The Employment Effects of the Social Security Earnings Test*

Alexander Gelber
UC San Diego and NBER

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

Daniel Sacks
Indiana University Kelley School of Business

Jae Song
Social Security Administration

January 2020

Abstract

We investigate the impact of the Social Security Annual Earnings Test (AET) on the employment decisions of older Americans. The AET reduces Social Security benefits by one dollar for every two dollars earned above the exempt amount. Using a differences-in-differences design, we find that the employment rate of those predicted to become subject to the AET decreases substantially relative to those not predicted to become subject to it. The point estimates suggest that the AET reduces the employment rate of Americans aged 63-64 by at least 1.2 percentage points.

* This research was supported by the U.S. Social Security Administration (SSA) through grant #RRC08098400-09 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium, as well as by the Alfred P. Sloan Foundation. This research was completed partly while Gelber was on leave at the Stanford Institute for Economic Policy Research, partly funded by a grant from the Alfred P. Sloan foundation. We thank Gary Engelhardt as well as seminar participants at SSA and the Retirement Research Consortium for comments, and we thank Jeff Shrader for sharing his code for two-sample two-stage least squares. The findings and conclusions expressed are solely those of the author(s) and do not represent the views of SSA, any agency of the Federal Government, or the NBER.

I. Introduction

As the economy-wide employment-to-population ratio has fallen by around four percentage points since its peak in 2000, many policy makers are interested in ways to affect employment rates (Abraham and Kearney, 2017). For older Americans, the Social Security Annual Earnings Test (AET) has a large effect on Social Security Old Age and Survivor Insurance (OASI) benefits and could therefore affect employment substantially. The AET reduces OASI claimants' current OASI benefits as a proportion of earnings, once a claimant earns in excess of an exempt amount. For example, for OASI claimants aged 62 to 65 in 2018, current OASI benefits are reduced by 50 cents for every extra dollar earned above $17,040. The bite of the AET can be substantial. For example, we estimate that in 2003—the last year for which we have the relevant data—the AET reduced the current OASI benefits of 538,000 individuals. The AET reduced their benefits by over half on average (51.4 percent), or $4.3 billion in aggregate.[1]

This large benefit reduction rate may reduce OASI claimants' incentives for additional work. 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 Normal Retirement Age (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, which does not apply to older ages.

[1] If earnings and/or employment are reduced by the AET, then the 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. Thus, this strengthens our case that the AET substantially affects current OASI benefits. As discussed below, when current OASI benefits are reduced due to the AET, future benefits are typically enhanced; however, our point here is simply that the AET greatly affects benefits, at least in their timing, and therefore has important potential implications.

The prior literature on the AET has tended to focus on the policy's intensive margin effect, *i.e.* the choice of how much to earn, given that an individual earns a positive amount (*e.g.* Burtless and Moffitt, 1985; Friedberg, 1998; Friedberg, 2000; Song and Manchester, 2007; Gelber *et al.*, 2013; Engelhardt and Kumar 2009; Engelhardt and Kumar 2014). For example, the AET could reduce the incentive for additional earnings above the exempt amount and cause an individual to choose part-time work rather than full-time work. This literature generally finds moderate substitution elasticities at the intensive margin with respect to the AET. However, nearly all prior studies find little evidence of extensive margin responses, *i.e.* little evidence that older workers respond to the AET on the margin of whether to work or not. The earlier empirical literature on the AET largely concludes that the policy has little meaningful effect on the labor supply of older men (Viscusi 1979; Burtless and Moffit 1985; Gustman and Steinmeier 1985; Vroman 1985; Honig and Reimers 1996; Leonesio 1990).

More recently, researchers have examined the effect of the AET on employment decisions using a difference-in-differences framework to study the effects of the year 2000 elimination of the AET for those over the NRA, comparing the evolution of employment rates of those over *vs.* under the NRA. Using this design, several studies find little evidence for an effect on the employment rate (Gruber and Orszag, 2003; Song, 2003; Song and Manchester, 2007; Haider and Loughran, 2008), while Friedberg and Webb (2009) find significant but modest effects in some specifications in Current Population Survey data.[2]

In this paper, we revisit the effects of the AET on the employment rate of older workers

[2] 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 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.

using administrative data and a different methodological approach. We propose a new strategy for investigating the effects of the AET on the employment rate by examining how employment outcomes vary with the change in the perceived incentive to work that occurs when one earns above the AET exempt amount. At this exempt amount the perceived incentives to work change notably due to the AET and could therefore reduce employment among those affected in this population. Using a differences-in-differences design, we compare employment rates subsequent to reaching the Social Security retirement age of those with predicted earnings above and below the AET exempt amount, who form the "treatment" and "control" groups, respectively. We implement this design using earnings histories derived from Social Security Administration (SSA) data on a 1 percent random sample of the U.S. population in birth cohorts 1931 to 1943 over the years 1986 to 2006. Our sample includes 240,000 people with 2.4 million annual observations.

Our results show much larger effects on employment than the bulk of previous literature had indicated. We find that the employment rate of those subject to the AET decreases significantly relative to those not subject to it. This effect occurs suddenly at the ages individuals become subject to the AET, bolstering the credibility of the estimates. Our point estimates suggest that the AET reduces the employment rate of affected older Americans aged 63-64 by at least 1.2 percentage points. This conclusion is robust to a battery of placebo and robustness checks.

These findings indicate that the AET is an important factor affecting the work decisions of older Americans and therefore should be a key focus of policy-makers. The combination of a large administrative dataset with individual-level microdata (also used in various forms in Song, 2003, Song and Manchester, 2007, and Haider and Loughran, 2008) and our novel identification strategy leads to estimates of sizeable elasticities. However, the effect we estimate applies to a younger group than those studied in the difference-in-differences literature cited above: we focus on those locating near the exempt amount in the policy-relevant 63-to-64-year-old group to which the AET currently applies. Our findings are therefore not directly comparable to this previous literature, but provide relatively novel estimates for a younger group of current policy relevance.

Recent work by Gelber *et al.* (2018), which also studies workers below and above the exempt amount, represents an additional exception to the conclusion that the AET has little effect on older workers' employment rate.[3] Whereas the Gelber *et al.* (2018) study relies on the Regression Kink Design smoothness assumptions, the current paper relies on the differences-in-differences parallel trends assumptions. Gelber *et al.* (2018) study a group that is local to the exempt amount, whereas the current paper's estimates apply both to those close to and farther from the exempt amount—a broader group of significant policy relevance to evaluating the more general employment effect of the AET.

The paper proceeds as follows. Section II describes the policy environment. Section III describes a framework for interpreting our results. Section IV explains the empirical strategy. Section V describes our data. Section VI proceeds to the results. Section VII discusses the results and relates them to prior estimates and Section VIII concludes.

II. Policy environment

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

[3] Because the policy environment, corresponding theoretical framework, data, and outcomes are similar to Gelber *et al.* (2018), the corresponding sections of this paper have overlap with that previous paper.

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. To illustrate how the AET works, consider a 63-year-old earning \$23,640 in 2019, receiving \$1,000 in monthly benefits, and facing a \$17,640 exempt amount and a 50 percent benefit reduction rate (BRR). Her current annual benefits would be reduced by \$3,000=(\$23,640-\$17,640)×50%, equal to 3 months of benefits. If that same worker instead earned \$50,000, annual benefits would be reduced by \$12,000, rather than being reduced by \$16,180=(\$50,000-\$17,640)×50%, because benefits are not reduced below zero.

Both the exempt amount and the BRR have changed over time. In Figure 1 we show the exempt amount for the years we study. For those whose age is at the NRA and over, the exempt amount is substantially higher than for those below the NRA. Over the main years we study, 1994 to 2006, the exempt amount for those under NRA ranged from \$11,830 to \$13,499, while the exempt amount for those NRA and over was on average around \$13,500 higher than the amount for those under NRA, with those older than the NRA no longer facing the AET after 2000. For those under the NRA but above the EEA—the main group that our empirical work studies—the BRR was 50 percent throughout the period we study, 1994 to 2006. The AET applied to earnings from ages 62 to 69 between 1994 and 2000, and from ages 62 to the NRA (roughly age 65) between 2000 and 2006.[4]

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. If both spouses earn above the exempt amount, the reduction in the total of both spouses' benefits is equal to the

[4] Starting in 2002, the NRA increased by 2 months per cohort, reaching 65 and 8 months in 2006.

sum of the reductions attributable to each spouse's earnings. For example, suppose the woman above earning $23,640 in 2018 and receiving $12,000 in annual benefits is married to a man earning $22,640 and receiving $15,000 in annual benefits. The AET would then reduce couple's total of $27,000 (=$15,000+$12,000) in current annual benefits by $5,500=($23,640-$17,640)×50%+($22,640-$17,640)×50%. We can interpret our estimates as the effect of the AET applying to a given spouse's earnings on that spouse's employment.

When current OASI benefits are lost to the AET, future scheduled benefits increase in most cases. This is sometimes referred to as "benefit enhancement." The potential increase in future benefits depends on one's age. 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 (Social Security Administration Program Operations Manual System 2019). Future benefits are raised by 0.55 percent per month of benefits withheld for the first three years of AET assessment. In the budget set, this adjustment creates a notch in future benefits as well as a kink in current benefits at the AET threshold. 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). However, the data show no response to these incentives. For those under NRA, Gelber, Jones and Sacks (2019) show that there is disproportionate bunching *under* the exempt amount – the opposite of the pattern we would expect if people reacted to the notch.

Perhaps the reason for this lack of reaction is that the earnings test is widely viewed as a pure tax. Most popular guides do not note the subsequent adjustment in benefits under the earnings test (Gruber & Orzag (2003)). During the period that we study, the popular guide Your Income Tax (J.K. Lasser Institute (1997)), for example, warned readers that if "you are under age 70, Social Security benefits are reduced by earned income," but did not note the subsequent benefit adjustment. Consistent with this explanation, when we refer in this paper to "incentives" created by the AET, we mean the *perceived* incentives.[5]

III. Framework for interpreting the empirical results

This section sketches a basic framework for interpreting the empirical strategy and results. We describe how the AET impacts an individual's decision of whether or not to have positive earnings, *i.e.* the "employment" decision. The Appendix contains a more formal setup.

Following previous literature (*e.g.* Friedberg, 1998; Friedberg, 2000), we model the AET as creating a positive benefit reduction rate (BRR) for some individuals above the exempt amount. For simplicity we suppose that separate from the AET a marginal tax rate τ_0 is imposed on earnings. In the presence of the AET, the AET reduces benefits at the margin when earnings increase above the exempt amount z_1^*, until the point at which benefits have been phased out entirely at earnings level z_2^*. At the extensive margin, to capture the realistic pattern of potential entry to or exit from non-trivial levels of earnings, we can assume a fixed cost of employment (Cogan, 1981). Extensive margin decisions are then a function of the average net-of-benefit-reduction rate (ANBRR), defined as the fraction of an individual's gross income that she keeps, net of both taxes and benefits, if she is employed at earnings level z rather than earning zero.[6]

---

[5] Aside from the influence of misperceptions, individuals could correctly perceive the AET as penalizing current earnings, for several potential reasons. For liquidity-constrained individuals, those whose expected lifespan is shorter than average, or those who discount particularly quickly, the AET is more punitive—and such individuals could also choose to reduce work or stop working in response to the AET.

[6] Gelber *et al.* (2018) give an extended discussion of employment decisions in the presence of a kinked budget set.

Figure 2 illustrates the extensive margin incentives created by the AET. We plot the ANBRR as a function of counterfactual earnings, meaning earnings conditional on working in the counterfactual state where there is only a linear tax and no AET. We distinguish between these earnings and *realized* earnings, z, which incorporate the extensive margin decision and can be zero. In the absence of the AET, the ANBRR is constant at $1-\tau_0$. In the presence of the AET, the ANBRR begins to decrease at the exempt amount as Social Security benefits are phased out by the AET. However, after the Social Security benefit has been entirely phased out due to the AET, the ANBRR begins to increase in counterfactual earnings as the importance of the AET benefit reductions diminish relative to counterfactual earnings, and for large enough counterfactual earnings the ANBRR asymptotes back to $1-\tau_0$.

Figure 3 illustrates how the AET impacts extensive margin decisions. The dashed line of Panel A represents a presumed smooth relationship between the employment rate in the absence of the AET (*y*-axis) and counterfactual earnings (*x*-axis). The solid line shows that the AET begins to reduce the employment rate around the exempt amount. This motivates our difference-in-difference empirical strategy: people with desired earnings below the exempt amount are essentially unaffected, but people with desired earnings above the exempt amount are. Around z_2^* where Social Security benefits are phased out entirely, the ANBRR begins to increase again and the difference between the two employment rates begins to decrease.[7] Panel B shows the effect of the AET on employment, defined as the difference between the employment rate in the dashed

[7] More precisely, if the marginal effect of the ANBRR on the employment rate does not depend on counterfactual earnings, the magnitude of the effect of the AET on the employment rate should decrease beginning at the non-convex kink. However, if the marginal effect of the ANBRR on the employment rate varies depending on counterfactual earnings, then the effect of the AET on the employment rate may be maximized at a different counterfactual earnings level. Regardless, the effect of the AET on employment follows a U-shaped pattern.

and solid lines from Panel A. The AET has a U-shaped effect on employment, where the maximum reduction in employment due to the AET occurs around z_2^*.[8]

B. Specific features of the policy setting

Our policy setting has a number of specific features that can be incorporated into the framework. 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. However, the results should still be interpreted as reflecting an "observed" elasticity (in the terminology of Chetty 2012). Our observed elasticity remains of interest regardless of the effect on claiming, in the sense that policy-makers are interested in the raw employment effects of changing the AET.

Second, we considered 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 Gelber *et al.* (2018), we present a fully dynamic, multi-period model with a joint decision over saving and earnings; we show that under a set of empirically relevant assumptions our results still hold, and we discuss the interpretation of our estimated results in a dynamic model. If agents act as if they do not anticipate the change in an initial period—consistent with the empirical findings in Gelber *et*

[8] The discussion here focuses on the pattern of employment over a broad range of earnings. By contrast, Gelber *et al.* (2018) focus on the patterns of employment in a small neighborhood around the exempt amount and develop conditions under which the employment rate exhibits a discontinuous change in slope at the exempt amount.

al. (2013, 2018)—then we can interpret the results as reflecting the impact of an unanticipated change in policy.

Third, Gelber *et al.* (2018) also extend this basic dynamic model to tailor it to our particular policy setting. They show that when reductions in current benefits due to the AET can lead to increases in later benefits, this can still lead to reductions in current labor supply if the benefit enhancement is not actuarially fair. Furthermore, if people do not understand or know about benefit enhancement, then they will respond to the AET as a pure tax.

IV. Empirical strategy

A. Regression specification

Our approach involves a difference-in-difference estimator of the effect of the AET on employment. Specifically, we exploit the fact that the AET only affects those who would earn above the exempt amount in a given year if they choose to work. Furthermore, the AET only affects those who claim Social Security, which can be claimed as early as age 62. We therefore compare those who would potentially earn above and below the exempt amount, before and after turning 62. We define "potential" earnings as the earnings individuals would earn in the absence of the incentives created by the AET (Gelber, Jones, and Sacks 2013; Gelber *et al.* 2018).

Given positive earnings z_{ai}, for individual i at "base" age a, we examine the probability that one continues to have positive earnings at age $a + t$, where t =3 or 4. Thus, an individual's "base age" or "base year" refers to a period three to four years before their "age." We explain below why we consider a minimum lead of 3. We denote the AET exempt amount by z^* and compare the difference in the probability of positive earnings for those with $z_{ia} \geq z^*$ to those with $z_{ia} < z^*$ when $a + t < 63$ and when $a + t \geq 63$.[9]

Thus, our approach represents a differences-in-differences analysis. Our outcome is the probability of having nonzero earnings t years in the future. Our treatment group is individuals with earnings above the exempt amount at age a, and our control group is individuals with earnings below the exempt amount at age a. Our pre-period (i.e. prior to the potential effect of the AET) is when $a + t < 63$, and our post-period (i.e. after the potential effect of the AET) is when $a + t \geq 63$.[10] We let $post$ indicate $a + t \geq 63$ and $treat$ indicate $z_{ia} \geq z^*$.

In regression form, our difference-in-difference model is estimated using the following specification:

$$E_{i,a+t} = \beta_1 treat_{ia} + \beta_2 treat_{ia} post_{ia} + \beta_3 b_i + \beta_4 b_i post_{ia}$$
$$+\beta_5 female + \alpha_a + \theta_{yob} + \varepsilon_{i,a+t} \tag{1}$$

where $E_{i,a+t}$ is a binary variable that equals one if an individual is employed t years after age a. We can estimate our main specification in equation (1) using ordinary least squares (OLS). The key coefficient of interest is β_2, the difference-in-difference coefficient for the interaction of "base earnings above exempt amount" and "base age + t above 63." It gives the average difference in the probability of having positive earnings at age 63 or older, for those who would

[9] We treat age 63 as the "post period" because that is the first year one would expect to observe an extensive margin effect of the AET. 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. If an individual claims OASI at age 62, the AET only applies to earnings in the 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.
[10] Our strategy implicitly assumes that desired earnings change modestly from age 60 to age 63. Indeed, our results show that for a large percentage of workers in a placebo sample three years apart (ages 57 and 60), there is little change in real earnings.

earn above the exempt amount, z^*, relative to those who would earn below the exempt amount, and relative to the pre-age 63 difference.

In our preferred specifications, we include demographic controls in the form of a female dummy variable, as well as fixed effects for age and year-of-birth (which implicitly control for year as well). We further control for b_i, a time-invariant measure of benefits, which we interact with the "post" indicator, $1\{a + t \geq 63\}$. We control for benefits because employment decisions may be sensitive to benefits (e.g. Coile and Gruber 2007), and because benefits, which depend on lifetime earnings, are higher for people with base age earnings above the exempt amount. We interact the benefit measure with the post dummy because we expect benefits to matter only once people become eligible for to claim OASI benefits. To measure b_i, we do not use actual benefits, as these are endogenous to earnings history. Rather, we impute benefits based on earnings history up to the age when people enter our analysis sample. We provide more detail in the data section. In implementation, we consider $t = 3$ or $t = 4$ as alternative robustness checks. We cluster our standard errors at the individual level to account for intra-individual correlations, such as those arising from serial correlation.

B. Validity and implementation of method

Our method relies on two key assumptions. First, we make the "parallel trends" assumption, common to all difference-in-difference estimators. That is, we assume that in the absence of our key policy variation, i.e. the introduction of the AET, the difference in the probability of positive earnings between those with earnings below and above the exempt amount would remain constant across ages. Formally, we impose the restriction in equation (1) that the parameter β_1 is time-invariant. We cannot technically test this assumption directly, but we can assess the merits of this assumption using data from ages younger than 63. Our assumption implies that labor force participation during these ages follows a similar trend for those earnings below versus above the exempt amount.

Second, we assume that earnings at age a are a reasonable proxy for potential earnings at age $a + t$, conditional on earning a positive amount in the counterfactual scenario where the AET is not in effect. We would ideally model the probability of positive earnings at age $a + t$ as a function of whether or not an individual is affected by the AET in that same year. However, we cannot observe potential earnings at age $a + t$ for those who do not work, and for those who do work, earnings may decrease in response to the AET.

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.

To circumvent this problem of endogenous earnings decisions at the ages directly subject to the AET, we will use lagged earnings as a proxy for potential earnings in the present period. Instead, we will use a lagged measure of earnings from age a, as a proxy for the earnings that would be realized at age $a + t$. This assumption will be violated if individuals tend to adjust their earnings in anticipation of the AET. However, our prior research has indicated that during the ages used in our analysis, there is little evidence of anticipatory adjustment to the AET (Gelber *et al.* 2013, 2018).

To support the validity of using lagged earnings as a proxy for future desired earnings, we show that desired earnings remain relatively 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. Panel A of Figure 4 shows that from age 58 to age 59—a placebo set of ages during which our sample is not subject to the AET—a large mass of real earnings growth does occur at or near zero percent growth. Examining three-year growth rates in Panel B of the figure shows a considerable mass around zero, although these growth rates are more diffuse than those in Panel A. Another way to understand the validity of our measure is to consider how well $treat_a$ predicts $treat_{a+3}$. We illustrate this relationship in Figure 5, which plots the average of $treat_{a+3}$ against age a earnings (relative to the exempt amount. Overall, we see that those treated at a are 57 percentage points more likely to be treated at $a + 3$. At very low earnings there is a low but positive probability of switching from untreated to treated, about 20 percentage points. This increases rapidly to about 80 percentage points once earnings reach about $20,000 above the exempt amount, and the probability levels off thereafter. Thus, $treat_a$ is highly predictive of $treat_{a+3}$, but not perfectly so. This creates measurement error in our main regressor which, if anything, attenuates our estimates.

Correcting for this measurement error would require an estimate of the signal-to-noise ratio in our proxy. With classical measurement error, plus the assumption that the earnings dynamics are stable from ages 55-61 to ages 59-64, the signal-to-noise ratio is 0.57 (i.e. the difference in probability of $treat_{a+3}$ given $treat_a = 1$ or $treat_a = 0$), and correcting for measurement error simply requires us to scale up our estimate by 1/0.57. However, we report unadjusted estimates throughout the text, for two reasons. First, we err on the side of the conservative, unadjusted estimate. Second, adjusting for measurement error requires extrapolating the earnings dynamics to older ages, but the dynamics could be quite different at those ages.

While investigating a lead of $t=4$ as a robustness check, in the baseline we choose to use a three-year gap between base age earnings and the employment outcome we investigate. We have to use at least a two-year lag to look at the age 63 effect of the AET (because age 62 earnings could respond to the AET). We use a three-year lag to guard against the possibility of anticipatory adjustment in age 61 earnings to the expected future imposition of the AET (and demonstrate that there is no evidence of anticipatory adjustment at age 60). A further advantage of using a 3-year lag (rather than 2) is that we can examine age 64 employment (which would be impossible with a 2-year lag because age 62 earnings respond to the AET). As our baseline we use the shortest lag possible, *i.e.* $t=3$, that satisfies these conditions. We avoid using a lag that is greater than $t=4$ because this would involve examining employment outcomes at ages above 65, when the exempt amount is much higher.

Our differences-in-differences method will control for the possibility that high earners have higher labor force attachment by removing the difference attributable to the constant effect of being in the low- or high-earning group at age a. The method relies on the assumption that in the absence of the AET, the difference in the probability of employment would be constant between those earning above and below the exempt amount, before and after age 63. While we cannot directly test this assumption, we can perform basic checks on its validity.

First, a testable prediction is that we should see the difference in employment probabilities between these two groups evolving similarly prior to age 63. Second, in our differences-in-differences specification, we can add separate trends in age in the below-z^* and above-z^* groups, to demonstrate that a significant difference occurs precisely around the time when the AET is imposed, relative to this trend.

Third, if those with prior earnings below the AET exempt amount have weaker labor force attachment and hypothetically drop out of the labor force more when they reach retirement age, then we would expect a larger decrease in the employment probability at retirement age among those with earnings below the exempt amount than among those above the exempt amount. This is the opposite of what we find in our results, namely a smaller decrease in the employment probability at retirement age among those with earnings below the exempt amount than among those above the exempt amount. Thus, if anything this hypothesis would push against our finding of a large response.

V. Data

To administer benefits, the Social Security Administration (SSA) collects earnings histories on all Americans in the Master Earnings File (MEF). Every year since 1978, the MEF records total pretax W-2 earnings as well as self-employment earnings for each Social Security Number. As W-2s are mandatory information returns filed with the Internal Revenue Service for each employee for whom the firm withholds taxes and/or to whom remuneration exceeds a modest threshold, the MEF contains earnings information regardless of whether an individual files taxes. SSA has made available two public-use datasets derived from the MEF: the Earnings Public Use File (EPUF) and the Benefits and Earnings Public Use File (BEPUF). Both data sets contain earnings histories and limited demographic information: year of birth and sex. They differ in the years covered and sampling frame. The EPUF covers earnings histories for 1951-2006, and it is a 1 percent sample of all Social Security Numbers ever issued as of 2006. The BEPUF covers earnings for 1951-2003, and it is a 1 percent sample of all current claimants as of

2004. Because the BEPUF conditions on claiming benefits and surviving to 2004, it is not necessarily representative of older cohorts. We therefore implement our estimation strategy using the EPUF. However, in Appendix E we show that we obtain very similar results in the BEPUF.[11]

We limit the EPUF to relatively recent cohorts: those born 1931-1943. We end in 1943 because the last year of the EPUF is 2006, so this is the last cohort that we can observe respond to the AET by age 63. Our choice to start in 1931 balances two considerations: more cohorts give us more power, but at the same time we would like to retain a focus on more recent cohorts that are of greater policy interest. We chose the 1931-1943 range because it represents the second half of our data, as the earliest we could start is 1918.[12, 13] We emphasize, however, that none of our results depends on the specific set of cohorts studied, and we show in robustness tests that using the full range of cohorts yields similar (in fact larger) responses.

We classify each age in a given calendar year as the highest age an individual attains in that calendar year. Individuals in our cohorts reach ages 63 to 64, the main ages at which we investigate the effect of the AET, in 1994 to 2006. Because our identification strategy requires us to observe base age earnings, we limit our estimation sample to people with positive earnings in the base age. We limit the sample to base ages 55-61 when we look $t = 3$ years ahead and base ages 55-60 when we look $t = 4$ years ahead (both corresponding to outcome ages 58-64). (We begin the analysis sample at age 55 so that we can use earnings through age 54 to impute

[11] In an earlier version of this paper, we used the 1917-1923 birth cohorts in the MEF itself rather than public use files. Although we have since lost access to those data, we show in Appendix E that the results are also similar in the MEF and in the EPUF (when we look at the same range of cohorts).

[12] 1918 is the earliest cohort we can start with because the AET applied to annual earnings only starting in 1978. As our "treated" group at age 63 has their earnings measured (for determining treatment) in age 60, this means that people turning 60 before 1978 will have potentially the wrong measure of treatment. This rules out earlier cohorts than 1918.

[13] We considered selecting the cohorts that are exposed to the current policy regime. However, the AET rules for 63-64 year olds have remained essentially unchanged since 1978, when the AET began applying to annual instead of quarterly earnings, so this sample would include many earlier cohorts.

benefits, as described below.) Our estimation sample ends up with 3.58 million annual observations on 240,181 people.

We choose 64 as an 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. We cannot use earnings at ages 62, 63, or older as a proxy for desired earnings at even older ages because of potential intensive margin responses to the AET once individuals have reached EEA. 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 shown in Figure 1—for only part of the year. This consideration applies *a fortiori* to those over 65.

Several features of the data merit discussion. First, the underlying administrative data are subject to little measurement error. However, the public use files degrade the data to prevent re-identification: specifically, they randomly round earnings up or down to the nearest $100. 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.

Fourth, we do not observe date of death, and in the years after an individual dies, earnings and employment appear in the dataset as zeroes. However, we would not exclude deceased individuals even if we observed date of death, because the AET could in principle affect mortality, and we wish to avoid selecting the sample according to an outcome variable.

Some of an effect on employment could in principle be mediated through an effect on mortality. Indeed, Fitzpatrick and Moore (2018) find that just after individuals become eligible for Social Security at age 62 aggregate mortality rates rise by around 1.5 percent – a two percent increase for males, and a one percent increase for females. However, SSA actuarial life tables show that annual mortality rates at 62 (as well as 63 and 64) are around 1 percent; a back-of-the-envelope calculation shows that an annual rise in the fraction of the population that dies on the order of 1.5 percent of 1 percent (i.e. 0.015%) should mechanically be associated with far smaller employment effects than those we estimate. Such mortality effects would affect the interpretation, but not the validity, of our results; in other words, some of the effect on employment we estimate could be mediated through an effect on mortality, but the identification strategy would still be valid. The effects we estimate are relevant to policy, in the sense that they reflect the overall effect on employment.

In our main specifications, we control for a measure of benefits, because benefits are potentially correlated with treatment status, and benefits would confound our main estimates if there is a liquidity or income effect of benefit receipt. As benefits are unobserved in the EPUF, we must impute them.[14] Specifically, we impute the benefits that an individual would receive if she claimed benefits at age 62 and had no age 62 earnings. We focus on age 62 benefits because we are interested in income effects among early claimers. One difficulty in controlling for benefits is that they are mechanically endogenous to employment: people with positive earnings at ages 55-61 will have higher imputed benefits (as well as actual benefits). To avoid the resulting bias from conditioning on an intermediate outcome, we only use earnings through age 54. Specifically, we simulate earnings up to age 61 by fixing age 55-61 earnings at their age 54

[14] We use a benefit calculator provided by Social Security for this imputation. We downloaded the executable version of the calculator available at https://www.ssa.gov/OACT/anypia/anypia.html.

earnings. This procedure introduces a small amount of measurement error in benefits. However, using observed benefits in the BEPUF, we show in Appendix B that our imputed benefits are very close to actual benefits among people who claim at age 62. [15] We show in robustness tests that our results are very similar even if we instrument for simulated benefits, to correct the limited remaining measurement error.

Table 1 shows summary statistics for our sample. In the first column we report statistics for all observations in our cohort range. Column (2) is limited to our estimation sample, and columns (3) and (4) report statistics separately for our control and treatment groups, i.e. people with base age earnings below or above the exempt amount, respectively. In our estimation sample, the mean yearly employment rate among 55 to 64-year-olds, *i.e.* the percent of the corresponding calendar years when the individual has positive earnings, is 74.11 percent. Mean earnings (including zeroes) at these ages is $25,972 in our main sample. 53.38 percent of the sample is female. For comparison we also show the full sample, not restricted to those with positive base age earnings. Throughout the paper, all dollar figures are expressed in real 2010 dollars.

VI. Results

A. Initial results

As an initial exercise, in Figure 6 we show the density of earnings at ages 60 and 62, relative to the exempt amount. Figure 6 shows that the density of earnings at age 60 appears smooth near the exempt amount, and that the amount of bunching, calculated using the method

[15] We do not simply use the benefits reported in BEPUF because these data report benefits as of 2004, so they include cost-of-living adjustments, delayed retirement credits, and actuarial adjustment for early claiming. They do not measure what we seek, benefits at age 62. However, we can still use the BEPUF to validate our measure: among people who claimed at age 62 in 2004, actual and imputed benefit should coincide closely, and they do. See Appendix B.

of Chetty, Friedman, Olsen, and Pistaferri (2011), is statistically insignificant. This supports the validity of our identification strategy: if the density hypothetically showed evidence of a reaction at age 60 to the future imposition of the AET, this could confound the validity of comparing those under and over the exempt amount at age 60. For comparison, Figure 6 also shows the earnings distribution at age 62, 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 (as documented in Friedberg 2000 or Gelber, Jones, and Sacks 2013).

B. Main results

Figure 7 shows the main results graphically. On the *x*-axis is an individual's age in year $a+3$, where a is the base age. The figure shows the difference in employment rates between the treatment and control groups as a function of age, after adjusting for out controls.[16] After essentially remaining stable from outcome ages 53 to 62, the figure shows a sharp decrease from 62 to 63; age 63 is exactly when individuals will first be able to show an employment reaction to the AET three years later, when they are age 63. This is followed by a further substantial decrease from 63 to 64, consistent with a lagged adjustment to the AET (Gelber, Jones, and Sacks 2013, 2018).

Figure 7 shows that the trends in employment for those earning above and below z^* during ages prior to 63 are very similar. Thus, we have reason to believe that anticipatory adjustment to the AET is not a significant issue in our context, as those who are likely to not face the AET have a similar trend in outcomes as those who are most likely to face the AET. Echoing

[16] The controls are sex, age and cohort fixed effects, imputed benefits, and benefits interacted with post. To adjust for these controls, we regress a positive earnings at $a + 3$ dummy on demeaned values of these controls, plus a set of fixed effects for age and age x treat. We plot the coefficients on age x treat.

these results, Appendix Figure 1 shows that for those with earnings above z^* in the base year, the probability of positive earnings falls sharply and substantially from ages 62 to 63, while for those earning below z^* in the base year, the probability of having positive earnings falls to a much smaller extent between ages 62 and 63.

Table 2 shows the regression results corresponding to regression (1) above. We begin with uncontrolled DID estimates in columns (1) and (4), add the benefit controls in columns (2) and (5), and further include demographic controls in columns (3) and (6). Our preferred specification is in column (3), as it is the most tightly controlled and has the least distance between the age at which treatment is measured and the age at which the outcome is measured. In this specification, the DID coefficient is about -3, meaning the probability of positive earnings falls by about 3 percentage points more for people with base age earnings above the exempt amount than for people with lower base age earnings, relative to pre-age 63 trends and adjusting for benefits, age, cohort, and sex. This estimate is not sensitive to the demographic controls, but the benefit controls do matter; our estimate would be about 20% higher without these controls.

To help gauge the magnitude of this coefficient, we transform it into a participation elasticity with respect to the average net-of-tax rate. To do so, we assume that everyone faces a marginal tax rate of 25 percent below the exempt amount and before claiming (corresponding to state, federal, local, and payroll taxes), and a marginal tax rate of 75 percent above the exempt amount. Ignoring actuarial adjustment and eventual benefit exhaustion, our estimates imply an elasticity of about 0.1.[17] We scale this elasticity up by 1/0.4 because only about 40% of people claim before turning 63, so our baseline elasticity is 0.26. This participation elasticity is on the in the middle of the extensive margin elasticities summarized in Chetty (2012), which range from

[17] We calculate this elasticity as $\frac{\beta^{PE}}{\beta^{ANTR}} \cdot \frac{\mu^{ANTR}}{\mu^{PE}}$, where β^X is the DID estimate for X and μ^X is the sample mean of X.

0.15 to 0.43, although those elasticities are generally for prime age women. Our baseline elasticity does not account for benefit exhaustion or measurement error in the treatment variable. Accounting for benefit exhaustion would raise our elasticity to 0.37 (using our benefit imputation measure, which likely overstates exhaustion and hence overstates the elasticity). Accounting for measurement error would raise the elasticity to 0.45. Accounting for both would raise the elasticity to 0.64. Fully accounting for claiming, actuarial adjustment, and benefit exhaustion, among other details, is beyond the scope of this paper, so we view this calculation as a useful illustration rather than a definitive estimate of a structural parameter.

The coefficient we estimate has implications for the effect of the AET on the overall employment rate. Our baseline linear specification without controls shows that the AET reduces the employment rate of those in our treatment group by 2.9 percentage points. However, this does not mean that the AET reduces the overall employment rate at ages 63 to 64 by this amount, for two reasons. First, at ages 60 and 61 averaged, 46 percent of individuals are not employed and therefore are not in our sample. Second, among employed 60- to 61-year-olds in our sample, 25 percent of have earnings that are below the exempt amount, implying that they are not subject to treatment. After deflating due to both of these factors, our estimates imply that the AET reduces employment by 1.2 percentage points in the group we study at ages 63-64 (against a base of 39 percentage points). If anything, this reflects a lower bound, for example because we observe desired earnings with error.[18] As we estimate a signal-to-noise ratio in the pre-period of about 0.57, a less conservative estimate would entail scaling up our estimate by 1/.57, resulting in a population-wide effect of 2.1 percentage points.

[18] One caveat is that our measure of employment may not count a shift from the formal labor market to "off-the-books" employment (Christensen 1990).

Figure 8 provides a further key piece of evidence that the AET has a strong effect on employment. Figure 8 shows how the estimated treatment effect on employment at age $t + 3$ varies by distance to exempt amount. To construct the figure, we replace the “treat” dummy in our main specification with a set of dummy variables for binned distance to the exempt amount. (We omit the bin just below the exempt amount as a normalization.) Aside from the lowest two bins (in which people have low earnings and little labor force attachment), there is no treatment effect for bins below the exempt amount. The treatment effect initially increases in absolute magnitude with distance to exempt amount. At higher distances above the exempt amount, the treatment effect begins to fall in absolute magnitude. This is the U-shaped pattern we would expect to see, as shown in Figure 3. Theory predicts that the effect of the AET will be greatest at the point where benefits exhausted, the “nonconvex kink.” We do not know where benefits are exhausted (because we lack data on family structure and spousal benefits), but we can make a crude approximation with some simple assumptions; we report in the figure the points where imputed individual benefits would be exhausted on average, as well as where an estimate of family benefits would be exhausted.[19] We find that the treatment effect indeed appears to be most negative in the range where benefits are exhausted, on average. We view this U-shaped pattern as helpful confirmation that the employment effects we estimate are due to the effects of the AET, which could be expected to generate exactly such a pattern of effects. It is arguably difficult to construct alternative explanations for this U-shaped pattern.

C. Heterogeneity in responses to the earnings test

[19] To precisely determine where benefits are exhausted, we would need data on spousal benefits; as described above, when one spouse is a dual or secondary beneficiary, the AET reduces the combined benefits of both spouses at the BRR. Our data do not contain information on spousal benefits. The calculation in Figure 8 assumes that 95% of claimants are married (Kreider, 2005), that each male beneficiary’s spousal benefit is the average women’s benefit, and vice versa for women.

The overall responses to the earnings documented in Table 2 mask at least two dimensions of heterogeneity. In Table 3 we re-estimate our main DID models, but stratifying on sex (in columns (1) and (2)) and permanent income (in columns (3) and (4)). Men and women both respond to the earnings test, but women appear more responsive. Interestingly, people with high permanent income appear to drive all the response that we observe. We define permanent income as average income (when positive) at age 35-54, and high permanent income people as people whose permanent income is above their sex-cohort median.[20] (The "high permanent income" group represents more than half the estimation sample because we use the full data set to calculate median permanent income.) For high permanent income people, we estimate a DID of -5.6, but for low permanent income people, the estimate is wrong signed, small, and insignificant. Although low permanent income people have lower income at ages 55-61, and hence they are less exposure to the AET (i.e. they face a smaller bite), this difference does not account for their lower responsiveness, as the estimated elasticity is also much smaller (and wrong signed when $t = 3$). We argue in Section 5.F below that these heterogeneous responses are helpful for understanding the mechanisms behind the overall response to the AET.

D. Responses at older ages

Our main analysis focuses on people younger than NRA, because that group is still subject to the AET and because at older ages there is a greater gap between the timing when treatment is measured and when outcomes are observed. People NRA and above face a higher exempt amount under the AET. In 2000, the AET was eliminated for people over NRA. Nonetheless in Table 4 we change our sample to include people at NRA and just above.

[20] We have also estimated models that define high permanent income based on the overall median rather than the cohort and sex specific median. The results are nearly identical.

Specifically, in column (1) we look with a lead of $t = 4$ and look at base ages 55-61, so we include up to age 65; in column (2) we look with a lead of $t = 5$ and so include up to age 66.[21] In the top panel we simply expand our sample to include these older ages. The point estimates are larger, indicating a larger response at these ages. There are several reasons we might see a larger response at these ages. First, employment at older ages may be more marginal and therefore more sensitive to incentives. Second, individuals may take time to react to the Earnings Test, consistent with the results in Gelber, Jones, and Sacks (2019). Third, as claiming increases with age, more and more people's earnings are actually subject to the earnings test.

E. Robustness checks

Our estimates are robust to several alternative specification choices. We present the results of several robustness checks in Table 5. For comparison, the first column is our baseline, corresponding to the results in columns (3) and (6) in Table 2. Our robustness tests begin in column (2), where we consider an alternative measure of *treat*. Specifically, rather than use base age earnings, we measure average earnings in the base age and the prior two years, and we now define *treat* as having three-year average earnings over the exempt amount. The idea behind this robustness test is that our baseline measure of *treat* is potentially measured with error, as earnings vary from year to year. This alternative measure may reduce the noise in *treat*, albeit at the cost of using earnings information that is more years behind the outcome age. This alternative measure has very little effect on our estimates.

In our main specification we use all earnings ranges, but a concern with this specification is that the control group includes people earning just a few thousand dollars or less, while the

[21] The exempt amount differs between those below and above NRA, but the difference is not large for most of our sample period; the greatest divergence happens post-2000 with the elimination of the AET for those older than NRA. In Table 4 we continue to define $treat$ as an indicator for base age earnings above the pre-NRA exempt amount.

treatment group includes people earning over $100,000. One might worry that such groups have very different employment patterns in the older years. To address this possibility, in column (3) we limit the sample to observations with base age earnings within $5,000 of the exempt amount, and in column (4) we instead limit the sample to observations with base age earnings between $5,000 and $50,000. In both specifications we estimate smaller DID coefficients than in the baseline, around -1.3 with the bandwidth of $5,000 and -2 with the wider (but still restricted) range. However, the reason for these smaller estimates is that the treatment is less intense for treatment group. That is, for people with earnings relatively close to the exempt amount, the bite of the earnings test is less severe. We therefore should expect a smaller treatment effect. We should also expect, however, a fairly similar elasticity after we adjust for this smaller first stage, and that is what we see: using the narrow bandwidth we have an elasticity of 0.42, and using the wider range we have an elasticity of 0.18, both compare to our baseline of 0.23.

In column (5) we show that our results are robust to using the full set of cohorts, including people born as early as 1918. The point estimates in fact rise by about 50 percent, as people born earlier respond more to the AET. Finally, in column (6) we attempt to correct for measurement error in a control variable, imputed benefits. We believe this is a potentially important control, but it is measured with error because as the imputation defining it holds income fixed at age 54 level. Because we observe actual benefits in the BEPUF, we can use two-sample two-stage least squares (TS2SLS) to correct for this measurement error. A limitation here is that we only observe actual benefits for people born in 1942 and claiming at age 62, so we cannot control for year-of-birth fixed effects. Correcting for measurement error does not appreciably change our estimates.

F. Placebo tests

We conduct two sets of placebo tests based on outcome ages younger than 63, and therefore not subject to the AET for the full year. In the first, we consider a wide range of possible exempt amounts. For a grid of exempt amounts between -7,000 and +7,000 (relative to the true exempt amount), we re-estimate our main DID model, but with treatment defined relative to the placebo exempt amount, the "post" period defined as outcome ages 60-61, and looking only at outcome ages 58-61 (to avoid picking up true effects). We plot the results in Figure 9. The placebo estimates are almost all positive (i.e. wrong-signed), small in magnitude, and all but one is statistically insignificant. As this exercise somewhat arbitrarily sets the treatment ages to 58 and 59, as a complementary placebo test, in Figure 10 we show the results of re-estimating our main specification, but setting the treatment ages to each two year window 59-60, 60-61, 61-62, and 62-63 (and excluding people older than the maximum treatment age, again to avoid picking up true effects). The placebo estimates are again positive at the younger treatment ages. They turn negative at 62 but this treatment bin includes outcome age 63, which is in fact treated. These placebo tests show that differential trends in retirement for lower earners, relative to higher earners, do not explain our results.[22]

VII. Discussion

Our finding of a substantial employment response to the AET complement those in Gelber *et al.* (2018), who focus on patterns within $3,000 of the exempt amount, and therefore leave open the questions of whether the results generalize to a broader group and what the overall effect on employment is in this age range. Thus, the current study examines a broader

[22] In principle it is possible to construct placebo tests using the actual AET ages, but looking only at people with income below the exempt amount and using placebo exempt amounts that are even lower. This approach would therefore compare very low earners to people with income just below the exempt amount. A difficulty with this approach is that people with income just below the exempt amount are in fact much more likely to end up treated at $a + 3$ than are people with lower earnings; see Figure 5. This "placebo" test therefore runs the risk of picking up true treatment effects.

group of interest to policy-makers, and indicates that the local estimates in Gelber *et al.* (2018) appear to apply to a broader group, in the sense that in a broader sample we still estimate large and highly significant effects of the AET. In the online Appendix, we briefly recapitulate some of the key findings from Gelber *et al.* (2018), which bolster the credibility of the current results further.

There are several reasons we might expect the RKD and DID responses to differ. As Gelber *et al.* (2018) note, the RKD strategy identifies a local effect around the exempt amount; the treated group therefore includes people with only a fairly small change in ANTR due to the AET. This would create a small overall treatment effect of the AET in the RKD analysis. On the other hand, the population average treatment effect that we report here for the DID analysis includes many people with zero earnings who bring down the average effect. In the end, the DID estimates are fairly similar to the RKD estimates. The DID estimates imply a decrease in aggregate employment of 1.2 percentage points, whereas the RKD of Gelber *et al.* (2018) estimates imply a of 1.4 percentage points among people with earnings near the exempt amount. Thus it turns out that the local response provides a reasonable approximation to the population-wide response. We caution, however, that the estimator is not the only difference between the approach here and in Gelber *et al.* (2018); that paper looked at earlier cohorts (1918-1923) and excluded people with self-employment income. When we look at the same range of cohorts, we get estimates that are roughly 25% higher than our main estimates, closely in line with Gelber *et al*. (2018).

Since the previous literature examined an older age group, it is difficult to formally reconcile our results directly with those in previous literature, as the effect of the AET could simply be larger in the group we examine than in the older age group. However, it is possible to see a foreshadowing of our results in previous literature, assuming that the impact of the AET is comparable in the older and younger groups. First, it has been widely noted that employment rates for older Americans (particularly those over 65) increased substantially after the mid-1980s (Goldin and Katz 2018); Song and Manchester (2007) note that employment rates did increase more in their treatment group than in their control group, though it is difficult to tell this trend apart from the relative rise in employment among those over 65 that pre-dated the year 2000. Our results suggest that some of the rise in employment over age 65 in the post-2000 period is due to the elimination of the AET, as opposed to being a continuation of pre-existing trends.

Second, Friedberg and Webb (2009) found significant effects of the AET in the over-65 population in some specifications in the survey data of the Current Population Survey; it is possible that with the greater sample size and precision of the EPUF, we are able to uncover more robust impacts than those found in their setting. Third, the standard errors in previous literature generally did not rule out substantial impacts of the AET on employment (and in some cases did indicate some evidence of an impact, as discussed above). Overall, our results solidify the case for a substantial impact of the AET on employment.

VIII. Conclusion

We show that the AET plays a substantial role in determining older workers' labor force participation decisions, reducing the employment rate of the workers we study by 1.2 percentage points under conservative assumptions, or as much as 2.1 percentage points if we explicitly adjust for measurement error. Our results are more generally consistent with the view that the retirement decisions of older individuals are rather sensitive to perceived incentives (*e.g.* Laitner and Silverman 2012).

Our results also imply that the planned increases in the Normal Retirement Age, to 67 in

2026, may reduce labor force participation by exposing seniors to the Earnings Test for longer. If individuals react to Social Security incentives only after learning about them (Gelber, Jones, and Sacks 2013), then the impacts we find could increase over time at older ages as more individuals learn about the Earnings Test over time after ages 63-64.

This paper is complementary to Gelber *et al.* (2019), who find that individuals' response to the AET is consistent with mis-perceptions of the AET. In particular, individuals disproportionately "bunch" just below the exempt amount rather than just above it, and the employment rate discontinuously drops just above the exempt amount relative to just below it. Both patterns are consistent with individuals believing the AET applies to *all* earnings below the exempt amount, rather than applying only to marginal earnings above the exempt amount; individuals believe that the benefit reductions due to the AET are much more severe than they are in reality. These misperceptions help to explain the large employment reactions to the AET that we document in this paper.

It is important to caution that our results do not necessarily imply that the AET policy is undesirable. By lowering AET benefits earlier in an individual's period of claiming, and thus raising later benefits through benefit adjustment, the AET on average shifts the profile of OASI benefits later in the lifecycle. All else equal, this could lead to (1) welfare gains if the optimal time profile of Social Security benefits slopes upward (Feldstein 1990); (2) more redistribution in terms of yearly income—as older claimants on average have lower income and assets (as well as worse health) than younger claimants; and (3) lower poverty rates among older Americans (see Figinski and Neumark 2018 on older women). Nonetheless, the substantial employment effects we have documented in our work represent an important input into a calculation of the welfare effects of the AET, in particular by suggesting important distortions from the AET.

Evaluating the full welfare and redistributive consequences of the AET, including both the employment and earnings impacts, as well as the impacts through the lifecycle profile of OASI benefits, is an important topic left to future work.

References

Abraham, Katharine, and Melissa Kearney. 2017. "Explaining the Decline in the U.S. Employment-to-Population Ratio: A Review of the Evidence." NBER Working Paper 24333.

Baker, Michael, and Dwayne A. Benjamin. 1999. "How do retirement tests affect the labor supply of older men?" *Journal of Public Economics*, 71, 27-51.

Brown, Jeffrey, Arie Kapteyn, Olivia Mitchell, and Teryn Mattox. 2013. "Framing the Social Security Earnings Test." Wharton Pension Research Council Working Paper.

Burtless, Gary, and Robert A. Moffitt. 1985. "The Joint Choice of Retirement Age and Postretirement Hours of Work." *Journal of Labor Economics*. 3: 209-236.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *The Quarterly Journal of Economics* 126(2): 749-804.

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

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

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

Coile, Courtney C. and Jonathan Gruber. "Future Social Security Entitlements and the Retirement Decision." *The Review of Economics and Statistics* 89(2):234-246, May 2007.

Diamond, Peter, and Jonathan Gruber. 1999. "Social Security and Retirement in the United States." In *Social Security and Retirement Around the World* (Jonathan Gruber and David A. Wise, eds), 437-473. Chicago: University of Chicago Press.

Disney, Richard, and Sarah Smith. 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, Nada, Henrik Kleven, and Claus Kreiner. 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.

Engelhardt, Gary, and Anil Kumar. 2009. "The Repeal of the Retirement Earnings Test and the Labor Supply of Older Men." *Journal of Pension Economics and Finance* 8(4), 429-450.

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

Feldstein, Martin. 1990. "Imperfect Annuity Markets, Unintended Bequests, and the Optimal Age Structure of Social Security Benefits." *Journal of Public Economics* 41(4): 31-43.

Fetter, Daniel, and Lee Lockwood. "Government Old-Age Support and Labor Supply: Evidence from the Old Age Assistance Program." Forthcoming, *American Economic Review*.

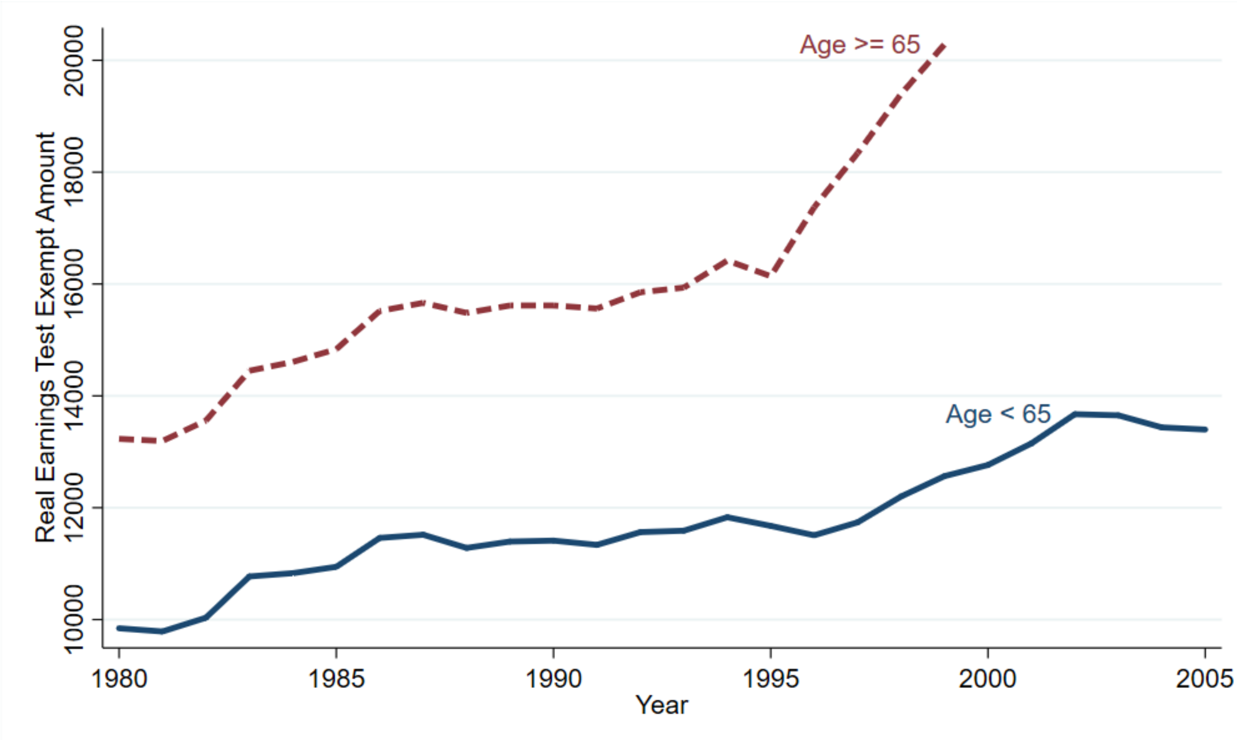
Figinski, Theodore, David Neumark, 2018. "Does Eliminating the Earnings Test Increase the Incidence of Low Income Among Older Women?" *Research on Aging*, 40(1): 27-53.

Fitzpatrick, Maria, and Timothy Moore, 2018. "The Mortality Effects of Retirement: Evidence from Social Security Eligibility at Age 62." *Journal of Public Economics*, 157: 121-137.

French, Eric. 2005. “The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour.” *The Review of Economic Studies* 72: 395-427.
Friedberg, Leora. 1998. “The Social Security Earnings Test and Labor Supply of Older Men.” In *Tax Policy and the Economy* (James M. Poterba, ed.), 121-150. Chicago: University of Chicago Press.
Friedberg, Leora. 2000. “The Labor Supply Effects of the Social Security Earnings Test.” *The Review of Economics and Statistics.* 82: 48-63.
Friedberg, Leora, and Anthony Webb. 2009. “New Evidence on the Labor Supply Effects of the Social Security Earnings Test.” In *Tax Policy and the Economy* (Jeffrey R. Brown and James M. Poterba, eds.), 1-35. Chicago: University of Chicago Press.
Gelber, Alexander, Damon Jones, and Daniel Sacks. 2013. “Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test.” NBER Working Paper No. 19491.
Gelber, Alexander, Damon Jones, Daniel Sacks, and Jae Song. 2018. “Using Kinked Budget Sets to Estimate Extensive Margin Responses: Evidence from the Social Security Earnings Test.” NBER Working Paper 23362.
Gelber, Alexander, Damon Jones, and Daniel Sacks. 2019. “Left-Bunching at Kinks: Evidence and Implications.” University of Chicago working paper.
Golden, Claudia and Lawrence Katz. 2018. *Women Working Longer*. Chicago: University of Chicago Press
Gruber, Jonathan, and Peter Orszag. 2003. “Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt?” *National Tax Journal.* 56: 755-773.
Gruber, Jonathan, and David Wise. 1999. "Introduction and Summary." In Gruber, Jonathan, and David Wise, *eds*., Social Security and Retirement Around the World (pp. 1-35). Chicago: University of Chicago Press.
Gustman, Alan, and Steinmeier, Thomas. 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.
Haider, Steven, and David Loughran. 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, Jerry. 1981. “Labor Supply.” In Aaron, Henry and Pechman, Joseph., *eds*., How Taxes Affect Economic Behavior (pp. 27-71). Washington, DC: Brookings Institution.
Heckman, James. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica*, 47(1), 153-161.
Kreider, Rose M. 2005. *Number, Timing, and Duration of Marriages and Divorces: 2001*. Current Population Reports, P70-97. U.S. Census Bureau, Washington, D.C.
Kleven, Henrik, and Mazhar Waseem. 2013. “Using Notches to Uncover Optimization Frcitions and Structural Elasticities: Theory and Evidence from Pakistan.” *Quarterly Journal of Economics*, 128(2), 669-723.
J.K. Lasser Institute. 1997. *J.K. Lasser’s Your Income Tax 1998*. New York: MacMillan.
Laitner, John, and Dan Silverman. 2012. “Consumption, Retirement, and Social Security: Evaluating the Efficiency of Reform that Encourages Longer Careers.” *Journal of Public Economics*, 96, 615-634.
Leonesio, Michael. 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.

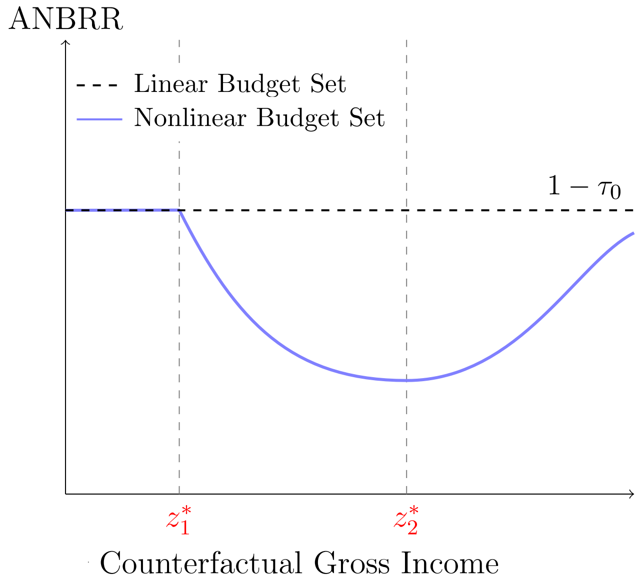
Liebman, Jeffrey B., and Erzo F.P. Luttmer. 2015. "Would People Behave Differently If They Better Understood Social Security? Evidence From a Field Experiment." *American Economic Journal: Economic Policy* 7(1): 275-299.
Meyer, Bruce, and Dan Rosenbaum. 2001. "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics*, 116(3), 1063-1114.
Powell, James. 1994. "Estimation of Semiparametric Models." In Engle, R. F., and McFadden, D., *eds*., *Handbook of Econometrics* (pp. 2443-2521). Amsterdam: Elsevier.
Reimers, Cordelia and Marjorie Honig. 1996. "Responses to Social Security by Men and Women: Myopic and Far-Sighted Behavior." *Journal of Human Resources* (1996): 359-382.
Rubin, Donald. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66(5), 688-701.
Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2, 180-212.
Social Security Administration, 2019. Program Operations Manual System. Available at https://secure.ssa.gov/apps10/poms.nsf/lnx/0300615482.
Song, Jae G. 2003. "Evaluating the Initial Impact of Eliminating the Retiremetn Earnings Test." *Social Security Bulletin* (2003): 1-16.
Song, Jae G., and Joyce Manchester. 2007. "New evidence on earnings and benefit claims following changes in the retirement earnings test in 2000." *Journal of Public Economics.* 91: 669-700.
Viscusi, W. Kip. 1979. *Welfare of the Elderly: An Economic Analysis and Policy Prescription*. New York: John Wiley & Sons.
Vroman, Wayne. 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. *Earnings Test Real Exempt Amount, 1901 to 2005*



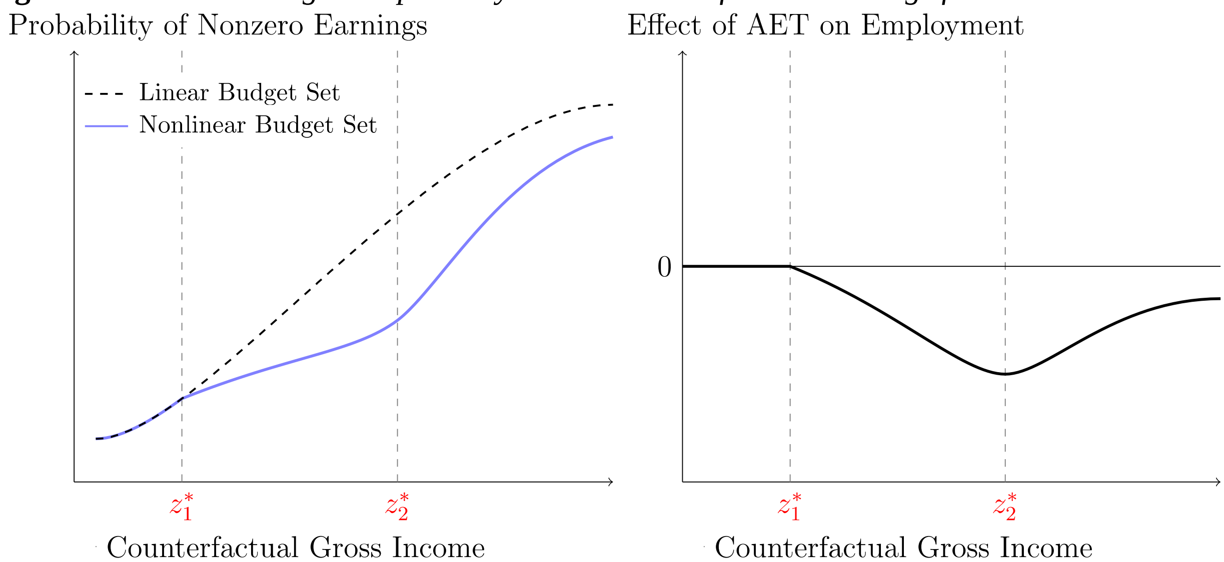
Notes: The figure shows the real value of the exempt amount over time among those 62-64 years old (labeled “Age<65” in the graph) and those 65 and above. The AET applied to earnings of claimants from ages 62 to 71 from 1981 to 1982, but only to claimants aged 62 to 69 from 1983 to 1999, and only claimants normal retirement age and younger from 2000 onward. For post-2000 we plot the exempt amount that applies to workers who are younger than normal retirement age. All dollar figures are expressed in real 2010 dollars.

Figure 2. *Extensive Margin Incentives*



Notes: The figure shows the ANBRR (*y*-axis) as a function of counterfactual gross income (*x*-axis). The ANBRR is defined as the fraction of gross income an individual keeps, net of benefit reduction and taxes. Incentives under a linear tax schedule in which the ANBRR is equal to $1 - \tau_0$ everywhere are represented by the dashed line. Incentives under a kinked tax schedule – in which the AET is imposed on earnings above $z_1^*$and Social Security benefits are phased out entirely by z_2^* – are represented by the solid line. For each individual, the earnings level z_2^* is a function of their Social Security benefit, as discussed in the Appendix.

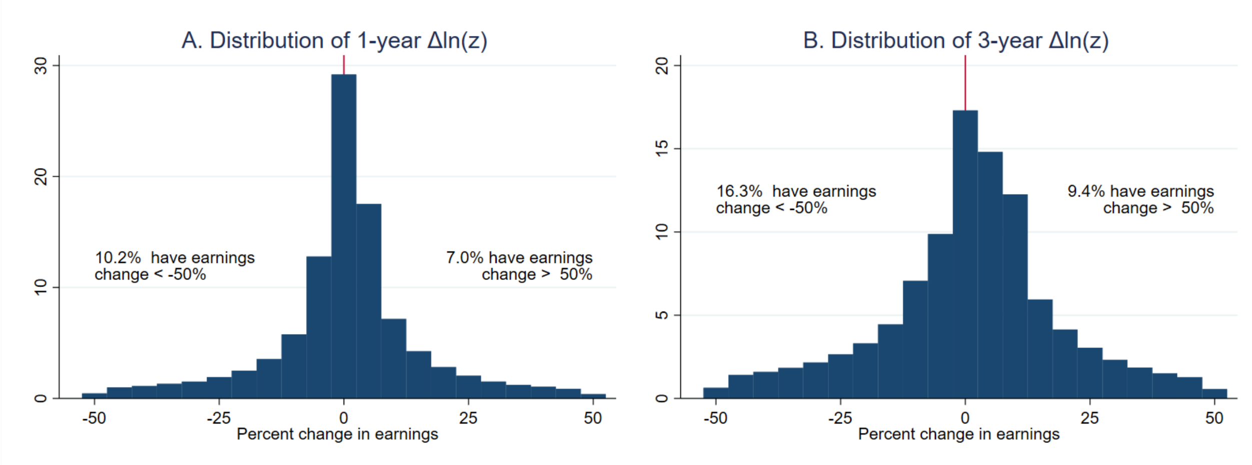
Figure 3. *Extensive Margin Response by State 0 Counterfactual Earnings for a Kink*



Panel A *Panel B*

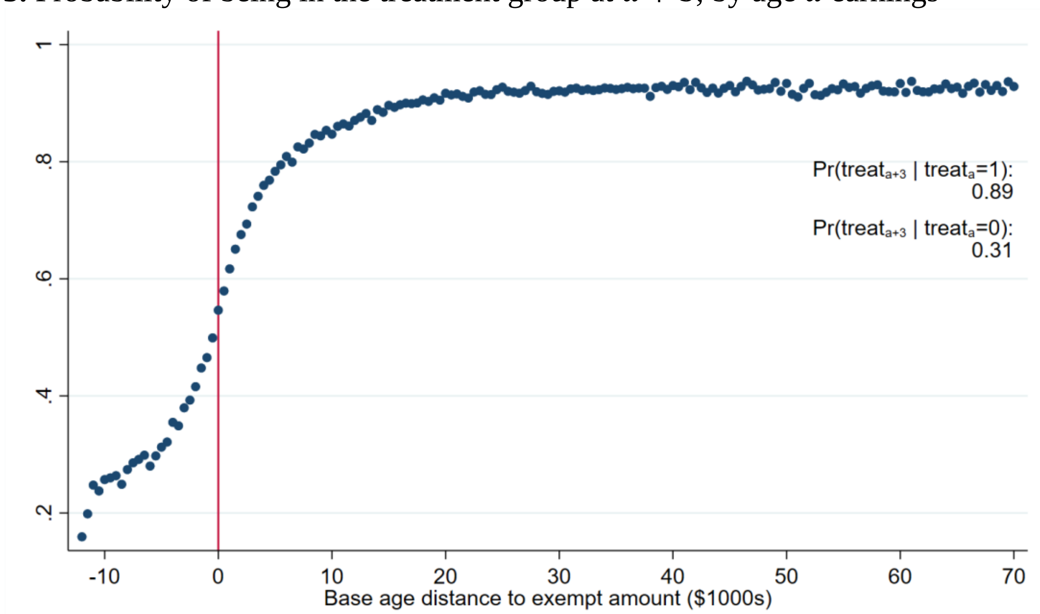
Notes: The figure illustrates the extensive margin response to the imposition of a kink, by counterfactual gross income. The *x*-axis shows desired gross income if employed on a linear budget set in the absence of the AET, $\tilde{z}_0$. The y-axis shows a hypothetical probability of employment, under two scenarios: a linear schedule in the absence of the AET (dashed line) and a kinked tax schedule in the presence of the AET (solid line). The figure illustrates that the effect of the AET on employment grows as the AET reduces the ANBRR above the convex kink, and shrinks above the non-convex kink, eventually asymptoting to zero. If the marginal effect of the ANBRR on employment does not depend on counterfactual earnings, the magnitude of the effect of the AET on the employment rate should decrease beginning at the non-convex kink, which is the case shown in Figure 3. However, if the marginal effect of the ANBRR on employment depends on counterfactual earnings, then the effect of the AET on the employment rate may be maximized at a counterfactual earnings level other than the level associated with the non-convex kink. In either case, the effect of the AET on employment follows a U-shaped pattern, as in Figure 3.

Figure 4. *Histogram of Percent Real Earnings Growth*



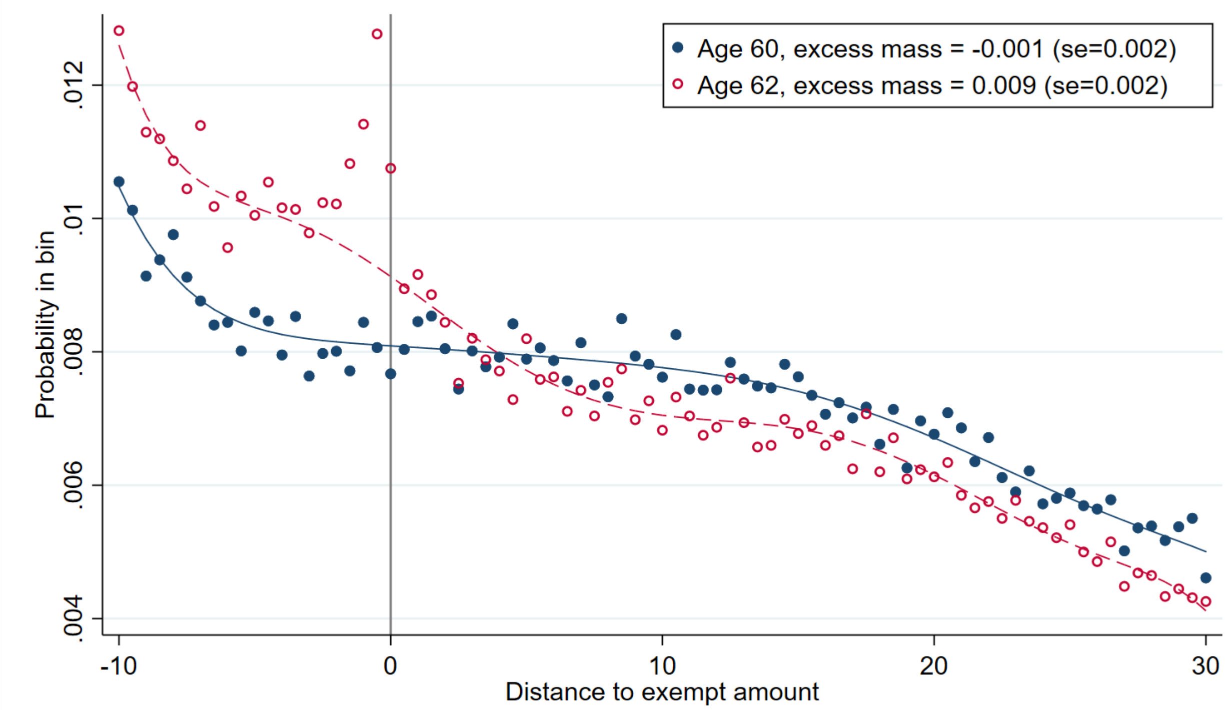
Notes: The figure shows the distribution of 1- and 3-year growth rates from base ages 55-58, among observations with growth rates less than 50%. This histogram shows that there is a large mass at or near zero percent real earnings growth across a "placebo" set of ages when individuals do not face the AET. This indicates that a substantial mass of individuals have no growth in desired real earnings, consistent with the assumptions necessary for our empirical strategy as described in the main text. Real earnings in each year are calculated using the CPI-U.

Figure 5. Probability of being in the treatment group at $a + 3$, by age a earnings



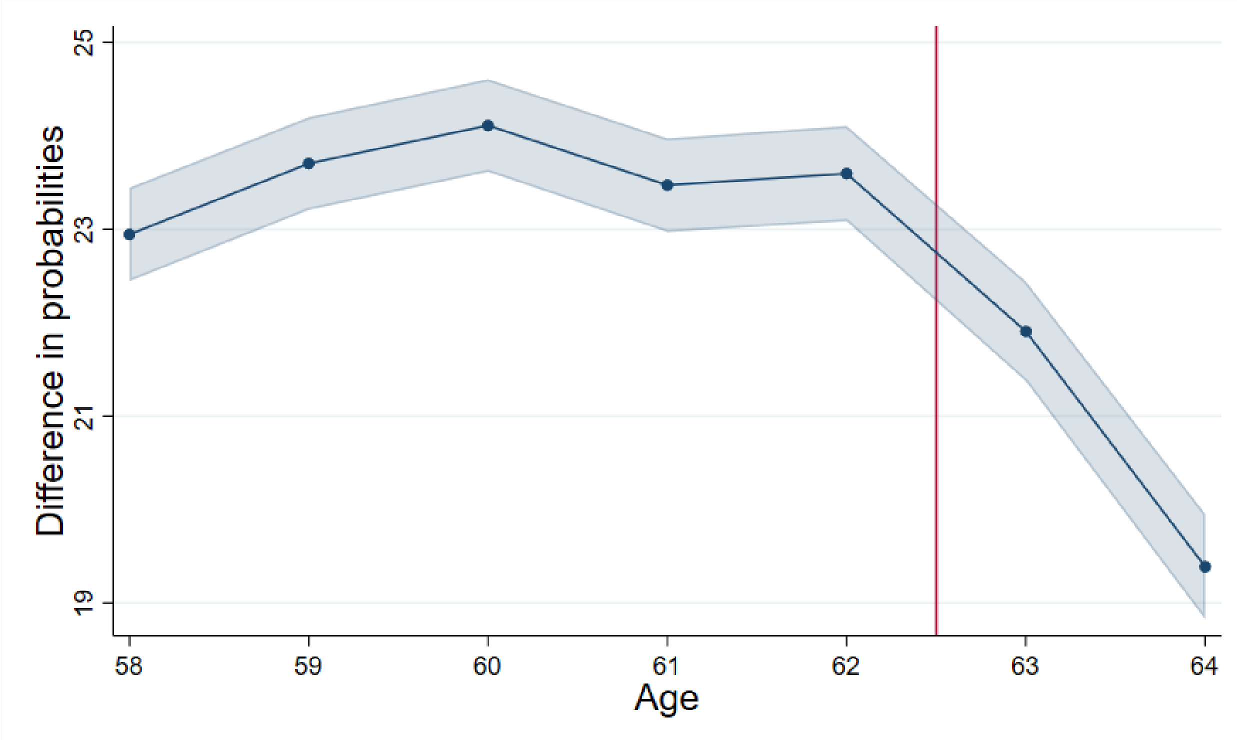
Notes: Figure plots, for each $500 bin of base age distance to exempt amount, the fraction of observations that are treated at $a + 3$, i.e. the fraction with age $a + 3$ earnings above the exempt amount. The sample is limited to people in base ages 55-58, so that $a + 3$ is not subject to the earnings test, and further limited to observations with positive earnings at a and $a + 3$ (so that observed earnings can proxy for desired earnings).

Figure 6. *Earnings Distributions at Ages 60 and 62*



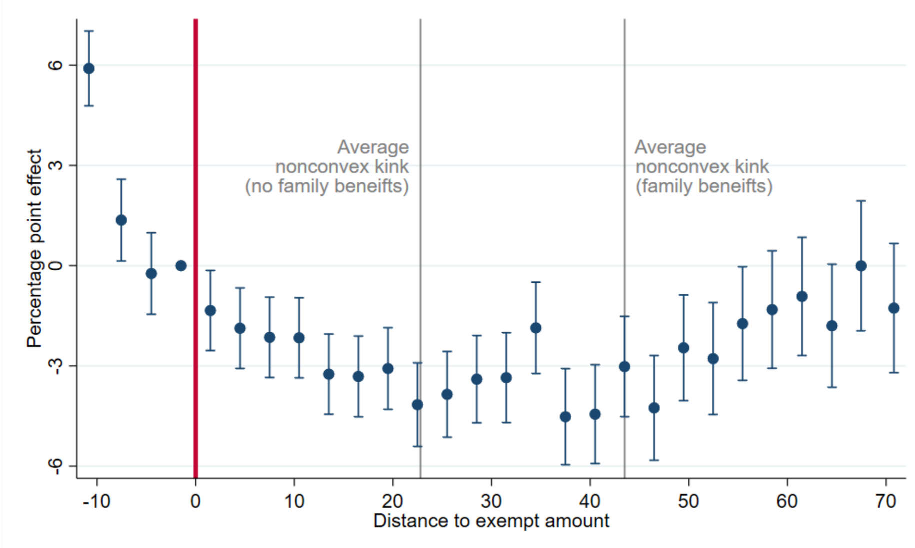
Note: Figure shows the earnings distributions at ages 60 and 62. We also report the excess mass (relative to the smoothed fit) within $3,000 of the exempt amount. We use the method of Chetty et al. (2011) to calculate this excess mass, although we do not impose the constraint that the counterfactual density sums to 1. The sample is drawn from the EPUF and consists of individuals with positive earnings at the indicated age, born 1931-1943.

Figure 7. *Difference between probability of positive earnings among those earning above and below the exempt amount, by age*



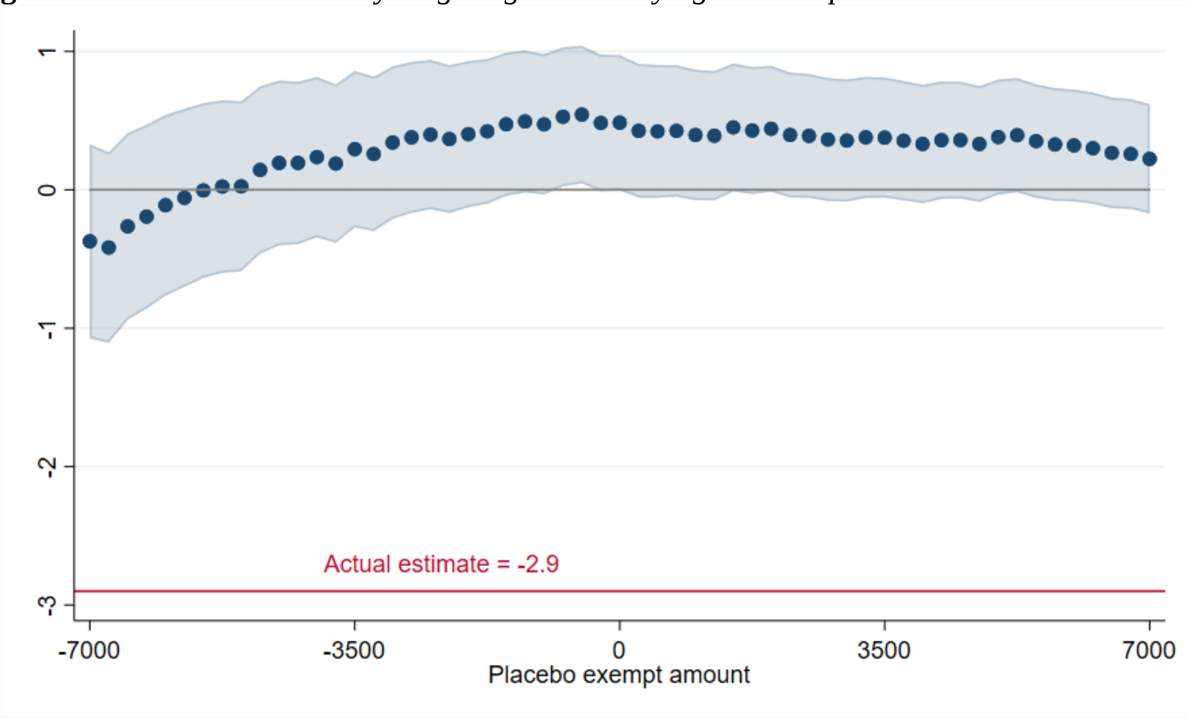
Notes: the figure shows (*y*-axis) the difference between those earning above *z** and those earning below *z** at age *a* in the probability of positive earnings at age *a*+3, as a function of age *a*+3 (*x*-axis). To parallel our main specification, we plot differences adjusting for our controls (female, age and cohort fixed effects, imputed benefits, and imputed benefits interacted with post). The shaded area is the 95% confidence region.

Figure 8. *Effect of AET on Employment, by Base Age Distance to the Exempt Amount*
by Base Age Distance to the Exempt Amount



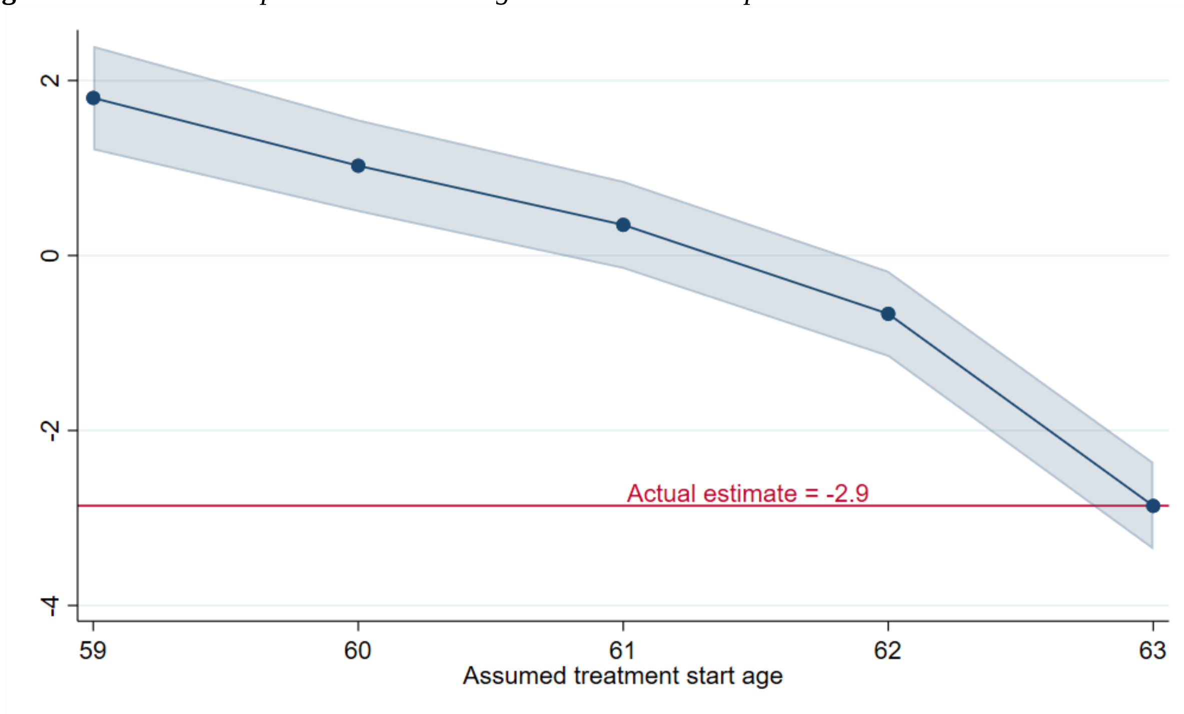
Notes: the figure shows the treatment effect of the AET on employment, for each bin in base year earnings relative to the exempt amount. We normalize the effect for the bin just below the exempt amount to zero. We report two estimates of the location of the average nonconvex kink, i.e. the earnings level where benefits are exhausted. The first location ignores family benefits and is based only on own imputed benefits. The second location is calculated as follows. First, we assume that everyone has a 95% chance of being married. We assume spousal benefits for men are the average female benefits in our sample, and likewise for women. We calculate family benefits as own benefits plus spousal benefits if married .The average nonconvex kink is the average distance at which family benefits are exhausted by the AET, i.e. the annual family benefit divided by the benefit reduction rate. The error bars represent 95% confidence intervals.

Figure 9. *Placebo estimates at younger ages and varying the exempt amount*



Notes: Each point in the figure is the estimate from a placebo regression in which the exempt amount is set as indicated, we exclude people older than base age 58, and we define the post period as base age older than 57. The shaded area is the 95% confidence region. The shaded area is the 95% confidence region.

Figure 10. *Estimates at placebo treatment ages but the true exempt amount*



Notes: Each point in the figure is the estimate from a DID regression in which treatment assumed to start at the indicated age. The sample underlying each point includes base ages from 55 to the treatment age – 2 (so that the "post" period lasts two years in all cases). The age 62 placebo includes one truly treated age (63) and the age 63 point corresponds to our main estimate. The shared area represents the 95% confidence region.

Table 1. *Summary statistics: means (standard deviations) of main variables*

Sample	All (1)	Estimation (2)	Control (3)	Treatment (4)
Positive earnings dummy, ages 55-64	55.45 (40.60)	74.11 (29.39)	66.56 (29.38)	80.49 (24.77)
Earnings, ages 55-64	19,290 (23,451)	25,972 (23,836)	18,139 (17,276)	30,755 (23,260)
Female	52.04 (49.96)	53.38 (49.89)	51.68 (49.97)	56.92 (49.52)
Year of birth	1937.36 (3.80)	1937.44 (3.80)	1937.30 (3.78)	1937.46 (3.80)
# Observations	3,206,255	2,377,854	1,064,122	1,313,732
# People	323,743	240,181	185,980	200,532

Notes: Table reports the mean and (in parentheses) standard deviation of the indicated variable. The sample is drawn from the EPUF. "All" sample consists of people aged 55-61 and born 1931-1943. Estimation sample is further restricted to people with positive earnings. The treatment group is person-years with earnings above the exempt amount, and the control group is everyone else.

Table 2. *Main estimates*

	t = 3 years ahead			t = 4 years ahead		
	(1)	(2)	(3)	(4)	(5)	(6)
DID coefficient	-3.7 (0.2)	-2.9 (0.3)	-2.9 (0.3)	-4.0 (0.3)	-3.1 (0.3)	-3.0 (0.3)
# Observations	1,351,688	1,351,688	1,351,688	1,181,301	1,181,301	1,181,301
# People	240,181	240,181	240,181	237,839	237,839	237,839
Controls						
Benefits		X	X		X	X
Sex			X			X
Age FE			X			X
Cohort FE			X			X

Notes: Table shows the DID coefficient. The outcome is a dummy for positive earnings (x100) in the indicated number of years ahead. The sample is drawn from the EPUF. The sample consists of people born 1931-1943 with positive base age earnings, and base age 55 to $64 - t$. All specifications control for main effects of *post* and *treat*. The benefit controls are imputed benefits plus their interaction with *post*. Standard errors (in parentheses) are clustered by individual.

Table 3. *Heterogeneity in estimates by sex and permanent income*

Split	Male (1)	Female (2)	High PI (3)	Low PI (4)
		A. $t = 3$ years ahead		
DID coefficient	-2.0 (0.4)	-3.6 (0.4)	-5.6 (0.4)	0.5 (0.4)
Elasticity	0.18	0.26	0.38	-0.05
# Observations	611,711	739,977	844,539	507,149
# People	111,962	128,219	143,127	97,054
		B. $t = 4$ years ahead		
DID coefficient	-2.5 (0.4)	-3.5 (0.4)	-5.4 (0.4)	-0.1 (0.4)
Elasticity	0.23	0.26	0.38	0.01
# Observations	535,116	646,185	737,408	443,893
# People	110,622	127,217	141,984	95,855

Notes: Table reports DID estimates for the indicated group. The outcome is a dummy for positive earnings the indicated number of years ahead (x100). The sample is drawn from the EPUF. The sample consists of people born 1931-1943 with positive earnings in base ages 55 to $64 - t$. All columns include controls for *treat*, sex, age, and cohort fixed effects, plus imputed benefits interacted with post. High PI means high permanent income. Permanent income is defined as average real earnings across ages 35-55 (counting only ages with positive income. High PI people have permanent income above their sex-cohort median. The elasticity assumes a 40% OASI claiming rate. Robust standard errors, clustered on person, in parentheses.

Table 4. *Estimates including outcome ages 65 and above*

	$t = 4$ years ahead (1)	$t = 5$ years ahead (2)
	A. Include outcome ages 63-64	
DID coefficient	-4.0	-5.0
	(0.3)	(0.3)
# Observations	1,085,274	1,070,751
# People	193,494	193,327
	B. Exclude outcome ages 63-64	
DID coefficient	-5.9	-6.8
	(0.4)	(0.3)
# Observations	789,445	763,214
# People	191,589	191,233

Notes: Table reports DID estimates. The outcome is a dummy for positive earnings the indicated number of years ahead (x100). The sample is drawn from the EPUF. The sample consists of people born 1931-1943 with positive earnings in base ages 55 to 61. Panel B excludes people with base age $63 - t$ and $64 - t$. Additional controls include controls for *treat*, sex, age, and cohort fixed effects, plus imputed benefits interacted with post. Robust standard errors, clustered on person, in parentheses.

Table 5. *Robustness checks*

Specification	Main	Different proxy	BW =$5,000	Earnings in [5k, 50k]	All Cohorts	TS2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
			A. $t = 3$ years ahead			
DID coefficient	-2.9	-3.1	-1.3	-2.0	-4.3	-2.8
	(0.3)	(0.3)	(0.5)	(0.3)	(0.2)	(0.2)
Elasticity	0.23	0.28	0.432	0.18	0.33	0.22
# Observations	1,334,493	1,334,493	206,171	803,294	2,521,147	1,334,493
# People	239,995	239,995	87,670	180,968	454,856	239,995
			B. $t = 4$ years ahead			
DID coefficient	-3.0	-3.0	-1.2	-1.8	-4.6	-3.9
	(0.3)	(0.3)	(0.5)	(0.3)	(0.2)	(0.3)
Elasticity	0.25	0.29	0.41	0.17	0.37	0.32
# Observations	1,163,216	1,163,216	177,821	698,491	2,202,262	1,163,216
# People	237,571	237,571	80,383	175,036	450,030	237,571

Notes: Table reports DID estimates. The outcome is a dummy for positive earnings the indicated number of years ahead (x100). Column (1) reports ours main estimates, in which the sample (drawn from the EPUF) consists of people born 1931-1943 with positive base age earnings, and base ages 55 to $64 - t$. The controls in column (1) include controls for *treat*, sex, age, and cohort fixed effects, plus imputed benefits interacted with post. The remaining columns all involve one change relative to this baseline. In column (2) we define *treat* using average earnings (when positive) in the base age and the prior two years. In column (3) we limit the sample to people with base age earnings within $5,000 of the exempt amount. In column (4) we limit the sample to people with base age earnings between $5,000 and $50,000. In column (5) we expand the sample to include birth cohorts 1918-1943. In column (6) we use TS2SLS and BEPUF data (for age 62 claimants in the 1942 cohort only) to correct for measurement error in the benefits measure; this necessitates dropping the year-of-birth fixed effects. The elasticity assumes a 40% OASI claiming rate. For all specification we report robust standard errors, clustered on person, in parentheses.

Online Appendix

A. Framework for interpreting the results

In this section, we sketch a basic framework that is helpful in interpreting the empirical strategy and results. We model how the AET impacts an individual's decision of whether or not to have a positive amount of earnings, which we refer to as the "employment" decision. Throughout, we make use of a potential outcomes framework (Rubin, 1974). We index two potential states of the world by $j \in \{0,1\}$.

To capture the real-world features of how OASI benefits, taxes, and the AET work, our framework incorporates all three.[23] Following previous literature (*e.g.* Friedberg, 1998; Friedberg, 2000), we model the AET as creating a positive benefit reduction rate (BRR) for some individuals above the exempt amount, consistent with the empirical finding in this previous literature that some individuals bunch at the exempt amount. Individuals receive a level of current benefits that is potentially a function of earnings, *i.e.* $B_j(z)$, where $B_j(\cdot)$ denotes their current benefit in state *j*, and *z* denotes their pre-tax and pre-benefits earnings. The "pre-reduction" level of benefits is b, which refers to the OASI benefits received before accounting for the effects of the AET or taxes.[24] Current benefits, $B_j(z)$, are determined both by b as well as by any reductions in benefits due to the AET. Finally, there is a linear tax on earnings, *i.e.* $T(z) = \tau_0 z$, which does not vary by state.[25] This tax, which reduces net earnings relative to gross earnings, is separate from the AET, which only acts to reduce OASI benefits.[26] Total post-tax and post-benefit resources are therefore:

$$z - T(z) + B_j(z) = (1 - \tau_0)z + B_j(z).$$

In state 0, when there is no AET, the current benefit level is independent of earnings, *i.e.* $B_0(z) = b$. Therefore, individuals face a flat net "benefit reduction rate." That is, as earnings increase, the marginal reduction in post-tax and post-benefit resources is simply τ_0 . In state 1, the AET BRR is τ_b where the AET reduces benefits at the margin, *i.e.* for earnings above the exempt amount but below the point at which benefits have been phased out entirely. The presence of the AET introduces two changes in slope to the budget set, one at z_1^* and another at $z_2^*(b)$, due to reductions in current benefits:

[23] It would alternatively be possible to model the effects of OASI benefits and taxes using a single function, but we have modeled them separately to capture the reality of how the tax system and the AET operate separately. They are administered by separate agencies: the Internal Revenue Service and the Social Security Administration, respectively.
[24] For notational simplicity, we have made the benefit constant across individuals. In reality, each individual may receive a different level of pre-reduction benefits. The main issue this affects for our purposes is the earnings level at which the benefit is phased out entirely, which in reality can be different across individuals.
[25] We do not model taxes on Social Security benefits for simplicity; adding taxes on benefits would not change the qualitative predictions of the framework. Social Security benefits were untaxed until 1984, so benefits escaped taxation fully in most of our sample years (from 1978 to 1983). Starting in 1984, benefits were only taxed above an income threshold that was well above the AET exempt amount, implying that most benefits still escaped taxation.
[26] Introducing non-linear taxes for each individual would not qualitatively affect the predictions of this section as long as they are linear on average in the relevant range.

$$B_1(z) = \begin{cases} b \text{ for } z \leq z_1^* \\ b - \tau_b(z - z_1^*) \text{ for } z_1^* < z \leq z_2^*(b) \\ 0 \text{ for } z_2^*(b) < z \end{cases}$$

The first change in slope occurs at the point at which the AET is imposed, z_1^*, while the second change in slope occurs at the point where OASI benefits are phased out entirely, $z_2^*(b)$. At the higher amount, $z_2^*(b)$, the net benefit reduction rate returns to its lower level, creating a non-convex kink in the budget set. This second threshold is a function of OASI benefits, and is defined as follows:

$$z_2^*(b) = z_1^* + b/\tau_b.$$

This second threshold varies at the individual level, based on the size of one's OASI annual benefit.

Following previous literature, we assume individuals have a smooth distribution of "ability," which governs the tradeoff between leisure and consumption. In the presence of a linear tax, this should result in a smooth distribution of earnings conditional on working (*e.g.* Hausman, 1981; Saez, 2010; Kleven and Waseem, 2013). In the presence of the AET, a standard model predicts an intensive margin response with excess mass in earnings, or "bunching," to be present at the convex kink created at z_1^* (Gelber *et al.* 2013). At the extensive margin, to capture the realistic pattern of potential entry to or exit from non-trivial levels of earnings, we can assume a fixed cost of employment (Cogan, 1981; Eissa *et al.*, 2008). In this case, extensive margin decisions are a function of the average net-of-benefit-reduction rate (ANBRR), defined as

$$ANBRR \equiv 1 - \frac{\left[\left(T(z) - B(z)\right) - \left(T(0) - B(0)\right)\right]}{z}.$$ [27]

The ANBRR reflects the fraction of an individual's gross income that she keeps, net of both taxes and benefits, if she is employed at earnings level *z* rather than earning zero.

To demonstrate the impact of a kink on the decision to work in this context, we illustrate the extensive margin incentives created by the AET in Figure 2. Here we plot the ANBRR as a function of counterfactual earnings, that is, earnings conditional on working, in the counterfactual state where there is only a linear tax. We denote these potential earnings as $\tilde{z}_0$. We distinguish between these earnings and *realized* earnings, *z*, which incorporate the extensive margin decision and can be zero. The ANBRR measures the share of pre-tax income that is kept after taxes when working and earning *z*. In state 0, the ANBRR is constant at $1 - \tau_0$. This is represented by a dashed line. The solid line in Figure 2 shows that with the nonlinear budget set created by the AET, the ANBRR is $1 - \tau_0$ below z_1^*, but becomes $1 - \tau_0 - \tau_b(\tilde{z}_0 - z_1^*)/\tilde{z}_0$ above z_1^*, and therefore begins to decrease in $\tilde{z}_0$. However, after the benefit has been entirely phased out, the ANBRR becomes $1 - \tau_0 - b/\tilde{z}_0$ at $z_2^*(b)$, begins to increase in $\tilde{z}_0$, and eventually asymptotes back to $1 - \tau_0$ for large enough $\tilde{z}_0$.

[27] Gelber *et al.* (2018) give an extended discussion of extensive margin earnings decisions in the presence of a kinked budget set.

B. Validating our Measure of Social Security Benefits

A key control variable in our analysis is the monthly benefit amount. Because we are interested in employment effects at ages 63 and 64, our interest is in what the benefit amounts would be if a person claimed OASI benefits. Our goal, therefore, is to impute the benefits a person would earn if she claimed at age 62. As d escribed in the text, we impute benefits using a calculator provided by the Social Security Administration, applied to observed earnings histories, but with earnings at ages 55-61 set equal to their age 55 level.

This procedure creates two potential sources of measurement error in our imputation. First, of course, we do not use actual earnings at ages 55-61. Second, we do not observe exact earnings; instead there is a slight amount of random rounding in principle leads to errors in the benefit imputation. To assess the severity of this measurement error, we use data from the BEPUF, in which we observe actual monthly benefits as of 2004. For most people in the BEPUF, benefits observed are not the object of interest, because most people are not 62 year-old claimants in 2004. However, for people who were born in 1942 and claim OASI benefits as a retired worker, actual benefits in the BEPUF are exactly the object of interest.

We validate our measure by estimating the following regression

$$b_i = \beta_0 + \beta_1 b_i^* + e_i,$$

where b is our imputed benefit and b^* is the observed benefit, and We use data from the BEPUF and we limit the sample to people born in 1942 and claiming at age 62. The slope coefficient in this regression is the signal-to-noise ratio, and if $\hat{b}$ were measured perfectly, we would expect a constant of 0 and a slope of 1. The results are in Appendix Table 1. The first column shows results for everyone and the second for retired workers only (e.g., excluding disabled beneficiaries). In the first column the slope is 0.844, close to 1 but clearly below it, and the constant is $232. However, the sample in the first column includes people who claimed benefits on a spouse's record or as a disabled beneficiary, and our imputation is likely inaccurate for them. When we limit the sample to retired workers only in column (2), the slope coefficient rises to 0.95 and the constant falls to $92. Only a small amount of measurement error remains. We conclude that our imputation process measures age 62 benefits faithfully.

C. *Discussion of appendix figures*

Appendix Figure 1 shows that for those with earnings above z^* in year a, the probability of positive earnings falls sharply and substantially from outcome ages 62 to 63, exactly the age threshold we would expect if individuals respond to the AET by earning zero once they begin to claim OASI and are subject to the AET. By contrast, for those initially earning below z^* in year a, the probability of having positive earnings at the outcome age falls to a much smaller extent, both in percentage point terms (shown in the figure) and in percent terms, and much less sharply (relative to the pre-trend) than for those initially earning above z^* in age a.[28] This is consistent with the hypothesis that the AET reduces employment, as it has particular "bite" among those with relatively high earnings who are disproportionately subject to the AET.

Appendix Figure 1 shows that the trends in employment for those earning above and below z^* during outcome ages prior to 63 are very similar. Thus, we have reason to believe that anticipatory adjustment to the AET is not a significant issue in our context, as those who are likely to not face the AET have a similar trend in outcomes as those who are most likely to face the AET.

Appendix Figure 2 presents our main Figure 7 in a different way. Figure 7 shows treatment effects that are specific to bins of base age earnings. The treatment effect is the differential change in the probability of positive earnings, relative to the pre-period change and relative to the omitted bin. An alternative way to show this is to calculate, for each bin of base age distance to exempt amount and for each base age a, the probability of positive earnings at $a + 3$, and then to difference out this probability relative to some pre-period age. In Appendix Figure 2 we difference out the age 55 probabilities. The figure shows three important patterns. First, holding fixed earnings, as age gets higher, the probability of positive earnings falls. Second, at younger ages, the relationship between earnings and future employment is near zero, but, third, at higher ages, it is u-shaped: at first flat, then decreasing, then increasing. This is the basic pattern implied by our model.

We also briefly recapitulate the key findings of in Gelber *et al.* (2018), which further help to bolster the credibility of the results of the current paper. Paralleling the sharp change at the exempt amount in the slope of the ANBRR shown in Figure 2, Gelber *et al.* show theoretically that there should be a corresponding sharp change in slope of the employment rate as a function of age 60 earnings if there are frictions at the intensive margin that prevent individuals from adjusting to the AET by bunching at the exempt amount. This pattern does arise in the data: We show in Appendix Figure 3 (from Gelber *et al.*, 2018) that there is no visible change in the slope of the employment rate at the exempt amount at ages 61 and 62—prior to the ages when we should start to see an effect—but that we begin to observe a visible change in slope at ages 63 and 64. Using a Regression Kink Design (RKD), Gelber *et al.* (2018) show that there is no statistically significant change in slope at the exempt amount at ages 61 or 62, but that the change in slope becomes statistically significant at ages 63 and 64.

We also show in Appendix Figure 4 that predetermined covariates do not noticeably change in slope or level around the exempt amount, as the regressions in Gelber *et al.* (2018) confirm. This is consistent with the assumptions necessary for the validity of the empirical design.

[28] It is not surprising that employment falls, albeit relatively smoothly, from ages 62 to 63 even for those initially earning below z^*; employment rates gradually fall at older ages (e.g. Maestas 2010). Moreover, pension programs could have income effects on employment that reduce employment substantially (Fetter and Lockwood forthcoming).

D. Interpretation of the results in light of benefit enhancement

One important question is whether the response to the AET is influenced by its impact on future OASI benefits, given that benefits may be enhanced in the future after they are initially reduced due to the AET. Several considerations point against the view that future benefits play a significant role in mediating responses to the Earnings Test, as we discuss in detail in this Appendix.

First, responses to the AET do not appear to be larger for those who have relatively short lifespans, for whom the Earnings Test is particularly punitive. As described above, literature has established that individuals "bunch" at the Earnings Test exempt amount (Friedberg 1998; Friedberg 2000; Gelber, Jones, and Sacks forthcoming). If expected future benefits matter for responses to the Earnings Test, then we would expect these bunchers to be disproportionately composed of those who have short average lifespans, for whom the Earnings Test is particularly punitive. If so, expected lifespan should fall sharply as a function of earnings, in a radius within approximately $3,000 of the exempt amount where individuals tend to bunch (as shown in Gelber, Jones, and Sacks forthcoming). However, we do not see this pattern in the data. Appendix Figures 5 and 6 show that the probability of living past 70 (Appendix Figure 5) and average realized lifespan conditional on dying by the end of the sample (Appendix Figure 6), in a one percent sample of the SSA data on the U.S. population that was used in Gelber, Jones, and Sacks (forthcoming). Both graphs are essentially flat around the exempt amount, suggesting that future benefits do not play a role in mediating these responses to the AET.[29]

Second, the data show no response to the incentives created by future benefit enhancement. Over time benefit enhancement has become more generous but, as we describe in greater detail in Gelber, Jones, and Sacks (2019), for those over NRA there is no evidence of systematic bunching reaction to changes in the DRC and little relationship between bunching and life expectancy. For those under NRA, future benefits are enhanced substantially if current earnings exceed the exempt amount by even a single dollar. This creates an *upward* notch in lifetime income. Gelber, Jones and Sacks (2019) show however that there is disproportionate bunching *under* the exempt amount – the opposite of the pattern we would expect if people reacted to the full incentives created by benefit enhancement..

Third, in Table 3 of the current paper, we show heterogeneity in the estimated effects with respect to average lifetime income. Although income is not a perfect proxy for longevity, the two variables are significantly correlated (and as shown in literature from Preston 1975 to Chetty et al. 2016). Those with below-average income will on average have lower lifespan and, all else equal, should therefore react more to the Earnings Test because it is more punitive. In fact, we observe a smaller reaction among those with low average prior lifetime income.

Fourth, Table 3 also shows heterogeneity with respect to sex. Men have shorter average lifetimes than women, so we might expect to react more to the Earnings Test. In fact, women respond more than men.

Fifth, in Gelber, Jones, and Sacks (2019) we provide additional evidence that individuals mis-perceive the Earnings Test as more punitive than it is. Perhaps the reason for this apparent lack of reaction to variation in future benefits, is that the earnings test is widely mis-perceived as

[29] We lack data on lifespan in the EPUF, and therefore we are not able to analyze the relationship between lifespan and extensive margin responses, given our current data constraints. However, the lack of a relationship between lifespan and intensive margin responses is telling and strongly suggestive that responses to the Earnings Test are not mediated by future benefits.

a pure tax. Most popular guides do not note the subsequent adjustment in benefits under the earnings test (Gruber & Orszag (2003)). During the period that we study, the popular guide Your Income Tax (J.K. Lasser Institute (1997)), for example, warned readers that if “you are under age 70, Social Security benefits are reduced by earned income,” but did not note the subsequent benefit adjustment. Many individuals also may not understand the AET benefit enhancement or other aspects of OASI (Liebman and Luttmer 2015; Brown, Kapteyn, Mitchell, and Mattox 2013). Previous literature has found significant bunching responses to the AET (e.g. Friedberg 2000; Gelber, Jones, and Sacks 2013), implying that some individuals act as if the AET is punitive.

E. Replication in the restricted-access data and in the BEPUF

We began this project with access to data from the Master Earning's File. The sample was a 25% random sample of people born 1918-1923. We limited that sample to people without self-employment income. The main advantage of the restricted-access data is that it contains exact earnings information, not top coded or rounded, and that it separates self-employment and other income; otherwise it is similar to the public access data sets. An additional advantage of the restricted data is its enormous size, although estimates with the public use files appear reasonably precise anyway.

Before we lost access to the restricted-access data, we estimated simple difference-in-differences models without controls. We present those results in Appendix Table 2. The coefficient when $t = 3$ is -7.1 and when $t = 4$ it is -9.4. These estimates are not comparable to our main results because of differences in the cohorts (1918-1923 vs. 1931-1943) and differences in the specifications. To make the estimates comparable, we present in Appendix Table 3 estimates from the EPUF that look at the same set of cohorts and, in columns (1) and (4), use the same set of controls. The estimates are quite similar, -8.2 and -8.6 when $t = 3$ and $t = 4$.

These results show very similar estimates in the restricted-access and public use files. The results do raise a different question, namely, why does responsiveness seem so much higher for the older cohorts? The answer is that the estimates in the restricted access data do not control for the confounding effect of benefits. When we control for imputed benefits in columns (2) and (5), the estimates fall dramatically and are now much closer to the estimates for later cohorts (although still 25 percent larger). Further controlling for demographics in the public use files does not much change the estimates. Thus we believe that we would obtain similar results in the restricted-access data, were we able to estimate our main specifications in them.

We also replicate our main results in the BEPUF. We follow our sample selection and specification as closely as possible. However, the BEPUF only contains earnings information through 2003, so we limit the sample to people born between 1931 and 1940 (rather than 1931 to 1943 in the EPUF, which runs through 2006), so that the last cohort in the BEPUF sample reaches age 63 at the end of the data. Because the BEPUF is a random sample of Social Security claimants, its sampling frame is quite different from the EPUF's. To make it more comparable, we further limit the BEPUF sample to people with retired worker benefits (as of 2004, the date at which benefits are recorded); otherwise the BEPUF would oversample disabled beneficiaries, who likely have different patterns of labor force attachment.

We present estimates from the BEPUF in Appendix Table 4. This table is exactly analogous to our Table 2 of the main text, our main results. The results are also highly similar. In our preferred specification, we estimate a DID coefficient of -2.5 when $t = 3$ and -2.8 when $t = 4$. Although the BEPUF's sampling frame is different from the EPUF's, we find it reassuring that both sets of estimates are within sampling error of each other. We conclude that our results are not sensitive to the choice data set.

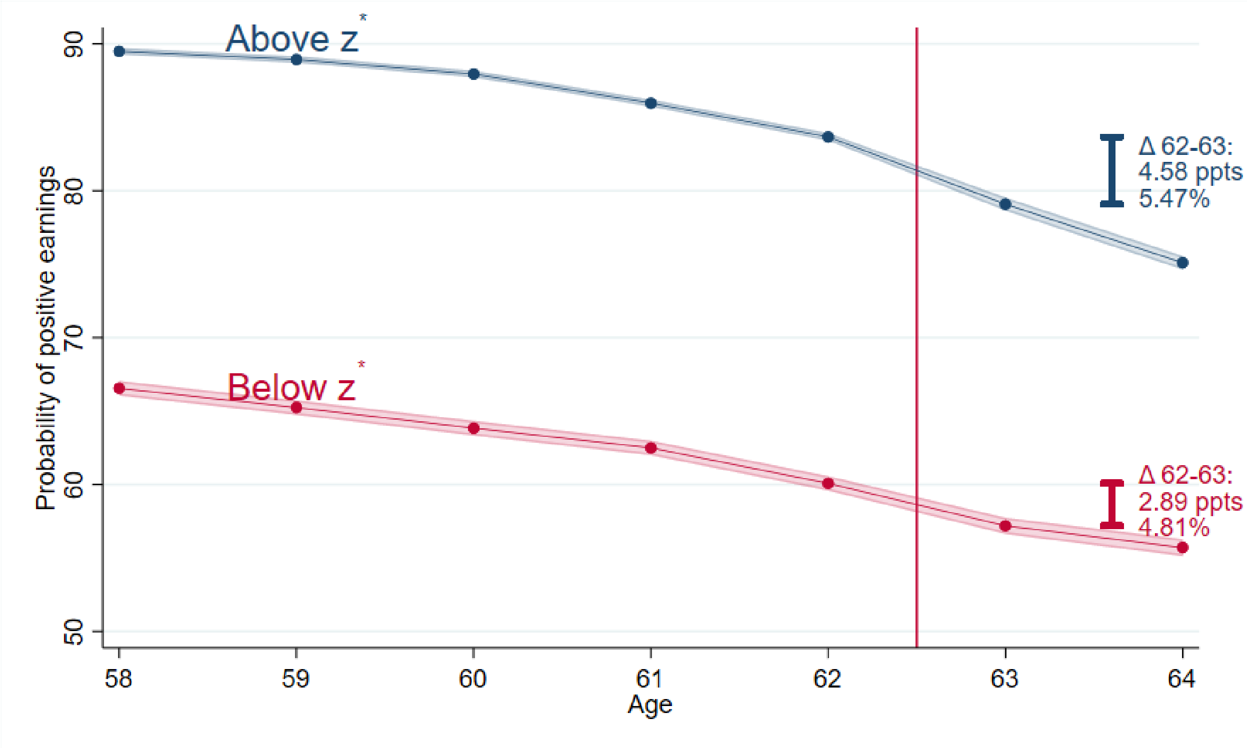
Appendix References not cited in the main text

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Association Between Income and Life Expectancy in the United States, 2001-2014." *JAMA*, 315(16): 1750-1766.

Maestas, Nichole. 2010. "Back to Work: Expectations and Realizations of Work After Retirement." *Journal of Human Resources*, 45(3): 718-748.

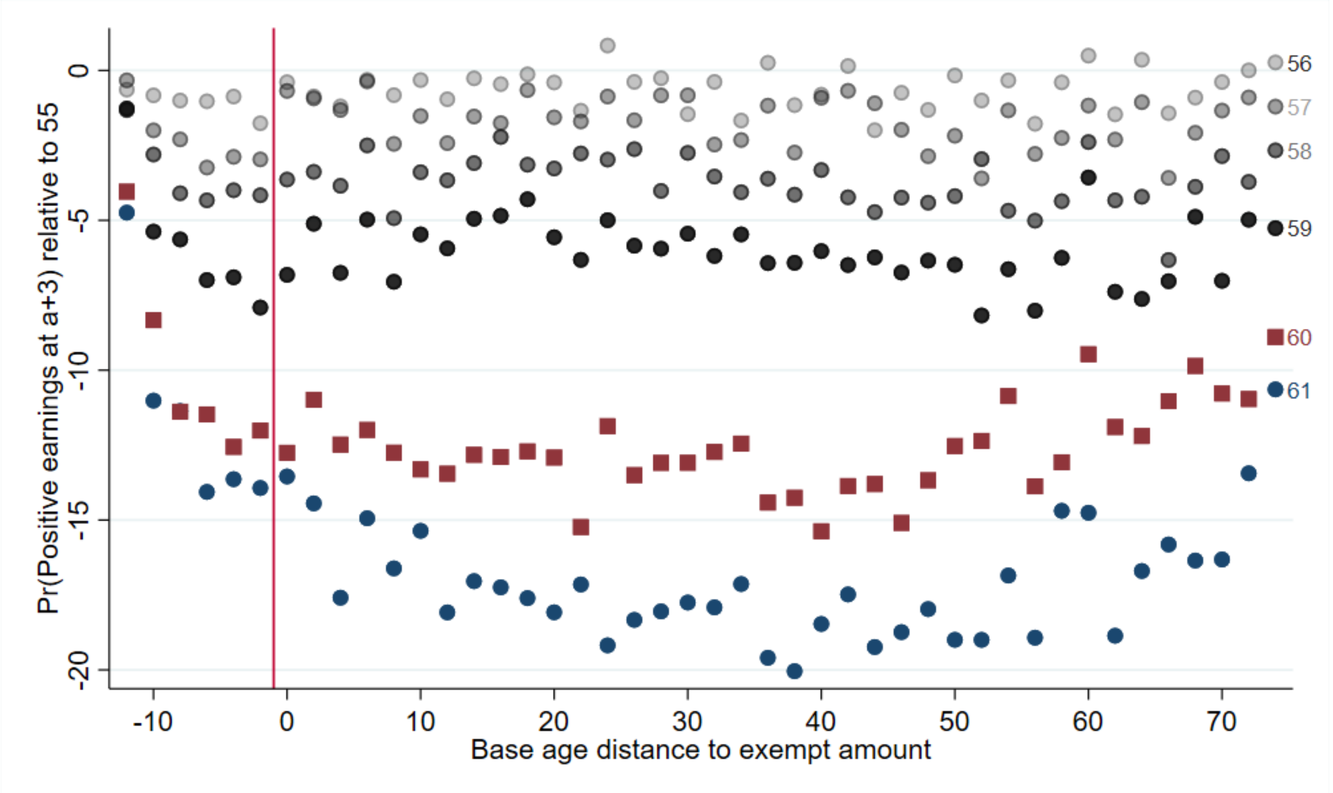
Preston, Samuel H. 1975. "The Changing Relation between Mortality and Level of Economic Development." *Population Studies*, 29(2): 231-248.

Appendix Figure 1. *Probability of positive earnings by age and earnings relative to exempt amount*



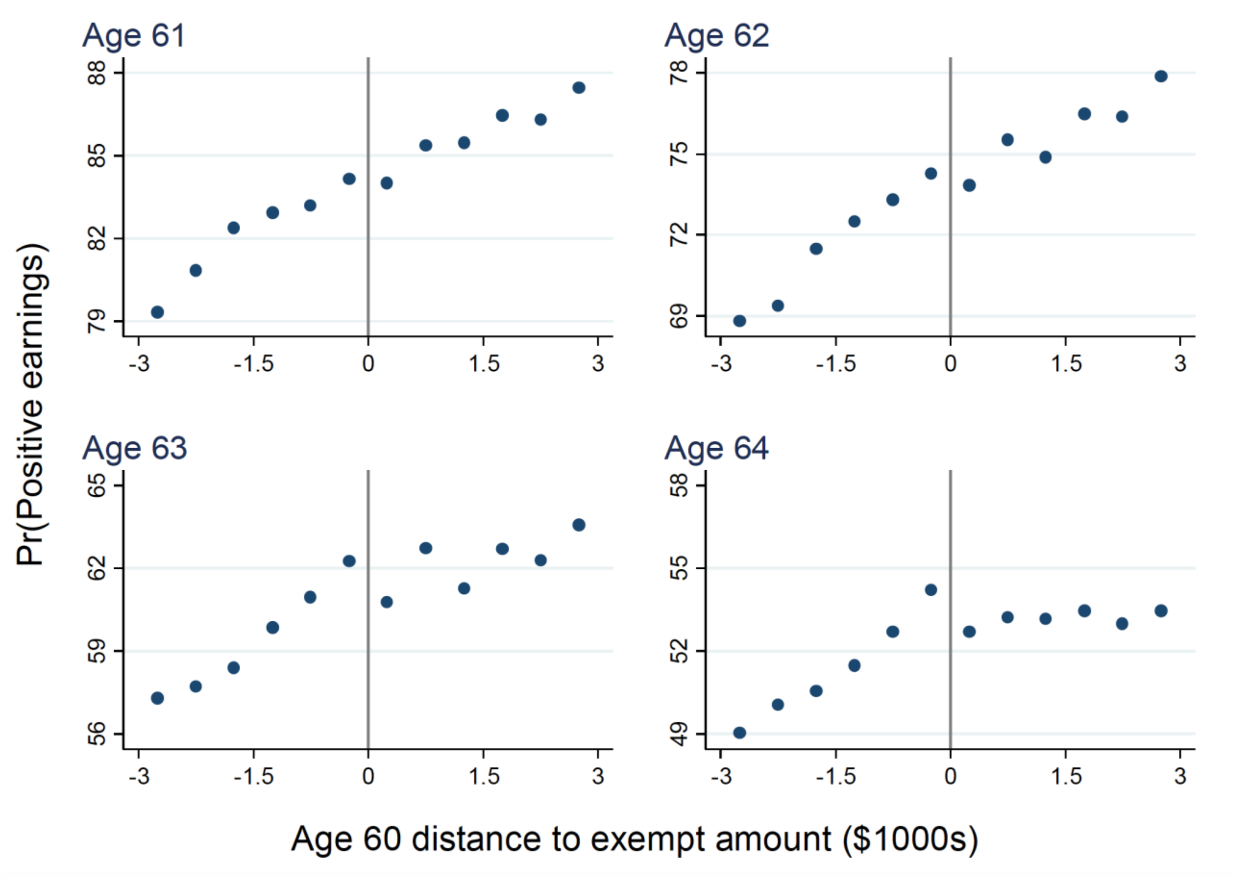
Notes: We show the employment probability by age, for the below-z* (above-z*) group that earned below (above) the exempt amount three years prior. To parallel our main specification, we plot means adjusting for our controls (female, age and cohort fixed effects, imputed benefits, and imputed benefits interacted with post). The shared area represents the 95% confidence region.

Appendix Figure 2 *Differential probability of positive earnings at $a + 3$, relative to age 55, as a function of base age and base age earnings*



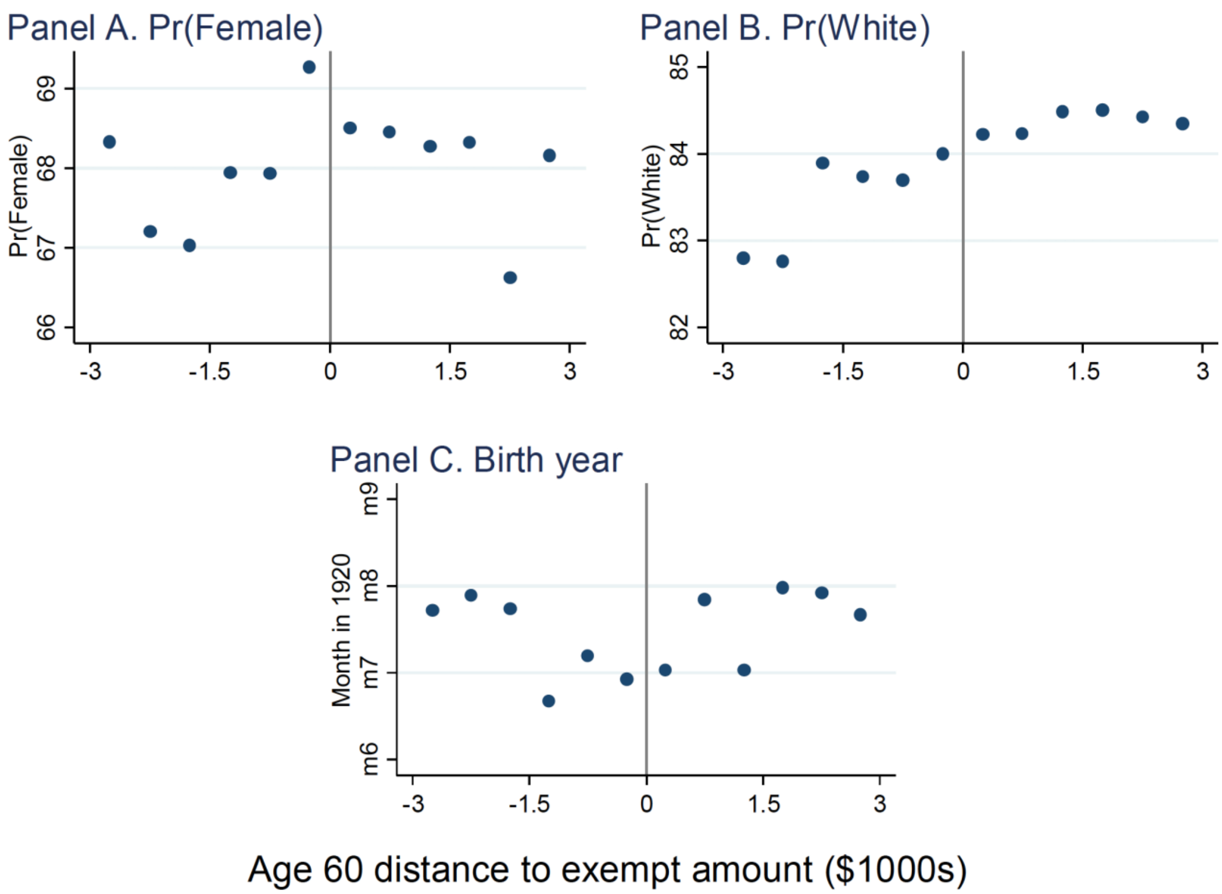
Notes: Figure plots, for each base age a and bin of earnings relative to the exempt amount, the difference in the fraction of observations with positive earnings at $a + 3$, relative to the fraction in base age 55. The sample is drawn from the EPUF and consists of people born 1931-1943, with positive earnings in the indicated base ages.

Appendix Figure 3. *Probability of Positive Earnings by Single Year of Age, Ages 61 to 64*



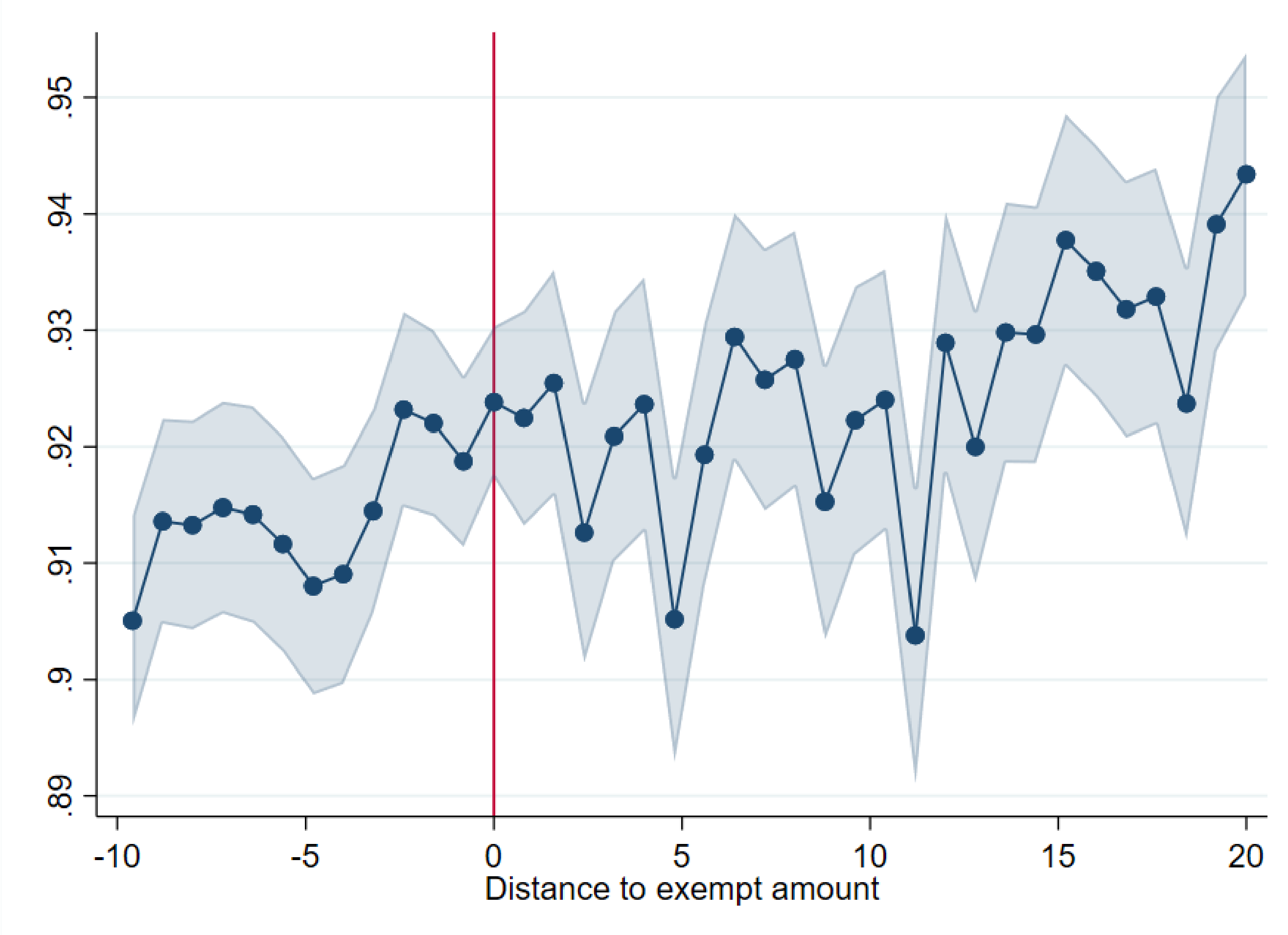
Notes: the source of the figure is Gelber *et al.* (2018). Each figure plots the mean annual employment rate, *i.e.* the probability of positive earnings, for each single year of age from 61 to 64, as a function of the distance to the exempt amount, which has been normalized to zero. The sample is individuals with positive age 60 earnings and no age 60 self-employment income, born 1918 to 1923.

Appendix Figure 4. *Predetermined covariates around the exempt amount*



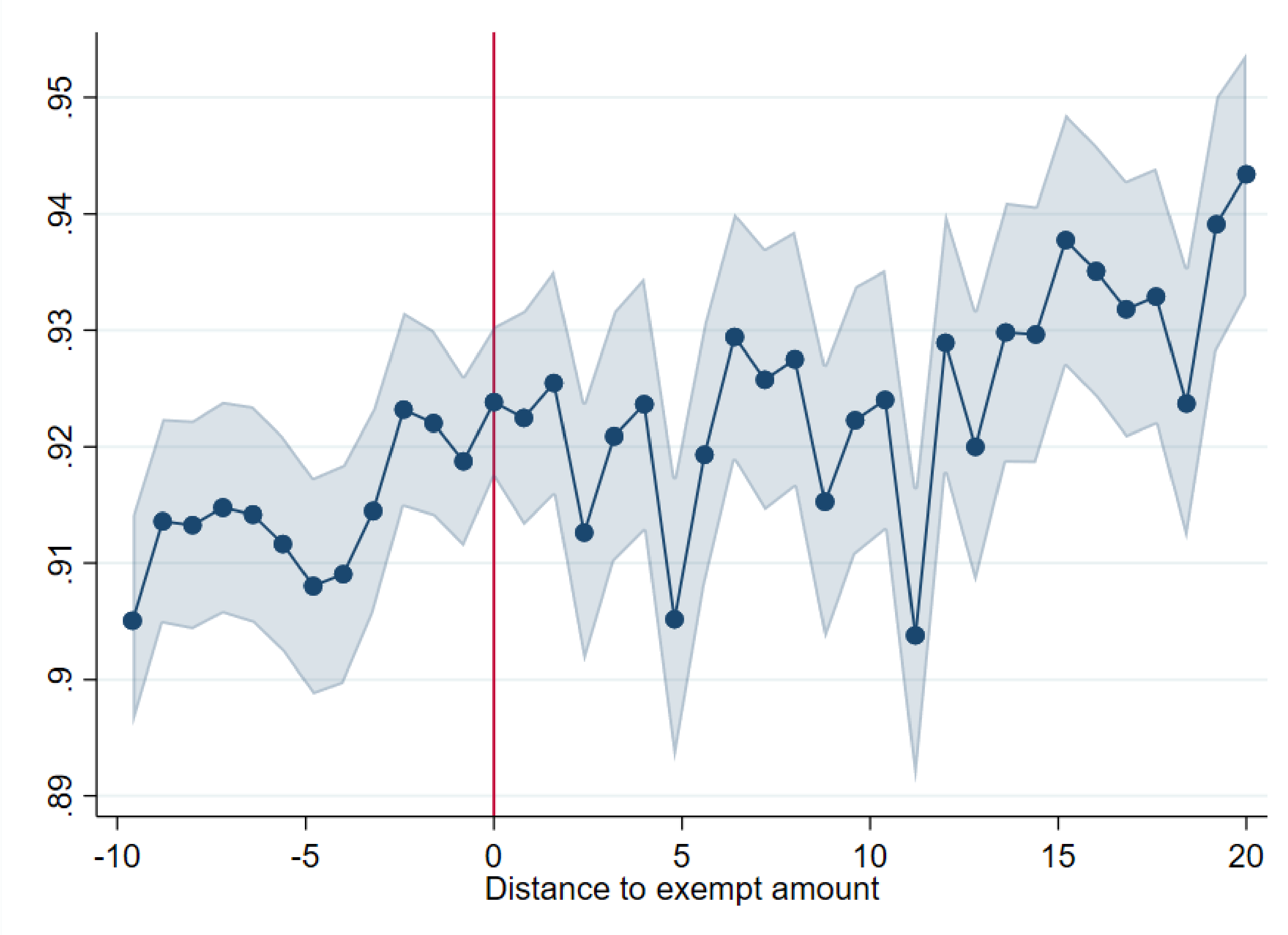
Notes: the source of the figure is Gelber *et al.* (2018). The figure shows the bin means of predetermined covariates as a function of the distance to the age 60 exempt amount. The figure demonstrates that there are no clear visual changes in slope in any of these covariates at the age 60 exempt amount, consistent with the assumptions necessary for the validity of the regression kink design employed in Gelber *et al.* (2018).

Appendix Figure 5. *Probability of living to age 70 as a function of age 62-64 earnings*



Notes: The sample consists of people aged 62-64 in 1990-1999 who claimed by age 65. The x-axis is earnings relative to the exempt amount. The y-axis shows the fraction of people living to age 70 or greater in each $800 bin. The shaded area is the 95% confidence interval. Source: SSA data.

Appendix Figure 6. *Age at death as a function of age 62-64 earnings*



Notes: The sample consists of people aged 62-64 in 1990-1999 who claimed by age 65. The x-axis is earnings relative to the exempt amount. The y-axis shows the average age at death (conditional on dying) in each $800 bin. The shaded area is the 95% confidence interval. Source: SSA data.

Appendix Table 1. *Validating benefit imputation*

Sample	All	Retired workers only
	(1)	(2)
Slope	0.844	0.953
	(0.005)	(0.004)
Constant	231.888	91.679
	(6.215)	(4.848)
R^2	0.822	0.934
# Observations	8,504	7,372

Notes: Table reports the estimates from a bivariate regression of monthly benefit amounts as reported in the BEPUF against imputed benefit amounts. Sample is limited to people born in 1942 and claiming benefits at age 62 because for this sample we observe actual benefits at age 62 in the BEPUF (i.e. the variable we impute). Robust standard errors in parentheses.

Appendix Table 2. *Results from restricted-access administrative data and specifications that do not control for benefits*

	$t = 3$ years ahead	$t = 4$ years ahead
	(1)	(2)
DID coefficient	-7.1	-9.4
	(0.05)	(0.05)
# Observations	48,580,452	48,580,452
# People	8,296,628	8,296,628

Notes: Table shows the DID coefficient. The outcome is a dummy for positive earnings (x100) in the indicated number of years ahead. The sample is a 25% random sample of people born 1918-1923 from the Master Earnings File. The sample is limited to observations with positive earnings and base age 55 to 61. Additional controls include *post* and *treat*. Standard errors (in parentheses) are clustered by individual.

Appendix Table 3. *Limiting the EPUF sample to earlier cohorts to match SSA data*

	t = 3 years ahead			t = 4 years ahead		
	(1)	(2)	(3)	(4)	(5)	(6)
DID Coefficient	-8.2	-3.8	-3.8	-8.6	-4.1	-4.1
	(0.4)	(0.47)	(0.47)	(0.4)	(0.5)	(0.5)
# Observations	542,581	542,581	542,581	475,796	475,796	475,796
# People	97,969	97,969	97,969	96,999	96,999	96,999
Controls						
Benefits		X	X		X	X
Sex			X			X
Age FE			X			X
Cohort FE			X			X

Notes: Table shows the DID coefficient. The outcome is a dummy for positive earnings (x100) in the indicated number of years ahead. The sample is drawn from the EPUF. It consists of people born 1918-1923 with positive base age earnings, and base age 55 to $64 - t$. All specifications control for main effects of *post* and *treat*. The benefit controls are imputed benefits plus their interaction with *post*. Standard errors (in parentheses) are clustered by individual.

Appendix Table 4. *Replicating the main estimates in the BEPUF*

	t = 3 years ahead			t = 4 years ahead		
	(1)	(2)	(3)	(4)	(5)	(6)
DID Coefficient	-3.5	-2.4	-2.5	-3.9	-2.7	-2.8
	(0.3)	(0.3)	(0.4)	(0.4)	(0.4)	(0.4)
Elasticity	0.24	0.19	0.19	0.27	0.22	0.22
# Observations	772.417	772.417	772.417	672,596	672,596	672,596
# People	133,240	133,240	133,240	132,093	132,093	132,093
Controls						
Benefits		X	X		X	X
Sex			X			X
Age FE			X			X
Cohort FE			X			X

Notes: Table shows the DID coefficient. The outcome is a dummy for positive earnings (x100) in the indicated number of years ahead. The sample is drawn from the BEPUF. It consists of people born 1931-1940 with positive base age earnings, and base age 55 to $64 - t$. All specifications control for main effects of *post* and *treat*. The benefit controls are imputed benefits plus their interaction with *post*. The elasticity assumes a 40% OASI claiming rate. Standard errors (in parentheses) are clustered by individual.