} = \begin{cases} g\left(\bar{q}_{n1}\middle|n\right)\frac{\partial v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} + \frac{\partial G(\bar{q}_{n1}|n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}}, & \text{if } \tilde{z}_{n0}<z^* \\ g\left(\bar{q}_{n1}\middle|n\right)\left[\lambda_n\left[(1-\tau_1)+\frac{v_z}{v_c}\right]+\frac{\partial v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}}\right]+\frac{\partial G(\bar{q}_{n1}|n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}}, & \text{if } \tilde{z}_{n0}\geq z^* \end{cases}$$
(A.7)

Note also that:

$$\lim_{\tilde{z}_{n0}\to z^{*+}}\frac{v_z\left(\tilde{z}_{n0}-T_1\left(\tilde{z}_{n0}\right),\tilde{z}_{n0};n\right)}{v_c\left(\tilde{z}_{n0}-T_1\left(\tilde{z}_{n0}\right),\tilde{z}_{n0};n\right)}=-\left(1-\tau_0\right)$$
(A.8)

where we have used the first order condition for $\tilde{z}_{n,0}$ and the fact that $\lim_{\tilde{z}_{n0}\to z^*}T_1\left(\tilde{z}_{n0}\right)=T_0\left(z^*\right)$. We now have the following expression for the difference in slopes at $z^*$:

$$\begin{aligned}\lim_{\tilde{z}_{n0}\to z^{*+}}\frac{d\Pr\left(z_{n1}>0\middle|\tilde{z}_{n0}\right)}{d\tilde{z}_{n0}}-\lim_{\tilde{z}_{n0}\to z^{*-}}\frac{d\Pr\left(z_{n1}>0\middle|\tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} &= \lim_{\tilde{z}_{n0}\to z^{*+}}g\left(\bar{q}_{n1}\middle|n\right)\cdot\lambda_n\left[(1-\tau_1)+\frac{v_z}{v_c}\right]\\ &= g\left(\bar{q}_{n^*1}\middle|n^*\right)\cdot\lambda_{n^*}\left[(1-\tau_1)-(1-\tau_0)\right]\\ &= -d\tau\cdot\lambda_{n^*}\cdot g\left(\bar{q}_{n^*1}\middle|n^*\right)\end{aligned}$$
(A.9)

where $\bar{q}_{n^*1}$, $n^*$, and $\lambda_{n^*}$ all refer the individual for whom $\tilde{z}_{n0}=z^*$.*

**Proposition 3:** To illustrate the first part of the result, note that $\tilde{z}_{n,1}=z^*$ for everyone with $\tilde{z}_{n,0}\in\left[z^*,z^*+\triangle z^N\right]$, i.e. the set of bunchers. Thus, as we cross $z^*$, there is no discontinuous change in after-tax income or optimal earnings, conditional on working, and thus no discontinuous change in the employment probability. To show the second result, our proof proceeds almost step-for-step as in the case of a kink. The key step of that proof is that the following term from (A.2) continues to vanish:

$$\frac{\partial v\left(\tilde{z}_{n1}-T_1\left(\tilde{z}_{n1}\right),\tilde{z}_{n1};n\right)}{\partial\tilde{z}_{n1}}\frac{d\tilde{z}_{n1}}{d\tilde{z}_{n0}}=0$$
(A.10)

For those with $\tilde{z}_{n0}<z^*$ or $\tilde{z}_{n0}>z^*+\Delta z^N$, $\partial v\left(\tilde{z}_{n1}-T_1\left(\tilde{z}_{n1}\right),\tilde{z}_{n1};n\right)/\partial\tilde{z}_{n1}=0$ due to the envelope theorem.[35] For those with $z^*\leq\tilde{z}_{n0}\leq z^*+\Delta z^N$, we have $d\tilde{z}_{n1}/d\tilde{z}_{n0}=0$, since $\tilde{z}_{n1}=z^*$ for everyone in this set—*i.e.* these agents bunch at $z^*$. The result then follows as in the case of a kink.

**Proposition 4**: The result follows immediately from the fact that utility just to the left and right of the notch will be $v\left(z^*-T_0\left(z^*\right),z^*;n^*\right)$ and $v\left(z^*-T_1^+\left(z^*\right),z^*;n^*\right)$, respectively. We also use the following first-order approximations:

$$\begin{aligned}G\left(\bar{q}_{n^*1}\middle|n^*\right)-G\left(\bar{q}_{n^*0}\middle|n^*\right) &\approx g\left(\bar{q}_{n^*,0}\middle|n^*\right)\left[\bar{q}_{n^*1}-\bar{q}_{n^*0}\right]\\ &= g\left(\bar{q}_{n^*,0}\middle|n^*\right)\left[v\left(z^*-T_1^+\left(z^*\right),z^*;n^*\right)-v\left(z^*-T_0\left(z^*\right),z^*;n^*\right)\right]\end{aligned}$$(A.11)

$$\begin{aligned}v\left(z^*-T_1^+\left(z^*\right),z^*;n^*\right)-v\left(z^*-T_0\left(z^*\right),z^*;n^*\right) &\approx \frac{\partial v}{\partial c}\left[z^*-T_1^+\left(z^*\right)-z^*+T_0\left(z^*\right)\right]\\ &= -\lambda_{n^*}\left[dT+d\tau\cdot z^*\right]\end{aligned}$$
(A.12)

**Proposition 5** *Assume that the utility function is quasi-linear in consumption and that the density of fixed costs is locally constant in the neighborhood of $z^*$, i.e. $g\left(\bar{q}_{n^*0}\middle|n^*\right)=g\left(\bar{q}_{n^*1}^+\middle|n^*\right)$. Then the employment probability, as a function of desired earnings in state 0, will exhibit a first-order change in slope (i.e. a kink)*

---

[35] In this and other similar expressions elsewhere we evaluate the partial derivative of $v$ with respect to $z$ allowing both earnings and consumption to change via the budget constraint, but holding $n$ constant.

*at $z^*$. The kink is given by:*

$$\lim_{\tilde{z}_{n0}\to z^{*+}} \frac{d\Pr\left(z_{n1}>0\middle|\tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} - \lim_{\tilde{z}_{n0}\to z^{*-}} \frac{d\Pr\left(z_{n1}>0\middle|\tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} = -d\tau\cdot g\left(\bar{q}_{n^*0}\middle|n^*\right)$$
$$-\left[\frac{\partial G\left(\bar{q}_{n^*0}\middle|n^*\right)}{\partial n} - \frac{\partial G\left(\bar{q}_{n^*1}^+\middle|n^*\right)}{\partial n}\right]\frac{dn}{d\tilde{z}_{n0}}$$

$$\text{(A.13)}$$

*Moreover, if we use a first-order approximation for $\partial G/\partial n$, we have:*

$$\lim_{\tilde{z}_{n0}\to z^{*+}} \frac{d\Pr\left(z_{n1}>0\middle|\tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} - \lim_{\tilde{z}_{n0}\to z^{*-}} \frac{d\Pr\left(z_{n1}>0\middle|\tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} \approx -d\tau\cdot g\left(\bar{q}_{n^*0}\middle|n^*\right)$$
$$- \left[dT + d\tau\cdot z^*\right]\frac{\partial g\left(\bar{q}_{n^*0}\middle|n^*\right)}{\partial n}\frac{dn}{d\tilde{z}_{n0}}\text{(A.14)}$$

**Proof 5**: Again, the proof proceeds as in the case of a kink above. However, the notch introduces discontinuities in a number of the elements of equation (A.7). In particular, we now have:

$$\lim_{\tilde{z}_{n0}\to z^{*+}} \bar{q}_{n1} = \bar{q}_{n^*1}^+ \neq \bar{q}_{n^*0} = \lim_{\tilde{z}_{n0}\to z^{*-}} \bar{q}_{n1} \tag{A.15}$$

However, if we assume that the density of fixed costs is locally constant in $q$ at $n^*$, then:

$$\lim_{\tilde{z}_{n0}\to z^{*+}} g\left(\bar{q}_{n1}\middle|n\right) = g\left(\bar{q}_{n^*1}\middle|n^*\right) = g\left(\bar{q}_{n^*0}\middle|n^*\right) = \lim_{\tilde{z}_{n0}\to z^{*-}} g\left(\bar{q}_{n1}\middle|n\right) \tag{A.16}$$

Furthermore, if we assume quasilinearity, we have:

$$\lambda_n = v_c = 1$$
$$\lim_{\tilde{z}_{n0}\to z^{*+}} \frac{\partial v\left(\tilde{z}_{n0} - T_1\left(\tilde{z}_{n0}\right), \tilde{z}_{n0}; n\right)}{\partial n} = \lim_{\tilde{z}_{n0}\to z^{*-}} \frac{\partial v\left(\tilde{z}_{n0} - T_1\left(\tilde{z}_{n0}\right), \tilde{z}_{n0}; n\right)}{\partial n}$$
$$\lim_{\tilde{z}_{n0}\to z^{*+}} v_z\left(\tilde{z}_{n0} - T_1\left(\tilde{z}_{n0}\right), \tilde{z}_{n0}; n\right) = -\left(1 - \tau_0\right) \tag{A.17}$$

The second and third lines follow because the quasilinearity assumption breaks the relationship between the level of consumption and the disutility of labor. Thus, a discontinuity in after-tax income doesn't effect the marginal disutility of labor. In the third line, we also use the first-order condition for $\tilde{z}_{n0}$. Using these above results, the first part of the proposition follows. The second part of the proposition can be shown using the following first-order approximation:

$$\frac{\partial G\left(\bar{q}_{n^*1}^+\middle|n^*\right)}{\partial n} - \frac{\partial G\left(\bar{q}_{n^*0}\middle|n^*\right)}{\partial n} \approx \frac{\partial^2 G\left(\bar{q}_{n^*0}\middle|n^*\right)}{\partial q\partial n}\left[\bar{q}_{n^*1}^+ - \bar{q}_{n^*0}\right]$$
$$= -\frac{\partial g\left(\bar{q}_{n^*0}\middle|n^*\right)}{\partial n}\left[dT + d\tau\cdot z^*\right]$$

**Estimation of Kink with Measurement Error:** In Section 4.2, we argue that in the case of measurement error, i.e. $\pi\left(\tilde{z}_0, v_{it}\right) < 1$, we estimate a lower bound on our elasticity. Here we formally demonstrate this result. Recall that our relationship between earnings at age 60 and desired earnings at ages 63-64 is $z^{60} = \tilde{z}_{0it} + p_{it}v_{it}$. We assume that the joint distribution of $(p, v)$, conditional on $z^{60}$, is continuous and continuously differentiable in $z^{60}$ at $z^{60} = z^*$. Consider the estimated kink in employment, as a function of earnings at age 60, *i.e.* the numerator in equation (17), using $z^{60}$ as the running variable. Denote this as $\beta^{RKD}$:

$$\beta^{RKD} \equiv \lim_{z_i^{60}\to z^{*+}} \frac{\partial\Pr\left(L_{it}=1\middle|z_i^{60}\right)}{\partial z_i^{60}} - \lim_{z_i^{60}\to z^{*-}} \frac{\partial\Pr\left(L_{it}=1\middle|z_i^{60}\right)}{\partial z_i^{60}}$$

Let the potential deviation in desired earnings, $v$, have a distribution with the CDF $M\left(v\middle|z\right)$ and pdf

$m\left(v\,|\,z\right)$, conditional on $z^{60} = z$. We have the following:

$$
\begin{aligned}
\beta^{RKD} &= \int \pi\left(z^*, v\right) \left[ \lim_{z_i^{60} \to z^{*+}} \frac{\partial \Pr\left(L_{it} = 1\,|\, z_i^{60}\right)}{\partial z_i^{60}} - \lim_{z_i^{60} \to z^{*-}} \frac{\partial \Pr\left(L_{it} = 1\,|\, z_i^{60}\right)}{\partial z_i^{60}} \right] m\left(v\,|\, z^*\right) dv \\
&\quad + \int \left(1 - \pi\left(z^* - v, v\right)\right) \left[ \lim_{z_i^{60} \to z^{*+}} \frac{\partial \Pr\left(L_{it} = 1\,|\, z_i^{60}\right)}{\partial z_i^{60}} - \lim_{z_i^{60} \to z^{*-}} \frac{\partial \Pr\left(L_{it} = 1\,|\, z_i^{60}\right)}{\partial z_i^{60}} \right] m\left(v\,|\, z^*\right) dv \\
&= \int \pi\left(z^*, v\right) \left[ \lim_{\tilde{z}_{i0t} \to z^{*+}} \frac{\partial \Pr\left(L_{it} = 1\,|\, \tilde{z}_{i0t}\right)}{\partial \tilde{z}_{i0t}} - \lim_{z_i^{60} \to z^{*-}} \frac{\partial \Pr\left(L_{it} = 1\,|\, \tilde{z}_{i0t}\right)}{\partial \tilde{z}_{i0t}} \right] m\left(v\,|\, z^*\right) dv \\
&\quad + \int \left(1 - \pi\left(z^* - v, v\right)\right) \left[ \lim_{\tilde{z}_{i0t} \to z^{*+} - v_{it}} \frac{\partial \Pr\left(L_{it} = 1\,|\, \tilde{z}_{i0t}\right)}{\partial \tilde{z}_{i0t}} - \lim_{\tilde{z}_{i0t} \to z^{*-} - v_{it}} \frac{\partial \Pr\left(L_{it} = 1\,|\, \tilde{z}_{i0t}\right)}{\partial \tilde{z}_{i0t}} \right] m\left(v\,|\, z^*\right) dv \\
&= \bar{\pi} \cdot \hat{\beta}\left(z^*\right) + \left(1 - \bar{\pi}\right) \int \frac{1 - \pi\left(z^* - v, v\right)}{1 - \bar{\pi}} \hat{\beta}\left(z^* - v\right) m\left(v\,|\, z^*\right) dv
\end{aligned}
$$

where

$$
\bar{\pi} = \int \pi\left(z^*, v\right) m\left(v\,|\, z^*\right) dv
$$

$$
\hat{\beta}\left(z\right) = \lim_{\tilde{z}_{i0t} \to z} \frac{\partial \Pr\left(L_{it} = 1\,|\, \tilde{z}_{i0t}\right)}{\partial \tilde{z}_{i0t}} - \lim_{\tilde{z}_{i0t} \to z} \frac{\partial \Pr\left(L_{it} = 1\,|\, \tilde{z}_{i0t}\right)}{\partial \tilde{z}_{i0t}}
$$

That is, $\bar{\pi}$ is the probability that $p_{it} = 0$ among individuals with $z_i^{60} = z^*$ and $\hat{\beta}\left(z\right)$ is a the kink in employment at $z$ that would estimated if desired earnings, $\tilde{z}_{i0t}$, were observed.

Given our assumptions on smoothness in heterogeneity and assuming the only relevant kink in the tax schedule is at $z^*$, we have:

$$
\hat{\beta}\left(z\right) = \left\{ \begin{array}{ll} \beta & \text{if } z = z^* \\ 0 & \text{if } z \neq z^* \end{array} \right.
$$

Here $\beta$ is the kink in employment among those who face a kink, as defined in equation (9). Thus, we have:

$$
\beta^{RKD} = \bar{\pi} \cdot \beta
$$

It follows that if $\pi\left(z, v\right) < 1$, *i.e.* we have measurement error in desired earnings when using $z^{60}$ as a proxy, then we estimate a lower bound on our elasticity when using the formula in equation (9). In other words, the estimated kink among those with age 60 earnings near $z^*$ reflects a weighted average of the subset of individuals who will actually face the kink at ages 63-64 and the subset who will not. Furthermore, when $\pi\left(z, v\right) = 0$, *i.e.* there is no persistence in earnings from age 60 to ages 63-64, we expect no employment kink. Note, we only require that the distribution of $v$ be smooth as a function of $z^{60}$. Thus, we can allow for cases where $v$ is not mean zero, *e.g.* where there is a systematic change in mean desired earnings from age 60 to ages 63-64. Note as well that we can relax the assumption that the only kink in the tax schedule is at $z^*$ and can instead assume that the set of other possible kinks is of measaure zero.

# B  Model Extensions

## B.1  Extension to Two or More Discrete Job Choices

We begin by focusing on two key options within the menu of positive earnings: one at the interior optimum in state 0, $\tilde{z}_{n0}$, and another "next-best" job at an alternative level of earnings, $\tilde{z}_n^{nb}$. This allows for an arbitrary number (whether finite or infinite) of discrete choices that are less preferred than the "next-best" job. The model can also be easily extended to allow for the possibility that multiple earnings levels give the same "next best" utility level as $\tilde{z}_n^{nb}$.

Before showing this result more formally, we briefly illustrate the intuition. Consider an individual whose optimal earnings under a linear tax, i.e. in the absence of the kink, is just above $z^*$. For the moment, suppose this person can either earn just above $z^*$ or exit. Now, we introduce a kink at $z^*$. The effect of the

kink on the average tax rate for the person just above $z^*$ is very vanishingly small. So, if this person now decides to exit, they must have been virtually indifferent between $z^*$ and exiting prior to the kink. Now, lets suppose instead that a part-time job is available at some earnings partway between 0 and $z^*$. Also, suppose that in the absence of the kink, the person continues to earn just above $z^*$, rather than exiting or taking the part-time job. By revealed preference, we learn that the job earning just above $z^*$ is preferred to the part-time job. Furthermore, since we have establised that this person is virtually indifferent between earning just above $z^*$ or exiting, it follows that exiting is also preferred to the part-time job. Thus, when we again introduce the kink, the features of the part-time job are irrelevant for the decision to exit or not. If, in the absence of the kink, exiting is preferred to the part-time job, it will continue to be preferred once the kink is introduced.

Formally, let $v_n^{nb}$ be the utility level associated with the next-best level of earnings:

$$v_n^{nb} \equiv v\left(\tilde{z}_n^{nb} - T\left(\tilde{z}_n^{nb}\right), \tilde{z}_n^{nb}; n\right) \tag{B.18}$$

As it is possible that the kink lowers utility at $\tilde{z}_{n0}$ while leaving utility at $\tilde{z}_n^{nb}$ unaffected, the probability of working in state 1 is one minus the probability that non-employment is preferred to both the earnings level $\tilde{z}_{n0}$ and the earnings level $\tilde{z}_n^{nb}$ of the next-best job:

$$
\begin{aligned}
\Pr\left(\left.z_{n1} > 0\right| \tilde{z}_{n0}\right) &= 1 - \Pr\left(v^0 \geq v_n^{nb} - q_{n1} \text{ and } v^0 \geq v\left(\tilde{z}_{n0} - T_1\left(\tilde{z}_{n0}\right), \tilde{z}_{n0}; n\right) - q_{n1}\right) \\
&= 1 - \Pr\left(\left.v^0 \geq v_n^{nb} - q_{n1}\right| \bar{q}_{n1} < q_{n1}\right) \cdot \Pr\left(\bar{q}_{n1} < q_{n1}\right) \\
&= 1 - \Pr\left(\left.v^0 \geq v_n^{nb} - q_{n1}\right| \bar{q}_{n1} < q_{n1}\right) \cdot \left[1 - G\left(\left.\bar{q}_{n1}\right| n\right)\right] \tag{B.19}
\end{aligned}
$$

The slope of the employment function is now a more complex expression:

$$
\begin{aligned}
\frac{d\Pr\left(\left.z_{n1} > 0\right|\tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} &= -\frac{d\Pr\left(\left.v^0 \geq v_n^{nb} - q_{n1}\right| \bar{q}_{n1} < q_{n1}\right)}{d\tilde{z}_{n0}}\left[1 - G\left(\left.\bar{q}_{n1}\right| n\right)\right] \\
&\quad + \Pr\left(\left.v^0 \geq v_n^{nb} - q_{n1}\right| \bar{q}_{n1} < q_{n1}\right) \cdot \left[\frac{dG\left(\left.\bar{q}_{n1}\right| n\right)}{d\tilde{z}_{n0}}\right] \tag{B.20}
\end{aligned}
$$

We now explore under what conditions this slope reduces to that of our earlier model in Section 2.5, in which intensive margin earnings in state 1 are constrained at their state 0 level. We will show in general that this is true for those with state 0 earnings below $z^*$. Next, we show that for those with state 0 earnings just above $z^*$, the slope is likewise unaffected relative to the model in Section 2.5, as long as the next-best job offers a level of earnings that is discretely different than the new interior optimum $z^*$.

Consider individuals earning below $z^*$ in state 0. We first focus on the term $\Pr\left(\left.v^0 \geq v_n^{nb} - q_{n1}\right|\bar{q}_{n1} < q_{n1}\right)$. We can show the following for the agents in this set for whom $\bar{q}_{n1} < q_{n1}$:

$$
\begin{aligned}
v^0 &> v\left(\tilde{z}_{n0} - T_1\left(\tilde{z}_{n0}\right), \tilde{z}_{n0}; n\right) - q_{n1} \\
&= v\left(\tilde{z}_{n0} - T_0\left(\tilde{z}_{n0}\right), \tilde{z}_{n0}; n\right) - q_{n1} \\
&> v_n^{nb} - q_{n1} \tag{B.21}
\end{aligned}
$$

where in the first line we used the fact that $\bar{q}_{n1} < q_{n1}$ and the definition of $\bar{q}_{n1}$. In the second line we used the fact that $T_1\left(\tilde{z}_{n0}\right) = T_0\left(\tilde{z}_{n0}\right)$ for individuals with $\tilde{z}_{n0} < z^*$. In the third line we used the fact that $v\left(\tilde{z}_{n0} - T_0\left(\tilde{z}_{n0}\right), \tilde{z}_{n0}; n\right) \geq v_n^{nb}$ due to revealed preference in state 0. It follows that:

$$
\begin{aligned}
&\Pr\left(\left.v^0 \geq v_n^{nb} - q_{n1}\right| \bar{q}_{n1} < q_{n1}, \tilde{z}_{n0} < z^*\right) = 1 \\
\Rightarrow &\frac{d\Pr\left(\left.v^0 \geq v_n^{nb} - q_{n1}\right| \bar{q}_{n1} < q_{n1}, \tilde{z}_{n0} < z^*\right)}{d\tilde{z}_{n0}} = 0 \tag{B.22}
\end{aligned}
$$

In other words, if an individual with state 0 earnings below $z^*$ prefers the outside option in the absence of the next-best job, she would continue to prefer it in the presence of the next-best job.

Using the results in (B.22), we can simplify the expression in (B.20), for those with $\tilde{z}_{n0} < z^*$:

$$
\begin{aligned}
\frac{d\Pr\left(z_{n1} > 0 \mid \tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} &= \frac{dG\left(\bar{q}_{n1} \mid n\right)}{d\tilde{z}_{n0}} \\
&= g\left(\bar{q}_{n1} \mid n\right) \frac{\partial v\left(\tilde{z}_{n0} - T_1\left(\tilde{z}_{n0}\right), \tilde{z}_{n0}; n\right)}{\partial n} \frac{dn}{d\tilde{z}_{n0}} + \frac{\partial G\left(\bar{q}_{n1} \mid n\right)}{\partial n} \frac{dn}{d\tilde{z}_{n0}}
\end{aligned}
\tag{B.23}
$$

where the second line follows from equation (A.7). Thus, the presence of a menu of discrete options does not affect the results for those with $\tilde{z}_{n0} < z^*$. Intuitively, after the introduction of a kink, the individual's state 0 optimal earnings amount $\tilde{z}_{n0}$ is still available and is the only positive earnings level relevant for extensive margin decisions.

Now consider individuals with $\tilde{z}_{n0} \geq z^*$ and recall that we are ultimately interested in the change in slope of the employment rate at $z^*$. Any change in the slope of the employment function at $z^*$ will depend on the following limit:

$$
\begin{aligned}
\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d\Pr\left(z_{n1} > 0 \mid z_{n0}\right)}{dz_{n0}} &= \left[ -\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d\Pr\left(v^0 \geq v_n^{nb} - q_{n1} \mid \bar{q}_{n1} < q_{n1}\right)}{d\tilde{z}_{n0}} \right] \left[1 - G\left(\bar{q}_{n^*1} \mid n^*\right)\right] \\
&\quad + \left[ \lim_{\tilde{z}_{n0} \to z^{*+}} \Pr\left(v^0 \geq v_n^{nb} - q_{n1} \mid \bar{q}_{n1} < q_{n1}\right) \right] \left( \frac{dG(\bar{q}_{n^*1} \mid n^*)}{d\tilde{z}_{n0}} \right)
\end{aligned}
\tag{B.24}
$$

Note the following:

$$
\begin{aligned}
\lim_{\tilde{z}_{n0} \to z^{*+}} \Pr\left(v^0 \geq v_n^{nb} - q_{n1} \mid \bar{q}_{n1} < q_{n1}\right) &= \lim_{\tilde{z}_{n0} \to z^{*+}} \Pr\left(v^0 \geq v_n^{nb} - q_{n1} \mid v^0 \geq v\left(\tilde{z}_{n0} - T_1\left(\tilde{z}_{n0}\right), \tilde{z}_{n0}; n\right) - q_{n1}\right) \\
&= \Pr\left(v^0 \geq v_{n^*}^{nb} - q_{n^*1} \mid v^0 \geq v\left(z^* - T_0\left(z^*\right), z^*; n^*\right) - q_{n^*1}\right) \\
&= 1
\end{aligned}
\tag{B.25}
$$

In the second line, we used the fact that $\lim_{\tilde{z}_{n0} \to z^*} \tilde{z}_{n0} = z^*$ and $T_1\left(z^*\right) = T_0\left(z^*\right)$, and the final line follows from revealed preference: $z^*$ was initially chosen over the next-best job.

We require that in the neighborhood of $z^*$ the earnings level offered at the next-best job be discretely different than that of the state 0 job; thus, we assume that:

$$
\lim_{\tilde{z}_{n0} \to z^*} \tilde{z}_n^{nb} \neq z^*
\tag{B.26}
$$

This assumption rules out alternative jobs that can be made arbitrarily close to the level of state 0 earnings. Intuitively, if this were not so then individuals earning just above $z^*$ in state 0 would be able to replicate intensive margin adjustment, which we have shown smooths out any kink in the employment function that would otherwise exist.

The assumption in (B.26) implies that the limit in (B.25) is reached at some level of state 0 earnings strictly above $z^*$. That is, as we approach $z^*$ from above, the probability that preferring the outside option without an alternative job implies preferring it in the presence of the next-best job plateaus at 1 at some point before reaching $z^*$. Thus, we have:

$$
\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d\Pr\left(v^0 \geq v_n^{nb} - q_{n1} \mid \bar{q}_{n1} < q_{n1}\right)}{d\tilde{z}_{n0}} = 0
\tag{B.27}
$$

As before, (B.25) and (B.27) can be used to simplify (B.24) as follows:

$$
\begin{aligned}
\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right)}{d\tilde{z}_{n0}} &= \lim_{\tilde{z}_{n0} \to z^{*+}} \frac{dG\left(\bar{q}_{n1} \middle| n\right)}{d\tilde{z}_{n0}} \\
&= g\left(\bar{q}_{n*1} \middle| n^*\right)\left[-d\tau \cdot \lambda_{n^*} + \frac{\partial v\left(\tilde{z}_{n*0} - T_1\left(\tilde{z}_{n*0}\right), \tilde{z}_{n*0}; n^*\right)}{\partial n} \frac{dn}{d\tilde{z}_{n0}}\right] \\
&\quad + \frac{dG\left(\bar{q}_{n*1} \middle| n^*\right)}{dn} \frac{dn}{d\tilde{z}_{n0}}
\end{aligned}
\tag{B.28}
$$

where the second line follows from equation (A.7). Combining (B.28) and (B.23), we have the same result as equation (8) of Section 2.5, *i.e.* our earlier model with $\tilde{z}_{n1} = \tilde{z}_{n0}$. Note that this result features as a special case the scenario in which the agent has the choice among a full-time job at $\tilde{z}_{n0}$, a part-time job as some lower level of earnings, or not working. Furthermore, this result can be extended beyond a setting with multiple, discrete earnings options to allow for some continuity in the choice set, as long as the options all strictly differ from $z^*$ among those with state 0 earnings near $z^*$.

## B.2 Jointly Estimating the Structural Elasticity among Constrained Types and $\pi_B^*$

In the model presented in Section 2.7, we may wish to estimate the structural elasticity among constrained types, $\eta_B^*$. We may be able to use data on extensive margin responses between states 0 and 1 along with evidence on intensive margin responses in state 1 to perform this decomposition. In particular, we continue to draw on the kink in employment in state 1. We also use the amount of bunching in the second period, which is related to intensive margin response among those who are not constrained. In addition, we estimate a second kink, this time in the average tax rate in state 1. The idea is that we have an analytical expression for the kink under complete friction. The extent to which the observed kink in the averate tax rate deviates from that quantity is a function of the share of the sample that faces intensive margin frictions. Finally, this method will require more restrictive assumptions on the underlying primitives, as explained below.

For notational convenience, define the set $R = \{n \mid z^* < z_{n,0} < z^* + \triangle z^*\}$ as the set of individuals in state 0 with earnings in the range that bunches under a kink in the absence of intensive margin frictions. Define $N_{R,0}$ as the number of individuals in this range in state 0. As before, denote Type $A$ earners as those who can adjust on the intensive margin and Type $B$ earners as those who cannot, in state 1. Define $N_{RA,0}$ and $N_{RB,0}$ as the number of Type $A$ and Type $B$ earners in the set $R$, respectively. It follows that:

$$N_{R,0} = N_{RA,0} + N_{RB,0}$$

Similarly, define $N_{R,1}$ as the number of individuals in the set $R$ that are still employed in state 1. That is $\{n \mid z^* < z_{n,0} < z^* + \triangle z^*, z_{n,1} > 0\}$. Again, define $N_{RA,1}$ and $N_{RB,1}$ as the number of Type $A$ and Type $B$ earners in the set $R$ that are also employed in state 1. Again, we have:

$$N_{R,1} = N_{RA,1} + N_{RB,1}$$

Finally, define $N_{Bunch}$ as the number of individuals in the set $R$ that bunch in state 1. Note that:

$$N_{Bunch} = N_{RA,1}$$

We can show the following:

$$
\begin{aligned}
\Pr\left(z_{n,1} > 0 \middle| z^* < z_{n,0} < z^* + \triangle z^*\right) &= \Pr\left(z_{n,1} > 0 \middle| n \in R\right) \\
&= \left(1 - \pi_B^*\right)\Pr\left(z_{n,1} > 0 \middle| n \in R, A\right) \\
&\quad + \pi_B^* \Pr\left(z_{n,1} > 0 \middle| n \in R, B\right)
\end{aligned}
$$

Define $N_0$ as the total number of people in the labor force in state 0 and define $N_1$ as the number of these people also in the labor force in state 1. If we define $B$ as the share of all such earners bunching in state 1,

then we have:

$$
\begin{aligned}
B &= \frac{N_{Bunch}}{N_1} \\
&= \frac{N_{RA,1}}{N_1} \\
&= \frac{N_{RA,1}}{N_{RA,0}} \frac{N_{RA,0}}{N_{R,0}} \frac{N_{R,0}}{N_0} \frac{N_0}{N_1} \\
&= \frac{\Pr\left(z_{n,1} > 0 \mid n \in R, A\right)\left(1 - \pi_B^*\right) \times \Pr\left(z^* < z_{n,0} < z^* + \triangle z^*\right)}{\Pr\left(z_{n,1} > 0 \mid z_{n,0} > 0\right)}
\end{aligned}
$$

In addition, if we assume that Type $A$ and Type $B$ individuals have the same preferences and differ only in the ability to adjust to $z^*$, then the only difference in employment exit between the two groups is due to the lack of intensive margin adjustment on the part of the Type $B$ agents. In this case, we can show the following:

$$
\begin{aligned}
\Pr\left(z_{n,1} > 0 \mid n \in R, B\right) &= \int_{z^*}^{z^* + \triangle z^*} \frac{d\Pr\left(z_{n,1} > 0 \mid z_{n,0} = \zeta, B\right)}{dz_{n,0}} h_0(\zeta)\, d\zeta \\
&= \int_{z^*}^{z^* + \triangle z^*} \left[\frac{d\Pr\left(z_{n,1} > 0 \mid z_{n,0} = \zeta, A\right)}{dz_{n,0}} + \lambda_n \left[(1 - \tau_1) + \frac{v_z}{v_c}\right] g\left(\bar{q}_{n,1} \mid n\right)\right] h_0(\zeta)\, d\zeta \\
&= \Pr\left(z_{n,1} > 0 \mid n \in R, A\right) + \int_{z^*}^{z^* + \triangle z^*} \left[\lambda_n \left[(1 - \tau_1) + \frac{v_z}{v_c}\right] g\left(\bar{q}_{n,1} \mid n\right)\right] h_0(\zeta)\, d\zeta,
\end{aligned}
$$

where the second line follows from equation (A.7). We will use a first-order approximation, assuming $\lambda_n \left[(1 - \tau_1) + \frac{v_z}{v_c}\right] g\left(\bar{q}_{n,1} \mid n\right)$ is constant in the set $R$. Then we have:

$$
\begin{aligned}
\int_{z^*}^{z^* + \triangle z^*} \left[\lambda_n \left[(1 - \tau_1) + \frac{v_z}{v_c}\right] g\left(\bar{q}_{n,1} \mid n\right)\right] h_0(\zeta)\, d\zeta &\approx -d\tau \cdot \lambda_{n^*} \cdot g\left(\bar{q}_{n^*,1} \mid n = n^*\right) \int_{z^*}^{z^* + \triangle z^*} h_0(\zeta)\, d\zeta \\
&= \frac{\beta^{RKD}}{\pi_B^*} \cdot \Pr\left(z^* < z_{n,0} < z^* + \triangle z^*\right)
\end{aligned}
$$

Generally, $\lambda_n$ and $v_z/v_c$ are decreasing over this range. If $g\left(\bar{q}_{n,1} \mid n\right)$ is also weakly decreasing, then we will be overstating the difference between $\Pr\left(z_{n,1} > 0 \mid n \in R, B\right)$ and $\Pr\left(z_{n,1} > 0 \mid n \in R, A\right)$.

Now, consider estimating the following:

$$
\lim_{z_{n,0} \downarrow z^*} \frac{d\mathbb{E}\left[1 - [T_1(z_{n,1}) - T_1(0)]/z_{n,1} \mid z_{n,0}, z_{n,1} > 0\right]}{dz_{n,0}} - \lim_{z_{n,0} \uparrow z^*} \frac{d\mathbb{E}\left[1 - [T_1(z_{n,1}) - T_1(0)]/z_{n,1} \mid z_{n,0}, z_{n,1} > 0\right]}{dz_{n,0}}
$$

which is the difference at the exempt amount $z^*$ in the slope of the average net-of-tax rate in state 1, as a function of state 0 earnings, among the set of individuals who do not exit employment between states 0 and 1. Note that since Type $A$ individuals bunch, the average tax rate is constant for this group. Thus, the difference in the slope will be zero for this group. For the Type $B$ individuals, the difference will be:

$$
\begin{aligned}
\left[\frac{d\left(1 - [T_1(z) - T_1(0)]/z\right)}{dz} - \frac{d\left(1 - [T_0(z) - T_0(0)]/z\right)}{dz}\right]\Bigg|_{z=z^*} &= -\frac{T_1'(z^*)}{z^*} + \frac{T_1(z^*)}{(z^*)^2} + \frac{T_0'(z^*)}{z^*} - \frac{T_0(z^*)}{(z^*)^2} \\
&= -\frac{T_1'(z^*) - T_0'(z^*)}{z^*} \\
&= -\frac{d\tau}{z^*}
\end{aligned}
$$

As a result, the average in the difference in slopes for the total group will be:

$$
\begin{aligned}
\beta^{ATR} &= \frac{N_{RA,1}}{N_{R,1}} \cdot 0 + \frac{N_{RB,1}}{N_{R,1}} \cdot \left(-\frac{d\tau}{z^*}\right) \\
&= \frac{N_{R,0}}{N_{R,1}} \frac{N_{RB,0}}{N_{R,0}} \frac{N_{RB,1}}{N_{RB,0}} \cdot \left(-\frac{d\tau}{z^*}\right) \\
&= \frac{\Pr\left(z_{n,1} > 0 \mid n \in R, B\right) \pi_B^*}{\Pr\left(z_{n,1} > 0 \mid z^* < z_{n,0} < z^* + \triangle z^*\right)} \left(-\frac{d\tau}{z^*}\right)
\end{aligned}
$$

We thus have four equations:

$$
\Pr\left(z_{n,1} > 0 \mid z^* < z_{n,0} < z^* + \triangle z^*\right) = (1 - \pi_B^*)\Pr\left(z_{n,1} > 0 \mid n \in R, A\right) + \pi_B^* \Pr\left(z_{n,1} > 0 \mid n \in R, B\right)
$$

$$
B = \frac{\Pr\left(z_{n,1} > 0 \mid n \in R, A\right)(1 - \pi_B^*) \times \Pr\left(z^* < z_{n,0} < z^* + \triangle z^*\right)}{\Pr\left(z_{n,1} > 0 \mid z_{n,0} > 0\right)}
$$

$$
\beta^{RKD} = \frac{\Pr\left(z_{n,1} > 0 \mid n \in R, B\right) - \Pr\left(z_{n,1} > 0 \mid n \in R, A\right)}{\Pr\left(z^* < z_{n,0} < z^* + \triangle z^*\right) / \pi_B^*}
$$

$$
\beta^{ATR} = \frac{\Pr\left(z_{n,1} > 0 \mid n \in R, B\right) \pi_B^*}{\Pr\left(z_{n,1} > 0 \mid z^* < z_{n,0} < z^* + \triangle z^*\right)} \left(-\frac{d\tau}{z^*}\right)
$$

and four unknowns: $\triangle z^*$, $\pi_B^*$, $\Pr\left(z_{n,1} > 0 \mid n \in R, A\right)$ and $\Pr\left(z_{n,1} > 0 \mid n \in R, B\right)$. We do not have a closed form solutions for either $\Pr\left(z_{n,1} > 0 \mid z^* < z_{n,0} < z^* + \triangle z^*\right)$ or $\Pr\left(z^* < z_{n,0} < z^* + \triangle z^*\right)$.

However, if we estimate a flexible polynomial for the employment rate and for the density in state 0, we can numerically solve for $\pi_B^*$. This can be combined with $\hat{\eta}$ to recover $\eta_B^*$. However, this requires additional assumptions relative to our non-parametric estimate of the observed elasticity, which is a non-parametric lower bound on the structural elasticity; for example, as noted above, to recover $\eta_B^*$ we would need to assume that types $A$ and $B$ have identical preferences.

## B.3 Fully Dynamic Extension of the Model

In this section we briefly demonstrate under what conditions our results continue to hold once our model is extended to a multi-period setting with forward-looking agents. We again have two states of the world, state 0 and state 1. In this multi-period model, the tax schedule is the same across the two states for periods $1, \ldots, t-1$. However, in period $t$, the tax schedule is linear in state 0 and has a convex kink at $z^*$ in state 1. The tax schedules are once again the same across the two states during periods $t+1, \ldots, T$. (For simplicity we assume here that the tax schedules across the two states are once again the same during periods $t+1, \ldots, T$, but the model can be easily extended to assume that in each of these periods the tax schedule is linear in state 0 and has a kink at $z^*$ in state 1.)

We assume that preferences and the economic environment yield a dynamic programming problem as follows. In each period, individuals maximize:

$$
u_t\left(c_{njt}, z_{njt}; n\right) = v\left(c_{njt}, z_{njt}; n\right) - q_{njt} \cdot 1\left\{z_{njt} > 0\right\} + V_t\left(A_{njt}, z_{njt}; n\right) \tag{B.29}
$$

subject to a dynamic budget constraint:

$$
c_{njt} = (1 + r_{t-1}) A_{nj,t-1} + z_{njt} - T_j\left(z_{njt}\right) - A_{njt} \tag{B.30}
$$

where $A_{njt}$ is the level of assets at the end of period $t$. The value function for the next period, $V_t\left(\cdot\right)$, may depend on the level of assets passed forward and potentially the level of earnings in the current period. For example, working today may have some effect on the choice set in the next period.

We once again index individuals by their counterfactual earnings in period $t$ in state 0, and we focus on the probability of having positive earnings in period $t$ in state 1, conditional on the counterfactual earnings level in state 0: $\Pr\left(z_{n1t} > 0 \middle| \tilde{z}_{n0t}\right)$. In addition to the assumptions we have made above in Section 2.4, we assume that the value function $V_t\left(\cdot\right)$ is $C^1$ in $A$, $z$, and $n$. Agents choose $c$, $z$, and $A$ to maximize utility. The outside value of not working in period $t$, $v^0\left(A_{nj,t-1}\right)$, depends on the current level of assets and includes the continuation value of future periods. Finally, the distribution of fixed costs of working, $G\left(q \middle| n, t\right)$, now depends on the time period as well.

The first-order conditions when earnings are positive are now:

$$
\begin{aligned}
v_z + V_z &= -\lambda\left(1 - T'\left(z\right)\right) \\
v_c &= V_A = \lambda
\end{aligned}
\tag{B.31}
$$

where $\lambda$ is the marginal utility of wealth. Using these conditions, we can show that there will still be bunching in response to a kink among those who can adjust on the intensive margin. As before, individuals will work if the utility conditional on working exceeds that of not working:

$$
v\left(\tilde{c}_{njt}, \tilde{z}_{njt}; n\right) + V_t\left(\tilde{A}_{njt}, \tilde{z}_{njt}; n\right) - v^0\left(A_{nj,t-1}\right) - q_{njt} > 0
\tag{B.32}
$$

where the "$\sim$" denotes optimal levels conditional on working. The probability of working in period 1 is still:

$$
\Pr\left(z_{n1t} > 0 \middle| \tilde{z}_{n0t}\right) = G\left(\bar{q}_{n1t} \middle| n, t\right)
\tag{B.33}
$$

where now:

$$
\bar{q}_{n1t} \equiv v\left(\tilde{c}_{n1t}, \tilde{z}_{n1t}; n\right) + V_t\left(\tilde{A}_{n1t}, \tilde{z}_{n1t}; n\right) - v^0\left(A_{n1,t-1}\right)
\tag{B.34}
$$

We now show under what conditions our main results still hold in this dynamic setting. First, the slope of the employment rate will still be:

$$
\frac{d\Pr\left(z_{n1t} > 0 \middle| \tilde{z}_{n0t}\right)}{d\tilde{z}_{n0t}} = g\left(\bar{q}_{n1t} \middle| n, t\right)\frac{d\bar{q}_{n1t}}{d\tilde{z}_{n0t}} + \frac{dG\left(\bar{q}_{n1t} \middle| n, t\right)}{dn}\frac{dn}{d\tilde{z}_{n0t}}
\tag{B.35}
$$

We will have a new expression for the first term on the right of equation (B.35). After substituting for $\tilde{c}_{n1t}$ in (B.34) using the dynamic budget constraint in (B.30) we have:

$$
\begin{aligned}
\frac{d\bar{q}_{n1t}}{d\tilde{z}_{n0t}} &= \left[\frac{\partial v\left(\left(1 + r_{t-1}\right)\tilde{A}_{n1,t-1} + \tilde{z}_{n1t} - T_1\left(\tilde{z}_{n1t}\right) - \tilde{A}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial \tilde{z}_{n1t}} + \frac{\partial V_t\left(\tilde{A}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial \tilde{z}_{n1t}}\right]\frac{d\tilde{z}_{n1t}}{d\tilde{z}_{n0t}} \\
&\quad + \left[\frac{\partial v\left(\left(1 + r_{t-1}\right)\tilde{A}_{n1,t-1} + \tilde{z}_{n1t} - T_1\left(\tilde{z}_{n1t}\right) - \tilde{A}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial \tilde{A}_{n1t}} + \frac{\partial V_t\left(\tilde{A}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial \tilde{A}_{n1t}}\right]\frac{d\tilde{A}_{n1t}}{d\tilde{z}_{n0t}} \\
&\quad + \left[\frac{\partial v\left(\left(1 + r_{t-1}\right)\tilde{A}_{n1,t-1} + \tilde{z}_{n1t} - T_1\left(\tilde{z}_{n1t}\right) - \tilde{A}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial n} + \frac{\partial V_t\left(\tilde{A}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial n}\right]\frac{dn}{d\tilde{z}_{n0t}} \\
&\quad + \left[\frac{\partial v\left(\left(1 + r_{t-1}\right)\tilde{A}_{n1,t-1} + \tilde{z}_{n1t} - T_1\left(\tilde{z}_{n1t}\right) - \tilde{A}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial \tilde{A}_{n1,t-1}} - \frac{\partial v^0\left(\tilde{A}_{n1,t-1}\right)}{\partial A_{n1,t-1}}\right]\frac{dA_{n1,t-1}}{d\tilde{z}_{n0t}} \\
&= \left[\left(1 - T_1'\left(z\right)\right)v_c + v_z + V_z\right]\frac{d\tilde{z}_{n1t}}{d\tilde{z}_{n0t}} + \left[-v_c + V_A\right]\frac{d\tilde{A}_{n1t}}{d\tilde{z}_{n0t}} + \left[v_n + V_n\right]\frac{dn}{d\tilde{z}_{n0t}} + \left[\left(1 + r_{t-1}\right)v_c - v_A^0\right]\frac{dA_{n1,t-1}}{d\tilde{z}_{n0t}}
\end{aligned}
\tag{B.36}
$$

For those with $\tilde{z}_{n0t} < z^*$ or $\tilde{z}_{n0t} > z^* + \Delta z^*$, we can use the first-order conditions in (B.31) to show that the first term in (B.36) equals zero when agents are able to adjust on the intensive margin as in Section 2.4. For

those with $z^* \leq \tilde{z}_{n0t} \leq z^* + \Delta z^*$, the first term in (B.36) equals zero: $d\tilde{z}_{n1t}/d\tilde{z}_{n0t} = 0$ since $\tilde{z}_{n1t} = z^*$ for everyone in this latter set due to bunching. Additionally, the second term in (B.36) equals zero for everyone, due to the first-order condition in (B.31) for saving.

Thus, when agents are able to adjust on both the intensive and extensive margin we have:

$$
\begin{aligned}
\frac{d\Pr\left(z_{n1t} > 0 \mid \tilde{z}_{n0t}\right)}{d\tilde{z}_{n0t}} &= g\left(\overline{q}_{n1t} \mid n, t\right)\left(\left[\frac{\partial v\left(\tilde{c}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial n} + \frac{\partial V_t\left(\tilde{A}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial n}\right]\frac{dn}{d\tilde{z}_{n0t}}\right. \\
&\quad \left. + \left[(1 + r_{t-1})\frac{\partial v\left(\tilde{c}_{n1t}, \tilde{z}_{n1t}; n\right)}{\partial \tilde{c}_{n1t}} - \frac{\partial v^0\left(\tilde{A}_{n1,t-1}\right)}{\partial A_{n1,t-1}}\right]\frac{dA_{n1,t-1}}{d\tilde{z}_{n0t}}\right) + \frac{dG\left(\overline{q}_{n1t} \mid n, t\right)}{dn}\frac{dn}{d\tilde{z}_{n0t}}
\end{aligned}
$$

$$\text{(B.37)}$$

Note that $n$, $\tilde{q}_{n,1}$, $\tilde{z}_{n,1}$, $\tilde{A}_{n,1}$, and $T_1(\cdot)$ are all continuous in $\tilde{z}_{n0t}$ at $z^*$. Furthermore, our smoothness assumptions imply that $g\left(\cdot\right)$, $\partial v/\partial n$, $\partial V/\partial n$, $\partial n/\partial \tilde{z}_{n0t}$, and $\partial G\left(\overline{q}_{n,1} \mid n, t\right)$ are likewise continuous in their arguments. We additionally assume that $\partial A_{n1,t-1}/\partial \tilde{z}_{n0t}$ is continuous at $z^*$; we discuss below the conditions under which this assumption holds and argue that they are satisfied in our setting. Given these assumptions, our original result follows when there are intensive margin adjustments:

$$
\lim_{\tilde{z}_{n0t} \to z^{*+}} \frac{d\Pr\left(z_{n1t} > 0 \mid \tilde{z}_{n0t}\right)}{d\tilde{z}_{n0t}} - \lim_{\tilde{z}_{n0t} \to z^{*-}} \frac{d\Pr\left(z_{n1t} > 0 \mid \tilde{z}_{n0t}\right)}{d\tilde{z}_{n0t}} = 0
$$

$$\text{(B.38)}$$

We now turn to the case in which we make the same assumptions, except that individuals are not able to adjust on the intensive margin, as in Section 2.5. We now have:

$$
\begin{aligned}
\frac{d\overline{q}_{n1t}}{d\tilde{z}_{n0t}} &= \left[\frac{\partial v\left((1 + r_{t-1})\tilde{A}_{n1,t-1} + \tilde{z}_{n0t} - T_1(\tilde{z}_{n0t}) - \tilde{A}_{n1t}, \tilde{z}_{n0t}; n\right)}{\partial \tilde{z}_{n0t}} + \frac{\partial V_t\left(\tilde{A}_{n1t}, \tilde{z}_{n0t}; n\right)}{\partial \tilde{z}_{n0t}}\right] \\
&\quad + \left[\frac{\partial v\left((1 + r_{t-1})\tilde{A}_{n1,t-1} + \tilde{z}_{n0t} - T_1(\tilde{z}_{n0t}) - \tilde{A}_{n1t}, \tilde{z}_{n0t}; n\right)}{\partial \tilde{A}_{n1t}} + \frac{\partial V_t\left(\tilde{A}_{n1t}, \tilde{z}_{n0t}; n\right)}{\partial \tilde{A}_{n1t}}\right]\frac{d\tilde{A}_{n1t}}{d\tilde{z}_{n0t}} \\
&\quad + \left[\frac{\partial v\left((1 + r_{t-1})\tilde{A}_{n1,t-1} + \tilde{z}_{n0t} - T_1(\tilde{z}_{n0t}) - \tilde{A}_{n1t}, \tilde{z}_{n0t}; n\right)}{\partial n} + \frac{\partial V_t\left(\tilde{A}_{n1t}, \tilde{z}_{n0t}; n\right)}{\partial n}\right]\frac{dn}{d\tilde{z}_{n0t}} \\
&\quad + \left[\frac{\partial v\left((1 + r_{t-1})\tilde{A}_{n1,t-1} + \tilde{z}_{n0t} - T_1(\tilde{z}_{n0t}) - \tilde{A}_{n1t}, \tilde{z}_{n0t}; n\right)}{\partial \tilde{A}_{n1,t-1}} - \frac{\partial v^0\left(\tilde{A}_{n1,t-1}\right)}{\partial A_{n1,t-1}}\right]\frac{dA_{n1,t-1}}{d\tilde{z}_{n0t}} \\
&= \left[(1 - T_1'(z))v_c + v_z + V_z\right] + \left[-v_c + V_A\right]\frac{d\tilde{A}_{n1t}}{d\tilde{z}_{n0t}} + \left[v_n + V_n\right]\frac{dn}{d\tilde{z}_{n0t}} + \left[(1 + r_{t-1})v_c - v_A^0\right]\frac{dA_{n1,t-1}}{d\tilde{z}_{n0t}}
\end{aligned}
$$

$$\text{(B.39)}$$

where the primary difference from (B.36) is that earnings are fixed at $\tilde{z}_{n0t}$. We still have the second term dropping out of the expression in (B.39), as assets, $\tilde{A}_{n1t}$, are optimally chosen even when earnings are fixed. Furthermore, for those with $\tilde{z}_{n0t} \leq z^*$ we have $T_1 = T_0$, and thus the first term also drops out among this group due to the envelope theorem. Note that for those with $\tilde{z}_{n0t} \geq z^*$ we can rewrite:

$$
(1 - T_1'(z))v_c + v_z + V_z = \lambda\left[(1 - \tau_1) + \frac{v_z + V_z}{v_c}\right]
$$

$$\text{(B.40)}$$

59

Similar to our less dynamic model above in Section 2.5, we therefore have:

$$
\frac{d\Pr\left(z_{n1t}>0\middle|\tilde{z}_{n0t}\right)}{d\tilde{z}_{n0t}} =
\begin{cases}
\begin{aligned}
& g\left(\overline{q}_{n1t}\middle|n,t\right)\left([v_n+V_n]\frac{dn}{d\tilde{z}_{n0t}}\right. \\
& \left.+\left[(1+r_{t-1})v_c-v_A^0\right]\frac{dA_{n1,t-1}}{d\tilde{z}_{n0t}}\right)+\frac{dG(\overline{q}_{n1t}|n,t)}{dn}\frac{dn}{d\tilde{z}_{n0t}}
\end{aligned}
& \text{if } \tilde{z}_{n0t}<z^* \\[2em]
\begin{aligned}
& g\left(\overline{q}_{n1t}\middle|n,t\right)\left(\lambda_n\left[(1-\tau_1)+\frac{v_z+V_z}{v_c}\right]+[v_n+V_n]\frac{dn}{d\tilde{z}_{n0t}}\right. \\
& \left.+\left[(1+r_{t-1})v_c-v_A^0\right]\frac{dA_{n1,t-1}}{d\tilde{z}_{n0t}}\right)+\frac{dG(\overline{q}_{n1t}|n,t)}{dn}\frac{dn}{d\tilde{z}_{n0t}}
\end{aligned}
& \text{if } \tilde{z}_{n0t}\geq z^*
\end{cases}
\tag{B.41}
$$

We note the following limit, making use of the first-order condition in (B.31):

$$
\lim_{\tilde{z}_{n0t}\to z^{*+}}\frac{v_z\left((1+r_{t-1})\tilde{A}_{n1,t-1}+\tilde{z}_{n0t}-T_1\left(\tilde{z}_{n0t}\right)-\tilde{A}_{n1t},\tilde{z}_{n0t};n\right)+V_z\left(\tilde{A}_{n1t},\tilde{z}_{n0t};n\right)}{v_c\left((1+r_{t-1})\tilde{A}_{n1,t-1}+\tilde{z}_{n0t}-T_1\left(\tilde{z}_{n0t}\right)-\tilde{A}_{n1t},\tilde{z}_{n0t};n\right)}=-(1-\tau_0)\tag{B.42}
$$

Maintaining our smoothness assumptions for this section, we thus have our original result when earnings in state 1 are constrained ($\tilde{z}_{n1t}=\tilde{z}_{n0t}$):

$$
\begin{aligned}
\lim_{\tilde{z}_{n0t}\to z^{*+}}\frac{d\Pr\left(z_{n1t}>0\middle|\tilde{z}_{n0t}\right)}{d\tilde{z}_{n0t}}-\lim_{\tilde{z}_{n0t}\to z^{*-}}\frac{d\Pr\left(z_{n1t}>0\middle|\tilde{z}_{n0t}\right)}{d\tilde{z}_{n0t}} &= \lim_{\tilde{z}_{n0t}\to z^{*+}}g\left(\overline{q}_{n1t}\middle|n,t\right)\cdot\lambda_n\left[(1-\tau_1)+\frac{v_z+V_z}{v_c}\right] \\
&= -d\tau\cdot\lambda_{n^*}\cdot g\left(\overline{q}_{n^*1t}\middle|n^*,t\right)
\end{aligned}
\tag{B.43}
$$

As in Appendix B.1, this result easily generalizes to the case of multiple, discrete job options away from $z^*$. The kink in this case can be used to calculate an extensive margin elasticity, as in equation (9).

Returning to our assumption that $\partial A_{n1,t-1}/\partial\tilde{z}_{n0t}$ is continuous at $\tilde{z}_{n0t}=z^*$, we view this as a natural assumption in our setting. First, if agents act as if the tax change is unanticipated, then this is assumption holds: in this case the slope of expected lifetime wealth is continuous at the kink. Many contexts will feature unanticipated changes in taxes. In our empirical application, Gelber *et al.* (2013) show that prior to being subject to the AET, individuals do not appear to act as if they anticipate the later imposition of the AET (consistent with the myopia suggested in Gelber, Isen, and Song, 2016). If the tax change were anticipated they would bunch at $z^*$ in anticipation of the later imposition of the kink (due to the fixed cost of adjustment), but empirically we do not observe such behavior. Thus, they do not appear to anticipate its imposition, so our empirical application appears consistent with this case. Moreover, in our setting, we test for kinks in predetermined or placebo outcomes, including measures of the employment rate at ages 56, 57, 60, and 61, as well as demographic variables. We show in Section 6.3 that predetermined variables as a function of age 60 earnings do not exhibit systematic discontinuities in their slopes at the exempt amount, consistent with our assumption of continuity in $\partial A_{n1,t-1}/\partial\tilde{z}_{n0t}$ at $\tilde{z}_{n0t}=z^*$.

Second, even if some individuals in our context acted as if the imposition of the AET were anticipated, if the AET is also actuarially fair and other sources of lifetime wealth are also smooth, then $\partial A_{n1,t-1}/\partial\tilde{z}_{n0t}$ should be smooth at $\tilde{z}_{n0t}=z^*$. It is commonly understood that the AET is approximately actuarially fair (*e.g.* Diamond and Gruber, 1999), and thus this assumption should approximately hold. Note that even if the AET is actuarially fair on average—but better than actuarially fair in expectation for some types and worse for others—it is possible that $\partial A_{n1,t-1}/\partial\tilde{z}_{n0t}$ is continuous at $\tilde{z}_{n0t}=z^*$, even though we also we observe a non-zero substitution effect of the incentives created by the AET when individuals are later subject to it. For example, if the AET is worse than actuarially fair for those who are particularly responsive to the substitution incentives created by the AET, then we could see a reduction in earnings or employment due to a bunching or extensive margin response to the substitution incentives created by the AET once individuals have claimed, even though $\partial A_{n1,t-1}/\partial\tilde{z}_{n0t}$ is continuous at at $\tilde{z}_{n0t}=z^*$.

Third, in the case that the imposition of the tax change is anticipated—which could be consistent with

the data if the AET is actuarially fair on average—we could interpret our estimated elasticity as a marginal-utility-of-wealth-constant elasticity. In this case our paper then can fit into a broader literature estimating Frisch elasticities (see Chetty *et al.*, 2013, for a review), and provides a novel estimate of a Frisch elasticity using transparent variation from the RKD. (Following Chetty *et al.* (2012) we call this extensive margin elasticity, holding the marginal utility of lifetime wealth constant, a Frisch elasticity, while recognizing that extensive margin Frisch elasticities are technically not defined because non-participants do not locate at an interior optimum.)

Fourth, even in the case in which the AET is not actuarially fair, we can estimate an elasticity that represents the response to a parametric shift in the entire wage profile; as Blundell and MaCurdy (1999) and others point out, this is the most relevant for policy evaluation of the impacts of permanent shifts in the entire wage profile and will reflect a combination of income and substitution effects. We have run several numerical simulations to gauge the quantitative importance of anticipatory savings for the continuity of $\partial A_{n1,t-1}/\partial \tilde{z}_{n0t}$ at $\tilde{z}_{n0t} = z^*$, and we find that extensive margin estimates using this method are little affected by a lack of smoothness in $A_{n1,t-1}$.[36] Furthermore, since predetermined variables do not show systematic discontinuities in their slopes as a function of age 60 earnings, this issue again does not appear to affect the validity of the smoothness assumption underlying our empirical design. We view our method as most easily applicable in settings in which the data are consistent with such interpretations of the parameters and the smoothness assumptions appear to be satisfied.

### B.3.1 Explicitly Modeling Actuarial Adjustment in the Context of the Dynamic Model

Thus far, we have abstracted from the impact of actuarial adjustment of future benefits on the behavioral response to the AET. In this section, we explicitly incorporate this feature of the AET and highlight how our estimator is affected. We maintain the setup of the dynamic model above with the following alterations. First, the dynamic program now involves choosing earnings to maximize:

$$u_t\left(c_{njt}, z_{njt}; n\right) = v\left(c_{njt}, z_{njt}; n\right) - q_{njt} \cdot 1\left\{z_{njt} > 0\right\} + V_t\left(A_{njt} + B_{njt}\left(z_{njt}\right), z_{njt}; n\right) \tag{B.44}$$

Relative to our previous case, the value function, $V_t\left(\cdot\right)$, is now a function of total assets, *i.e.* the sum of savings, $A_{njt}$, and the present discounted value of all future OASI benefits, $B_{njt}\left(z_{njt}\right)$. These benefits are potentially affected by the current level of earnings. Additionally, the new dynamic budget constraint is as follows:

$$c_{njt} = \left(1 + r_{t-1}\right)A_{nj,t-1} + z_{njt} - T\left(z_{njt}\right) + b_{jt}\left(z_{njt}\right) - A_{njt} \tag{B.45}$$

where we have now decoupled the flow of OASI benefits from the tax schedule. The tax schedule no longer varies across states and is simply $T\left(z_{njt}\right) = \tau_0 z_{njt}$. The flow of benefits, $b_{jt}\left(z_{njt}\right)$, does potentially vary across states, due to the presence of the AET in state 1. The first-order conditions are now:

$$\begin{aligned} v_z + V_z &= -\lambda\left(1 - T'\left(z\right) + \left[b'\left(z\right) + B'\left(z\right)\right]\right) \\ v_c &= V_A = \lambda \end{aligned} \tag{B.46}$$

where $b'_{jt}\left(z\right)$ is the marginal effect of earning more on the current flow of benefits, *i.e.* the benefit reduction rate (BRR). It takes the following form in state 1:

$$b'\left(z\right) = \begin{cases} 0 & \text{if } z < z^* \\ -d\tau & \text{if } z \geq z^* \end{cases} \tag{B.47}$$

Finally, $B'\left(z\right)$ is the effect of increasing current earnings on the future stream of benefits from OASI. $B'\left(z\right)$ similarly takes the form:

$$B'\left(z\right) = \begin{cases} 0 & \text{if } z < z^* \\ db & \text{if } z \geq z^* \end{cases} \tag{B.48}$$

---

[36]Results from our simulations are available upon request. We continue to recover the extensive margin elasticity in a dynamic model when intensive margin adjustment is constrained and we observe the correct, counterfactual running variable. We estimate an attenuated version of the true elasticity when intensive margin adjustment is constrained and we use lagged earnings as a proxy for the running variable, though our simulations show that the extent of attenuation is slight. If anything, this slight attenuation again strengthens our case that the elasticity is large, as the lower bound we estimate is itself large.

Note that when the adjustment is actuarially fair, we have $db = d\tau$. In this version of the model, we also express the outside utility as a function of savings and OASI benefits:

$$v^0 = v^0 \left( A_{nj,t-1}, B_{nj,t-1} \right) \tag{B.49}$$

Using similar steps as in the basic dynamic model above, we have the following expression for the (potential) kink in the employment rate at $z^*$, when intensive margin earnings are constrained:

$$\lim_{\tilde{z}_{n0t} \to z^{*+}} \frac{d \Pr \left( z_{n1t} > 0 \middle| \tilde{z}_{n0t} \right)}{d\tilde{z}_{n0t}} - \lim_{\tilde{z}_{n0t} \to z^{*-}} \frac{d \Pr \left( z_{n1t} > 0 \middle| \tilde{z}_{n0t} \right)}{d\tilde{z}_{n0t}} = - (d\tau - db) \cdot \lambda_{n^*} \cdot g \left( \bar{q}_{n^*1t} \middle| n^*, t \right)$$

Here we make explicit that an additional necessary condition for finding a kink in the employment rate is that $db \neq d\tau$, either because the AET is not actually fair, or perhaps, because individuals do not pay attention to actuarial adjustment. The sign of $db$ is non-negative, as the actuarial adjustment never reduces future benefits, but $db$ can be smaller than $d\tau$ if adjustment is not full. It can also be larger than $d\tau$ for those with high life expectancies, in which case there will be a positive kink in the slope of the employment function. In our applications above, we have effectively assumed $db = 0$ for illustrative purposes and following previous literature, although our method can easily accomodate alternative assumptions. Since we do find a kink in the employment rate empirically (and also find intensive margin bunching), this assumption is validated.

## B.4 Model Extension: Joint Claiming and Employment Decision

A specific feature of our empirical application to an OASI context—which will not always be true when applying our method for estimating elasticities in other contexts—is that whether or not individuals face a kink is endogenous. That is, if an agent does not claim, she faces the linear budget set. We now extend the model to our particular empirical application, so that the model includes this trade-off between facing a kink or delaying claiming. To model claiming, we introduce additional notation. We model claiming in a somewhat "reduced form" fashion, but our model easily generalizes to the previous dynamic setting, in which we can explicitly incorporate the effects of claiming OASI on the timing and magnitude of OASI benefits in different periods, as well as the resulting effects on wealth and savings. In our case, when $j = 0$ an individual faces a linear tax schedule whether or not she claims. However, when $j = 1$ an individual faces a kinked tax schedule when claiming OASI or a linear schedule when not claiming OASI. We focus on our version of the model in which intensive margin adjustments are constrained, i.e. $\tilde{z}_{n1} \equiv \tilde{z}_{n0}$, and we will determine whether a kink in the probability of working still occurs when claiming is endogenous.

First, we will denote $v_n^0 \equiv v(\tilde{z}_{n0} - T_0(\tilde{z}_{n0}), \tilde{z}_{n0}, n)$ as the flow utility (net of the fixed cost of working) when earning $\tilde{z}_{n0}$, facing a linear tax schedule $T_0(\cdot)$, and claiming in state 0. Next, we denote $v_n^1 \equiv v(\tilde{z}_{n0} - T_1(\tilde{z}_{n0}), \tilde{z}_{n0}, n)$, i.e. the net flow utility when earning $\tilde{z}_{n0}$, facing the kinked tax schedule $T_1(\cdot)$, and claiming in state 1. As before, $v^0$ is the utility received when claiming OASI and not working, in either state 0 or state 1. Next, we specify utility when claiming is delayed. The payoff when not working and not claiming OASI is $v^0 + \theta_n$, in both state 0 and state 1. Likewise, the payoff when working and not claiming OASI is $v_n^0 + \gamma_n$, again in either state 1 or state 0. Thus, the parameters $\theta$ and $\gamma$ capture the relative change in utility when claiming is delayed. These can be considered to capture the income effect of delaying claiming. We leave these parameters relatively unrestricted, and thus they may represent an increase in lifetime wealth when delaying claiming increases lifetime benefits (for example due to the actuarial adjustment being better than actuarially fair), or they may represent a utility decrease for those facing liquidity constraints, for example.

In addition to $n$ we now have three parameters that capture unobserved heterogeneity, $(\gamma, \theta, q)$. We allow these variables to have a relatively unrestricted, joint distribution, which we represent with a conditional, joint density of $\gamma$ and $\theta$, $m(\gamma, \theta | q, n)$ and our previous marginal density of $q$, $g(q|n)$. The joint density of $(\gamma, \theta, q)$ is therefore $m(\gamma, \theta | q, n) \cdot g(q|n)$. We extend our assumption of smoothness in unobserved heterogeneity to the joint distribution of $(\gamma, \theta, q)$, conditional on $n$ .

In State 1, our model will now feature comparisons among four discrete choices: (1) working while claiming, $v_n^1 - q_n$; (2) not working while claiming, $v^0$; (3) working while not claiming $v_n^0 + \gamma_n - q_n$; and

(4) not working while not claiming, $v^0 + \theta_n$. We will additionally denote critical values of our unobservable parameters, which arise when comparing the various discrete options. As before, when comparing working while claiming to not working while claiming, the expression for indifference is $v_n^1 - q_n = v^0$, which implies a critical value for $q$, $\bar{q}_n \equiv v_n^1 - v^0$. Similarly, when comparing working while not claiming to not working while not claiming, the expression for indifference is $v_n^0 + \gamma_n - q_n = v^0 + \theta_n$. We can write this in terms of $\gamma_n$ as $\gamma_n = \theta_n + q_n - (v_n^0 - v^0) = \theta_n + q_n - \bar{q}_n^0$, where $\bar{q}_n^0 \equiv v_n^0 - v^0$ is an analogous critical value for $q$ in state 0 when claiming.

Furthermore, for those who are indifferent between working while claiming and not working while not claiming, we have $v_n^1 - q_n = v^0 + \theta_n$. This implies a critical value for $\theta_n$ : $\theta_n = \bar{q}_n - q_n$. Symmetrically, for those who are indifferent between not working while claiming and working while not claiming, we have $v^0 = v_n^0 + \gamma_n - q_n$, with a critical value for $\gamma_n$ : $\gamma_n = \bar{q}_n^0 - q_n$. Indifference between working while claiming and working while not claiming implies $v_n^1 - q_n = v_n^0 + \gamma_n - q_n$, and a critical value for $\gamma_n$ is: $\gamma_n = v_n^1 - v_n^0 \equiv \Delta v_n$. Finally, indifference between not working while claiming and not working while not claiming implies $v^0 = v^0 + \theta_n$ or $\theta_n = 0$.

Using our previous results, the envelope theorem implies the following:

$$\frac{\partial \bar{q}_n^0}{\partial \tilde{z}_{n0}} = \frac{\partial v_n^0}{\partial \tilde{z}_{n0}} = 0 \tag{B.50}$$

Additionally, we have:

$$\frac{\partial \bar{q}_n}{\partial \tilde{z}_{n0}} = \frac{\partial \Delta v_n}{\partial \tilde{z}_{n0}} = \frac{\partial v_n^1}{\partial \tilde{z}_{n0}} = \begin{cases} \lambda_n \left[ (1 - \tau_0) + \frac{v_z}{v_c} \right] & \text{if } \tilde{z}_{n0} < z^* \\[2ex] \lambda_n \left[ (1 - \tau_1) + \frac{v_z}{v_c} \right] & \text{if } \tilde{z}_{n0} \geq z^* \end{cases} \tag{B.51}$$

### B.4.1 Claiming Decision

We now characterize the claiming probability and derive an expression for a kink in the claiming probability at $\tilde{z}_{n0} = z^*$. The probability of claiming can be expressed as follows:

$$\begin{aligned} \Pr\left( claim | \tilde{z}_{n0} \right) &= \int_{-\infty}^{\bar{q}} \int_{-\infty}^{\bar{q}-q} \int_{-\infty}^{\Delta v} m\left( \gamma, \theta | q, n \right) g\left( q | n \right) d\gamma d\theta dq \\ &+ \int_{\bar{q}}^{\infty} \int_{-\infty}^{0} \int_{-\infty}^{q - \bar{q}^0} m\left( \gamma, \theta | q, n \right) g\left( q | n \right) d\gamma d\theta dq \end{aligned} \tag{B.52}$$

The first term integrates over the values of $q$ for which working while claiming is preferred to not working while claiming, $q \in [-\infty, \bar{q}]$. The next two integrals restrict attention to values of $\theta$ that render not working while not claiming dominated by working while claiming, $i.e.$ $\theta \in [-\infty, \bar{q} - q]$ and values of $\gamma$ that similarly render working while not claiming dominated by working while claiming, $i.e.$ $\gamma \in [-\infty, \Delta v]$. The second term integrates over values of $q$ for which not working while claiming is preferred to working while claiming, $q \in [\bar{q}, \infty]$. Over this range we restrict analysis to $\theta \in [-\infty, 0]$ and $\gamma \in [-\infty, \bar{q} - q]$, $i.e.$ values that render not claiming dominated by not working while claiming.

Consider the slope of $\Pr\left( claim | \tilde{z}_{n0} \right)$, which can now be expressed as:

$$\frac{d\Pr\left( claim | \tilde{z}_{n0} \right)}{d\tilde{z}_{n0}} = \frac{\partial \Pr\left( claim | \tilde{z}_{n0} \right)}{\partial \tilde{z}_{n0}} + \frac{\partial \Pr\left( claim | \tilde{z}_{n0} \right)}{\partial n} \frac{dn}{d\tilde{z}_{n0}} \tag{B.53}$$

Given our assumptions regarding smoothness and our results above, we know that the second term will be

continuous at $\tilde{z}_{n0} = z^*$. We therefore focus on the first term in this expression. Leibniz's rule implies:

$$
\begin{aligned}
\frac{\partial \Pr\left(claim\middle|\tilde{z}_{n0}\right)}{\partial \tilde{z}_{n0}} =\ & \int_{-\infty}^{\overline{q}} \int_{-\infty}^{\overline{q}-q} \frac{\partial \Delta v_n}{\partial \tilde{z}_{n0}} m\left(\Delta v_n, \theta\middle| q, n\right) g\left(q\middle| n\right) d\gamma d\theta dq \\
& + \int_{-\infty}^{\overline{q}} \frac{\partial \overline{q}_n}{\partial \tilde{z}_{n0}}\left[\int_{-\infty}^{\Delta v} m\left(\gamma, \overline{q}-q\middle| q, n\right) d\gamma\right] g\left(q\middle| n\right) dq \\
& + \frac{\partial \overline{q}_n}{\partial \tilde{z}_{n0}}\left[\int_{-\infty}^{0} \int_{-\infty}^{\Delta v} m\left(\gamma, \theta\middle| \overline{q}, n\right) d\gamma d\theta\right] g\left(\overline{q}\middle| n\right) \\
& - \int_{\overline{q}}^{\infty} \int_{-\infty}^{0} \frac{\partial \overline{q}_n^0}{\partial \tilde{z}_{n0}} m\left(q-\overline{q}^0, \theta\middle| q, n\right) g\left(q\middle| n\right) d\theta dq \\
& - \frac{\partial \overline{q}_n}{\partial \tilde{z}_{n0}}\left[\int_{-\infty}^{0} \int_{-\infty}^{\Delta v} m\left(\gamma, \theta\middle| \overline{q}, n\right) d\gamma d\theta\right] g\left(\overline{q}\middle| n\right). \qquad \text{(B.54)}
\end{aligned}
$$

Rearranging terms, and using the fact that $\partial \overline{q}^0 / \partial \tilde{z}_{n0} = 0$ and $\partial \overline{q}/\partial \tilde{z}_{n0} = \partial \Delta v/\partial \tilde{z}_{n0} = \lambda\left[(1-\tau) + \frac{v_z}{v_c}\right]$, we have:

$$
\begin{aligned}
\frac{\partial \Pr\left(claim\middle|\tilde{z}_{n0}\right)}{\partial \tilde{z}_{n0}} =\ & \lambda\left[(1-\tau)+\frac{v_z}{v_c}\right] \times \\
& \left\{\int_{-\infty}^{\overline{q}}\left[\int_{-\infty}^{\overline{q}-q} m\left(\Delta v, \theta\middle| q, n\right) d\theta + \int_{-\infty}^{\Delta v} m\left(\gamma, \overline{q}-q\middle| q, n\right) d\gamma\right] g\left(q\middle| n\right) dq\right\}
\end{aligned}
$$
$$\text{(B.55)}$$

Finally, the kink in the probability of claiming can be expressed as:

$$
\lim_{\tilde{z}_{n0}\to z^{*+}} \frac{\partial \Pr\left(claim\middle|\tilde{z}_{n0}\right)}{\partial \tilde{z}_{n0}} - \lim_{\tilde{z}_{n0}\to z^{*-}} \frac{\partial \Pr\left(claim\middle|\tilde{z}_{n0}\right)}{\partial \tilde{z}_{n0}} = \delta \qquad \text{(B.56)}
$$

where

$$
\begin{aligned}
\delta =\ & \lim_{\tilde{z}_{n0}\to z^{*+}} \lambda\left[(1-\tau)+\frac{v_z}{v_c}\right] \cdot \left\{\int_{-\infty}^{\overline{q}}\left[\int_{-\infty}^{\overline{q}-q} m\left(\Delta v, \theta\middle| q, n\right) d\theta + \int_{-\infty}^{\Delta v} m\left(\gamma, \overline{q}-q\middle| q, n\right) d\gamma\right] g\left(q\middle| n\right) dq\right\} \\
=\ & -d\tau \cdot \lambda_{n^*} \cdot \left\{\int_{-\infty}^{\overline{q}}\left[\int_{-\infty}^{\overline{q}-q} m\left(0, \theta\middle| q, n^*\right) d\theta + \int_{-\infty}^{0} m\left(\gamma, \overline{q}-q\middle| q, n^*\right) d\gamma\right] g\left(q\middle| n^*\right) dq\right\} \qquad \text{(B.57)}
\end{aligned}
$$

and where we used the results from Section 2.5 and the fact that $\lim_{\tilde{z}_{n0}\to z^*} \Delta v_n = 0$.

The two integrals in the expression can be interpreted as joint probabilities. The first is the joint probability that (1) working while claiming is preferred to not working while claiming, (2) working while claiming is preferred to not working while not claiming, and (3) the individual is indifferent between working while claiming and working while not claiming, *i.e.* $\gamma_{n^*} = \Delta v_{n^*} = 0$. The second term is the joint probability that (1) working while claiming is preferred to not working while claiming, (2) working while claiming is preferred to working while not claiming, and (3) the individual is indifferent between working while claiming and not working while not claiming, *i.e.* $\theta_{n^*} = \overline{q}_{n^*} - q_{n^*}$. Note that the probabilities are conditional on $n = n^*$.

Thus, there will be a downward kink in the probability of claiming, which increases with the size of the kink and the size of two key sets of marginal claimants. The first set are on the margin of moving from working while claiming to working while not claiming, and the second set are on the margin of moving from working while claiming to not working while not claiming. Finally, note that the kink only affects claiming

among those for whom working while claiming is optimal. Therefore, the only relevant shifting is from working while claiming to either state of not claiming.

### B.4.2 Extensive Margin Response with Endogenous Claiming

We now consider the extensive margin choice of whether to work, allowing for an endogenous claiming response. The probability of having positive earnings in state 1 is now:

$$
\Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right) = \int_{-\infty}^{\bar{q}} \left[ \int_{-\infty}^{\bar{q}-q} \int_{-\infty}^{\infty} m\left(\gamma, \theta \middle| q, n\right) d\gamma d\theta + \int_{\bar{q}-q}^{\infty} \int_{\theta+q-\bar{q}^0}^{\infty} m\left(\gamma, \theta \middle| q, n\right) d\gamma d\theta \right] g\left(q \middle| n\right) dq
$$
$$
+ \int_{\bar{q}}^{\infty} \left[ \int_{-\infty}^{0} \int_{q-\bar{q}^0}^{\infty} m\left(\gamma, \theta \middle| q, n\right) d\gamma d\theta + \int_{0}^{\infty} \int_{\theta+q-\bar{q}^0}^{\infty} m\left(\gamma, \theta \middle| q, n\right) d\gamma d\theta \right] g\left(q \middle| n\right) dq
$$

(B.58)

The probability is comprised of four terms. The first two terms correspond to values of $q$ that render working while claiming preferable to not working while claiming. The first term in this set captures individuals for whom working while claiming also dominates not working while not claiming. In this case, the individual will always work, regardless of the value of $\gamma$. The second term captures individuals for whom not working while not claiming dominates working while claiming. In this case, only those who prefer working while not claiming to not working while not claiming will work. The second set of terms likewise covers the two settings in which not working while claiming dominates working while claiming, but the individual still decides to work. Put another way, an agent will work when $\max(v_n^1 - q, v_n^0 + \gamma - q) > \max(v^0, v^0 + \theta)$.

As in the case of claiming above, we focus on the discontinuity in the partial derivative of this probability, $\partial \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right) / \partial \tilde{z}_{n0}$. Leibniz's rule again implies:

$$
\frac{\partial \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right)}{\partial \tilde{z}_{n0}} = \int_{-\infty}^{\bar{q}} \left\{ \frac{\partial \bar{q}_n}{\partial \tilde{z}_{n0}} \left[ \int_{-\infty}^{\infty} m\left(\gamma, \bar{q}-q \middle| q, n\right) d\gamma - \int_{\Delta v}^{\infty} m\left(\gamma, \bar{q}-q \middle| q, n\right) d\gamma \right] \right.
$$
$$
\left. + \int_{\bar{q}-q}^{\infty} \frac{\partial \bar{q}_n^0}{\partial \tilde{z}_{n0}} m\left(\theta + q - \bar{q}^0, \theta \middle| q, n\right) d\theta \right\} g\left(q \middle| n\right) dq
$$
$$
+ \frac{\partial \bar{q}}{\partial \tilde{z}_{n0}} \left[ \int_{-\infty}^{0} \int_{-\infty}^{\infty} m\left(\gamma, \theta \middle| \bar{q}, n\right) d\gamma d\theta + \int_{0}^{\infty} \int_{\theta+\Delta v}^{\infty} m\left(\gamma, \theta \middle| \bar{q}, n\right) d\gamma d\theta \right] g\left(\bar{q} \middle| n\right)
$$
$$
+ \int_{\bar{q}}^{\infty} \left\{ \frac{\partial \bar{q}^0}{\partial \tilde{z}_{n0}} \left[ \int_{-\infty}^{0} m\left(q - \bar{q}^0, \theta \middle| q, n\right) d\gamma d\theta + \int_{0}^{\infty} m\left(\theta + q - \bar{q}^0, \theta \middle| q, n\right) d\gamma d\theta \right] \right\} g\left(q \middle| n\right) dq
$$
$$
- \frac{\partial \bar{q}}{\partial \tilde{z}_{n0}} \left[ \int_{-\infty}^{0} \int_{\Delta v}^{\infty} m\left(\gamma, \theta \middle| \bar{q}, n\right) d\gamma d\theta + \int_{0}^{\infty} \int_{\theta+\Delta v}^{\infty} m\left(\gamma, \theta \middle| \bar{q}, n\right) d\gamma d\theta \right] g\left(\bar{q} \middle| n\right).
$$

(B.59)

Once again relying on the fact that $\partial \bar{q}^0/\partial \tilde{z}_{n0} = 0$ and $\partial \bar{q}/\partial \tilde{z}_{n0} = \partial \Delta v/\partial \tilde{z}_{n0} = \lambda \left[ (1-\tau) + \frac{v_z}{v_c} \right]$, we can rearrange terms to yield:

$$
\frac{\partial \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right)}{\partial \tilde{z}_{n0}} = \lambda \left[ (1-\tau) + \frac{v_z}{v_c} \right] \times \left\{ \left[ \int_{-\infty}^{0} \int_{-\infty}^{\Delta v} m\left(\gamma, \theta \middle| \bar{q}, n\right) d\gamma d\theta \right] g\left(\bar{q} \middle| n\right) \right.
$$
$$
\left. + \int_{-\infty}^{\bar{q}} \int_{-\infty}^{\Delta v} m\left(\gamma, \bar{q}-q \middle| q, n\right) g\left(q \middle| n\right) d\gamma dq \right\}.
$$

(B.60)

The kink in the probability of having positive earnings can now be expressed as:

$$
\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right)}{d \tilde{z}_{n0}} - \lim_{\tilde{z}_{n0} \to z^{*-}} \frac{d \Pr\left(z_{n1} > 0 \middle| \tilde{z}_{n0}\right)}{d \tilde{z}_{n0}} = \beta,
$$

(B.61)

where

$$\beta = \lim_{\tilde{z}_{n0} \to z^{*+}} \lambda \left[(1-\tau) + \frac{v_z}{v_c}\right] \times \left\{ \left[\int_{-\infty}^{0} \int_{-\infty}^{\triangle v} m\left(\gamma, \theta \mid \bar{q}, n\right) d\gamma d\theta\right] g\left(\bar{q} \mid n\right) \right.$$

$$\left. + \int_{-\infty}^{\bar{q}} \int_{-\infty}^{\triangle v} m\left(\gamma, \bar{q} - q \mid q, n\right) g\left(q \mid n\right) d\gamma dq \right\}$$

$$= -d\tau \cdot \lambda_{n^*} \cdot g\left(\bar{q}_{n^*} \mid n^*\right) \times \underbrace{\left[\int_{-\infty}^{0} \int_{-\infty}^{0} m\left(\gamma, \theta \mid \bar{q}_{n^*}, n^*\right) d\gamma d\theta\right]}_{\Pr(\,claim \mid n^*, q_{n^*} = \bar{q}_{n^*})}$$

$$- d\tau \cdot \lambda_{n^*} \cdot \int_{-\infty}^{\bar{q}} \int_{-\infty}^{0} m\left(\gamma, \bar{q} - q \mid q, n^*\right) g\left(q \mid n^*\right) d\gamma dq, \tag{B.62}$$

and where again we have used the fact that $\lim\limits_{\tilde{z}_{n0} \to z^*} \Delta v_n = 0$.

The first term in this expression is an attenuated version of our previous kink in the probability of working. The extra term can be interpreted as the probability of claiming among the marginal labor force participants, *i.e.* those for whom $q_{n^*} = \bar{q}_{n^*}$. The integral covers the range of values for $\gamma$ and $\theta$ that render working while claiming preferable to either not working while not claiming or working while not claiming. Since the individual is indifferent between working while claiming and not working while claiming, this also implies that claiming is preferable to not claiming among this set of agents. The second term may be recognized as one component of the kink in claiming. In particular, the integral captures individuals who prefer working while claiming to not working while claiming, and prefer working while claiming to working while not claiming, but are indifferent between working while claiming and not working while not claiming. The kink in the budget set while working and claiming shifts this marginal individual to not claiming, at which point not working becomes optimal.

Thus, endogenizing claiming has two effects on the kink in the probability of positive earnings. First, it attenuates this behavioral response, because the kink only affects those who claim. Second, it amplifies the behavioral response if there are individuals who are shifted to not claiming, and prefer to not work conditional on not claiming, but to work conditional on claiming. In this case, we cannot generally conclude whether our estimate of the participation elasticity is a lower bound or an upper bound. However, we detail below two approaches for establishing an upper bound on the (negative) behavioral response of participation under certain restrictions on the parameters (*i.e.* a lower bound on the absolute value of the participation response).

### B.4.3 Bounding the Behavioral Response

One method for establishing a bound involves imposing an additional restriction on unobserved heterogeneity. In particular, we can assume that $\gamma_n \cdot \theta_n \geq 0$, *i.e.* the signs of $\gamma$ and $\theta$ are the same. In this case, delaying claiming must cause utility when working and not working to either both increase or both decrease, although the absolute value of the change in utility can be different across the two states. This assumption makes intuitive sense, as we are restricting analysis to the case in which earnings are fixed at $\tilde{z}_{n0}$. In this case, delaying claiming does not affect employment, and delaying claiming only affects utility through its direct effect on either current disposable income, or (in a dynamic setting) lifetime benefits. This would be the case, for example, if delaying claiming has the same effect on wealth, regardless of working status. The assumption rules out cases in which, for example, working affects life expectancy in a way that changes the sign of the effect of delaying claiming on lifetime benefits, or a case in which working relaxes a borrowing constraint and the effect of delaying claiming on utility therefore changes sign, conditional on working.

If our assumption holds, we can show the following:

$$\int_{-\infty}^{\bar{q}} \int_{-\infty}^{0} m\left(\gamma, \bar{q} - q \mid q, n^*\right) g\left(q \mid n^*\right) d\gamma dq = 0 \tag{B.63}$$

66

The reason why this probability becomes zero is as follows. First, we are restricting attention to values of $q \leq \bar{q}$. In addition, we are fixing the value of $\theta = \bar{q} - q \geq 0$. However, we are also restricting analysis to individuals for whom $\gamma \leq 0$. Given our assumption that the signs of $\gamma$ and $\theta$ must be the same, this set must be of measure zero.

In this case, we can simplify our two kinks above as follows:

$$\delta = -d\tau \cdot \lambda_{n^*} \times \int_{-\infty}^{\bar{q}} \int_{-\infty}^{\bar{q}-q} m\left(0, \theta \mid q, n^*\right) g\left(q \mid n^*\right) d\theta dq \tag{B.64}$$

$$\beta = -d\tau \cdot \lambda_{n^*} \times g\left(\bar{q} \mid n^*\right) \underbrace{\left[\int_{-\infty}^{0} \int_{-\infty}^{0} m\left(\gamma, \theta \mid \bar{q}_{n^*}, n^*\right) d\gamma d\theta\right]}_{\Pr\left(claim \mid n^*, q_{n^*}=\bar{q}_{n^*}\right)} \tag{B.65}$$

The kink in participation we estimate is now a weak upper bound on a negative kink (*i.e.* the absolute value of the kink in participation we estimate is a lower bound on the kink that would obtain absent a claiming response). This implies that our observed elasticity will be a lower bound. Note that the probability of claiming that attenuates the behavioral response is local to marginal labor force participants, *i.e.* $q = \bar{q}$, and thus is not the same as the population probability of claiming at $\tilde{z}_{n0} = z^*$.

If we are not willing to impose the above restrictions on the joint distribution of $\gamma$ and $\theta$, we can still achieve a lower bound. Using the results from the previous two sections, we have:

$$\beta - \delta = -d\tau \cdot \lambda_{n^*} \times g\left(\bar{q}_{n^*} \mid n^*, t\right) \left[\int_{-\infty}^{0} \int_{-\infty}^{0} m\left(\gamma, \theta \mid \bar{q}_{n^*}, n^*\right) d\gamma d\theta\right]$$

$$+ d\tau \cdot \lambda_{n^*} \times \int_{-\infty}^{\bar{q}} \int_{-\infty}^{\bar{q}-q} m\left(0, \theta \mid q, n^*\right) g\left(q \mid n^*\right) d\theta dq \tag{B.66}$$

Again, this difference in kinks provides an upper bound on the negative kink in participation. Note that this bound is not guaranteed to be negative. However, in cases in which the kink in claiming is relatively negligible—as appears to be the case in our empirical application—it is possible to establish a non-trivial upper bound on the kink and, by extension, a lower bound on the employment elasticity. Furthermore, if there is no detectable kink in claiming, then simply rescaling the kink in employment by the share claiming is sufficient to adjust for endogenous claiming.

## C Procedure for estimating excess mass

As we explain in Gelber, Jones, and Sacks (2013), we seek to estimate the "excess mass" at the kink, *i.e.* the fraction of the sample that locates at the kink under the kinked tax schedule but not under the linear tax schedule. Following a standard procedure in the literature (*e.g.* Saez, 2010; Chetty, Friedman, Olsen, and Pistaferri, 2011), we estimate the counterfactual density (*i.e.* the density in the presence of a linear budget set) by fitting a smooth polynomial to the earnings density away from the kink, and then estimating the "excess" mass in the region of the kink that occurs above this smooth polynomial.

Specifically, for each earnings bin $z_i$, we calculate $p_i$, the proportion of the sample with earnings in the range $[z_i - k/2, z_i + k/2)$. The earnings bins are normalized by distance-to-kink, so that for $z_i = 0$, $p_i$ is the fraction of all individuals with earnings in the range $[0, k)$. To estimate bunching, we assume that $p_i$ can be written as:

$$p_i = \sum_{d=0}^{D} \beta_d \cdot (z_i)^d + \sum_{j=-k}^{k} \gamma \cdot 1\{z_i = j\} + \varepsilon_i \tag{C.67}$$

and run this regression (where 1 denotes the indicator function and $j$ denotes the bin). This equation expresses the earnings distribution as a degree $D$ polynomial, plus a set of indicators for each bin within $k\delta$ of the kink, where $\delta$ is the bin width. In our empirical application, we choose $D = 7$, $\delta = 500$ and $k = 6$ as our baseline (so that six bins are excluded from the polynomial estimation). We control for a baseline

seventh-degree polynomial through the density following Chetty, Friedman, Olsen, and Pistaferri (2011). The parameter $\gamma$ reflects the excess density near the kink.

Our measure of excess mass is $\hat{M} = 2k\gamma$, the estimated excess probability of locating at the kink (relative to the polynomial term). This measure depends on the counterfactual density near the kink, so to obtain a measure of excess mass that is comparable at the kink, we scale by the predicted density that we would obtain if there were a linear budget set. This is just the constant term in the polynomial, since $z_i$ is the distance to zero. Thus, our estimate of normalized excess mass is $\hat{B} = \frac{\hat{M}}{\hat{\beta}_0}$. We calculate standard errors using the delta method. We calculate the density in each bin by dividing the number of beneficiaries in the bin by the total number of beneficiaries within the bandwidth; note that this normalization should not affect the excess normalized mass or the estimated density, because dividing by the total number of beneficiaries within the bandwidth affects the numerator (*i.e.* $\hat{M}$) and denominator (*i.e.* $\hat{\beta}_0$) of the expression for $\hat{B}$ $\left( = \frac{\hat{M}}{\hat{\beta}_0} \right)$ in equal proportions and therefore should have no impact on $\widehat{B}$.

# D  Appendix Figures and Tables

Appendix Figure D.1: Earnings Test Real Exempt Amount, 1978 to 1987
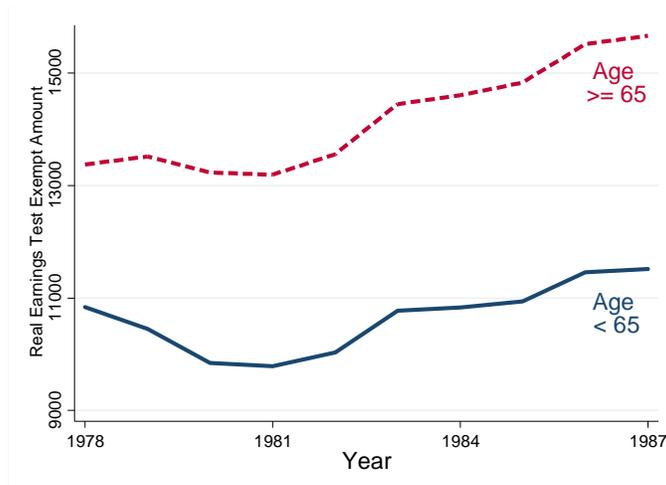


Notes: The figure shows the real value of the exempt amount over time among those 62-64 years old (labeled "Age<65" in the graph) and those above (labeled "Age>=65"). The AET applied to earnings of claimants from ages 62 to 71 from 1978 to 1982, but only to claimants aged 62 to 69 from 1983 to 1989. All dollar figures are expressed in real 2010 dollars.

Appendix Figure D.2: Histogram of Percent Earnings Growth, Ages 28 to 59
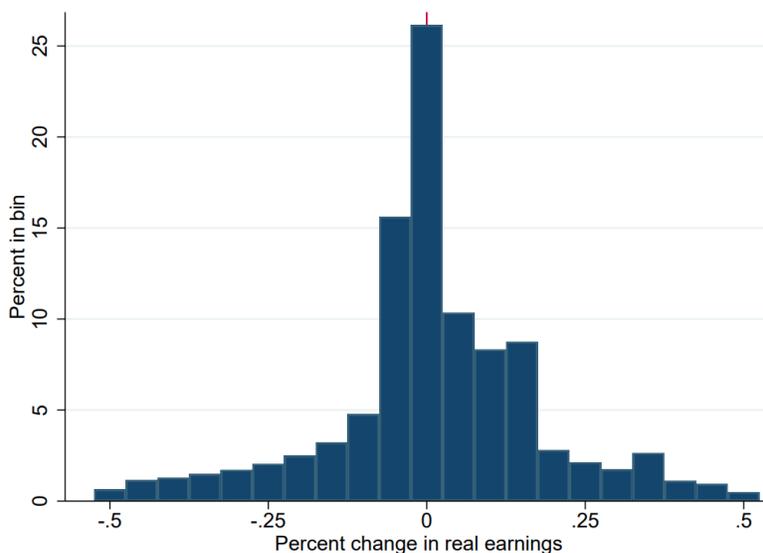


Notes: This histogram shows that there is a large mass near zero percent nominal earnings growth from one year to the next in a wider set of ages, 28 to 59, than we focus on in the main analysis. This indicates that among a broad set of ages, a substantial mass of individuals have no growth in desired nominal earnings, consistent with the assumptions necessary for our RKD to estimate a lower bound on the elasticity as described in the main text. This suggests that it should be possible to use our method, implemented in this case through using lagged earnings to proxy for desired earnings, when studying extensive margin responses to other policies applying in other age ranges. The figure uses the SSA data we have, covering the 1918 to 1923 cohorts in calendar years from 1951 to 1984. See other notes to Figure 4.

Appendix Figure D.3: Probability of Positive Earnings at Ages 63 to 64, wider x-axis
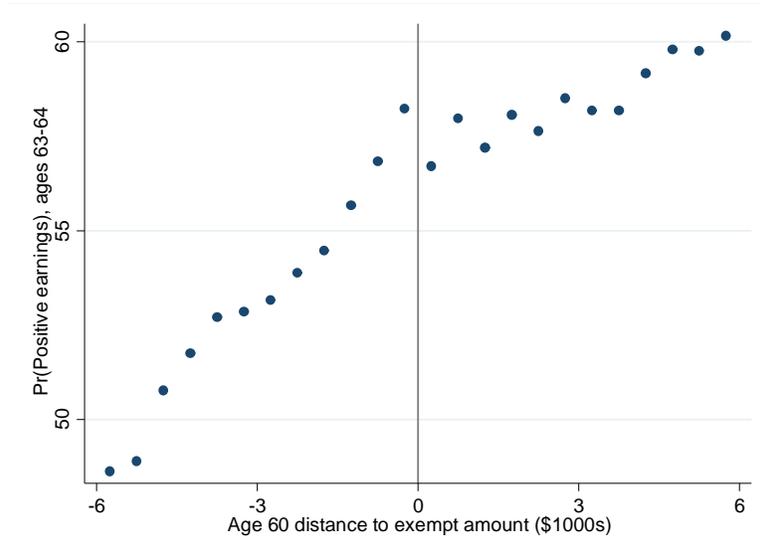


Notes: See the notes to Figure 7. This figure is identical to Figure 7, except that the range of the $x$-axis on this figure runs from -$6,000 to $6,000. Like Figure 7, this figure also shows a clear, discontinuous change in slope at the exempt amount.

Appendix Figure D.4: Mean Probability of Claiming at Ages 63 to 64
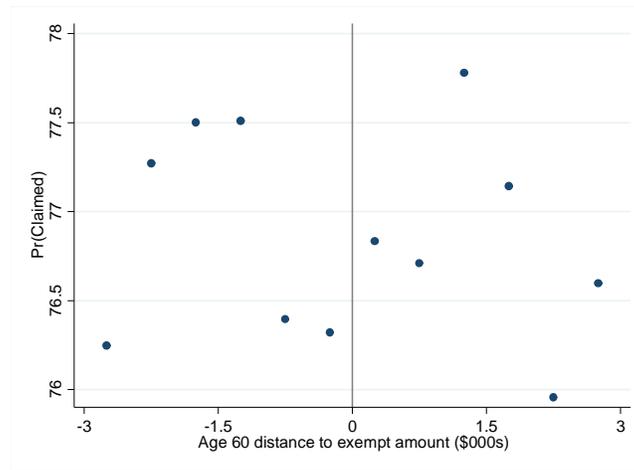


Notes: The figure plots the mean claiming rate, *i.e.* the probability someone has claimed by the calendar year of reaching age $t$, at ages 63 to 64 averaged, as a function of the distance of age 60 earnings from the exempt amount. The figure shows that there is no clear visual change in the slope of the claiming rate, and regression evidence supports the same conclusion: a placebo test in the spirit of Ganong and Jäger (2015) shows p=0.15 for the two-sided test of equality of the coefficient with zero. See other notes to Figure 4.

Appendix Table D.1: Robustness of Elasticity Estimates

| | (1) | (2) |
|---|---|---|
| | Baseline Specification | With Controls |
| Linear | 0.49 | 0.45 |
| | (0.19)*** | (0.18)*** |
| N | 95,960 | 104,665 |
| Quadratic | 0.66 | 0.60 |
| | (0.27)*** | (0.24)*** |
| N | 160,785 | 172,979 |
| Cubic | 0.82 | 0.68 |
| | (0.28)*** | (0.28)*** |
| N | 273,421 | 326,782 |

Notes: The table presents robustness checks on the main elasticity estimates in Table 5. The "with controls" column shows the kink in the employment probability when we control for dummies for year of birth, sex, and race. Robust standard errors, using the procedure of Calonico, Cattaneo, and Titiunik (2014), are reported in parentheses. The number of individuals included in each "reduced form" regression (18) is shown below the standard error. See other notes to Table 2.

Appendix Table D.2: RKD Elasticity Estimates using Regression-Based First Stage

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Full Sample | Men | Women | White | Non-White | High Prior Earnings | Low Prior Earnings |
| First Stage Kink | -4.00 | -3.25 | -3.99 | -3.88 | -3.09 | -3.76 | -3.82 |
| | (0.01)*** | (0.01)*** | (0.01)*** | (0.01)*** | (0.01)*** | (0.01)*** | (0.01)*** |
| Elasticity | 0.59 | 0.37 | 0.60 | 0.59 | 0.64 | 0.62 | 0.52 |
| | (0.23)*** | (0.29) | (0.26)*** | (0.22)*** | (0.68) | (0.35)* | (0.11)*** |
| N | 95,960 | 68,971 | 66,251 | 93,722 | 39,271 | 19,574 | 101,709 |

Notes: See notes to Table 5. Relative to Table 5, the elasticities differ here because we use a linear RKD to estimate the first stage kink in the ANTR (reported in the first row), rather than using the analytic expression as in Table 5.

Appendix Table D.3: Elasticity Estimates Accounting for Claiming

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Full Sample | Men | Women | White | Non-White | High Prior Earnings | Low Prior Earnings |
| Elasticity | 0.63 | 0.32 | 0.65 | 0.61 | 0.53 | 0.62 | 0.53 |
| | (0.25)*** | (0.25) | (0.28)** | (0.23)*** | (0.56) | (0.35)* | (0.12)*** |
| N | 95,960 | 68,971 | 66,251 | 93,722 | 39,271 | 19,574 | 101,709 |

Notes: See notes to Table 5. As explained in the main text, we calculate these elasticities by inflating the Table 5 elasticities by 29.9 percent, to account for claiming behavior. Among those with age 60 earnings below the kink, but not more than $2,797 below the kink, 74.5 percent of the sample claims by age 63, and 79.5 by age 64. We calculate 29.9 percent as $100 \cdot (1/[(0.795+0.745)/2]-1)$.