

NBER WORKING PAPER SERIES

ESTIMATING EARNINGS ADJUSTMENT FRICTIONS:
METHOD AND EVIDENCE FROM THE SOCIAL SECURITY EARNINGS TEST

Alexander M. Gelber
Damon Jones
Daniel W. Sacks

Working Paper 19491
<http://www.nber.org/papers/w19491>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138
October 2013, Revised January 2019

Previously circulated as "Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test" and "Estimating Earnings Adjustment Frictions: Method and Evidence from the Social Security Earnings Test." We thank Raj Chetty, Jim Cole, Mark Duggan, John Friedman, Bill Gale, Hilary Hoynes, Henrik Kleven, Camille Landais, Olivia Mitchell, Daniel Reck, Emmanuel Saez, Chris Walters, and numerous seminar participants for helpful comments. 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 the Steven H. Sandell Grant Program through the Boston College Center for Retirement Research, from NIH grant #1R03 AG043039-01, from a National Science Foundation Graduate Research Fellowship under Grant No. DGE-0822219, 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. All results have been reviewed to ensure that no confidential information is disclosed. All errors are our own.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w19491.ack>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Alexander M. Gelber, Damon Jones, and Daniel W. Sacks. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimating Earnings Adjustment Frictions: Method and Evidence from the Social Security Earnings Test

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

NBER Working Paper No. 19491

October 2013, Revised January 2019

JEL No. H20,H31,J14,J26

ABSTRACT

We introduce a method for estimating the cost of adjusting earnings, as well as the earnings elasticity. Our method uses information on bunching in the earnings distribution at convex budget set kinks before and after policy-induced changes in the magnitude of the kinks: the larger is the adjustment cost, the smaller is the absolute change in bunching from before to after the policy change. In the context of the Social Security Earnings Test, our results demonstrate that the short-run impact of changes in the effective marginal tax rate can be substantially attenuated.

Alexander M. Gelber
Department of Economics
University of California at San Diego
9500 Gilman Dr.
San Diego, CA 92093
and NBER
gelber@gmail.com

Daniel W. Sacks
Indiana University
1309 E. 10th Street
Bloomington, IN 47405
dansacks@indiana.edu

Damon Jones
Harris School of Public Policy
University of Chicago
1155 East 60th Street
Chicago, IL 60637
and NBER
damonjones@uchicago.edu

1 Introduction

When a policy changes, do frictions prevent economic behavior from adjusting? How long does it take for economic agents to overcome these frictions? Does it appear that long-run responses are larger than short-run responses? How do we measure the magnitude of these frictions, as well as agents' underlying responsiveness in the absence of frictions? This paper develops and implements methods to answer these questions.

We must account for such frictions to estimate the earnings or labor supply response to taxation, our key context of interest. Adjustment frictions in this context could encompass several factors, including a lack of knowledge of a tax regime, the cost of negotiating a new contract with an employer, or the time and financial cost of job search. In a cross-section of data, frictions attenuate the response to taxation (Chetty *et al.*, 2011; Chetty, 2012; Chetty *et al.*, 2013; Kleven and Waseem, 2013). For example, wage-earners typically do not “bunch” in the earnings distribution at many convex budget set kinks, as they should in the absence of frictions (Saez, 2010).

Looking over time, it has been postulated that long-run responses are significantly larger than the short-run responses that are typically measured, due to frictions that impede adjustment in the short run (Saez 2010; Saez *et al.*, 2012). This could help explain patterns in the data like the slow rise in retirement at age 62 subsequent to the introduction of the Social Security Early Retirement Age (Gruber and Wise, 2013). Such attenuated and slow responses matter to policy-makers, who often wish to estimate the timing of the earnings or labor supply reaction to changes in tax and transfer policies, as well as the magnitude of long-run responses beyond the short-run empirical estimation windows typically examined (*e.g.* Congressional Budget Office, 2009). However, the existing literature has not yet developed a method for estimating earnings adjustment frictions, or their implications for the speed of adjustment or the estimation of long-run elasticities.

We make three main contributions to understanding adjustment frictions in the earnings context. First, we introduce a method for documenting adjustment frictions and estimating the amount of time it takes to adjust fully to policy changes. In the absence of adjustment frictions, the removal of a convex kink in the effective tax schedule should result in the

immediate dissolution of bunching at the former kink; thus, any observed delay in reaching zero bunching should reflect adjustment frictions. The time delay reveals the speed of adjustment. We implement this in the context of a kink, but the method applies equally to the context of a notch.

Second, formalizing and generalizing this insight, we specify a model of earnings adjustment that allows us to estimate adjustment costs and the elasticity of earnings with respect to the effective net-of-tax rate.¹ Adding adjustment frictions to the model of Saez (2010), we develop tractable methods that allow the estimation of elasticities and adjustment costs. Our starting point is the context of a kinked budget set. When tax rates change around a kink in our framework, *ceteris paribus* the absolute change in the amount of bunching is decreasing in the adjustment cost, while the initial amount of bunching is increasing in the elasticity. We extend our method to the dynamic case, to estimate the speed of arrival of adjustment opportunities along with the elasticity and adjustment cost. We focus on the special case of fixed adjustment costs, but we address how to estimate adjustment costs with any polynomial functional form.

Third, we apply our methods to estimate these parameters and document adjustment frictions in the context of the U.S. Social Security Annual Earnings Test (“Earnings Test”). The Earnings Test reduces Social Security Old Age and Survivors Insurance (“Social Security”) benefits in a given year as a proportion of a Social Security claimant’s earnings above an exempt amount in that year. For example, for Social Security claimants under age 66 in 2019, current Social Security benefits are reduced by one dollar for every two dollars earned above \$17,640. Previous literature has found that Social Security claimants bunch at this convex kink (Burtless and Moffitt, 1985; Friedberg, 1998, 2000; Song and Manchester, 2007; Engelhardt and Kumar, 2014). In addition to providing a laboratory for studying adjustment costs and earnings elasticities, the Earnings Test is important to policy-makers in its own right. In the latest year of the available micro-data in 2003, the Earnings Test led to an estimated total of \$4.3 billion in current benefit reductions for around 538,000 beneficiaries, thus substantially affecting benefits and their timing. The importance of the Earnings Test

¹For 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. The net-of-tax rate (or equivalently, net-of-tax share) is defined as one minus the effective marginal tax rate.

is now increasing as the affected age range expands gradually to encompass age 67 for those born in 1960 and later.

Using Social Security Administration (SSA) administrative tax data on a one percent sample of the U.S. population, we document clear evidence of adjustment frictions: after individuals no longer face the Earnings Test, they continue to bunch around the location of the former exempt amount. In a baseline specification, we estimate that the fixed adjustment cost within one year of the policy change is around \$280 (in 2010 dollars). We also estimate that the earnings elasticity with respect to the net-of-tax rate is 0.35. When we allow for dynamic adjustment, these parameter estimates are comparable—the long-run elasticity is 0.36, the adjustment cost is around \$245—and we also estimate that full adjustment occurs only after three years.

Our estimates demonstrate that incorporating adjustment costs can change earnings elasticity estimates significantly. The frictionless Saez (2010) method estimates an average elasticity of 0.19 in our Earnings Test context; our method’s estimate is nearly twice as large. Moreover, simulations based on our parameter estimates show that the adjustment frictions we estimate can greatly attenuate the short-run earnings reaction even to a large change in the effective marginal tax rate, frustrating the goal of affecting short-run earnings as envisioned in many discussions and projections of the effects of tax and transfer policies. The results also suggest that the time frame of three years often used to assess earnings responses to taxation (Gruber and Saez 2002; Saez et al. 2012) appears sufficient to capture long-run responses in our context, in contrast to hypotheses that long-run responses may be much larger.

This paper builds on previous literature that has documented the importance of adjustment frictions but has not yet developed methods for estimating them (Chetty *et al.*, 2011; Chetty, 2012; Chetty *et al.*, 2013). Our method complements Kleven and Waseem (2013), who innovate a static method to estimate elasticities and the share of the population that is inert in the presence of a notch in the budget set. Our method is different in three primary ways. First, our method allows estimation of adjustment cost rather than an inert population share. The adjustment cost is necessary for welfare calculations in many applications (Chetty *et al.*, 2009), and is a structural parameter that can be used to perform counterfac-

tual exercises across different contexts. Second, our basic method developed here applies to kinks (see also the applications of our method in He, Peng, and Wang, 2016; Schächtele 2016; and Mortenson *et al.* 2017), and has been adapted to the case of notches as well to estimate adjustment costs (Gudgeon and Trenkle, 2016; Zaresani, 2016). Third, our dynamic method allows us to estimate the parameters of the gradual adjustment process over time, as well as the speed of adjustment.

Our paper also follows a large existing literature on adjustment costs in areas outside labor and public economics. For example, adjustment costs have long been studied in inventory theory (*e.g.* Arrow *et al.*, 1951, and subsequent literature), macroeconomics (*e.g.* Baumol, 1952, and subsequent literature), firm investment (*e.g.* Abel and Eberly, 1994), durable good consumption (*e.g.* Grossman and Laroque, 1990), pricing and inflation (*e.g.* Sheshinski and Weiss, 1977), and other settings including the “s-S” literature (see literature reviews in Leahy, 2008, or Stokey, 2008). In our paper, changes in non-linear budget sets generate clear changes in bunching that can be mapped to our parameter estimates in a manner that transparently follows the patterns in the data.

The rest of the paper proceeds as follows. Section 2 describes the policy environment. Section 3 presents the method for quantifying bunching. Section 4 describes our data. Section 5 documents on adjustment frictions empirically. Section 6 specifies our model. Section 7 presents our parameter estimates. Section 8 describes simulations based on the estimates. Section 9 concludes. The Appendix contains additional results. More results are available in an earlier working paper version of the present paper (Gelber, Jones, and Sacks, 2013).

2 Policy Environment

Social Security provides annuity income to the elderly and to survivors of deceased workers. Individuals with sufficient years of eligible earnings can claim Social Security benefits through their own earnings history as early as age 62. Individuals in our sample reach the Normal Retirement Age at 65, when they can claim their full Social Security benefits.

Individuals who claim Social Security may keep working, but their earning are subject to the Earnings Test. For each dollar they earn above an exempt amount, their benefits are reduced. Figure 1 shows that the Earnings Test became less stringent over 1961-2009. Prior

to 1989, the benefit reduction rate above the exempt amount was 50 percent. In 1990 and after, the benefit reduction rate fell to 33.33 percent for beneficiaries at or older than 65; this change had been scheduled since the 1983 Social Security Amendments. During our period of interest from 1983 to 1999 period, the Earnings Test applied to Social Security beneficiaries aged 62-69 (prior to 1983, it applied to those 62-71). Starting in 1978, beneficiaries younger than 65 faced a lower exempt amount than those at 65 or above.

When current Social Security benefits are lost to the Earnings Test, future scheduled benefits are increased in some circumstances, which is sometimes called “benefit enhancement.” This can reduce the effective tax rate associated with the Earnings Test. For beneficiaries subject to the Earnings Test aged Normal Retirement Age and older, a one percent Delayed Retirement Credit was introduced in 1972, meaning that each year of foregone benefits led to a one percent increase in future yearly benefits. The Delayed Retirement Credit was raised to three percent in 1982 and gradually rose to eight percent for cohorts reaching Normal Retirement Age from 1990 to 2008. 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 claiming benefits now or delaying claiming and receiving higher benefits once she begins to collect Social Security (Diamond and Gruber, 1999).

The Delayed Retirement Credit only raises claimants’ future benefits when annual earnings are high enough that the Earnings Test reduces at least an entire month’s worth of benefits (Friedberg, 1998; Social Security Administration, 2012a). In particular, an entire month’s benefits are lost—and benefit enhancement occurs—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 Earnings Test benefit reduction rate. 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 ($=\$1,000/0.3333$) above the exempt amount. Although the Earnings Test withholds benefits at the monthly level, the Earnings Test is generally applied based on *annual* earnings—the object we observe in our data. We model the Earnings Test as creating a positive implicit marginal tax rate for some individuals—reflecting the reduction in current benefits—consistent with both the empirical bunching at

Earnings Test kinks and with the practice in previous literature.

For individuals considering earning in a region well above the Earnings Test exempt amount, thus triggering benefit enhancement, the Earnings Test could also affect decisions for several reasons. The Earnings Test was roughly actuarially fair only beginning in the late 1990s. Those whose expected life span is shorter than average should expect to collect Social Security benefits for less long than average, implying that the Earnings Test is more financially punitive. Liquidity-constrained individuals or those who discount faster than average could also reduce work in response to the Earnings Test. Finally, some may not understand the Earnings Test benefit enhancement or other aspects of Social Security (Liebman and Luttmer 2012; Brown, Kapteyn, Mitchell, and Mattox, 2013). We follow previous work and do not distinguish among these potential reasons in our main analysis (Gelber, Jones, and Sacks 2013 analyze certain reasons for the response).

For beneficiaries under Normal Retirement Age, the actuarial adjustment raises future benefits whenever an individual earns over the Earnings Test exempt amount (Social Security Administration, 2012, Section 728.2; Gruber and Orszag, 2003), by 0.55 percent per month of benefits withheld. Thus, beneficiaries in this age range do not face a pure kink in the budget set at the exempt amount. To address this, we limit the sample to ages above Normal Retirement Age in our estimates of elasticities and adjustment costs.

3 Initial Bunching Framework

To understand the effects of the kink created by the Earnings Test, we begin with a model with no frictions to illustrate our technique for estimating bunching at kinks (Saez 2010). Agents maximize utility $u(c, z; a)$ over consumption c and pre-tax earnings z , subject to a budget constraint $c = (1 - \tau)z + R$, where R is virtual income.² Greater earnings are associated with greater disutility due to the cost of effort. The first-order condition, $(1 - \tau)u_c + u_z = 0$, implicitly defines an earnings supply function $z((1 - \tau), R; a)$.

The parameter a reflects heterogeneous “ability,” *i.e.* the trade-off between consumption and earnings supply. Following previous literature, we assume rank preservation in earnings

²We can write $c = z - T(z)$, where $T(z)$ is a general, nonlinear tax schedule. As in the public finance literature (*e.g.* Hausman, 1981) we rewrite the budget constraint in linearized form, $c = (1 - \tau)z + R$, where $\tau \equiv T'(z)$ is the marginal tax rate and $R \equiv T'(z) \cdot z - T(z)$ is virtual income, *i.e.* the intercept of a linear budget set passing through the point $(z, T(z))$.

as a function of a . Thus, a is isomorphic to the level of earnings that would occur in the absence of any tax. a is distributed according a smooth CDF. Under a constant marginal tax rate of τ_0 , this implies a smooth distribution of earnings $H_0(\cdot)$, with pdf $h_0(\cdot)$.

Starting with a linear tax at a rate of τ_0 , suppose the Earnings Test is additionally introduced, so that the marginal net-of-tax rate decreases to $1 - \tau_1$ for earnings above a threshold z^* , where $\tau_1 > \tau_0$. Individuals earning in the neighborhood above z^* reduce their earnings due to the higher tax. If ability is smoothly distributed, a range of individuals initially locating between z^* and $z^* + \Delta z^*$ will “bunch” exactly at z^* , due to the reduced incentive to earn above z^* . In practice, previous literature finds empirically that individuals locate in the neighborhood of z^* , rather than exactly at z^* .

To quantify the amount of bunching, or “excess mass,” we use a technique similar to Chetty *et al.* (2011) and Kleven and Waseem (2013). For each earnings bin z_i of width δ we calculate p_i , the proportion of all people with annual earnings in the range $[z_i - \delta/2, z_i + \delta/2)$. We estimate this regression:

$$p_i = \sum_{d=0}^D \beta_d (z_i - z^*)^d + \sum_{j=-k}^k \gamma_j 1\{z_i - z^* = j \cdot \delta\} + u_i \quad (1)$$

This expresses the annual earnings distribution as a degree D polynomial, plus a set of indicators for each bin with a midpoint within $k\delta$ of the kink.

Our measure of bunching is $\hat{B} = \sum_{j=-k}^k \hat{\gamma}_j$, the estimated excess probability of locating at the kink, relative to the polynomial fit. To obtain a measure of excess mass that is comparable across different kinks, we scale by the counterfactual density at z^* , *i.e.* $\hat{h}_0(z^*) = \hat{\beta}_0/\delta$. We refer to the density of earnings in the absence of the earnings test, under a linear tax schedule with a constant marginal tax rate, as the “counterfactual” or “initial” earnings density. Thus, our estimate of “normalized excess mass” is $\hat{b} = \hat{B} / \hat{h}_0(z^*) = \delta \hat{B} / \hat{\beta}_0$. In our empirical application, we choose $D = 7$, $\delta = 800$ and $k = 4$ as a baseline, implying that our estimate of bunching is driven by individuals with annual earnings within \$3,600 of the kink. We also show our results under alternative choices of D , δ , and k . We estimate bootstrapped standard errors.

4 Data

We apply this bunching framework on a one percent random sample of Social Security numbers from the restricted-access Social Security Administration Master Earnings File, linked to the Master Beneficiary Record. The data contain a complete longitudinal earnings history with information on earnings in each calendar year since 1951; year of birth; the year (if any) that claiming began; date of death; and sex. In a calendar year, “age” is defined as the highest age an individual attains in that calendar year.

Starting in 1978, the earnings measure reflects total wage compensation, as reported on W-2 tax forms. Earnings are not subject to manipulation through tax deductions, credits, or exemptions, and are subject to third-party reporting among the non-self-employed. Separate information is available on self-employment earnings and non-self-employment earnings. The data do not contain information on hours worked or job amenities.

Our main sample at each age and year consists of individuals who have ultimately claimed at an age less than or equal to 65, which allows us to investigate a constant sample across ages. We exclude person-years with positive self-employment income. Because we focus on the intensive margin response, we further limit the sample in a given year to observations with positive earnings in that year.

Table 1 shows summary statistics in our main sample, 62 to 69 year-olds in 1990 to 1999. The sample has 376,431 observations. The sample is 57 percent male. Median earnings, \$14,555.56, is not far from the Earnings Test exempt amount, which averages \$16,738 for those 65 and older and \$11,650 for those younger than 65 over this period. Conditional on positive earnings, mean earnings is \$28,892.63.

Our second data source is the Longitudinal Employer Household Dynamics (LEHD) of the U.S. Census (Abowd *et al.*, 2009), which longitudinally follows the earnings of around nine-tenths of workers in covered states. We use a 20 percent random subsample of these individuals from 1990 to 1999. We use these data only in one figure, for which the large sample size in the LEHD is helpful.

To generate the effective marginal tax rate, in our baseline we incorporate the Earnings Test benefit reduction rate as well as the average federal and state income and payroll

marginal tax rates. We calculate marginal tax rates using TAXSIM (Feenberg and Coutts, 1993) and information on individuals within \$2,000 of the kink in the Statistics of Income data in the years we examine.

In our estimates and model we abstract from the claiming decision by examining those who have already claimed Social Security. This is only a trivial abstraction here because nearly everyone (over 90 percent) has claimed by the ages we study in our main evidence, 66 to 71.

5 Documenting Earnings Adjustment Frictions

Using the administrative data, we document several pieces of evidence for adjustment frictions by examining the pattern of bunching across ages. We focus on the period 1990 to 1999, when the Earnings Test applied from ages 62 to 69. The policy changes at ages 62 and 70—when the Earnings Test is imposed and removed, respectively, for Social Security claimants—would be anticipated by those who have knowledge of the relevant policies. Figure 2 Panel A plots earnings histograms for each age from 59 to 73, along with the estimated smooth counterfactual polynomial density.

First, we show that “de-bunching”—movement away from the former kink among those initially bunching at the kink—does not occur immediately for some individuals. Figure 2 Panel A shows clear visual evidence of substantial bunching from ages 62 to 69, when the Earnings Test applies to claimants’ Social Security benefits, and no excess mass at earlier ages. At ages 70 and 71, which are not subject to the Earnings Test, there is still clear visual evidence of bunching in the region of the kink.

We estimate that there is substantial and significant excess mass at ages 70 and 71. Figure 2 Panel B shows that normalized excess mass is statistically significantly different from zero at each age from 62 to 71 ($p < 0.01$ at each age). Normalized excess mass rises from 62 to 63 and remains around this level until age 69 (with a dip at age 65 that we discuss below). When we pool data from 1983 to 1999 in Figure 4—giving us more power than in our baseline sample over 1990 to 1999 when the Earnings Test does not change—bunching above age 70 is even more visually apparent, and excess mass at age 71 is highly significant and clearly positive.

Second, we exploit the panel dimension of the data to demonstrate inertia near the kink at the individual level. Figure 5 shows that conditional on earnings at ages 70 or 71 within \$1,000 of the exempt amount, the density of earnings at age 69 spikes at the exempt amount. Similarly, conditional on earnings at age 69 within \$1,000 of the exempt amount, the density of age 70 or age 71 earnings spikes near the exempt amount. It is notable that we document adjustment frictions even among those who were flexible enough to bunch at the kink initially.

Third, Figure 3 shows spikes near the exempt amount in the mean percentage change in earnings from ages 69 to 70 and 70 to 71, consistent with de-bunching from age 69 to 70, and from age 70 to 71, among those initially near the kink in the LEHD. This shows that bunchers are returning to higher earnings, as predicted by theory, and that this process continues at least until age 71.³

Fourth, Figure 2 Panel B shows that bunching is substantially lower at age 65 than surrounding ages. The location of the kink changes substantially from age 64 to age 65 because the exempt amount rises greatly (Figure 1). Individuals may have difficulty adjusting to the new location of the kink within one year. This delay suggests that individuals also face adjustment frictions in this context.⁴

Fifth, the amount of bunching rises from age 62 to 63, suggesting gradual adjustment. Appendix Figure B.2 shows that when the sample at a given age consists of those who have claimed by that age, we still find a substantial increase in bunching from 62 to 63.

Each of these several pieces of evidence points to adjustment frictions. In Appendix Table B.2, we probe the robustness of our results by varying the bandwidth, the degree of the polynomial, and the excluded region when we estimate bunching. We also conduct

³We classify claimants as age 70 when they attain age 70 during that calendar year. As a result, some individuals will be classified as age 70 but will have been subject to the Earnings Test for a portion of the year (in the extreme case of a December 31 birthday, for all but one day). 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; (2) the continued adjustment away from the kink from age 70 to age 71 in Figure 3; and (3) the spike in the elasticity estimated using the Saez (2010) approach in 1990, documented in Figure 8 and explained below. Moreover, Appendix Table B.1 shows that those born in January to March—who are subject to the Earnings Test for only a small portion of the calendar year when they turn age 70—also show significant bunching at ages 70 ($p < 0.05$) and 71 ($p < 0.10$) from 1983 to 1999.

⁴This interpretation of the patterns around ages 64 and 65 is consistent with Figure B.1, which shows that conditional on age 64 earnings near the age 64 exempt amount, the age 65 earnings density shows a large spike at the kink that prevailed at age 64 and a smaller spike at the current, age 65 kink. Also, conditional on age 65 earnings near the age 65 exempt amount, the density of age 64 earnings shows a spike near the exempt amount for age 64. In principle, our coarse measure of age could affect these patterns: individuals turning 65 in a given calendar year face the age-65 exempt amount for only the part of the calendar year after they turn 65, which could serve as a partial explanation for continued bunching at age 65 at the exempt amount applying to age 64. However, we would then expect the age 64 and age 65 exempt amounts to display equal amounts of bunching, which is not the case.

several additional analyses in Gelber, Jones, and Sacks (2013), including varying the time period examined. Overall, these additional analyses generally show similar patterns.

6 Model Underlying Estimation

The results thus far suggest a role for adjustment frictions in individuals' earnings choices. To estimate such adjustment costs as well as earnings elasticities, we build on the frictionless Saez (2010) model described in Section 3. There we considered a transition from a linear tax schedule with a constant marginal tax rate τ_0 to a schedule with a convex kink, where the rate below the kink earnings level z^* is τ_0 , and the rate above z^* is $\tau_1 > \tau_0$. We refer to this kink at z^* as K_1 . Next, as in our empirical context, we consider a decrease in the higher marginal tax rate above z^* to $\tau_2 < \tau_1$.⁵ We refer to this less sharply bent kink as K_2 . In the presence of a kink K_j with marginal tax rate τ_0 below z^* and τ_j above z^* , $j \in \{1, 2\}$, the share of individuals bunching at z^* in the frictionless model is:

$$B_j^* = \int_{z^*}^{z^* + \Delta z_j^*} h_0(\zeta) d\zeta \quad (2)$$

For small tax rate changes, we can relate the elasticity to the earnings change Δz_j^* for the individual with the highest *ex ante* earnings who bunches *ex post*:

$$\varepsilon = \frac{\Delta z_j^*/z^*}{d\tau_j/(1 - \tau_0)} \quad (3)$$

where $d\tau_j = \tau_j - \tau_0$ and ε is the elasticity, $\varepsilon \equiv -(\partial z/z)/(\partial \tau/(1 - \tau))$. The higher the elasticity and the change in taxes at the kink, the larger is the range Δz_j^* of bunchers.

6.1 Bunching in a Single Cross-Section with Adjustment Costs

We now extend the model to include a cost of adjusting earnings. In a basic version of the model, individuals must pay a fixed utility cost of ϕ similar to Chetty *et. al.* (2011); we discuss later how this can be extended to any polynomial adjustment cost function with any number of parameters. The fixed cost could represent the information costs associated with navigating a new tax-and-transfer regime if, for example, individuals only make the effort to understand their earnings incentives when the utility gains from doing so are sufficiently large

⁵The case of $d\tau_2 > d\tau_1$ is governed by an analogous set of formulas.

(*e.g.* Simon, 1955; Chetty *et al.*, 2007; Hoopes, Reck, and Slemrod, 2015). Alternatively, the fixed 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.⁶

Our model of fixed costs relates to labor economics literature on constraints on hours worked, as well as public finance literature that explores frictions in earnings. One common feature of models of earnings frictions in labor economics (*e.g.* Cogan, 1981; Altonji and Paxson, 1990, 1992; Dickens and Lundberg, 1993) and public finance (*e.g.* Chetty *et al.*, 2011; Chetty, 2012) is that the decision-making setting is generally static. We begin by adopting this modeling convention.

Figure 6 Panel A illustrates how a fixed adjustment cost attenuates the level of bunching, relative to equation (2), and obscures the estimation of ε in a single cross-section that is possible in the Saez (2010) model. Consider the individual at point 0, who initially earns \underline{z}_1 along the linear budget constraint with tax rate τ_0 . This individual faces a higher marginal tax rate τ_1 after the kink is introduced. Because she faces an adjustment cost, she may decide to keep her earnings at \underline{z}_1 and locate at point 1. Alternatively, with a sufficiently low adjustment cost, she incurs the adjustment cost and reduces her earnings to z^* (point 2).

We assume that the benefit of relocating to the kink is increasing in the distance from the kink for initial earnings in the range $[z^*, z^* + \Delta z_1^*]$. This requires that the size of the optimal adjustment in earnings increases in a at a rate faster than the decrease in the marginal utility of consumption.⁷ This is true, for example, if utility is quasilinear, as in related recent public finance literature (*e.g.* Saez, 2010; Chetty *et al.*, 2011; Kleven and Waseem, 2013; Kleven, Landais, Saez, and Schultz, 2014). This implies 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. In Figure 6, this individual is the marginal buncher

⁶Inattention or the difficulty of negotiating new contracts should be associated with positive adjustment costs, consistent with the empirical patterns in Section 5, including continued bunching at former kinks. That could distinguish this context from others such as the firm context in Garicano *et al.* (2016), who find negative fixed costs attributed to within-firm positive spillovers from information collection, which seem less applicable in our context.

⁷To see this, note that the utility gain from reoptimizing is $u((1 - \tau_1)z_1 + R_1, z_1; a) - u((1 - \tau_1)z_0 + R_1, z_0; a) \approx u_c \cdot (1 - \tau_1)[z_1 - z_0] + u_z \cdot [z_1 - z_0] = u_c \cdot (\tau_1 - \tau_0)[z_0 - z_1]$, 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 first term, u_c , is decreasing as a (and therefore initial earnings z_0) increases. Thus, in order for the gain in utility to be increasing in a , we need the size of earnings adjustment $[z_0 - z_1]$ to increase at a rate that dominates.

who is indifferent between staying at the initial level of earnings z_1 (point 1) and moving to the kink earnings level z^* (point 2) by paying the adjustment cost ϕ .

Panel B of Figure 6 illustrates the degree of attenuation of bunching due to the adjustment cost. With the adjustment cost, only individuals with initial earnings in the range $[z_1, z^* + \Delta z_1]$ bunch at the kink K_1 . Bunching is given by the integral of the initial earnings density, $h_0(\cdot)$, over this range:

$$B_1(\boldsymbol{\tau}_1, z^*; \varepsilon, \phi) = \int_{z_1}^{z^* + \Delta z_1} h_0(\zeta) d\zeta, \quad (4)$$

where $\boldsymbol{\tau}_1 = (\tau_0, \tau_1)$ reflects the tax rates below and above z^* . The threshold level of earnings z_1 is an increasing function of ϕ , because larger adjustment costs attenuate the earnings of a greater range of individuals. The lower limit of the integral, z_1 , is implicitly defined by the indifference condition shown in Figure 6, Panel A:

$$\phi = u((1 - \tau_1)z^* + R_1, z^*; \underline{a}_1) - u((1 - \tau_1)z_1 + R_1, z_1; \underline{a}_1) \quad (5)$$

where R_1 is virtual income and \underline{a}_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, $\boldsymbol{\tau}_1 = (\tau_0, \tau_1)$, and the density $h_0(\cdot)$ near the exempt amount z^* . With only one kink and without further assumptions, 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 a change in $d\tau$. We assume that ability a is fixed over time from K_1 to K_2 , described above. Some individuals will remain bunching at the kink, even though they would prefer to move away from the kink in the absence of an adjustment cost, because the gain from de-bunching is not large enough to overcome the adjustment cost. The adjustment cost therefore attenuates the reduction in bunching, relative to a frictionless case.

Attenuation in the change in bunching is driven by those in area *iii* of Panel B in Figure 6. Under a frictionless model, individuals in this range do not bunch under K_2 . However, when moving from K_1 to K_2 in the presence of frictions, those in area *iii* continue to bunch

because their gains from adjusting from K_1 to K_2 are smaller than the adjustment cost, as shown in Panel C of Figure 6. At point 0, we show an individual's initial earnings \bar{z}_0 under a constant marginal tax rate of τ_0 . The individual responds to K_1 by bunching at z^* (point 1), since $\bar{z}_0 > \underline{z}_1$. Under K_2 , this individual would have chosen earnings $\bar{z}_2 > z^*$ (point 2) in a frictionless setting; we have illustrated the marginal buncher who, due to the fixed cost, is indifferent between staying at z^* and moving to \bar{z}_2 .

Thus, bunching under K_2 is:

$$\tilde{B}_2(\tilde{\tau}_2, z^*; \varepsilon, \phi) = \int_{\underline{z}_1}^{\bar{z}_0} h_0(\zeta) d\zeta, \quad (6)$$

where $\tilde{\tau}_2 = (\tau_0, \tau_1, \tau_2)$, and the “ \sim ” indicates that K_2 was preceded by a larger kink K_1 . The critical earnings levels for the marginal buncher, \bar{z}_0 and \bar{z}_2 , are implicitly defined by:⁸

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

The earnings elasticity is related to the adjustment of the marginal buncher: $\varepsilon = \frac{\bar{z}_0 - \bar{z}_2}{\bar{z}_2} \frac{(1 - \tau_0)}{d\tau_2}$.

The equations in (4), (5), (6) and (7) together pin down four unknowns (Δz_1^* , \underline{z}_1 , \bar{z}_0 and \bar{z}_2), each of which is a function of ε and ϕ . In our “comparative static method,” we draw on two empirical moments in the data, B_1 and \tilde{B}_2 , to identify our two key parameters, ε and ϕ .

The features of the data that help drive our estimates of the elasticity and adjustment cost are intuitive. In the frictionless model of Saez (2010), bunching at a convex kink is approximately proportional to $d\tau$; when $d\tau$ falls in this model, bunching at the kink falls proportionately. As we move from the more pronounced kink to the less pronounced kink in our model, 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. In the extreme case in which a kink has been eliminated, we can attribute any residual bunching to adjustment costs. Moreover, we show in Gelber, Jones, and Sacks (2013) that the absolute

⁸We additionally require that $\bar{z}_0 \leq z^* + \Delta z_1^*$. When this inequality is binding, none of the bunchers move away from the kink at z^* when the kink is reduced from K_1 to K_2 . Since we observe a reduction in bunching in our empirical setting, we ignore this inequality.

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 iv is decreasing in the adjustment cost. As in the frictionless case, the amount of bunching at K_1 is still increasing in the elasticity.

By applying our approach thus far to study adjustment over a given time frame, the resulting parameters should be interpreted as meaning that bunching in this time frame can be predicted if individuals behaved as if they faced the indicated adjustment cost and elasticity, in the spirit of Friedman (1953). This framework may be applied to yield “as if” estimates separately for each period.

6.3 Dynamic Version of Model

To account for how bunching evolves over time, as in the lagged adjustment shown in Section 5, we can nest our comparative static model within a framework incorporating more dynamic elements. We use a Calvo (1983) or “CalvoPlus” framework (*e.g.* Nakamura and Steinsson, 2010), in which there is a positive probability in each period of facing a finite, fixed adjustment cost.

We assume that the adjustment cost in any period is drawn from a discrete distribution $\{0, \phi\}$. This generates a gradual response to policy, as agents may adjust only when a sufficiently low value of the fixed cost is drawn. Such variation over time in the size of the adjustment cost from this discrete distribution could capture, for example, the stochastic arrival of available jobs or information about the policy.

How we model dynamics is also influenced by a key feature observed in the data: the lack of an anticipatory response to policy changes. In Appendix A.1 we solve a completely forward-looking model in which agents anticipate a policy change. This model nests the models presented in the main text. If agents were to place weight on the future in our forward-looking model, they should begin to bunch in anticipation of facing a kink, and they should begin to de-bunch in anticipation of the disappearance of a kink—neither of which we have observed in the data. Meanwhile, we observe a degree of delayed response to policy changes. We can capture both of these features of the data by assuming that a stochastic process determines whether an agent faces the cost of adjustment, but agents do not anticipate the policy change.

Formally, our main dynamic model without anticipatory behavior extends the notation from above as follows. As before, we assume that agents begin with their optimal frictionless level of earnings in period 0. Flow utility in each period is $v(c_{a,t}, z_{a,t}; a, z_{a,t-1}) = u(c_{a,t}, z_{a,t}; a) - \tilde{\phi}_t \cdot \mathbf{1}(z_{a,t} \neq z_{a,t-1})$, where $\mathbf{1}(\cdot)$ is the indicator function for changing earnings, which incurs a cost $\tilde{\phi}_t$. In each period, an agent draws $\tilde{\phi}_t$ from a discrete distribution, which equals ϕ with probability π_{t-t^*} and equals 0 with probability $1 - \pi_{t-t^*}$. To capture the observed features of the data, in which the probability of adjusting (conditional on initially locating at the kink) appears to vary over time, we allow the probability π_{t-t^*} to be a function of the time elapsed since the most recent policy change, $t - t^*$. Individuals are again indexed by a time-invariant heterogeneity parameter, a , which captures ability.

Individuals make decisions over a finite horizon. In period 0, individuals face a linear tax schedule, $T_0(z) = \tau_0 z$, with marginal tax rate τ_0 . In period 1, a kink, K_1 , is introduced at the earnings level z^* . This tax schedule is implemented for \mathcal{T}_1 periods, after which the tax schedule features a less pronounced kink, K_2 , at the earnings level z^* . For simplicity, we assume quasilinear utility, $u(c, z; a) = c - \frac{a}{1+1/\varepsilon} \left(\frac{z}{a}\right)^{1+1/\varepsilon}$, to abstract from income effects and focus on the dynamics created by the presence of adjustment costs. In each period, individuals draw $\tilde{\phi}_t$ and then maximize flow utility subject to a per-period budget constraint $z_{a,t} - T_j(z_{a,t}) - c_{a,t} \geq m$, where m reflects a borrowing constraint.⁹

These assumptions generate a simple decision rule. Let $\tilde{z}_{a,t}$ be the optimal frictionless level of earnings for an individual with ability a in period t . An agent will choose this level of earnings provided that the flow utility gain of moving from last-period earnings $z_{a,t-1}$ to the frictionless optimum $\tilde{z}_{a,t}$ exceeds the currently-drawn cost of adjustment, $\tilde{\phi}_t$. Otherwise, the agent remains at $z_{a,t-1}$.

We can now generalize our earlier expressions for bunching under K_1 and K_2 . Denote B_1^t as bunching at K_1 in period $t \in [1, \mathcal{T}_1]$. We have the following dynamic version of (4):

$$\begin{aligned} B_1^t &= \int_{z_1}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta + (1 - \prod_{j=1}^t \pi_j) \int_{z^*}^{z_1} h_0(\zeta) d\zeta \\ &= \prod_{j=1}^t \pi_j \cdot B_1 + (1 - \prod_{j=1}^t \pi_j) B_1^* \end{aligned} \tag{8}$$

⁹The quasilinearity assumption implies that the borrowing constraint does not directly affect the earnings decision. However, when agents are not forward looking, the borrowing constraint is necessary to rule out infinite borrowing.

where B_1^* is the frictionless level of bunching defined in (2) when $j = 1$. The first line of (8) shows that bunching in period t at K_1 is composed of two components added together. The first integral represents those who immediately adjust in period 1—the same group as in Section 6.3, areas ii through iv in Figure 6 Panel B. The second integral represents those in area i of the figure, who only adjust if they draw a zero cost of adjustment. The probability that this occurs by period t is $1 - \prod_{j=1}^t \pi_j$. The second line of (8) shows that as t grows, bunching converges to the frictionless level of bunching B_1^* .

We can similarly derive an expression for B_2^t , bunching at K_2 in period $t > \mathcal{T}_1$:

$$\begin{aligned}
B_2^t &= \int_{z_1}^{z^* + \Delta z_2^*} h_0(\zeta) d\zeta + \prod_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \int_{z^* + \Delta z_2^*}^{\bar{z}_0} h_0(\zeta) d\zeta \\
&\quad + (1 - \prod_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \prod_{j=1}^{\mathcal{T}_1} \pi_j) \int_{z^*}^{z_1} h_0(\zeta) d\zeta \\
&= \prod_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot [\tilde{B}_2 + (1 - \prod_{j=1}^{\mathcal{T}_1} \pi_j) [B_1^* - B_1]] + (1 - \prod_{j=1}^{t-\mathcal{T}_1} \pi_j) B_2^* \tag{9}
\end{aligned}$$

where B_2^* is the frictionless level of bunching at K_2 . In the first two lines, bunching in period t at K_2 consists of three components added together. First, individuals in area ii in the figure immediately bunch in period 1, and remain bunching at the smaller kink. Second, in area iii of the figure, excess bunchers who immediately bunch in period 1 now de-bunch when a zero cost of adjustment is drawn. Third, those in area i of the figure would like to bunch under both K_1 and K_2 , but only do so once a zero cost of adjustment is drawn. On the third line, we again see that as the time between period t and \mathcal{T}_1 grows, the level of bunching converges to the frictionless amount, B_2^* , shown in areas i and ii of the figure.

Relative to the dynamic model, the comparative static model from Sections 6.1-6.2 has both strengths and weaknesses. The comparative static model has the strength of transparently illustrating the basic forces determining the elasticity and adjustment cost. We assume that ability is fixed throughout the window of estimation, which may be more plausible in the case of the comparative static model—when we only use two cross-sections from adjacent time periods—than when we use a dynamic model and study a longer time frame. The estimation of the more dynamic model requires more moments from the data to estimate additional parameters. However, the dynamic model has the strength of allowing us to account for the time pattern of bunching. The comparative static model corresponds to the

special case of the dynamic model in which individuals never draw zero adjustment cost, so that $\pi_j = 1$.

6.4 Extensions

Our framework can be extended in a number of ways. First, we can extend the model to accommodate heterogeneity in elasticities and adjustment costs. In Appendix A.2 we derive generalized formulae for bunching that allow us to interpret our comparative static model estimates as average parameters among the set of bunchers. Appendix A.2.2 further discusses how the dynamic model can be interpreted in the presence of heterogeneity in these parameters and the vector π_i .¹⁰

Second, 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 A.3, we extend the model to allow 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 report estimates under this method below.

Third, the model with a fixed cost of adjustment can be generalized to any polynomial functional form of the adjustment. In Appendix A.0.1, we discuss estimation of a model with a cost of adjustment that has both a fixed cost component and a component that is linear in the size of the earnings adjustment. In general, we can allow for the adjustment cost to be an arbitrary polynomial function of order n that depends on the size of the adjustment; this requires $n + 1$ moments for estimation.

6.5 Econometric Estimation of the Model

We estimate the model we have described using a minimum distance estimator. As explained in Appendix A.5, to estimate (ε, ϕ) in the static setting, we seek the values of the parameters that make predicted bunching and actual (estimated) bunching as close as possible on

¹⁰Special cases of our model have implications for other moments of the earnings distribution. However, with heterogeneity in the parameters it is not possible to use these moments without more stringent distributional assumptions.

average.

Equations (8) and (9) illustrate how we estimate the elasticity and adjustment cost in the dynamic setting. We require as many observations of bunching as the parameters, $(\varepsilon, \phi, \pi_1, \dots, \pi_J)$, and these moments must span a change in $d\tau$. Suppose we observe the pattern of bunching over time around two or more different policy changes. Loosely speaking, the π 's are estimated relative to one another from the time pattern of bunching: a delay in adjustment in a given period will generally correspond to a higher probability of facing the adjustment cost (all else equal). This relationship is linear, as the degree of “inertia” in bunching in each period increases linearly in π_1 . Meanwhile, a higher ϕ implies a larger amount of inertia in *all* periods until bunching has fully dissipated (in a way that depends on the earnings distribution, the elasticity, and the size of the tax change). Finally, a higher ε will correspond to a larger amount of bunching once bunching has had time to adjust fully to the policy changes. Intuitively, these features of the data help us to identify the parameters using our dynamic model.

In our baseline, we use a non-parametric density for the counterfactual earnings distribution, H_0 . Once H_0 is known, in the comparative static model we use (4) and (6) to obtain predicted bunching from the model. To recover H_0 non-parametrically we use the empirical earnings distribution for 72 year-olds in \$800 bins as the counterfactual distribution. 72 year-olds’ earnings density represents a reasonable counterfactual because they no longer face the Earnings Test, no longer show bunching, and are close in age to those aged 70 or 71.¹¹

Our estimator assumes a quasilinear utility function, $u(c, z; a) = c - \frac{a}{1+1/\varepsilon} \left(\frac{z}{a}\right)^{1+1/\varepsilon}$, as in previous literature. Without the quasi-linearity assumption, there are income effects on labor supply. To estimate the parameters of the model, we would need additional parametric assumptions, as well as additional data: in a static model we would need individual-level data on unearned income, and in a lifecycle earnings supply model we would need data on lifetime wealth including assets (Blundell and MaCurdy, 1999). We do not have data on unearned income or assets, though these may be available in other applications. Given such data, our

¹¹Because we use the age-72 density as our counterfactual density, our method is not subject to the Blomquist and Newey (2017) point that preference heterogeneity cannot be simultaneously estimated with the taxable income elasticity.

estimates of the parameters could be performed with two moments under any one-parameter utility function that satisfies the single-crossing property, which will generate a unique cutoff level of counterfactual earnings above which individuals adjust to the kink. With an n -parameter utility function, we would require $n + 1$ moments. Assuming that unearned income is not changing over time across counterfactual earnings levels due to factors other than the Earnings Test, the estimated elasticity in the comparative static model will be a weighted average of the compensated and uncompensated elasticity (Kleven 2016, footnote 5).¹²

For each bootstrap sample, generated using the procedure of Chetty *et al.* (2011), we compute the estimated values of the parameters. We determine whether an estimate of the adjustment cost $\hat{\phi}$ is significantly different from zero by assessing how frequently the constraint $\phi \geq 0$ binds in our estimation. In Appendix A.4 we demonstrate identification more formally.

7 Estimates of Elasticity and Adjustment Cost

7.1 Estimates using the Comparative Static Method

To estimate ε and ϕ using our “comparative static” method, we first examine the reduction in the rate in 1990 as a baseline and next turn to the elimination of the Earnings Test at ages 70 and older. No other key policy changes occurred in 1990 that would have materially affected bunching near the kink.

Figure 7 shows the patterns driving the parameter estimates for the 1990 change. Figure 7 shows bunching among 66-68 year-olds, for whom the benefit reduction rate fell from 50 percent to 33.33 percent in 1990. Bunching fell negligibly from 1989 to 1990 but fell more subsequent to 1990.

Table 2 presents estimates of our static model, examining 66-68 year-olds in 1989 and 1990. We estimate an elasticity of 0.35 and an adjustment cost of \$278, both significantly different from zero ($p < 0.01$). This estimated adjustment cost represents the cost of adjusting

¹² Assuming that leisure is a normal good—so that increases in unearned income decreases earnings—the implied compensated elasticity will be larger than the observed policy elasticity (Hendren 2016). The presence of income effects would have an ambiguous effect on the magnitude of our estimated adjustment costs. In our dynamic model, the income effects would add a savings decision and a new state variable, assets, unless we continue to assume myopia.

earnings in the first year after the policy change.

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 0.58. Consistent with our discussion above, the estimated elasticity is higher when we do not allow for adjustment costs than when we do, because 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 — 66 percent higher in the constrained case — and statistically significant ($p < 0.01$). Similarly, when we apply the frictionless Saez method over the years 1982 to 1993 (excluding the transitional year of 1990), the average elasticity we estimate is 0.19 ($p < 0.01$) — just over half our baseline elasticity — because adjustment frictions attenuate the degree of bunching and elasticity estimate.

Other specifications in Table 2 show similar results. We adjust the marginal tax rate to take account of benefit enhancement, following the calculations of the effective Social Security tax rate net of benefit enhancement in Coile and Gruber (2001). This raises the estimated elasticity but yields similar qualitative patterns across the constrained and unconstrained estimates. The next rows show that our estimates are similar under other specifications: excluding FICA taxes from the baseline tax rate; using a locally uniform density; other bandwidths; and other years of analysis.

Returning to the baseline specification, the point estimates in Appendix Table B.3 show that across groups, elasticities tend to be similar, but women have higher adjustment costs than men, those with low prior lifetime real earnings have higher adjustment costs than those with high prior earnings, and those with high and low volatility of prior earnings have similar adjustment costs. In Appendix Table B.4 we find similar results when 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 A.3.

We believe that three factors make the identification strategy in Table 2 credible. First, Figure 7 shows that in a “control group” of 62-64 year-olds who do not experience a policy change in 1990, bunching is very stable in the years before and after 1990, suggesting that the 66-68 year-old group will be sufficient to pick up changes in bunching due to the policy change. Appendix Table B.5 verifies that in a “differences-in-differences specification” comparing 66-

68 year-olds to 62-64 year-olds, bunching among 66-68 year-olds falls insignificantly in 1990 relative to before 1990, bunching is significantly smaller among 66-68 year-olds in years after 1990, and these estimates are very similar to the time series estimates comparing only 66-68 year-olds over time.

Second, Figure 8 shows that the elasticity we estimate among 66-68 year-olds using the frictionless Saez (2010) method shows a sudden upward spike in bunching in 1990 but subsequently reverts to near its previous level. This relates directly to our theory, which predicts that following a reduction in the change in the marginal tax rate at the kink, there may be excess bunching due to inertia reflected in area *iv* in Figure 6, Panel B. Once we allow for an adjustment cost, this excess bunching is attributed to optimization frictions.

Third, Table 3 shows comparable evidence of frictions when we examine the removal of the kink at age 70 (pooling years 1990-1999). When comparing adjustment at age 70 to adjustment in 1990, a key pattern consistent with our model is that the decrease in normalized excess mass from 1989 to 1990 in Figure 7 is much smaller in absolute and percentage terms than the decrease in normalized excess mass from age 69 to age 70 in Figure 2 Panel B. With an adjustment cost preventing immediate adjustment, normalized excess mass should fall less when the jump in marginal tax rates at the kink falls less (in the change from a 50 percent to a 33.33 percent benefit reduction rate in 1990) than when the jump in marginal tax rates at the kink falls more (in the change from a 33.33 percent to a 0 percent benefit reduction rate at age 70). Table 4 shows that we estimate similar results when we pool data from the ages 69 to 71 transition with the 1989 to 1990 transition.

Our estimates of elasticities and adjustment costs, and our earlier descriptive evidence documenting the speed of adjustment, are local to the population that is observed bunching at the kinks. Local estimates are a general feature of quasi-experimental settings. With enough variation in the location of kinks, the set of bunchers—and the resulting parameter estimates—could in principle jointly cover much of the earnings distribution and population. With respect to external validity in our specific context, it is encouraging that the local parameter estimates are similar in both the context of the change in the benefit reduction rate in 1990 from 50 percent to 33.33 percent, and the change at age 70 from 33.33 percent to zero percent.

7.2 Estimates using the Dynamic Method

Table 5 shows the estimates of the dynamic model. There are several parameters to estimate— ε , ϕ , and the vector of observed π_{t-t^*} 's—but a limited number of years in the data with useful variation: bunching varies little from year to year prior to the policy changes in 1990 or at age 70, and bunching fully dissipates by at most three years after the policy changes. So that we have a sufficient number of moments to estimate the parameters, as in Table 4 we pool data on bunching from 1990 to 1999 at ages 67, 68, 69, 70, 71, and 72, with data on bunching among 66-68 year-olds in 1987, 1988, 1989, 1990, 1991, and 1992. This gives us twelve moments (six moments for each of two policy changes) with which to estimate seven parameters (ε , ϕ , π_1 , $\pi_1\pi_2$, $\pi_1\pi_2\pi_3$, $\pi_1\pi_2\pi_3\pi_4$, and $\pi_1\pi_2\pi_3\pi_4\pi_5$).

We estimate $\varepsilon = 0.36$ and $\phi = \$243$ in the baseline dynamic specification. The estimates of ε are remarkably similar under the static and dynamic models applied to comparable data in Tables 4 and 5, respectively. The estimates of ϕ are also in the same range. The point estimate of π_1 varies across specifications from 0.64 in the baseline to 1, indicating that a minority of individuals are able to adjust in the year of the policy change. This mirrors our earlier finding that while some individuals adjust in the year of a policy change, many do not. The point estimate of $\pi_1\pi_2$ varies across specifications from 0.00 to 0.47, indicating that a majority of individuals are able to adjust by the year following a policy change. This mirrors our earlier finding that substantial adjustment occurs with a lag. In all specifications, $\pi_1\pi_2\pi_3$ is estimated to be zero, indicating that individuals are fully able to adjust by the third year after a policy change. This mirrors our earlier finding that adjustment fully occurs by three years after the policy change.

Given our estimates of the π_j 's, it makes sense that we estimate comparable results from the static and dynamic models. If hypothetically adjustment were completely constrained in years 1 and 2 after the policy change and subsequently completely unconstrained, then we should estimate essentially identical results in the static and dynamic models because the static model effectively assumes that the only barrier to adjustment is the adjustment cost ϕ —similar to assuming that $\pi_j = 1$ for the periods over which adjustment is estimated. The estimates of the dynamic model are not very different from this hypothetical scenario: π_1 is

well over 50 percent, and $\pi_1\pi_2$ is substantial but under 50 percent.

8 Simulations of the Effect of Policy Changes

Our parameter estimates imply that incorporating adjustment costs into the analysis can have important implications for predicting the short-run impact of policy changes on earnings, as policy-makers often seek to do. In particular, the adjustment costs we estimate greatly attenuate the predicted short-run impact of policy changes on earnings.

We use our estimates of the static model, using the year before through the year after a policy change as in our baseline, to simulate the effect in our data of two illustrative policy changes. Details are provided in Appendix A.6 and Appendix Table B.6. Reducing the marginal tax rate above the kink by 50 percentage points—as could be implied by a policy like eliminating the Earnings Test for 62-64 year-olds—would cause a large, 23.4 percent rise in earnings at the intensive margin. However, a less large change—in particular, any cut in the marginal tax rate above the exempt amount of 17.22 percentage points or smaller—would cause no change in earnings within a one-year time horizon because the potential gains from adjusting are not large enough to overcome the adjustment cost.

This illustrates a principle: because the gains to relocation are second-order near the kink, even a modest adjustment cost around \$280 can prevent adjustment in the short run—and even following a substantial cut in marginal tax rates. Moreover, the lack of immediate response predicted with a change of 17.22 percentage points makes sense in light of the empirical patterns we observe, in particular the negligible change in bunching seen in the data from 1989 to 1990 when the marginal tax rate falls by 17 percentage points. Similarly, this sheds light on why our estimated adjustment cost is small despite significant attenuation. The Appendix shows this conclusion is robust to other assumptions. Under our estimates of the dynamic model we would still find that the short-run reaction even to large taxes changes is greatly attenuated, since the dynamic model estimates show that most individuals are constrained from adjusting immediately.

9 Conclusion

We introduce a method for documenting adjustment frictions: examining the speed of adjustment to the disappearance of convex kinks in the effective tax schedule. We document delays in earnings adjustment to large changes in the Social Security Earnings Test. The lack of immediate response suggests that the short-run impact of changes in the effective marginal tax rate can be substantially attenuated, even with large policy changes.

Next, we develop a method to estimate earnings elasticities and adjustment costs relying on bunching at convex budget set kinks. Examining data in the year of a policy change, we estimate that the elasticity is 0.35 and the adjustment cost is around \$280. When we estimate a frictionless model with zero adjustment cost, the elasticity is quite different. We extend our methods to a dynamic context and estimate that full adjustment takes three years.

Even modest fixed adjustment costs—like the \$280 cost we estimate in our baseline—can greatly impede short-run adjustment to large reforms because the costs of deviating from the frictionless optimum are second order. Our simulations confirm that adjustment costs can make a dramatic difference in the predictions. This could frustrate the goal of immediately impacting short-run earnings, as envisioned in many recent policy discussions, and could have important implications for policy-makers’ projections of the magnitude and timing of the earnings reaction to changes in tax and transfer policies.

We find bunching among wage earners, whereas previous studies in the U.S. have found substantial earnings bunching only among the self-employed (Saez 2010, Chetty, Friedman, and Saez 2013). Our study suggests a possible reason for this: adjustment costs can imply that only large kinks should generate bunching, at least before individuals have an opportunity to adjust. Our results could be uncovering a positive and substantial labor supply elasticity that can be obscured in other contexts, in which kinks are usually smaller. It is also possible that elasticities are larger, that adjustment costs are smaller, or that the time needed to make an adjustment is shorter in our context than in others. Bunching does occur in many settings, and our method can be, and has been, used in such settings, both within and outside the labor supply context (He, Peng, and Wang, 2016; Schächtele 2016;

Mortenson *et al.* 2017; Gudgeon and Trenkle, 2016; Zaresani, 2016).

Further analysis could enrich our findings. First, further work distinguishing among the possible reasons for reaction to the Earnings Test, including misperceptions, remains an important issue, as is understanding the mechanisms that underlie adjustment costs. Our graphs show that more individuals “bunch” under the exempt amount than over it; it is worth investigating whether this relates to mis-perceptions of the Earnings Test. Second, if labor supply adjustments are sluggish more broadly, then it would be interesting to study whether forecasters such as the Congressional Budget Office systematically over-estimate the near-term employment and revenue effects of changes in effective tax rates. Third, most empirical specifications have related an individual’s tax rate in a given year to the individual’s earnings in that year. Our methods and findings could be used in selecting the time horizon for estimating responses to policy. If our results on the speed of adjustment generalize, this would also suggest that relatively short time frames can capture long-run responses. Investigating the speed of adjustment in other contexts would be valuable.

Finally, kinked budget sets are common across a wide variety of economic applications, including electricity demand (*e.g.* Ito, 2014), health insurance (*e.g.* Einav, Finkelstein, and Schrimpf, 2015), and retirement savings (*e.g.* Bernheim, Fradkin, and Popov, 2015). Our method could be adapted to estimate elasticities and adjustment frictions in the context of other consumption decisions.

References

- Abel, A. B. and J. C. Eberly (1994), “A Unified Model of Investment Under Uncertainty,” *American Economic Review*, 84, 1369-1384.
- Abowd, J. M., B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock (2009), “The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators.” In *Producer Dynamics: New Evidence from Micro Data*, 149–230, University of Chicago Press.
- Altonji, J. G. and C. H. Paxson (1988), “Labor Supply Preferences, Hours Constraints, and Hours-Wage Trade-Offs,” *Journal of Labor Economics*, 6, 254-276.
- Arrow, K., T. Harris and J. Marschak (1951), “Optimal Inventory Policy,” *Econometrica*, 19, 205-272.
- Bernheim, B. D., A. Fradkin, and I. Popov (forthcoming), “The Welfare Economics of Default Options in 401(k) Plans,” *American Economic Review*.
- Baumol, W. J. (1952), “The Transactions Demand for Cash,” *Quarterly Journal of Economics*, 66, 545-556.

- Blomquist, S., and W. Newey** (2017), “The Bunching Estimator Cannot Identify the Taxable Income Elasticity.” MIT working paper.
- Blundell, R., and T. MaCurdy** (1999), “Labor Supply.” In *Handbook of Labor Economics*, vol. 3, Elsevier.
- Brown, J., A. Kapteyn, O. Mitchell, and T. Mattox** (2013), “Framing the Social Security Earnings Test,” Wharton Pension Research Council Working Paper 2013-06.
- Burtless, G. and R. A. Moffitt** (1985), “The Joint Choice of Retirement Age and Postretirement Hours of Work,” *Journal of Labor Economics*, 3, 209–236.
- Calvo, G.** (1983), “Staggered Prices in a Utility-Maximizing Framework,” *Journal of Monetary Economics*, 12, 383-398.
- Chetty, R.** (2012), “Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply,” *Econometrica*, 80, 969–1018.
- Chetty, R., J. N. Friedman, T. Olsen, and L. 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, R., A. Guren, D. Manoli, and A. 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*, vol. 27, University of Chicago Press.
- Chetty, R., A. Looney, and K. Kroft** (2007), “Salience and Taxation: Theory and Evidence,” NBER Working Paper No. 13330.
- Chetty, R., A. Looney, and K. Kroft** (2009), “Salience and Taxation: Theory and Evidence,” *American Economic Review*, 99, 1145–1177.
- Chetty, R., J. N. Friedman, and E. Saez** (2013), “Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings,” *American Economic Review*, 103, 2683-2721.
- Coile, C., and J. Gruber** (2001), “Social security incentives for retirement.” In *Themes in the Economics of Aging* (David Wise, ed.), 311-354, University of Chicago Press.
- Congressional Budget Office** (2009), “Estimated Macroeconomic Impacts of the American Recovery and Reinvestment Act of 2009.” Washington, D.C.: Government Printing Office.
- Diamond, P., and J. 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.
- Einav, L., A. Finkelstein, and P. Schrimpf** (2015), “The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D,” *Quarterly Journal of Economics*, 130(2), 841-899.
- Engelhardt, G., and A. 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.
- Feenberg, D., and E. Coutts** (1993), “An Introduction to the TAXSIM Model,” *Journal of Policy Analysis and Management*, 12, 189-194.
- Friedman, M.** (1953), *Essays in Positive Economics*. Chicago: University of Chicago Press.
- Friedberg, L.** (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, L.** (2000), “The Labor Supply Effects of the Social Security Earnings Test,” *Review of Economics and Statistics*, 82, 48–63.
- Garicano, L., C. Lelarge, and J. Van Reenen** (2016), “Firm Size Distortions and the Productivity Distribution: Evidence from France,” *American Economic Review*, 106(11), 3439-79.

- Gelber, A., D. Jones, and D. Sacks** (2013), “Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test,” NBER Working Paper 19491.
- Grossman, S. and G. Laroque** (1990), “Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods,” *Econometrica*, 58, 25-51.
- Gruber, J. and D. Wise** (1999), “Social Security Programs and Retirement Around the World: Introduction and Summary.” In *Social Security and Retirement Around the World*, J. Gruber and D. Wise, eds. Chicago: University of Chicago Press, 1-36.
- Gruber, J. and P. Orszag** (2003), “Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt?” *National Tax Journal*, 56, 755–773.
- Gudgeon, M., and S. Trenkle** (2016), “Frictions in Adjusting Earnings: Evidence from Notches in German Mini Jobs,” Boston University Working Paper.
- Hausman, J. A.** (1981), “Labor Supply.” In *How Taxes Affect Economic Behavior* (Henry J. Aaron and Joseph A. Pechman, eds.), 27–71, Brookings Institution.
- He, D., L. Peng, and X. Wang** (2016), “Understanding China’s Personal Income Tax,” UCSD Working Paper.
- Hoopes, J., D. Reck, and J. Slemrod** (2015), “Taxpayer Search for Information: Implications for Rational Attention,” *American Economic Journal: Economic Policy*, 7, 177-208.
- Kleven, H.** (2016), “Bunching,” *Annual Review of Economics*, 8, 435-464.
- Kleven, H., M. Knudsen, C. Kreiner, S. Pedersen, and E. Saez** (2011), “Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark,” *Econometrica*, 79, 651-692.
- Kleven, H., C. Landais, E. Saez, and E. Schultz** (2014), “Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners’ Tax Scheme in Denmark,” *Quarterly Journal of Economics*, 129, 333-378.
- Kleven, H. and M. Waseem** (2013), “Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan,” *Quarterly Journal of Economics*, 128, 669-723.
- Kline, P., and C. Walters** (2016), “Evaluating Public Programs with Close Substitutes: The Case of Head Start,” forthcoming, *Quarterly Journal of Economics*.
- Leahy, J.** (2008), “s-S Models.” In *The New Palgrave Dictionary of Economics, Second Edition* (Steven N. Durlauf and Lawrence E. Blume, eds.), Palgrave MacMillan.
- Liebman, J., and E. F. P. Luttmer** (2012), “The Perception of Social Security Incentives for Labor Supply and Retirement: The Median Voter Knows More Than You’d Think.” In *Tax Policy and the Economy* 26, 1-42.
- Mortenson, J., H. Schramm, and A. Whitten** (2017), “The Effects of Required Minimum Distribution Rules on Withdrawals from Traditional IRAs.” Joint Committee on Taxation working paper.
- Nakamura, E. and J. Steinsson** (2010), “Monetary Non-Neutrality in a Multisector Menu Cost Model,” *Quarterly Journal of Economics*, 125, 961-1013.
- Newey, W., and D. McFadden** (1994), “Large Sample Estimation and Hypothesis Testing.” In R.F. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4.
- Saez, E.** (2010), “Do Taxpayers Bunch at Kink Points?” *American Economic Journal: Economic Policy*, 2, 180–212.
- Saez, E., J. Slemrod, and S. H. Giertz** (2012), “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review,” *Journal of Economic Literature*, 50, 3–50.
- Schächtele, S.** (2016). “Deduction Responses to the Income Tax: Bunching Evidence from Germany,” Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Tax Responses, No. E10-V2.

Simon, H. A. (1955), “A Behavioral Model of Rational Choice,” *Quarterly Journal of Economics*, 69, 99–118.

Sheshinski, E. and Y. Weiss (1977), “Inflation and the cost of price adjustment,” *Review of Economic Studies*, 44, 287-303.

Social Security Administration (2012a), *Annual Statistical Supplement*. Washington, D.C.

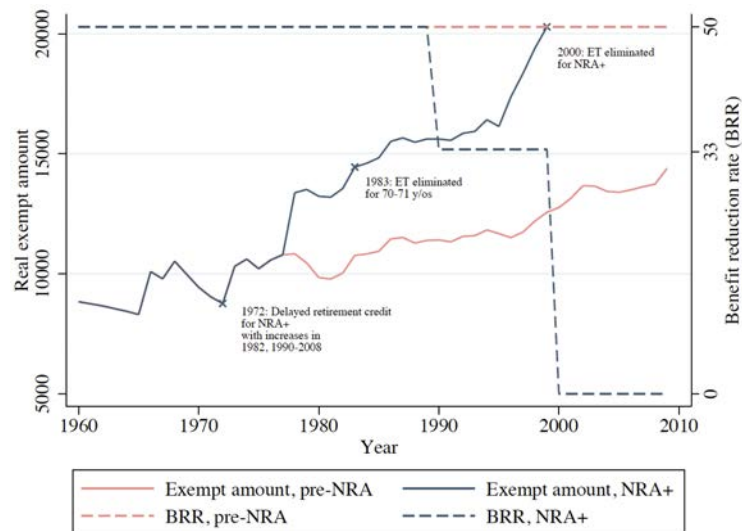
Social Security Administration (2012b), *Social Security Handbook*. Washington, D.C.

Song, J. G. and J. 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.

Stokey, N. (2008), *The Economics of Inaction*, Princeton University Press.

Zaresani, A. (2016), “Adjustment Costs and Incentives to Work: Evidence from a Disability Insurance Program.” University of Calgary working paper.

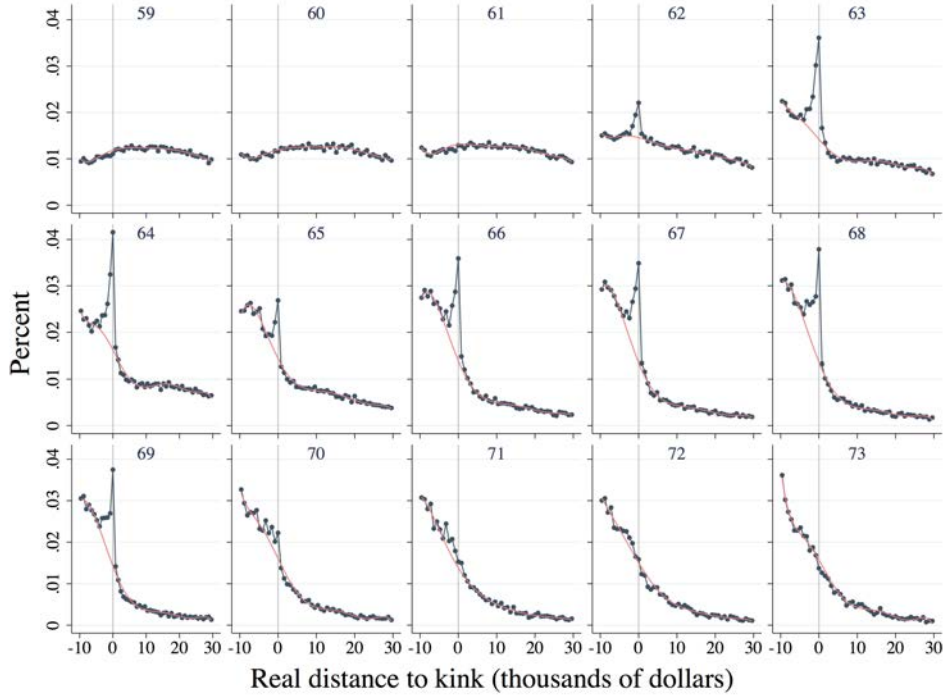
Figure 1: Key Earnings Test Rules, 1961-2009



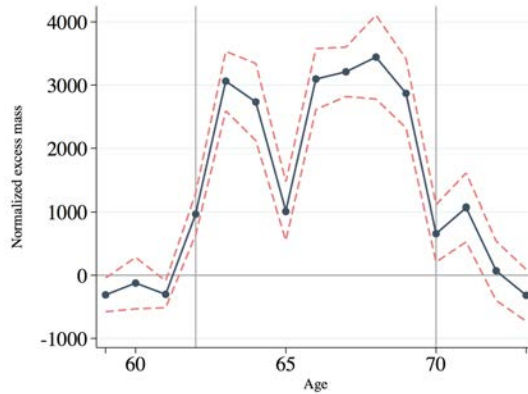
Notes: The right vertical axis measures the benefit reduction rate in Social Security payments for every dollar earned beyond the exempt amount. The left vertical axis measures the real value of the exempt amount over time.

Figure 2: Earnings Histograms and Normalized Excess Mass by Age

A. Histograms by Age

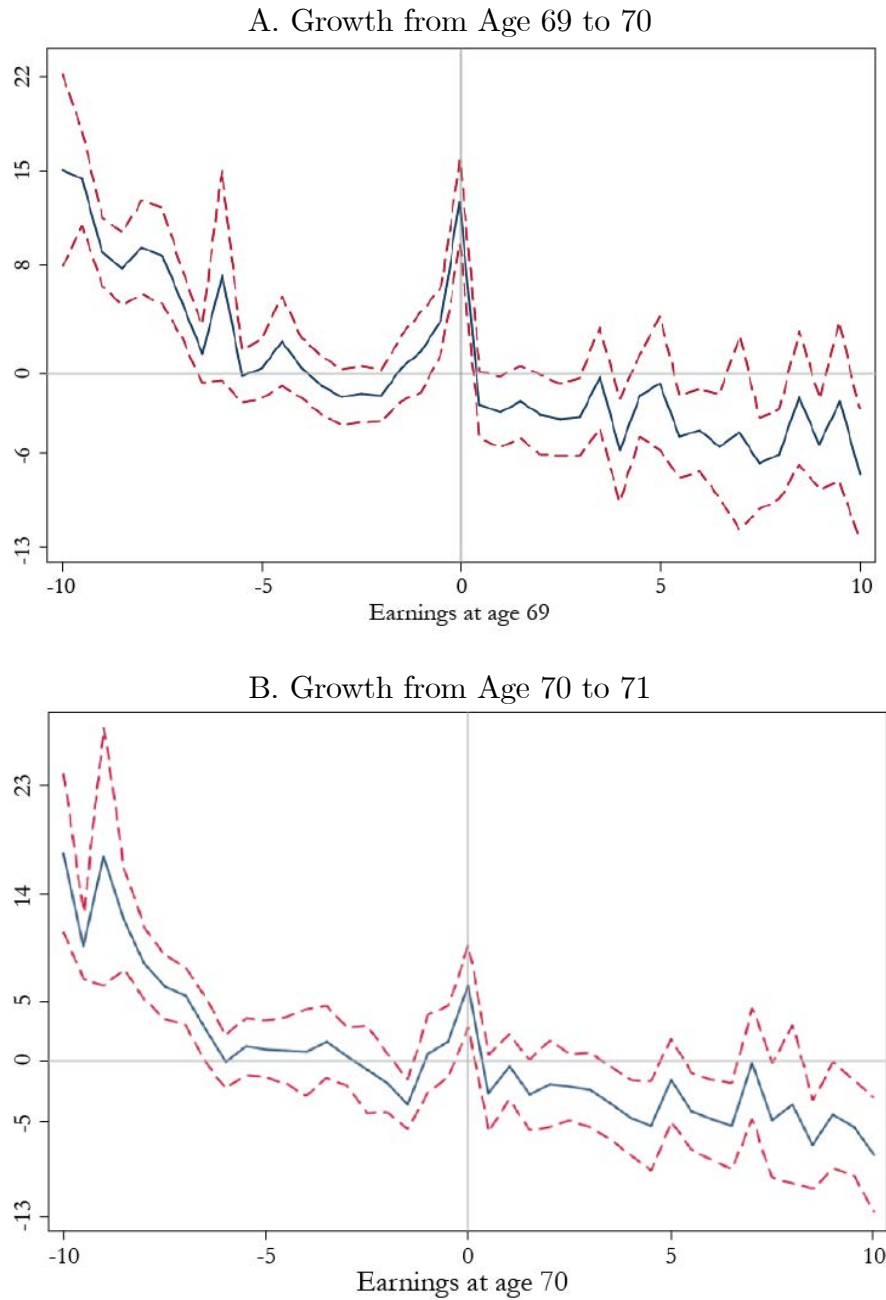


B. Normalized Excess Mass by Age



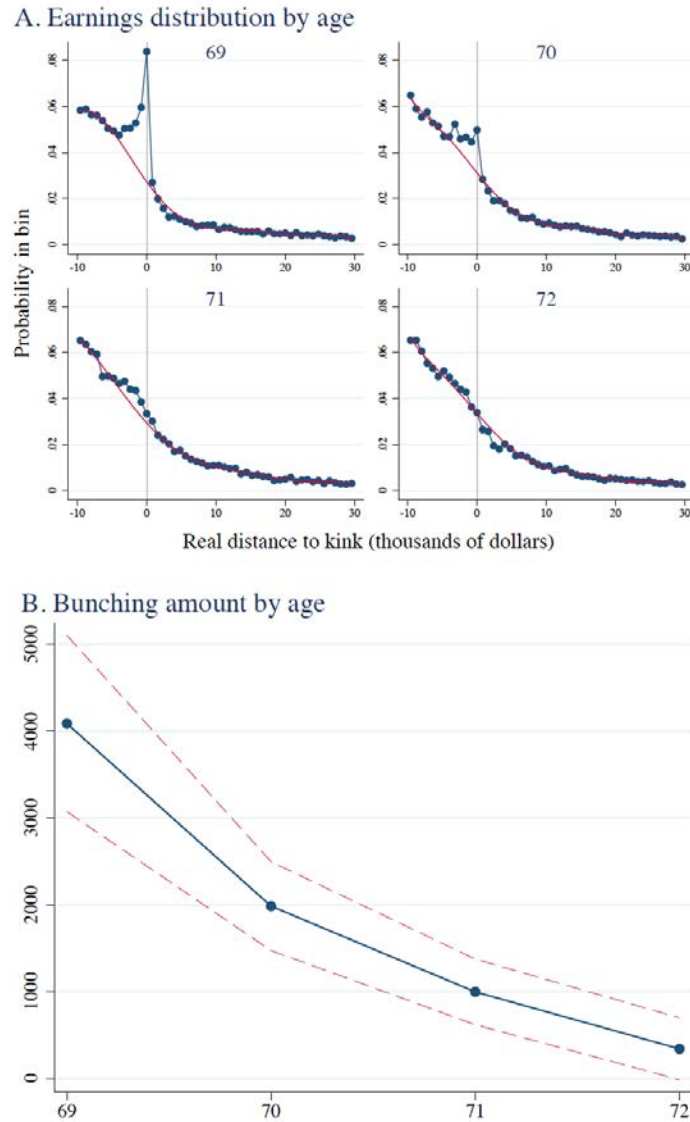
Notes: The sample is a one percent random sample of all Social Security numbers, among individuals who claim Social Security benefits by age 65, over calendar years 1990 to 1999. We exclude person-years with self-employment income or with zero non-self-employment earnings. The bin width is \$800. In Panel A, 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. Panel B shows normalized bunching at the Earnings Test kink, calculated as described in Section 3. Dashed lines denote 95% confidence intervals. The vertical lines show the ages at which the Earnings Test first applies (62) and ceases to apply (70). For ages younger than 62 (70 and older), we define the “placebo” kink in a given year as the kink that applies to pre-Normal Retirement Age (post-NRA) claimants in that year.

Figure 3: Mean Percentage Change in Earnings from Age t to $t+1$, by Earnings at Age t , 1990-1998



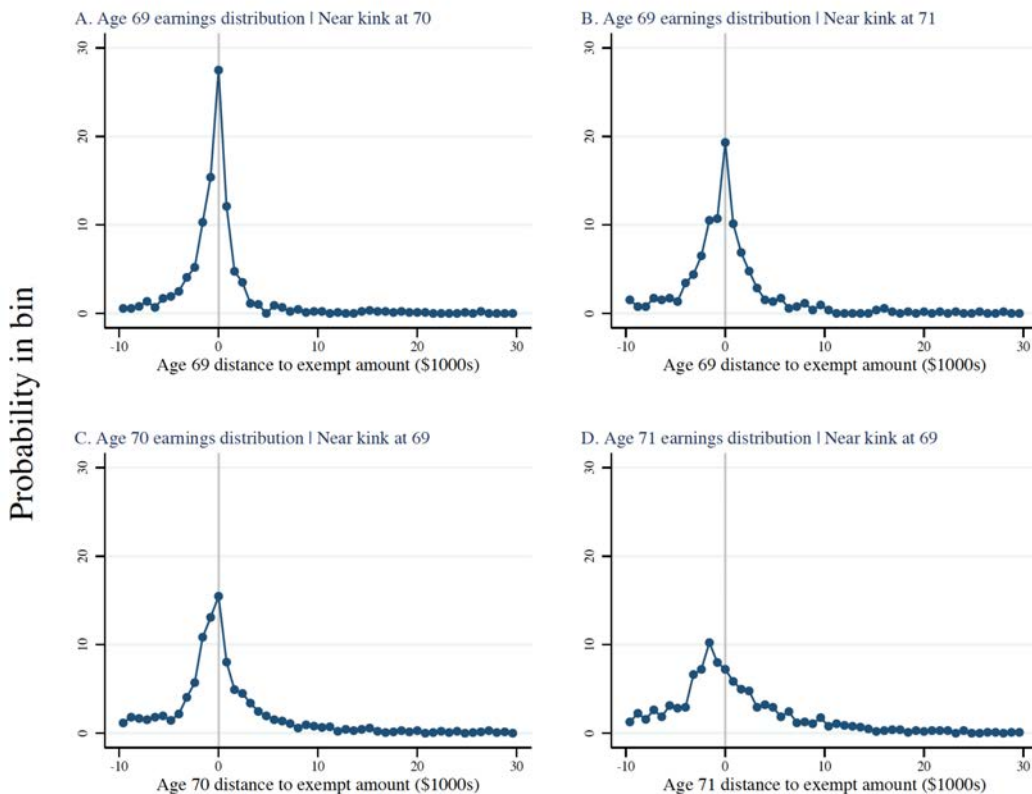
Notes: The figure shows the mean percentage change in earnings from age t to age $t+1$ (y-axis), against earnings at age t (x-axis). In Panel A, $t=69$, and in Panel B, $t=70$. Dashed lines denote 95 percent confidence intervals. Earnings are measured relative to the kink, shown at zero on the x-axis. 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 Earnings Test is eliminated for those older than Normal Retirement Age in 2000. Higher earnings growth far below the kink reflects mean reversion visible in this part of the earnings distribution at all ages.

Figure 4: Normalized Excess Mass of Claimants, Ages 69 to 72, 1983 to 1999



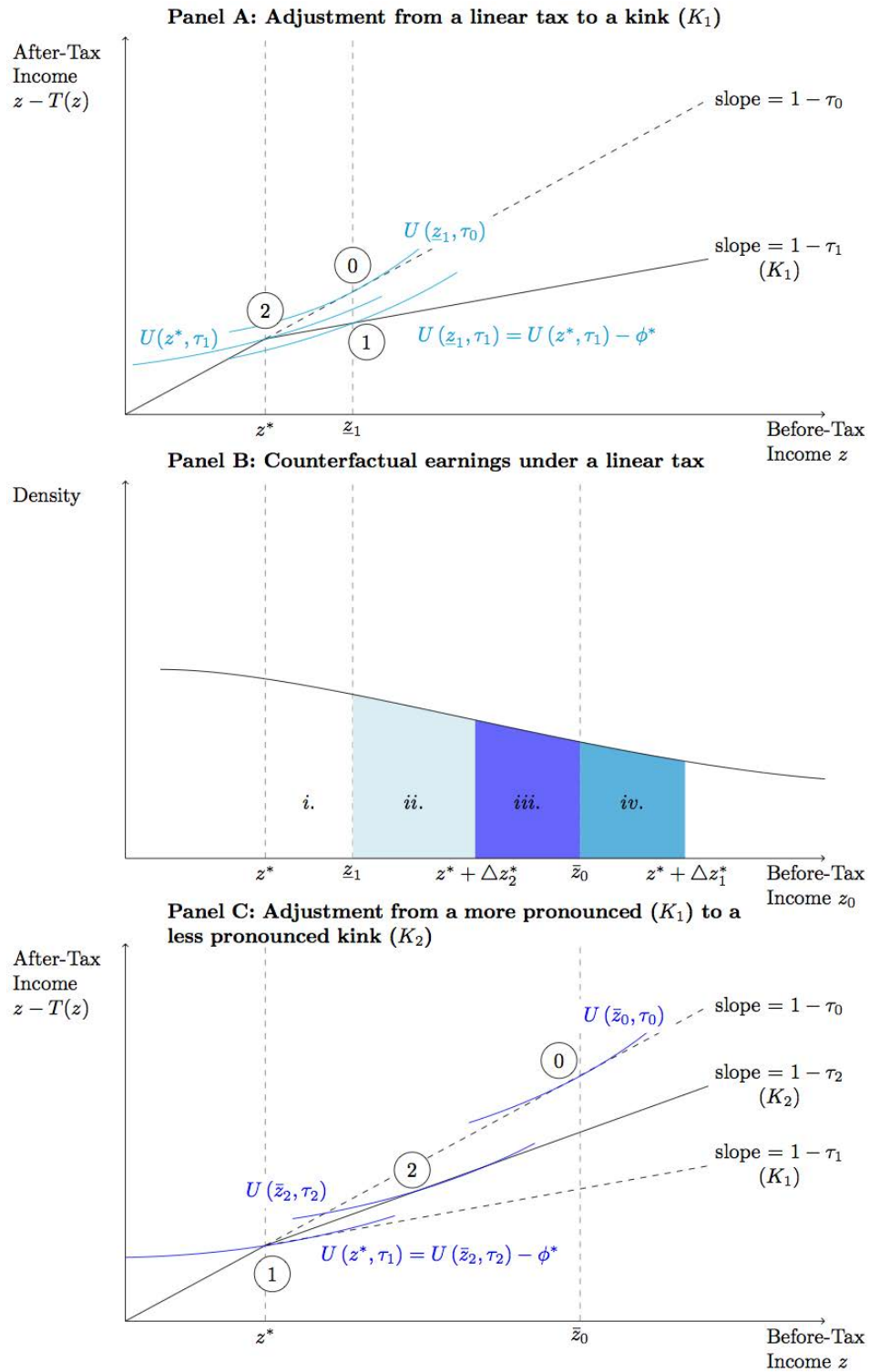
Note: See notes from Figure 2. Panel A of this figure differs from Figure 2 because here we pool 1983 to 1999 to gain extra statistical power. The continued bunching at age 71 is more evident. In the main sample, we pool only 1990 to 1999 because the benefit reduction rate was constant over this period, avoiding issues relating to the transition to a lower rate in 1990. Panel B of the figure shows normalized excess mass by age, demonstrating that excess normalized mass remains significant until age 71 and smoothly decreases from age 69 to age 72.

Figure 5: Inertia in Bunching from 69 to 70 and 71



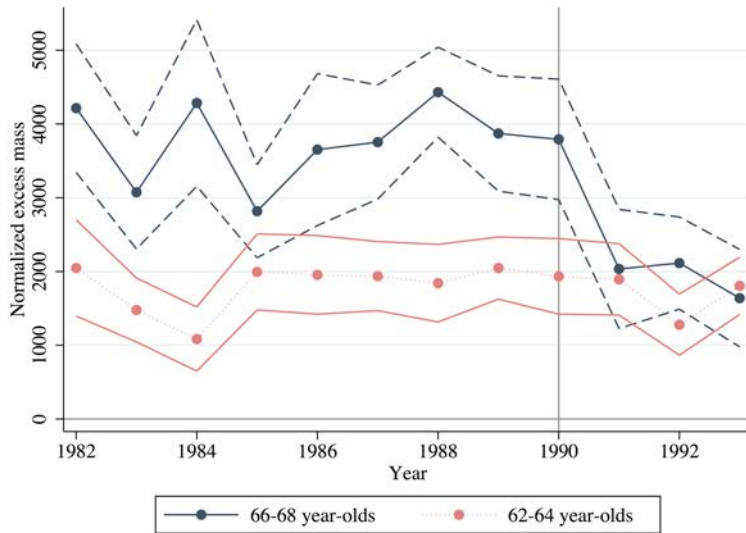
Notes: Using data from 1990 to 1999, the figure shows that those bunching at age 69 tend to remain near the kink at ages 70 and 71, and that those bunching at ages 70 and 71 were also bunching at age 69. Specifically, the figure shows the density of earnings at age 69 conditional on having earnings near the kink at age 70 (Panel A), the density of earnings at age 69 conditional on having earnings near the kink at age 71 (Panel B), the density of earnings at age 70 conditional on having earnings near the kink at age 69 (Panel C), and the density of earnings at age 71 conditional on having earnings near the kink at age 69 (Panel D). Having earnings “near the kink” is defined as having earnings within \$1,000 of the exempt amount applying to that age. See also notes from Figure B.1.

Figure 6: Bunching Responses to a Convex Kink, with Fixed Adjustment Costs



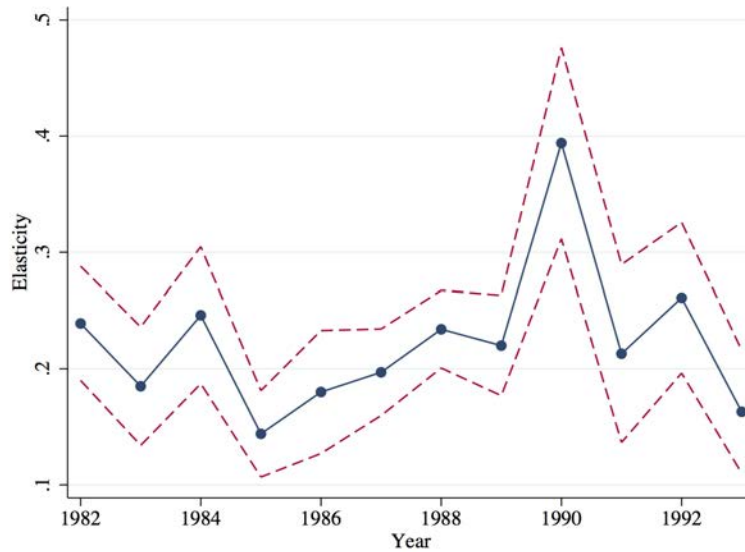
Note: See Section 6 for an explanation of the figures.

Figure 7: Comparison of Normalized Excess Mass Among 62-64 Year-Olds and 66-68 Year-Olds, 1982-1993



Notes: The figure shows normalized bunching among 62-64 year-olds and 66-68 year-olds in each year from 1982 to 1993. See other notes from Figure 2.

Figure 8: Elasticity Estimates by Year, Saez (2010) Method, 1982-1993



Notes: The figure shows elasticities estimated using the Saez (2010) method, by year from 1982 to 1993, among 66-68 year-old Social Security claimants. Dashed lines denote 95 percent 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	0.57
Observations	376,431

Notes: 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: Variation Around 1990 Policy Change

	(1)	(2)	(3)	(4)
	ε	ϕ	$\varepsilon \phi = 0$	
			1990	1989
Baseline	0.35 [0.31, 0.43]***	\$278 [58, 391]***	0.58 [0.45, 0.73]***	0.31 [0.24, 0.39]***
Uniform Density	0.21 [0.18, 0.24]***	\$162 [55, 211]***	0.36 [0.30, 0.43]***	0.19 [0.16, 0.23]***
Benefit Enhancement	0.58 [0.50, 0.72]***	\$151 [17, 226]***	0.87 [0.69, 1.11]***	0.52 [0.41, 0.66]***
Excluding FICA	0.49 [0.44, 0.59]***	\$318 [60, 364]***	0.74 [0.58, 0.94]***	0.42 [0.33, 0.54]***
Bandwidth = \$400	0.45 [0.36, 0.58]***	\$103 [0, 478]*	0.62 [0.47, 0.81]***	0.43 [0.32, 0.56]***
Bandwidth = \$1,600	0.33 [0.29, 0.43]***	\$251 [34, 407]***	0.55 [0.43, 0.72]***	0.30 [0.23, 0.40]***

Notes: The table shows estimates of the elasticity and adjustment cost using the method described in Section 6.2, investigating the 1990 reduction in the Earnings Test benefit reduction rate from 50 percent to 33.33 percent. This is our baseline because it facilitates a comparison of our estimates to the Saez (2010) method. We report bootstrapped 95 percent confidence intervals in parentheses. The baseline specification uses a nonparametric density taken from the age 72 earnings distribution, calculates the effective marginal tax rate by including the effects of the Earnings Test and federal and state income and FICA taxes, uses data from 1989 and 1990, and calculates bunching using a bin width of \$800. The estimates that include benefit enhancement use effective marginal tax rates due to the Earnings Test based on the authors' calculations relying on Coile and Gruber (2001) (assuming that individuals are considering earning just enough to trigger benefit enhancement), which imply the benefit reduction rate falls from 36% to 24% due to the 1990 policy change. Columns (1) and (2) report joint estimates with $\phi \geq 0$ imposed (consistent with theory), while Columns (3) and (4) impose the restriction $\phi = 0$. The constrained estimate in Column (3) only uses data from 1990, Column (4) uses only data from 1989. *** indicates that the left endpoint of the 99% confidence interval (CI) is greater than zero, ** the 95% CI and * the 90% CI.

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

	(1)	(2)	(3)
	ε	ϕ	$\varepsilon \phi = 0, \text{ Age } 69$
Baseline	0.42 [0.35, 0.53]***	\$90 [20, 349]***	0.38 [0.32, 0.47]***
Uniform Density	0.28 [0.24, 0.33]***	\$90 [21, 238]***	0.25 [0.22, 0.30]***
Benefit Enhancement	0.62 [0.53, 0.77]***	\$59 [13, 205]***	0.58 [0.49, 0.71]***
Excluding FICA	0.53 [0.45, 0.66]***	\$83 [19, 305]***	0.49 [0.42, 0.61]***
Bandwidth = \$400	0.39 [0.31, 0.48]***	\$62 [25, 133]***	0.36 [0.28, 0.45]***
Bandwidth = \$1,600	0.45 [0.37, 0.56]***	\$100 [20, 444]***	0.41 [0.33, 0.49]***
68-70 year-olds	0.44 [0.38, 0.58]***	\$42 [0.49, 267]**	0.43 [0.37, 0.50]***
69, 71 year-olds	0.45 [0.36, 0.86]***	\$175 [30, 1053]***	0.38 [0.32, 0.47]***
Born January-March	0.48 [0.36, 0.76]***	\$86 [10, 1008]***	0.49 [0.37, 0.71]***

Notes: The table estimates parameters using the removal of the Earnings Test at age 70, using data on 69-71 year-olds in 1990-1999. The estimates of bunching at age 70 are potentially affected by the coarse measure of age that we use, as explained in the main text. Thus, we use both age 70 and age 71 in estimating these results, and alternatively use only ages 69 and 71, which shows very similar results. The final row shows the results only for those born in January to March, again to address this issue. For this sample, we pool 1983-1989 and 1990-1999 (accounting for the different benefit reduction rates in each period) to maximize statistical power. See also notes from Table 2.

Table 4: Estimates of Elasticity and Adjustment Cost: Pooling 69/70 Transition and 1989/1990 Transition

	(1)	(2)
	ε	ϕ
Baseline	0.39 [0.34, 0.46]***	\$160 [59, 362]***
Uniform Density	0.22 [0.20, 0.25]***	\$105 [47, 185]***
Benefit Enhancement	0.62 [0.55, 0.75]***	\$100 [33, 211]***
Excluding FICA	0.41 [0.37, 0.56]***	\$67 [9, 192]***
Bandwidth = \$400	0.46 [0.39, 0.56]***	\$94 [25, 399]***
Bandwidth = \$1,600	0.37 [0.32, 0.45]***	\$135 [43, 299]***

Notes: This table implements our “comparative static” method, applied to pooled data from two policy changes: (1) around the 1989/1990 transition analyzed in Table 2, and (2) around the age 69/70 transition analyzed in Table 3. The table shows extremely similar results to the dynamic specification in Table 5, where we also pool data from around these two policy changes. See also notes from Tables 2 and 3.

Table 5: Estimates of Elasticity and Adjustment Cost Using Dynamic Model

	(1)	(2)	(3)	(4)	(5)
	ε	ϕ	π_1	$\pi_1\pi_2$	$\pi_1\pi_2\pi_3$
Baseline	0.36 [0.34, 0.40]***	\$243 [34, 671]***	0.64 [0.39, 1.00]***	0.22 [0.00, 0.94]*	0.00 [0.00, 0.14]*
Uniform Density	0.21 [0.20, 0.23]***	\$81 [31, 183]***	1.00 [0.72, 1.00]***	0.31 [0.00, 0.92]***	0.00 [0.00, 0.16]***
Benefit Enhancement	0.59 [0.54, 0.64]***	\$53 [18, 169]***	1.00 [0.76, 1.00]***	0.37 [0.00, 1.00]***	0.00 [0.00, 0.084]***
Excluding FICA	0.40 [0.37, 0.43]***	\$55 [9, 165]***	1.00 [1.00, 1.00]***	0.00 [0.00, 0.00]***	0.00 [0.00, 0.00]***
Bandwidth = \$400	0.40 [0.36, 0.44]***	\$74 [20, 271]***	1.00 [0.74, 1.00]***	0.47 [0.094, 0.94]***	0.00 [0.00, 0.20]***
Bandwidth = \$1,600	0.36 [0.34, 0.39]***	\$99 [19, 401]***	0.88 [0.40, 1.00]***	0.52 [0.043, 1.00]***	0.00 [0.00, 0.071]***

Notes: The table shows estimates of the elasticity and adjustment cost using the dynamic method described in Section 7. The table reports the elasticity ε , the adjustment cost ϕ , and the cumulative probability in each period t of having drawn $\phi_t > 0$ for each period following the policy change, i.e. π_1 as well as $\pi_1\pi_2$. π_1 , π_2 , π_3 , π_4 , and π_5 are not all separately identified; only the cumulative probabilities are identified, i.e. π_1 , $\pi_1\pi_2$, $\pi_1\pi_2\pi_3$, $\pi_1\pi_2\pi_3\pi_4$, and $\pi_1\pi_2\pi_3\pi_4\pi_5$. The reason is that once one of π_1 , π_2 , π_3 , π_4 , or π_5 equals zero, none of the subsequent probabilities is identified. The model is estimated by matching predicted and observed bunching, using bunching on 66-68 year-olds (pooled) for each year 1987-1992, and bunching on 1990-1999 (pooled) for each age 67-72. Estimates of $\pi_1\pi_2\pi_3$ are statistically significantly different from zero, even though the reported point estimates are 0.00, because the point estimates are positive but round to zero. $\pi_1\pi_2\pi_3\pi_4$ and $\pi_1\pi_2\pi_3\pi_4\pi_5$ are always estimated to 0.00, with a confidence interval that rules out more than a small value (results available upon request). The results are comparable when we investigate only the 1989/1990 or 69/70 policy changes alone using the dynamic specification (results available upon request). *** indicates $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A Appendix (for online publication)

A.0.1 Polynomial Adjustment Costs

We now extend the adjustment cost to encompass a polynomial adjustment cost, allowing for greater generality than a fixed cost. We begin with an adjustment cost that increases linearly in the size of the adjustment, which illustrates how the method generalizes for higher-order polynomials. 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 $d\tau_1$:

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

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:

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

In this case, we need at least three kinks to separately identify $(\varepsilon, \phi^F, \phi^L)$. A similar argument generalizes this to the case of any polynomial adjustment cost: for a polynomial adjustment cost of order n , we need $n + 1$ moments to identify these parameters as well as the elasticity.

A.1 Dynamic Model with Forward-Looking Behavior

We present in this appendix a version of the dynamic model in Section 6.3 in which we allow for forward-looking behavior. The key difference in implications is that in addition to a gradual, lagged response to policy changes, this version of the model also predicts anticipatory adjustment by agents when policy changes are anticipated in advance. We have essentially the same setting as in Section 6.3, except that we will alter three of the assumptions. First, in each period, an individual draws a cost of adjustment, $\tilde{\phi}_t$, from a discrete distribution, which takes a value of ϕ with probability π and a value of 0 with probability $1 - \pi$.¹³ Second, individuals make decisions over a finite horizon, living until Period \bar{T} . In period 0, the individuals face a linear tax schedule, $T_0(z) = \tau_0 z$, with marginal tax rate τ_0 . In period 1, a kink, K_1 , is introduced at the earnings level z^* . This tax schedule is implemented for \mathcal{T}_1 periods, after which the tax schedule features a smaller kink, K_2 , at the earnings level z^* . The smaller kink is present until period \mathcal{T}_2 , after which we return to the linear tax schedule, T_0 . As before, the kink K_j , $j \in \{1, 2\}$, features a top marginal tax rate of τ_j for earnings above z^* .¹⁴ Finally, in each period, individuals solve this maximization problem:

$$\max_{(c_{a,t}, z_{a,t})} v(c_{a,t}, z_{a,t}; a, z_{a,t-1}) + \delta V_{a,t+1}(z_{a,t}, A_{a,t}), \quad (\text{A.1})$$

where $v(c_{a,t}, z_{a,t}; a, z_{a,t-1}) \equiv u(c_{a,t}, z_{a,t}; a) - \tilde{\phi}_t \cdot \mathbf{1}(z_{a,t} \neq z_{a,t-1})$, δ is the discount factor, and $V_{a,t+1}$ is the value function moving forward in Period $t + 1$:

$$V_{a,t+1}(\zeta, A_{a,t}) = \mathbb{E}_\phi \left[\max_{(c_{a,t+1}, z_{a,t+1})} v(c_{a,t+1}, z_{a,t+1}; a, \zeta) + \delta V_{a,t+2}(z_{a,t+1}, A_{a,t+1}) \right]. \quad (\text{A.2})$$

$V_{a,t+1}$ is a function of where the individual has chosen to earn in Period t and assets $A_{a,t}$. The expectation $\mathbb{E}_\phi[\cdot]$ is taken over the distribution of $\tilde{\phi}_t$. The intertemporal budget constraint

¹³For expositional purposes, we constrain the probability of drawing a nonzero fixed costs to be π in all periods. Thus, the terms from Section 6.3 of the form $\prod \pi_j$ simplify to π^j in this appendix. All results go through with the more flexible distribution of adjustment costs in Section 6.3.

¹⁴In Section 6.3, we do not specify time \mathcal{T}_2 , when the smaller kink, K_2 , is removed, as it is not relevant to the case where individuals are not forward-looking.

is:

$$A_{a,t} = (1+r)(A_{a,t-1} + z_{a,t} - T(z_{a,t}) - c_{a,t}). \quad (\text{A.3})$$

We assume that $\delta(1+r) = 1$. Because individuals have quasilinear preferences, this implies that consumption can be set to disposable income in each period: $c_{a,t} = z_{a,t} - T(z_{a,t})$. We therefore use the following shorthand:

$$\begin{aligned} u_a^j(z) &= u(z - T_j(z), z; a) \\ V_{a,t}(z) &= V_{a,t}(z, A_{a,t-1}) \end{aligned} \quad (\text{A.4})$$

Next, we define two operators that measure the utility gain (or loss) following a discrete change in earnings:

$$\begin{aligned} \Delta u_a^j(z, z') &= u_a^j(z) - u_a^j(z') \\ \Delta V_{a,t}(z, z') &= V_{a,t}(z) - V_{a,t}(z') \end{aligned} \quad (\text{A.5})$$

In each case above, the utility and utility differential depend on the tax schedule. We define z_a^j as the optimal level of earnings under a frictionless, static optimization problem, facing the tax schedule T_j . We will refer to the frictionless, *dynamic* optimum in any given period as $\tilde{z}_{a,t}$.¹⁵ This is the optimal level of earnings when there is a fixed cost of zero drawn in the *current* period, but a nonzero fixed cost may be drawn in future periods. We will also make a distinction between two types of earnings adjustments: *active* and *passive*. An *active* earnings adjustment takes place in the presence of a nonzero fixed cost, while a *passive* earnings adjustment takes place only when a fixed cost of zero is drawn. We solve the model recursively, beginning in the regime after time \mathcal{T}_2 , when the smaller kink, K_2 , has been removed, continuing with the solution while the kink K_2 is present between times \mathcal{T}_1 and \mathcal{T}_2 , and finally considering the first regime when the kink K_1 is present between time period 1 and \mathcal{T}_1 .¹⁶

A.1.1 Earnings between \mathcal{T}_2 and $\bar{\mathcal{T}}$

We will now derive the value function $V_{a,\mathcal{T}_2+1}(z)$. We begin with the following result: If an individual with initial earnings z makes an active adjustment in period $t > \mathcal{T}_2 + 1$, then it must be the case that

$$\frac{1 - (\delta\pi)^{\mathcal{T}_1+1-t}}{1 - \delta\pi} \Delta u_a^0(z_a^0, z) \geq \phi. \quad (\text{A.6})$$

We demonstrate this result with a constructive proof, showing the result for periods $\bar{\mathcal{T}}$ and $\bar{\mathcal{T}} - 1$. Because the tax schedule is constant throughout this terminal period, the frictionless, dynamic optimum is equal to the static optimum: $\tilde{z}_{a,t} = z_a^0$. First, consider an agent in period $\bar{\mathcal{T}}$, with initial earnings z , who is considering maintaining earnings at z or paying the fixed cost ϕ and making an active adjustment to z_a^0 , the frictionless, dynamic

¹⁵In a model with no forward-looking behavior, $z_a^j = \tilde{z}_{a,t}$.

¹⁶Our recursive method can be extended to the case of multiple, successive kinks. The effect on bunching of a sequence of more kinks depends on the relative size of the successive kinks.

optimum in period \bar{T} . The agent will make the adjustment if:

$$\begin{aligned}\Delta u_a^0(z_a^0, z) &\geq \phi \\ &= \frac{1 - \delta\pi}{1 - \delta\pi} \phi.\end{aligned}\tag{A.7}$$

Rearranging terms, we have satisfied the inequality in (A.6).

Now consider agents in period $\bar{T} - 1$ with initial earnings z . There are two types, those who would make an active adjustment to z_a^0 in period \bar{T} if the earnings z are carried forward and those who would not. Consider those who would not. If the agent remains with earnings of z , then utility will be $u_a^0(z) + \delta V_{a,\bar{T}}(z) = u_a^0(z) + \delta [\pi (u_a^0(z)) + (1 - \pi) u_a^0(z_a^0)]$. If the agent actively adjusts to z_a^0 , then utility will be $u_a^0(z_a^0) - \phi + \delta u_a^0(z_a^0)$. The agent will actively adjust in period $\bar{T} - 1$ if:

$$\begin{aligned}\Delta u_a^0(z_a^0, z) &\geq \frac{1}{1 + \delta\pi} \phi \\ &= \frac{1 - \delta\pi}{1 - (\delta\pi)^2} \phi.\end{aligned}\tag{A.8}$$

Once again, rearranging terms confirms that (A.6) holds. Finally, consider agents who would actively adjust from z to z_a^0 if earnings level z is carried forward. In this case, the agent's utility when remaining at z is:

$$\begin{aligned}u_a^0(z) + \delta V_{a,\bar{T}}(z) &= u_a^0(z) + \delta [\pi (u_a^0(z_a^0) - \phi) + (1 - \pi) u_a^0(z_a^0)] \\ &= u_a^0(z) + \delta (u_a^0(z_a^0) - \pi\phi).\end{aligned}\tag{A.9}$$

Intuitively, the agent will receive the optimal level of utility in the next period, and with probability π the agent will have to pay the fixed cost to achieve it. Similarly, the agent's utility after actively adjusting to z_a^0 in period $\bar{T} - 1$ is $u_a^0(z_a^0) - \phi + \delta u_a^0(z_a^0)$. The agent will therefore adjust in period \bar{T} if:

$$\Delta u_a^0(z_a^0, z) \geq (1 - \delta\pi) \phi.\tag{A.10}$$

However, we know from (A.7) that this already holds for the agent who actively adjusts in period \bar{T} . Finally, note that (A.7) implies (A.8). It follows that in period $\bar{T} - 1$, adjustment implies (A.7). We can similarly show the result for earlier periods by considering separately: (a) those who would actively adjust in the current period, but not in any future period; and (b) those who would adjust in some future period. Both types will satisfy the key inequality. As a corollary, note that if an individual with initial earnings z makes an active adjustment in period $t > \mathcal{T}_2 + 1$, then she will also find it optimal to do so in any period t' , where $\mathcal{T}_2 < t' < t$. To see this, note that if (A.6) holds for t , then it also holds for $t' < t$. It follows that the agent would also actively adjust in period t' .

Now consider an agent who earns z in period \mathcal{T}_2 . Note that our results above imply that any active adjustment that takes place after \mathcal{T}_2 will only happen in period $\mathcal{T}_2 + 1$. These agents will receive a stream of discounted payoffs of $u_a^0(z_a^0)$ for $\bar{T} - \mathcal{T}_2$ periods, *i.e.*

$\sum_{j=0}^{\bar{T}-\mathcal{T}_2-1} \delta^j u_a^0(z_a^0) = \frac{1-\delta^{\bar{T}-\mathcal{T}_2}}{1-\delta} u_a^0(z_a^0)$, and pay a fixed cost of ϕ in period \mathcal{T}_2 with probability π . Otherwise, an agent will adjust to the dynamic frictionless optimum z_a^0 only when a fixed cost of zero is drawn. In the latter case, the agent receives a payoff of $u_a^0(z)$ until a fixed cost of zero is drawn, after which, the agent receives $u_a^0(z_a^0)$. We can therefore derive the following value function:¹⁷

$$V_{a,\mathcal{T}_2+1}(z) = \begin{cases} \frac{1-\delta^{\bar{T}-\mathcal{T}_2}}{1-\delta} u_a^0(z_a^0) - \pi\phi & \text{if } \frac{1-(\delta\pi)^{\bar{T}-\mathcal{T}_2}}{1-\delta\pi} \Delta u_a^0(z_a^0, z) \geq \phi \\ \frac{1-\delta^{\bar{T}-\mathcal{T}_2}}{1-\delta} u_a^0(z_a^0) - \pi \frac{1-(\delta\pi)^{\bar{T}-\mathcal{T}_2}}{1-\delta\pi} \Delta u_a^0(z_a^0, z) & \text{otherwise} \end{cases}. \quad (\text{A.11})$$

To gain some intuition for (A.6), note that the left side of (A.6) is the net present value of the stream of the utility differential once the agent adjusts from z to z_a^0 . If this exceeds the up-front cost of adjustment, ϕ , then the agent actively adjusts. The discount factor for j periods in the future, however, is $(\delta\pi)^j$, instead of only δ^j . The reason is that current adjustment only affects future utility j periods from now if j consecutive nonzero fixed costs are drawn, which happens with probability π^j . To better understand our second result regarding the timing of active changes, note that if the gains from adjustment over $\bar{T} - t$ periods exceed the up-front cost, then the agent should also be willing to adjust in period $t' < t$ and accrue $\bar{T} - t'$ periods of this gain, for the same up-front cost of ϕ .

A.1.2 Earnings between \mathcal{T}_1 and \mathcal{T}_2

We now derive the value function $V_{a,\mathcal{T}_1+1}(z)$. In this case, the dynamic frictionless optimum in each period, $\tilde{z}_{a,t}$, is not constant. Intuitively, the agent trades off the gains from adjusting earnings in response to K_2 with the effect of this adjustment on the value function V_{a,\mathcal{T}_2+1} . In general, the optimum is defined as:

$$\tilde{z}_{a,t} = \arg \max_{z \in [z_a^2, z_a^0]} \frac{1 - (\delta\pi)^{\mathcal{T}_2+1-t}}{1 - \delta\pi} u_a^2(z) + \delta^{\mathcal{T}_2+1-t} \pi^{\mathcal{T}_2-t} V_{a,\mathcal{T}_2+1}(z). \quad (\text{A.12})$$

We restrict the maximization to the interval $[z_a^2, z_a^0]$, since reducing earnings below z_a^2 or raising earnings above z_a^0 weakly reduces utility in any current and all future periods for $t > \mathcal{T}_1$. From (A.11), we know that V_{a,\mathcal{T}_2+1} is continuous, and thus the solution in (A.12) exists.¹⁸ We present two results analogous to those in Section A.1.1, without proof. The proofs, nearly identical to those in the previous section, are available upon request. First, if an individual with initial earnings z makes an active adjustment in period t , $\mathcal{T}_1 < t \leq \mathcal{T}_2$, then:

$$\frac{1 - (\delta\pi)^{\mathcal{T}_2+1-t}}{1 - \delta\pi} \Delta u_a^2(\tilde{z}_{a,t}, z) + \delta^{\mathcal{T}_2+1-t} \pi^{\mathcal{T}_2-t} \Delta V_{a,\mathcal{T}_2+1}(\tilde{z}_{a,t}, z) \geq \phi. \quad (\text{A.13})$$

Furthermore, if an individual with initial earnings z makes an active adjustment in period t , $\mathcal{T}_1 < t \leq \mathcal{T}_2$, then she will also find it optimal to do so in any period t' , where $\mathcal{T}_1 < t' < t$.

The condition in (A.13) differs from that in (A.6) because the effect of adjustment on the

¹⁷The expected utility for passive adjusters is constructed recursively, working backward from period \bar{T} to period $\mathcal{T}_2 + 1$.

¹⁸Technically, we can see from (A.11) that while the function V_{a,\mathcal{T}_2+1} is continuous, it is kinked, which creates a nonconvexity. Thus, the solution in (A.12) may not always be single-valued. In such cases, we define $\tilde{z}_{a,t}$ as the lowest level of earnings that maximizes utility.

utility beyond period \mathcal{T}_2 is taken into account, in addition to the up-front cost of adjustment, ϕ . Any adjustment in this time interval, active or passive, will be to the dynamic, frictionless optimum for the current period, $\tilde{z}_{a,t}$. As before, (A.13) implies that all active adjustment occurring between $\mathcal{T}_1 + 1$ and \mathcal{T}_2 takes place in period $\mathcal{T}_1 + 1$. Those who adjust in period $\mathcal{T}_1 + 1$ will earn $\tilde{z}_{a,\mathcal{T}_1+1}$. Thereafter, they only adjust to $\tilde{z}_{a,t}$ when a fixed cost of zero is drawn. Likewise, those who only adjust passively earn z_{a,\mathcal{T}_1} in period $\mathcal{T}_1 + 1$, and thereafter adjust to $\tilde{z}_{a,t}$ when a fixed cost of zero is drawn. We can therefore derive the following value function:

$$V_{a,\mathcal{T}_1+1}(z) = \left\{ \begin{array}{l} \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-1} \delta^j u_a^2(\tilde{z}_{a,\mathcal{T}_1+1+j}) + \delta^{\mathcal{T}_2-\mathcal{T}_1} \Delta V_{a,\mathcal{T}_2+1}(\tilde{z}_{a,\mathcal{T}_2}) \\ - \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-2} \frac{(\delta\pi)^{\mathcal{T}_2-\mathcal{T}_1}}{\pi^{j+1}} \Delta V_{a,\mathcal{T}_2+1}(\tilde{z}_{a,\mathcal{T}_1+2+j}, \tilde{z}_{a,\mathcal{T}_1+1+j}) \\ - \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-2} \frac{1-(\delta\pi)^{\mathcal{T}_2-\mathcal{T}_1-1-j}}{1-\delta\pi} \delta^{j+1} \pi \Delta u_a^2(\tilde{z}_{a,\mathcal{T}_1+2+j}, \tilde{z}_{a,\mathcal{T}_1+1+j}) \\ - \pi \phi \\ \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-1} \delta^j u_a^2(\tilde{z}_{a,\mathcal{T}_1+1+j}) + \delta^{\mathcal{T}_2-\mathcal{T}_1} \Delta V_{a,\mathcal{T}_2+1}(\tilde{z}_{a,\mathcal{T}_2}) \\ - \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-2} \frac{(\delta\pi)^{\mathcal{T}_2-\mathcal{T}_1}}{\pi^{j+1}} \Delta V_{a,\mathcal{T}_2+1}(\tilde{z}_{a,\mathcal{T}_1+2+j}, \tilde{z}_{a,\mathcal{T}_1+1+j}) \\ - \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-2} \frac{1-(\delta\pi)^{\mathcal{T}_2-\mathcal{T}_1-1-j}}{1-\delta\pi} \delta^{j+1} \pi \Delta u_a^2(\tilde{z}_{a,\mathcal{T}_1+2+j}, \tilde{z}_{a,\mathcal{T}_1+1+j}) \\ - \pi \left\{ \begin{array}{l} \sum_{j=0}^{\mathcal{T}_2-\mathcal{T}_1-1} (\delta\pi)^j \Delta u_a^2(\tilde{z}_{a,\mathcal{T}_1+1}, z) \\ - \delta^{\mathcal{T}_2-\mathcal{T}_1} \pi^{\mathcal{T}_2+1-\mathcal{T}_1} \Delta V_{a,\mathcal{T}_2+1}(\tilde{z}_{a,\mathcal{T}_1+1}, z) \end{array} \right. \end{array} \right. \begin{array}{l} \text{if (A.13) is satisfied} \\ \text{when } t = \mathcal{T}_1 + 1 \\ \\ \\ \\ \text{otherwise} \end{array} \quad (\text{A.14})$$

The first case in (A.14) applies to those who actively adjust in period $\mathcal{T}_1 + 1$ and passively adjust thereafter. The first line is the utility that would accrue if a fixed cost of zero were drawn in each period. The next two lines represent the deviation from this stream of utility, due to nonzero fixed costs potentially drawn in periods $\mathcal{T}_1 + 1$ through \mathcal{T}_2 . The final line represents the fixed cost that is paid in period $\mathcal{T}_1 + 1$ with probability π . The second case in (A.14) applies to those who only passively adjust. The first three lines remain the same. The final two lines represent a loss in utility attributed to fact that earnings in period $\mathcal{T}_1 + 1$ may not be $\tilde{z}_{a,\mathcal{T}_1+1}$. Note that earnings in period \mathcal{T}_1 can only affect utility through this last channel.

A.1.3 Earnings between Period 1 and \mathcal{T}_1

Earnings during the first period, when the kink K_1 is present, can be derived similarly. The dynamic, frictionless optimum is now defined as:

$$\tilde{z}_{a,t} = \arg \max_{z \in [z_a^1, z_a^0]} \frac{1 - (\delta\pi)^{\mathcal{T}_1+1-t}}{1 - \delta\pi} u_a^1(z) + \delta^{\mathcal{T}_1+1-t} \pi^{\mathcal{T}_1-t} V_{a,\mathcal{T}_1+1}(z).^{19} \quad (\text{A.15})$$

Similar to the other cases, if an individual with initial earnings z makes an active adjustment in period t , $0 < t \leq \mathcal{T}_1$, then it must be the case that

$$\frac{1 - (\delta\pi)^{\mathcal{T}_1+1-t}}{1 - \delta\pi} \Delta u_a^1(\tilde{z}_{a,t}, z) + \delta^{\mathcal{T}_1+1-t} \pi^{\mathcal{T}_1-t} \Delta V_{a,\mathcal{T}_1+1}(\tilde{z}_{a,t}, z) \geq \phi. \quad (\text{A.16})$$

Furthermore, if an individual with initial earnings z makes an active adjustment in period t , $0 < t \leq \mathcal{T}_1$, then she will also find it optimal to do so in any period t' , where $0 < t' < t$. Again, this implies that all active adjustment will take place in period 1. Since individuals begin with earnings of z_a^0 , we know that all active adjustment will be downward. Thereafter, it can be shown that $\tilde{z}_{a,t}$ is weakly increasing, and upward adjustment will occur passively.

A.1.4 Characterizing Bunching

Given these results, we can now derive expressions for excess mass at z^* analogous to (8) and (9). For notational convenience, we define $\mathcal{A}_j(z)$ as the set of individuals, a , with initial earnings z who actively adjust in period j . Again, denote B_1^t as bunching at K_1 in period $t \in [1, \mathcal{T}_1]$. We have the following generalized version of (8):

$$\begin{aligned} B_1^t = & \int_{z^*}^{z^* + \Delta z_1^*} \left[\mathbf{1} \{ \tilde{z}_{a,1} = z^*, a \in \mathcal{A}_1(\zeta) \} \right. \\ & + \sum_{j=1}^t (1 - \pi^j) \pi^{t-j} \mathbf{1} \{ \sup \{ l \mid l \leq t, \tilde{z}_{a,l} = z^* \} = j, a \notin \mathcal{A}_1(\zeta) \} \\ & \left. - \sum_{j=1}^{t-1} (1 - \pi^{t-j}) \mathbf{1} \{ \sup \{ l \mid l \leq t, \tilde{z}_{a,l} = z^* \} = j, a \in \mathcal{A}_1(\zeta) \} \right] h_0(\zeta) d\zeta. \end{aligned} \quad (\text{A.17})$$

We have partitioned the set of potential bunchers into three groups in (A.17). In the first line, we have the set of active bunchers in period 1. In the second line, we capture individuals who are passive bunchers, *i.e.* $a \notin \mathcal{A}_1(z_a^0)$. For $j \in [1, t-1]$, the indicator function selects the individual who has $\tilde{z}_{a,j} = z^*$ but $\tilde{z}_{a,j+1} \neq z^*$. Since $\tilde{z}_{a,t}$ is weakly increasing, the optimal earnings for this individual is z^* in periods 1 through $j-1$. The probability that the individual bunches by period j is $1 - \pi^j$. Thereafter, the individual will de-bunch if a fixed cost of zero is drawn. The probability of only drawing nonzero fixed costs thereafter is π^{t-j} . For $j = t$, the indicator function selects agents for whom $\tilde{z}_{a,t} = z^*$. Their probability

¹⁹Note, the objective function now features two potential nonconvexities. In cases where the solution is multi-valued, we again define $\tilde{z}_{a,t}$ as the lowest earnings level from the set of solutions.

of passively bunching by period t is $1 - \pi^t$. The third line captures the outflow of active bunchers, for whom $\tilde{z}_{a,t}$ ceases to be z^* starting in period j . The probability of having drawn a nonzero fixed cost and de-bunching since period j is $1 - \pi^{t-j}$.

Equation (A.17) differs from (8) in three key ways. First, the set of active bunchers in period 1 is different, as can be seen by comparing (A.16) and the relevant condition for active bunchers in Section 6.3, $\Delta u_a^1(z^*, z_a^0) \geq \phi$. The utility gain accrues for multiple periods in the forward-looking case, increasing the probability of actively bunching, but the effect of adjustment on future payoffs via V_{a, \mathcal{T}_1+1} may either reinforce or offset this incentive. Furthermore, passive bunchers are (weakly) less likely to remain bunching, as they de-bunch in anticipation of policy changes in future periods. To see this, note that the π^{t-j} factor is decreasing in t . Finally, the set of active bunchers similarly de-bunch passively, in anticipation of future policy changes. The model therefore predicts a gradual outflow from the set of bunchers, in anticipation of the shift from K_1 to K_2 . Nonetheless, the overall net change in bunching over time is ambiguous.

We now turn to bunching starting in period $\mathcal{T} + 1$. It can be shown, similarly to the cases above, that if an agent would be willing to actively bunch in period $\mathcal{T}_1 + 1$, she will also be willing to actively bunch in earlier periods. Thus, the only active adjustment occurring that affects bunching will be de-bunching. The set of individuals who actively de-bunch, $\mathcal{A}_{\mathcal{T}_1+1}(z^*)$, are those for whom (A.13) is satisfied, when evaluated at $t = \mathcal{T}_1 + 1$ and $z = z^*$. The remaining changes in bunching between \mathcal{T}_1 and \mathcal{T}_2 consist of passive adjustment among those who were bunching at the end of period \mathcal{T}_1 . We can thus characterize B_2^t , bunching at K_2 in period $t \in [\mathcal{T}_1 + 1, \overline{\mathcal{T}}]$, in a manner analogous to (9):²⁰

$$\begin{aligned}
B_2^t &= \int_{z^*}^{z^* + \Delta z_1^*} \left[\mathbf{1} \{a \notin \mathcal{A}_{\mathcal{T}_1+1}(z^*)\} \right. \\
&\times \left\{ \pi^{t-\mathcal{T}_1} \mathbf{1} \{\tilde{z}_{a, \mathcal{T}_1+1} \neq z^*\} + \sum_{j=\mathcal{T}_1+1}^t \pi^{t-j} \mathbf{1} \{\sup \{l \mid l \leq t, \tilde{z}_{a,l} = z^*\} = j\} \right\} \\
&\times \left\{ \mathbf{1} \{\tilde{z}_{a,1} = z^*, a \in \mathcal{A}_1(\zeta)\} \right. \\
&+ \sum_{j=1}^{\mathcal{T}_1} (1 - \pi^j) \pi^{\mathcal{T}_1-j} \mathbf{1} \{\sup \{l \mid l \leq \mathcal{T}_1, \tilde{z}_{a,l} = z^*\} = j, a \notin \mathcal{A}_1(\zeta)\} \\
&\left. \left. - \sum_{j=1}^{\mathcal{T}_1-1} (1 - \pi^{\mathcal{T}_1-j}) \mathbf{1} \{\sup \{l \mid l \leq \mathcal{T}_1, \tilde{z}_{a,l} = z^*\} = j, a \in \mathcal{A}_1(\zeta)\} \right\} \right] h_0(\zeta) d\zeta.
\end{aligned} \tag{A.18}$$

The first line of this expression selects only those agents who do not actively de-bunch immediately in period $\mathcal{T}_1 + 1$. The second line selects the set of agents who would like to passively de-bunch beginning at some period $j > \mathcal{T}_1 + 1$. They are weighted by the probability of continuing to bunch due to consecutive draws of nonzero fixed costs. The final three lines

²⁰When $\mathcal{T}_1 = 1$, we set the very last summation to zero.

select agents from the set of bunchers at the end of period T_1 . As with our simpler model in Section 6.3, bunching gradually decreases following a reduction in the size of the kink from K_1 to K_2 . However, in this case, the reduction is due to both fixed costs of adjustment and anticipation of the removal of the kink K_2 in period $T_2 + 1$.

As in Section 6.3, the richer model in this appendix nests the dynamic model without forward looking behavior when we set $\delta = 0$, collapses to the comparative static model of Sections 6.1-6.2 if we additionally assume that $\pi = 1$ and is equivalent to the frictionless model when either $\phi = 0$ or $\pi = 0$.

A.2 Derivation of Bunching Formulae with Heterogeneity

A.2.1 Comparative Static Model

Under heterogenous preferences, our estimates can be interpreted as reflecting average parameters among the set of bunchers (as in Saez, 2010, and Kleven and Waseem, 2013). As described in the main text, suppose $(\varepsilon_i, \phi_i, a_i)$ 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 $h_0^*(z, \varepsilon, \phi)$ under a linear tax of τ_0 . Assume that the density of earnings is constant over the interval $[z^*, z^* + \Delta z^*]$, conditional on ε and ϕ . When moving from no kink to a kink, we derive a formula for bunching at K_1 in the presence of heterogeneity as follows:

$$\begin{aligned}
B_1 &= \iiint_{\underline{z}_1}^{z^* + \Delta z_1^*} h_0^*(\zeta, \varepsilon, \varphi) d\zeta d\varepsilon d\varphi \\
&= \iint [z^* + \Delta z_1^* - \underline{z}_1] h_0^*(z^*, \varepsilon, \varphi) d\varepsilon d\varphi \\
&= h_0(z^*) \cdot \iint [z^* + \Delta z_1^* - \underline{z}_1] \frac{h_0^*(z^*, \varepsilon, \varphi)}{h_0(z^*)} d\varepsilon d\varphi \\
&= h_0(z^*) \cdot \mathbb{E}[z^* + \Delta z_1^* - \underline{z}_1], \tag{A.19}
\end{aligned}$$

where we have used the assumption of constant $h_0^*(\cdot)$ in line two, $h_0(z^*) = \iint h_0^*(z^*, \varepsilon, \varphi) d\varepsilon d\varphi$, 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:

$$b_1 = z^* + \mathbb{E}[\Delta z_1^*] - \mathbb{E}[\underline{z}_1]. \tag{A.20}$$

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.

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:

$$\begin{aligned}
\tilde{B}_2 &= \iiint_{\underline{z}_1}^{\bar{z}_0} h_0^*(\zeta, \epsilon, \varphi) d\zeta d\epsilon d\varphi \\
&= \iint [\bar{z}_0 - \underline{z}_1] h_0^*(z^*, \epsilon, \varphi) d\epsilon d\varphi \\
&= h_0(z^*) \cdot \iint [\bar{z}_0 - \underline{z}_1] \frac{h_0^*(z^*, \epsilon, \varphi)}{h_0(z^*)} d\epsilon d\varphi \\
&= h_0(z^*) \cdot \mathbb{E}[\bar{z}_0 - \underline{z}_1].
\end{aligned} \tag{A.21}$$

Similarly, normalized bunching can now be expressed as follows:

$$\tilde{b}_2 = \mathbb{E}[\bar{z}_0] - \mathbb{E}[\underline{z}_1]. \tag{A.22}$$

Once again, the expectations are taken over the population of bunchers.

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 (5) and (7) and the identities $\Delta z_1^* = \varepsilon z^* d\tau_1 / (1 - \tau_0)$ and $\bar{z}_0 - \bar{z}_2 = \varepsilon \bar{z}_2 d\tau_2 / (1 - \tau_0)$.

A.2.2 Dynamic Model

A similar interpretation of our results holds when we turn to our more dynamic framework in Section 6.3. Suppose now that $(\varepsilon_i, \phi_i, a_i, \boldsymbol{\pi}_i)$ is jointly distributed according to a smooth CDF, which results in a smooth, joint distribution of elasticities, fixed costs, earnings, and probabilities of drawing a positive fixed cost. In order to gain tractability, we assume that the profile $\boldsymbol{\pi}_i$ is independent of the parameters $(\varepsilon_i, \phi_i, a_i)$. The result is that the joint density of these parameters, under a linear tax of τ_0 , can be expressed as a product of two densities: $h_0^*(z, \varepsilon, \phi) g(\boldsymbol{\pi}_i)$. We maintain the assumption that the density of earnings is constant over the interval $[z^*, z^* + \Delta z^*]$, conditional on ε and ϕ . Bunching at K_1 in period $t \in [1, \mathcal{T}_1]$ will

now be:

$$\begin{aligned}
B_1^t &= \iiint\limits_{\underline{z}_1}^{z^*+\Delta z_1^*} h_0^*(\zeta, \epsilon, \varphi) g(\boldsymbol{\pi}) d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&+ \iiint\limits_{z^*}^{\underline{z}_1} (1 - \prod_{j=1}^t \pi_j) h_0^*(\zeta, \epsilon, \varphi) g(\boldsymbol{\pi}) d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&= \iint [z^* + \Delta z_1^* - \underline{z}_1] h_0^*(z^*, \epsilon, \varphi) \left(\int g(\boldsymbol{\pi}) d\boldsymbol{\pi} \right) d\epsilon d\varphi \\
&+ \iint [\underline{z}_1 - z^*] h_0^*(z^*, \epsilon, \varphi) \left(\int (1 - \prod_{j=1}^t \pi_j) g(\boldsymbol{\pi}) d\boldsymbol{\pi} \right) d\epsilon d\varphi \\
&= h_0(z^*) \left\{ \iint [z^* + \Delta z_1^* - \underline{z}_1] \frac{h_0^*(z^*, \epsilon, \varphi)}{h_0(z^*)} d\epsilon d\varphi \right. \\
&\quad \left. + (1 - \mathbb{E} [\prod_{j=1}^t \pi_j]) \iint [\underline{z}_1 - z^*] \frac{h_0^*(z^*, \epsilon, \varphi)}{h_0(z^*)} d\epsilon d\varphi \right\} \\
&= h_0(z^*) \left\{ z^* + \mathbb{E} [\Delta z_1^*] - \mathbb{E} [\underline{z}_1] + (1 - \mathbb{E} [\prod_{j=1}^t \pi_j]) (\mathbb{E} [\underline{z}_1] - z^*) \right\} \\
&= h_0(z^*) \left\{ \mathbb{E} [\Delta z_1^*] - \mathbb{E} [\prod_{j=1}^t \pi_j] (\mathbb{E} [\underline{z}_1] - z^*) \right\}, \tag{A.23}
\end{aligned}$$

where now $h_0(z^*) = \iiint h_0^*(z^*, \epsilon, \varphi) g(\boldsymbol{\pi}) d\epsilon d\varphi d\boldsymbol{\pi}$. In the second line, we have again made use of a constant $h_0^*(\cdot)$ and also the independence of $\boldsymbol{\pi}_i$. Normalized bunching at K_1 in period t will then be:

$$b_1^t = \mathbb{E} [\Delta z_1^*] - \mathbb{E} [\prod_{j=1}^t \pi_j] (\mathbb{E} [\underline{z}_1] - z^*). \tag{A.24}$$

Using similar steps, we can show that bunching in period $t > \mathcal{T}_1$ at K_2 , when moving sequentially from K_1 , can be written as:

$$\begin{aligned}
B_2^t &= \iiint\limits_{\underline{z}_1}^{z^*+\Delta z_2^*} h_0^*(\zeta, \epsilon, \varphi) g(\boldsymbol{\pi}) d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&+ \iiint\limits_{z^*+\Delta z_2^*}^{\bar{z}_0} (\prod_{j=1}^{t-\mathcal{T}_1} \pi_j) h_0^*(\zeta, \epsilon, \varphi) g(\boldsymbol{\pi}) d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&+ \iiint\limits_{z^*}^{\underline{z}_1} (1 - \prod_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \prod_{j=1}^{\mathcal{T}_1} \pi_j) h_0^*(\zeta, \epsilon, \varphi) g(\boldsymbol{\pi}) d\zeta d\epsilon d\varphi d\boldsymbol{\pi} \\
&= h_0(z^*) \left\{ (1 - \mathbb{E} [\prod_{j=1}^{t-\mathcal{T}_1} \pi_j]) \mathbb{E} [\Delta z_2^*] + \mathbb{E} [\prod_{j=1}^{t-\mathcal{T}_1} \pi_j] \mathbb{E} [\bar{z}_0] \right. \\
&\quad \left. - \mathbb{E} [\prod_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \prod_{j=1}^{\mathcal{T}_1} \pi_j] \mathbb{E} [\underline{z}_1] - (\mathbb{E} [\prod_{j=1}^{t-\mathcal{T}_1} \pi_j] - \mathbb{E} [\prod_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \prod_{j=1}^{\mathcal{T}_1} \pi_j]) z^* \right\}. \tag{A.25}
\end{aligned}$$

Likewise, normalized bunching at K_2 will be:

$$\begin{aligned}
b_2 &= (1 - \mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j]) \mathbb{E} [\Delta z_2^*] + \mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j] \mathbb{E} [\bar{z}_0] - \mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1} \pi_j] \mathbb{E} [\underline{z}_1] \\
&\quad - (\mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j] - \mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1} \pi_j]) z^*.
\end{aligned} \tag{A.26}$$

The levels of bunching at the kink before and after the transition are now functions of average behavioral responses, $(\Delta z_1^*, \Delta z_2^*)$, the average thresholds for marginal bunchers, $(\underline{z}_1, \bar{z}_0)$, and average survival probabilities, $(\Pi_{j=1}^t \pi_j, \Pi_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1} \pi_j)$. Relative to our baseline dynamic model in Section 6.3, the number of intermediate parameters to be identified is increasing in the number of post-transition periods, due to the terms of the form $\mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1} \pi_j]$. A sufficient condition that allows us to retain identification while only using two transitions in kinks is that the expectation of this product simplifies to a product of expectations: $\mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1} \pi_j] = \mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j] \mathbb{E} [\Pi_{j=1}^{\mathcal{T}_1} \pi_j]$. There are two cases of interest that satisfy this condition. First, if $\pi_j = 0$ for some $j < \mathcal{T}_1$, then $\Pi_{j=1}^{\mathcal{T}_1} \pi_j = 0$, and the condition holds. This empirically appears to be the case in our context: adjustment takes roughly two years, while $\mathcal{T}_1 \geq 3$ in our two main applications. Second, if there is no heterogeneity in $\boldsymbol{\pi}$ across agents, the condition also holds.

If we relax the assumption that $\mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j \cdot \Pi_{j=1}^{\mathcal{T}_1} \pi_j] = \mathbb{E} [\Pi_{j=1}^{t-\mathcal{T}_1} \pi_j] \mathbb{E} [\Pi_{j=1}^{\mathcal{T}_1} \pi_j]$, we will require additional transitions in kinks in order to achieve identification. Furthermore, if we relax the assumption that the profile $\boldsymbol{\pi}_i$ is independent of $(\varepsilon_i, \phi_i, a_i)$, identification is more complicated, as the expectations in the above expressions will then feature weights that vary with t . In that case, more parametric structure on the joint distribution of $(\varepsilon_i, \phi_i, a_i, \boldsymbol{\pi}_i)$ is needed to achieve identification. We discuss identification further in section A.4 of the Appendix.

A.3 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 in subsequent periods 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 in Section 6.2. That is, we demonstrate this extension in the context of the ‘‘comparative static’’ model. 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 a . Let $z_{a,j}$ be actual earnings for individual a in period when facing tax schedule $T_j(z)$, and let $\tilde{z}_{a,j}$ 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:

$$\begin{aligned}
z_{a,j}(\tilde{z}_{a,0}) &\in [z_{a,j}^-(\tilde{z}_{a,0}), z_{a,j}^+(\tilde{z}_{a,0})], \\
&\text{where } z_{a,t}^- \leq \tilde{z}_{a,j} \leq z_{a,t}^+ \\
\text{and } u(\tilde{z}_{a,j} - T_j(\tilde{z}_{a,j}), \tilde{z}_{a,j}; a) - \phi^* &= u(z_{a,j}^- - T_j(z_{a,j}^-), z_{a,j}^-; a) \\
&= u(z_{a,j}^+ - T_j(z_{a,j}^+), z_{a,j}^+; a)
\end{aligned} \tag{A.27}$$

where $T_j(\cdot)$ 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, $z_{a,j}^-$ and $z_{a,j}^+$ 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 $z_{a,j}(\tilde{z}_{a,0})$ as a function of the optimal level of earnings for individual a 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 $F_{a,j}(z_{a,j} | \tilde{z}_{a,0})$, with probability density function $f_{a,j}(z_{a,j} | \tilde{z}_{a,0})$. 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 6 Panel B area i did not bunch in the baseline because they were sufficiently close to the kink. These are agents for whom $z^* < \tilde{z}_{a,0} < \underline{z}_1$. 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 $z_{a,0} \in [z_{a,1}^+, z_{a,0}^+]$. 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 , and iv in Figure 6, 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 $z_{a,0} < z_{a,1}^+$. These cases are illustrated in Appendix Figure B.3.

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

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

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 $\bar{z}_0 \leq \tilde{z}_{a,0} \leq z^* + \Delta z_1^*$, i.e. area *iv* in Figure 6, 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 $z^* \leq \tilde{z}_{a,0} \leq \bar{z}_0$ who actually bunched in period 1. Formally, bunching in period 2 under K_2 can be expressed as follows:

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

We can rewrite the level of bunching in this setting in terms of bunching amounts derived above:

$$\begin{aligned}B'_1 &= \int_{z^*}^{z^* + \Delta z_1^*} \Pr(z_{a,0} \geq z_{a,1}^+ | \tilde{z}_{a,0} = \zeta) h(\zeta) d\zeta \\ &= \int_{z^*}^{z^* + \Delta z_1^*} h(\zeta) d\zeta \cdot \int_{z^*}^{z^* + \Delta z_1} \Pr(z_{a,0} \geq z_{a,1}^+ | \tilde{z}_{a,0} = \zeta) \frac{h(\zeta)}{\int_{z^*}^{z^* + \Delta z_1} h(\zeta) d\zeta} d\zeta \\ &= B_1^* \cdot \int_{z^*}^{z^* + \Delta z_1^*} \Pr(z_{a,0} \geq z_{a,1}^+ | \tilde{z}_{a,0} = \zeta) h(\zeta | z^* < \zeta \leq z^* + \Delta z_1^*) d\zeta \\ &= B_1^* \cdot \mathbb{E}[\Pr(z_{a,0} \geq z_{a,1}^+) | z^* < \tilde{z}_{a,0} \leq z^* + \Delta z_1^*]\end{aligned}$$

where $B_1^* = \int_{z^*}^{z^* + \Delta z_1^*} h_0(\zeta) d\zeta$ is defined in equation (2) when $j = 1$. This is the bunching that would occur in a model of no frictions under K_1 , i.e. areas *i* – *iv* in Figure 6, Panel B. Likewise, we have:

$$\begin{aligned}\tilde{B}'_2 &= \int_{z^*}^{\bar{z}_0} \Pr(z_{a,0} \geq z_{a,1}^+ | \tilde{z}_{a,0} = \zeta) h_0(\zeta) d\zeta \\ &= [\tilde{B}_2 + B_1^* - B_1] \cdot \mathbb{E}[\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < \tilde{z}_{a,0} \leq \bar{z}_0]\end{aligned}$$

where \tilde{B}_2 is defined in equation (6), and B_1 is defined in equation (4). It follows that $\tilde{B}_2 + B_1^* - B_1 = \int_{z^*}^{\bar{z}_0} h_0(\zeta) d\zeta$, i.e. areas *i* – *iv* in Figure 6.

Without further restrictions on the distribution of optimal earnings under a linear tax, $H_0(z)$, or distribution of earnings about the frictionless optimum in period 0, $F_{a,j}(z_{a,j} | \tilde{z}_{a,0})$, we cannot make further simplifications of these expressions. However, if we assume that the initial actual earnings level is distributed uniformly about optimal earnings in period 0, following Chetty *et al.* (2011) or Kleven and Waseem (2013), then we have:

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

which implies that:

$$\Pr(z_{a,0} \geq z_{a,1}^+ | \tilde{z}_{a,0} = \zeta) = \min\left(\frac{z_{a,0}^+(\zeta) - z_{a,1}^+(\zeta)}{z_{a,0}^+(\zeta) - z_{a,0}^-(\zeta)}, 1\right)$$

Using our definitions above for $z_{a,0}^+(\cdot)$, $z_{a,0}^-(\cdot)$ and $z_{a,1}^+(\cdot)$ 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. $[z_{a,0}^-, z_{a,0}^+]$, will tend to be different distances from the frictionless optimum. We can also calculate B_1^* , B_1 , and \tilde{B}_2 , conditional on the counterfactual distribution $H_0(z)$ and a value of ε and ϕ . We are therefore able to calculate predicted values for B_1' and \tilde{B}_2' and use these in a modified version of the estimation procedure outlined in Section 6.5.

Although it is not necessary for our estimation procedure, if we further 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:

$$\begin{aligned} B_1' &= B_1^* \cdot \mathbb{E} [\Pr(z_{a,0} \geq z_{a,1}^+) | z^* < \tilde{z}_{a,0} \leq z^* + \Delta z_1^*] \\ &= \Delta z_1^* h_0(z^*) \cdot \mathbb{E} [\Pr(z_{a,0} \geq z_{a,1}^+) | z^* < \tilde{z}_{a,0} \leq z^* + \Delta z_1^*] \end{aligned}$$

and likewise:

$$\begin{aligned} B_2' &= [\tilde{B}_2 + B_1^* - B_1] \cdot \mathbb{E} [\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < \tilde{z}_{a,0} \leq \bar{z}_0] \\ &= [\bar{z}_0 - z^*] h_0(z^*) \cdot \mathbb{E} [\Pr(z_{n,0} \geq z_{n,1}^+) | z^* < \tilde{z}_{a,0} \leq \bar{z}_0] \end{aligned}$$

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

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

A.4 Identification

Our estimator is a minimum distance estimator (MDE); Newey and McFadden (1994) give conditions for identification, consistency, and asymptotic normality. An MDE is defined as:

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta} \hat{Q}(\theta) \\ \hat{Q}(\theta) &= [B - m(\theta)]' \hat{W} [B - m(\theta)] \end{aligned}$$

In our case, B is a vector of L estimated bunching amounts from before and after a policy change, and $m(\theta)$ is a vector of predicted bunching amounts. \hat{W} is a weighting matrix. We consider our comparative static, and dynamic, models, in turn.

A.4.1 Comparative Static Model

We focus on the exactly identified case with two bunching moments, which is relevant in our empirical application of the comparative static model. We have:

$$\begin{aligned} m(\theta) &= (B_1(\varepsilon, \phi), \tilde{B}_2(\varepsilon, \phi)) \\ B_1 &= \int_{\underline{z}_1}^{z^* + \Delta z_1^*} h(\xi) d\xi \\ \tilde{B}_2 &= \int_{\underline{z}_1}^{\bar{z}_0} h(\xi) d\xi \end{aligned}$$

where B_1 and \tilde{B}_2 refer to bunching before and after the policy change, and $\theta \equiv (\varepsilon, \phi)$.

The upper cutoff in B_1 is defined as

$$z^* + \Delta z_1^* = z^* \left(\frac{1 - \tau_0}{1 - \tau_1} \right)^\varepsilon.$$

A necessary condition for identification is that solutions for \underline{z}_1 and \bar{z}_0 exist; if they do not, then no bunching occurs. It is straightforward to show that a solution for \underline{z}_1 exists if

$$z^* \left[(1 - \tau_1) - \left(\frac{1 - \tau_0}{1 - \tau_1} \right)^\varepsilon ((1 - \tau_1) - \varepsilon (\tau_1 - \tau_0)) \right] > \phi(\varepsilon + 1).$$

This ensures that the “top” buncher wants to adjust to the kink. A solution for \bar{z}_0 exists as long as some debunching occurs. It is straightforward to show that this requires that:

$$z^* \left[\frac{(1 - \tau_2)^{\varepsilon+1} - (1 - \tau_1)^{\varepsilon+1}}{(1 - \tau_1)^\varepsilon} \right] > \phi(\varepsilon + 1).$$

As long as $\tau_0 < \tau_2 < \tau_1$, $\varepsilon > 0$, and $\phi > 0$, there exists a range of values of ε and ϕ for which these inequalities hold.

Provided that \bar{z}_0 and \underline{z}_1 exist, identification requires that $m(\theta) = B$ has a unique solution. Following previous literature (*e.g.* Kline and Walters 2016), we establish local uniqueness by linearizing $m(\cdot)$ around a solution $m(\theta_0) = B$. Let θ_0 be a solution to $m(\theta) = B$. Linearizing $m(\cdot)$ around θ_0 , we have:

$$m(\theta) \approx m(\theta_0) + \nabla m(\theta_0)(\theta - \theta_0).$$

It follows that a unique solution requires $\mathbf{J}_m(\theta_0)$ to have full rank, where $\mathbf{J}_m(\theta_0)$ is the Jacobian of $m(\cdot)$ evaluated at θ_0 :

$$\mathbf{J}_m(\theta_0) = \begin{bmatrix} \frac{\partial B_1}{\partial \varepsilon} & \frac{\partial B_1}{\partial \phi} \\ \frac{\partial \tilde{B}_2}{\partial \varepsilon} & \frac{\partial \tilde{B}_2}{\partial \phi} \end{bmatrix}.$$

We calculate the elements of this matrix analytically by differentiating the expressions above

for B_1 and \tilde{B}_2 , which is straightforward.²¹ Thus, given $\hat{\theta}$, \underline{z}_1 , and \bar{z}_0 , we can calculate the Jacobian analytically (although \underline{z}_1 and \bar{z}_0 must be found numerically).

\mathbf{J}_m has full rank only if it has a non-zero determinant. We find in all of our bootstrap iterations that $\det(\mathbf{J}_m) < 0$, demonstrating that the determinant is significantly different from zero. We have also shown analytically that the determinant is generically non-zero (results available upon request).

A.4.2 Dynamic Model

To identify the dynamic model, we need to observe at least as many moments as the number of parameters we seek to estimate. In our case this means that we must observe bunching across multiple policy changes, specifically the reductions in the benefit reduction rate above the exempt amount in 1990 and at age 70. Let l index different such policy changes (in our case, $l \in \{1990, 70\}$). Let $B_{1,l}^t$ be bunching at kink l and period t *before* the policy change, let $B_{2,l}^t$ be bunching at kink l and period t *after* the policy change, let time t measure the time since the introduction of the first kink, $K_{1,l}$, and let the policy change at kink l take place at time $\mathcal{T}_{1,l}$. The parameter vector θ now consists of $(\varepsilon, \phi, \pi_1, \pi_2, \dots, \pi_5)$. We match 12 bunching amounts in our estimates: 1987 to 1992 (pooling 66 to 68 year olds) and ages 67 to 72 (pooling years 1990 to 1999).

Bunching before the policy change is

$$B_{1,l}^t = \prod_{j=1}^t \pi_j \cdot B_{1,l} + (1 - \prod_{j=1}^t \pi_j) B_{1,l}^*$$

where $B_{1,l} = \int_{\underline{z}_{1,l}}^{z_l^* + \Delta z_{1,l}^*} h(\xi) d\xi$ and $B_{1,l}^* = \int_{z_l^*}^{z_l^* + \Delta z_{1,l}^*} h(\xi) d\xi$, and the limits of integration are defined similarly to the static case (but with the additional subscript l to allow for analysis across multiple policy changes, as in our empirical application of the dynamic model). If the policy change happens $\mathcal{T}_{1,l}$ periods after the kink is initially introduced, then bunching under the new policy in period t is

$$B_{2,l}^t = \prod_{j=1}^{t-\mathcal{T}_{1,l}} \pi_j \cdot \tilde{B}_{2,l} + \left(1 - \prod_{j=1}^{t-\mathcal{T}_{1,l}} \pi_j\right) B_{2,l}^* + \prod_{j=1}^{t-\mathcal{T}_{1,l}} \pi_j \left(1 - \prod_{j=1}^{\mathcal{T}_{1,l}} \pi_j\right) (B_{1,l}^* - B_{1,l})$$

where $\tilde{B}_{2,l} = \int_{\underline{z}_{1,l}}^{\bar{z}_{0,l}} h(\xi) d\xi$, $B_{2,l}^* = \int_{z_l^*}^{z_l^* + \Delta z_{2,l}^*} h(\xi) d\xi$, and the limits of integration again are defined similarly to the static case but with the additional subscript l .

We calculate the elements of the resulting Jacobian analytically by differentiating the expressions above for $B_{1,l}^t$ and $B_{2,l}^t$ with respect to $\varepsilon, \phi, \pi_1, \pi_2, \pi_3, \pi_4$, and π_5 , which is again straightforward. Thus, given $\hat{\theta}$, $\underline{z}_{1,l}$ and $\bar{z}_{0,l}$, we can again calculate the Jacobian analytically.

Identification requires that this Jacobian have full rank. To test for full rank of the Jacobian, we use the method of Kleibergen and Papp (2006). We use the bootstrap to obtain an estimate of $Var[\mathbf{J}_m(\hat{\theta})]$. In each iteration of our bootstrap, we also calculate $\mathbf{J}_m(\hat{\theta})$, and we estimate $Var[\mathbf{J}_m(\hat{\theta})]$ from the bootstrap variance-covariance matrix. The RK test easily rejects under-identification, with $p < 0.001$.

²¹We can specify functions implicitly defining the lower and upper cutoffs \underline{z}_1 and \bar{z}_0 , respectively, as functions of the other parameters, given our quasilinear and isoelastic case. These enter the expressions for each element of the Jacobian (more details are available upon request).

A.5 Econometric Estimation

We begin by describing our econometric estimation procedure under our basic comparative static model of Sections 6.1 and 6.2. Let $B = (B_1, B_2, \dots, B_L)$ be a vector of (estimated) bunching amounts, using the method described in Section 3. Let $\tau = (\tau_1, \dots, \tau_L)$ be the tax schedule at each kink. The triplet $\tau_l = (\tau_{0,l}, \tau_{1,l}, \tau_{2,l})$ denotes the tax rate below the kink ($\tau_{0,l}$), above the kink ($\tau_{1,l}$), and the *ex post* marginal tax rate above the kink after it has been reduced ($\tau_{2,l}$), as in Section 6.2. Let $\mathbf{z}^* = (z_1^*, \dots, z_L^*)$ be the earnings levels associated with each kink. In principle, it would be possible to estimate bunching separately for each age group at a given kink. In practice and for simplicity, we pool across a constant set of ages to estimate bunching at a given kink—for example, when examining the 1990 policy change we examine 66-68 year-olds both before and after the change. Thus, the bunching amounts are not indexed by age.²²

In our baseline, we use a non-parametric density for the counterfactual earnings distribution, H_0 . Once H_0 is known, we use (4) and (6) to obtain predicted bunching from the model. To recover H_0 non-parametrically we take the empirical earnings distribution for 72 year-olds in \$800 bins as the counterfactual distribution. 72 year-olds' earnings density represents a reasonable counterfactual because they no longer face the Earnings Test, no longer show bunching, and are close in age to those aged 70 or 71. 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 midpoints of the bins, so we use linear interpolation for other values of z . In a robustness check, we instead assume that the earnings distribution over the range $[z^*, z^* + \Delta z]$ is uniform, a common assumption in the literature (*e.g.* Chetty *et al.*, 2011, Kleven and Waseem, 2013). Using the nonparametrically-estimated distribution of earnings from age 72 is helpful because it does not entail distributional assumptions, but relative to assuming a uniform distribution, using the age-72 distribution comes at the cost of using a different age (*i.e.* 72) to generate the earnings distribution.²³

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 $\hat{B}(\varepsilon, \phi) \equiv (\hat{B}(\tau_1, z_1^*, \varepsilon, \phi), \dots, \hat{B}(\tau_L, z_L^*, \varepsilon, \phi))$, our estimator is:

$$\left(\hat{\varepsilon}, \hat{\phi}\right) = \operatorname{argmin}_{(\varepsilon, \phi)} \left(\hat{B}(\varepsilon, \phi) - B\right)' W \left(\hat{B}(\varepsilon, \phi) - B\right), \quad (\text{A.28})$$

where W is a $K \times K$ identity matrix. This estimation procedure runs parallel to our theoretical model, as the bunching amounts \hat{B} are those predicted by the theory (and the estimated counterparts B are found using the procedure outlined in Section 3).²⁴ When we pool data

²²Analogously, when we examine bunching at each age around 70 when the AET is eliminated, we pool across calendar years (namely 1990-1999) to estimate bunching, so that we do not also have to index the bunching amounts by calendar year. We find comparable results when we estimate bunching separately at each age and year.

²³Because we use the age-72 density as our counterfactual density – unlike most bunching papers bunching that estimate the counterfactual from the same density that is used to estimate bunching – our method is not subject to the Blomquist and Newey (2017) point that the functional form of preference heterogeneity cannot be simultaneously estimated with the taxable income elasticity.

²⁴Without loss of generality, we use normalized bunching, $\hat{b} = \delta \hat{B} / h_0(z^*)$, so that the moments are identical to what is reported elsewhere in the text.

across multiple time periods, we assume that ε and ϕ are constant across these time periods.

We obtain our estimates by minimizing (A.28) numerically. Solving this problem requires evaluating \hat{B} at each trial guess of (ε, ϕ) .²⁵ Our estimator assumes a quasilinear utility function, $u(c, z; a) = c - \frac{a}{1+1/\varepsilon} \left(\frac{z}{a}\right)^{1+1/\varepsilon}$, following Saez (2010), Chetty *et al.* (2011) and Kleven and Waseem (2013). Note that because we have assumed quasilinearity, $\Delta z_{1,l} = z_l^* \left(\left(\frac{1-\tau_{1,l}}{1-\tau_{0,l}} \right)^\varepsilon - 1 \right)$ and $a = z(\tau) / (1-\tau)^\varepsilon$, where $z(\tau)$ are the optimal, interior earnings under a linear tax of τ . Typically there is no closed form solution for $\underline{z}_{1,l}$ or $\bar{z}_{0,l}$. Instead, given ε and ϕ , we find $\underline{z}_{1,l}$ and $\bar{z}_{0,l}$ numerically as the solution to the relevant indifference conditions in (5) and (7). For example, $\underline{z}_{1,l}$ is defined implicitly by:

$$\underbrace{u((1-\tau_{1,l})z_l^* + R_{1,l}, z_l^*; \underline{z}_{1,l}/(1-\tau_{0,l})^\varepsilon)}_{\text{utility from adjusting to kink}} - \underbrace{u((1-\tau_{1,l})\underline{z}_{1,l} + R_{1,l}, \underline{z}_{1,l}; \underline{z}_{1,l}/(1-\tau_{0,l})^\varepsilon)}_{\text{utility from not adjusting}} = \phi, \quad (\text{A.29})$$

This equation is continuously differentiable and has a unique solution for $\underline{z}_{1,l}$.²⁶

A.5.1 Dynamic Model

Our estimation method is easily amended to accommodate the dynamic extension of our model in Section 6.3. As in (8) and (9), the bunching expressions in the dynamic model are weighted sums of B_1 and \tilde{B}_2 , which are calculated as in Section 6.5, and two measures of frictionless bunching, B_1^* and B_2^* . Frictionless bunching under either kink can be calculated conditional on H_0 and ε using (2).

We must also estimate the probability of drawing a positive fixed cost as a function of the time since the last policy shock, π_{t-t^*} .²⁷ For given values of ε , ϕ , and the vector $\boldsymbol{\pi}$ of π_{t-t^*} 's, we can evaluate (8) and (9). Our vector of predicted bunching, \hat{B} , will now be a function of these additional parameters, as well as the relevant time indices: $\hat{B}(\varepsilon, \phi, \boldsymbol{\pi}) \equiv (\hat{B}(\boldsymbol{\tau}_1, z_1^*, t_1, \mathcal{T}_{1,l}, \varepsilon, \phi, \boldsymbol{\pi}), \dots, \hat{B}(\boldsymbol{\tau}_L, z_L^*, t_L, \mathcal{T}_{1,L}, \varepsilon, \phi, \boldsymbol{\pi}))$, where t_l is the time elapsed since the first kink, $K_{1,l}$, was introduced, and $\mathcal{T}_{1,l}$ is the length of time before the second kink, $K_{2,l}$, is introduced. Once again we use the minimum distance estimator (A.28).

Equations (8) and (9) illustrate how we estimate the elasticity and adjustment cost in this richer setting. We require as many observations of bunching as the parameters, $(\varepsilon, \phi, \pi_1, \dots, \pi_J)$, and these moments must span a change in $d\tau$.²⁸ Suppose we observe the pattern of bunching over time around two or more different policy changes. Loosely speaking, the $\boldsymbol{\pi}$'s are estimated relative to one another from the time pattern of bunching over time: a delay in adjustment in a given period will generally correspond to a higher probability of facing the adjustment cost (all else equal). Note that the relationship is linear; the degree of

²⁵In solving (A.28), 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. Inattention or the difficulty of negotiating new contracts should be associated with positive adjustment costs (which could distinguish this context from the firm context studied in Garicano *et al.*, 2016).

²⁶Note that some combinations of $\boldsymbol{\tau}_l$, z_l^* , ε , and ϕ imply $\underline{z}_{1,l} > z_l^* + \Delta z_{1,l}$. In this case, the lowest-earning adjuster does not adjust to the kink. Whenever this happens, we set $\hat{B}_l = 0$.

²⁷We have also tried using a flexible, logistic functional form, $\pi_j = \exp(\alpha + \beta \cdot j) / (1 + \exp(\alpha + \beta \cdot j))$, and we found comparable results (available upon request).

²⁸The number of moments is not itself sufficient. We also require non-trivial variation in bunching before and after the tax change in order to point identify ϕ . As in footnote 8, this requires $\bar{z}_0 < z^* + \Delta z_1^*$.

“inertia” in bunching in (for example) period 1 increases linearly in π_1 . Meanwhile, a higher ϕ implies a larger amount of inertia in *all* periods until bunching has fully dissipated (in a way that depends on the earnings distribution, the elasticity, and the size of the tax change). Finally, a higher ε will correspond to a larger amount of bunching once bunching has had time to adjust fully to the policy changes. Intuitively, these features of the data help us to identify the parameters using our dynamic model.

A.6 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 not 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 Earnings Test. 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 Earnings Test 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 Earnings Test 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 Social Security claimants.²⁹ In period 2, the tax rate below the exempt amount remains 27.21 percent, but the tax rate above the exempt amount changes according

²⁹As 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.

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 0.35 and the adjustment cost ϕ is \$280 (our baseline estimates); or (b) the elasticity ε is 0.35 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.

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 B.6 below.

We calculate that if the marginal tax rate above the exempt amount were reduced by 17.22 percentage points, so that the tax rate above the exempt amount were reduced from 77.21 percent to 59.99 percent, mean earnings in the population under consideration would be unchanged at \$9,371.9 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 17.22 percentage points; 17.22 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 of \$280 prevents adjustment with an 17.22 percentage point (or smaller) cut in marginal tax rates. By contrast, when assuming $\varepsilon=0.35$ and $\phi=0$, we predict that mean earnings would rise from \$9,340.3 to \$10,166.3, an increase of 8.84 percent.

At the same time we calculate that if the 50 percent Earnings Test 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 \$9,371.9 to \$11,566.7, or 23.4 percent, under our baseline estimates of the elasticity and adjustment cost. When assuming $\varepsilon=0.35$ and $\phi=0$, we predict that mean earnings would rise from \$9,340.3 to \$11,639.2, a nearly identical increase of 24.6 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.

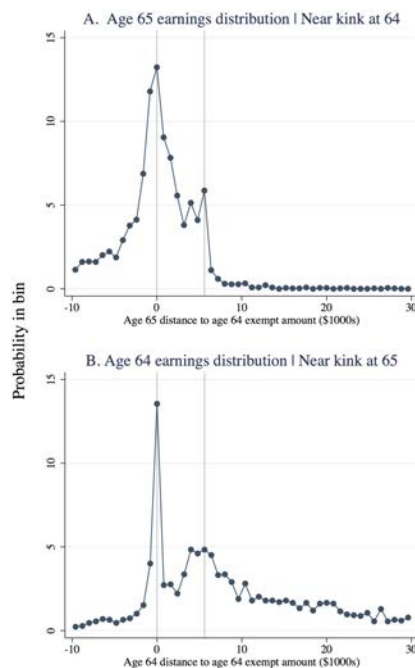
It is worth noting an additional caveat to these results: they apply to those with counterfactual earnings in the range from \$10,000 below to \$10,000 above the exempt amount. If we allowed unbounded counterfactual earnings, there would be some individuals with very large counterfactual earnings for whom the gains from adjustment would outweigh the adjustment

cost, even in the presence of adjustment costs. However, this is less relevant to the Earnings Test because as we have noted, the Social Security 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 qualitatively robust to considering other earnings ranges within the range we measure in our study, such as the range of individuals earning from \$10,000 below to \$30,000 above the exempt amount. 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 (0.58) when performing the simulations (which actually leads to still starker results).

All of these simulations use the static model. If we were to use our estimates of the dynamic model instead to perform these simulations, we would still find that the immediate reaction even to large taxes changes is greatly attenuated, since the estimates of the dynamic model still show that most individuals are constrained from adjusting immediately.

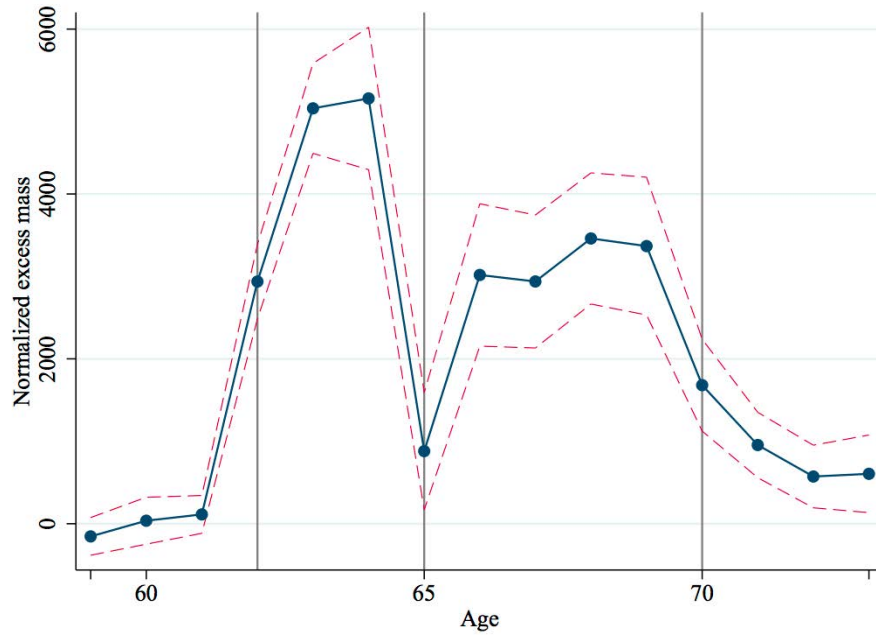
B Appendix: Additional Empirical Results

Figure B.1: Inertia in Bunching from 64 to 65



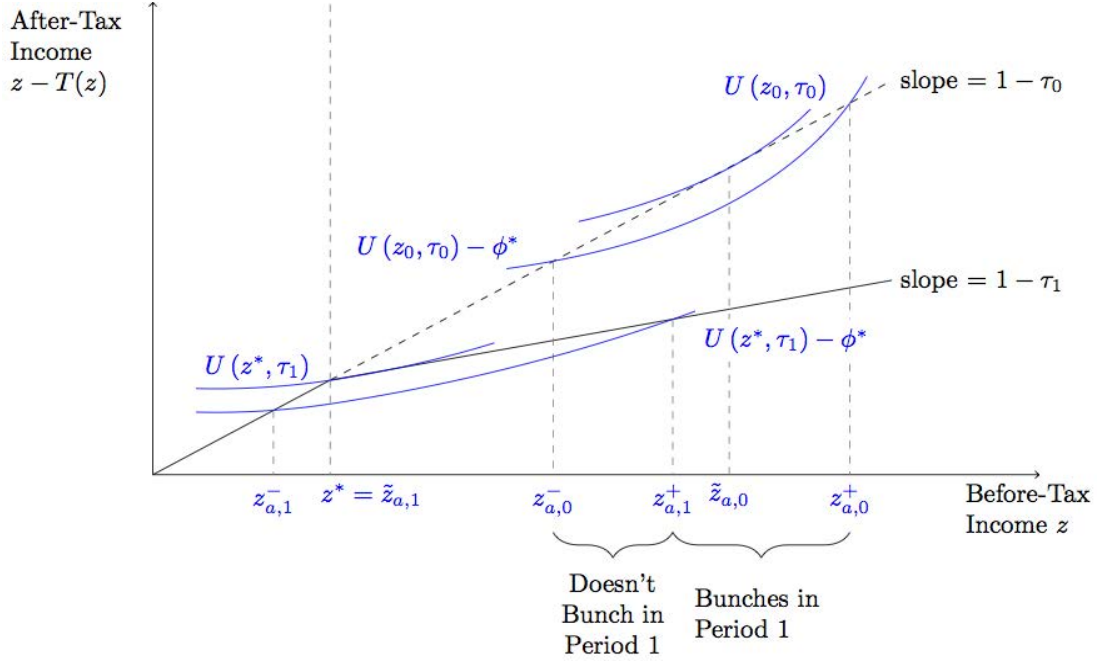
Notes: Using data from 1990 to 1999, Panel A of the figure shows that when they are age 65, those previously bunching at age 64 tend to either (a) remain near the age 64 exempt amount or (b) move to the age 65 exempt amount. Panel B of the figure shows that those bunching at age 65 were usually bunching at age 64 in the previous year, or were near the age 65 exempt amount in the previous year. Having earnings “near the kink” at a given age is defined as having earnings within \$1,000 of the kink at that age. The first vertical line at zero shows the age 64 exempt amount, and the second vertical line shows the average location of the age 65 exempt amount.

Figure B.2: Normalized Excess Mass of Claimants, Ages 59 to 73, 1990 to 1999



Note: See notes to Figure 2 Panel B. This figure differs from Figure 2 Panel B because here the sample in year t consists only of people who have claimed Social Security in year t or before (whereas in Figure 2 Panel B it consists of those who claimed by age 65).

Figure B.3: Bunching Response to a Convex Kink, with Frictions in Initial Earnings



Note: See Section A.3 for an explanation of the figure.

Table B.1: Robustness of normalized bunching to alternative birth month restrictions

	b_{68}	b_{69}	b_{70}	b_{71}	b_{72}
A) Born January-March	3545.4 [2681.1, 4409.7]***	4036.2 [2934.2, 5138.2]***	565 [42, 1088.1]**	881.7 [-14.3, 1777.6]*	-236.8 [-890.4, 416.8]
B) Born any month	3992.2 [3386.8, 4597.7]***	3552.3 [3092.4, 4012.2]***	1203.9 [929, 1478.9]***	941.4 [453, 1429.8]***	-231.4 [-510.7, 47.8]

Notes: The table shows excess normalized bunching and its confidence interval at each age from 68 to 72 for two samples: those born January to March (Row A), and those born in any month (Row B). The data are pooled over the period from 1983-1999. The table shows that we continue to estimate significant bunching at age 70 (and in some cases 71) when the sample is restricted to those born in January to March. Limiting the sample only to those born in January yields insignificant results, with little statistical power. *** indicates $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table B.2: Robustness to alternative empirical choices

Binsize	Degree	Excluded Bins	b_{68}	b_{69}	b_{70}	b_{71}	b_{72}
Panel A: Baseline							
\$800	7	4	3442.3 [2763.5, 4121.2]***	2868.4 [2763.5, 4121.2]***	657.9 [195.5, 1120.4]***	1068.8 [527.2, 1610.4]***	70.1 [-328.5, 468.7]
\$400	7	8	3107.8 [2653.6, 3561.9]***	2606.3 [2090.4, 3122.2]***	462.2 [111.1, 813.2]***	923.0 [541.5, 1304.4]***	-55.3 [-391.4, 280.7]
\$1,600	7	2	3047.0 [2362.4, 3731.7]***	2941.0 [2458.5, 3423.5]***	601.4 [48.2, 1154.5]**	1210.9 [581.5, 1840.3]***	241.2 [-363.6, 846.1]
Panel B: Robustness to bin size							
Panel C: Robustness to degree							
\$800	6	4	3677.2 [3117.4, 4237.0]***	3267.3 [2810.5, 3724.0]***	1310.6 [853.3, 1767.9]***	993.7 [527.8, 1459.6]***	224.6 [-162.0, 611.2]
\$800	8	4	3535.1 [2944.7, 4125.6]***	2948.0 [2529.1, 3366.9]***	710.8 [296.3, 1125.2]***	1084.3 [554.5, 1614.1]***	82.4 [-393.4, 558.2]
Panel D: Robustness to excluded region							
\$800	7	3	2170.2 [1605.1, 2735.4]***	2182.4 [1697.1, 2667.7]***	202.2 [-126.9, 531.4]	191.2 [-243.9, 626.2]	-55.0 [-390.8, 280.8]
\$800	7	5	3610.6 [2672.5, 4548.6]***	2651.3 [1972.2, 3330.4]***	298.1 [-304.8, 901.0]	1103.3 [229.6, 1977.1]**	579.9 [-161.8, 1321.6]

Notes: The table shows the estimated bunching amount at each age from 68 to 72, varying the bin size, degree of the polynomial of the smooth density, or number of excluded bins around the exempt amount. Note that varying the bin size but fixing the number of excluded bins automatically changes the width of the excluded region, so to (approximately) fix the width of the excluded region when changing the bin size, we also change the number of excluded bins. *** indicates $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table B.3: Heterogeneity in Estimates of Elasticity and Adjustment Cost across Samples

	(1)	(2)	(3)	(4)
	ε	p -value for ε equality	ϕ	p -value for ϕ equality
Men	0.44 [0.38, 0.52]***	0.39	\$62 [14, 167]***	0.00
Women	0.42 [0.32, 0.50]***		\$489 [165, 720]***	
High lifetime earnings	0.48 [0.41, 0.58]***	0.05	\$24 [2, 90]***	0.00
Low lifetime earnings	0.44 [0.32, 0.51]***		\$538 [217, 688]***	
High lifetime earnings variability	0.39 [0.35, 0.46]***	0.25	\$116 [37, 315]***	0.16
Low lifetime earnings variability	0.38 [0.33, 0.46]***		\$178 [55, 378]***	

Notes: This table implements our “comparative static” method separately in each of several groups shown in each row. “High/low lifetime earnings” refers to the group of individuals with mean real earnings from 1951 (when the data begin) to 1989 that are above/below the median level in our study population. “High/low lifetime earnings variability” refers to the group of individuals for whom the standard deviation of real earnings from 1951 to 1989 is above/below the median level in our study population. Columns 2 and 4 show the p -values for the two-sided test of equality in the estimates between each set of groups (i.e. men vs. women, high vs. low lifetime earnings, and high vs. low earnings variability), for ε and ϕ , respectively. We pool data from two policy changes: (a) around the 1989/1990 transition analyzed in Table 2, and (b) around the age 69/70 transition analyzed in Table 3. We pool the transitions because this gives us the maximum power to detect differences across groups. The results are generally comparable when we investigate each transition separately. See also notes from Tables 2 and 3.

Table B.4: 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	0.28 [0.25, 0.43]***	\$187.78 [58.97, 1303.68]***	0.43 [0.36, 0.54]***	0.24 [0.20, 0.28]***
Uniform Density	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 = \$400	0.26 [0.22, 0.33]***	\$104.10 [7.01, 440.80]***	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 may not be at the frictionless optimum, as described in the text. See also notes to Table 2.

Table B.5: Estimates of Changes in Bunching Around 1990

Sample	Old only	Old only, linear trend	DD	DD, separate linear trend
old x 1990 dummy	28.9 (249.1)	-165.1 (411.0)	-107.3 (306.7)	-69.2 (411.7)
old x 1991 dummy	-1728.9 (249.1)***	-1966.0 (500.6)***	-1824.5 (306.7)***	-1777.9 (481.3)***
old x 1992 dummy	-1648.8 (249.1)***	-1928.9 (594.9)***	-1130.2 (306.7)***	-1075.1 (558.1)*
old x 1993 dummy	-2123.8 (249.1)***	-2447.1 (692.1)***	-2131.2 (306.7)***	-2067.6 (639.7)***
Ages	66-68	66-68	62-64, 66-68	62-64, 66-68
Year FE?	No	No	Yes	Yes
Linear time trend (in year)	No	Yes	No	No
Separate linear trend for “old”	No	No	No	Yes

Notes: The table shows that the estimated change in bunching amounts from before to after 1990 in the age 66-68 age group are similar under several specifications. The dummy variable “old” indicates the older age group (66-68). The sample in Columns (1) and (2) includes only 66-68 year-olds, and in Columns (3) and (4) it also includes 62-64 year-olds. Additional controls include a linear time trend (in year) in column (2), year fixed effects in columns (3) and (4), and the linear time trend interacted with the “old” dummy in column (4). Robust standard errors are in parentheses. Under all the specifications, the coefficient on old x 1990 is insignificantly different from zero: bunching in 1990 is not significantly different from prior bunching, indicating that adjustment does not immediately occur. However, the coefficients on old x 1991, old x 1992, old x 1993 are negative and significant, indicating that bunching falls significantly after 1990—i.e. a reduction in bunching does eventually occur (but not immediately in 1990). The fact that the results are similar under all these various specifications indicates that the results are little changed by controlling for a linear trend (Column 2), comparing 66-68 year-olds to a reasonable control group of 62-64 year-olds (Column 3), and additionally controlling for a separate linear trend for the older group (Column 4). In Columns 1 and 3, the standard errors are the same across all of the interaction coefficients shown because there is only one observation underlying each dummy, and the dummies are exactly identified. See also notes from Table 2.

Table B.6: Policy Simulations

	(1)	(2)
Panel A: Eliminate Earnings Test for 62-64 year olds		
	With adjustment costs	Without adjustment costs
Period 1 mean earnings	\$9,371.9	\$9,340.1
Mean earnings change	\$2,194.8	\$2,298.9
Share affected	50.4	50.4
Share who adjust	41.9	50.4
Mean change among adjusters	\$5,239.6	\$4,563.7
Percent change among adjusters	42.6	37.3
Panel B: Reduce Earnings Test BRR by 17.22 percentage points		
	With adjustment costs	Without adjustment costs
Period 1 mean earnings	\$9,371.9	\$9,340.3
Mean earnings change	\$0	\$826.0
Percent earnings change	50.4	50.4
Share who adjust	0.0	37.4
Mean change among adjusters	0.0	\$2,207.6
Percent change among adjusters	0.0	17.7

Note: Each panel shows the results of a different policy simulation. Column 1 shows the results when we assume $\varepsilon = 0.35$ and $\phi = \$280$, and Column 2 shows the results when we assume $\varepsilon = 0.35$ 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. “BRR” is the benefit reduction rate. See Appendix A.6 for further explanation.