

Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test

Alexander M. Gelber, Damon Jones, and Daniel W. Sacks*

August 2014

Abstract

We study frictions in adjusting earnings in response to changes in the Social Security Annual Earnings Test (AET), using a panel of Social Security Administration microdata on one percent of the U.S. population from 1961 to 2006. Individuals continue to "bunch" at the convex kink the AET creates even when they are no longer subject to the AET, demonstrating that adjustment frictions help drive behavior in a new and important context. We develop a novel framework for estimating an earnings elasticity and an adjustment cost using information on the amount of bunching at kinks before and after policy changes in earnings incentives around the kinks. We apply this method in settings in which individuals face changes in the AET benefit reduction rate, and we estimate in a baseline case that the earnings elasticity with respect to the implicit net-of-tax share is 0.23, and the fixed cost of adjustment is \$152.08. Our results demonstrate that the short-run impact of changes in the effective marginal tax rate can be substantially attenuated.

*Gelber: UC Berkeley Goldman School of Public Policy and NBER, agelber@berkeley.edu; Jones: University of Chicago, Harris School of Public Policy and NBER, damonjones@uchicago.edu; Sacks: Indiana University, Kelley School of Business, dansacks@indiana.edu. We thank Raj Chetty, Jim Cole, Jim Davis, Mark Duggan, Jonathan Fisher, Richard Freeman, John Friedman, Bill Gale, Hilary Hoynes, Adam Isen, Henrik Kleven, Olivia Mitchell, Emmanuel Saez, and numerous seminar participants. We are extremely grateful to David Pattison for generously running the code on the data. We acknowledge financial support from the Wharton Center for Human Resources and the Wharton Risk and Decision Processes Center, from NIH grant #1R03 AG043039-01, from a National Science Foundation Graduate Research Fellowship, and from support from the U.S. Social Security Administration through grant #5RRC08098400-05-00 to the National Bureau of Economic Research (NBER) as part of the SSA Retirement Research Consortium. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, or the NBER. The research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants: SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. All results have been reviewed to ensure that no confidential information is disclosed. All errors are our own.

1 Introduction

In a traditional model of workers' earnings or labor supply choices, individuals optimize their behavior frictionlessly. However, several recent papers have suggested that individuals face frictions in adjusting behavior to policy (Chetty, Looney, and Kroft 2009; Chetty, Friedman, Olsen, and Pistaferri 2011; Chetty, Guren, Manoli, and Weber 2012; Chetty, Friedman, and Saez 2012; Chetty 2012; Kleven and Waseem 2013). Adjustment frictions could reflect a variety of factors. For example, frictions could reflect lack of knowledge of a tax regime, for example if individuals only make the effort to understand their earnings incentives when the utility gains from doing so are sufficiently large (*e.g.* Simon 1955; Chetty *et al.* 2007; Hoopes, Reck, and Slemrod 2013), or they could reflect the cost of negotiating a new contract with an employer or the time and financial cost of job search. Costs of adjusting behavior help to govern the welfare cost of taxation (Chetty *et al.* 2009), and they also help to explain heterogeneity across contexts in the observed elasticity of earnings with respect to the net-of-tax rate (Chetty *et al.* 2011, 2012b; Chetty 2012).¹ Frictions in adjusting earnings could help to explain patterns in the data, such as the slow rise in retirement at age 62 subsequent to the introduction of the Social Security Early Retirement Age (Gruber 2013), or the lack of bunching at many kink points in budget sets (Chetty, Friedman, Olsen, and Pistaferri 2011).

This paper develops evidence on the existence and size of frictions in adjusting earnings in response to policy. The U.S. Social Security Annual Earnings Test (AET) represents a promising environment for studying these questions. This setting provides a useful illustration of many issues—such as the development and application of a methodology for estimating elasticities and adjustment costs simultaneously—that are applicable to studying earnings responses to policy more broadly. The AET reduces Social Security Old Age and Survivors Insurance (OASI) benefits in a given year as a proportion of an OASI claimant's earnings above an exempt amount in that year. For example, for OASI claimants aged 62-65 in 2013, current OASI benefits are reduced by 50 cents for every extra dollar earned above \$15,120. The AET may lead to very large effective benefit reduction rates (BRRs)

¹The net-of-tax rate is defined as one minus the marginal tax rate (MTR). Literature including Altonji and Paxson (1988) examines hours constraints in the context of labor supply.

on earnings above the exempt amount, creating a strong incentive for many individuals to “bunch” at the convex kink in the budget constraint located at the exempt amount (Burtless and Moffitt 1985; Friedberg, 1998, 2000; Song and Manchester 2007; Engelhardt and Kumar 2014). Reductions in current benefits due to the AET sometimes lead to increases in later benefits; nonetheless, as we discuss in detail in Section 2, several factors may explain why individuals’ earnings still respond to the AET.

The AET is an appealing context for studying earnings adjustment for at least three reasons. First, bunching at the AET kink is easily visible on a graph, allowing credible documentation of behavioral responses.² Second, the AET represents one of the few known kinks at which bunching occurs; indeed, our paper represents the first study to find robust evidence of sharp bunching at the intensive margin among the non-self-employed at any kink in the U.S.³ Third, the AET is important to policy-makers in its own right, as it is a significant factor that affects the earnings of the elderly in the U.S.

We make two main contributions to understanding adjustment frictions. First, we provide new evidence on earnings adjustment frictions and document that such frictions exist in the U.S., by showing that in some contexts individuals do not adjust immediately to changes in AET. We focus particularly on cases in which a kink in the effective tax schedule disappears. In the absence of adjustment frictions, the removal of a convex kink in the effective tax schedule should immediately lead to a complete lack of bunching at the former kink; thus, any observed delay in reaching zero bunching should reflect adjustment frictions.⁴ We observe clear evidence of delays in some contexts: individuals continue to bunch around the location of a former kink, consistent with the existence of adjustment frictions. Nonetheless, the vast majority of individuals’ adjustment occurs within at most three years. Thus, we interpret the frictions we observe as evidence of barriers to making an *immediate* adjustment in response to changes in incentives.

Second, we specify a model of earnings adjustment that allows us to estimate a fixed

²Other papers have examined bunching in the earnings schedule, including Blundell and Hoynes (2004) and Saez (2010). Saez (2010) shows that the amount of bunching can be related to the elasticity of earnings with respect to the net-of-tax rate.

³Chetty *et al.* (2012) find evidence of diffuse earnings responses to the Earned Income Tax Credit among the non-self-employed.

⁴For consistency with the previous literature on kink points that has focused on the effect of taxation, we sometimes use “tax” as shorthand for “tax-and-transfer,” while recognizing that the AET reduces Social Security benefits and is not administered through the tax system. The “effective” marginal tax rate is potentially affected by the AET BRR, among other factors.

adjustment cost and the elasticity of earnings with respect to the effective net-of-tax rate. Recent work demonstrating the importance adjustment costs has raised the question of how to estimate both the elasticity and adjustment cost simultaneously. We develop tractable methods that allow estimation of elasticities and adjustment costs with kinked budget sets. Our method complements Kleven and Waseem (2013), who innovate a method to estimate elasticities and the share of the population that is inert in the presence of a notch in the budget set (but does not estimate adjustment costs). To our knowledge, our method is the first to allow estimation of both elasticities and adjustment costs. As discussed in Chetty *et al.* (2009), the elasticity and adjustment cost are both necessary for welfare calculations in many applications, and our estimates of these parameters allow us to perform counterfactual simulations of the effects of alternative policies that have not been performed using other recent methods for addressing adjustment frictions. Our method is also applicable in a different context than Kleven and Waseem’s: ours relies on bunching at kinks, rather than notches, to perform the estimates.

Our method relies on clear patterns in the data. In our model, all else equal the amount of bunching at a newly introduced kink increases with the elasticity (as a higher elasticity will induce more individuals to locate at the kink) but decreases with the adjustment cost (as the adjustment cost prevents bunching among some individuals). This prevents estimation of both parameters using a single cross-section—since a small amount of bunching, for example, could be consistent with either a low elasticity or a high adjustment cost—but with two or more cross-sections of individuals facing different tax rates in the region of the kink, we can specify two or more equations and find the values of two variables (the elasticity and the adjustment cost). In a frictionless model (Saez 2010), bunching at a convex kink is approximately proportional to the jump in the net-of-tax rate at the kink; thus, when a kink becomes (for example) less sharply bent in the frictionless model (i.e. the jump in the net-of-tax rate at the kink falls), the degree of bunching at the kink falls proportionately. In our model, adjustment costs help to explain deviations from this pattern. As we move from the more sharply bent kink to the less sharply bent kink in our model with adjustment costs, bunching falls by a less-than-proportional amount—consistent with our empirical observation that individuals continue to bunch at the location of a former kink. Moreover, the absolute

value of the change in bunching is decreasing in the adjustment cost. Intuitively, these patterns help us to identify the adjustment cost, as well as the elasticity.

We apply our method to data spanning the decrease in the AET benefit reduction rate from 50 percent to 33.33 percent in 1990 for those aged 66 to 69, as well as a setting in which the AET ceases to apply, when moving from age 69 to ages 70 and older in the 1990-1999 period. In a baseline specification examining the 1990 change, we estimate that the fixed adjustment cost is \$152.08 (in 2010 dollars)—if the gains exceed this level, then the individual adjusts earnings—and that the earnings elasticity with respect to the net-of-tax share is 0.23. This specification examines data on individuals in 1989 and 1990; thus, our estimated adjustment cost represents the cost of adjusting earnings in the first year after the policy change. Other empirical strategies show results in the same range. Just following the reduction in the kink in 1990, if we constrain the adjustment cost to be zero (and thus fail to account for excess bunching following the policy change due to inertia), we estimate a statistically significantly higher earnings elasticity of 0.39 in the baseline specification (69 percent larger than the unconstrained estimate).

Our estimates suggest that while adjustment costs are modest in our setting, they have the potential to change elasticity estimates significantly, thus illustrating that it can be important to incorporate adjustment costs when estimating elasticities. While the methodology we develop applies more generally, our numerical estimates are specific to our setting, and adjustment costs and elasticities may be substantially different (larger or smaller) in other contexts. Our estimates apply to the population bunching at kinks; we believe it is particularly striking that we find evidence of adjustment frictions even among those initially bunching at the kink, whose initial bunching indicates flexibility (enough to locate at the kink initially). By demonstrating that even in this setting, earnings adjustment frictions exist and substantially change elasticity estimates, our results suggest the importance of taking them into account in other settings. As a key example of the application of our estimates, our simulations show that adjustment costs can substantially attenuate short-run earnings reactions to changes in the effective marginal tax rate, frustrating the goal of affecting short-run earnings as envisioned in many recent discussions of tax policy. While the adjustment cost we estimate is modest, and the constrained and unconstrained elasticities are not very

different in absolute value, the adjustment cost does imply that even large changes in policy can have little or no effect within a year or more of the implementation of a new policy, thus implying that it makes an important difference to include adjustment costs in the analysis.

While the primary focus of the paper relates to drawing lessons about earnings frictions and adjustment, as opposed to answering questions about the AET specifically, a secondary contribution of the paper is to provide new evidence on the effects of the AET in particular. First, we use SSA administrative data with a sample of 376,431 observations in our main period, building on certain previous studies on the AET that use survey data. Second, our study is the first to estimate bunching in the context of the AET through a method similar to Saez (2010). Third, we investigate a variety of major AET policy changes over a period spanning decades. Fourth, we investigate whether mortality expectations help drive individuals' earnings responses to the AET by estimating the pattern of life expectancy around the exempt amount. Finally, our simulations show that a large reduction in the AET BRR could lead to a large rise in earnings.

The remainder of the paper is structured as follows. Section 2 describes the policies we examine. Section 3 presents a framework for analyzing the behavioral response to these policies and describes our empirical strategy for quantifying bunching. Section 4 describes our data. Section 5 presents empirical evidence on the earnings response to changes in the AET. Section 6 specifies a tractable model of earnings adjustment. Section 7 estimates the fixed adjustment cost and elasticity simultaneously. Section 8 concludes with discussion and avenues for future work.

2 Policy Environment

Figure 1 shows key features of the AET rules from 1961 to 2009. The AET became less stringent over this period. The dashed line and right vertical axis show the benefit reduction rate. From 1961 to 1989, every dollar of earnings above the exempt amount reduced OASI benefits by 50 cents (until OASI benefits reached zero). In 1990 and after, the benefit reduction rate fell to 33.33 percent for beneficiaries older than the Normal Retirement Age (NRA). The NRA, the age at which workers can claim their full OASI benefits, is 65 for those born 1937 and before that we focus on. During the period 1983-1999 that we focus

on, the AET applied to individuals aged 62-69. The solid line and left vertical axis show the real exempt amount. Starting in 1978, beneficiaries younger than NRA faced a lower exempt amount than those at least NRA but younger than the maximum age subject to the AET.

We discuss below how, in some cases, reductions in current benefits by the AET are offset by future benefit increases. We nonetheless model the AET as creating a positive implicit marginal tax rate for some individuals — reflecting the reduction in current benefits — consistent with the empirical finding that some individuals bunch at AET kinks, certain theoretical considerations we describe below, and previous literature.

When current OASI benefits are lost to the AET, future scheduled benefits are increased in some circumstances. This is sometimes called "benefit enhancement." Benefit enhancement can reduce the effective tax rate associated with the AET, in particular for those individuals considering earning enough to trigger the enhancement in the post-1972 period, as we describe in detail in Appendix A and briefly in this section.

Prior to 1972, the AET caused a pure loss in benefits for those NRA and older, as there was no benefit enhancement for these individuals. For beneficiaries subject to the AET aged NRA and older, a one percent DRC was introduced in 1972, meaning that each year of benefits foregone led to a one percent increase in future yearly benefits. The DRC was raised to three percent in 1982 and gradually rose to eight percent for cohorts reaching NRA from 1990 to 2008 (though the AET was eliminated in 2000 for those older than the NRA). An increase in future benefits between seven and eight percent is approximately actuarially fair on average, meaning that an individual with no liquidity constraints and average life expectancy should be indifferent between either claiming benefits now or delaying claiming and receiving higher benefits once she begins to collect OASI (as Diamond and Gruber 1999 show with respect to the actuarial adjustment for early claiming).

As we describe further in Appendix A, OASI claimants' future benefits are only raised due to the DRC when annual earnings are sufficiently high that the individual loses an entire month's worth of OASI benefits due to the reductions associated with the AET (Friedberg 1998; Social Security Administration 2012a). In particular, an entire month's benefits are lost once the individual earns $z^* + (MB/\tau)$ or higher, where z^* is the annual exempt amount,

MB is the monthly benefit, and τ is the AET benefit reduction rate. Note that as we describe in detail in the Appendix, although the AET withholds benefits at the monthly level, the AET is generally applied based on *annual* earnings—the object we observe in our data. With a typical monthly benefit of \$1,000 and a benefit reduction rate of 33.33 percent, one month’s benefit enhancement occurs when the individual’s annual earnings are \$3,000 ($=\$1000/0.3333$) above the exempt amount. For example, if an individual born in 1933-1934 earned at or just this amount (i.e. $z^* + \$3,000$), she was subject to the DRC, and future benefits were raised by 0.46 percent (but no increase occurs if the individual earns below this amount). As a result, at or just above the AET threshold, earning an extra dollar does not affect subsequent OASI benefits. Thus, benefit enhancement is only relevant to an individual considering earning substantially in excess of the exempt amount. Later, our empirical specification alternatively assumes that benefit enhancement does not (or does) affect the AET implicit marginal tax rate, and we find similar patterns in both specifications.

Thus, the AET could affect the earnings decisions of those NRA and older for a number of reasons. As we have discussed, for those to whom benefit enhancement is effectively irrelevant (because they are only considering earning sufficiently near to the AET that they would not receive benefit enhancement through increasing earnings), the marginal incentives they face are not affected by benefit enhancement. For those to whom benefit enhancement is relevant (because they are considering earning in a region well above the AET exempt amount, thus triggering benefit enhancement), the AET could also affect decisions, for several reasons. The AET was on average roughly actuarially fair only beginning in the late 1990s. Furthermore, those whose expected lifespan is shorter than average should expect to collect OASI benefits for less long than average, implying that the AET is more financially punitive (though we ultimately find no evidence consistent with this hypothesis). Liquidity-constrained individuals or those who discount faster than average could also reduce work in response to the AET. Finally, many individuals also may not understand the AET benefit enhancement or other aspects of OASI (Liebman and Luttmer 2011; Brown, Kapteyn, Mitchell, and Mattox 2013).

For beneficiaries under NRA, the actuarial adjustment raises future benefits whenever an

individual earns any amount over the AET exempt amount (Social Security Administration 2012, Section 728.2; Gruber and Orszag 2003). 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 creates an upward notch in future benefits as well as a kink in current benefits at the AET threshold—as opposed to a simple kink, whose properties we explore in our theory sections. However, as we show in Figure 2 below, the histograms show no evidence that the earnings density increases from just below to just above the exempt amount (as one would predict if they respond to the incentives created by this notch that raises future benefits when an individual earns above the exempt amount). Our discussion of the effects of kinks therefore does not directly apply to pre-NRA ages. Thus, in our estimates of elasticities and adjustment costs, we limit the sample to ages NRA and older, for which the budget set (in the region of the exempt amount) is a kink rather than a notch.

3 Initial Bunching Framework

As a preliminary step, we begin with a model with no frictions. This model is well-known and described in detail elsewhere, but we briefly describe it here and in more detail in Appendix E.⁵ After we have presented our empirical results, we specify a model with frictions that is consistent with the descriptive patterns we document.

Appendix Figure F.1 shows the budget constraint and incentives created by the AET for those NRA and older in the frictionless case. Start with a linear tax (Panel A) at a rate of τ_0 . Now, suppose the AET is introduced (on top of pre-existing taxes), so that the marginal net-of-tax rate decreases to $1 - \tau_1$ (where $\tau_1 > \tau_0$) for earnings above a threshold z^* (Panel B). Individuals earning in the neighborhood above z^* reduce their earnings. If ability is smoothly distributed, a range of individuals initially locating between z^* and $z^* + \Delta z_1^*$ (as depicted in the density in Panel C) will "bunch" exactly at z^* , due to the discontinuous jump in the marginal net-of-tax rate at z^* . In fact, we find empirically that these individuals locate in the neighborhood of z^* , as shown in Panel D.

To quantify the amount of bunching, i.e. excess mass, we use a technique similar to Chetty *et al.* (2011) and Kleven and Waseem (2013), which we illustrate in Appendix Figure F.2

⁵Saez (2010) describes this model in greater detail. This work follows earlier work on estimation of labor supply responses on nonlinear budget sets, including Burtless and Hausman (1978) and Hausman (1981). Moffitt (1990) surveys these methods.

and describe further in Appendix B. The x-axis measures before-tax income, z , while the y-axis measures the density of earnings. In Panel A, we show that the ex-post density of earnings in the presence of a kink is comprised of a number of groups. Those in the region labeled X in the figure ("bunchers") have optimal earnings above z^* under the lower rate of τ_0 and at z^* under the higher rate of τ_1 . Those in the region labeled Y in the figure consist of individuals whose optimal earnings are below z^* under a lower marginal tax rate of τ_0 , as well as other individuals whose optimal earnings are above z^* under the higher marginal tax rate of τ_1 . Panel B shows that to estimate the size of region X, we must estimate the *ex post* density and subtract the mass associated with Group Y.

As described further in Appendix B, we divide the earnings distribution into bins of width $\delta = \$800$ and estimate a seven-degree polynomial through the densities associated with the bins, controlling for dummies for being in the seven bins nearest to the kink. Our estimate of bunching, B , is the sum of the coefficients on these dummies. This implies that our estimate of bunching is driven by individuals locating within \$2800 of the kink (as the central bin runs from \$400 under the kink to \$400 above the kink). We therefore measure bunching as the excess mass in the \$5600-wide region around the kink, relative to the prediction from a smooth polynomial counterfactual distribution estimated away from the kink. We estimate confidence intervals through a bootstrap procedure that we describe further in Appendix E.10 (and the results are similar under the delta method). We report our bunching amount, B , normalized by the share of individuals in the neighborhood $[z^* - \delta/2, z^* + \delta/2]$ who belong to Group Y (which we approximate as the area under our polynomial over this range); we refer to this as "normalized excess mass." While we show this bunching at the kink as arising in a frictionless model here, this technique is also suited to measuring the bunching at the kink arising in a model with frictions.

Some apparent limitations of our approach are worth discussion. First, following previous literature on earnings responses to kinks, we do not take into account other choices that could affect earnings in the long run, such as human capital accumulation. However, human capital accumulation is likely to be less important for the older workers we study than it is for the population as a whole. Second, other programs—such as Medicaid, Supplemental Security Income, Disability Insurance, or taxes such as unemployment insurance payroll taxes—create

earnings incentives near the bottom of the earnings distribution. The kinks created by such programs—which in principle pose an issue for most of the literature on bunching at kinks—in practice are typically inapplicable or safely far away from the AET convex kink. Third, we follow previous work and largely do not distinguish among the potential reasons for a response to the AET. Following previous literature, our bunching framework presumes that certain individuals treat the AET as creating *some* effective marginal tax rate above the exempt amount, consistent with the empirical evidence documenting clear responses to the incentives created by the AET.

Fourth, in this paper, we focus on the marginal incentives created by the AET and intensive margin responses, following previous literature based on the technique of Saez (2010). Other important decisions could include the choice of whether to earn a positive amount, or the decision to claim OASI (Gruber and Orszag 2003). We abstract from the claiming decision by examining a sample of OASI claimants, following previous literature such as Friedberg (1998, 2000); however, it is worth noting that if the AET affects the claiming decision, there is no *a priori* reason that this change in claiming should increase or decrease the magnitude of the bunching responses we document among claimants. Moreover, we add to previous literature by showing in Appendix Figure F.15 that the hazard of claiming at year $t + 1$ is smooth around the exempt amount at year t , indicating no evidence that claimants come disproportionately from close to or far from the kink. We discuss the claiming decision further in the Appendix. Examining the extensive margin response goes beyond previous papers estimating intensive margin responses to incentives using bunching techniques like Saez (2010) and is a valuable topic for future research, which we have explored in Gelber, Jones, and Sacks (2014).

Finally, the results are specific to the AET and may not generalize outside of this context. We estimate the speed of adjustment among those initially bunching at a kink. We believe it is all the more interesting that we still find evidence of modest adjustment frictions among this group whose initial bunching indicates flexibility (enough to locate at the kink initially). Furthermore, our estimation procedure that we describe later relies on estimating bunching at more than one kink, and therefore it has the potential to incorporate information on the responses of individuals across a wide range of the income distribution (across multiple

kinks).

4 Data

We primarily rely on the restricted-access Social Security Administration Master Earnings File (MEF) and Master Beneficiary Record (MBR), described more fully in the Appendix. The data contain a complete longitudinal earnings history with yearly information on earnings since 1951; the type and amount of yearly Social Security benefits an individual receives; year of birth; the year (if any) that claiming began; and sex (among other variables). Separate information is available on self-employment earnings and non-self-employment earnings. Starting in 1978, the earnings measure reflects total wage compensation, as reported on Internal Revenue Service forms. Our dataset reflects a one percent random sample of all Social Security numbers in the MEF, keeping all available years of data for each individual sampled.

In choosing our main sample, we take into account a number of considerations. It is desirable to show a constant sample in making comparisons of earnings densities. Meanwhile, the AET only affects people who claim OASI, and thus we wish to focus on claimants. However, many individuals claim OASI at ages over the Early Entitlement Age (62), implying that they have not claimed at younger ages but have claimed by older ages. This implies that to investigate a constant sample, we cannot simply limit the sample to claimants at each age (because many people move from not claiming to claiming). To balance these considerations, our main sample at each age and year consists of individuals who have ultimately claimed by the year they turn 65 (i.e. they ultimately claim at an age less than or equal to 65). We exclude person-years with positive self-employment income (and examine the self-employed in Gelber, Jones, and Sacks 2013). We show that the results are robust to other sample definitions, including defining the sample at each age as people who claimed by that age. Because we focus on the intensive margin response (consistent with Saez (2010) and subsequent papers on bunching), in our main analysis we further limit the sample in a given year to observations with positive earnings in that year.⁶

Several features of the data are worth discussion. First, these administrative data allow large sample sizes and are subject to little measurement error. Second, earnings (as measured

⁶When we examine an outcome in year $t + 1$ as a function of a variable in t , the sample consists of those with positive earnings in year t (i.e. we do not further select the sample on positive earnings in $t + 1$).

in the dataset) are taken from W-2 tax forms and are not subject to manipulation through tax deductions, credits, or exemptions. Third, because earnings are taken from the W-2 form, they are subject to third-party reporting (among the non-self-employed), which has been found in the literature to greatly reduce evasion (Kleven et al. 2011). Fourth, the data do not contain information on hours worked or amenities at individuals' jobs.

Table 1 shows summary statistics in our main sample for 62-69 year-olds in 1990-1999, the age and year range that we examine most closely. The sample has 376,431 observations. 50 percent of the sample is male. Mean earnings (conditional on having positive earnings) is \$28,843.11. Median earnings, \$14,555.56, is not far from the AET exempt amount, which averages \$16,738 for those NRA and older, and \$11,650 for those younger than NRA over this time period.

The second dataset we use is the Longitudinal Employer Household Dynamics (LEHD) dataset of the U.S. Census (McKinney and Vilhuber, 2008; Abowd et al., 2009), described further in the Appendix. The data longitudinally follow workers' earnings over time. The data have information on around nine-tenths of workers in covered states and their employers, though we are only able to use data on a 20 percent random subsample of these individuals. We use these data primarily because the sample size in the LEHD is much larger than in the SSA data. We use the LEHD only in the context of two appendix figures (F.4 and F.5); all other analysis is based on the SSA data.

5 Earnings Response to Policy

5.1 Descriptive Evidence from Policy Variation Across Ages

We first examine the pattern of bunching across ages, in order to determine how quickly individuals respond to changes in policy across ages and whether they face delays in responding (consistent with the existence of adjustment frictions). Empirical work often estimates only short-run responses to changes in policy (see Saez *et al.*, 2012, for a review of literature on earnings responses to taxation). If individuals are able to respond more (less) in the long run than in the short run, then this large body of work would under-estimate (over-estimate) long-run responses. Moreover, most empirical specifications have related an individual's tax

rate in a given year to the individual's earnings or taxable (labor) income in that year. In order to choose the appropriate time horizon over which to study behavioral responses to policy, it is necessary to establish how long it takes to respond to policy changes.

Subsequent to 1982, the AET applies to ages 62-69. The policy changes at ages 62 and 70—the imposition and removal of the AET—are "anticipated," in the sense that these changes would be anticipated by those who have knowledge of the relevant policies. We begin by examining the period 1990-1999. Figure 2 plots earnings histograms for each age from 59 to 73 (connected dots), along with the estimated smooth counterfactual polynomial density (smooth line). Earnings are measured along the x-axis, relative to the exempt amount, which is shown by a vertical line. For ages younger than 62, we define the (placebo) kink in a given year as the kink that applies to pre-NRA individuals in that year. For individuals 70 and older, we define the (placebo) kink in a given year as the kink that applies to post-NRA individuals in that year.

Figure 2 shows clear visual evidence of substantial bunching from ages 62-69, when the AET is in effect, and no excess mass at earlier ages. At ages 70 and 71, which are not subject to the AET, there is still clear visual evidence of bunching in the region of the kink.

Figure 3 plots the estimates of normalized excess mass at each age. Bunching is statistically significantly different from zero at each age from 62 to 71 ($p < 0.01$ at all ages). Normalized excess mass is the ratio of excess density in the region of the kink to the prediction from the smooth counterfactual density; at age 62 there are more than twice as many people in this region as predicted by the smooth density, and by age 63 the excess mass rises to nearly four times the prediction from the smooth density. Excess mass remains around this level until age 70 (with a dip at age 65 that we discuss below).

As Figure 3 shows, we continue to find statistically significant bunching among 70 and 71 year-olds, even though they are not subject to the AET. At these ages, we estimate that there are around twice as many individuals in the region of the kink as predicted by the smooth density. We conclude that for some individuals, "unbunching" does not occur immediately. It is possible that a small amount of bunching occurs at ages 72 or older, but this is statistically insignificant.

Three additional considerations also indicate that the earnings densities at ages 70 and

71 reflect delayed adjustment. First, we show later that the statistically significant positive estimates are robust to varying the degree of the polynomial, the excluded region, and the bandwidth used in the estimates. Second, the distributions at other ages not affected by the AET that represent reasonable counterfactuals (such as 61 or 73), show nearly perfectly smooth earnings distributions, suggesting that the excess mass near the kink at ages 70 and 71 would not arise in the absence of the AET. Indeed, as we discuss later, when we use the earnings distribution at age 72 as the counterfactual, we estimate comparable results. Third, Appendix Figure F.4 shows that the mean percentage change in earnings from age 70 to 71 shows a modest spike near the exempt amount, consistent with continued "unbunching" from age 70 to age 71 among those near the kink at age 70 (similar to the spike in earnings growth at the kink in the graph of earnings growth from age 69 to age 70 against age 69 earnings, shown in Appendix Figure F.5). We find it striking that even among the group bunching prior to age 70—that (the data reveal) are able to adjust earnings to the kink—we still find evidence of adjustment frictions.⁷

Figure 3 shows that bunching is substantially lower at age 65 than surrounding ages.⁸ The location of the kink changes substantially from age 64 to age 65; as Figure 1 shows, during this period the exempt amount is much higher for individuals NRA and older than for individuals younger than NRA. Individuals may have difficulty adjusting to the new location of the kink within one year. Prior to the divergence of the exempt amount for those younger and older than the NRA in 1978, we find no such dip in bunching at age 65; this "placebo" evidence further supports the hypothesis that the dip in bunching at age 65 arises from delayed adjustment to the increase in the exempt amount from ages 64 to 65 that emerges after 1978. This delay suggests that individuals also face adjustment frictions in this context.

In our context, the only "appearance" of a new kink that we observe is the appearance

⁷We classify claimants as age 70 based on the highest age they attain in the calendar year. As a result, some individuals (in the extreme case, born December 31) will be classified as age 70 but will be subject to the AET for only a small amount of time (in the extreme case of a December 31 birthday, for only one day). Thus, in principle this is one potential explanation for continued bunching at age 70 that does not rely on earnings adjustment frictions. However, other evidence is sufficient to document earnings adjustment frictions, namely: (1) the continued bunching at age 71, which cannot be explained through the coarse measure of age and must be due to earnings adjustment frictions; (2) the continued adjustment away from the kink from age 70 to age 71 documented in F.4; and (3) the spike in the elasticity estimated using the Saez (2010) approach in 1990 documented in Figure 6 and explained below. Moreover, for those born in January—who are subject to the AET for (nearly) the entire year at age 70—the distribution of earnings at age 70 shows a spike at age 70.

⁸Note that this finding cannot be explained by the appearance of new claimants at age 65, who in principle could bunch less than others as they adjust to being subject to the AET. The reason is that we investigate the sample of people who claimed *by* 65, so the sample of 65 year-olds is identical to the sample of 64 year-olds.

of a kink at age 62. The amount of time since the appearance of the kink at age 62 is correlated with age, and elasticities and adjustment costs could also be correlated with age—thus confounding analysis of the time necessary to adjust to appearance of a kink. While recognizing these caveats, it is worth noting that the amount of bunching slowly rises from age 62 to 63, which suggests gradual adjustment. In principle, this could also relate to the fact that these graphs show the sample of those who have claimed by age 65, and the probability of claiming at a given age (conditional on claiming by age 65) rises from age 62 to 63. To address this issue, in Appendix Figure F.3 we show the results when the sample at a given age consists of those who have claimed by that age, which still shows a substantial increase in bunching from 62 to 63.

Each of these several pieces of evidence points to delayed adjustment, though we consider continued bunching at age 71 to be the single most convincing source of evidence. Similar patterns of adjustment occur when looking at the periods 1972-1982, 1983-1989 and 2000-2006 (Appendix Figures F.6 to F.8). We find evidence of adjustment delays, as individuals continue to bunch at the kink at ages older than the highest age to which the AET applies. However, in no case does adjustment appear to take more than three years.

5.2 Other Evidence Relating to Bunching

We conduct a variety of robustness tests. Appendix Figure F.9 uses a bandwidth of \$500 instead of \$800, which changes our estimates little (as have other bandwidths we have chosen). With a typical monthly benefit of \$1,000, the DRC began applying within \$2,000 of the kink in years with a benefit reduction rate of 50 percent, so it is notable that our results are robust to the smaller bandwidth of \$500 (encompassing a region of only \$1,750 on either side of the kink). In Appendix Figure F.10, we vary the degree of the polynomial we use between 6 and 8, which shows similar results; other sufficiently rich polynomials we have tried have also shown similar results. In Appendix Figure F.11, we vary the region near the kink we exclude when estimating the amount of bunching (from \$2,000 to \$3,000 to \$4,000) and again estimate similar results. Limiting the sample to those who have substantial benefits (such as those with \$1,000 or higher in benefits)—so that they are safely far from the concave kink in the budget set created when the AET reduces OASI benefits to zero—also

yields very similar results. Appendix Figure F.12 shows that both men and women bunch at the kink (though interestingly, men show more bunching than women).

We find no evidence of adjustment in anticipation of future changes in policy, as those younger than 62 do not bunch. If the cost of adjustment in each year rose with the size of adjustment and this relationship were convex, we would expect anticipatory adjustment.

Gelber, Jones, and Sacks (2013) also discuss adjustment to the elimination of the AET in 2000 for those older than the NRA. This episode shows some evidence of delayed adjustment to the policy change, though the evidence is not as clear as in the case of adjustment to the removal of the AET after age 69. However, it is clear that in at least some contexts—*i.e.* when aging out of the AET at age 70, after the policy change in 1990 that we discuss later, and quite possibly after the policy change in 2000—earnings adjustment frictions prevent some individuals from reacting immediately to the removal of a kink.

Gelber, Jones, and Sacks (2013) also discuss additional findings, including evidence relating to whether employers or employees drive the degree of bunching, bunching among the self-employed, whether expected mortality helps drive adjustment to the kink, and whether the AET influences earnings by inducing earnings changes within or across employers. They also perform a longitudinal analysis showing that individuals locating at the kink in one year tend to locate at the kink again in the next year; thus, those who show delayed responses to the removal of the kink generally located at the kink the prior year.

6 Method for Estimating Elasticities and Adjustment Costs

The results thus far suggest a role for adjustment frictions in individuals' earnings choices. As a first step in incorporating frictions into an estimable model of earnings supply, we build on the Saez (2010) model described briefly in Section 3, which uses bunching to identify the elasticity of (taxable) earnings with respect to the net-of-tax rate, $\varepsilon \equiv -(\partial z / z) / (\partial \tau / (1 - \tau))$. We extend this model to allow for a cost of adjusting to tax changes, and we develop a method to estimate elasticities and adjustment costs jointly.

As described further in Appendix E, agents maximize utility $u(c, z; n)$ over consumption

c and pre-tax earnings z (where greater earnings are associated with greater disutility at the margin), subject to a budget constraint $c = (1 - \tau)z + R$, where R is virtual income. The parameter n reflects "ability" (i.e. the tradeoff between consumption and earnings supply) and creates heterogeneity in utility given consumption and earnings.⁹ We first consider a transition from a linear tax schedule with a constant MTR τ_0 to a schedule with a convex kink, where the MTR below the kink earnings level z^* is τ_0 , and the MTR above z^* is $\tau_1 > \tau_0$. Second, we consider a decrease in the higher MTR to $\tau_2 < \tau_1$, which mirrors our empirical context.

We assume that in order to change earnings from an initial level, individuals must pay a fixed utility cost of ϕ^* . This cost could represent the information costs associated with navigating a new tax regime if, for example, individuals only make the effort to understand their earnings incentives when the utility gains from doing so are sufficiently large (*e.g.* Simon 1955; Chetty *et al.* 2007; Hoopes, Reck, and Slemrod 2013). Alternatively, this cost may represent frictions such as the cost of negotiating a new contract with an employer or the time and financial cost of job search, assuming that these costs do not depend on the size of the desired earnings change. All of these factors may play a role even when individuals consider moving from one positive earnings level to another positive level.

Our fixed cost model follows recent literature that has focused on fixed costs (*e.g.* Chetty *et al.*, 2011; Chetty, 2012). Gelber, Jones, and Sacks (2013) compare earnings distributions at ages subject to the AET (*e.g.* age 62 or 69) to those not subject to the AET (*e.g.* age 61, 70, or 71). Compared to ages not subject to the AET, the earnings distributions at ages subject to the AET show a sharp spike at the kink; a higher density immediately to the right of the kink; generally a lower density at earnings levels starting several thousand dollars above the kink; and eventually reach a similar earnings density. This is consistent with a simple model with fixed adjustment costs that lead to a region of inaction and a region of adjustment (though we discuss later the difficulty of using such features of the data to identify the parameters, as such features are sensitive to the joint distribution of elasticities

⁹Recent literature on bunching—including Saez (2010), Chetty *et al.*, 2009, 2011, 2012a,b, Chetty (2012), and Kleven and Waseem (2013)—specifies a static model of earnings choice in each period. As we describe, our analysis incorporates the dynamic consideration that the transition from the initial kink to the final kink matters, as individuals face an adjustment cost in making any possible earnings change from the initial period to the final period so that the temporal ordering of these budget sets matters.

and adjustment costs).

In Appendix E.5.1, we extend our model to a case in which the cost of adjustment is linear in the size of the adjustment. In this case, we need at least three kinks to estimate the three parameters of interest (the elasticity, fixed cost, and slope of the variable cost). Because we do not examine a setting in which one can compare bunching for a given group under three different successive positive tax rates, we do not have a credible source of variation to estimate all of these parameters.

In our main model (i.e. with a fixed cost but no variable cost), we first show that the introduction of a fixed adjustment cost attenuates bunching relative to a frictionless model. In this case, the elasticity can no longer be estimated using a single cross section of data. We then show that adjustment costs also attenuate the change in bunching in response to a change in the size of the kink. Finally, we demonstrate that both the elasticity and adjustment cost may be estimated using data from bunching at kinks before and after this policy change. We additionally extend our model to accommodate heterogeneity in ε and ϕ^* , a flexible earnings distribution in the neighborhood of the kink, and a richer earnings adjustment process.

6.1 Bunching in a Single Cross-Section with Adjustment Costs

We first describe behavior following the introduction of a kink – with a jump in MTR from τ_0 to τ_1 at z^* – when previously there was only a linear tax schedule with rate τ_0 . As shown in Appendix E and described in Section 3, in a frictionless model, a range of individuals will locate at the exempt amount in the presence of a kinked budget set—those whose *ex ante* earnings (i.e. their earnings when faced with a linear tax schedule of τ_0) lie in the range $[z^*, z^* + \Delta z_1^*]$. Saez (2010) illustrates how the level of bunching can be related to the parameter Δz_1^* , and, therefore, can be used to estimate ε .

Figure 4 Panel A illustrates how a fixed adjustment cost attenuates the level of bunching and obscures the estimation of ε in a single cross-section that is possible in the Saez (2010) model. Consider the person with initial earnings z_1 along the linear budget constraint with tax rate τ_0 at point 0. This individual faces a higher marginal tax rate τ_1 after the kink is introduced, which increases the marginal tax rate to τ_1 above earnings level z^* . Because she

faces an adjustment cost, she could decide to keep her earnings at \underline{z}_1 and locate at point 1. Alternatively, with a sufficiently low adjustment cost, she would like to pay the adjustment cost and reduce her earnings to z^* , marked by point 2.

We assume that the benefit of relocating to the kink is increasing in distance from the kink for initial earnings in the range $[z^*, z^* + \Delta z_1^*]$. In general, this requires that the size of the optimal adjustment in earnings increases in n at a rate faster than the decrease in the marginal utility of consumption. We explore the implications of this assumption in Appendix E.4. This assumption is true, for example, if utility is quasilinear, which is assumed in related recent literature (e.g. Saez 2010, Chetty et al. 2011, Kleven, Landais, Saez, and Schultz 2012, and Kleven and Waseem 2013).

These assumptions imply that above a threshold level of initial earnings, \underline{z}_1 , individuals adjust their earnings to the kink, and below this threshold individuals remain inert. We have drawn this individual as the marginal buncher who is indifferent between staying at the initial level of earnings \underline{z}_1 (at point 1) and moving to the kink earnings level z^* (point 2) by paying the adjustment cost ϕ^* .

In Panel B, we show that the level of bunching is attenuated due to the adjustment cost. Panel B plots the counterfactual density of earnings, i.e. under a linear tax τ_0 . Only individuals with initial earnings in the range $[\underline{z}_1, z^* + \Delta z_1]$ bunch at the kink (areas *ii*, *iii*, *iv*, and *v*)—whereas in the absence of an adjustment cost, individuals with initial earnings in the range $[z^*, z^* + \Delta z_1^*]$ bunch (areas *i*, *ii*, *iii*, *iv*, and *v*). The amount of bunching is equal to the integral of the initial earnings density over the range $[\underline{z}_1, z^* + \Delta z_1^*]$:

$$B_1(\boldsymbol{\tau}_1, z^*; \varepsilon, \phi^*) = \int_{\underline{z}_1}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta, \quad (1)$$

where $h_0(\cdot)$ is the density of counterfactual earnings, i.e. earnings in the presence of a linear tax τ_0 , $\boldsymbol{\tau}_1 = (\tau_0, \tau_1)$ measures the tax rates below and above z^* , and ζ is a dummy variable of integration. The lower limit of the integral, \underline{z}_1 , is implicitly defined by the indifference condition drawn in Figure 4, Panel A:

$$\phi^* \equiv u((1 - \tau_1)z^* + R_1, z^*; \underline{n}_1) - u((1 - \tau_1)\underline{z}_1 + R_1, \underline{z}_1; \underline{n}_1) \quad (2)$$

where R_1 is virtual income, and \underline{n}_1 is the "ability" level of this marginal buncher.

Bunching therefore depends on the preference parameters ε and ϕ^* , the tax rates below and above the kink, τ_1 , and the density $h_0(\cdot)$ near the exempt amount z^* . It is clear that with only one kink, we cannot estimate both ε and ϕ^* , as the level of bunching depends on both parameters.

6.2 Estimation Using Variation in Kink Size

We can estimate elasticities and adjustment costs when we observe bunching at a kink both before and after the jump in the marginal tax rate at the kink earnings level changes (which may involve the kink disappearing), as we observe in our empirical applications. Suppose we observe a population that moves from facing a more pronounced kink K_1 (i.e. a kink with a larger jump in the marginal tax rate when moving from below to above z^*), with a marginal tax rate τ_1 above z^* , to facing a less pronounced kink K_2 (i.e. a kink with a smaller jump in the marginal tax rate when moving from below to above z^*), with a marginal tax rate of $\tau_2 < \tau_1$ above z^* . Adjustment costs prevent some individuals from unbunching from the kink, even though they would prefer to move away from the kink in the absence of an adjustment cost, because the gain from unbunching is not large enough to outweigh the adjustment cost. The fixed adjustment cost therefore attenuates the reduction in bunching, relative to a frictionless case.¹⁰

Attenuation in the change in bunching is driven by individuals in area *iv* of Panel B. Under a frictionless model, individuals in this range do not bunch under the smaller kink K_2 . To see this, note that their counterfactual earnings are greater than $z^* + \Delta z_2^*$, i.e. the highest level of initial earnings among bunchers at K_2 when there are no frictions. However, when moving from K_1 to K_2 in the presence of frictions, those in area *iv* continue to bunch. Panel C of Figure 4 demonstrates this. At point 0, we show an individual's initial earnings $\bar{z}_0 \in [z^*, z^* + \Delta z_1^*]$ under a constant marginal tax rate of τ_0 . We now introduce the first kink, K_1 . The individual responds by bunching at z^* at point 1. Next, we transition to the less pronounced kink K_2 . Note that since $\bar{z}_0 > z^* + \Delta z_2^*$, this individual would have

¹⁰If $d\tau_2 > d\tau_1$ instead – i.e. the kink becomes larger – then additional individuals will be induced to bunch, but the change in bunching will in general be still be attenuated (due to the adjustment cost). This is governed by an analogous set of formulas to the case $d\tau_2 < d\tau_1$ that we explore.

chosen earnings $\bar{z}_2 > z^*$ (marked as point 2) under τ_2 in a frictionless setting. However, in order to move to point 2, this individual must pay a fixed cost of ϕ^* . We have drawn this individual as the marginal buncher who is indifferent between staying at z^* and moving to \bar{z}_2 . All individuals with initial earnings in the range $[z^* + \Delta z_2^*, \bar{z}_0]$ will remain at the kink.

Thus, bunching under K_2 is:

$$\tilde{B}_2(\tilde{\tau}_2, z^*; \varepsilon, \phi^*) = \int_{z_1}^{\bar{z}_0} h_0(\zeta) d\zeta, \quad (3)$$

where $\tilde{\tau}_2 = (\tau_0, \tau_1, \tau_2)$ measures the tax rate below z^* , the initial tax rate above z^* , and the final tax rate above z^* , respectively. As discussed in Appendix E.6, ε is still identified by the adjustment of the top-most buncher: $\varepsilon = \frac{\bar{z}_0 - \bar{z}_2}{\bar{z}_2} \frac{(1 - \tau_0)}{d\tau_2}$. The critical earnings level \bar{z}_2 is defined implicitly by the indifference condition in Panel C:

$$\phi^* \equiv u((1 - \tau_2)\bar{z}_2 + R_2, \bar{z}_2; \bar{n}_2) - u((1 - \tau_0)z^* + R_0, z^*; \bar{n}_2). \quad (4)$$

The four equations (1), (2), (3) and (4) together pin down four unknowns (Δz_1^* , z_1 , \bar{z}_0 and \bar{z}_2), each of which is in turn a function of ε and ϕ^* . ε and ϕ^* therefore jointly determine bunching at each kink. Ultimately, we draw on two empirical moments in the data, B_1 and \tilde{B}_2 , to identify our two key parameters, ε and ϕ^* .

When we later perform our main estimates, we make use of a minimum distance estimator described in Appendix E.10 to find the values of ε and ϕ^* so that the bunching predicted by the model matches the bunching in the data. Intuitively, we rely on a before-and-after comparison of bunching at the same kink, once the jump in marginal tax rates has been reduced. Inertia generates an excess amount of bunching in the period after the policy change. As we explain below, the greater the adjustment cost, the greater the inertia; thus, the degree of inertia helps in estimating the adjustment cost. In the extreme case in which a kink has been eliminated, we can attribute any residual bunching to adjustment costs. The amount of residual bunching at the kink, in combination with the amount of bunching prior to the change in the jump in MTRs at the kink, therefore helps to identify both the elasticity and the adjustment cost. An approximation explained below in Section 6.3 also

helps to build intuition.

Relative to the frictionless case represented by the Saez model, the change in bunching from the more pronounced kink K_1 to the less pronounced kink K_2 is now attenuated by the adjustment cost. As noted above, in the Saez model, bunching decreases by areas iv and v in Figure 4 when moving from K_1 to K_2 . When moving sequentially from K_1 to K_2 in the presence of an adjustment cost, areas ii , iii , iv , and v bunch under K_1 , whereas areas ii , iii , and iv bunch under K_2 . Thus, bunching decreases only by area v , rather than by both areas iv and v as in the frictionless case. The absolute value of the decrease in bunching from K_1 to K_2 is decreasing in the adjustment cost— \bar{z}_0 is increasing in the adjustment cost, and therefore area v is decreasing in the adjustment cost. As in the frictionless case, the amount of bunching at K_1 is still increasing in the elasticity (*ceteris paribus*). Such observations help to provide intuition for the features of the data that help drive our estimates of the elasticity and adjustment cost.

The key assumption underlying this method is that utility is quasi-linear and isoelastic, which is common in the bunching literature (see Saez 2010, Chetty *et al.* 2011, Kleven, Landais, Saez, and Schultz 2012 and Kleven and Waseem 2013, for example). If we were to relax the assumption of quasilinearity, we would need to observe wealth, which is not available in the data. We also note that income effects are second-order just above a kink; while we investigate a substantial kink, small income effects may be more relevant in other contexts in which our method is applicable.

6.3 Intuition and Tractable Approximation

To build intuition regarding the impact of ε and ϕ^* on the level of bunching, we derive an expression relating the elasticity and adjustment cost to the level of bunching that can be easily solved in closed form. We make use of a series of approximations to specify a simple system of linear equations. Let $b \equiv B/h_0(z^*)$, *i.e.* the amount of bunching scaled by the density of earnings at z^* when there is no kink. Also assume that $h_0(z)$ is uniform and equal to $h_0(z^*)$ in the range between \underline{z}_1 and $[z^* + \Delta z_1^*]$. This assumption is commonly made in the literature on bunching (Chetty *et al.* 2011, Kleven and Waseem 2013), but we discuss below in Section 6.4.2 how we can relax this assumption. We show in Appendix E.5 that

scaled bunching when moving from no kink to K_1 is approximately:

$$b_1(\tau_1, z^*; \varepsilon, \phi) = \varepsilon \left(z^* \frac{d\tau_1}{1 - \tau_0} \right) - \phi \left(\frac{1}{d\tau_1} \right), \quad (5)$$

where $d\tau_1 = \tau_1 - \tau_0$ and $\phi = \phi^*/u_c$ is the dollar equivalent of the adjustment cost.¹¹ This equation shows intuitive comparative statics: All else equal, bunching at the initial kink is increasing in the elasticity, decreasing in the adjustment cost, and increasing in the size of the tax change at the kink. This generalizes and nests the formula developed in Saez (2010), which is equivalent in the case in which there is no adjustment cost. Because the amount of bunching is decreasing in the adjustment cost, constraining $\phi = 0$ and using the Saez (2010) method will weakly underestimate the elasticity in the case in which a kink is introduced to a linear tax schedule, since attenuation in bunching is attributed to a small elasticity rather than to the adjustment cost.¹² Note that the expository derivation in (5) does not impose quasilinearity but uses the uniform density assumption and a first-order approximation for utility in the neighborhood of the kink.

We additionally show in Appendix E.6 that once the kink is reduced from K_1 to K_2 , the resulting level of scaled bunching is approximately:

$$\tilde{b}_2(\tau_2, z^*; \varepsilon, \phi) = \varepsilon \left(\bar{z}_2 \frac{d\tau_2}{1 - \tau_0} \right) - \phi \left(\frac{1}{d\tau_1} \right) + (\bar{z}_2 - z^*) \quad (6)$$

where the \tilde{b}_2 denotes bunching at K_2 when transitioning directly from K_1 and \bar{z}_2 is defined above in equation (4). Again, the amount of bunching is increasing in ε . In comparison with equation (5), additional bunching at the second kink due to inertia is captured in equation (6) by (a) the appearance of $\bar{z}_2 > z^*$ in the first term, (b) the $d\tau_1 > d\tau_2$ in the denominator of the second term and (c) an additional third term $(\bar{z}_2 - z^*) > 0$.

The model of Saez (2010) describes how bunching should vary between two different kinks in a frictionless setting, and the extent to which observed bunching deviates from this pattern is attributed to the adjustment cost. Further intuition is gained by examining the three policy changes that we have considered. In the first case, we transition from a linear tax schedule to

¹¹In the quasi-linear case $\phi^* = \phi$, i.e. the utility cost of adjustment is expressed in dollar unit. Thus, we hereafter refer to the adjustment cost simply as ϕ .

¹²As we describe later, when a kink becomes less pronounced, the Saez (2010) formula will weakly over-estimate the elasticity.

a large kink K_1 . In the second case, we transition from a linear tax schedule to a smaller kink K_2 . In the final case, relevant to our setting, we transition from a linear tax schedule to the more pronounced kink K_1 and then directly to the less pronounced kink K_2 . Denote scaled bunching (i.e. $b = B/h_0(z^*)$) in the final period in each of the three cases as b_1 , b_2 and \tilde{b}_2 , respectively. Under the frictionless model, we have $b_2/b_1 = \tilde{b}_2/b_1 = d\tau_2/d\tau_1$; bunching scales with the change in the tax rate at the kink. However, in a model with a fixed adjustment cost, we have $b_2/b_1 < d\tau_2/d\tau_1 < \tilde{b}_2/b_1$. The inequality $d\tau_2/d\tau_1 < \tilde{b}_2/b_1$ is driven by inertia in bunching when the kink sharply reduces. The ratio \tilde{b}_2/b_1 is increasing in ϕ , holding $d\tau_2/d\tau_1$ constant. The reduction in scaled bunching is attenuated by the adjustment cost. Thus, the estimation of the adjustment cost relies on the less-than-proportionate reduction in scaled bunching. This is equivalent to showing that under a fixed cost of adjustment, an elasticity estimated according to a misspecified, frictionless model will actually *increase* following a sharp reduction in the kink (as we show in Appendix E.7). This generates a testable prediction of our model that we verify empirically below. We similarly show in Appendix E.6 that the absolute value of the change in bunching when moving from K_1 to K_2 , $|\tilde{b}_2 - b_1|$, is decreasing in ϕ .

6.4 Extensions

6.4.1 Heterogeneity in Elasticities and Fixed Costs of Adjustment

The previous analysis assumed homogeneous elasticities and adjustment costs, but we can extend the model to accommodate heterogeneity, as we show more fully in Appendix E.8. Suppose $(\varepsilon_i, \phi_i, n_i)$ is jointly distributed according to a smooth CDF, which translates into a smooth, joint distribution of elasticities, fixed adjustment costs, and earnings in the presence of a linear tax. Again assuming that the density of earnings, conditional on ε and ϕ , is constant over the interval $[\underline{z}, z^* + \Delta z^*]$, we derive generalized formulae for bunching analogous to equations (1) and (3) that allow us to interpret our estimates as the average behavioral response and attenuation due to adjustment costs among the set of bunchers.¹³ Note that we make no assumptions on the joint distribution of $(\varepsilon_i, \phi_i, n_i)$ other than smoothness.

Our estimates of elasticities and adjustment costs are specific to the population observed

¹³We are grateful to Henrik Kleven for suggesting the approach that led to this derivation.

bunching at the kinks. At the same time, note that for any value of ϕ_i , there exists a value of ε_i that generates positive bunching. Thus, while our estimates are local to the observed set of bunchers, they need not be confined to a subpopulation with small values of ϕ_i —for example, if ε_i and ϕ_i are positively correlated. Nevertheless, there may be a set of individuals for whom ε_i is small enough relative to ϕ_i to preclude bunching under either K_1 or K_2 , and who therefore do not contribute to our parameter estimates. It is important to note that while this may be a limitation of our particular policy setting, it is not a general methodological limitation in the sense that with sufficiently large variation in tax rates it may be possible to estimate population average parameters. Loosely speaking, the greater the variation in tax rates, the more of the population we will observe who bunch at kinks and therefore contribute to our estimates. In that light, our policy variation is useful in varying over a large range of BRRs (from 50 percent to 33 percent to 0 percent). Moreover, it is perhaps reassuring that we will find similar elasticity and adjustment cost estimates when we examine a larger change in the BRR (from 33 percent to zero percent) than when we examine a smaller change (from 50 percent to 33 percent). Extrapolating our estimates from bunchers to non-bunchers would require assumptions on the joint distribution of ε and ϕ . It is striking that our estimates will find positive adjustment costs even among those initially bunching at the kink, whose initial bunching indicates enough flexibility to locate at the kink initially.

6.4.2 Flexible Distribution of Counterfactual Earnings

As explained in Appendix E.10, in our simplest estimation strategy we assume that the density of initial earnings $h_0(z)$ is uniform over the range $[z^*, z^* + \Delta z^*]$. Although this assumption is common to the literature (see, e.g., Chetty *et al.* 2011 or Kleven and Waseem 2013), we show in Appendix E.10 that we may alternatively use the nonparametric density of earnings estimated in our data (using data on 72 year-olds, who represent a reasonable counterfactual earnings density because they no longer face the AET, no longer show bunching, and are close in age to those aged 70 or 71). Using the nonparametrically-estimated distribution of earnings from age 72 is helpful because it does not entail distributional assumptions, but this comes at the cost of using a different age (i.e. 72) to generate the

earnings distribution. Below we report estimates across the various methods and show that our results are extremely similar regardless of the method chosen.

6.4.3 Frictions in Initial Period

Our model above assumes that initial earnings under τ_0 are located at the frictionless optimum. However, it is also possible to assume that individuals may find themselves away from their frictionless optimum in period 0, due to the same adjustment costs that attenuate bunching under K_1 and K_2 . In Appendix E.9, we specify an extended model that allows for individuals to be arbitrarily located in a neighborhood of their frictionless optimum. As in Chetty (2012), we only require that earnings are close enough to the optimum to preclude any further utility gains that outweigh the adjustment cost ϕ^* . We derive generalized versions of equations (1) and (3) that hold for arbitrary initial conditions consistent with a fixed adjustment cost. We then show that estimation now requires structure on the distribution of initial earnings within the neighborhood of the frictionless optimum. We demonstrate how our estimator is amended in the case of a uniform distribution of initial earnings about the frictionless optimum and report estimates under this method below.

6.4.4 Using Other Moments of the Data

While we focus on the estimation of our key parameters using the amount of bunching, the model has implications for other moments of the earnings distribution. For example, the set of non-bunchers (e.g. area i in Figure 4, Panel B) will tend to overlap with relatively high earners (those to the right of area v) who locate just to the right of the kink K_1 after it is introduced. This will tend to generate extra mass just above z^* . Our model in Section 6.2 also implies that under certain parameter values, there will be a hole in the *ex post* density just to the right of \underline{z}_1 once K_2 is introduced (when $\bar{z}_2 > \underline{z}_1$). As another example, with homogeneous elasticities one would expect that the distribution of earnings on the period 0 linear tax schedule, conditional on locating at the kink K_1 in period 1, would show a hole above the kink (corresponding to area i in Figure 4, Panel B).

However, in the case of heterogenous preferences (Section 6.4.1) or initial earnings away from the optimum (Section 6.4.3), these features of the data will tend to be smoothed or eliminated. For example, if individuals with low fixed costs of adjustment tend to have low

elasticities, then the period 0 earnings distribution, conditional on locating at the kink in period 1, should tend to be closer to the kink on average than if individuals with low fixed costs of adjustment tend to have high elasticities. Additionally, if individuals are located in the neighborhood of their frictionless optimum in Period 0 (as in Section 6.4.3), they will tend to fill in the holes predicted in the *ex post* density after moving from K_1 to K_2 . Thus, such features of the data will not allow us to estimate ε and ϕ without more stringent assumptions on the joint distribution of ε and ϕ (which is left unrestricted in our method above for addressing heterogeneity).¹⁴

6.4.5 Unrelated Cross Sections

The estimation method described above relies on a comparison of bunching before and after a policy change. In Gelber, Jones, and Sacks (2013), we develop an alternative method based on bunching at kinks in two or more cross sections that need not be related in this way. In other words, under this method, we need not observe data from both before and after a policy change; we could, for example, compare bunching in a given population and at a given kink to bunching in another population at an unrelated kink. Under certain approximations, this method also yields simple, closed-form expressions for the elasticity and adjustment cost in terms of the amount of normalized excess mass observed at two kinks. Gelber, Jones, and Sacks (2013) contains details and estimation results.

7 Estimates of Elasticity and Adjustment Cost

To estimate ε and ϕ , we first examine the reduction in the rate in 1990, and next we turn to the elimination of the AET from ages 69 to 70 that our graphical analysis above considers. We use 1990 as a baseline because we can apply the Saez (2010) method in 1990 as a comparison to our method (whereas as we explain, we cannot apply the Saez method at age 70 or later because the benefit reduction rate is zero above the exempt amount); however, it is important to emphasize our results are similar when examining both contexts.

Estimating these parameters requires estimates of the implicit marginal tax rate that

¹⁴Kleven and Waseem (2013) show that in the case of a notch, both the excess mass at the notch and the hole in the density to the right of the notch may be used to quantify adjustment costs. In contrast to their case, the hole in our density is not guaranteed to show up in the data. The reason is that unlike the dominated region in their case, there is generally no common interval at the intersection of our holes, under various values of ε and ϕ .

individuals face. This requires estimates of both the "baseline" marginal tax rate, τ_0 —the rate that individuals near the AET threshold face in the absence of the AET due to federal and state taxes—and estimates of the implicit marginal tax rate associated with the AET. We begin by using a marginal tax rate that incorporates the effects of the AET BRR, as well as the average marginal income and FICA tax rates (including federal and state taxes), as described in the Appendix. Recall that Appendix A shows that benefit enhancement is not relevant to an individual's marginal incentives for earning an extra dollar near the AET exempt amount. This assumption is also consistent with the methodology of Friedberg (1998, 2000), who treats the AET as a pure tax. Moreover, the time series of bunching shows little systematic bunching reaction around the time of changes in the DRC. We vary these assumptions in various dimensions, which show similar results to the baseline: we exclude FICA taxes in the calculation of the baseline tax rate (following Liebman, Luttmer, and Seif (2009), who suggest that individuals may not perceive FICA taxes as pure taxes), and we alternatively assume that the benefit enhancement corresponds to a reduction in the effective marginal tax rate.

Before turning to our empirical estimates, we begin with graphical depictions of the patterns driving the parameter estimates for the 1990 change. Figure 5 shows bunching among 66-69 year-olds, for whom the BRR fell from 50 percent to 33.33 percent in 1990 (the patterns around 1990 are extremely similar for the 67-68 year-old group that we focus on in our estimates). Bunching fell slightly from 1989 to 1990 but fell more subsequent to 1990. For comparison, Figure 5 also shows that bunching stayed relatively constant—both in 1990 and subsequently—for the 62-64 year-old group that experienced no policy change in 1990. The relative comparison demonstrates that bunching fell only slightly among 66-69 year-olds relative to the "control" group of 62-64 year-olds in 1990, but fell more after 1990—suggesting an incomplete reaction to the policy change among the 66-69 year-old group in 1990.

Figure 6 shows that the elasticity we estimate using the Saez (2010) method—constraining the adjustment cost to be zero—rises sharply from 1989 to 1990. This relates directly to our theory, which predicts that following a reduction in the change in the MTR at the kink, there

may be excess bunching due to inertia (corresponding to area *iv* in Figure 4, Panel B).¹⁵ Once we allow for an adjustment cost, this excess bunching is attributed to optimization frictions. Indeed, in Appendix E.7 we explain that a rise (from just before to just after the policy change) in the elasticity that we estimate using the Saez (2010) method is a telltale sign that we face an adjustment cost as modeled above. Specifically, we show that if agents actually face an adjustment costs but we mis-specify the model as a frictionless (Saez 2010) model, and we face a decrease in the jump in the marginal tax rate at the kink, the estimated Saez (2010) elasticity will weakly rise from just before to just after this policy change. In a context in which individuals have not yet had a chance to adjust (and the effective marginal tax rate has fallen), frictions may lead to larger elasticity estimates. Interestingly, this case yields the opposite of the usual presumption that adjustment frictions should always lead to attenuation of elasticity estimates. Our finding also contrasts with the usual presumption that in the presence of adjustment frictions, smaller variation in taxes (i.e. smaller kinks) always yield smaller elasticity estimates.

Table 2 presents results from our estimation method, examining the 1990 change. We follow a group of 67-68 year-olds, so that we can examine an age group that moved over time from being affected by the 50 percent tax rate before the policy change to the 33 percent tax rate after the policy change, as required under our method. An individual who is 69 in 1989, for example, would have turned 70 by 1990 and therefore would not have been affected by the 33 percent tax rate in 1990—which prevents us from examining those age 69 if we wish to examine a constant age group (which is crucial because different age groups have persistently different amounts of bunching, which we do not wish to confound with the effect of the policy change).

It is worth discussing the primary variation that drives identification of the parameters in Table 2. The variation driving our identification is *not* confounded with age variation, because we hold ages constant from before to after the policy change (examining 67-68 year-olds in both cases). The variation instead comes from the time series shown in Figure 5 for 66-69 year-olds—specifically, we compare the amount of bunching in 1990 to the amount in

¹⁵In the Section 8, we discuss dynamic considerations that might subsequently cause residual bunching to disappear. Our focus is limited to the period just after a policy change, before residual bunching has dissipated.

1989. While this particular observation is in principle indistinguishable from time dummies, we believe that a number of considerations make our identification credible. First, in the "control group" of 62-64 year-olds who do not experience a policy change in 1990, bunching is *extremely* stable in the years before and after 1990. Since this counterfactual shows no discernable changes in bunching, this suggests that the amount of bunching in the 66-69 year-old group will be sufficient to pick up changes due to the policy change. Second, Figure 6 shows a sharp, sudden spike in bunching as estimated through the Saez (2010) method in 1990. This is consistent with our interpretation that the excess bunching in 1990 reflects delayed adjustment to the policy change, as opposed to some other shock; if it were due to another shock, we might expect similar year-to-year variation in bunching in Figure 6 across other years, which we do not see. Third, we estimate similar results when we examine the removal of the kink at age 70 in Table 3; these results cannot be driven by shocks across ages, because there is no reason for bunching at ages over 70 except delayed adjustment to the removal of the kink.

We estimate an elasticity of 0.23 in Column (1) of Table 2 and an adjustment cost of \$152.08 in Column (2), both significantly different from zero ($p < 0.01$). This specification examines data in 1989 and 1990; thus, our estimated adjustment cost represents the cost of adjusting earnings in the first year after the policy change. We interpret our estimates as meaning that when considering a given time frame (in this particular case, from 1989 to 1990), bunching amounts can be predicted if individuals behaved as if they had the estimated elasticity and adjustment cost (in this particular case, 0.23 and \$152.08, respectively).

When we constrain the adjustment cost to zero using 1990 data in Column (3), as most previous literature has implicitly done, we estimate a substantially larger elasticity of 0.39.¹⁶ Consistent with our discussion above, it makes sense that the estimated elasticity is higher when we do not allow for adjustment costs than when we do, as adjustment costs keep individuals bunching at the kink even though tax rates have fallen. The difference in the constrained and unconstrained estimates of the elasticity is substantial (69 percent higher in the constrained case) and statistically significant ($p < 0.01$).

¹⁶Friedberg (2000) finds uncompensated elasticity estimates of 0.22 and 0.32 in different samples. However, differences in the estimation strategies imply that these results are not directly comparable to ours.

We also apply alternative specifications in the context of the 1990 change. Using a nonparametrically-estimated distribution of earnings (using the earnings distribution among 72 year-olds) rather than a locally uniform density (as described in Appendix E.10) yields very similar estimates. Adjusting the marginal tax rate to take account of benefit enhancement (applicable to those individuals to whom benefit enhancement is relevant to their earnings choices) raises the estimated elasticity but yields similar qualitative patterns across the constrained and unconstrained estimates. This makes sense: for the same behavioral response, if we assume a less pronounced percentage change in the net-of-tax rate, we infer a larger elasticity. The next rows show other specifications: excluding FICA taxes from the baseline tax rate; other bandwidths; and other years of analysis. Our results are similar under these and other variations.

In Table 3, we apply our method to the disappearance of the kink at age 70 in 1990-1999 (in which context we find residual bunching in our main graphical analysis in Figure 2) and find similar elasticity estimates and adjustment costs. When comparing adjustment at age 70 to adjustment in 1990, a key pattern in the data consistent with our model is that the fall in normalized excess mass from 1989 to 1990 shown in Figure 5 is much smaller (in absolute and percentage terms) than the fall in normalized excess mass from age 69 to age 70 shown in Figure 3. With an adjustment cost preventing immediate adjustment as in our model, normalized excess mass should fall less when the jump in marginal tax rates at the kink falls less (as in the change from a 50 percent to a 33 percent BRR in 1990) than when the jump in marginal tax rates at the kink falls more (as in the change from a 33 percent to a 0 percent BRR at age 70).

In Appendix Table G.1, we apply our method to the 1990 policy change but assume that bunching in 1989 is not attenuated by adjustment frictions (under the rationale that bunching could have reached a "steady state" in 1989 that is not attenuated by adjustment frictions). We estimate results similar to the baseline. Finally, in Appendix Table G.2, we apply our method to the 1990 policy change but allow individuals to be initially located away from their frictionless optimum, as described above and in Appendix E.9. We again estimate similar results to the baseline.

It is also worth considering the relationship of our results to the Saez (2010) estimates.

Since we estimate that it takes a few years to react to changes in the AET, it may be possible to estimate the elasticity by estimating the elasticity using the method of Saez (2010) a few years after a kink has disappeared. In that light, it is reassuring that our elasticity estimate of 0.23 is similar to the elasticities estimated through the Saez approach two and three years after the 1990 policy change (which are 0.26 and 0.16 in 1992 and 1993, respectively, in Figure 6) and are also similar to the Saez elasticities estimated just before the policy change (which are 0.23 and 0.22 in 1988 and 1989, respectively). In fact, the average Saez elasticity estimated over these four years is 0.22, which is nearly identical to our estimate of 0.23. Although this suggests one may be able to estimate the elasticity through the Saez (2010) method within two to three years of a policy change in this particular context, our approach is useful in a number of respects. First, we are able to estimate adjustment costs faced in the short run (in this context, within one or two years of facing a change in policy). Second, without verifying that excess bunching dissipates within two to three years—as we do—one would not know that the Saez formula may be applicable in our setting once two to three years have passed. Finally, our approach can in principle be used in contexts in which adjustment costs prevent adjustments in a longer time horizon than we examine; for example, if hypothetically bunching persisted for 5 or 10 years after a kink disappeared, one could use our approach to measure elasticities and adjustment costs over that time horizon.

8 Conclusion

We investigate earnings adjustment frictions in the context of the Social Security Annual Earnings Test. We develop several related findings. First, we examine the speed of adjustment to the disappearance of convex kinks in the effective tax schedule. We document delays in adjustment, consistent with the existence of earnings adjustment frictions in the U.S. Despite the presence of frictions, we find that adjustment is rapid, as the vast majority of adjustment occurs within at most three years of budget set changes. We therefore interpret the observed frictions as reflecting a cost of making an immediate adjustment to policy. Since responses occur within a few years, this suggests that long-run elasticities are similar to those estimated in a medium-run time frame of a few years.

Next, we specify a model of employees' earnings adjustment consistent with these findings

and use it to estimate the earnings elasticity and the fixed adjustment cost. The model relies on the observation of the degree of bunching at two or more kinks of different sizes to estimate the two parameters of interest. We exploit the observations that all else equal, (1) the amount of bunching at an initial kink should be increasing in the elasticity and decreasing in the adjustment cost, and (2) the change in bunching when we vary the size of the kink should be decreasing in the adjustment cost. We extend our basic model to accommodate a number of additional features, including heterogeneity in the parameters and the possibility of adjustment frictions in the initial period. When we consider the change in bunching associated with the reduction in the AET benefit reduction rate from 50 percent to 33.33 percent for those older than NRA from 1989 to 1990, we estimate that the elasticity is 0.23 and the adjustment cost is \$152.08. We interpret our adjustment cost as meaning that bunching in this time frame can be predicted if individuals behaved as if they faced an adjustment cost of \$152.08. The results are typically in the same range with other samples or specifications. When we constrain adjustment costs to zero in the baseline specification, the elasticity we estimate in 1990 (0.39) is substantially (69 percent) larger, demonstrating the potential importance of taking account of adjustment costs.

Our estimates demonstrate the applicability of the methodology and the potential importance of allowing for adjustment costs when estimating elasticities. The finding that adjustment costs matter even among those flexible enough to bunch at the kink initially suggests the importance of taking frictions into account in other contexts. While the difference in the constrained and unconstrained elasticities we estimate is large in percentage terms (69 percent), the absolute difference between the elasticities (0.39 and 0.23) is moderate; however, as we show next, the modest adjustment cost does imply that even large changes in policy can have little or no immediate effect, suggesting that adjustment costs can make a dramatic difference to the analysis.

As a key example of the implications of our findings, consider whether decreases or increases in marginal tax rates can substantially affect earnings in the short run, as envisioned in many recent discussions of tax policy. If individuals face an adjustment cost in the short run, the short-run earnings response could be substantially attenuated. Indeed, from 1989 to 1990 in our data, the amount of normalized excess mass changes very little, suggesting that

short-run adjustment may be greatly attenuated in this context. To illustrate the potential for our estimation results to shed light on such issues, as described in detail in Appendix E.11 we use our estimates of the elasticity and adjustment cost to simulate the effect in our data of two policy changes. Reducing the marginal tax rate above the kink by 50 percentage points (as could be implied by a policy like eliminating the AET for 62-64 year-olds) would cause a large, 17.57 percent rise in earnings at the intensive margin. However, a less large change—in particular, any cut in the marginal tax rate of 14.67 percentage points or smaller above the exempt amount—would cause no change in earnings because the potential gains from adjusting are not large enough to outweigh the adjustment cost. While this result relies on the admittedly extreme assumption of a homogeneous elasticity and adjustment cost, it illustrates the principle that because the gains to relocation are second-order near the kink, even a modest adjustment cost of \$152.08 can prevent individuals’ adjustment—even following a substantial cut in marginal tax rates (up to 14.67 percentage points). Moreover, the lack of response predicted with a change of 14.67 percentage points makes sense in light of the empirical patterns we observe, in particular the very small decrease in bunching seen in the data from 1989 to 1990 when the marginal tax rate falls by slightly more (17 percentage points). It is important to emphasize that these estimates are not intended to be an exhaustive account of the implications of our estimates of adjustment costs for earnings responses to taxation under different policies; rather, they are intended simply to suggest the potential for adjustment costs to be important in the analysis of the earnings effects of tax changes in such contexts. This conclusion is robust to other assumptions that we discuss in the Appendix.

The analysis leaves open a number of avenues of further inquiry. First, dynamics are beyond the scope of this paper, but it would further enrich the framework to extend the analysis here to a model that incorporates dynamic features beyond the transition from a kink of one size to another that we have already modeled. We consider our framework for estimating elasticities and adjustment costs to be a natural first step in understanding estimation of these parameters (in the spirit of other static papers such as Saez 2010, Chetty *et al.*, 2009, 2011, 2012a,b, Chetty 2012, and Kleven and Waseem 2013), but incorporating further dynamic considerations is an important next step. The speed at which individuals

respond to changes in the AET, the nature of income effects, and the distinction between anticipated and unanticipated changes—all of which our work has begun exploring—are three of several possible pieces of evidence that could help in specifying a model of this sort (perhaps including a stochastic wage arrival process, and incorporating the benefits over time to adjustment at a given time). Note that in considering a model that incorporates more dynamic features, income effects (that drive some of the salient features of many dynamic models) should matter only far away from the kink; in the region of the kink, income effects are second-order. Even without developing such a model, we interpret our parameter estimates as implying that over the time period in question, individuals behave as if they face the elasticity and adjustment cost we estimate.

Second, further work distinguishing among the possible reasons for reaction to the AET remains an important issue (such as misperceptions, or behavioral factors (Diamond and Kőszegi 2003) that could lead the elderly to put particular weight on current OASI benefits). Relatedly, following most previous literature, we have treated the adjustment cost as a "black box," without modeling the process that underlies this cost, such as information acquisition or job search. Future research could model such processes and distinguish these explanations using data. Third, a dataset with information on hours worked (and a large enough sample to generate sufficient power) would allow us to determine whether these earnings responses correspond to responses of hours worked. Finally, studying earnings adjustment to other policies is a high priority.

References

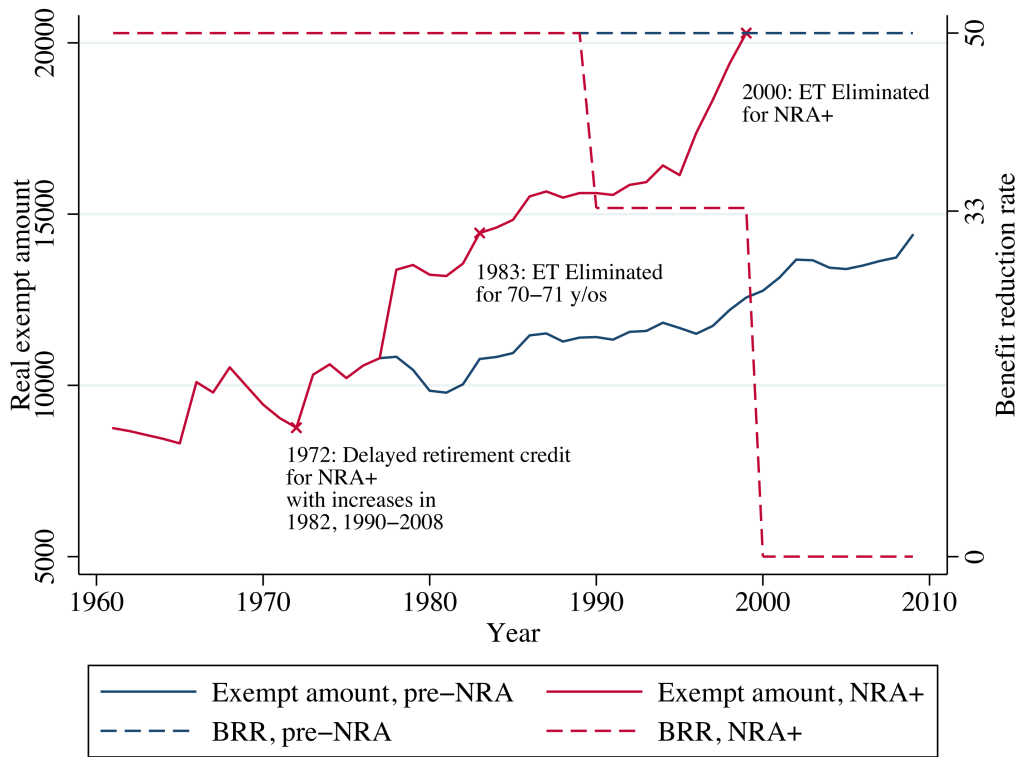
- Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock (2009), "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In *Producer Dynamics: New Evidence from Micro Data* (Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, eds.), 149–230, University of Chicago Press.
- Altonji, Joseph G. and Christina H. Paxson (1988), "Labor Supply Preferences, Hours Constraints, and Hours-Wage Trade-Offs," *Journal of Labor Economics*, 6, 254–276.
- Blundell, Richard, and Hilary Hoynes (2004), "Has 'In-Work' Benefit Reform Helped the Labor Market?" In *Seeking a Premier Economy: The Economic Effects of British Economic reforms, 1980–2000*, (David Card, Richard Blundell, and Richard B. Freeman, eds.), 411–59, University of Chicago Press.
- Brown, Jeffrey, Arie Kapteyn, Olivia Mitchell, and Teryn Mattox (2013), "Framing the Social Security Earnings Test," Wharton Pension Research Council Working Paper 2013-06.
- 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 (2012), "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica*, 80, 969–1018.
- 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." *Quarterly Journal of Economics*, 126, 749–804.
- Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber (2012), "Does Indivisible Labor Explain the Difference Between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities." In *NBER Macroeconomics Annual 2012* (Daron Acemoglu, Jonathan Parker, and Michael Woodford, eds.), vol. 27, University of Chicago Press.
- Chetty, Raj, Adam Looney, and Kory Kroft (2007), "Salience and Taxation: Theory and Evidence." NBER Working Paper No. 13330.
- Chetty, Raj, Adam Looney, and Kory Kroft (2009), "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99, 1145–1177.
- Chetty, Raj, John N. Friedman, and Emmanuel Saez (2012), "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." Forthcoming, *American Economic Review*.
- Coile, Courtney, and Jonathan Gruber (2001), "Social security incentives for retirement." In *Themes in the Economics of Aging* (David Wise, Ed.), 311–354, University of Chicago Press.
- Diamond, Peter, and Botond Köszegi (2003), "Quasi-Hyperbolic Discounting and Retirement."

Journal of Public Economics, 87, 1839-1872.

- 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 Wise, eds.), 437-473, University of Chicago Press.
- 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.
- Eissa, Nada, and Jeffrey B. Liebman (1996), "Labor Supply Response to the Earned Income Tax Credit." *Quarterly Journal of Economics*, 111, 605-37.
- 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, University of Chicago Press.
- Friedberg, Leora (2000), "The Labor Supply Effects of the Social Security Earnings Test." *Review of Economics and Statistics*, 82, 48-63.
- Gelber, Alexander, Damon Jones, and Daniel Sacks (2013), "Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test." NBER Working Paper 19491.
- Gelber, Alexander, Damon Jones, and Daniel Sacks (2013), "Extensive Margin Responses to the Social Security Earnings Test." Mimeo, Indiana University.
- Gruber, Jonathan (2013), *Public Finance and Public Policy*. New York: Worth Publishers.
- Gruber, Jonathan and Peter Orszag (2003), "Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt?" *National Tax Journal*, 56, 755-773.
- Hausman, Jerry A. (1981), "Labor Supply." In *How Taxes Affect Economic Behavior* (Henry J. Aaron and Joseph A. Pechman, Eds.), 27-71, Brookings Institution.
- Hoopes, Jeffrey, Daniel Reck, and Joel Slemrod (2013), "Taxpayer Search for Information: Implications for Rational Attention." University of Michigan Working Paper.
- Kleven, Henrik, Martin Knudsen, Claus Kreiner, Søren Pedersen, and Emmanuel Saez (2011), "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark." *Econometrica*, 79, 651-692.
- Kleven, Henrik, Camille Landais, Emmanuel Saez, and Esben Schultz (2012), "Taxation and International Migration of Top Earners: Evidence from the Foreigner Tax Scheme in Denmark." Forthcoming, *American Economic Review*.
- Kleven, Henrik and Mazhar Waseem (2013), "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics*, 128, 669-723.

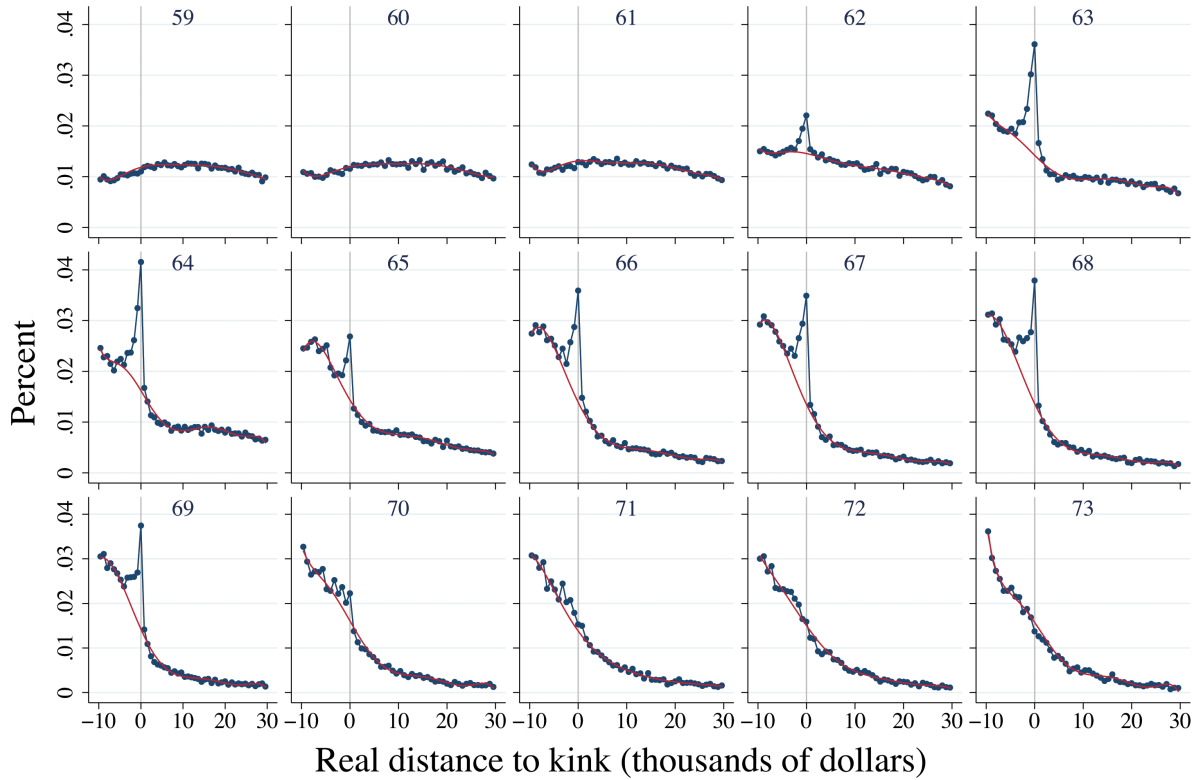
- Liebman, Jeffrey B., and Erzo F.P. Luttmer (2011), "Would People Behave Differently If They Better Understood Social Security? Evidence From a Field Experiment." NBER Working Paper 17287.
- Liebman, Jeffrey B., Erzo F.P. Luttmer, and David Seif (2009), "Labor Supply Responses to Marginal Social Security Benefits: Evidence from Discontinuities," *Journal of Public Economics*, 93, 1208-1223.
- McKinney, Kevin L. and Lars Villhuber (2008), "LEHD Infrastructure Files in the Census RDC - Overview Revision: 219." U.S. Census Bureau, LEHD Program.
- Meyer, Bruce D and Dan T. Rosenbaum (2001), "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics*, 116, 1063–1114.
- Moffitt, Robert (1990), "The Econometrics of Kinked Budget Constraints." *Journal of Economic Perspectives*, 4, 119–39.
- Saez, Emmanuel (2010), "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2, 180–212.
- Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz (2012), "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature*, 50, 3–50.
- Simon, Herbert A. (1955), "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*, 69, 99–118.
- Social Security Administration (2012a), *Annual Statistical Supplement*. Washington, D.C.
- Social Security Administration (2012b), *Social Security Handbook*. Washington, D.C.
- 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.

Figure 1: Key Earnings Test Rules, 1961-2009



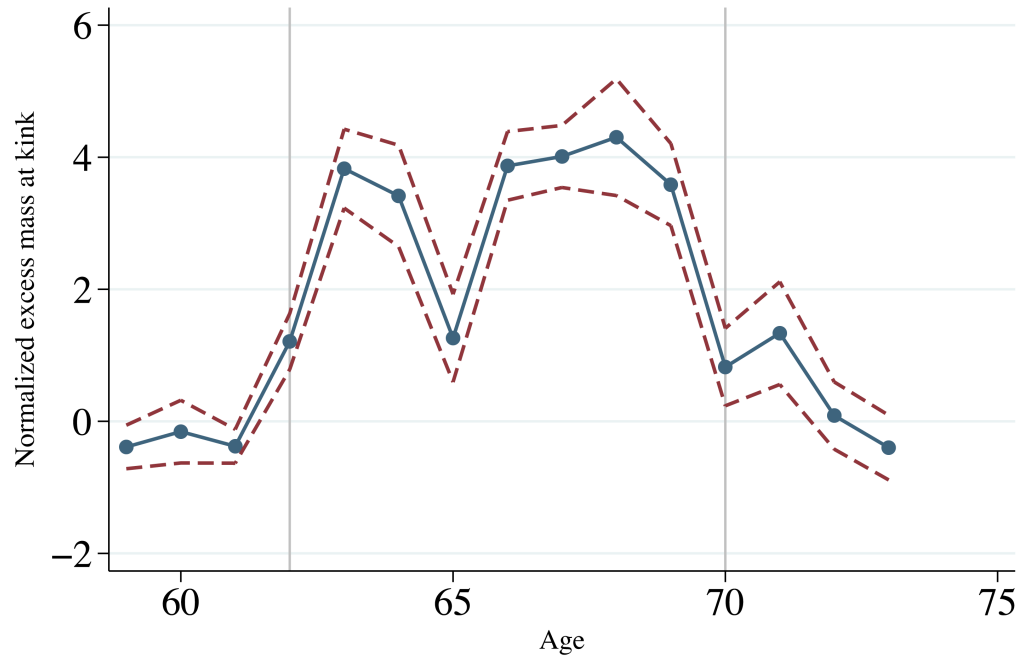
Note: The right vertical axis measures the benefit reduction rate in OASI payments for every dollar earned beyond the exempt amount. The left vertical axis measures the real value of the exempt amount over time.

Figure 2: Histograms of Earnings, 59-73-year-olds Claiming OASI by Age 65, 1990-1999



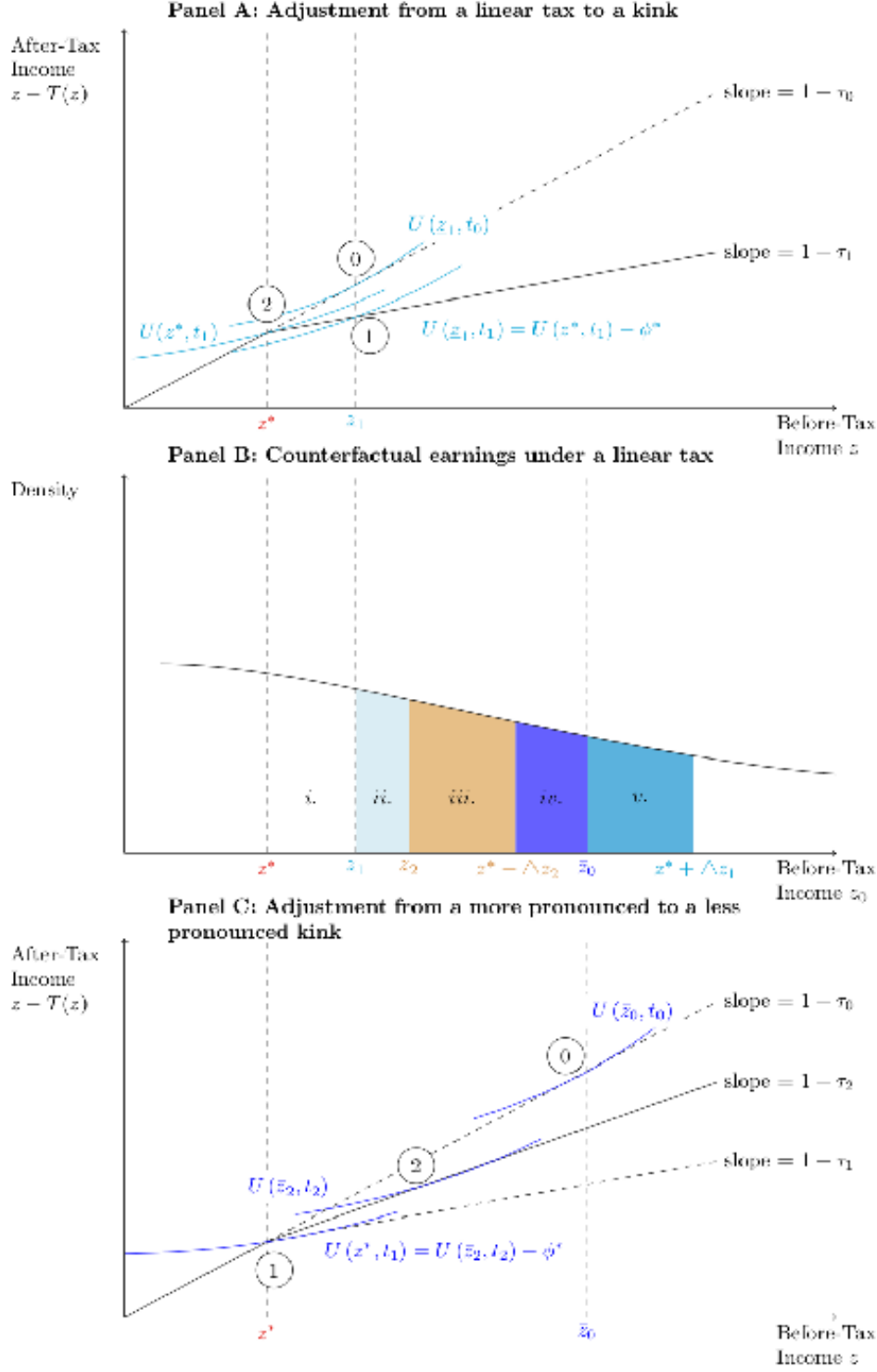
Notes: The sample is a one-percent random sample of all Social Security numbers, limited to individuals who claim OASI benefits by age 65. We exclude person-years with self-employment income or with zero non-self-employment earnings. The bin width is \$800. The earnings level zero, shown by the vertical lines, denotes the kink. The dots show the histograms using the raw data, and the polynomial curves show the estimated counterfactual densities estimated using data away from the kink.

Figure 3: Adjustment Across Ages: Normalized Excess Mass, 59-73-year-olds Claiming OASI by Age 65, 1990-1999



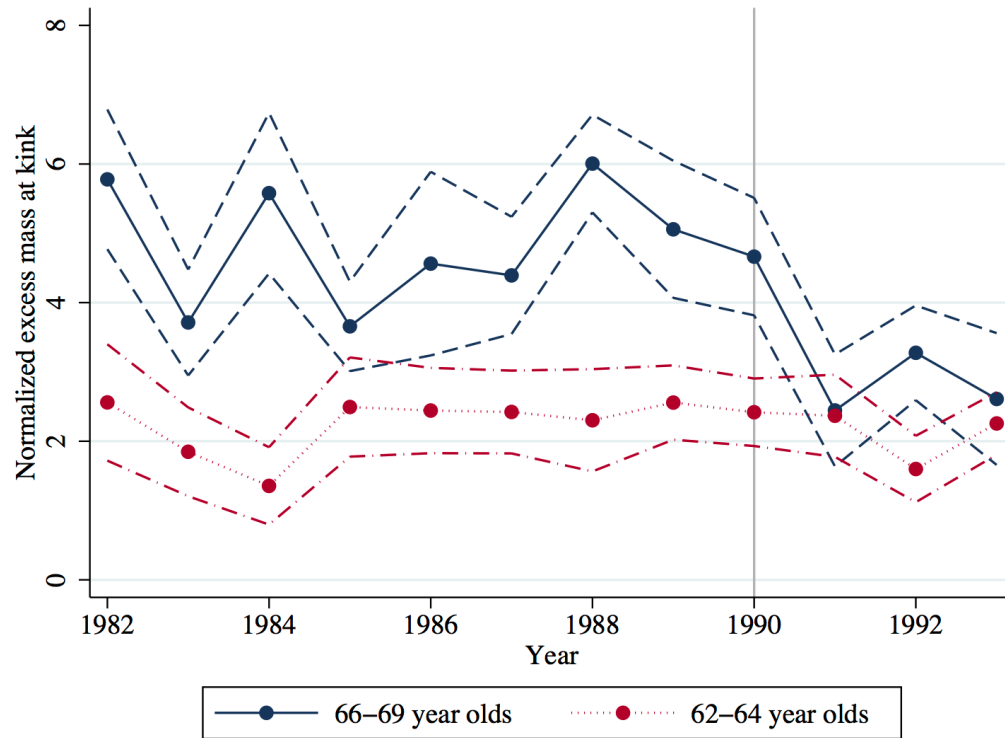
Note: The figure shows normalized bunching at the AET kink, calculated as described in the text. Dashed lines denote 95% confidence intervals. The vertical lines show the ages at which the AET first applies (62) and the age at which the AET ceases to apply (70). See other notes to Figure 2.

Figure 4: Bunching Responses to a Convex Kink, with Fixed Adjustment Costs



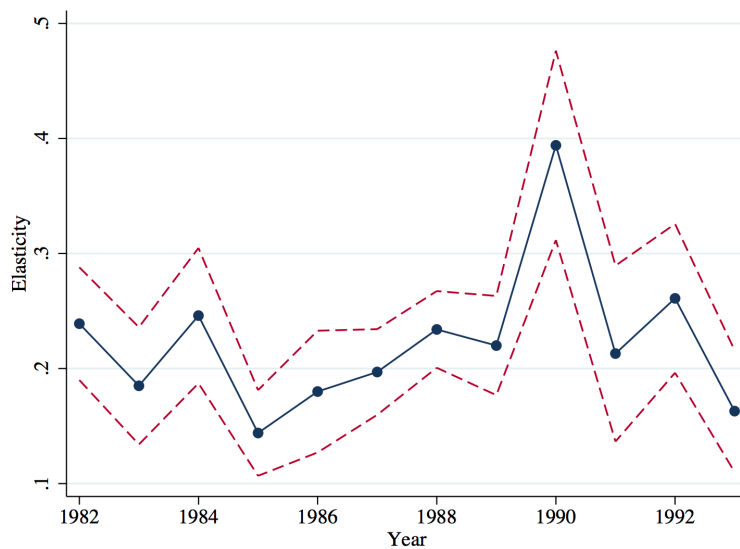
Note: See Section 6 for an explanation of the figures.

Figure 5: Comparison of Normalized Excess Mass Among 62-64 Year-Olds and 66-69 Year-Olds, 1982-1993



Note: the figure shows normalized bunching among 62-64 year-olds and 66-69 year-olds in each year from 1982 to 1993. Note the caveat that the 62-64 year-old group faces a notch at the exempt amount, as opposed to the kink faced by those 66-69. See other notes to Figure 2.

Figure 6: Elasticity Estimates by Year, Saez (2010) Method, 1982-1993



Note: The figure shows elasticities estimated using the Saez (2010) method, by year from 1982 to 1993, among 67-68 year-old OASI claimants. Dashed lines denote 95% confidence intervals. We use our methods for estimating normalized excess mass but use Saez' (2010) formula to calculate elasticities, under a constant density. This method yields the following formula:

$$\varepsilon = \left[\log \left(\frac{b}{z^*} + 1 \right) \right] / \left[\log \left(\frac{1 - \tau_0}{1 - \tau_1} \right) \right].$$

Table 1: Summary Statistics, Social Security Administration Master Earnings File

Ages 62-69	
Mean Earnings	28,892.63 (78,842.99)
10th Percentile	1,193.64
25th Percentile	5,887.75
50th Percentile	14,555.56
75th Percentile	35,073.00
90th Percentile	64,647.40
Fraction Male	0.50
Observations	376,431

Note: The data are taken from a one percent random sample of the SSA Master Earnings File and Master Beneficiary Record. The data cover those in 1990-1999 who are aged 62-69, claim by age 65, do not report self-employment earnings, and have positive earnings. Earnings are expressed in 2010 dollars. Numbers in parentheses are standard deviations.

Table 2: Estimates of Elasticity and Adjustment Cost

	(1)	(2)	(3)	(4)
	ε	ϕ	$\varepsilon \phi = 0$	
			1990	1989
Baseline:	0.23 [0.21, 0.30]***	\$152.08 [49.54, 710.48]***	0.39 [0.33, 0.48]***	0.22 [0.18, 0.25]***
Nonparametric Density:	0.26 [0.23, 0.37]***	\$174.60 [52.27, 940.93]***	0.43 [0.36, 0.54]***	0.24 [0.20, 0.28]***
Benefit Enhancement :	0.39 [0.34, 0.82]***	\$81.52 [19.79, 2558.85]***	0.59 [0.49, 0.73]***	0.37 [0.30, 0.43]***
Excluding FICA :	0.32 [0.28, 0.42]***	\$129.14 [34.18, 884.25]***	0.50 [0.42, 0.62]***	0.30 [0.25, 0.35]***
Bandwidth = \$500:	0.25 [0.21, 0.31]***	\$90.65 [6.04, 339.56]***	0.38 [0.30, 0.50]***	0.24 [0.19, 0.30]***

Note: The table shows estimates of the elasticity and adjustment cost using the method described in Section 6.2. We report bootstrapped 95 percent confidence intervals shown in parentheses. We investigate the 1990 reduction in the AET BRR from 50 percent to 33.33 percent. The baseline specification assumes a locally uniform density in the neighborhood of the kink, calculates the effective MTR by including the effects of the AET BRR and federal and state income and FICA taxes, uses data from 1989 and 1990, and calculates bunching using a bin width of \$800. Alternative specifications deviate from the baseline as noted. The nonparametric density specification uses the age 72 earnings distribution as the counterfactual density for 68-69 year olds. The estimates that include benefit enhancement use effective marginal tax rates due to the AET based on the authors' calculations relying on Coile and Gruber (2001) (assuming that individuals are considering earning just enough to trigger benefit enhancement). This translates the BRR before and after the 1990 policy change to 36% and 24%, respectively. Columns (1) and (2) report joint estimates with $\phi \geq 0$ imposed (consistent with theory, as described in the Appendix), while Columns (3) and (4) impose the restriction $\phi = 0$. The constrained estimate in Column (3) only uses data from 1990, whereas that in Column (4) uses only data from 1989. In some specifications we estimate very wide confidence intervals on the adjustment cost (but note that we always estimate tight confidence intervals on the elasticity). This is because in some of our bootstrap samples, there is more bunching in 1990 than in 1989. The model cannot generate this pattern, but it generates the best possible fit by making bunching in 1990 the same as in 1989, and this requires a very high adjustment cost. The confidence intervals never contain zero because in none of our bootstrap samples do we observe more than proportional adjustment to the change in the jump at the kink. *** indicates that the left endpoint of the 99 percent confidence interval is greater than zero.

Table 3: Estimates of Elasticity and Adjustment Cost Using Disappearance of Kink at Age 70

	(1)	(2)	(3)
	ε	ϕ	$\varepsilon \phi = 0, \text{ Age } 69$
Baseline	0.28 [0.25, 0.34]***	\$89.93 [25.51, 246.79]***	0.25 [0.22, 0.30]***
Nonparametric Density	0.31 [0.28, 0.38]***	105.19 [28.62, 266.43]***	0.29 [0.24, 0.34]***
Benefit Enhancement	0.42 [0.37, 0.65]***	58.01 [16.70, 2094.40]***	0.38 [0.33, 0.45]***
Excluding FICA	0.36 [0.31, 0.46]***	\$83.01 [23.79, 304.56]***	0.33 [0.28, 0.38]***
Bandwidth = \$500	0.20 [0.16, 0.48]***	\$17.05 [4.13, 1542.70]***	0.20 [0.16, 0.23]***
68-70 year-olds	0.29 [0.26, 0.61]***	\$42.13 [4.78, 1992.64]***	0.28 [0.25, 0.32]***
69, 71 year-olds	0.30 [0.25, 0.95]***	\$154.72 [34.11, 1968.13]***	0.25 [0.22, 0.30]***

Note: The table estimates elasticities and adjustment costs using the removal of the AET at age 70, using data on 69-71 year-olds in 1990-1999. We cannot estimate the constrained elasticity using only data on age 70 because the benefit reduction rate is zero at that age. We show these results in the Appendix, rather than the main text, because the estimates of bunching at age 70 are potentially affected by the coarse measure of age that we use, as explained in the Appendix. To address this issue, we use both age 70 and age 71 in estimating these results. To further address this issue, in the final row of the table, we use only ages 69 and 71, which shows very similar results—this is unsurprising because Figure 3 shows that normalized excess mass is similar at ages 70 and 71. The row labeled "68-70 year-olds" uses data from ages within this range. See also notes to Table 2.

A Appendix: Additional Features of Annual Earnings Test (for online publication)¹⁷

When current benefits are lost to the AET, future scheduled benefits are increased in some circumstances. This is sometimes called "benefit enhancement." As we describe, for workers NRA or older in the pre-2000 period (when they faced the AET), benefit enhancement attenuates the effective AET BRR for individuals considering earning enough to trigger the benefit enhancement, but it does not attenuate the effective AET BRR for those considering earning less than this amount.

The benefit enhancement rules have varied over time, and they depend on whether the beneficiary is older or younger than the NRA. Prior to 1972, there was no benefit enhancement for people aged NRA and older. In these years, the AET represented a pure loss in benefits for those NRA and older (equivalent to a pure tax). For beneficiaries NRA and older, a one percent Delayed Retirement Credit (DRC) was introduced in 1972. The DRC was intended to compensate beneficiaries who delayed claiming beyond age 65, but they also apply to earnings lost to the AET. For individuals older than the NRA, benefits are increased 1/12 of 1 percent for each month between ages 65 and 72 for which no benefits received after 1972 (Social Security, 2012, Table 2.A.20).¹⁸

This language indicates that each month's worth of foregone benefits—either because of delayed claiming or because of the AET—results in increased future benefits. A beneficiary has to forego an *entire* month of benefits in order to receive the DRC; if, for example, she earns slightly over the exempt amount and loses only a small amount of benefits to the AET, then her future benefits are not adjusted. Thus, the DRC provides no marginal relief from the AET for a claimant who is considering earning near the exempt amount: no benefit enhancement occurs when she earns a marginal dollar at or near the AET earnings threshold. Meanwhile, if she earns enough to forego an entire month's worth of benefits (but not when a smaller amount of benefits is lost due to the AET), future benefits are increased by 1/12 of 1 percent.

As a result of these rules, future benefits are enhanced when the individual's yearly earnings are over $z^* + (MB/\tau)$, where z^* is the annual exempt amount, MB is the monthly benefit, and τ is the AET benefit reduction rate.¹⁹ For example, with a typical monthly benefit of \$1,000 and a benefit reduction rate of 33.33 percent, benefit enhancement occurs when the individual's yearly earnings are \$3,000 ($=\$1000/0.3333$) above the exempt amount. Benefit enhancement corresponding to one more month of reduced earnings occurs once annual earnings reaches \$6,000 above the exempt amount, and so forth. Thus, benefit enhancement is only relevant to an individual considering earning substantially in excess of the exempt amount and is therefore not relevant to marginal earnings decisions at the

¹⁷This section is based on table 2.A.20 of the Annual Statistical Supplement of the Social Security bulletin, as well as extensive email correspondence with numerous officials at the Social Security Administration.

¹⁸The size of the DRC was increased to three percent per year in 1982, and then increased steadily throughout the 1990s, reaching eight percent for each year of foregone benefits in 2008. Starting in 1983, benefit enhancement only applied through age 69.

¹⁹Another month's benefit enhancement would occur if the individual earns more than $z^* + (2MB/\tau)$; a third month's benefit enhancement would occur if she earns more than $z^* + (3MB/\tau)$; and so on. Note that this creates 12 notches in the budget set, the final one at $z^* + (12MB/\tau)$.

exempt amount.

This theoretical presumption is consistent with suggestive evidence that indicates little systematic bunching reaction to changes in the DRC and that mean age at death is smooth near the exempt amount. Appendix Figure F.14 shows that there is no sharp change in the amount of bunching around the increases in the Delayed Retirement Credit in 1972 or 1982. We consider this suggestive—but not definitive—evidence of little discernible reaction to policy changes in benefit enhancement (particularly in light of our other results suggesting fast adjustment). A general downward trend in the amount of bunching is discernible in the 1990s—with the notable exception of a number of years, including 1995—which is coincident in the rise in the DRC through this period. However, we cannot conclusively attribute this potential trend to the influence of the DRC, as it could be due to other factors that changed over this period.

The AET is implemented in a number of stages. First, SSA must determine that a claimant is expected to exceed the exempt amount, or that she has already done so. Claimants can notify SSA in advance if they expect to exceed the exempt amount, or they can report their earnings *ex post facto* at any point in the year. In addition, SSA uses W-2 records at the end of the calendar year to determine if the AET threshold has been crossed (for those who have a W-2). Second, SSA withholds OASI benefits in monthly increments, until enough benefits have been withheld to cover the AET penalty amount. For example, assume an individual aged 66 with a monthly benefit of \$1,200 earns \$1,800 dollars beyond the AET exempt amount in 1992, when the benefit reduction rate was 0.3333. This individual should receive a yearly benefit reduction of $\$1,800 \times 0.3333 = \600 . SSA withholds an entire month's check, \$1,200, in order to collect the \$600. Finally, at the end of the calendar year, SSA refunds any overwithheld benefits. In the same example, at the end of the year SSA would return \$600 in overwithheld benefits. Importantly, the DRC is not applied to future benefits in this case—less than a month's worth of benefits, \$600, was ultimately collected by SSA, after factoring overwithholding and refunds.

After considering both withholding and refunds, the AET is ultimately applied at a yearly level—much in the same way that the Earned Income Tax Credit (EITC) is applied at the annual level but the receipt of the credit depends on one's withholding patterns and when the income tax return is filed. Note that although the AET withholds benefits at the monthly level, the AET is generally applied based on *annual* calendar year earnings—the object we observe in our data.

In sum, for people NRA and older, the AET effectively acts as a kink for those earning close enough to the exempt amount, and benefit enhancement does not attenuate the marginal work disincentives associated with the AET in this range of earnings. However, the DRC is relevant to an individual considering earning enough to reduce her OASI benefit by at least a month's worth (*i.e.* at least $z^* + (MB/\tau)$). Empirically, we find that limiting the sample to those with substantial OASI benefits—for whom this earnings level is several thousand dollars above the AET exempt amount, and for whom the notch created by the DRC is therefore less relevant—yields very similar results to those we have shown. Our empirical specification alternatively assumes that benefit enhancement does not affect the AET implicit marginal tax rate (or does), and we find similar patterns in both specifications.

Note that in our empirical estimation, the region we "dummy out" near the kink z^* is $[z^* - 2800, z^* + 2800]$. Thus, if all of the "bunchers" arrived at the kink from initial

earnings levels between z^* and $z^* + 2800$, we could find zero bunching despite substantial actual bunching. However, the earnings densities clearly show that the polynomial does not substantially overpredict bunching in the region above the exempt amount. Moreover, the densities also show no evidence of bunching near notches in the budget set created by the DRC. We alternatively use a bandwidth of \$500, corresponding to a "dummied out" region $[z^* - 1750, z^* + 1750]$, and find similar results.

Individuals younger than NRA are subject to different rules for benefit adjustment, called "actuarial adjustment." The rule for this younger group was introduced in the legislation allowing people younger than 65 to claim early benefits (in 1956 for women and in 1961 for men). For those younger than NRA, future benefits are reduced 5/9 of 1 percent for each month under age 65 in which an individual claims benefits (Social Security, 2012, Table 2.A.20). This implies that if a beneficiary has any income withheld under the AET in a given month, then she receives a full benefit enhancement for that month. On the other hand, if a beneficiary does not have any income withheld under the AET in a given month, then she receives no benefit enhancement for that month. When considering lifetime OASI benefits, this creates a notch at the exempt amount—a discontinuous increase in future benefits when moving from just under the exempt amount to just over it—creating incentives to bunch just above the exempt amount in order to receive the monthly (or yearly) benefit adjustment. We find no evidence for such behavior. Because benefit enhancement occurs at exactly the same earnings level that the AET begins to apply for those younger than NRA, we focus on the group NRA and older in our estimates of elasticities and adjustment costs.

Formally, the number of months' worth of benefit enhancement enjoyed by OASI recipients is therefore $\text{floor}(\tau \cdot (z - z^*)/MB)$ for those NRA and older, and $\text{ceiling}(\tau \cdot (z - z^*)/MB)$ for those younger than NRA.

Benefit enhancement is actuarially fair if the net present value of the benefit enhancement equals the benefits lost due to the AET. The actuarial adjustment is approximately actuarially fair *in the sense that delaying OASI claiming an extra year is approximately actuarially fair*; however, this does *not* imply that benefit enhancement is actuarially fair when an additional dollar of benefits is withheld due to the AET. For example, actuarial adjustment is not actuarially fair for (among others) those with positive OASI benefits considering earning an additional amount above the AET exempt amount, because this does not result in additional benefit enhancement. Similar considerations apply to the DRC: additional marginal increments of earnings are *not* compensated through benefit enhancement (except in the case when an individual goes from earning just under to just over $z^* + (MB/\tau)$ (or one of the other 12 thresholds)).

The AET applies to an individual's earnings; spouses' earnings do not count in the earnings total to which the AET is applied. For a retired worker (*i.e.* primary) beneficiary whose spouse collects spousal benefits, the AET reduces the family's OASI benefit by the amounts we have described. The family benefit is also reduced when the spouse (separately) earns more than the AET threshold. For a retired worker beneficiary whose spouse is collecting benefits on his or her own earnings record, the AET reduces the retired worker beneficiary's benefits by the amounts described while not affecting the spouse's benefits. Thus, following previous literature (*e.g.* Friedberg 1998, 2000), we model the AET as creating the MTRs associated with the BRRs described, because the AET reduces family benefits by these amounts (all else equal). Our data do not contain the information necessary to link spouses

(except when one spouse is claiming OASI benefits on the other spouse's record).

It is also worth noting how the actuarial adjustment and DRC interact with incentives for claiming OASI. Under the actuarial adjustment, the full benefit enhancement occurs when the individual earnings over the threshold level. Thus, the individual could in principle claim OASI; earn just over this threshold level; collect nearly her entire OASI benefit in this year (since the AET only reduces current OASI benefits at the margin); and later enjoy full benefit enhancement. This illustrates the more general point that it can be in an individual's interest to claim OASI even if the individual faces the AET. More generally, for individuals for whom the AET reduces OASI benefits sufficiently little, and for whom current OASI benefits are sufficiently important, it can be in their interest to claim OASI even if they face the AET. Appendix Figure F.15 shows that among the sample of individuals who have not claimed by year t , the hazard of claiming at year $t + 1$ is smooth near the kink, indicating no evidence that claimants come disproportionately from close to or far from the kink.

In addition to the AET threshold we investigate, until 1972 there was a second, higher earnings threshold over which the benefit reduction rate was 100 percent (Social Security Annual Statistical Supplement 2012). The second threshold is well above the first threshold, ranging from 25 percent to 80 percent higher depending on the year.

The exempt amount has not been a "focal" earnings level—such as \$1,000, \$5000, or \$10,000—that could lead to bunching at the exempt amount even in the absence of AET. Indeed, in our main period of study we find no evidence of bunching at the exempt amount among those younger than the ages to which the AET applies.

As noted above, the AET also potentially creates other distortions not explicitly modeled, including: a slight notch for those NRA age and older every time an entire month's worth of benefits are lost due to the Delayed Retirement Credit; an additional non-convex kink in the budget constraint at the point at which OASI benefits are fully phased out; and a notch for those younger than the NRA for every month of withheld benefits that triggers the actuarial adjustment described above. However, in the case of those NRA and older that we focus on, these incentives are not likely to be relevant for potential bunchers and do not appear to be empirically relevant (as we discuss elsewhere). For these reasons, we abstract from these additional features but discuss such incentives when we present our empirical evidence.

B Appendix: Procedure for Estimating Normalized Excess Mass (for online publication)

In order to estimate normalized excess mass, we use the following procedure. For each earnings bin z_i , we calculate p_i , the proportion of all people with earnings in the range $[z_i - k/2, z_i + k/2)$ (in a given time period and for a given age group). For example, underlying the first panel in Figure 2 is the probability p_i of earning in various bins z_i for 62 year-olds in the 1990-1999 period. The earnings bins are normalized by distance-to-kink, so that for $z_i = 0$, p_i is the fraction of people with earnings in the range $[-k/2, k/2)$. To estimate

bunching, we assume that p_i can be written as

$$p_i = \sum_{d=0}^D \beta_d (z_i)^d + \sum_{n=-k}^k \gamma_n 1\{z_i = n\delta\} + \varepsilon_i \quad (\text{B.1})$$

and run this regression. This equation expresses the earnings distribution as a degree D polynomial, plus a set of indicators for each bin within $k\delta$ of the kink, where δ is the binwidth. In our empirical application, we choose $D = 7$, $\delta = 800$ and $k = 3$ (so that seven bins are excluded from the polynomial estimation, including the bin centered at the kink). We show that our results to alternative choices of D , δ , and k .

Our measure of excess mass is $\hat{EM} = \sum_{n=-k}^k \hat{\gamma}_n$, the estimated excess probability of locating at the kink (relative to the polynomial term). This measure depends on the counterfactual density near the kink, so to obtain a measure of excess mass that is comparable at the kink, we scale by the predicted density that would obtain if there were no bunching. This is just the constant term in the polynomial, since the z_i is distance to zero. So our estimate of normalized excess mass is

$$\hat{B} = \frac{\hat{EM}}{\hat{\beta}_0}. \quad (\text{B.2})$$

We consider two approaches for constructing standard errors. First, from Equation (B.2), it is straightforward to apply the Delta method. Second, we employ the parametric bootstrap procedure of Chetty, Friedman, Olsen and Pistaferri (2011). This bootstrap draws with replacement from the estimated distribution of errors ε_i from Equation (B.1). For each set of draws, we get a new value of p_i and use these new values to re-estimate B . The standard deviation across draws of B is our measure of the standard error \hat{B} . In practice these two procedures produced extremely similar results, so we only report standard errors from the bootstrap.

C Appendix: Social Security Data and Tax Calculations (for online publication)

Our data come from the Social Security Master Earnings File (MEF), which is described more extensively in Song and Manchester (2007). The MEF is a longitudinal history of Social Security taxable earnings for all Social Security Numbers (SSNs) in the U.S. Our data are a one percent random sample of SSNs; we randomly extract SSNs from the database and follow each of these individuals over the full time period. The AET is based on earnings as measured in this dataset. Prior to 1978, the data have information on annual FICA earnings; since 1978, the data have information on uncapped wage compensation. Before 1978, the data do not clearly distinguish between earnings from self-employment and non-self-employment earnings, but we are able to distinguish them in the data starting in 1978. The data also contain information on date of birth, date of death, and sex.

We supplement the MEF with information from the Master Beneficiary Record (MBR) file, which contains data on the day, month, and year that people began to claim Social Security (and other variables). The majority of workers excluded from OASDI coverage are

in four main categories: (1) federal civilian employees hired before January 1, 1984; (2) agricultural workers and domestic workers whose earnings do not meet certain minimum requirements; (3) individuals with very low net earnings from self-employment (generally less than \$400 per year); and (4) employees of several state and local governments. However, civil service and other government workers are covered by Medicare and are therefore present in the MBR.

Information on AET parameters is from table 2.A.20 and 2.A.29 of the Annual Statistical Supplement to the Social Security Bulletin. Friedberg (1998, 2000) provides a thorough description of these rules. All dollar amounts are deflated to 2010 dollars using the CPI-U.

In 1983-1999, the AET is assessed on earnings until the month in which the individual turns age 70. For simplicity, in our baseline sample we measure age as calendar year minus year of birth. Thus, if an individual turns age 70 later in the year—in the extreme case, on December 31—she will have had an incentive to bunch at the kink during nearly the entire year when she is classified as age 70 in our data. As a result, her yearly earnings may appear to be located at or near the kink even though she is bunching at the kink applicable to 69-year-olds through almost all of the calendar year over which her earnings are observed. However, the figure shows that significant bunching occurs at age 71, which cannot be due to this coarse measure of birth dates. Thus, the results do show a delay in complete adjustment. We have also found substantial and significant ($p < 0.01$) bunching at age 70 among those born in January, who no longer face the AET immediately in January of the year they turn 70 and therefore should not show bunching at this age in the absence of adjustment frictions. Likewise, we find a spike in mean earnings growth from age 70 to age 71 among those born in January. In our sample period, the AET applied to ages 62-71 before 1983, and it applied to ages under NRA in 2000 and after. In these time periods, examining only those born in January also shows a delay in responding to the removal of the AET.

Since 1978, the earnings test has been assessed on yearly earnings, implying that we analyze the appropriate time period, *i.e.* earnings in a calendar year. Prior to 1978, the earnings test was assessed on quarterly earnings. While there is likely some error in measuring the amount of bunching pre-1978, we believe that this is not a major issue: the patterns of bunching in the pre-1978 period are visually clear and appear unlikely to be changed in a qualitative sense by an examination of quarterly data. Moreover, Figure F.14 shows that the amount of bunching falls from 1977 to 1978 and subsequent years, rather than rising as we might expect if we hypothetically measured bunching more accurately starting in 1978.

Using TAXSIM and the Statistics of Income individual tax return files, we calculated the mean of the sum of federal and state marginal income and FICA tax rates for people with positive Social Security benefits and earnings within \$2000 of the kink, in the same years as the data we examine. For example, when we examine data from 1989 and 1990, we calculate marginal tax rates in these years. The results are not sensitive to other such choices. We have found that incentives, including income tax rates, are smooth on average around the AET convex kink.

D Appendix: Longitudinal Employer Household Dynamics (for online publication)

We use the Longitudinal Employer Household Dynamics (LEHD) dataset, which contains wage data available from state-level unemployment insurance (UI) programs. These data measure uncapped quarterly earnings for employees covered by state unemployment insurance systems, estimated to cover over 95 percent of private sector employment. Although coverage laws vary slightly from state to state, UI programs do not cover federal employees, the self-employed, and many agricultural workers, domestic workers, churches, nonprofits, and state and local government employees. We examine a 20 percent random sample of the original LEHD file, as this was the largest amount of data that our available server space could handle.

These administrative earnings records are linked across quarters to create individual work histories. In addition to earnings, information on gender and date of birth are available. We select data from 1990-1999. During this period, the AET explicit benefit reduction rate was constant. 1990-1999 also represent natural years to investigate because large sample sizes are not available in the LEHD prior to 1990. When we include other years and age groups in the LEHD sample, we find similar results to those reported here. Note that the population we investigate is not constant over this period, because (among other reasons) an increasingly broad set of states is included in the LEHD over time. Data are available on 13 states in 1990, climbing to 28 states by 1999. In a given quarter, we include in our sample all states whose data are available. Holding the sample constant yields very similar results.

The LEHD lacks information on whether a given individual is claiming OASI. Nonetheless, the ultimate importance of this shortcoming is limited. In our SSA data, 97 percent of people claim by age 69, so it is a safe assumption that the great majority of the individuals observed in the LEHD data of the ages we are interested in (primarily ages 69 and 70) have claimed OASI. The magnitude of the bunching we observe is likely to be slightly understated relative to the magnitude we would measure in the population of OASI claimants, as the results include non-claimants in the sample. However, our primary interest concerns the patterns of responses to the AET across ages and over time, which prove to be visually and statistically clear in the LEHD.

E Appendix: Model of Earnings Response (for online publication)

E.1 Baseline Model

We start with a baseline, frictionless model of earnings, following Saez (2010). We briefly sketch the key features of this model for comparison to our model with a fixed cost of adjustment. In Saez (2010), individuals maximize utility over consumption, c , and costly earnings, z :

$$u(c, z; n)$$

Heterogeneity is parameterized by an "ability" parameter n , which is distributed according to the smooth CDF $F(\cdot)$. Individuals maximize utility subject to the following budget

constraint:

$$c = (1 - \tau) z + R$$

where R is virtual income. This leads to the first order condition:

$$-\frac{u_z(c, z; n)}{u_c(c, z; n)} = (1 - \tau),$$

which implicitly defines an earnings supply function: $z(1 - \tau, R, n)$.

When necessary, we will use a quasi-linear and isoelastic utility function:

$$u(c, z; n) = c - \frac{n}{1 + 1/\varepsilon} \left(\frac{z}{n}\right)^{1+1/\varepsilon}$$

Under this assumption, the first order condition simplifies to:

$$(1 - \tau) - \left(\frac{z}{n}\right)^{\frac{1}{\varepsilon}} = 0,$$

which implies this earnings supply function:

$$z = n(1 - \tau)^\varepsilon.$$

E.2 Linear Tax Schedule

Consider first a linear tax schedule with a constant marginal tax rate τ_0 . Observe that with a smooth distribution of skills n , we have a smooth distribution of earnings that is monotonic in skill, provided we make the typical Spence-Mirrlees assumption. Let $H_0(\cdot)$ denote the cumulative distribution function (CDF) of earnings under the constant marginal tax rate, and let $h_0(\cdot) = H'_0(\cdot)$ denote the density of this distribution. Under quasilinear utility, we have:

$$H_0(z) = F\left(\frac{z}{(1 - \tau_0)^\varepsilon}\right).$$

Define $H_1(\cdot)$ and $h_1(\cdot)$ as the smooth CDF and density of earnings under a higher, constant marginal tax rate τ_1 ; H_1 is defined similarly as a function of τ_1 .

E.3 Kinked Tax Schedule

Now consider a piecewise linear tax schedule with a convex kink: the marginal tax rate below earnings level z^* is τ_0 , and the marginal tax rate above z^* is $\tau_1 > \tau_0$, as illustrated in Appendix Figure F.1 and described in Section 3. Given the tax schedule, individuals bunch at the kink point z^* ; as explained in Saez (2010), the realized density in earnings has an excess mass at z^* . Denote the realized distribution of earnings once the kink has been introduced at z^* as $H(\cdot)$:

$$H(z) = \begin{cases} H_0(z) & \text{if } z < z^* \\ H_1(z) & \text{if } z \geq z^* \end{cases}$$

Denote the density of this realized distribution as $h(\cdot) = H'(\cdot)$. In general there is now a discrete jump in the earnings density at z^* :

$$h(z) = \begin{cases} h_0(z) & \text{if } z < z^* \\ h_1(z) & \text{if } z > z^* \end{cases}$$

The share of people who relocate to the kink is:

$$B = \int_{z^*}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta$$

These "bunchers" are those whose *ex ante* earnings lie in the range $[z^*, z^* + \Delta z_1^*]$, who are induced to locate at the kink by the rise in the MTR above the kink point. For relatively small changes in the tax rate, we can relate the elasticity of earnings with respect to the net-of-tax rate to the earnings change Δz_1^* for the individual with the highest *ex ante* earnings who bunches *ex post*:

$$\varepsilon = \frac{\Delta z_1^*/z^*}{d\tau_1/(1 - \tau_0)}$$

where $d\tau_1 = \tau_1 - \tau_0$.²⁰

E.4 Fixed Adjustment Costs

We now extend the model to include a fixed cost of adjusting earnings. We assume that the adjustment cost reflects a disutility of ϕ^* of increasing or decreasing earnings from some initial earnings level. We begin by analyzing the response to a change in the marginal tax rate from τ_0 to τ_1 , where the tax schedule is linear in both cases, in order to build intuition for the case with a kinked budget set. We assume that following a change in tax rates from τ_0 to τ_1 , the gain (absent adjustment costs) to reoptimizing is increasing in n . In general, this requires that the size of the optimal earnings adjustment increases in n at a rate faster than the decrease in the marginal utility of consumption.²¹ This is true, for example, if utility is quasilinear.

If the gain in utility is monotonically increasing in initial earnings, and the cost of adjustment is fixed, there exists a unique level of initial earnings at which the agent is indifferent between adjusting and staying at the initial earnings level. We formally state the implications in the following result:

Remark 1 (Linear Tax Change and Adjustment Costs)

²⁰This formula holds if there is a single elasticity ε in the population. Under heterogeneity, the method returns $\bar{\varepsilon}$, the average elasticity among bunchers. We investigate cases with heterogeneity below.

²¹To see this, note that the utility gain from reoptimizing is $u((1 - \tau_1)z_1 + R_1, z_1; n) - u((1 - \tau_1)z_0 + R_1, z_0; n) \approx u_c \cdot (1 - \tau_1)[z_1 - z_0] + u_z \cdot [z_1 - z_0] = u_c \cdot (\tau_1 - \tau_0)[z_0 - z_1]$, where in the first expression, we have used a first-order approximation for utility at $((1 - \tau_0)z_0 + R_0, z_0)$ and in the second expression we have used the first order condition $u_z = -u_c(1 - \tau_0)$. The gain in utility is approximately equal to an expression that depends on the marginal utility of consumption, the change in tax rates, and the size of the earnings adjustment. The first term, u_c , is decreasing as n (and therefore initial earnings z_0) increases. Thus, in order for the gain in utility to be increasing in n , we need the size of earnings adjustment $[z_0 - z_1]$ to increase at a rate that dominates.

After a change in linear tax rates from τ_0 to τ_1 , if there is a constant adjustment cost of ϕ^* and the size of the optimal earnings adjustment increases in n at a rate faster than the decrease in the marginal utility of consumption, then there is a unique threshold of initial earnings, $z_{0,\phi}$, above which all individuals will adjust their earnings in response to the tax change. Those initially locating below the threshold will not adjust. The threshold level of earnings satisfies the following identity:

$$u((1 - \tau_1) z_{1,\phi} + R_1, z_{1,\phi}) - u((1 - \tau_1) z_{0,\phi} + R_1, z_{0,\phi}) \equiv \phi^*$$

where $z_{1,\phi}$ is the *ex post* earnings level of the individual who initially locates at $z_{0,\phi}$. In other words, at the threshold level, the gain in u from adjusting earnings is exactly equal to the adjustment cost ϕ^* .

Next, consider choices in the presence of adjustment costs on a budget set with a convex kink. Consider again an initial linear tax schedule with marginal tax rate τ_0 . Now, introduce a higher MTR $\tau_1 > \tau_0$ for earnings above z^* . We again assume that the gain to reoptimizing is increasing in initial earnings over the range $[z^*, z^* + \Delta z^*]$. Using the same logic as above—the gain in utility is monotonically increasing in initial earnings, and the cost of adjustment is fixed—there exists a unique level of initial earnings at which the agent is indifferent between adjusting and staying at the initial earnings level. Thus, we have the following result:

Remark 2 (Non-Linear Tax and Adjustment Costs)

When a kink is introduced in the budget set (i.e. a jump in marginal tax rates from τ_0 below z^* to τ_1 above z^*), there is a fixed adjustment cost of ϕ^* , and $z^* \geq z_{1,\phi}$,²² then:

1. Individuals with initial earnings below a unique threshold \underline{z} do not adjust their earnings.
2. The threshold level of earnings is implicitly defined by the following:

$$u((1 - \tau_1) z^* + R_1, z^*) - u((1 - \tau_1) \underline{z}_1 + R_1, \underline{z}_1) \equiv \phi^* \quad (\text{E.3})$$

$$z^* \leq \underline{z}_1 \leq z^* + \Delta z_1^*.$$

3. Individuals with initial earnings in $[\underline{z}_1, z^* + \Delta z_1^*]$ bunch at the kink point z^* .
4. Individuals with initial earnings above $z^* + \Delta z_1^*$ reduce their earnings to a new level of earnings higher than z^* .
5. Excess mass, or bunching, at z^* is given by:

$$B = \int_{\underline{z}_1}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta$$

²² $z_{1,\phi}$ is again the *ex post* level of earnings for the individual who initially locates at $z_{0,\phi}$ – where $z_{0,\phi}$ is the initial earnings level over which individuals adjust their earnings defined above in Remark (1). Note that \underline{z} denotes this threshold in the non-linear budget set case, whereas $z_{0,\phi}$ denotes this threshold in the linear budget set case.

If the kink point z^* is lower than $z_{1,\phi}$, then:

1. Individuals only adjust their earnings if their initial earnings level is above the threshold $z_{0,\phi}$.
2. There is no bunching at z^* .

E.5 Derivation of Closed-Form Solution for Elasticity and Adjustment Cost

As we discuss in Remark (2) above and in Section 6 in the text, the amount of bunching in the presence of a fixed adjustment cost is equal to the integral of the initial earnings density over the range $[\underline{z}_1, z^* + \Delta z_1^*]$:

$$B(\boldsymbol{\tau}, z^*; \varepsilon, \phi^*) = \int_{\underline{z}_1}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta, \quad (\text{E.4})$$

where $\boldsymbol{\tau} \equiv (\tau_0, \tau_1)$. If the density is locally uniform, the integral in (E.4) is:

$$B(\boldsymbol{\tau}, z^*; \varepsilon, \phi^*) \approx h_0(\underline{z})(z^* + \Delta z_1^* - \underline{z}_1) \quad (\text{E.5})$$

Taking a first-order Taylor approximation of $u((1 - \tau_1)\underline{z}_1 + R_1, \underline{z}_1, \underline{n}_1)$ and $u((1 - \tau_1)z^* + R_1, z^*, \underline{n}_1)$ at $((1 - \tau_0)\underline{z}_1 + R_0, \underline{z}_1, \underline{n}_1)$, and using the first order condition for initial earnings, $(1 - \tau_0)u_c = -u_z$, we have from (E.3):

$$\begin{aligned} \phi^* &\approx u_c \cdot (1 - \tau_1) [z^* - \underline{z}_1] + u_z \cdot (z^* - \underline{z}_1) \\ &= u_c \cdot (\tau_1 - \tau_0) [\underline{z}_1 - z^*] \\ \Rightarrow \underline{z}_1 &\approx z^* + \frac{\phi^*/u_c}{(\tau_1 - \tau_0)} \\ &= z^* + \frac{\phi}{d\tau_1}, \end{aligned}$$

where $d\tau_1 = \tau_1 - \tau_0$ and $\phi = \phi^*/u_c$ is the dollar equivalent of the disutility associated with adjusting earnings. Substituting this expression for \underline{z}_1 into (E.5), we have

$$B(\boldsymbol{\tau}, z^*; \varepsilon, \phi) = h_0(\underline{z})(\Delta z_1^* - \phi/d\tau_1),$$

where bunching now depends on the dollar-denominated cost of adjusting, rather than the utility cost. Finally, for small $d\tau_1$, Δz_1^* is small and $h_0(\underline{z}) \approx h_0(z^* + \Delta z_1^*) \approx h_0(z^*)$. Let $b \equiv B/h_0(z^*)$, and note that $\Delta z_1^* = z^*(d\tau_1/(1 - \tau_0))\varepsilon$. The excess mass at the kink can now be expressed as a linear function of the parameters:

$$b(\boldsymbol{\tau}, z^*; \varepsilon, \phi) = \varepsilon \left(z^* \frac{d\tau_1}{1 - \tau_0} \right) - \phi \left(\frac{1}{d\tau_1} \right). \quad (\text{E.6})$$

E.5.1 Linear Adjustment Costs

We now introduce an adjustment cost that increases linearly in the size of the adjustment. Assume that given an initial level of earnings z_0 , agents must pay a cost of $\phi^* \cdot |z - z_0|$ when they change their earnings to a new level z . Utility \tilde{u} at the new earnings level can be represented as:

$$\tilde{u}(c, z; n, z_0) = u(c, z; n) - \phi^* \cdot |z - z_0|.$$

The first order condition for earnings can be characterized as:

$$\begin{aligned} -\frac{u_z(c, z; n)}{u_c(c, z; n)} &= (1 - \tau - \phi^*/\lambda^* \cdot \text{sgn}(z - z_0)) \\ &= \begin{cases} (1 - \tau - \phi) & \text{if } z > z_0 \\ (1 - \tau + \phi) & \text{if } z < z_0 \end{cases}, \end{aligned}$$

where $\lambda^* = u_c(c^*, z^*; n)$ is the Lagrange multiplier and $\phi = \phi^*/\lambda^*$ is the dollar equivalent of the linear adjustment cost ϕ^* .

The individual chooses earnings as if he faces an effective marginal tax rate of $\tilde{\tau} = \tau + \phi \cdot \text{sgn}(z - z_0)$. It follows that our predictions about earnings adjustment are similar to our previous predictions, except that the effective marginal tax rate $\tilde{\tau}$ appears, rather than τ . Thus, we can solve for the elasticity of earnings as a function of the change in earnings Δz^* due to introduction of a kink in the tax schedule and the jump in marginal tax rate $d\tau_1$:

$$\begin{aligned} \varepsilon &= \frac{\Delta z_1^*/z^*}{d\tilde{\tau}_1/(1 - \tilde{\tau}_0)} \\ &= \frac{\Delta z_1^*/z^*}{(d\tau_1 - 2\phi)/(1 - \tau_0 - \phi)}. \end{aligned}$$

Since the right-hand side is increasing in ϕ , the estimate of the elasticity increases as the linear adjustment cost increases. This makes intuitive sense: the adjustment cost attenuates bunching, so holding constant the level of bunching, the elasticity must be higher as the adjustment cost increases.

Now assume that when an individual adjusts his earnings, he incurs a linear adjustment cost ϕ^{*L} for every unit of change in earnings, as well as a fixed cost ϕ^{*F} associated with any change in earnings. Consider again bunching at z^* , with a tax rate jump of $d\tau_1 = \tau_1 - \tau_0$ at earnings level z^* . We have the following set of expressions for excess mass:

$$\begin{aligned} B &= \int_{\underline{z}}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta \\ \varepsilon &= \frac{\Delta z_1^*/z^*}{(d\tau_1 - 2\phi^L)/(1 - \tau_0 - \phi^L)} \\ \phi^{*F} + \phi^{*L} \cdot (z_1 - z^*) &= u\left((1 - \tau_1)z^* + R', z^*; \underline{n}_1\right) - u\left((1 - \tau_1)\underline{z} + R', \underline{z}_1; \underline{n}_1\right). \end{aligned}$$

Using a left rectangle approximation for the integral, we have:

$$\begin{aligned}
b &\equiv B/h_0(z^*) \\
&= z^* + \Delta z_1^* - \underline{z}_1 \\
&= z^* \left(\frac{d\tau_1 - 2\phi^L}{1 - \tau_0 - \phi^L} \varepsilon + 1 \right) - \underline{z}_1.
\end{aligned}$$

We can further apply an approximation for \underline{z} similar to the approximation we used in Section 6, *i.e.* $\underline{z} = z^* + \phi^F / (d\tau_1 - 2\phi^L)$. Thus, the expression for bunching can be simplified to:

$$b = \varepsilon \left(z^* \frac{d\tau_1 - 2\phi^L}{1 - \tau_0 - \phi^L} \right) - \frac{\phi^F}{(d\tau_1 - 2\phi^L)},$$

where $(\phi^F, \phi^L) = (\phi^{*F}/\lambda^*, \phi^{*L}/\lambda^*)$. In this case, we need at least three kinks to separately identify $(\varepsilon, \phi^F, \phi^L)$. Because we do not examine a setting in which one can compare bunching under three different positive tax rates, we are not able to estimate these parameters using data (or to fruitfully estimate the parameters in the non-linear case since we do not have a credible source of variation to identify them).

E.6 Derivation of Formula for Bunching with a Pre-Existing Kink

In our context, we conduct analysis using data just before and just after the benefit reduction rate was decreased (from 50 percent to 33.33 percent for 66-69 year olds in 1990, or from 33.33 percent for 69-year-olds to zero for 70-year-olds in 1990-1999). These changes involve moving from an initial state with a kink to a new state with a smaller kink. In a frictionless model, the distinction is immaterial. However, as we show, this matters in the presence of a fixed adjustment cost. In particular, when the kink becomes more muted, the change in bunching will be attenuated due to the fixed adjustment cost.

We will assume that in the initial state, bunching is characterized as in Remark (2), with a fixed adjustment cost. Let the initial kink, K_1 , be characterized by a lower marginal tax rate, τ_0 , under z^* , and a higher marginal tax rate, τ_1 , above z^* . The initial level of bunching is:

$$B_1 = \int_{\underline{z}_1}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta$$

Now, consider a change in the kink to K_2 , which retains the lower marginal tax rate τ_0 under z^* but reduces the marginal tax rate above z^* to $\tau_2 < \tau_1$. Had we begun with no kink and introduced K_2 , bunching would have been:

$$B_2 = \int_{\underline{z}_2}^{z^* + \Delta z_2^*} h_0(\zeta) d\zeta$$

Note that relative to K_1 , K_2 provides a weaker incentive to bunch, *when starting from a baseline tax schedule with no kink*. Formally, we have $\underline{z}_2 \geq \underline{z}_1$, $\Delta z_2^* < \Delta z_1^*$ and $B_2 \leq B_1$. However, as we show next, the fact that we move sequentially from K_1 to K_2 matters and will cause bunching under K_2 to in general differ from the amount predicted by formula E.6

above. In Gelber, Jones, and Sacks (2013), we show alternative estimates based on applying E.6 in both the cross sections in Period 1 and Period 2.

E.6.1 Characterizing Bunching

In characterizing bunching when moving sequentially from K_1 to K_2 , individuals may be separated into several groups based on their optimal level of earnings z_0 in the absence of a kink. First, there are individuals with $z_0 < z^*$. They will locate to the left of the kink under both K_1 and K_2 .

Second, we have individuals with $z^* < z_0 \leq \underline{z}_1$ (area *i* in Figure 4). These individuals would optimize in the presence of K_1 by moving to z^* , were it not for the adjustment cost. Now, with a smaller kink K_2 , these individuals continue to remain at the initial earnings level $z_0 > z^*$, as the utility gain to reoptimizing to z^* is even smaller than it was under K_1 .

Third, we have those with $\underline{z}_1 < z_0 \leq \underline{z}_2$ (area *ii* in Figure 4). When moving from no kink to K_1 , these individuals locate at the kink, z^* . If the budget set had hypothetically transitioned from no kink to K_2 , these individuals would have chosen to remain at z_0 , due to the fixed adjustment cost. However, when moving from K_1 to K_2 , these agents remain at the kink z^* . The reason is that the frictionless optimum under K_2 is z^* for everyone initially earning in the range $[z^*, z^* + \Delta z_2^*]$.

Fourth, we have agents with $\underline{z}_2 < z_0 \leq z^* + \Delta z_2^*$ (area *iii* in Figure 4). These individuals bunch at z^* when moving from no kink to either K_1 or K_2 . Thus, they remain bunching at z^* when moving from K_1 to K_2 .

Fifth, we have agents with $z^* + \Delta z_2^* < z_0 \leq z^* + \Delta z_1^*$ (areas *iv* and *v* in Figure 4). When starting from a budget set with no kink, these agents bunch under K_1 , but not under K_2 . Starting instead from K_1 , they must choose between remaining at the kink z^* or moving to the frictionless optimum under K_2 , $z_2 > z^*$. We know that at least some of these individuals will remain bunching. To see this, consider an individual with earnings under no kink $z_0 = z^* + \Delta z_2^* + \delta_0$. For small enough δ_0 optimal earnings under K_1 is z^* , and optimal earnings under K_2 tends to z^* as δ_0 tends to zero. Likewise, the net utility gain from relocating from z^* to z_2 under K_2 tends to zero as δ_0 tends to zero. However, the fixed adjustment cost remains strictly positive. Therefore, this individual will remain at z^* when moving from K_1 to K_2 for small enough δ_0 . In Figure 4, area *iv* shows those with initial earnings $z^* + \Delta z_2^* < z_0 < \bar{z}_0$ who remain bunching at the kink when transitioning from K_1 to K_2 . Area *v* shows those with initial earnings $\bar{z}_0 < z_0 < z^* + \Delta z_1^*$, who "unbunch" from the kink when moving from K_1 to K_2 .

When reoptimizing is beneficial for at least some agents in this final group, we will have a reduction in bunching when transitioning from K_1 to K_2 .²³ Empirically, we observe such a reduction over time, so this is the case relevant to our setting. In this case, the marginal

²³In certain cases, it is possible that reoptimizing away from the kink is not optimal for anyone in the initial earnings range $[z^* + \Delta z_2, z^* + \Delta z_1]$. In that case, there is no change in bunching when moving from K_1 to K_2 .

"unbuncher" will be defined by the following conditions:

$$\begin{aligned}
-\frac{u_z(c_2, \bar{z}_2; \bar{n}_2)}{u_c(c_2, \bar{z}_2; \bar{n}_2)} &= (1 - \tau_2) \\
u((1 - \tau_2)\bar{z}_2 + R_2, \bar{z}_2; \bar{n}_2) - u((1 - \tau_2)z^* + R_2, z^*; \bar{n}_2) &\equiv \phi^* \\
-\frac{u_z(c_0, \bar{z}_0; \bar{n}_2)}{u_c(c_0, \bar{z}_0; \bar{n}_2)} &= (1 - \tau_0) \\
\bar{z}_0 &\leq z^* + \Delta z_1^* \quad (\text{E.7})
\end{aligned}$$

In words, the first line indicates that $\bar{z}_2 > z^*$ is the optimal, frictionless level of earnings chosen by the top buncher in the presence of K_2 . The second line requires that when facing K_2 , this agent is indifferent between remaining at z^* and moving to \bar{z}_2 through paying the adjustment cost. The third line defines \bar{z}_0 as the initial level of earnings that this individual chooses when facing a constant marginal tax rate of τ_0 and no kink. The fourth line requires that this individual is initially bunching at z^* in response to K_1 . If this last inequality is binding, then when moving from K_1 to K_2 , none of the bunchers "unbunch" and the fraction bunching is unchanged. In that case, we have no variation available to identify both ε and ϕ . As noted above, empirically we do observe that bunching falls around 1990 when the BRR falls from 50 percent to 33.33 percent (as well as in the other cases we examine empirically, in which the BRR falls from a positive level to zero). Thus, we restrict attention to the case in which $\bar{z}_0 < z^* + \Delta z_1^*$.

Bunching at K_2 following K_1 , when $\bar{z}_0 < z^* + \Delta z_1^*$, can therefore be expressed as:

$$\tilde{B}_2 = \int_{\underline{z}_1}^{\bar{z}_0} h_0(\zeta) d\zeta$$

We can again solve this system of equations for ϕ^* and ε . Note that ε is still identified by the potential adjustment of the top-most buncher:

$$\varepsilon = \frac{\bar{z}_0 - \bar{z}_2}{\bar{z}_2} \frac{(1 - \tau_0)}{d\tau_2}$$

To build intuition for this, note that when moving from K_1 to K_2 , the change in bunching is smaller than it would be if we had started with steady state bunching at K_1 following no kink (B_1) and then moved to steady state bunching K_2 following no kink (B_2). That is:

$$\begin{aligned}
B_1 - \tilde{B}_2 &= \int_{\underline{z}_1}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta - \int_{\underline{z}_1}^{\bar{z}_0} h_0(\zeta) d\zeta \\
&\leq \int_{\underline{z}_1}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta - \int_{\underline{z}_2}^{z^* + \Delta z_2^*} h_0(\zeta) d\zeta \\
&= B_1 - B_2,
\end{aligned}$$

where the second line follows from the fact that $\bar{z}_0 \geq z^* + \Delta z_2^*$ and $\underline{z}_1 \leq \underline{z}_2$. The adjustment cost attenuates the change in bunching.

In our empirical application, we apply our method using variation in the size of the kink

in estimating elasticities and adjustment costs using data on individuals in different years (in the baseline specification, 1989 and 1990). Our empirical approach is applicable in the case in which individuals make year-by-year static earnings decisions; in the baseline, this effectively assumes that individuals weigh the cost of adjustment against the benefits in 1990. If instead individuals compare the costs of adjustment in 1990 to the benefits of adjustment in 1990 and subsequent years, then the benefits and therefore the estimated cost of adjustment would likely be larger. In this case, our estimated cost of adjustment could be considered a lower bound. Our estimates demonstrate the applicability of the methodology, including in settings in which the benefits may be realized over more years (as these discounted benefits would then be weighed against the costs). As we discuss, we view our approach as a natural first step toward estimating elasticities and adjustment costs, but incorporating more dynamic features in the model represents an important next step.

E.6.2 Simplified Approximation

We can again build intuition for this result by simplifying the formula for bunching in the second period. Assuming the density is constant over the range $[z^*, z^* + \Delta z_1^*]$, we have:

$$\begin{aligned}\tilde{b}_2 &= \bar{z}_0 - \underline{z}_1 \\ &= (\bar{z}_0 - \bar{z}_2) + \bar{z}_2 - \underline{z}_1 \\ &= \varepsilon \left(\bar{z}_2 \frac{d\tau_2}{1 - \tau_0} \right) + \bar{z}_2 - \underline{z}_1 \\ &= \varepsilon \left(\bar{z}_2 \frac{d\tau_2}{1 - \tau_0} \right) + (\bar{z}_2 - z^*) - \frac{\phi}{d\tau_1}\end{aligned}$$

where $\tilde{b}_2 \equiv \tilde{B}_2/h_0(z^*)$. On the third line, we have used the definition of the elasticity and on the fourth line, we used a first-order approximation to solve for \underline{z}_1 as before in Section (E.5). Thus, we can see why bunching after moving from a larger to smaller kink is greater than would be predicted by comparing bunching across two separate kinks using equation (E.4). First, the term multiplying ε has a \bar{z}_2 instead of a z^* and there is an additional term $\bar{z}_2 - z^*$ – both of which increase bunching, since $\bar{z}_2 > z^*$. Both of these capture the excess bunching from above (i.e. area *iv*) due to inertia. Finally, the third term has a $d\tau_1$ in the denominator instead of a $d\tau_2$. The larger denominator increases bunching – $d\tau_1 > d\tau_2$ – and captures the fact that there is less attenuation in bunching from below (i.e. area *ii*), also due to inertia.

We can also show that $|\tilde{b}_2 - b_1|$ is decreasing in ϕ :

$$\begin{aligned}|\tilde{b}_2 - b_1| &= |(\bar{z}_0 - \underline{z}_1) - (z^* + \Delta z_1^* - \underline{z}_1)| \\ &= z^* + \Delta z_1^* - \bar{z}_0.\end{aligned}$$

Since \bar{z}_0 is increasing in ϕ , the difference $|\tilde{b}_2 - b_1|$ is decreasing in ϕ (as Δz_1^* and z^* do not depend on ϕ).

E.6.3 Pre-Existing Kink with Linear Adjustment Cost

Note that with a reduction in the size of a kink, with a linear adjustment cost we can derive the following approximation for bunching (analogously to Section E.5.1):

$$\tilde{b}_2 = \varepsilon \left(\bar{z}_2 \frac{d\tau_2 - 2\phi^L}{1 - \tau_0 - \phi^L} \right) - (\bar{z}_2 - z^*) - \frac{\phi^F}{d\tau_1 - 2\phi^L}.$$

E.7 Elasticities Under Frictionless (Saez 2010) Formula

We investigate the results when applying the Saez (2010) formula for estimating elasticities—applicable to a frictionless setting—in a setting in which there are in fact adjustment costs. In other words, we answer the question: if there are adjustment costs and we mis-specify our estimate of the elasticity by assuming that we are in a frictionless setting, in what way do we mis-estimate the elasticity? As we show, if we face an adjustment cost, but we estimate the elasticity using the Saez (2010) formula applicable to a frictionless setting, we will see the elasticity estimate increase when we move from a larger kink to a smaller kink (as in our empirical application, and as we observe empirically in Figure 6). Thus, an increase in the Saez (2010) estimate of the elasticity—of the sort that we observe empirically—is a telltale sign that we are operating under a model with an adjustment cost.

We first present the formula for the elasticity in a frictionless model, as in Saez (2010). Next, we present the formulas for the elasticities we would estimate if we mis-specified the model as the frictionless (Saez 2010) model, even though we in fact face an adjustment cost.

E.7.1 Saez (2010) model

We assume that tax changes are relatively small and that we can therefore treat the density is constant. (We derive analogous results if we instead use exact formulas under quasilinearity.) Assume that we begin with a more pronounced kink K_1 and then move to a less pronounced kink K_2 , by lowering the jump in marginal tax rates at exempt amount from $d\tau_1$ to $d\tau_2$. Assume in each year, we can estimate normalized bunching: $b \equiv B/h_0(z^*)$. In a frictionless (Saez 2010) model, we have:

$$\begin{aligned} b_1 &= \Delta z_1^* = \varepsilon z^* \frac{d\tau_1}{1 - \tau_0} \\ b_2 &= \Delta z_2^* = \varepsilon z^* \frac{d\tau_2}{1 - \tau_0} \end{aligned}$$

where we have used the fact that: $\Delta z^* = \varepsilon z^* d\tau / (1 - \tau_0)$. A natural estimator of the elasticity is the Saez estimator e^S :

$$e^S = \frac{b}{z^*} \frac{(1 - \tau_0)}{d\tau} = \frac{b}{a \cdot d\tau}$$

where $a \equiv z^* / (1 - \tau_0)$.

In each period (denoted by the subscript), we have the following for the Saez estimator

when there are no frictions:

$$\begin{aligned} e_1^S &= \frac{b_1}{a \cdot d\tau_1} = \varepsilon \\ e_2^S &= \frac{b_2}{a \cdot d\tau_2} = \varepsilon \end{aligned}$$

Thus, $e_1^S = e_2^S$. Here e_1^S denotes the Saez (2010) estimate of the elasticity in period 1 (under K_1), and e_2^S denotes this elasticity in period 2 (under K_2).

E.7.2 Model with Fixed Adjustment Costs

By contrast, in the presence of a fixed adjustment cost, we start at kink K_1 and then move straight to K_2 , once again estimating normalized bunching. We have the following results here:

$$\begin{aligned} b_1 &= z^* + \Delta z_1^* - z_1 \\ b_2 &= \bar{z}_0 - z_1 \end{aligned}$$

Rewrite b_1 and b_2 as follows:

$$\begin{aligned} b_1 &= \Delta z_1^* - (z_1 - z^*) \\ b_2 &= \Delta z_2^* + (\bar{z}_0 - \Delta z_2^* - z_1) \end{aligned}$$

The Saez estimators now return:

$$\begin{aligned} e_1^S &= \frac{b_1}{a \cdot d\tau_1} = \varepsilon - \frac{(z_1 - z^*)}{a \cdot d\tau_1} \\ e_2^S &= \frac{b_2}{a \cdot d\tau_2} = \varepsilon + \frac{(\bar{z}_0 - \Delta z_2^* - z_1)}{a \cdot d\tau_2} \end{aligned}$$

The change in Saez estimators is:

$$\begin{aligned} e_2^S - e_1^S &= \frac{(\bar{z}_0 - \Delta z_2^* - z_1)}{a \cdot d\tau_2} + \frac{(z_1 - z^*)}{a \cdot d\tau_1} \\ &\geq \frac{(\bar{z}_0 - \Delta z_2^* - z_1)}{a \cdot d\tau_1} + \frac{(z_1 - z^*)}{a \cdot d\tau_1} \\ &= \frac{\bar{z}_0 - \Delta z_2^* - z_1 + z_1 - z^*}{a \cdot d\tau_1} \\ &= \frac{\bar{z}_0 - z^* - \Delta z_2^*}{a \cdot d\tau_1} \\ &\geq \frac{z^* + \Delta z_2^* - z^* - \Delta z_2^*}{a \cdot d\tau_1} \\ &= 0 \end{aligned}$$

where in the second line, we use the fact that $d\tau_1 > d\tau_2$ and in the second-to-last line, we use the fact that $\bar{z}_0 \geq z^* + \Delta z_2^*$.

Thus, $e_2^S - e_1^S$ is weakly greater than zero: the Saez (2010) frictionless elasticity estimate weakly increases (as we observe empirically). The pattern we observe empirically—an upward jump in the Saez (2010) estimate of the elasticity when the policy change occurs in 1990, as shown in Figure 6—is a telltale sign that we are operating in the presence of an adjustment friction.

E.8 Derivation of Formula for Bunching with Heterogeneity

Under heterogenous preferences, our estimates can be interpreted as identify average parameters among the set of bunchers (as in Saez 2010 and Kleven and Waseem 2013). As described in Section E, suppose $(\varepsilon_i, \phi_i, n_i)$ is jointly distributed according to a smooth CDF, which translates to a smooth, joint distribution of elasticities, fixed costs and earnings. Let the joint density of earnings, adjustment costs and elasticities be $\tilde{h}_0(z, \varepsilon, \phi)$ under a linear tax of τ_0 . Assume that the density of earnings is again constant over the interval $[\underline{z}_1, z^* + \Delta z^*]$, conditional on ε and ϕ . When moving from no kink to a kink, we derive a formula for bunching B in the presence of heterogeneity as follows:

$$\begin{aligned} B_1 &= \iiint_{\underline{z}_1}^{z^* + \Delta z^*} \tilde{h}_0(\zeta, \epsilon, \varphi) d\zeta d\epsilon d\varphi \\ &= \iint [z^* + \Delta z_1^* - \underline{z}_1] \tilde{h}_0(z^*, \epsilon, \varphi) d\epsilon d\varphi \\ &= h_0(z^*) \cdot \iint [z^* + \Delta z_1^* - \underline{z}_1] \frac{\tilde{h}_0(z^*, \epsilon, \varphi)}{h_0(z^*)} d\epsilon d\varphi \\ &= h_0(z^*) \cdot \mathbb{E}[z^* + \Delta z_1^* - \underline{z}_1], \end{aligned}$$

where we have used the assumption of constant $\tilde{h}_0(\cdot)$ in line two, $h_0(z^*) = \iint \tilde{h}_0(z^*, \epsilon, \varphi) d\epsilon d\varphi$ and ζ, ϵ and φ are dummies of integration. The expectation $\mathbb{E}[\cdot]$ is taken over the set of bunchers, under the various combinations of ε and ϕ throughout the support. It follows that normalized bunching can be expressed as follows:

$$b_1 = z^* + \mathbb{E}[\Delta z_1^*] - \mathbb{E}[\underline{z}_1].$$

Under heterogeneity, the level of bunching identifies the average behavioral response, Δz^* , and threshold earnings, \underline{z}_1 , among the marginal bunchers under each possible combination of parameters ε and ϕ . Under certain parameter values, there is no bunching, and thus, the values of the elasticity and adjustment cost in these cases do not contribute our estimates. Using the approximation for \underline{z} from (E.6) above, we can further show the following:

$$\begin{aligned} b_1 &= \mathbb{E}[z^* + \Delta z_1^* - \underline{z}_1] \\ &= \mathbb{E}\left[\varepsilon \left(z^* \frac{d\tau_1}{1 - \tau_0}\right) - \phi \left(\frac{1}{d\tau_1}\right)\right] \\ &= \bar{\varepsilon} \left(z^* \frac{d\tau_1}{1 - \tau_0}\right) - \bar{\phi} \left(\frac{1}{d\tau_1}\right) \end{aligned}$$

where $\bar{\varepsilon}$ and $\bar{\phi}$ are the average elasticity and adjustment cost, respectively.

When we move sequentially from a larger kink, K_1 to a smaller kink, K_2 , our formula for bunching under K_2 in the presence of heterogeneity is likewise derived as follows:

$$\begin{aligned}
\tilde{B}_2 &= \iiint_{\underline{z}_1}^{\bar{z}_0} \tilde{h}_0(\zeta, \epsilon, \varphi) d\zeta d\epsilon d\varphi \\
&= \iint [\bar{z}_0 - \underline{z}_1] \tilde{h}_0(z^*, \epsilon, \varphi) d\epsilon d\varphi \\
&= h_0(z^*) \cdot \iint [\bar{z}_0 - \underline{z}_1] \frac{\tilde{h}_0(z^*, \epsilon, \varphi)}{h_0(z^*)} d\epsilon d\varphi \\
&= h_0(z^*) \cdot \mathbb{E}[\bar{z}_0 - \underline{z}_1].
\end{aligned}$$

Similarly, normalized bunching can now be expressed as follows:

$$b_2 = \mathbb{E}[\bar{z}_0] - \mathbb{E}[\underline{z}_1].$$

Once again, the expectations are taken over the population of bunchers and can be related to the parameters ε and ϕ using (E.7).

Following the approach in Kleven and Waseem (2013, pg. 682), the average value of the parameters Δz_1^* , \underline{z}_1 and \bar{z}_0 can then be related to ε and ϕ , assuming a quasi-linear utility function and using equations (2) and (4) and the identities $\Delta z_1^* = \varepsilon z^* d\tau_1 / (1 - \tau_0)$ and $\bar{z}_0 - \bar{z}_2 = \varepsilon \bar{z}_2 d\tau_2 / (1 - \tau_0)$.

E.9 Allowing for Frictions in Initial Earnings

In the initial period 0 (prior to the policy change), under a linear tax of τ_0 , we have assumed that individuals are located at their frictionless optimum, while we have assumed elsewhere that adjustment costs may preclude individuals from reaching their exact, interior optimum. Here, we extend the model to allow for agents to be away from their optimum in period 0, in a way that is consistent with our model of a fixed adjustment cost.

We now analyze the thought experiment previously discussed. From a linear tax of τ_0 in period 0, in period 1 we introduce a kink, K_1 , at z^* , and let the marginal tax rate increase to τ_1 for earnings above z^* . Finally, in period 2 we replace the first kink with a second, smaller kink, K_2 , at z^* , where the marginal tax rate only increases to τ_2 .

Again, agents are indexed by n . Let $z_{n,t}$ be actual earnings for individual n in period t , and let $z_{n,t}^{Opt}$ be the optimal level of earnings she would choose in the absence of adjustment frictions. As in Chetty (2012), assume that earnings are not "too far" from the frictionless optimum; that is, assume that earnings are within a set such that the utility gain of adjusting to the optimum does not exceed the adjustment cost. Formally:

$$\begin{aligned}
z_{n,t} \left(z_{n,0}^{Opt} \right) &\in \left[z_{n,t}^- \left(z_{n,0}^{Opt} \right), z_{n,t}^+ \left(z_{n,0}^{Opt} \right) \right], \\
&\text{where } z_{n,t}^- \leq z_{n,t}^{Opt} \leq z_{n,t}^+ \\
\text{and } u \left(z_{n,t}^{Opt} - T_t \left(z_{n,t}^{Opt} \right), z_{n,t}^{Opt}; n \right) - \phi^* &= u \left(z_{n,t}^- - T_t \left(z_{n,t}^- \right), z_{n,t}^-; n \right) \\
&= u \left(z_{n,t}^+ - T_t \left(z_{n,t}^+ \right), z_{n,t}^+; n \right)
\end{aligned} \tag{E.8}$$

where $T_t(\cdot)$ represents a linear tax of τ_0 in period 0, reflects the kink K_1 in period 1, and reflects the kink K_2 in period 2. In words, $z_{n,t}^-$ and $z_{n,t}^+$ are the lowest and highest level of earnings, respectively, that would be acceptable before an individual chooses to adjust to their optimal earnings level. Note that we have defined $z_{n,t} \left(z_{n,0}^{Opt} \right)$ as a function of the optimal level of earnings for individual n in period 0 for notational convenience. Let the actual earnings, conditional on optimal earnings in period 0, be distributed according to the cumulative distribution function $F_{n,t} \left(z_{n,t} | z_{n,0}^{Opt} \right)$, with probability density function $f_{n,t} \left(z_{n,t} | z_{n,0}^{Opt} \right)$. Thus, individuals are distributed around their frictionless optimum in period 0.

First, consider the level of bunching at K_1 . Relative to our baseline model with frictions (that assumes individuals are initially located at their frictionless optimum), there will be two differences in who bunches. First, individuals in Figure (4) Panel B area i did not bunch in the baseline because they were sufficiently close to the kink. These are agents for whom $z^* < z_{n,0}^{Opt} < \bar{z}_1$. Now, with some probability, a fraction of these agents will be sufficiently far from z^* in period 0 to justify moving to the kink in Period 1—formally, those for whom $z_{n,0} \in [z_{n,1}^+, z_{n,0}^+]$. Their initial earnings are above their interior optimum in period 0, but not far enough to outweigh the fixed cost of adjustment in Period 0. Now that the optimum in period 1 has moved to z^* , the utility gain to readjusting exceeds the fixed cost of adjustment. These individuals will now bunch under K_1 . The second difference in this version of the model relative to our baseline model is that some individuals who had bunched under K_1 in the baseline model, i.e. areas ii, iii, iv , and v in Figure 4, may find themselves already close enough to z^* in period 0 that they do not bunch at z^* in period 0 (because relocating to z^* in period 0 does not have sufficient benefit to outweigh the fixed adjustment cost). Formally, these are individuals for whom $z_{n,0} < z_{n,1}^+$. These cases are illustrated in Appendix Figure F.16.

Define bunching under this modified model as B'_1 . Bunching under K_1 can be expressed as:

$$\begin{aligned} B'_1 &= \int_{z^*}^{z^* + \Delta z_1^*} \left[\int_{z_{n,1}^+}^{z_{n,0}^+} f_{n,0}(v | \zeta) dv \right] h_0(\zeta) d\zeta \\ &= \int_{z^*}^{z^* + \Delta z_1^*} [1 - F_{n,0}(z_{n,1}^+ | \zeta)] h_0(\zeta) d\zeta \\ &= \int_{z^*}^{z^* + \Delta z_1^*} \Pr \left(z_{n,0} \geq z_{n,1}^+ | z_{n,0}^{Opt} = \zeta \right) h_0(\zeta) d\zeta \end{aligned}$$

where ν and ζ are dummies of integration.

We now turn to bunching in period 2, under K_2 . Note that because this kink is smaller, anyone sufficiently close to z^* that they did not bunch under K_1 will continue not to bunch under K_2 . Thus, the only change in bunching in period 2 will be those who now move away from the kink. Under the baseline model, these were individuals for whom $\bar{z}_0 \leq z_{n,0}^{Opt} \leq z^* + \Delta z_1^*$, i.e. area v in Figure 4, Panel B. These individuals will still find it worthwhile to move away from the kink, but the difference from the baseline model is that only a subset of them bunched in period 1. Thus, the decrease in bunching will be related to the share of

people in area v who actually bunched under K_1 . What remains are those individuals with $z^* \leq z_{n,0}^{Opt} \leq \bar{z}_0$ who actually bunched in period 1. Formally, bunching in period 2 under K_2 can be expressed as follows:

$$\begin{aligned} B'_2 &= \int_{z^*}^{\bar{z}_0} \left[\int_{z_{n,1}^+}^{z_{n,0}^+} f_{n,0}(v|\zeta) dv \right] h_0(\zeta) d\zeta \\ &= \int_{z^*}^{\bar{z}_0} [1 - F_{n,0}(z_{n,1}^+|\zeta)] h_0(\zeta) d\zeta \\ &= \int_{z^*}^{\bar{z}_0} \Pr(z_{n,0} \geq z_{n,1}^+ | z_{n,0}^{Opt} = \zeta) h_0(\zeta) d\zeta \end{aligned}$$

We can rewrite the level of bunching in terms of counterfactual bunching amounts that would occur in the absence of frictions in previous periods:

$$\begin{aligned} B'_1 &= \int_{z^*}^{z^* + \Delta z_1^*} \Pr(z_{n,0} \geq z_{n,1}^+ | z_{n,0}^{Opt} = \zeta) h(\zeta) d\zeta \\ &= \int_{z^*}^{z^* + \Delta z_1^*} h(\zeta) d\zeta \cdot \int_{z^*}^{z^* + \Delta z_1} \Pr(z_{n,0} \geq z_{n,1}^+ | z_{n,0}^{Opt} = \zeta) \frac{h(\zeta)}{\int_{z^*}^{z^* + \Delta z_1} h(\zeta) d\zeta} d\zeta \\ &= B''_1 \cdot \int_{z^*}^{z^* + \Delta z_1^*} \Pr(z_{n,0} \geq z_{n,1}^+ | z_{n,0}^{Opt} = \zeta) h(\zeta | z^* < \zeta \leq z^* + \Delta z_1^*) d\zeta \\ &= B''_1 \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < z_{n,0}^{Opt} \leq z^* + \Delta z_1^* \right] \end{aligned}$$

where $B''_1 = \int_{z^*}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta$ is the bunching that would obtain in a model of no frictions under K_1 , i.e. areas $i - v$ in Figure 4, Panel B. Likewise, we have:

$$\begin{aligned} B'_2 &= \int_{z^*}^{\bar{z}_0} \Pr(z_{n,0} \geq z_{n,1}^+ | z_{n,0}^{Opt} = \zeta) h_0(\zeta) d\zeta \\ &= B''_2 \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < z_{n,0}^{Opt} \leq \bar{z}_0 \right] \end{aligned}$$

where $B''_2 = \int_{z^*}^{\bar{z}_0} h_0(\zeta) d\zeta$ is the level of bunching that would obtain if earnings decisions in period 0 and period 1 were frictionless, but bunching under period 2 was subject to the adjustment cost. These are individuals in areas $i - iv$ in Figure 4.

Without further restrictions on the distribution of heterogeneity or distribution of earnings about the frictionless optimum in period 0, we cannot make further, significant simplifications of these expressions. However, if we assume that the optimal earnings density, $h_0(\cdot)$, is constant over the range $[z^*, z^* + \Delta z_1^*]$, as is common in the literature (e.g., Chetty *et al.* 2011 or Kleven and Waseem 2013), then we have the following:

$$\begin{aligned} B'_1 &= B''_1 \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < z_{n,0}^{Opt} \leq z^* + \Delta z_1^* \right] \\ &= \Delta z_1^* h_0(z^*) \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < z_{n,0}^{Opt} \leq z^* + \Delta z_1^* \right] \end{aligned}$$

and likewise:

$$\begin{aligned} B_2' &= B_2'' \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < z_{n,0}^{Opt} \leq \bar{z}_0 \right] \\ &= [\bar{z}_0 - z^*] h_0(z^*) \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < z_{n,0}^{Opt} \leq \bar{z}_0 \right] \end{aligned}$$

It also follows that bunching normalized by the height of the density at the kink will be:

$$\begin{aligned} b_1' &= \Delta z_1^* \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < z_{n,0}^{Opt} \leq z^* + \Delta z_1^* \right] \\ b_2' &= [\bar{z}_0 - z^*] \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < z_{n,0}^{Opt} \leq \bar{z}_0 \right] \end{aligned}$$

If we further assume that the initial actual earnings level is distributed uniformly about optimal earnings in period 0, then we have:

$$z_{n,0} \sim U[z_{n,0}^-, z_{n,0}^+]$$

which implies that:

$$\Pr(z_{n,0} \geq z_{n,1}^+ | z_{n,0}^{Opt} = \zeta) = \min \left(\frac{z_{n,0}^+(\zeta) - z_{n,1}^+(\zeta)}{z_{n,0}^+(\zeta) - z_{n,0}^-(\zeta)}, 1 \right)$$

Using our definitions above for $z_{n,0}^+(\cdot)$, $z_{n,0}^-(\cdot)$ and $z_{n,1}^+(\cdot)$ we can calculate this probability conditional on initial frictionless earnings in period 0, the elasticity ε and the adjustment cost ϕ^* . Note that the uniform distribution of actual earnings is not generally centered at the optimal earnings level in period 0, since the lower and upper limits of the support in period 0, i.e. $[z_{n,0}^-, z_{n,0}^+]$, will tend to be different distances from the frictionless optimum.

E.10 Estimating the Elasticity and Adjustment Cost

In this section, we describe in more detail how we use data on the amount of bunching to estimate the elasticity and adjustment cost. Let $b = (b_1, b_2, \dots, b_K)$ be a vector of (estimated) bunching amounts normalized by the density at the kink, using the method described in Section (3). Let $\tau = (\tau_1, \dots, \tau_K)$ be the tax schedule at each kink. The triplet $\tau_k = (\tau_0^k, \tau_1^k, \tau_2^k)$ denotes the tax rate below (τ_0^k) and above (τ_1^k) the kink k and τ_2^k denotes the *ex post* marginal tax rate above the kink after it has been reduced, as in Section (E.6). Let $z^* = (z_1^*, \dots, z_K^*)$ be the earnings levels associated with each kink. To estimate (ε, ϕ) , we seek the values of the parameters that make predicted bunching \hat{b} and actual (estimated) bunching b as close as possible on average.

Letting $\hat{b}(\varepsilon, \phi) = (\hat{b}(\tau_1, z_1^*, \varepsilon, \phi), \dots, \hat{b}(\tau_K, z_K^*, \varepsilon, \phi))$, our estimator is:

$$(\hat{\varepsilon}, \hat{\phi}) = \operatorname{argmin}_{(\varepsilon, \phi)} \left(\hat{b}(\varepsilon, \phi) - b \right)' W \left(\hat{b}(\varepsilon, \phi) - b \right), \quad (\text{E.9})$$

where W is a $K \times K$ diagonal matrix whose diagonal entries are the inverse of the variances of the estimates of the b_k .

We obtain our estimates by minimizing equation (E.9) numerically. Solving this problem

requires evaluating \hat{b} at each trial guess of (ε, ϕ) .²⁴ Recall that in general bunching takes the form:

$$B_k(\tau_k, z_k^*; \varepsilon, \phi^*) = \int_{z_k^{lb}}^{z_k^{ub}} h_0(\zeta) d\zeta,$$

where (z_k^{lb}, z_k^{ub}) are the ex-ante earnings levels of the lowest and highest earning bunchers, in the presence of linear tax at the lower tax rate, τ_0^k . Define $z_k^* + \Delta z_1^k$ as the *ex ante* earnings level for the highest earning buncher – in the absence of frictions – when the size of the kink is $d\tau_1^k = \tau_1^k - \tau_0^k$. As in the main text, we continue to assume that $h(\cdot)$ is uniform in $[z_k^*, z_k^* + \Delta z_1^k]$, so that

$$b_k(\tau_k, z_k^*; \varepsilon, \phi^*) = z_k^{ub} - z_k^{lb},$$

where $b = B/h(z_k^*)$. The definitions of (z_k^{lb}, z_k^{ub}) vary depending on the setting and are defined as follows. In the frictionless case (Saez 2010), we have:

$$\begin{aligned} z_k^{lb} &= z_k^* \\ z_k^{ub} &= z_k^* + \Delta z_1^k. \end{aligned}$$

In the presence of a fixed adjustment cost, we have under the the first kink, K_1 :

$$\begin{aligned} z_k^{lb} &= \underline{z}_1^k \\ z_k^{ub} &= z_k^* + \Delta z_1^k, \end{aligned}$$

where \underline{z}_1^k is the ex-ante earnings of the marginal buncher from below given a fixed cost of adjustment. This is defined in the indifference condition above in equation (E.3). When we move sequentially from a larger kink to a smaller kink, K_2 , we have:

$$\begin{aligned} z_k^{lb} &= \underline{z}_1^k \\ z_k^{ub} &= \bar{z}_0^k, \end{aligned}$$

where \bar{z}_0^k is similarly the *ex ante* earnings of the marginal buncher from below (calculated using a kink with $d\tau_1^k = \tau_1^k - \tau_0^k$). The *ex ante* earnings of the marginal buncher from above, \bar{z}_0^k , is defined in Section (E.6) where $d\tau_1^k = \tau_1^k - \tau_0^k$ and $d\tau_2^k = \tau_2^k - \tau_0^k$.

Finally, if we allow for frictions in initial earnings in period 0, the formulae are slightly adjusted. We have under the first kink, K_1 :

$$\begin{aligned} z_k^{lb} &= z_k^* \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) \mid z_k^* < z_{n,0}^{Opt} \leq z_k^* + \Delta z_1^k \right] \\ z_k^{ub} &= (z_k^* + \Delta z_1^k) \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+) \mid z_k^* < z_{n,0}^{Opt} \leq z_k^* + \Delta z_1^k \right], \end{aligned}$$

²⁴In solving problem (E.9), we impose that $\phi \geq 0$. When $\phi < 0$, every individual adjusts her earnings by at least some arbitrarily small amount, regardless of the size of ϕ . This implies that ϕ is not identified if it is less than zero.

and under the second kink, K_2 :

$$\begin{aligned} z_k^{lb} &= z_k^* \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+ | z_k^* < z_{n,0}^{Opt} \leq \bar{z}_0^k) \right] \\ z_k^{ub} &= \bar{z}_0^k \cdot \mathbb{E} \left[\Pr(z_{n,0} \geq z_{n,1}^+ | z_k^* < z_{n,0}^{Opt} \leq \bar{z}_0^k) \right]. \end{aligned}$$

In this last case, we must calculate $\mathbb{E} [\Pr(z_{n,0} \geq z_{n,1}^+)]$, which requires knowledge of the CDF $F_{n,t}(z_{n,t} | z_{n,0}^{Opt})$ above. In the case of uniform distribution for $F(\cdot)$ and a constant value of $h_0(\cdot)$ over the range $[z_k^*, z_k^* + \Delta z_1^k]$, we can calculate the expectation as follows:

$$\begin{aligned} \mathbb{E} [\Pr(z_{n,0} \geq z_{n,1}^+ | z_k^* < z_{n,0}^{Opt} \leq z_k^* + \Delta z_1^k)] &\approx \frac{1}{L} \sum_{l=1}^L \min \left[\frac{z_{n,0}^+(z_l; \varepsilon, \phi) - z_{n,1}^+(z_l; \varepsilon, \phi)}{z_{n,0}^+(z_l; \varepsilon, \phi) - z_{n,0}^-(z_l; \varepsilon, \phi)}, 1 \right] \\ \mathbb{E} [\Pr(z_{n,0} \geq z_{n,1}^+ | z_k^* < z_{n,0}^{Opt} \leq \bar{z}_0^k)] &\approx \frac{1}{J} \sum_{j=1}^J \min \left[\frac{z_{n,0}^+(z_j; \varepsilon, \phi) - z_{n,1}^+(z_j; \varepsilon, \phi)}{z_{n,0}^+(z_j; \varepsilon, \phi) - z_{n,0}^-(z_j; \varepsilon, \phi)}, 1 \right] \end{aligned}$$

where the points z_l and z_j are drawn from the relevant ranges (i.e. $z_k^* < z_{n,0}^{Opt} \leq z_k^* + \Delta z_1^k$ and $z_k^* < z_{n,0}^{Opt} \leq \bar{z}_0^k$) and the values $z_{n,0}^-(\cdot; \varepsilon, \phi)$, $z_{n,0}^+(\cdot; \varepsilon, \phi)$ and $z_{n,1}^+(\cdot; \varepsilon, \phi)$ are evaluated using equation (E.8) above.

Our estimator assumes a quasilinear utility function, $u(c, z; n) = c - \frac{n}{1+1/\varepsilon} \left(\frac{z}{n}\right)^{1+1/\varepsilon}$, following Saez (2010), Chetty *et al.* (2011) and Kleven and Waseem (2013). In order to relax this assumption empirically, we would have to observe wealth, which is not available in the data. Note that because we have assumed quasilinearity, $\phi^* = \phi$, $\Delta z_1^k = z_k^* \left(\left(\frac{1-\tau_1^k}{1-\tau_0^k} \right)^\varepsilon - 1 \right)$ and $n = z(\tau) / (1-\tau)^\varepsilon$, where $z(\tau)$ are the optimal, interior earnings under a linear tax of τ . However, there typically is not a closed form solution for the (z_k^{lb}, z_k^{ub}) in other cases. Instead, given ε and ϕ , we find (z_k^{lb}, z_k^{ub}) numerically as the solution to relevant indifference condition. For example, \underline{z}_1^k is defined implicitly by:

$$\underbrace{u((1-\tau_1^k)z_k^* + R_1^k, z_k^*; \underline{z}_1^k / (1-\tau_0^k)^\varepsilon)}_{\text{utility from adjusting to kink}} - \underbrace{u((1-\tau_1^k)\underline{z}_1^k + R_1^k, \underline{z}_1^k; \underline{z}_1^k / (1-\tau_0^k)^\varepsilon)}_{\text{utility from not adjusting}} = \phi,$$

The equation is continuously differentiable and has a unique solution for \underline{z}_1^k . As such, Newton-type solvers are able to find \underline{z}_1^k accurately. Note that some combinations of $\tau_k, z_k^*, \varepsilon$, and ϕ imply $z_k^{lb} > z_k^{ub}$. In this case, the lowest-earning adjuster does not adjust to the kink, and whenever this happens we set $\hat{b}_k = 0$. The predicted amount of bunching is therefore:

$$\hat{b}_k(\tau_k, z_k^*; \varepsilon, \phi) = \max(z_k^{ub} - z_k^{lb}, 0).$$

We have also shown a robustness check in our tables in which we use a non-parametric density for the counterfactual earnings distribution, H_0 . Once H_0 is known, we use the general expression for B_k to get predicted bunching from the model. To recover H_0 non-parametrically we take the empirical earnings distribution for 72 year-olds as the counterfactual distribution. To estimate this distribution, we calculate for each \$800 bin the share

of 72 year-olds with earnings in that bin. Letting z_i index the bins, our estimate of the distribution is

$$\hat{H}_0(z_i) = \sum_{j \leq i} Pr(z \in z_j).$$

This function is only defined at the mid-points of the bins, so we use linear interpolation for other values of z .

We estimate bootstrapped standard errors. Observe that the estimated vector of parameters $(\hat{\varepsilon}, \hat{\phi})$ is a function of the estimated amount of bunching; call this function $\theta(b)$. To compute bootstrapped standard errors, we use the bootstrap procedure of Chetty *et al.* (2011) to obtain 200 bootstrap samples of b . For each bootstrap sample, we compute $\hat{\varepsilon}$ and $\hat{\phi}$ as the solution to (E.9). The standard deviation of $\hat{\varepsilon}$ and $\hat{\phi}$ across bootstrap samples is the bootstrap standard error, and we compute confidence intervals analogously. We estimate whether an estimate is significantly different from zero by assessing how frequently the constraint $\phi \geq 0$ binds in our estimation. Given this constraint, p-values are from a one-sided test of equality with zero. We have also estimated the standard errors using the delta method and obtained similar results.

E.11 Policy Simulations

In this Appendix, we describe how we simulate the effect of various policy changes on earnings. These calculations are designed to be illustrative of the attenuation of earnings responses to policy changes that can result from incorporating adjustment frictions in the analysis. Nonetheless, we highlight that these calculations are done in the context of a highly stylized model making a number of assumptions, as well as a particular sample of earners. One key (extreme) assumption is that everyone has the same elasticity and adjustment cost. Moreover, these estimates are specific to a particular context, and they are by no means intended to be an exhaustive account of the implications of adjustment costs for earnings responses to taxation. Rather, they are intended simply to illustrate the attenuation of earnings responses to policy changes that can result from incorporating adjustment frictions in the analysis in such contexts.

We assume that utility is isoelastic and quasi-linear with elasticity ε . Individuals must pay an adjustment cost ϕ to change their earnings. Individuals are heterogeneous in their ability n_i . Individuals are therefore distributed according to their "counterfactual" earnings z_{0i} that they would have under a linear tax schedule. (Despite the absence of heterogeneity in the elasticity and adjustment cost, there is still heterogeneity in the gains from re-optimizing earnings, due to heterogeneity in z_{0i} .) We use the 1989 earnings distribution for 60-61 year-olds (from the MEF data) as the counterfactual earnings distribution, i.e. the earnings distribution under a linear tax schedule in the region of the exempt amount. We incorporate the key features of the individual income tax code, including individual federal income taxes, state income taxes, and FICA (all from Taxsim applied in 1989), and the AET. Our estimates of elasticities and adjustment costs apply to a population earning near the exempt amount; to avoid extrapolating too far out of sample, our simulations examine only those whose counterfactual earnings is from \$10,000 under to \$10,000 over the exempt amount (and is greater than \$0). (While the AET should only affect people whose counterfactual earnings are over the exempt amount, we also include the group earning up to \$10,000 under the

exempt amount in order to illustrate the fact that some individuals could be unaffected by a policy change.)

We consider two periods, 1 and 2. In period 1, in the region of the AET exempt amount, the mean tax rate below the exempt amount is 27.21 percent, and the mean tax rate above the exempt amount is 77.21 percent. Note that these tax rates mimic those faced by 62-64 year-old OASI claimants.²⁵ In period 2, the tax rate below the exempt amount remains 27.21 percent, but the tax rate above the exempt amount changes according to the policy changes we specify below. (We assume that in the counterfactual individuals face a linear schedule with a mean tax rate of 27.21 percent.)

For a given counterfactual earnings level z_{0i} , we calculate optimal frictionless earnings z_{1i}^* in period 1, and we calculate whether the individual with counterfactual earnings z_{0i} wishes to adjust her earnings from the frictionless optimum because the gains from doing so outweigh the adjustment cost. (Optimal "frictionless" earnings refers to the individual's optimal earnings in the absence of adjustment costs.) We then determine the individual's optimal frictionless earnings z_{2i}^* under the new tax schedule in period 2. We assess whether given the adjustment cost, the individual obtains higher utility by staying at her period 1 earnings level, or by paying the adjustment cost and moving to a new earnings level in period 2.

We perform these calculations alternatively under the assumptions that (a) the elasticity ε is 0.23 and the adjustment cost ϕ is \$152.08 (our baseline estimates); or (b) the elasticity ε is 0.23 and the adjustment cost ϕ is zero. Thus, our simulations illustrate the difference between incorporating adjustment costs and not incorporating them, holding the elasticity constant. (As we have noted, the Saez (2010) estimate of the elasticity in the cross-sections prior to 1990 is very similar to our elasticity estimate; for example, the estimated Saez (2010) elasticity in 1989 is 0.22, which is very similar to our elasticity estimate of 0.23. Thus, the comparison of (a) and (b) above can be considered similar to the comparison between the estimates we would make if we used only information from a single cross-section prior to the policy change to estimate the elasticity using the Saez (2010) method, and compared the resulting prediction to the prediction of our model with adjustment costs.)

Under these alternative assumptions, we can perform a number of experiments to simulate the effects of changing the effective tax schedule. These calculations are shown in Appendix Table G.3 below.

We calculate that if the marginal tax rate above the exempt amount were reduced by 14.67 percentage points, so that the tax rate above the exempt amount were reduced from 77.21 percent to 62.71 percent, mean earnings in the population under consideration would be unchanged at \$9,873.1 under our baseline estimates of the elasticity and adjustment cost. In this case, adjustment is not optimal for anyone when we assume the adjustment cost. In fact, earnings would be unchanged for any reduction in the marginal tax rate above the exempt amount up to 14.67 percentage points; 14.67 percentage points is the largest percentage point marginal tax rate decrease above the exempt amount for which there is no adjustment. Since the gains are second-order near the kink, even a modest adjustment cost

²⁵As we note elsewhere, 62-64 year-olds technically face a notch in the budget constraint at the exempt amount, as opposed to a kink. However, we find no evidence that they behave as if they faced a notch, as the earnings distribution for this age group 1) does not show bunching just above the exempt amount and 2) does not show a "hole" in the earnings distribution just under the exempt amount.

of \$152.08 prevents adjustment with an 14.67 percentage point (or smaller) cut in marginal tax rates. By contrast, when assuming $\varepsilon=0.23$ and $\phi=0$, we predict that mean earnings would rise from \$9,873.1 to \$10,502.6, an increase of 6.38 percent.

At the same time we calculate that if the 50 percent AET above the exempt amount were eliminated, so that the tax rate above the exempt amount were reduced from 77.21 percent to 27.21 percent, mean earnings in the population under consideration would rise from \$9,873.10 to \$11,607.70, or 17.57 percent, under our baseline estimates of the elasticity and adjustment cost. When assuming $\varepsilon=0.23$ and $\phi=0$, we predict that mean earnings would rise from \$9,873.10 to \$11,639.30, a nearly identical increase of 17.89 percent. The slight discrepancy between the two estimates arises because there are individuals whose counterfactual earnings is just above the exempt amount who choose to adjust without adjustment costs, but for whom the gains from adjustment do not outweigh the adjustment cost when we assume the friction.

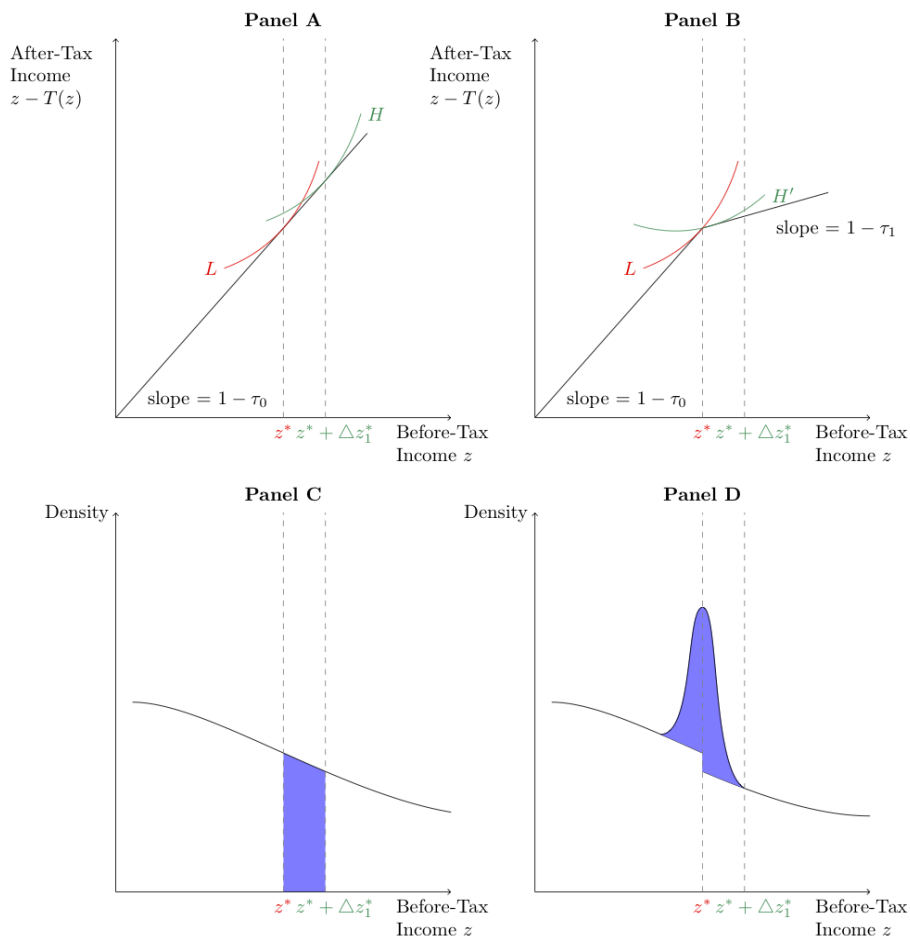
Note that thus far we have analyzed the effect of eliminating the additional 50 percent marginal tax rate above the exempt amount, given that it was previously imposed. An alternative experiment compares earnings if we never imposed the additional 50 percent tax above the kink, to earnings if this tax were imposed. We find that mean earnings are 15.17 percent lower if we impose the tax than if we never imposed it (both when we do and do not assume adjustment costs).

It is worth noting an additional caveat to these results: they apply to those with counterfactual earnings in the range from \$10,000 below to \$10,000 above the exempt amount. If we allowed unbounded counterfactual earnings, there would be some individuals with very large counterfactual earnings for whom the gains from adjustment would outweigh the adjustment cost, even in the presence of adjustment costs. However, this is less relevant to the AET because as we have noted, the OASI benefit phases out entirely at very high earnings levels. Moreover, considering such individuals would involve extrapolating the estimates much farther out of sample. Finally, the results are robust to considering other earnings ranges within the range we measure in our study; for example, when we examine the range of individuals earning from \$10,000 below to \$30,000 above the exempt amount, we still find that the maximum tax cut that leads to no earnings change is quite substantial (10.29 percentage points).

In fact, under all of the other choices we have explored, the results always show that the maximum tax cut that leads to no earnings change is quite substantial (and larger than the changes in marginal tax rates envisioned in most tax reform proposals)—including when we use other ages to specify the counterfactual earnings density; use a different baseline marginal tax rate; and use the constrained estimate of the elasticity (0.39) when performing the simulations (which actually leads to still starker results). When we use the elasticity and adjustment cost estimated from the disappearance of the kink at age 70 (as shown in 3), we find that the largest cut in the marginal tax rate above the exempt amount that leads to no change in earnings is 10.30 percentage points when we examine the earnings range from \$10,000 below to \$10,000 above the exempt amount (and is 7.28 percentage points when we examine the earnings range from \$10,000 below to \$30,000 above the exempt amount).

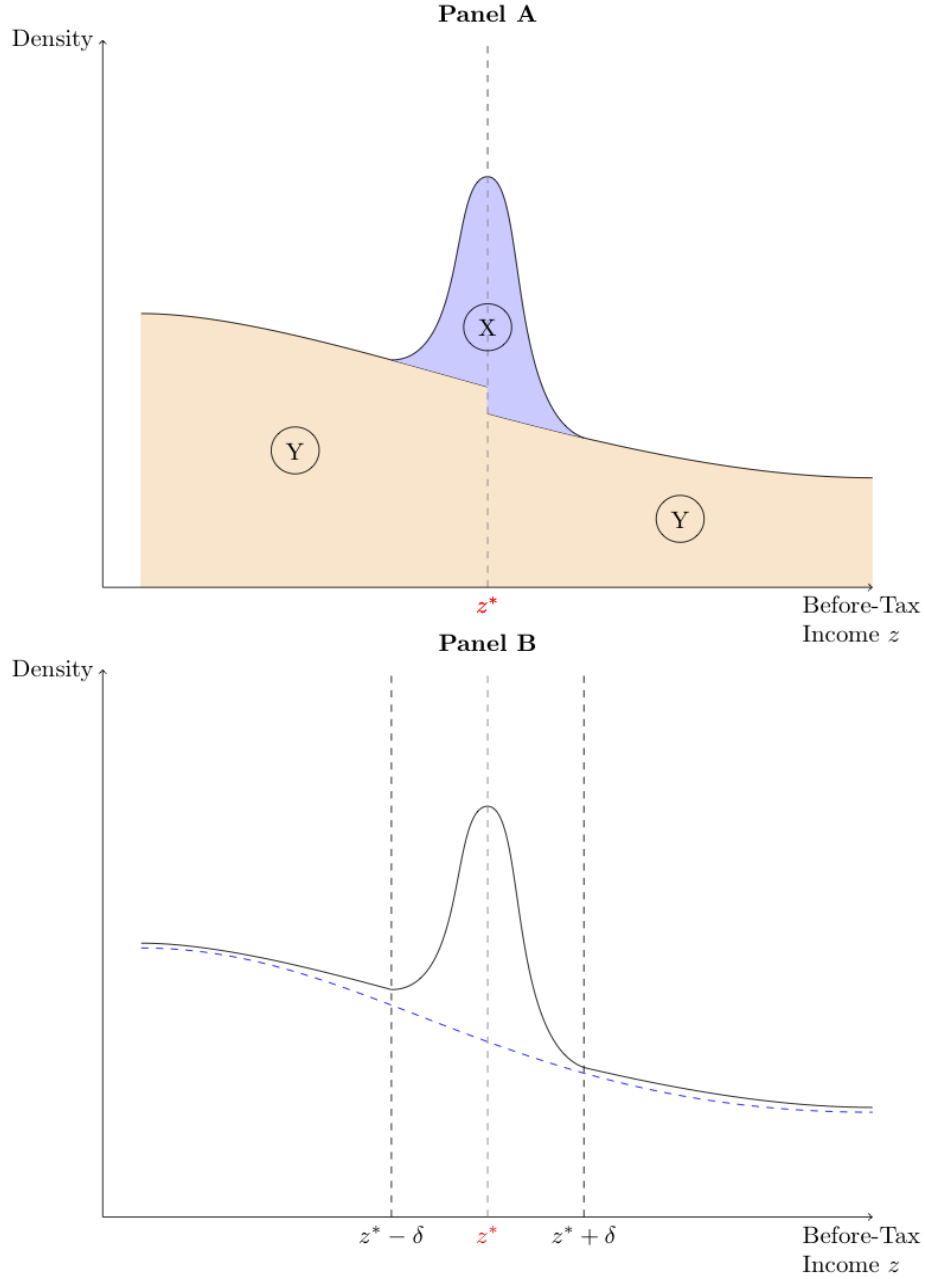
F Appendix: Additional Figures (for online publication)

Figure F.1: Bunching Response to a Convex Kink (Frictionless Case)



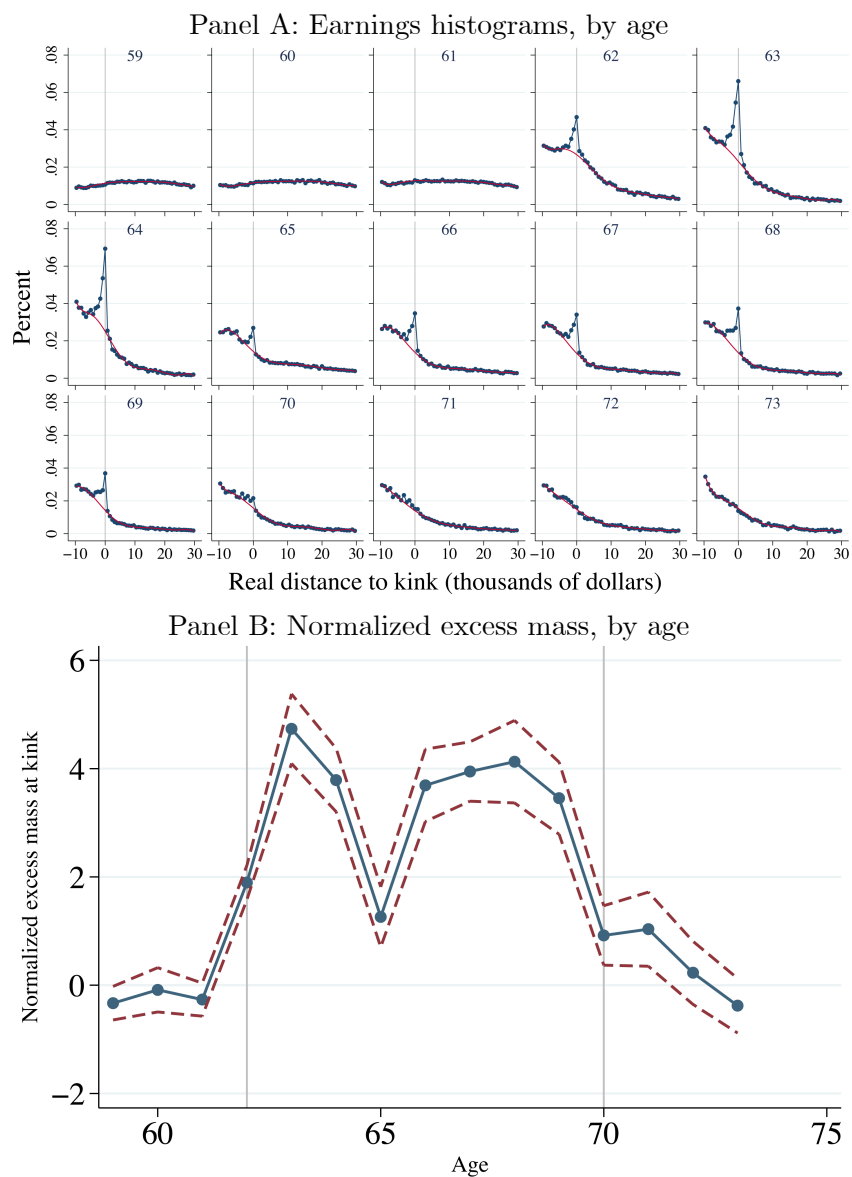
Note: When we move from a linear budget constraint (Panel A) to a convex kink (B), individuals with initial earnings between z^* and $z^* + \Delta z^*$ relocate to the kink. As we move from a linear budget constraint (Panel C) to a convex kink (Panel D), a spike in the earnings density appears at the kink, corresponding to the density that was initially located between z^* and $z^* + \Delta z^*$. The spike is spread out in the vicinity of the kink in Panel D; this may result from several factors discussed in Saez (2010), such as inability to control earnings precisely.

Figure F.2: Bunching Estimation Methodology



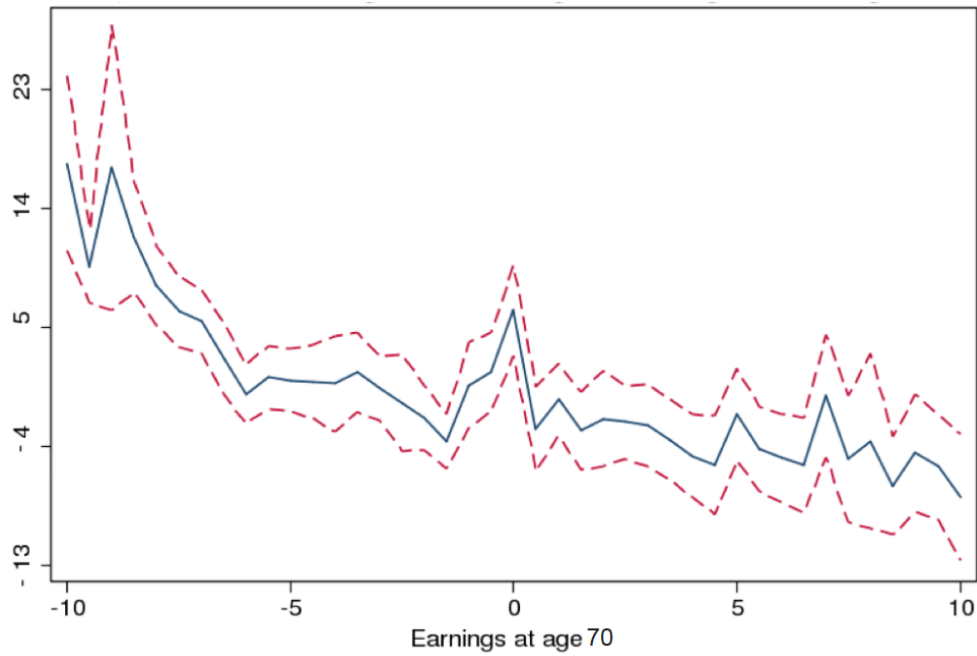
Note: Panel A decomposes the ex-post earnings distribution shown in Appendix Figure F.1 Panel D into two groups. The bunchers, group X , are those who bunch at the kink in the presence of the higher marginal tax rate $\tau + d\tau$ but not at the lower marginal tax rate τ . The non-bunchers, group Y , are comprised of those who locate to the left of the kink under the initial lower marginal tax rate τ , and those who locate to the right of the kink under the higher marginal tax rate $\tau + d\tau$. Panel B demonstrates how the distribution of earnings in the absence of the kink is estimated to recover the share of bunchers, by excluding data in a neighborhood of z^* .

Figure F.3: Adjustment Across Ages: Histograms of Earnings and Normalized Excess Mass, 59-73-year-old OASI Claimants, 1990-1999



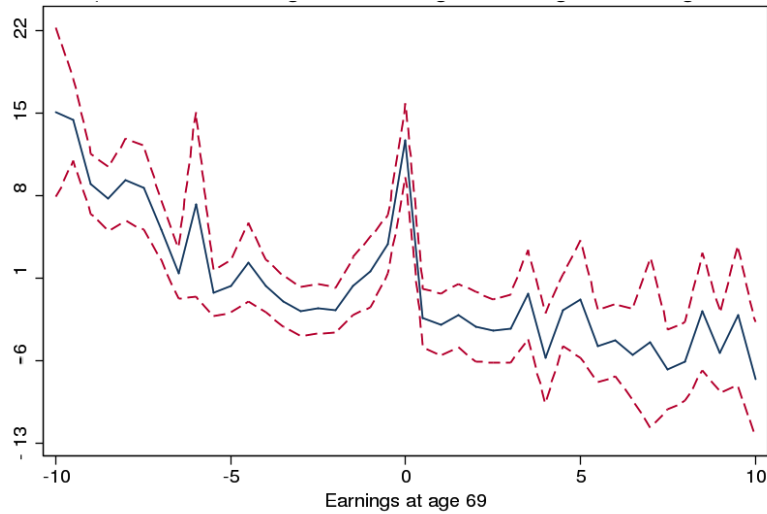
See notes to Figures 2 and 3. This figure differs from Figures 2 and 3 only because the sample in year t consists only of people who have claimed OASI in year t or before (whereas in Figures 2 and 3 it consists of all those who claimed by age 65).

Figure F.4: Mean Percentage Change in Earnings from Age 70 to 71, by Earnings at 70, 1990-1998



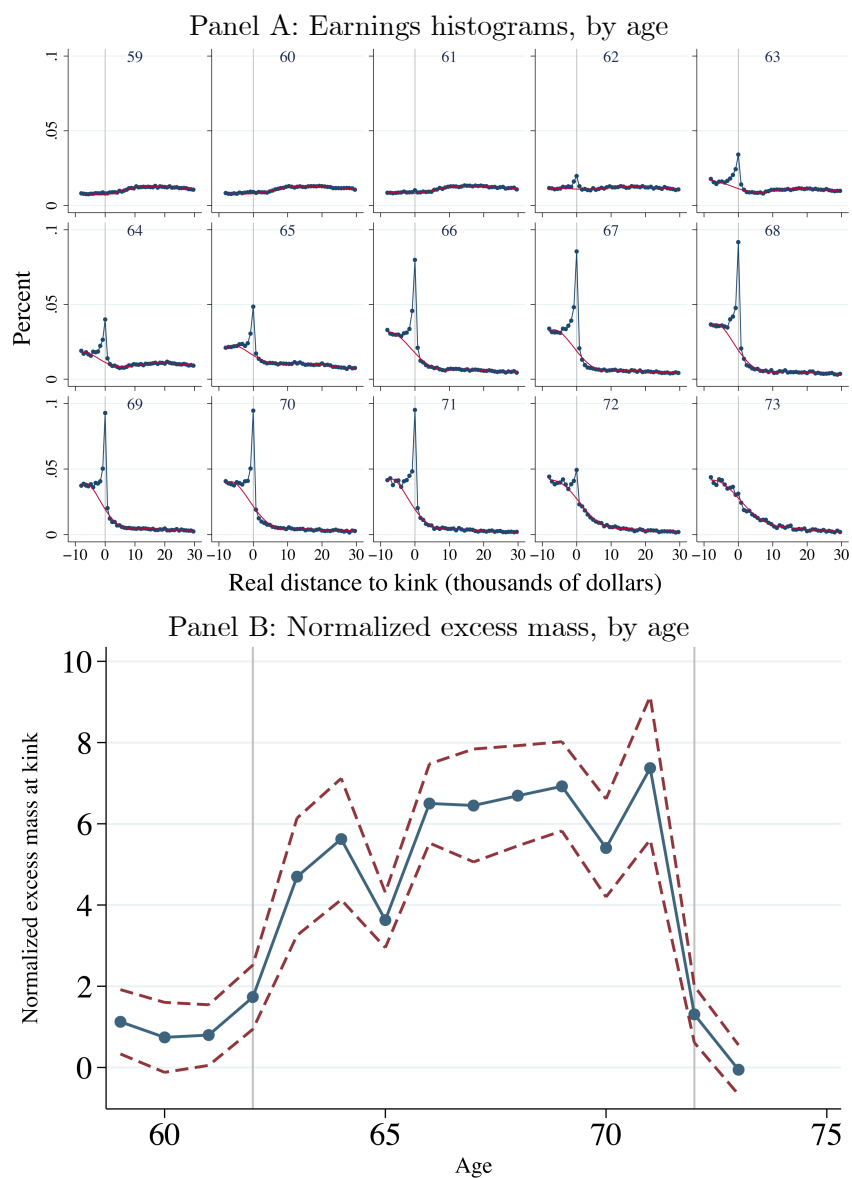
Note: The figure shows the mean percentage change in earnings from age 70 to age 71 (y-axis), against earnings at age 70 (x-axis). Dashed lines denote 95% confidence intervals. Earnings are measured relative to the kink, shown at zero on the x-axis. The data are a 20 percent random sample of 70-year-olds in the LEHD in 1990-1998. We exclude 1999 as a base year in this and similar graphs because the AET is eliminated for those older than NRA in 2000. Higher earnings growth far below the kink reflects mean reversion visible in this part of the earnings distribution at all ages. We also find a spike in mean earnings growth from age 70 to age 71 among those born in January.

Figure F.5: Mean Percentage Change in Earnings from Age 69 to 70, by Earnings at 69, 1990-1998



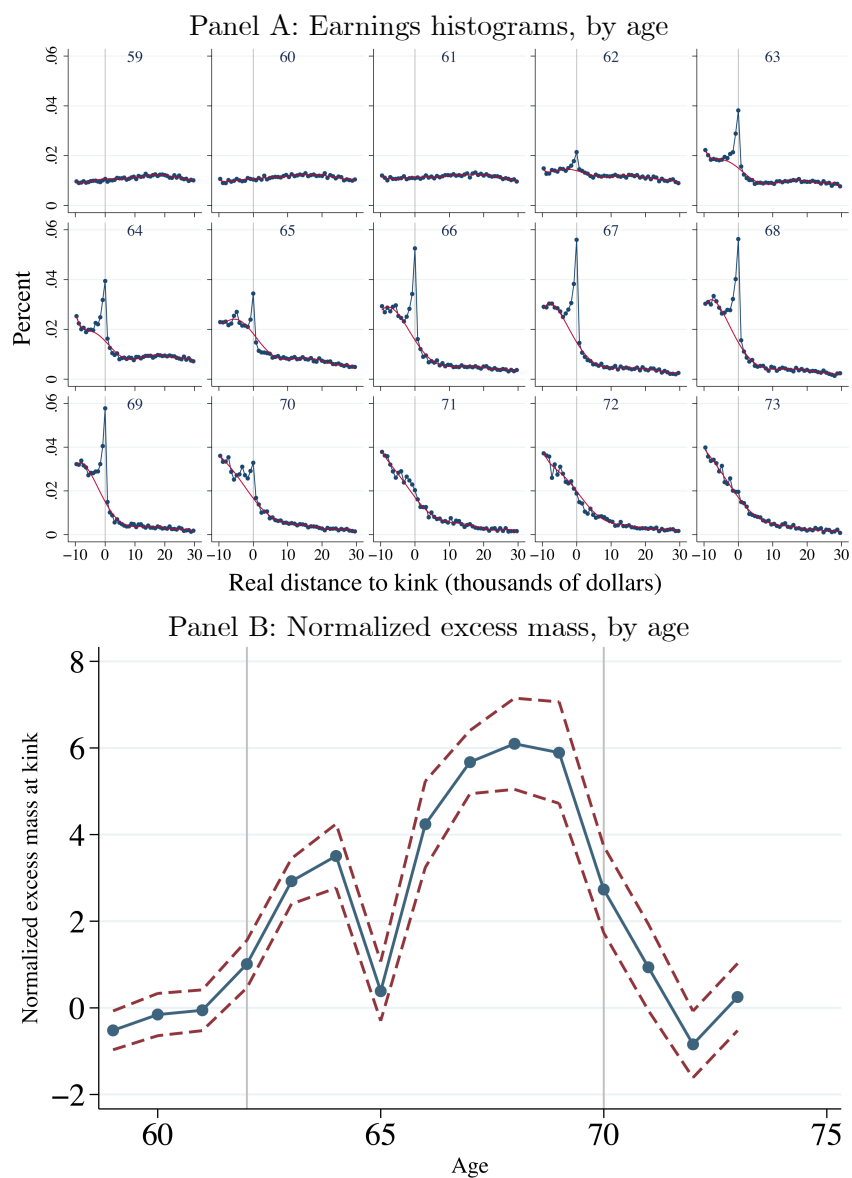
Note: The figure shows the mean percentage change in earnings from age 69 to age 70 (y-axis), against earnings at age 69 (x-axis). Dashed lines denote 95% confidence intervals. Earnings are measured relative to the kink, shown at zero on the x-axis. The data are a 20 percent random sample of 69-year-olds in the LEHD in 1990-1998. We exclude 1999 as a base year in this and similar graphs because the AET is eliminated for those older than NRA in 2000.

Figure F.6: Adjustment Across Ages: Histograms of Earnings and Normalized Excess Mass, 59-73-year-olds Claiming OASI by Age 65, 1972-1982



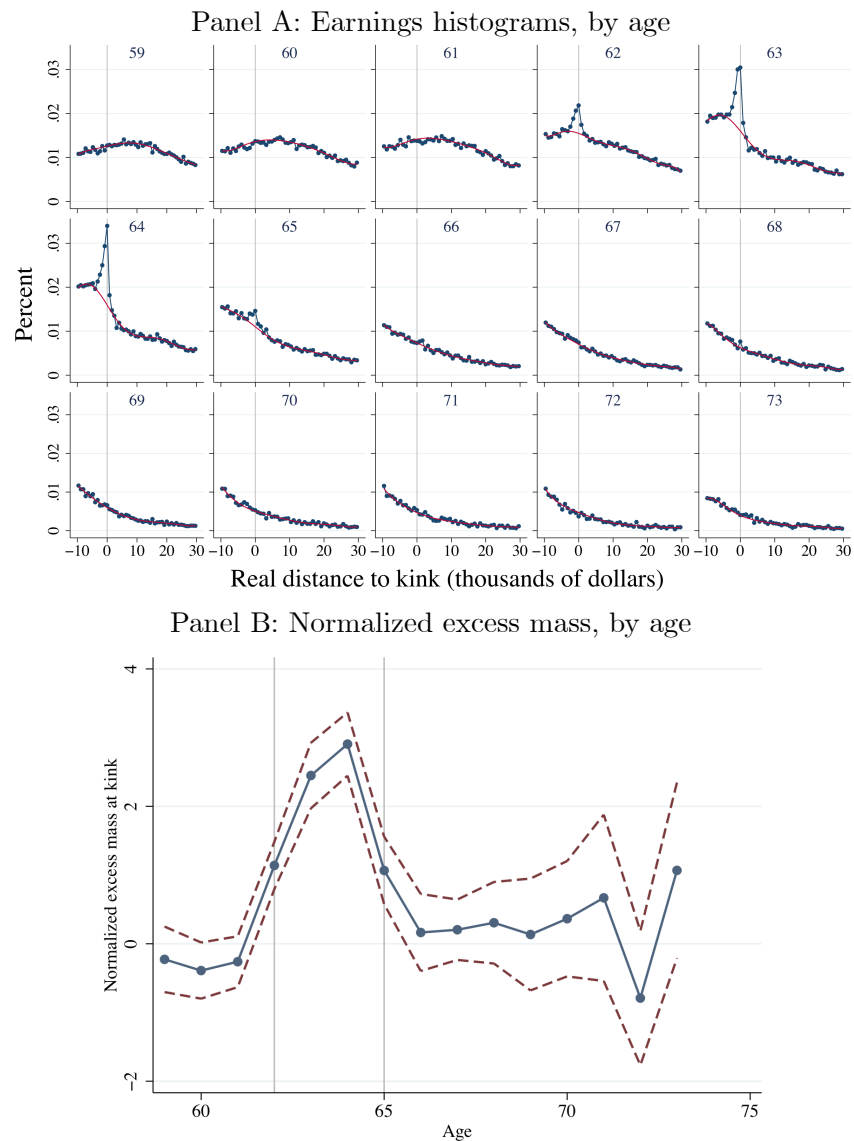
See notes to Figures 2 and 3. This figure differs from Figures 2 and 3 only because the years examined are 1972-1982 (whereas in Figures 2 and 3 the years examined are 1990-1999).

Figure F.7: Adjustment Across Ages: Histograms of Earnings and Normalized Excess Mass, 59-73-year-olds Claiming OASI by Age 65, 1983-1989



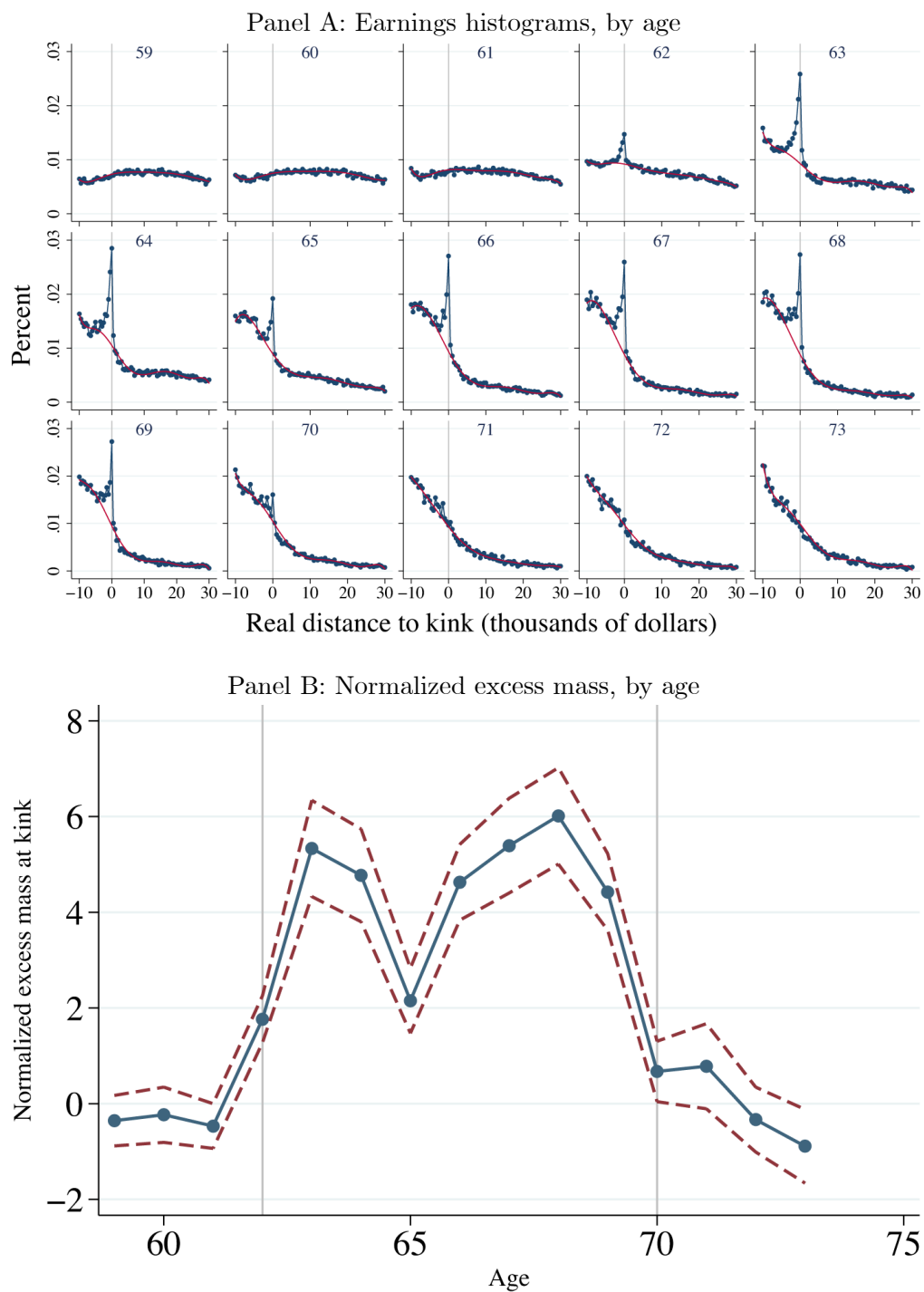
See notes to Figures 2 and 3. This figure differs from Figures 2 and 3 only because the years examined are 1983-1989 (whereas in Figures 2 and 3 the years examined are 1990-1999).

Figure F.8: Adjustment Across Ages: Histograms of Earnings and Normalized Excess Mass, 59-73-year-olds Claiming OASI by Age 65, 2000-2006



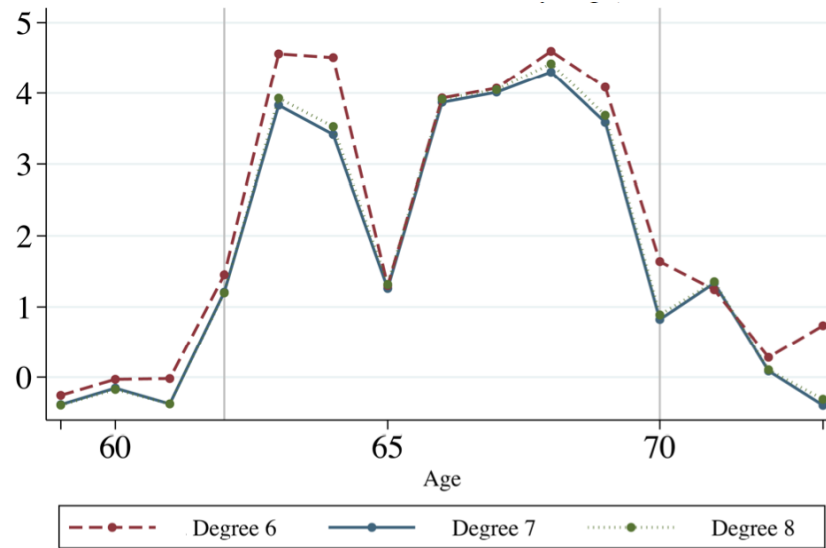
See notes to Figures 2 and 3. This figure differs from Figures 2 and 3 only because the years examined are 2000-2006 (whereas in Figures 2 and 3 the years examined are 1990-1999). As explained in the main text, the NRA slowly rose from 65 for cohorts that reached age 62 during this period; the results are extremely similar when the sample is restricted to those who claimed by 66, instead of 65. In the year of attaining NRA, the AET applies for months prior to such attainment.

Figure F.9: Adjustment Across Ages: Histograms of Earnings and Normalized Excess Mass, 59-73-year-olds Claiming OASI by Age 65, 1990-1999



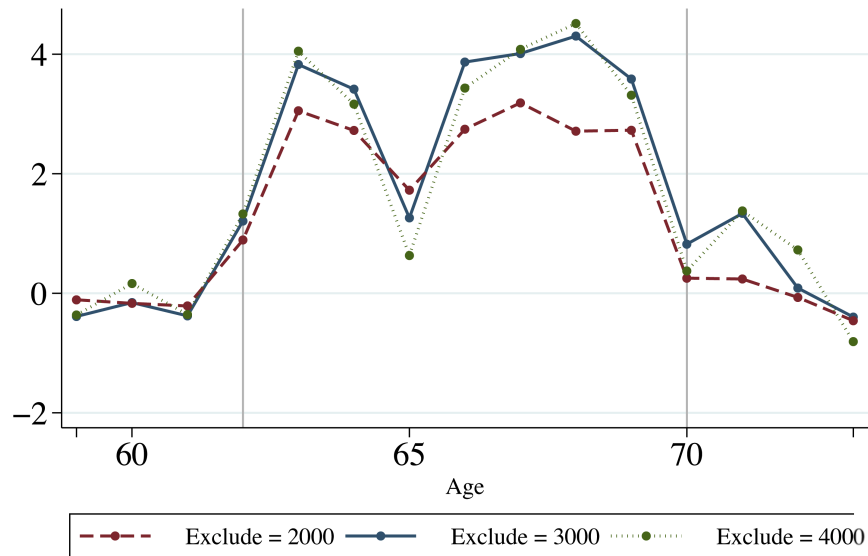
See notes to Figures 2 and 3. This figure differs from Figures 2 and 3 only because the bandwidth is \$500 (whereas in Figures 2 and 3 it is \$800).

Figure F.10: Robustness to Polynomial Degree: Normalized Excess Mass by Age, OASI Claimants by 65, 1990-1999



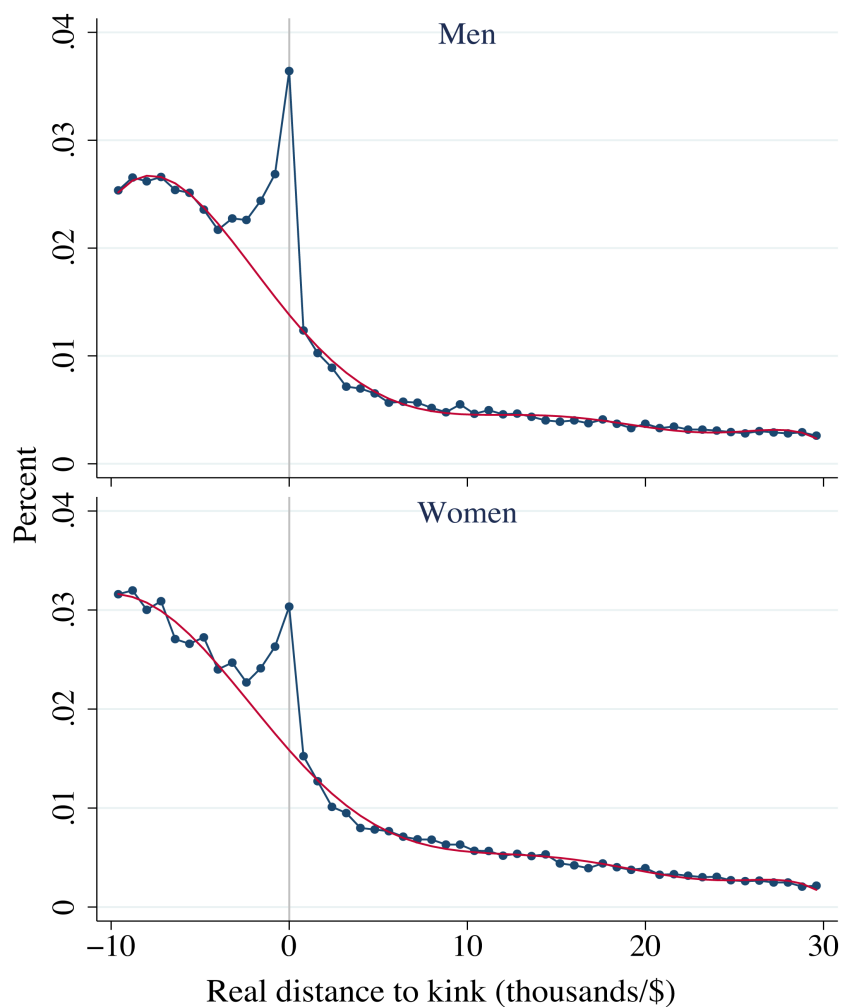
Notes: The figure shows the difference in estimates of normalized excess mass as we vary the degree of the polynomial used. For additional notes on the samples see Figure 3.

Figure F.11: Robustness to the Excluded Region: Normalized Excess Mass by Age, OASI Claimants by 65, 1990-1999



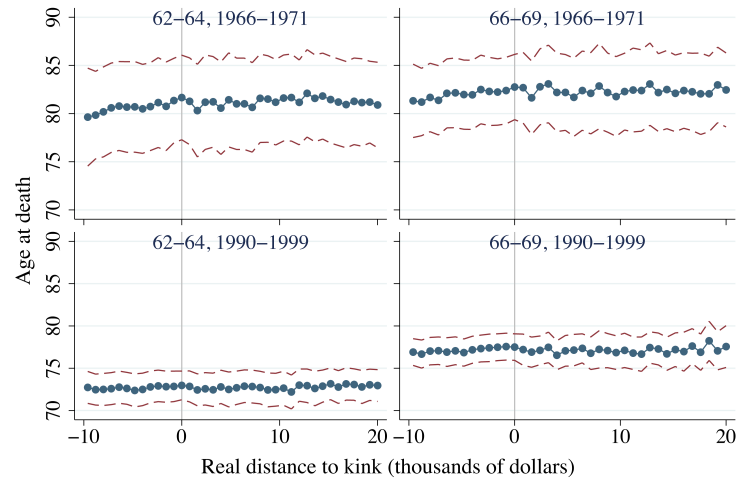
Notes: The figure shows the difference in estimates of normalized excess mass as we vary the region about the kink that is "dummied out" in the polynomial estimation. For additional notes on the samples see Figure 3.

Figure F.12: Adjustment by Sex: Histograms of Earnings, 59-73-year-olds Claiming OASI by Age 65, 1990-1999



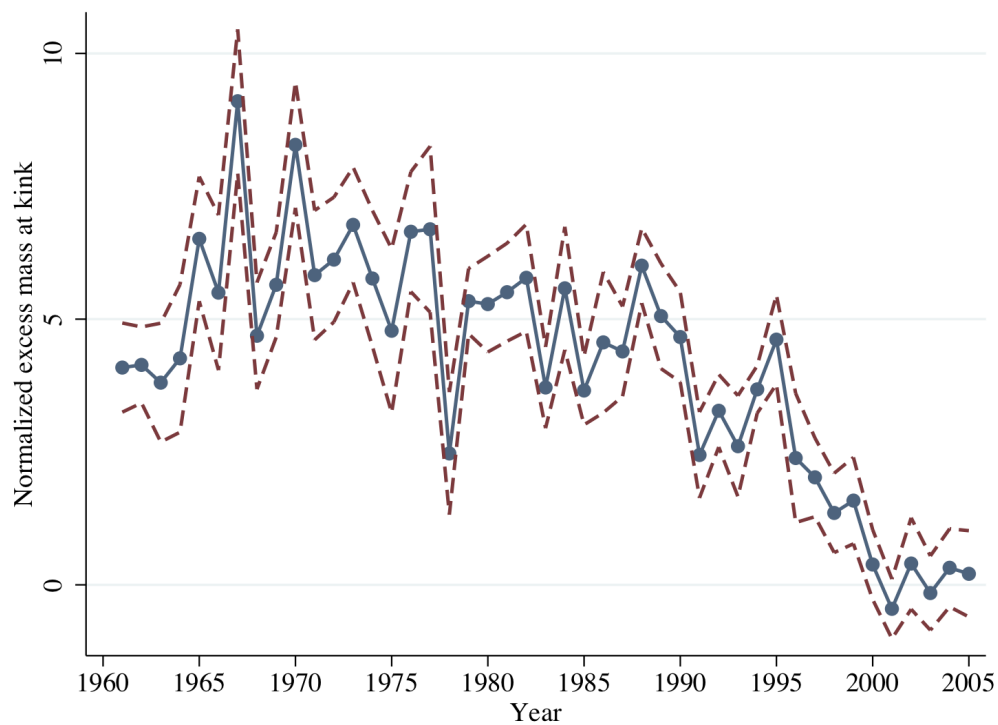
See notes to Figure 2. The sample examined is the same as in Figure 2 but examines men and women separately.

Figure F.13: Mortality Analysis: Mean Age at Death, 62-69-year-olds Claiming OASI by Age 65, 1966-1971 and 1990-1999



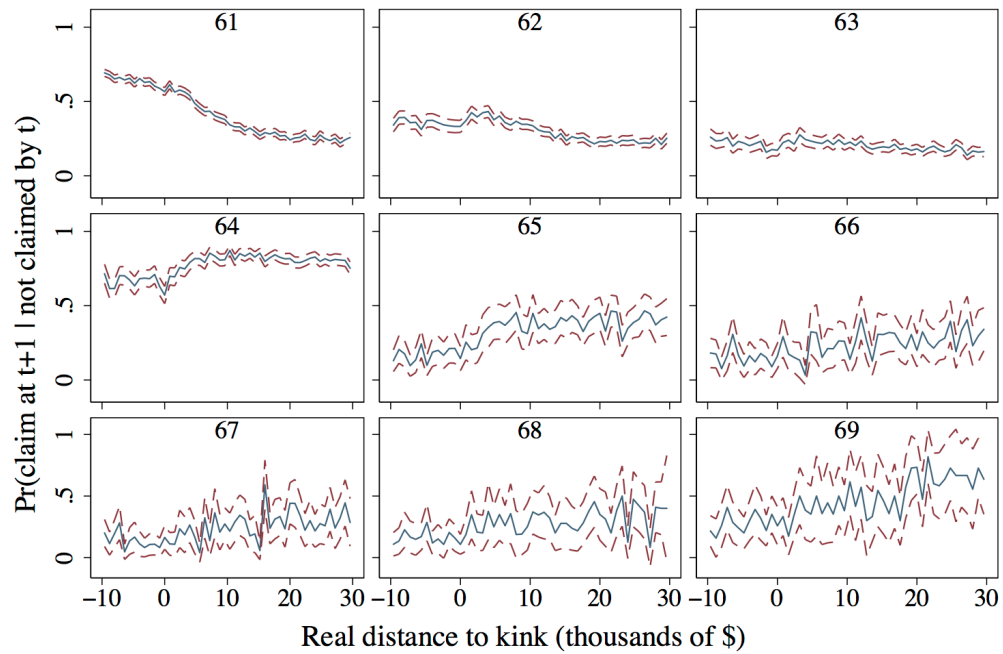
Note: The figure shows mean age at death from a one percent random sample of SSA administrative data on individuals aged 59-73, claiming OASI by age 65, between 1966 and 1971 (inclusive) in the top panel, and between 1990 and 1999 (inclusive) in the bottom panel. Dashed lines denote 95% confidence intervals. The figure shows no clearly noticeable patterns at the kink that are different from those away from the kink. This holds true for those 62-64 and 66-69, in a period prior to the introduction of the Delayed Retirement Credit (*i.e.* 1966-1971) and subsequent to its introduction (*i.e.* 1990-1999). Results are similar for other time periods.

Figure F.14: Normalized Excess Mass by Year, 1961-2005



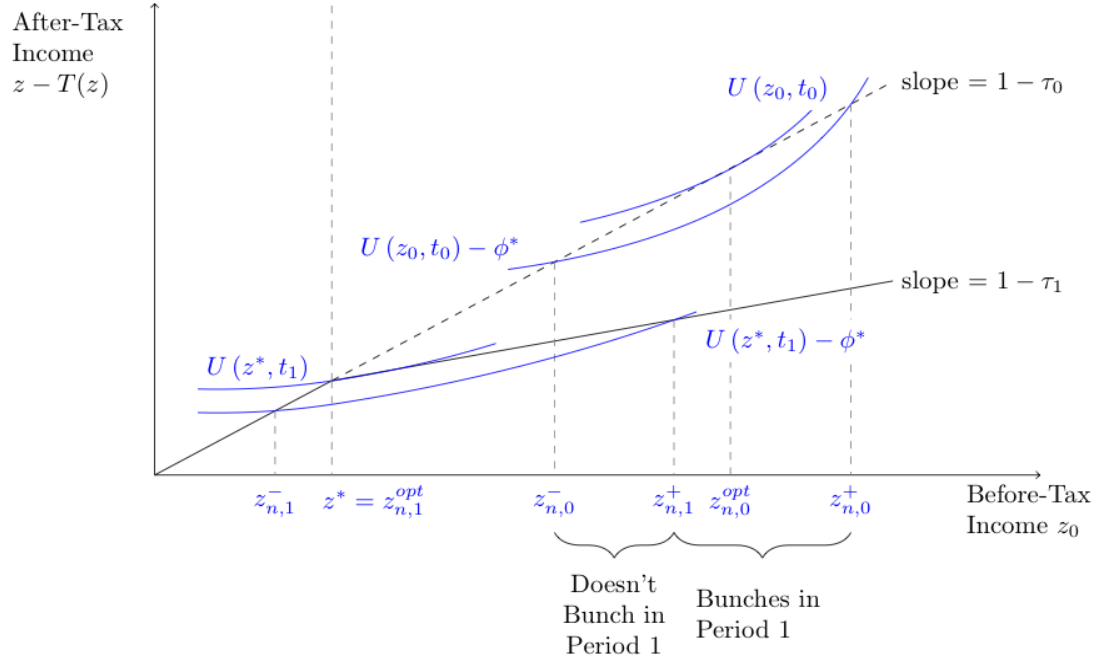
Note: The figure shows normalized excess mass from a one percent random sample of SSA administrative data on Social Security claimants aged 66-69 in each year between 1961 and 2005. See other notes to Figure 3.

Figure F.15: Probability of claiming OASI in year $t+1$ among 61-69 year-olds in year t who are not claiming, 1990-1998



Note: The figure shows the probability that an individual claims OASI in year $t + 1$, conditional on not claiming OASI in year t , for those ages 61-68 in year t from 1990-1998. Dashed lines denote 95% confidence intervals.

Figure F.16: Bunching Response to a Convex Kink, with Frictions in Initial Earnings



Note: See Section E.9 for an explanation of the figure.

G Appendix: Additional Estimates of ε and ϕ (for on-line publication)

Table G.1: Estimates of Elasticity and Adjustment Cost Using 1990 Policy Change, Assuming no Pre-Period Bunching Attenuation

	(1)	(2)	(3)	(4)
	ε	ϕ	$\varepsilon \phi = 0$	
			1990	1989
Baseline	0.22 [0.18, 0.25]***	\$139.53 [50.17, 2065.89]***	0.39 [0.33, 0.48]***	0.22 [0.18, 0.25]***
Nonparametric Density	0.24 [0.20, 0.29]***	\$153.32 [50.25, 1035.99]***	0.43 [0.36, 0.54]***	0.24 [0.20, 0.28]***
Benefit Enhancement	0.37 [0.31, 0.43]***	\$76.24 [21.07, 32655.05]***	0.59 [0.49, 0.73]***	0.37 [0.30, 0.43]***
Excluding FICA	0.30 [0.25, 0.35]***	\$119.84 [38.84, 23383.88]***	0.50 [0.42, 0.62]***	0.30 [0.25, 0.35]***
Bandwidth = \$500	0.24 [0.20, 0.30]***	\$83.80 [5.86, 278.82]***	0.38 [0.30, 0.50]***	0.24 [0.19, 0.30]***

Note: The table examines the 1990 policy change, using data from 1989 and 1990, but assumes that bunching in 1989 is not attenuated by adjustment frictions. The constrained estimate of bunching using data only from 1989 is mechanically the same as the unconstrained estimate, as both rely on the Saez (2010) formula for bunching. See also notes to Appendix Table 3.

Table G.2: Estimates of Elasticity and Adjustment Cost 1990 Policy Change, Assuming Pre-Period Bunching may not be at Frictionless Optimum

	(1)	(2)	(3)	(4)
	ε	ϕ	$\varepsilon \phi = 0$	
			1990	1989
Baseline	0.24 [0.22, 0.39]***	\$162.26 [55.27, 1556.01]***	0.39 [0.33, 0.48]***	0.22 [0.18, 0.25]***
Benefit Enhancement	0.40 [0.36, 0.53]***	\$86.45 [20.31, 561.81]***	0.59 [0.49, 0.73]***	0.37 [0.30, 0.43]***
Excluding FICA	0.33 [0.29, 0.41]***	\$139.03 [38.30, 525.91]***	0.50 [0.42, 0.62]***	0.30 [0.25, 0.35]***
Bandwidth = \$500	0.26 [0.22, 0.33]***	\$104.10 [7.01, 440.80]***	0.38 [0.30, 0.50]***	0.24 [0.19, 0.30]***
Nonparametric Density	0.28 [0.25, 0.43]***	\$187.78 [58.97, 1303.68]***	0.43 [0.36, 0.54]***	0.24 [0.20, 0.28]***

Note: The table examines the 1990 policy change, using data from 1989 and 1990, but assumes that bunching in 1989 may not be at the frictionless optimum, as described in the text. See also notes to Appendix Table 3.

Table G.3: Policy Simulations

	(1)	(2)
Panel A: Eliminate AET for 62-64 year olds		
	With adjustment costs	Without adjustment costs
Period 1 mean earnings	\$9,873.1	\$9,873.1
Period 2 mean earnings	\$11,607.7	\$11,639.3
Mean earnings change	\$1,734.6	\$1,766.2
Percent earnings change	17.57	17.89
Share who adjust	46.4	50.4
Mean change among adjusters	\$3,736.5	\$3,506.1
Percent change among adjusters	27.9	26.2
Panel B: Reduce AET BRR by 14.67 percentage points		
	With adjustment costs	Without adjustment costs
Period 1 mean earnings	\$9,873.1	\$9,873.1
Period 2 mean earnings	\$9,873.1	\$10,502.6
Mean earnings change	\$0	\$629.5
Percent earnings change	0	6.38
Share who adjust	0.0	41.9
Mean change among adjusters	0.0	\$1,502.9
Percent change among adjusters	0.0	11.0

Note: Each panel shows the results of a different policy simulation. Column 1 shows the results when we assume $\varepsilon = 0.23$ and $\phi = \$152.08$, and Column 2 shows the results when we assume $\varepsilon = 0.23$ and $\phi = 0$. "Mean earnings change" refers to the change in mean earnings from Period 1 to Period 2 predicted in the full study population (i.e. the population with counterfactual earnings between -\$10,000 below and \$10,000 above the exempt amount). "Percent earnings change" is the percent change in mean earnings predicted in the full study population. "Share who adjust" refers to the percent of the full study population whose earnings does not change in response to the policy change. Note that only 50.4 percent of the full study population has counterfactual earnings above the exempt amount and therefore has incentives that are potentially affected by the policy change in our model. "Mean change among adjusters" refers to the change in mean earnings predicted among those who change earnings in response to the policy change. "Percent change among adjusters" refers to the percent change in mean earnings among those who change earnings in response to the policy change. See Appendix E.11 for further explanation.