American Economic Journal: Applied Economics 2020, 12(1): 1–31
https://doi.org/10.1257/app.20170717

Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test†

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

We introduce a method for estimating the cost of adjusting earnings, as well as the earnings elasticity with respect to the net-of-tax share. 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. (*JEL* H24, H31, H55, J22, J31)

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

* Gelber: UC San Diego, 9500 Gilman Drive, La Jolla, CA 92093, and NBER (email: amgelber@ucsd.edu); Jones: Harris School of Public Policy, University of Chicago, 1307 E 60th Street, Chicago, IL 60637, and NBER (email: damonjones@uchicago.edu); Sacks: Kelley School of Business, Indiana University Bloomington, 1309 E 10th Street, Bloomington, IN 47405 (email: dansacks@indiana.edu). Alexandre Mas was coeditor for this article. We thank Raj Chetty, Jim Cole, Mark Duggan, John Friedman, Bill Gale, Hilary Hoynes, Henrik Kleven, Camille Landais, Olivia Mitchell, Daniel Reck, Emmanuel Saez, Joel Slemrod, 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 NIH grant 1R03 AG043039-01, from a National Science Foundation Graduate Research Fellowship under grant number DGE-0822219, and from support from the US 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.

† Go to https://doi.org/10.1257/app.20170717 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

(Chetty et al. 2011; Chetty 2012; Chetty, Friedman, and Saez 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, Slemrod, and Giertz 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 1999). Such attenuated and slow responses matter to policymakers, 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 US 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

[1] For consistency with the previous literature on kink points that has focused on the effect of taxation, we sometimes use "tax" as shorthand for "tax-and-transfer," 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.

Social Security claimants under age 66 in 2019, current Social Security benefits are reduced by $1 for every $2 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 policymakers 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 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 1 percent sample of the US 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, Slemrod, and Giertz 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, Friedman, and Saez 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, Looney, and Kroft 2009) and is a structural parameter that can be used to perform counterfactual 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; Mortenson, Schramm, and Whitten 2016; and Zaresani 2018) and has been adapted to the case of notches as well to estimate adjustment costs (Gudgeon and Trenkle 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, Harris, and Marschak 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 2009). In our paper, changes in nonlinear 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 I describes the policy environment. Section II presents the method for quantifying bunching. Section III describes our data. Section IV documents on adjustment frictions empirically. Section V specifies our model. Section VI presents our parameter estimates. Section VII describes simulations based on the estimates. Section VIII concludes. The online Appendix contains additional results. More results are available in an earlier working paper version of the present paper (Gelber, Jones, and Sacks 2013).

I. 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, 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 1 percent Delayed Retirement Credit was introduced in 1972, meaning that each year of foregone benefits led to a 1 percent increase in future yearly benefits. The Delayed Retirement Credit was raised to 3 percent in 1982 and gradually rose to 8 percent for cohorts reaching Normal Retirement Age from 1990 to 2008. An increase in future benefits between 7 and 8 percent is approximately actuarially fair

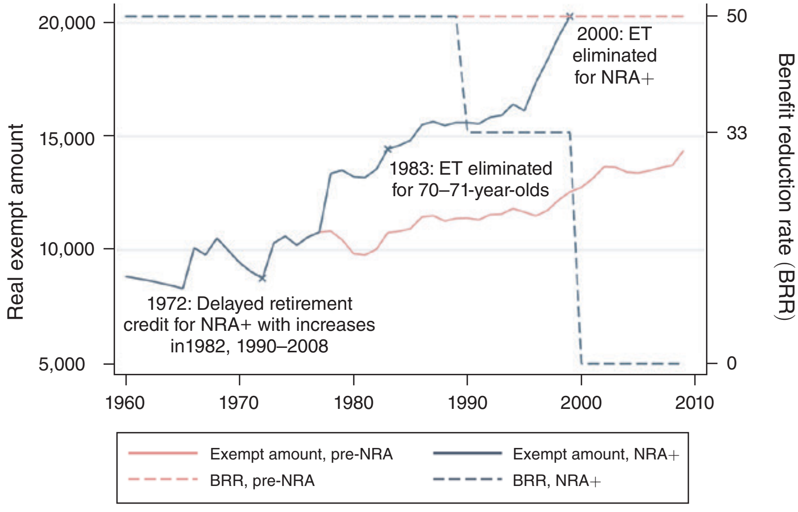

Figure 1. Key Earnings Test Rules, 1961–2009

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.

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 2013a). 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 et al. 2013). We follow previous work and do not distinguish among these potential reasons in our main analysis (Gelber, Jones, and Sacks 2013 analyzes 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.

II. 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 pretax 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 as a function of a. Thus, a is isomorphic to the level of earnings that would occur in the absence of any tax. The parameter a is distributed according to 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:

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

[2] We 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 $(z, T(z))$.

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.

III. Data

We apply this bunching framework on a 1 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 US 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

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

	Ages 62–69
Mean earnings	28,892.63
	(78,842.99)
Tenth percentile	1,193.64
Twenty-fifth percentile	5,887.75
Fiftieth percentile	14,555.56
Seventy-fifth percentile	35,073.00
Ninetieth percentile	64,647.40
Fraction male	0.57
Observations	376,431

Notes: The data are taken from a 1 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.

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.

IV. 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 later in this paper). When we pool data from 1983

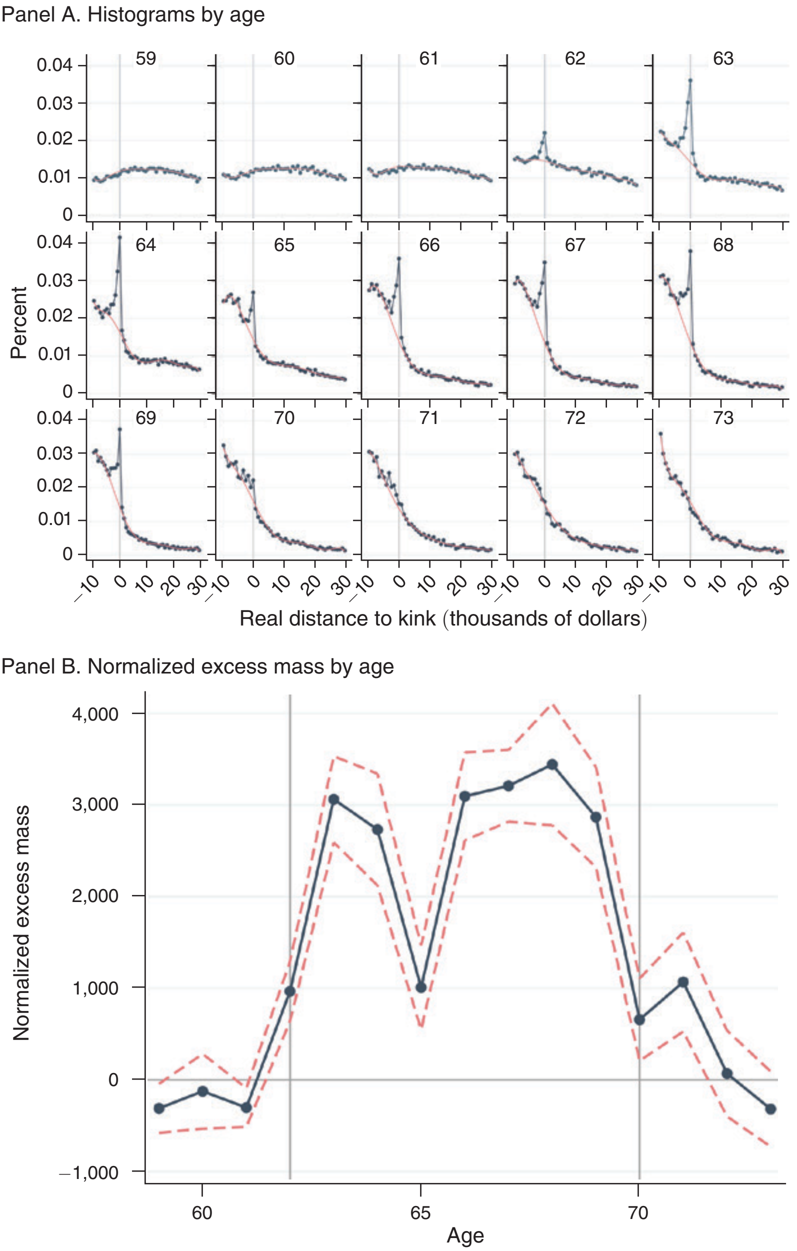

FIGURE 2. EARNINGS HISTOGRAMS AND NORMALIZED EXCESS MASS BY AGE

Notes: The sample is a 1 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 II. Dashed lines denote 95 pecent 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.

to 1999 in Figure 3—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 4 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 5 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. Online Appendix Figure B2 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 online Appendix Table B3, 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 several additional analyses in Gelber, Jones, and Sacks (2019), including varying the time period examined. Overall, these additional analyses generally show similar patterns.

[3] We 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 31st 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: (i) the continued bunching at age 71, which cannot be explained through the coarse measure of age; (ii) the continued adjustment away from the kink from age 70 to age 71 in Figure 5; and (iii) the spike in the elasticity estimated using the Saez (2010) approach in 1990, documented in Figure 8 and explained later in this paper. Moreover, online Appendix Table B1 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.

[4] This interpretation of the patterns around ages 64 and 65 is consistent with Figure B1, 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.

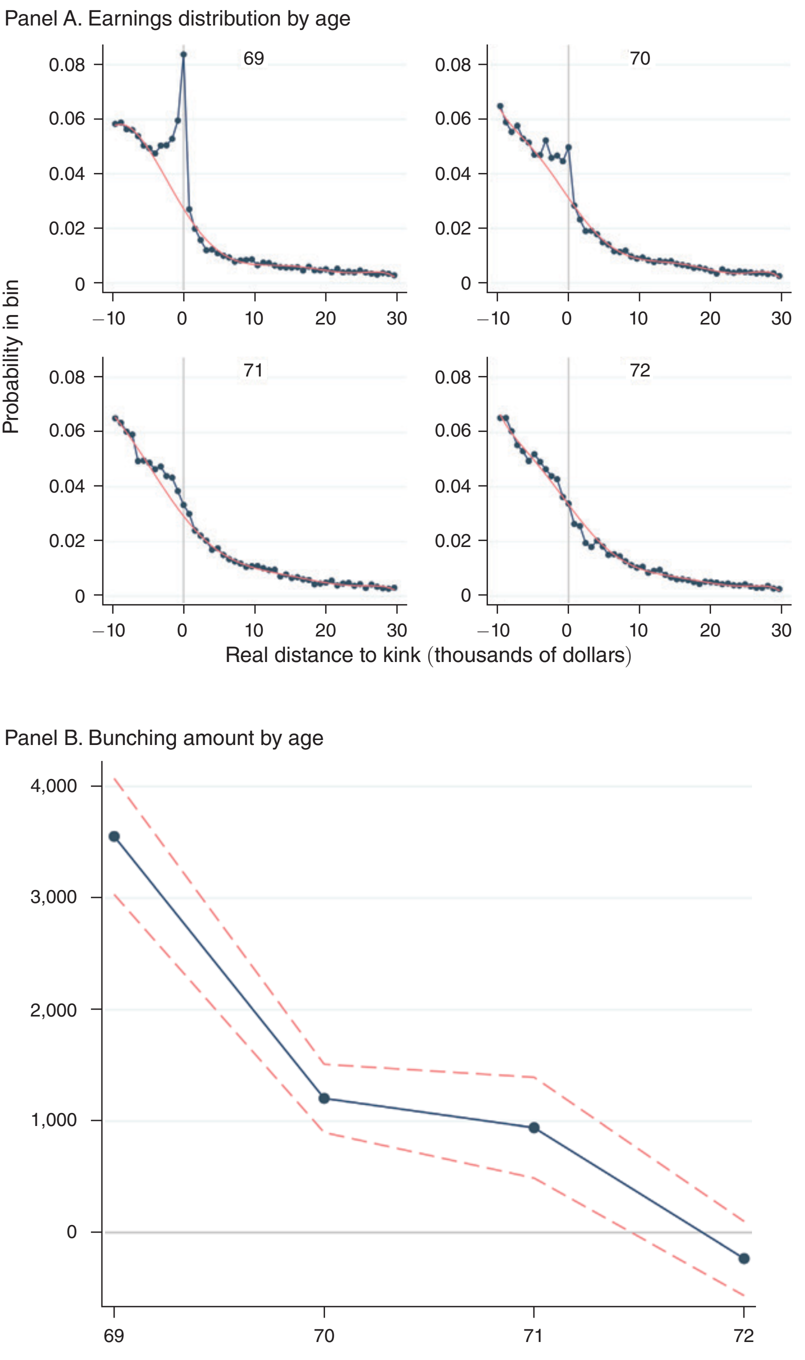

Figure 3. Normalized Excess Mass of Claimants, Ages 69 to 72, 1983 to 1999

Notes: 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.

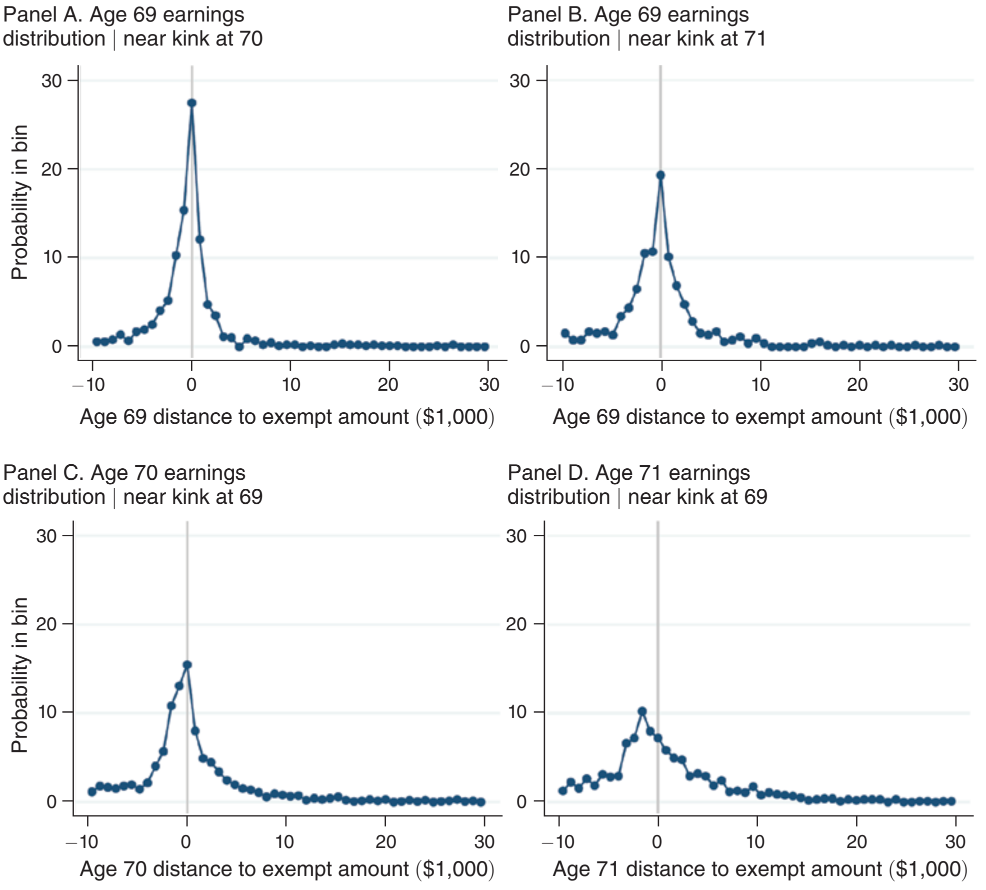

Figure 4. Inertia in Bunching from 69 to 70 and 71

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 online Appendix Figure B1.

V. 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 III. 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

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

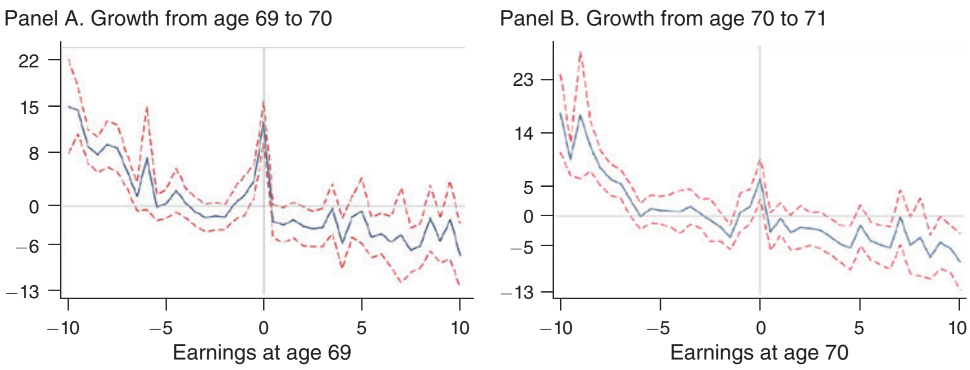

Figure 5. Mean Percentage Change in Earnings from Age t to $t+1$, by Earnings at Age t, 1990–1998

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.

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

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

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:

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

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.

A. *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 (e.g., Simon 1955; Chetty, Looney, and Kroft 2007; and 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 1988, and Dickens and Lundberg 1993) and public finance (e.g., Chetty et al. 2011 and 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 $\left[z^*, z^* + \Delta z_1^*\right]$. 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 quasi-linear, as in related recent public finance literature (e.g., Saez 2010, Chetty et al. 2011, Kleven and Waseem 2013, and Kleven et al. 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 who is indifferent between staying at the initial level of earnings $\underline{z}_1$ (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 $\left[\underline{z}_1, z^* + \Delta z_1\right]$ bunch at the kink K_1. Bunching is given by the integral of the initial earnings density, $h_0(\cdot)$, over this range:

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

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

[7] To see this, note that the utility gain from re-optimizing is $u\left((1-\tau_1)z_1 + R_1, z_1; a\right) - u\left((1-\tau_1)z_0 + R_1, z_0; a\right) \approx u_c \cdot (1-\tau_1)\left[z_1 - z_0\right] + u_z \cdot \left[z_1 - z_0\right] = u_c \cdot (\tau_1 - \tau_0)\left[z_0 - z_1\right]$, where in the first expression, we have used a first-order approximation for utility at $\left((1-\tau_0)z_0 + R_0, z_0\right)$, 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 $\left[z_0 - z_1\right]$ to increase at a rate that dominates.

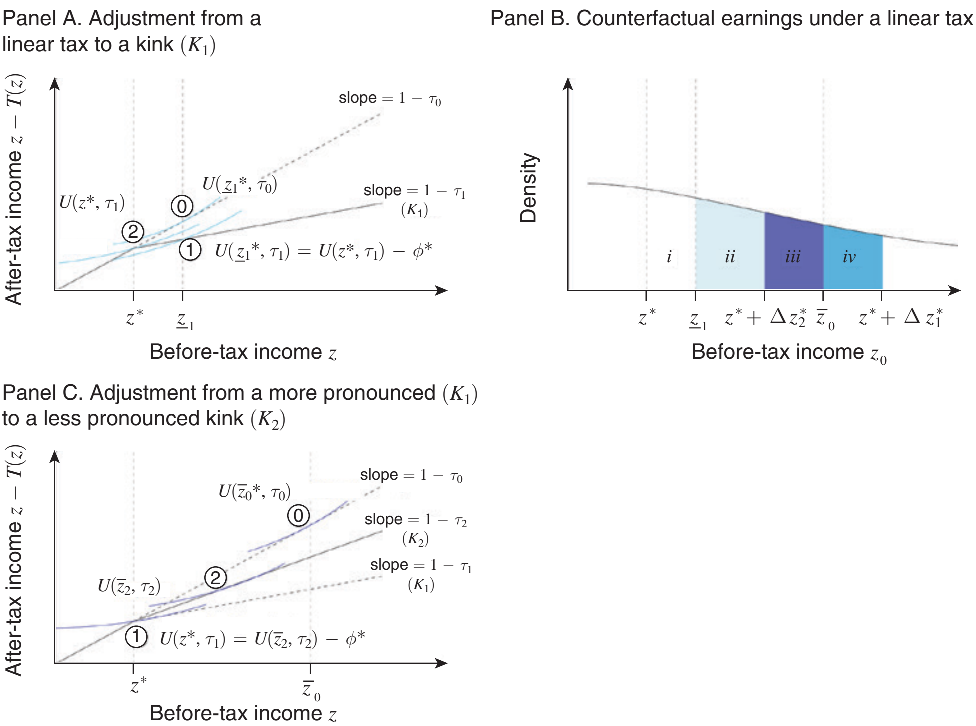

FIGURE 6. BUNCHING RESPONSES TO A CONVEX KINK, WITH FIXED ADJUSTMENT COSTS

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

where $\boldsymbol{\tau}_1 = (\tau_0, \tau_1)$ reflects the tax rates below and above z^*. The threshold level of earnings $\underline{z}_1$ 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, $\underline{z}_1$, is implicitly defined by the indifference condition shown in Figure 6, panel A:

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

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.

B. *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 earlier. 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

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

where $\tilde{\boldsymbol{\tau}}_2 = (\tau_0, \tau_1, \tau_2)$, 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]

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

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

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

[8] We 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.

eliminated, we can attribute any residual bunching to adjustment costs. Moreover, we show in Gelber, Jones, and Sacks (2013) that the absolute value of the decrease in bunching from K_1 to K_2 is decreasing in the adjustment cost: $\bar{z}_0$ is increasing in the adjustment cost, and therefore area *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.

C. *Dynamic Version of Model*

To account for how bunching evolves over time, as in the lagged adjustment shown in Section IV, 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 online Appendix A.A2, 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 earlier 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 quasi-linear utility, $u(c,z;a) = c - (a/(1+1/\varepsilon))(z/a)^{1+1/\varepsilon}$, 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):

$$
\begin{aligned}
(8) \qquad B_1^t &= \int_{\underline{z}_1}^{z^*+\Delta z_1^*} h_0(\zeta)\, d\zeta + \left(1 - \prod_{j=1}^{t} \pi_j\right) \int_{z^*}^{\underline{z}_1} h_0(\zeta)\, d\zeta \\
&= \prod_{j=1}^{t} \pi_j \cdot B_1 + \left(1 - \prod_{j=1}^{t} \pi_j\right) B_1^*,
\end{aligned}
$$

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 VB, 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 $t > \mathcal{T}_1$:

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

[9] The quasi-linearity 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_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 bunched in period 1 and remain bunching at the smaller kink. Second, in area *iii* of the figure, excess bunchers who immediately bunched 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 periods 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 VA–VB 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 $\pi_j = 1$.

D. *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 online Appendix A.A3, we derive generalized formulae for bunching that allow us to interpret our comparative static model estimates as average parameters among the set of bunchers. Online Appendix A.A3 further discusses how the dynamic model can be interpreted in the presence of heterogeneity in these parameters and the vector $\boldsymbol{\pi}_i$.[10]

Second, our model 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 online Appendix A.A4, 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 later in this paper.

Third, the model with a fixed cost of adjustment can be generalized to any polynomial functional form of the adjustment. In online Appendix A.A1, we discuss estimation of a model with a cost of adjustment that has both a fixed cost component and

[10] Special 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.

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.

E. *Econometric Estimation of the Model*

We estimate the model we have described using a minimum distance estimator. As explained in online Appendix A.A6, 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 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, $(\varepsilon, \phi, \pi_1, \ldots, \pi_J)$, 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 $\boldsymbol{\pi}$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 nonparametric 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 nonparametrically we use the empirical earnings distribution for 72-year-olds in \$800 bins as the counterfactual distribution. The earnings density of 72-year-olds 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 quasi-linear utility function, $u(c, z; a) = c - \big(a/(1 + 1/\varepsilon)\big)(z/a)^{1+1/\varepsilon}$, 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 life-cycle 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 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

[11] Because 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.

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 online Appendix A.A5, we demonstrate identification more formally.

VI. Estimates of Elasticity and Adjustment Cost

A. *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 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 earlier discussion, 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 implicit 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).

[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.

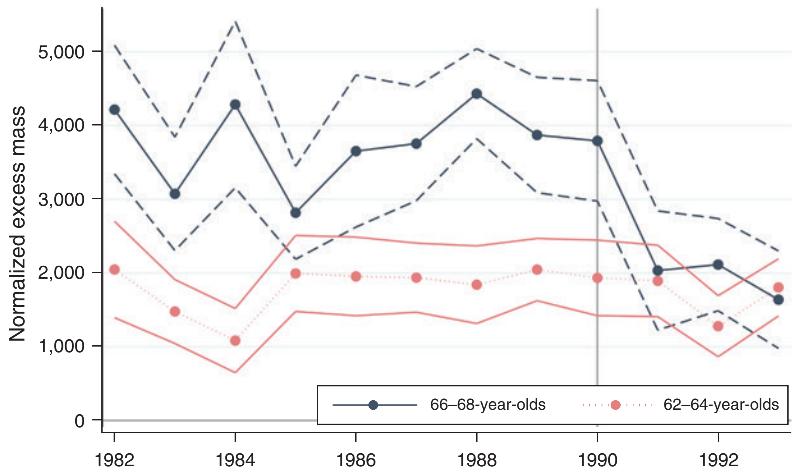

Figure 7. Comparison of Normalized Excess Mass among 62–64-Year-Olds and 66–68-Year-Olds, 1982–1993

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.

Table 2—Estimates of Elasticity and Adjustment Cost: Variation around 1990 Policy Change

	ε	ϕ	$\varepsilon \mid \phi = 0$	
			1990	1989
	(1)	(2)	(3)	(4)
Baseline	0.35	\$278	0.58	0.31
	[0.31, 0.43]	[58, 388]	[0.45, 0.73]	[0.24, 0.39]
Uniform density	0.21	\$162	0.36	0.19
	[0.18, 0.24]	[55, 211]	[0.30, 0.43]	[0.16, 0.23]
Benefit enhancement	0.58	\$151	0.87	0.52
	[0.50, 0.72]	[17, 220]	[0.69, 1.11]	[0.41, 0.66]
Excluding FICA	0.49	\$319	0.74	0.42
	[0.44, 0.59]	[60, 365]	[0.58, 0.94]	[0.33, 0.54]
Bandwidth = \$400	0.45	\$103	0.62	0.43
	[0.36, 0.58]	[0, 478]	[0.47, 0.81]	[0.32, 0.56]
Bandwidth = \$1,600	0.33	\$251	0.55	0.30
	[0.29, 0.43]	[34, 407]	[0.43, 0.72]	[0.23, 0.40]

Notes: The table shows estimates of the elasticity and adjustment cost using the method described in Section VB, 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 percent to 24 percent 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, and column 4 uses only data from 1989.

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 online Appendix Table B2 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 online Appendix Table B4, 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 earlier and in online Appendix A.A4.

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. Online Appendix Table B5 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

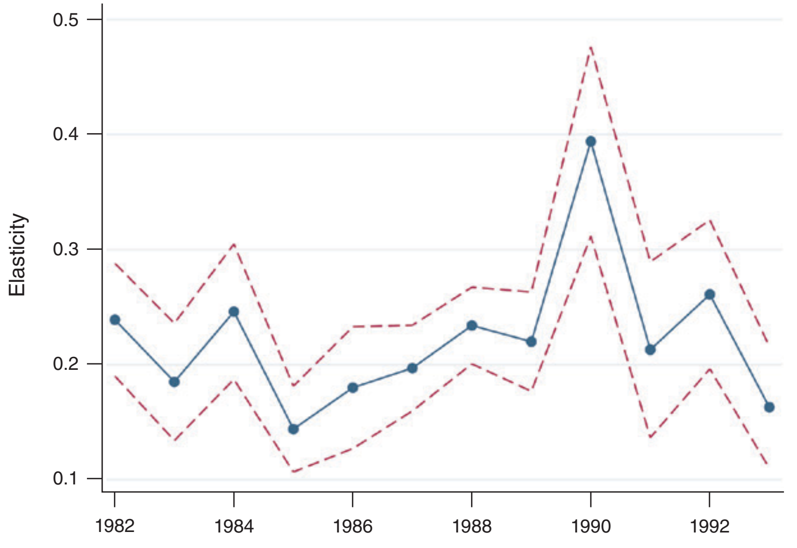

Figure 8. Elasticity Estimates by Year, Saez (2010) Method, 1982–1993

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

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.

B. *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^*} terms—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

TABLE 3—ESTIMATES OF ELASTICITY AND ADJUSTMENT COST: DISAPPEARANCE OF KINK AT AGE 70

	ε (1)	ϕ (2)	$\varepsilon \mid \phi = 0$, age 69 (3)
Baseline	0.42	$90	0.38
	[0.35, 0.53]	[20, 349]	[0.32, 0.47]
Uniform density	0.28	$90	0.25
	[0.24, 0.33]	[21, 238]	[0.22, 0.30]
Benefit enhancement	0.62	$59	0.58
	[0.53, 0.77]	[13, 205]	[0.49, 0.71]
Excluding FICA	0.53	$83	0.49
	[0.45, 0.66]	[19, 305]	[0.42, 0.61]
Bandwidth = $400	0.39	$62	0.36
	[0.31, 0.48]	[25, 133]	[0.28, 0.45]
Bandwidth = $1,600	0.45	$100	0.41
	[0.37, 0.56]	[20, 444]	[0.33, 0.49]
68–70-year-olds	0.44	$42	0.43
	[0.38, 0.58]	[0.49, 267]	[0.37, 0.50]
69- and 71-year-olds	0.45	$175	0.38
	[0.36, 0.86]	[30, 1,053]	[0.32, 0.47]
Born January–March	0.48	$86	0.49
	[0.36, 0.76]	[10, 1,008]	[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

	ε (1)	ϕ (2)
Baseline	0.39	$160
	[0.34, 0.46]	[59, 362]
Uniform density	0.22	$105
	[0.20, 0.24]	[47, 185]
Benefit enhancement	0.62	$100
	[0.55, 0.74]	[33, 211]
Excluding FICA	0.41	$67
	[0.37, 0.48]	[9, 192]
Bandwidth = $400	0.46	$94
	[0.39, 0.55]	[25, 399]
Bandwidth = $1,600	0.37	$135
	[0.32, 0.45]	[43, 299]

Notes: This table implements our "comparative static" method, applied to pooled data from two policy changes: (i) around the 1989/1990 transition analyzed in Table 2 and (ii) 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

	ε (1)	ϕ (2)	π_1 (3)	$\pi_1\pi_2$ (4)	$\pi_1\pi_2\pi_3$ (5)
Baseline	0.36	$243	0.64	0.22	0.00
	[0.34, 0.40]	[34, 638]	[0.39, 1.00]	[0.00, 1.00]	[0.00, 0.14]
Uniform density	0.21	$81	1.00	0.31	0.00
	[0.20, 0.23]	[31, 186]	[0.61, 1.00]	[0.00, 0.92]	[0.00, 0.15]
Benefit enhancement	0.59	$53	1.00	0.37	0.00
	[0.54, 0.64]	[17, 167]	[0.78, 1.00]	[0.00, 1.00]	[0.00, 0.04]
Excluding FICA	0.39	$55	1.00	0.00	0.00
	[0.37, 0.43]	[9, 165]	[1.00, 1.00]	[0.00, 0.00]	[0.00, 0.00]
Bandwidth = $400	0.40	$74	1.00	0.47	0.00
	[0.36, 0.44]	[20, 271]	[0.69, 1.00]	[0.09, 0.94]	[0.00, 0.10]
Bandwidth = $1,600	0.36	$99	0.88	0.52	0.00
	[0.34, 0.39]	[19, 494]	[0.37, 1.00]	[0.04, 1.00]	[0.00, 0.07]

Notes: The table shows estimates of the elasticity and adjustment cost using the dynamic method described in Section VC. 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. Also, $\pi_1\pi_2\pi_3\pi_4$ 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).

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_js, 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.

VII. 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 policymakers 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 online Appendix A.A7 and online 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 online 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.

VIII. 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 policymakers' 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 United States 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, Schramm, and Whitten 2016; Gudgeon and Trenkle 2016; Zaresani 2018).

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 misperceptions 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 overestimate 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, Andrew B., and Janice C. Eberly. 1994. "A Unified Model of Investment under Uncertainty." *American Economic Review* 84 (5): 1369–84.

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In *Producer Dynamics: New Evidence from Micro Data*, edited by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, 149–230. Chicago: University of Chicago Press.

Altonji, Joseph G., and Christina H. Paxson. 1988. "Labor Supply Preferences, Hours Constraints, and Hours-Wage Trade-Offs." *Journal of Labor Economics* 6 (2): 254–76.

Arrow, Kenneth J., Theodore Harris, and Jacob Marschak. 1951. "Optimal Inventory Policy." *Econometrica* 19 (3): 250–72.

Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov. 2015. "The Welfare Economics of Default Options in 401(k) Plans." *American Economic Review* 105 (9): 2798–2837.

Baumol, William J. 1952. "The Transactions Demand for Cash: An Inventory Theoretic Approach." *Quarterly Journal of Economics* 66 (4): 545–56.

Blomquist, Sören, and Whitney K. Newey. 2017. "The Bunching Estimator Cannot Identify the Taxable Income Elasticity." http://economics.mit.edu/files/13770.

Blundell, Richard, and Thomas Macurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In *Handbook of Labor Economics*, Vol. 3A, edited by Orley C. Ashenfelter and David Card, 1559–1695. Amsterdam: Elsevier.

Brown, Jeffrey R., Arie Kapteyn, Olivia S. Mitchell, and Teryn Mattox. 2013. "Framing the Social Security Earnings Test." University of Pennsylvania Pension Research Council Working Paper 2013-06.

Burtless, Gary, and Robert A. Moffitt. 1985. "The Joint Choice of Retirement Age and Postretirement Hours of Work." *Journal of Labor Economics* 3 (2): 209–36.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–98.

Chetty, Raj. 2012. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80 (3): 969–1018.

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

Chetty, Raj, John N. Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103 (7): 2683–2721.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2012. "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-analysis of Extensive Margin Elasticities." In *NBER Macroeconomics Annual* 27 (1): 1–56.

Chetty, Raj, Adam Looney, and Kory Kroft. 2007. "Salience and Taxation: Theory and Evidence." NBER Working Paper 13330.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77.

Cogan, John F. 1981. "Fixed Costs and Labor Supply." *Econometrica* 49 (4): 945–63.

Coile, Courtney, and Jonathan Gruber. 2001. "Social Security Incentives for Retirement." In *Themes in the Economics of Aging*, edited by David A. Wise, 311–54. Chicago: University of Chicago Press.

Congressional Budget Office. 2009. "Estimated Macroeconomic Impacts of the American Recovery and Reinvestment Act of 2009." Washington, DC: Government Printing Office.

Diamond, Peter, and Jonathan Gruber. 1999. "Social Security and Retirement in the United States." In *Social Security and Retirement around the World*, edited by Jonathan Gruber and David Wise, 437–73. Chicago: University of Chicago Press.

Dickens, William, and Shelly Lundberg. 1993. "Labor Supply with Hours Restrictions." *International Economic Review* 34 (1): 169–92.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2015. "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D." *Quarterly Journal of Economics* 130 (2): 841–99.

Engelhardt, Gary V., and Anil Kumar. 2014. "Taxes and the Labor Supply of Older Americans: Recent Evidence from the Social Security Earnings Test." *National Tax Journal* 67 (2): 443–58.

Feenberg, Daniel, and Elisabeth Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management* 12 (1): 189–94.

Friedman, Milton. 1953. *Essays in Positive Economics*. Chicago: University of Chicago Press.

Friedberg, Leora. 1998. "The Social Security Earnings Test and Labor Supply of Older Men." In *Tax Policy and the Economy*, Vol. 12, edited by James M. Poterba, 121–50. Chicago: University of Chicago Press.

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

Garicano, Luis, Claire Lelarge, and John Van Reenen. 2016. "Firm Size Distortions and the Productivity Distribution: Evidence from France." *American Economic Review* 106 (11): 3439–79.

Gelber, Alexander M., Damon Jones, and Daniel W. Sacks. 2013. "Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test." NBER Working Paper 19491.

Grossman, Sanford J., and Guy Laroque. 1990. "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods." *Econometrica* 58 (1): 25–51.

Gruber, Jon, and Emmanuel Saez. 2002. "The Elasticity of Taxable Income: Evidence and Implications." *Journal of Public Economics* 84: 1–32.

Gruber, Jonathan, and Peter Orszag. 2003. "Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt?" *National Tax Journal* 56 (4): 755–73.

Gruber, Jonathan, and David A. Wise, eds. 1999. "Introduction and Summary." In *Social Security and Retirement around the World*, 1–35. Chicago: University of Chicago Press.

Gudgeon, Matthew, and Simon Trenkle. 2016. "Frictions in Adjusting Earnings: Evidence from Notches in German Mini Jobs." http://conference.iza.org/conference_files/SUMS_2016/gudgeon_m24051.pdf.

Hausman, J.A. 1981. "Labor Supply." In *How Taxes Affect Economic Behavior*, edited by Henry J. Aaron and Joseph A. Pechman, 27–72. Washington, DC: Brookings Institution.

He, D., L. Peng, and X. Wang. 2016. "Understanding the Elasticity of Taxable Income: A Tale of Two Approaches." Unpublished.

Hoopes, Jeffrey L., Daniel H. Reck, and Joel Slemrod. 2015. "Taxpayer Search for Information: Implications for Rational Attention." *American Economic Journal: Economic Policy* 7 (3): 177–208.

Ito, Koichiro. 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review* 104 (2): 537–63.

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics* 8: 435–64.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thrustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark." *Econometrica* 79 (3): 651–92.

Kleven, Henrik Jacobsen, Camille Landais, Emmanuel Saez, and Esben Schultz. 2014. "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark." *Quarterly Journal of Economics* 129 (1): 333–78.

Kleven, Henrik J., and Mazhar Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics* 128 (2): 669–723.

Kline, Patrick, and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Quarterly Journal of Economics* 131 (4): 1795–1848.

Leahy, J. 2008. "s-S Models." In *The New Palgrave Dictionary of Economics*, Second Edition, edited by Steven N. Durlauf and Lawrence E. Blume. New York: Palgrave MacMillan.

Liebman, Jeffrey B., and Erzo F.P. Luttmer. 2012. "The Perception of Social Security Incentives for Labor Supply and Retirement: The Median Voter Knows More Than You'd Think." *Tax Policy and the Economy* 26 (1): 1–42.

Mortenson, J., H. Schramm, and A. Whitten. 2016. "The Effects of Required Minimum Distribution Rules on Withdrawals from Traditional IRAs." Unpublished.

Nakamura, Emi, and Jón Steinsson. 2010. "Monetary Non-neutrality in a Multisector Menu Cost Model." *Quarterly Journal of Economics* 125 (3): 961–1013.

Newey, Whitney K., and Daniel McFadden. 1994. "Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics*, Vol. 4, edited by Robert F. Engle and Daniel L. McFadden, 2111–2245.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): 180–212.

Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature* 50 (1): 3–50.

Schächtele, Simeon. 2016. "The Importance of Deductions in Response to the Personal Income Tax: Bunching Evidence from Germany." Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Tax Responses, No. E10-V2.

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99–118.

Sheshinski, Eytan, and Yoram Weiss. 1977. "Inflation and the Costs of Price Adjustment." *Review of Economic Studies* 44 (2): 287–303.

Social Security Administration. 2012. *Online Social Security Handbook*. https://www.ssa.gov/OP_Home/handbook/handbook.html (accessed October 10, 2019).

Social Security Administration. 2013. *Annual Statistical Supplement to the Social Security Bulletin, 2012*. Washington, DC: Social Security Administration.

Song, Jae G., and Joyce Manchester. 2007. "New Evidence on Earnings and Benefit Claims Following Changes in the Retirement Earnings Test in 2000." *Journal of Public Economics* 91 (3–4): 669–700.

Stokey, Nancy L. 2009. *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton: Princeton University Press.

Zaresani, Arezou. 2018. "Adjustment Costs and Incentives to Work: Evidence from a Disability Insurance Program." https://azaresani.com/wp-content/uploads/AISH2016.pdf.