NBER WORKING PAPER SERIES

MISPERCEPTIONS OF NONLINEAR BUDGET SETS: EVIDENCE FROM ADMINISTRATIVE TAX DATA

Alexander M. Gelber
Damon Jones
Ithai Lurie
Daniel W. Sacks

Working Paper 33496
http://www.nber.org/papers/w33496

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02134
February 2025

This research was completed partly while Gelber was on leave at the Stanford Institute for Economic Policy Research, partly funded by a grant from the Alfred P. Sloan foundation. This research was conducted while Ithai Lurie was an employee at the U.S. Department of the Treasury. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily reflect the views or the official positions of the U.S. Department of the Treasury. Any taxpayer data used in this research was kept in a secured IRS data repository, and all results have been reviewed to ensure that no confidential information is disclosed. For helpful comments we thank seminar participants at Stanford. All errors are our own. The research reported herein was also performed pursuant to grant RDR18000003 from the US Social Security Administration (SSA) funded as part of the Retirement and Disability Research Consortium. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA, any agency of the Federal Government, or NBER. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States Government or any agency thereof. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33496

Misperceptions of Nonlinear Budget Sets: Evidence from Administrative Tax Data
Alexander M. Gelber, Damon Jones, Ithai Lurie, and Daniel W. Sacks
NBER Working Paper No. 33496
February 2025
JEL No. H20, H24, H55, J14, J22, J26

ABSTRACT

Budget set kinks are much studied in economics, including in the context of "bunching" estimators that assume individuals react to the true marginal tax rate. We document that individuals disproportionately "left-bunch" below kinks in the context of the Social Security Earnings Test where incentives from its actuarial adjustments should instead push many rational agents to bunch above the kink. We show that the left bunching in this case cannot be explained through standard, rational reactions to the incentives. We demonstrate that these patterns represent the first empirical evidence consistent with "spotlighting," wherein individuals misperceive the local marginal tax rates as applying throughout the tax schedule and therefore treat the kink as a notch. In the context of the Earnings Test, this misperception provides an explanation for why literature has found large earnings responses despite the fact that the Earnings Test typically creates weak incentives for rational agents to adjust earnings. More generally, if individuals perceive kinks as notches, then elasticities estimated from bunching at kinks where this misperception may be at play may be significantly over-estimated.

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

Damon Jones
Harris School of Public Policy
University of Chicago
1307 E 60th Street
Chicago, IL 60637
and NBER
damonjones@uchicago.edu

Ithai Lurie
Department of the Treasury
1500 Pennsylvania Avenue NW
Washington, DC 20220
Ithai.Lurie@treasury.gov

Daniel W. Sacks
University of Wisconsin - Madison
975 University Ave
Madison, WI 53715
dan.sacks@wisc.edu

1 Introduction

Nonlinear price schedules are common across a wide variety of economic applications, including labor supply (e.g. Hausman, Hausman and Holl, 1981), health insurance (e.g. Einav and Levin, 2015), electricity demand (e.g. Ito, 2014), and retirement savings (e.g. Bernheim, Fradkin and Popov, 2015). In particular, many price schedules and policies feature kink points, which are discontinuous changes in the price of a good. For example, many means-tested government transfer programs reduce or eliminate transfer benefits as income rises above threshold levels, and the individual income tax in many countries generates a piecewise linear budget set with kinks at each point at which the marginal tax rate (MTR) jumps. Previous literature has developed methods for explicitly using kink points to estimate price or tax elasticities (e.g. Burtless and Hausman, 1978; Hausman, Hausman and Holl, 1981; Hoynes, 1996; Saez, 2010; Chetty et al., 2011*b*; Kleven and Waseem, 2013; Bernheim, Fradkin and Popov, 2015; Kleven, 2016, for a review).

These papers that estimate elasticities from kinks have generally assumed that individuals respond to the marginal tax rate, as in standard models. However, outside the literature on elasticities and bunching at kinks, other papers have suggested that individuals confuse average and marginal tax rates (Liebman and Zeckhauser, 2004; Rees-Jones and Taubinsky, 2019). The notion that individuals react to the average tax rate, instead of reacting to the marginal tax rate as in the standard models, has been proposed as an explanation for the lack of "bunching" observed at some kink points (Ito, 2014). This still leaves unexplained why bunching does occur at many other kink points that have been the focus of much literature. Puzzlingly, when bunching does occur, it is often asymmetric, with much more bunching below the kink point than above it, a phenomenon we call "left bunching."

In this paper, we perform the first systematic analysis of left bunching, and explore two classes of possible explanations. First, we note that if a kink is imposed in the presence of a downward-sloping density of outcomes, standard theory implies that individuals should left-bunch. The imposition of the kink causes the density to shift downward, which leads to fewer bunchers above than below the kink. In principle, this could explain the left bunching that has been observed in such circumstances. We call this the "standard" candidate explanation for left bunching. Second, individuals could left-bunch because of some "behavioral" deviation from standard theory, including misperceiving the tax schedule or other explanations.

We tease apart these classes of explanations in the key context of the U.S. Social Security Earnings Test. The Earnings Test reduces Social Security Old Age and Survivors Insurance ("Social Security") benefits in a given year as a proportion of a Social Security claimant's earnings above an exempt amount in that year. For example, for Social Security claimants under age 66 in 2019, current Social Security benefits are reduced by one dollar for every two dollars earned above $17,640. Previous literature has found that Social Security claimants bunch at this convex kink (Burtless and Moffitt, 1985; Friedberg, 1998, 2000; Song and Manchester, 2007; Engelhardt and Kumar, 2014; Gelber, Jones and Sacks, 2020), and that employment falls due to the Earnings Test (Friedberg and Webb, 2009; Gelber et al., 2021, 2022). In addition to providing a laboratory for studying left bunching, the Earnings Test is important to policy-makers in its own right. In the latest year of the available micro-data, 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 has increased as the affected age range – those at or below the Normal Retirement Age (NRA) – has expanded gradually from 65 for cohorts born before 1938, to age 67 for those born in 1960 and later.

Reductions in current benefits due to the Earnings Test sometimes lead to increases in later benefits through an actuarial adjustment of benefits. In particular, there is a little-understood "notch" in the budget set just over the exempt amount: when individuals earn just above this level, their benefits once they reach the NRA are adjusted upward by five-ninths of one percent. Thus, the incentives – understood properly – should lead Social Security beneficiaries to locate just above the exempt amount, i.e. they should "right-bunch." Moreover, benefits after reaching the Normal Retirement Age are adjusted upward by five-ninths of one percent for every month of Social Security benefits that experiences any reduction due to the Earnings Test, which significantly dulls the incentives to bunch or reduce earnings (Social Security Administration, Section 728.2, 2018; Gruber and Orszag, 2003). This has led to a longstanding puzzle in the Earnings Test literature: why do earnings respond strongly to the Earnings Test, despite the actuarial adjustment of benefits (Burtless and Moffitt, 1985; Gelber, Jones and Sacks, 2020)?

Using administrative tax data from the Internal Revenue Service (IRS) on a one-hundred percent sample of the U.S. residents with a Social Security Number (SSN), born between 1939 and 1954, we show that nearly all bunching at the exempt amount is left-bunching. Several pieces of evidence rule out the standard explanation as entirely accounting for this left bunching. First, left bunching does not only occur amid the downward-sloping densities postulated in the standard explanation; it occurs even at ages when the distribution of earnings is close to flat around the exempt amount. Second, an illustrative simulation of a rational model of bunching indeed yields far less left bunching than we observe. Third, in a panel of data we can proxy for individuals' desired earnings in the absence of the Earnings Test by examining their earnings in years just prior to reaching retirement age and facing the Earnings Test. We show that individuals overwhelmingly left-bunch, particularly those whose earnings just prior to reaching retirement age were substantially above the exempt amount. This contrasts with behavior in the standard explanation, wherein these individuals – especially those initially locating far above the exempt amount – would tend to right-bunch.

Having dispatched the standard explanation as the sole origin of these patterns, we explore other explanations. One possibility is that some individuals exhibit "spotlighting," in which individuals perceive the local marginal tax rate to apply everywhere in the budget set (Liebman and Zeckhauser, 2004)[1]. Individuals who spotlight believe that their average tax rate increases discontinuously as they earn just above the exempt amount. In other words, even though the Earnings

[1] Liebman and Luttmer (2015) and Brown et al. (2013) document that many individuals do not understand the Earnings Test or other aspects of Social Security, but do not specifically develop evidence on whether individuals exhibit spotlighting or other specific types of systematic misperceptions.

Test in fact applies only to marginal earnings above the exempt amount, in this explanation, spotlighters would perceive that the Earnings Test also applies to infra-marginal earnings below the exempt amount, creating a notch, i.e. a discrete loss of income at the exempt amount.

To substantiate this explanation, we document that there is a downward discontinuity in the employment probability as a function of desired earnings in the presence of a linear budget set. We specify a stylized model and show that in this framework, such a downward discontinuity should occur in the presence of a (perceived) notch in the budget set, but not in the presence of a kink. We also find that some individuals locate just above the exempt amount, implying that either some individuals are inert to the perceived notch (Kleven and Waseem, 2013), or that there is a mixture of types in which some spotlight and others react according to the standard model. As we find only a small discontinuity in employment, our results are most consistent with a relatively small share of individuals who both spotlight and are inert on the intensive margin. The downward discontinuity in the employment probability is also inconsistent with other non-rational explanations that could be posited for left bunching, such as loss aversion to the Earnings Test combined with "diminishing sensitivity" (see Rees-Jones et al. 2018 on loss aversion and bunching in a public finance context).

To our knowledge, this is the first evidence of spotlighting. Spotlighting contrasts with the "ironing" phenomenon documented in Ito (2014) and Taubinsky and Rees-Jones (2018). Under spotlighting individuals perceive the local marginal tax rate as applying everywhere (i.e. as being the average tax rate), whereas under ironing individuals react to the average tax rate, instead of reacting to the marginal tax rate as a rational agent would. Ironing can help explain the lack of bunching at many kinks, while leaving unexplained why bunching occurs at other kinks; meanwhile, spotlighting can help explain the bunching at kinks that have been used to estimate elasticities, which often features left bunching.

If bunching at other kinks is due to spotlighting, then estimates of elasticities based on bunching at such kinks will not apply more generally. Much of the literature estimates elasticities at kinks under the assumption of a standard reaction to the kink incentives. This can could be considered a "policy elasticity" (Hendren, 2016)—but if many individuals are instead reacting to a perceived notch, then the elasticity with respect to the perceived marginal tax rate is much lower. In other words, the elasticity necessary to rationalize a given amount of observed bunching is far lower with a (perceived) budget set notch than with a kink, because a notch creates much sharper incentives than a kink. This suggests that, for example, the elasticity with respect to a correctly-perceived tax rate on a linear budget set would be much lower than the elasticity estimated by erroneously applying standard methodology to bunching at a kink, rather than recognizing that some individuals perceive this kink as a notch. We argue that this could have broad implications for understanding bunching-based elasticity estimates at kinks. Within the specific context of the Earnings Test, these findings also provide a novel explanation for the strong earnings reaction to the Earnings Test despite the actuarial adjustment: individuals misperceive the incentives to reduce earnings as being much more severe than they are in reality.

We next describe the features of the Earnings Test in more detail, followed by a framework for interpreting left bunching. Subsequently, we describe our data and our empirical results.[2]

2 Policy Setting

Social Security Old-Age and Survivors Insurance (OASI)—hereafter referred to simply as "Social Security"—provides benefits to older Americans and survivors of deceased workers, but delivery of benefits can be affected by whether one is currently working. For those who are simultaneously working and claiming Social Security benefits, the Social Security Annual Earnings Test (AET) reduces current benefits in proportion to earnings above an exempt amount, while typically adjusting future benefits upward in an actuarial fair fashion. For example, consider a 63-year-old earning $23,640 in 2019, receiving $1,000 in monthly benefits, and facing a $17,640 exempt amount and a 50 percent benefit reduction rate (BRR). Their current annual benefits would be reduced by $3,000 = ($23,640 - $17,640)×50%, equal to 3 months of benefits. In general, the exempt amount and the benefit reduction rate depend on age. People can claim benefits on their own record as early as age 62, the Early Entitlement Age (EEA), and the Earnings Test applies until one reaches the Normal Retirement Age (NRA).

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

3 Conceptual Framework

We begin by noting evidence that individuals tend to left-bunch across a variety of contexts. Kleven (2016) notes that a number of papers find more bunching below the threshold of a kink than above – citing the examples of Devereux, Xing and Maffini (2016), Gelber, Jones and Sacks (2020), and Seim (2017) – though none of these papers notes the left bunching themselves, and neither these papers nor Kleven (2016) explores left bunching further. These papers span the disparate contexts of corporate taxes in the U.K., earnings responses to Social Security in the U.S., and wealth responses to wealth taxes in Sweden. Other examples are plentiful, including Peng, Wang and He (2019), studying income taxes in China, Le Barbanchon (2016), studying unemployment insurance in the U.S., and (Asatryan and Peichl, 2018) studying Armenian firms. Moreover, we have not

[2] Because the policy environment and data are similar to Gelber, Jones and Sacks (2020); Gelber et al. (2021, 2022), the corresponding sections of this paper have overlap with those previous papers.

[3] Social Security Administration (2012); Gruber and Orszag (2003).

yet observed a context with significant right-bunching. However, in all of the left bunching cases above, the density is downward-sloping, leading to the question of whether the left bunching is due to the shape of the density in a standard model or due to non-standard factors like misperceptions.[4]

In this section, we sketch a basic framework that is helpful in interpreting left bunching and highlighting the types of underlying models that could generate this phenomenon.

3.1 Baseline Case

We first begin with a standard model of bunching in the presence of a kink, as in Saez (2010). In anticipation of our empirical application, we adapt the setting to incorporate features of the Earnings Test, where a convex kink is created by a reduction in current benefits, as opposed to the standard jump in marginal tax rates. As in Saez (2010), agents draw utility from consumption c and disutility associated with earning income z (i.e. the disutility of labor supply). These two outcomes are linked through the budget constraint:

$$c(z) = z - T(z) + B(z)$$

where the function $T(\cdot)$ represents the tax and transfer system, which we will assume is locally linear. The function $B(\cdot)$ represents the flow of current Social Security benefits and captures the earnings test:

$$B(z) = \begin{cases} B_0 & \text{if } z \le z^* \\ B_0 - BRR \cdot (z - z^*) & \text{if } z > z^* \end{cases}$$

In particular, current benefits are reduced at the benefit reduction rate (BRR) for every dollar earned above the exempt amount z^*, creating a kink in the budget set where the marginal return to earning an additional dollar discontinuously reduces.

We add one additional feature to the baseline case to better capture the reality of earnings data: diffuse bunching. Empirically, bunching at kinks tends to be humped-shaped, rather than a sharp mass point exactly as the kink (Kleven, 2016), which could be due to a number of optimization frictions. Following Saez (1999), we model this with a two-step process: first, the agent chooses a level of earnings z, and disutility of earnings is realized. Second, the realized levels of earnings that determines the level of consumption, net tax and transfers, and benefit reduction will be $z+\epsilon$, where $\epsilon \sim N(0, \sigma_z)$ is an uncertain component of earnings. We now can characterize the baseline maximization problem:

$$\max_z \int u(c(z+\epsilon), z) f(\epsilon) \, d\epsilon$$
$$\text{where } c(z+\epsilon) = z + \epsilon - T(z+\epsilon) + B(z+\epsilon)$$

As has been covered in previous literature (Saez, 2010; Kleven, 2016), the kink in the budget set

[4] Other papers appear to find relatively symmetric bunching. Our conclusions in this paper do not appear to apply in such cases.

results in a set of agents who choose earnings at or near the kink, in response to the higher implicit marginal tax rate above the exempt amount. Importantly, under this model, there is no inherent incentive for agents to bunch to the left or right of the kink: the costs of small misoptimization in either direction symmetrically vanish as one gets close to the kink.

3.2 Extensions to the Baseline Case

We now introduce a number of behavioral extensions to the model that might induce asymmetric bunching near a kink.

Downward sloping ability distribution: First, and most simply, we can generate the appearance of left bunching in the case where the distribution of earnings, even in the absence of a kinked budget set, is steeply downward sloping. This creates more mass to the left than to the right of any point on a downward-sloping portion of the earnings distribution.

Reference dependent preferences: Second, there is the phenomenon of reference-dependent preferences, originally pioneered by Kahneman and Tversky (1979) and reviewed extensively by O'Donoghue and Sprenger (2018). In particular, agents with reference-dependent preferences may exhibit loss aversion, where the disutility of losses, relative to a reference point, loom larger than equally sized gains. We can introduce this feature into our baseline model, and also incorporate the phenomenon of diminishing sensitivity, where the curvature of the gain-loss utility function differs on either side of the reference point. Importantly, we additionally modify reference-dependence in our setting: rather than model a general aversion to losses in income, which amplify both marginal tax rates and benefit reduction rates, we instead only add a gain-loss utility function over benefits, relative to the reference point of baseline Social Security benefits. This is done via a modified, "effective" consumption function:

$$c(z) = z - T(z) + B(z) + \underbrace{g(B(z) - B_0)}_{\text{gain/loss utility}}$$

The gain/loss utility function is as follows:

$$g(x) = \begin{cases} x^{\alpha} & \text{if } x > 0 \\ -\lambda(-x)^{\alpha} & \text{if } x \leq 0 \end{cases}$$

where λ is the loss aversion parameter and α captures diminishing sensitivity. Reference-dependence has been related to bunching behavior in the setting of bunching in effort by marathon runners (Allen et al., 2017) and in the case of manipulation of reported taxable income (Rees-Jones et al., 2018).

Spotlighting: Finally, we consider the case of spotlighting, where agents misperceive marginal tax rates as average tax rates. In the case of a kink in the budget set, this implies that a discontinuous change in the marginal tax rate is mistaken for a discontinuous change in the level of tax liability. We represent this with a perceived benefit schedule $\hat{B}(z)$, defined as follows:

$$\hat{B}(z) = \begin{cases} B_0 & \text{if } z \leq z^* \\ (1 - BRR) \cdot B_0 & \text{if } z > z^* \end{cases}$$

Now, instead of reduction in benefits proportional to the marginal dollar, the agent believes a discrete fraction, BRR, of benefits will be loss when earnings exceed the exempt amount z^* by even a dollar. As explained in detail in Kleven and Waseem (2013), this induces a strong incentive for those just above the threshold to locate on the left side of the tax change.

As a tool to summarize the incentives created by both the kink and uncertainty in earnings, we can make use of an effective budget set, as in Saez (1999). That is, we find the nonlinear tax schedule, $\tilde{T}(z)$, that generates the same optimal earnings decision as our more complicated optimization problem with potentially behavioral preferences and a budget set that features taxes, transfers, and Social Security benefits, as well as income uncertainty:

$$u\left(z - \tilde{T}(z), z\right) \equiv \int u\left(z + \epsilon - T(z+\epsilon) + \hat{B}(z+\epsilon) + g\left(\hat{B}(z+\epsilon) - B_0\right), z\right) f(\epsilon)\, d\epsilon$$

By plotting the effective budget set and comparing it to the actual budget, we can get a sense of how different features of the model—misperceptions of the budget set, non-standard preferences, uncertainty, etc.—might cause behavior to deviate from a standard model of bunching.

3.3 Summary of Bunching Predictions

We now detail how the expected distribution of earnings varies as we introduce the various extensions to the baseline model. To illustrate our results, we carry out numerical simulations, featuring agents with a particular set of preferences, uncertainty in income, and perceptions of the budget set. We either assume a uniform distribution of ability, which results in a uniform distribution of earnings in the absence of a kink in the budget set, or, in one case, a downward sloping ability distribution, as discussed above. In each case, we have a set of agents who are completely inert and just choose the level of earnings that would be chosen in the absence of a kink. This captures the empirical fact that some subset of agents appears to not be able to respond on the margin to kinks and or notches in budget sets (Chetty, 2012; Kleven and Waseem, 2013; Gelber, Jones and Sacks, 2020).

The key object distinguishing different models is the effective budget set. We plot the actual and effective budget sets under our different models in the left set of panels in Figure 1, and we show the simulated earnings distributions in the right set.

We begin with simulations from the baseline model. In all models, the effective budget has a less pronounced kink than the actual budget set, due to the uncertainty in income. This effectively smooths out any nonlinearities, since agents on the lower marginal tax rate side of the kink can end up on the higher marginal rate side, and vice versa, once earnings are realized. This can be seen, for example, in Panel A of Figure 1. In Panel B, we show the resulting distribution of earnings.

There is diffuse bunching at the kink, with a slight skew toward right-bunching.[5] The baseline model does not generate left bunching, but a natural alternative—a downward sloping earnings density—does. We show this in panel D.

We next simulate a set of models with loss aversion. Loss aversion magnifies the size of the kink in budget constraint, as we show in Panel E of Figure 1. We start with a case we call "weak" loss aversion, and we shut down diminishing sensitivity to gains and losses. Weak here means that loss aversion is not strong enough to generate a negatively sloped effective budget constraint. The bunching response to weak loss aversion is qualitatively similar to the baseline case, as seen in Panel F. This makes intuitive sense, since weak loss aversion simply amplifies the kink—makes it steeper than it actually is—so for any loss aversion level, there is an equivalent benefit reduction rate that generates the same economic incentives. Thus, weak loss aversion does not generate left bunching.

We next turn to stronger forms of loss aversion which do generate left bunching, as we show in panels G-J of Figure 1. If we increase the amount of loss aversion enough, the budget set actually slops downward to the right of z^*, which generates behavior similar to that of a notch. This generates left bunching as we show in Panel D. Alternatively, we can also generate left bunching with by adding diminishing sensitivity to our model with weak loss aversion. This generates loss aversion because the convex loss function, which operates on losses above z^*, generates a segment of the effective budget that is non-convex, again generating behavior akin to that of a notch.

Our final model of left bunching is a model of spotlighting. Spotlighting means that people misperceive a kink as a notch, so the effective budget set here falls discontinuously at the kink point. The resulting earnings distribution in Panel L now features a disproportionate amount of bunching to the left of the earnings threshold z^*, because adjusting earnings just below the exempt amount is necessary to avoid perceived losses. Note that in this model, the earnings distribution is continuous; in a frictionless model, a notch would also be predicted to generate a hole in the earning distribution within a dominated region just above z^*. However, in practice, it is often observed that a subset of agents nonetheless locate in this dominated region (Kleven, 2016). As mentioned above, we achieve this by setting a fraction of agents to be inert.

3.4 Distinguishing between models with extensive margin responses

Thus far, we have presented multiple models that might generate left bunching. To distinguish among these models, we turn to an additional outcome: extensive margin behavior, i.e. the decision to work or not. We generate extensive margin responses by introducing a heterogeneous fixed cost of working, say q_0. Our key result is that spotlighting, along with intensive margin frictions, implies an extensive margin response at the the kink point z^*, i.e. a discontinuous drop in employment at z^*. Other models, alone or in combination with frictions, do not.

Here we present the basic set up and intuition. Appendix B contains the model and formal

[5] Saez (1999) shows that when earnings are uncertain, as in our model, the effective kink in the modified budget set moves slightly to the right, which actually pushes agents away from left bunching.

statements, Appendix C contains the proofs. The model extends Saez (2010) to accommodate frictions and extensive margin responses.[6] The key result from this model is that the combination of spotlighting and intensive margin adjustment frictions generate extensive margin responses to a kink.

The intuition is as follows. Introduce a kink at z^*, where the tax rate increases from τ_0 to τ_1. Consider an individual with baseline earnings (i.e., earnings under τ_0) that are close to but above z^*, say $z = z^* + dz$. This individual will want to adjust their earnings, and they can do so on the extensive margin—drop out of work entirely—or on the intensive margin. Which do they choose? In the absence of both spotlighting and intensive margin frictions, they will adjust on the intensive margin, to z^*. This is because if their baseline earnings are near z^*, they are roughly indifferent between that and z^* (by envelope theorem type logic), and so incur no welfare loss by adjusting on the intensive margin. But dropping out of employment does involve a first order loss, except in the probability zero case that they were indifferent between working and not.[7] Thus in the absence of both adjustment frictions and spotlighting, we do not get an extensive margin response.

Now let us introduce adjustment frictions, so that the individual cannot adjust her earnings to exactly z^* without incurring some cost. This individual will not choose an extensive margin response, because the kink changes their utility by $dz(\tau_1 - \tau_0)$. Since dz is small, this utility loss is small, whereas, again, exiting employment entails a large utility loss with high probability. So adjustment frictions alone do not generate an extensive margin response.

Finally introduce spotlighting, so that the individual perceives the kink as a notch. In the absence of intensive margin frictions, they will adjust to just below z^* instead of exiting employment—that is, they will left bunch—because they continue to prefer intensive adjustment to exit. However, with intensive margin frictions, adjustment to the z^* is impossible or costly. And they now perceive a first order cost of not adjusting, because a perceived notch reduces consumption by $z^*(\tau_1 - \tau_))$. Therefore there is a positive probability that they prefer to exit employment than to make a costly adjustment or to remain at their baseline earnings.

Thus spotlighting and intensive margin frictions combine to generate both left bunching and extensive margin responses, even among those with desired earnings near z^*. Other models of left bunching do not generate these local extensive margin responses, because they imply that the welfare cost of the kink is continuous in distance to z^*, and so extensive margin adjustment is not optimal.

[6]The material presented here on extensive margin responses to notched budget sets was originally presented in Gelber et al. (2017), but has not been published previously. Gelber et al. (2021) considers extensive margin responses to *kinked* budget sets, and Kleven and Waseem (2013) considers *intensive* margin responses to notched budget sets.

[7]The probability here and in the following paragraphs is with respect to the fixed cost of working. When we say probability zero here, we mean there is zero probability that their fixed working cost is exactly equal to their utility from working.

4 Data

We use data derived from the IRS population files. These collections of tax returns go back to 1999 and contain all information returns as well as income tax returns. We supplement the IRS files with birth and death information from Social Security's Death Master File. We start with a 100 percent extract of all people with a social security number, born between 1939 and 1953, limited to ages 60-64, with at least one year of positive earnings (necessary for defining our running variable). We focus on 1939-1953 birth cohorts because these are the cohorts for which we can construct a balanced panel of earnings between ages 60 and 64, i.e two years before and after age 62, when claiming Social Security is first possible.

The resulting sample, which we call the full sample, consists of 58 million people. In our main analyses, we limit the sample to people with no self-employment income at age 61. We limit the sample this way because self-employed people can more easily manipulate reported income, and we want to avoid confounding our estimates of labor supply behavior with income reporting behavior. The resulting main sample consists of 53 million people.

We construct several variables from these data, including age in each year as of December 31 and an indicator for being identified as female in the the Death Master File. We also observe several key outcomes related to retirement. In addition, we have several measures of annual income. Our primary measure is total W2 income, as reported across all employers. We measure self-employment earnings as total Medicare-taxable earnings reported on Schedule SE. We observe Old Age and Survivor's Insurance (OASI) income, and (separately) Social Security Disability Insurance (DI) income.

We emphasize several appealing features of our data. First, our earnings measures are administratively collected and not subject to reporting error as survey data would be. Second, our primary earnings measure is simply W2 income (not including self-employment income), which is reported by a third party and cannot be manipulated by claiming deductions. Variation in this measure therefore likely reflects true earnings variation and not reporting effects. Third, our sample consists of a balanced panel of people *ever* in the tax system. We do not drop observations with zero income, for example, nor do we condition on filing an information return. We therefore do not select a sample on the basis of reported income.

We report summary statistics on person-level variables in Table 1, and we report summary statistics of time-varying variables by age for the main sample in Table 2. Half the sample is identified as female, the average claim age is about 61.9,[8] and about a 6 percent of our main sample ever has self-employment income. Earnings and (especially) the probability of having any earnings fall with age, slowly from 57-61 and then rapidly from 62-66. DI income is relatively rare, with 10-12 percent of the sample having any. OASI income is rare until age 62 and then becomes common; a quarter of the sample has some OASI income at age 62 and 70 percent has some by age 66.

[8] It is possible for a person to have Social Security income at ages younger than 62 because they may be claiming as a surviving beneficiary.

5 Empirical Results

5.1 Left bunching is substantial

We begin our empirical analysis by documenting in Figure 2 left bunching at the Earnings Test Exempt Amount. The figure plots the distribution of income relative to the exempt amount, by age, for ages 60, 62, 63, and 64. As a comparison we overlay the age 61 earnings distribution in gray. At age 60, before anyone is eligible for OASI, the earnings distribution is smooth and no bunching is apparent. At ages 62, 63, and 64, there is a clear excess mass around the exempt amount. The distribution increases sharply beginning a few thousand dollars below the exempt amount, and then falls discontinuously above the exempt amount. There is clearly more mass just to the left of the exempt amount than just to the right. We note that, although the density drops off discontinuously above the exempt amount (especially at ages 63 and 64), there is otherwise not a strong trend in the density, and therefore it is unlikely that a downward sloping density drives the observed left bunching. Indeed the age 60 and 61 densities are actually slightly upward sloping over this range.

Quantifying left bunching—or right or total bunching—requires measuring excess mass relative to some counterfacutal density. Early approaches to measuring bunching calculated counterfatual densities by estimating a smooth polynomial using data away from the exempt amount, and interpolating near the exempt amount (Chetty et al., 2011*a*; Kleven and Waseem, 2011; Kleven, 2016). We show such a smooth fit as the black line in Figure 2. We exclude observations within $4,000 of the exempt amount. Given this counterfactual earnings distribution, it is straightforward to calculate left and right bunching as the actual fraction of observations in the excluded region (below or above the threshold), less the predicted value from the smooth fit. We report left and right bunching in Table 3. We estimate essentially zero left bunching at age 60, 1.8 percent at age 62, and over 4 percent at ages 63 and 64. We estimate zero or negative right bunching at all ages. Left bunching is statistically significant at ages 62-64.[9]

Although it is standard in the literature to estimate a counterfactual density using polynomial interpolation, recent research has cautioned against it. Blomquist, Bartolino and Waldo (2019) point out that any interpolation or extrapolation ends up identifying bunching in part form functional form assumptions, and in general bunching amounts are not well-identified from a single cross-section. As an alternative counterfactual, we therefore use the age 61 income distribution. We plot this distribution as gray circles in Figure 2, and we report left and right bunching relative to this counterfactual in the last two columns of Table 3. We continue to find substantial left bunching at ages 62-64. We acknowledge that this counterfactual is only valid if underlying ability distributions do not change too much from year to year and if there are no extensive margin responses to the Earnings Tests - both assumptions that are likely to be violated in this context, at least to some extent. Nonetheless we view this as a useful check on the functional form assumptions implicit in our polynomial fits. Overall, we find visually clear evidence that left bunching

[9] We estimate standard errors using the bootstrap procedure of Chetty et al. (2011*b*), resampling residuals.

is substantial, especially relative to right bunching. Quantitatively, we estimate large amounts of left bunching and small amounts of right bunching across a range of possible counterfactual densities.

5.2 Left bunchers come from far above the exempt amount

The evidence so far shows substantial left bunching, in the sense that the excess mass just below the exempt amount is much larger than the excess mass just above it. This left bunching is surprising from the perspective of standard models, which do not predict asymmetric bunching. Our interpretation is that this left bunching reflects misperceptions: some people confuse the kink for a notch. There are however simple, innocuous interpretations of left bunching. Perhaps people with income just above the exempt amount are reducing their incomes very slightly to end up just below it. Or perhaps lumpy adjustment causes people to over adjust and end up below the exempt amount rather than exactly it at. We rule out these interpretations by showing that left bunchers overwhelmingly come from far above the exempt amount, i.e. their age 61 earnings are quite high. This is inconsistent with small or lumpy adjustment, since small adjustment could not cause high earners to bunch at all, and lumpy adjustment of high earners should not bias bunching towards the left. We show this in two ways. First, we look at the distribution of age 64 earnings conditional on different levels of age 61 earnings, which shows left bunching is common among high earners. Second, we show age 61 earnings as a function of age 64 earnings, which shows where bunchers typically come from.

We show in Figure 3 the distribution of income relative to the exempt amount at ages 62-64, among people with high age 61 earnings. Specifically, we condition on having age 61 earnings at least $30,000 above the exempt amount. Relative to all earners, this group has much higher modal earnings, unsurprisingly. These high earners also exhibit bunching—a slight amount at age 62, and then more at ages 63 and 64. Perhaps surprisingly, this bunching is highly asymmetric: there is a clear excess mass just below the exempt amount, and much less mass below the exempt amount. Looking closely, we can see that at ages 63 and 64, the bins just below the exempt amount are slightly elevated relative to higher levels. Thus, the figure shows small amounts of right bunching and large amounts of left bunching. By construction, this left bunching comes exclusively from people with high earnings at age 61, and therefore is unlikely to be driven by lumpy adjustment. Figure 3 focuses on very high earners, similar patterns are evident at intermediate age 61 earnings amounts, as we show in Appendix Figures A.1-A.3.

Thus, left bunching is typical among high earners, measured by age 61 earnings. Next, we show the converse: left bunchers typically have high age 61 earnings, relative to right bunchers. We show this graphically in Figure 4. The figures plot average 61 earnings (relative to the exempt amount), given age 62, 63, 64, or pooled 63-64 earnings (relative to the exempt amount). As earnings are highly autocorrelated, we expect a strong relationship between these variables. Indeed, there is a clear, positive, and roughly linear relationship on either side of the exempt amount. There is also, however, a clear downward discontinuity in age 61 earnings as age 63 or 64 earnings cross the exempt amount. We do not see this discontinuity at age 62, but this reflects the fact that there is relatively little OASI claiming or bunching at age 62.

To quantify the discontinuity in age 61 earnings, we estimate linear regression discontinuity models of the following form:

$$Distance_{i,61} = \beta_0 + \beta_1 Distance_{i,a} + \beta_2 above_{i,a} + \beta_3 Distance_{i,a} \times above_{i,a} + \beta_4 AtExemptAmount_{i,a} + \varepsilon_{i,a}. \quad (1)$$

where $Distance_{i,a}$ is person i's earnings at age a relative to the exempt amount, $above_{i,a}$ is an indicator for $distance_{i,a} > 0$, and $AtExemptAmount_{i,a}$ is an indicator for having earnings almost exactly at the kink, specifically within either \$50 or \$100 of it. This is a standard regression discontinuity model, except that we include the $AtExemptAmount$ indicator. The reason for this indicator is that people with earnings exactly at the exempt amount appear unlike others—they have very high age 61 earnings. (Some indication of this can be seen in Figure 4, as the bins immediately above and immediately below the exempt amount are a bit off the fit line.) Including the $AtExemptAmount$ control ensures that our RD estimates are not driven by these very sharp bunchers. However we show results that omit $AtExemptAmount$.

We estimate Equation (1) using the Calonico, Cattaneo and Titiunik (2014) optimal bandwidth (the "CCT-optimal bandwidth"). We estimate separate models for ages 62, 63, and 64. The results are in Table 4. There is no discontinuity in age 61 earnings as age 62 earnings cross the exempt amount. However, ages 64 and 64 there is a clear discontinuity: average age 61 earnings fall by about \$1,000 as age 63 earnings cross the exempt amount, and \$1,500 as age 64 earnings cross the exempt amount. This indicates that left bunchers have higher earnings than right bunchers, and hence adjust from farther above the exempt amount. Sharp bunchers have substantially higher earnings than either group, but our RD estimates are not sensitive to the inclusion of the $AtExemptAmount$ control.

5.3 Employment responses

The results so far show clear left bunching, with left bunchers often adjusting from high levels of earnings. These results are inconsistent with standard models of bunching, but they could be rationalized either with misperceptions (in particular, spotlighting), or with either "strong" loss aversion or loss aversion with diminishing sensitivity, as show in Section (3.3). The key to distinguishing these explanation is the extensive margin response: spotlighting implies that employment responds discontinuously to the perceived notch in the budget set, but loss aversion does not.[10]

We therefore investigate how employment at ages 62-64 varies with age 61 earnings. Figure 5 plots the probability of employment (i.e. positive earnings) at ages 62, 63, and 64, given age

[10]Standard models imply an employment response as well, but not a discontinuity, see Gelber et al. (2021), Gelber et al. (2022), and Appendix B.

61 distance to the exempt amount. The figure plots a range of earnings corresponding to the CCT-optimal bandwidth. The idea behind this plot, which we lay out in more detail in Gelber et al. (2021), is that earnings at age 61 are a good proxy for earnings at age 62 and beyond (in part because not everyone can easily adjust their earnings from on year to the next, Gelber, Jones and Sacks (2020)). People with age 61 earnings just above the exempt amount are "treated" in the sense that they are subject to the Earnings Test if they do not adjust their earnings. People with age 61 earnings just below the exempt amount likely have similar employment propensities but are "untreated." Thus, this plot can give a clean indication of the effect of the Earnings Test on employment among people with age 61 earnings close to the exempt amount. A kink and a discontinuity are both evident in the plot, although the discontinuity appears small.

We quantify the discontinuity by estimating regression discontinuity models of the following form:

$$1\left\{Earnings_{i,a} > 0\right\} = \beta_0 + \beta_D istance_{i,61} + \beta_2 above_{i,61} + \beta_3 Distance_{i,61} \times above_{i,61} + \varepsilon_{i,a}. \quad (2)$$

The outcome is an indicator for positive earnings at age a, and the running variable is age 61 earnings relative to the exempt amount. We estimate the model separately for ages 62, 63, 64, and 63-64 pooled, and we use the CCT-optimal bandwidth. We do not expect employment responses at age 62 because our employment outcome is an indicator for having no earnings for the entire year, and whole-year responses would not manifest until people have claimed OASI for an entire year, which is not possible at age 62. In general, we would expect the discontinuity to grow with age as OASI claiming increases.

At age 62, we estimate an insignificant discontinuity of -0.2 percentage points. This grows to -0.4 percentage points at ages 63 and 64. Only this last estimate is statistically significant. When we pool ages 63-64 (the ages at which we would expect to see an employment response), we estimate a statistically significant employment discontinuity of -0.5 percentage points. Thus, we find statistically significant discontinuities in employment. We show graphical evidence of employment responses in Figure 5. Consistent with fairly small discontinuities, the graphs show only slight drops in earnings above the exempt amount.

Although these responses are small, arguably any discontinuity here is surprising. This is for two reasons. First, the RD estimate compares people with age 61 earnings just above and just below the exempt amount. The "treated" group has exactly the same work incentives as the untreated group, so we would expect no employment response. Second, even under misperceptions, we should expect employment discontinuities only for people who cannot adjust on the intensive margin (Gelber et al., 2017), so the employment discontinuity is scaled down by the fraction of the population that both mispercieves and cannot adjust on the intensive margin.

6 Conclusion

Using administrative data on earnings, we perform the first systematic exploration of the puzzling left bunching phenomenon that has been noted across a number of economic contexts. In this context, we find that left bunching cannot be explained through standard reactions to incentives. Instead, we find that mis-perceptions of incentives explain the data well, in the cases of both the intensive and extensive margin responses. We have shown several patterns that are consistent with a model of spotlighting at this kink: the existence of left bunching, the observation of left bunching even with an earnings distribution that is nearly flat, the fact that left-bunchers are drawn from every part of the counterfactual earnings distribution and especially those with high counterfactual earnings, and the downward discontinuity in employment at the exempt amount.

Our findings have two key implications. First, for Social Security claimaints, the misperception of the Earnings Test can be a costly one. For one, the true policy features a notch with a discontinuous *increase* in Social Security benefits at the exempt amount equal to 0.55 percent of benefits after reaching Normal Retirement Age. Beneficiaries are therefore leaving substantial "money on the table" by locating just below the exempt amount. Thus, the spotlighting is a mis-perception of the true benefit schedule in two senses, first a kink in current benefits is misconstrued as a notch, and second, an actual upward notch in benefits is apparently overlooked by the majority of claimants who bunch. These mis-perceptions may affect estimates of elasticities based on bunching in response to the AET (Gelber, Jones and Sacks, 2020; Gelber et al., 2021, e.g.), but may be less of a concern for estimates that rely on more general variation in incentives away from the kink (Gelber et al., 2022, e.g.).

Second, to the extent that such misperceptions also explain left bunching in other contexts, this has wide-ranging implications for our understanding of elasticities as they have been estimated from bunching at kinks. As explained in Kleven (2016), the implied elasticity for a given amount of bunching in repsonse to a notch is much smaller than the same amount of bunching in the presence of a kink. These estimates are typically used to evaluate the efficiency costs of various taxes and implicit taxes, and also are used to derive optimal tax and transfer systems. It is therefore important that estimates apporpriately consider the underlying model of behavior. Such a model should involve several factors. First, it will be necessary to estimate the fraction of the population that mis-perceives the kink as a notch, and the complementary fraction that correctly perceives the kink. In principle, each of these two groups could exhibit different elasticities, which could also be estimated. Moreover, within the mis-perceiving group, some individuals may be inert to the notch as in Kleven and Waseem (2013), and this fraction could also be estimated. Developing a model that could credibly estimate all of these parameters – given enough empirical moments – will be a worthy challenge for future work. Another challenge for future work will be to explore the reasons why, as discussed above, left bunching appears clear in many circumstances, but bunching appears relatively symmetric in others.

References

Allen, Eric J, Patricia M Dechow, Devin G Pope, and George Wu. 2017. "Reference-dependent preferences: Evidence from marathon runners." *Management Science*, 63(6): 1657–1672. 7

Asatryan, Zareh, and Andreas Peichl. 2018. "Responses of Firms to Tax, Administrative and Accounting Rules: Evidence from Armenia." *ZEW-Centre for European Economic Research Discussion Paper*, , (16-065). 5

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. 2

Blomquist, Johan, Valerio Bartolino, and Staffan Waldo. 2019. "Price premiums for eco-labelled seafood: effects of the MSC certification suspension in the Baltic Sea cod fishery." *European Review of Agricultural Economics*. 12

Brown, Jeffrey R, Arie Kapteyn, Teryn Mattox, and Olivia S Mitchell. 2013. "Framing the Social Security Earnings Test." 3

Burtless, Gary, and Jerry A Hausman. 1978. "The effect of taxation on labor supply: Evaluating the Gary negative income tax experiment." *Journal of political Economy*, 86(6): 1103–1130. 2

Burtless, Gary, and Robert A Moffitt. 1985. "The joint choice of retirement age and postretirement hours of work." *Journal of Labor Economics*, 3(2): 209–236. 2, 3

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica*, 82(6): 2295–2326. 14

Chetty, Raj. 2012. "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply." *Econometrica*, 80(3): 969–1018. 8

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2011*a*. "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins." *American Economic Review*, 101(3): 471–75. 12

Chetty, Raj, John N Friedman, Tore Olsen, and Luigi Pistaferri. 2011*b*. "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records." *The quarterly journal of economics*, 126(2): 749–804. 2, 12, 42

Cogan, John F. 1980. "Fixed costs and labor supply." *Econometrica*, 49: 945–963. 36

Devereux, Michael, Jing Xing, and Giorgia Maffini. 2016. "The impact of investment incentives: evidence from UK corporation tax returns." 5

Diamond, Peter, and Jonathan Gruber. 1999. "Social security and retirement in the United States." In *Social security and retirement around the world*. 437–473. University of Chicago Press. 5

Dickens, William, and Shelly J Lundberg. 1985. "Hours restrictions and labor supply." 42

Einav, Liran, and Jonathan Levin. 2015. "Managed competition in health insurance." *Journal of the European Economic Association*, 13(6): 998–1021. 2

Eissa, Nada, Henrik Jacobsen Kleven, and Claus Thustrup Kreiner. 2008. "Evaluation of four tax reforms in the United States: Labor supply and welfare effects for single mothers." *Journal of Public Economics*, 92(3-4): 795–816. 36

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. 2

Friedberg, Leora. 1998. "The Social Security earnings test and labor supply of older men." *Tax policy and the economy*, 12: 121–150. 2

Friedberg, Leora. 2000. "The labor supply effects of the social security earnings test." *Review of Economics and Statistics*, 82(1): 48–63. 2

Friedberg, Leora, and Anthony Webb. 2009. "New Evidence on the Labor Supply Effects of the Social Security Earnings Test." *Tax Policy and the Economy*, 23(1): 1–36. 2

Gelber, Alexander, Damon Jones, Daniel W Sacks, and Jae Song. 2022. "The employment effects of the social security earnings test." *Journal of Human Resources*, 57(2): 341–371. 2, 5, 14, 16

Gelber, Alexander M., Damon Jones, and Daniel W. Sacks. 2020. "Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test." *American Economic Journal: Applied Economics*, 12(1): 1–31. 2, 3, 5, 8, 15, 16, 42

Gelber, Alexander M., Damon Jones, Daniel W. Sacks, and Jae Song. 2017. "Using Non-Linear Budget Sets to Estimate Extensive Margin Responses: Method and Evidence from the Social Security Earnings Test." National Bureau of Economic Research 23362. 10, 15, 35

Gelber, Alexander M., Damon Jones, Daniel W. Sacks, and Jae Song. 2021. "Using Nonlinear Budget Sets to Estimate Extensive Margin Responses: Method and Evidence from the Earnings Test." *American Economic Journal: Applied Economics*, 13(4): 150–93. 2, 5, 10, 14, 15, 16, 35

Gruber, Jonathan, and David Wise. 1999. "Social Security and Retirement around the World." 1–35. 5

Gruber, Jonathan, and Peter Orszag. 2003. "Does the social security earnings test affect labor supply and benefits receipt?" 3, 5

Hausman, I, JA Hausman, and D Holl. 1981. "Labor supply." Citeseer. 2, 35

Hendren, Nathaniel. 2016. "The policy elasticity." *Tax Policy and the Economy*, 30(1): 51–89. 4

Hoynes, Hilary. 1996. "Welfare transfers in two-parent families: labor supply and welfare participation under AFDC-UP." National Bureau of Economic Research. 2

Ito, Koichiro. 2014. "Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing." *American Economic Review*, 104(2): 537–63. 2, 4

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263–292. 7

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics*, 8: 435–464. 2, 5, 6, 9, 12, 16

Kleven, Henrik Jacobsen, and Claus Thustrup Kreiner. 2005. "Labor supply behavior and the design of tax and transfer policy." *Danish Journal of Economics*, 143(2005): 321–358. 36

Kleven, Henrik Jacobsen, and Mazhar Waseem. 2011. "Tax notches in pakistan: Tax evasion, real responses, and income shifting." *Quarterly Journal of Economics, Forthcoming*. 12

Kleven, Henrik J, and Mazhar Waseem. 2013. "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan." *The Quarterly Journal of Economics*, 128(2): 669–723. 2, 4, 8, 10, 16, 35, 36, 42

Le Barbanchon, Thomas. 2016. "Partial unemployment insurance." *Manuscript, Bocconi University*. 5

Liebman, Jeffrey B, and Erzo FP Luttmer. 2015. "Would people behave differently if they better understood social security? Evidence from a field experiment." *American Economic Journal: Economic Policy*, 7(1): 275–99. 3

Liebman, Jeffrey B, and Richard J Zeckhauser. 2004. "Schmeduling." 2, 3

O'Donoghue, Ted, and Charles Sprenger. 2018. "Reference-dependent preferences." In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 1, 1–77. Elsevier. 7

Peng, Langchuan, Xiaxin Wang, and Daixin He. 2019. "How Do Minimum Wage Adjustments Affect Wages in China: Evidence Based on Administrative Personal Income Tax Data." *Contemporary Economic Policy*, 37(2): 349–365. 5

Rees-Jones, Alex, and Dmitry Taubinsky. 2019. "Measuring schmeduling." 2

Rees-Jones, Alex, et al. 2018. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies*, 85(2): 1251–1278. 4, 7

Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology*, 66(5): 688. 35

Saez, Emmanuel. 1999. "Do taxpayers bunch at kink points?" National Bureau of Economic Research. 6, 8, 9

Saez, Emmanuel. 2010. "Do taxpayers bunch at kink points?" *American Economic Journal: Economic Policy*, 2(3): 180–212. 2, 6, 10, 35, 36

Seim, David. 2017. "Behavioral responses to wealth taxes: Evidence from Sweden." *American Economic Journal: Economic Policy*, 9(4): 395–421. 5

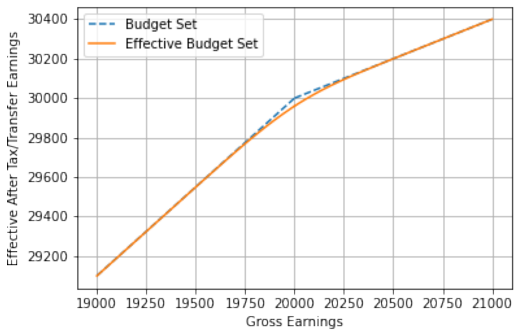
Social Security Administration. 2012. https://www.ssa.gov/OP_Home/handbook/handbook.07/handbook-0728.html. 5

Social Security Administration, Section 728.2. 2018. https://www.ssa.gov/OP_Home/handbook/handbook.07/handbook-0728.html. 3

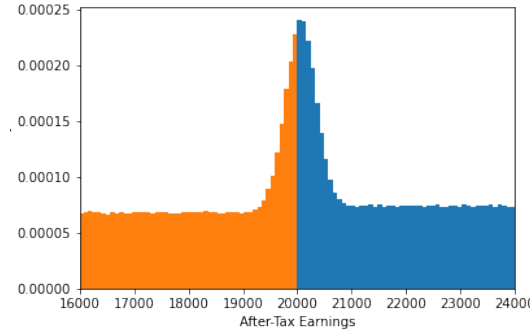
Song, Jae G, and Joyce Manchester. 2007. "New evidence on earnings and benefit claims following changes in the retirement earnings test in 2000." *Journal of Public Economics*, 91(3-4): 669–700. 2

Taubinsky, Dmitry, and Alex Rees-Jones. 2018. "Attention variation and welfare: theory and evidence from a tax salience experiment." *The Review of Economic Studies*, 85(4): 2462–2496. 4

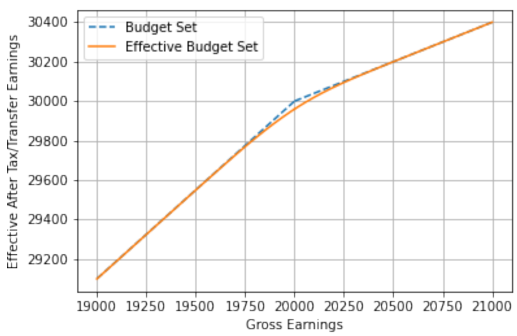
Figure 1: Models that do and do not predict left bunching



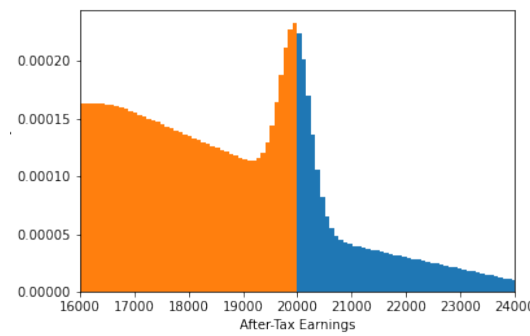
(a) Baseline



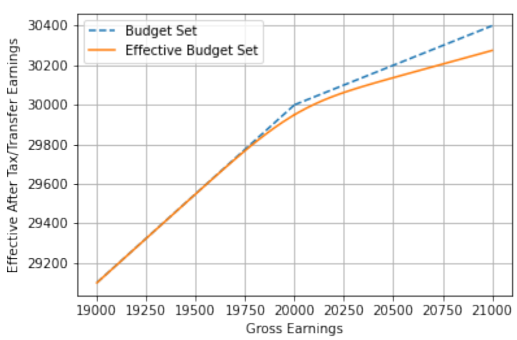
(b) Baseline



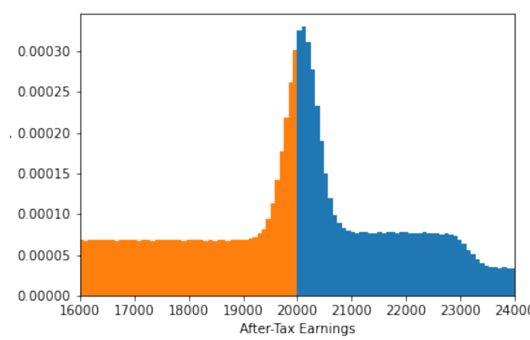
(c) Downward Sloping Density



(d) Downward Sloping Density



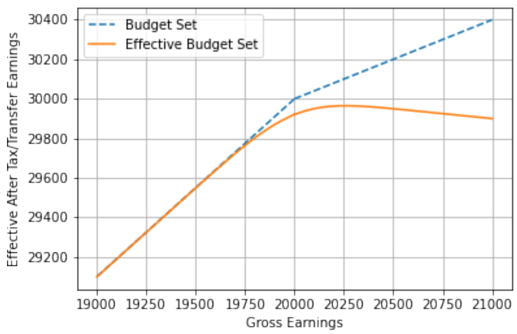
(e) Weak Loss Aversion



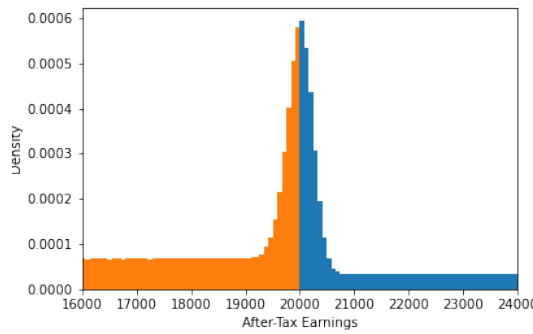
(f) Weak Loss Aversion

Notes: Figure plots the effective budget set (left side) and simulated earnings distribution (right side), for different models. All models include some uncertainty about exact earnings, generating diffuse bunching, and all models except "Downward Sloping Density." "Baseline" is otherwise standard. "Downward Sloping Density" has a downward sloping ability distribution. "Weak loss aversion incorporates weak loss aversion." Figure continues.

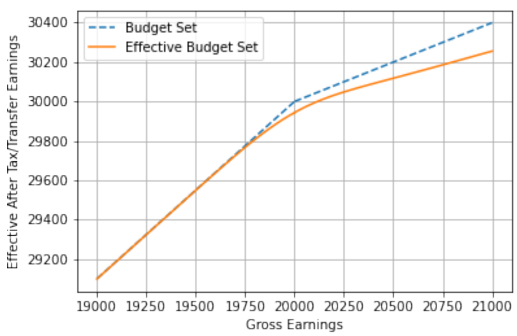
Figure 1: Models that do and do not predict left bunching (continued)



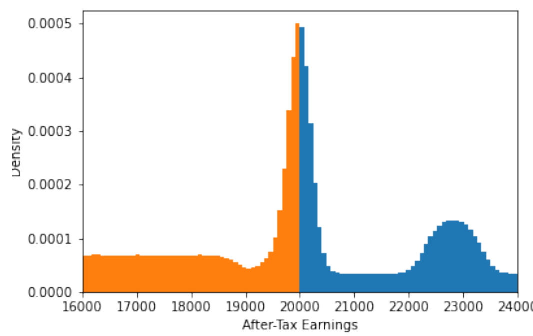
(g) Strong loss aversion



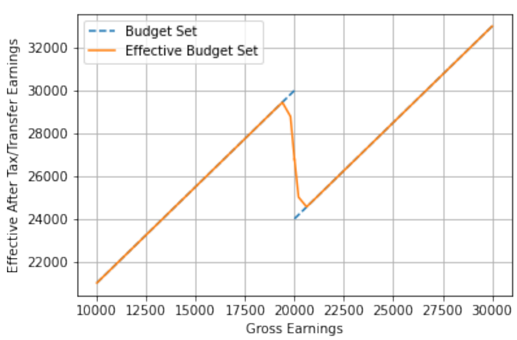
(h) Strong loss aversion



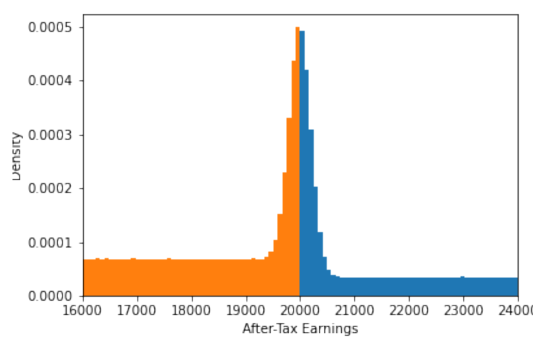
(i) Diminishing sensitivity



(j) Diminishing sensitivity



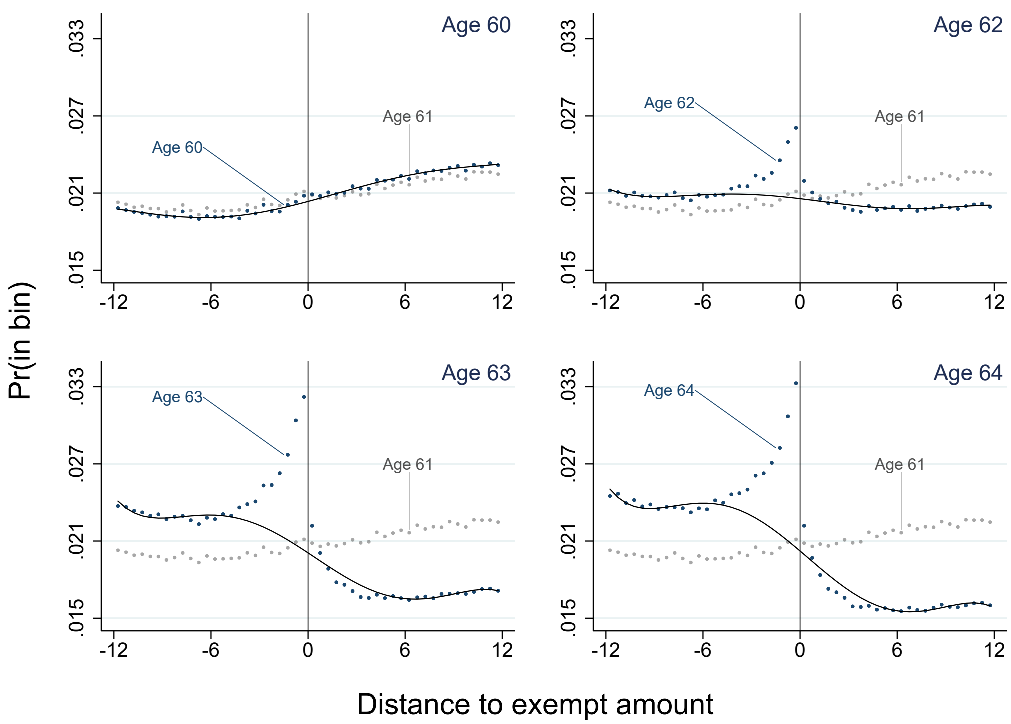
(k) Spotlighting



(l) Spotlighting

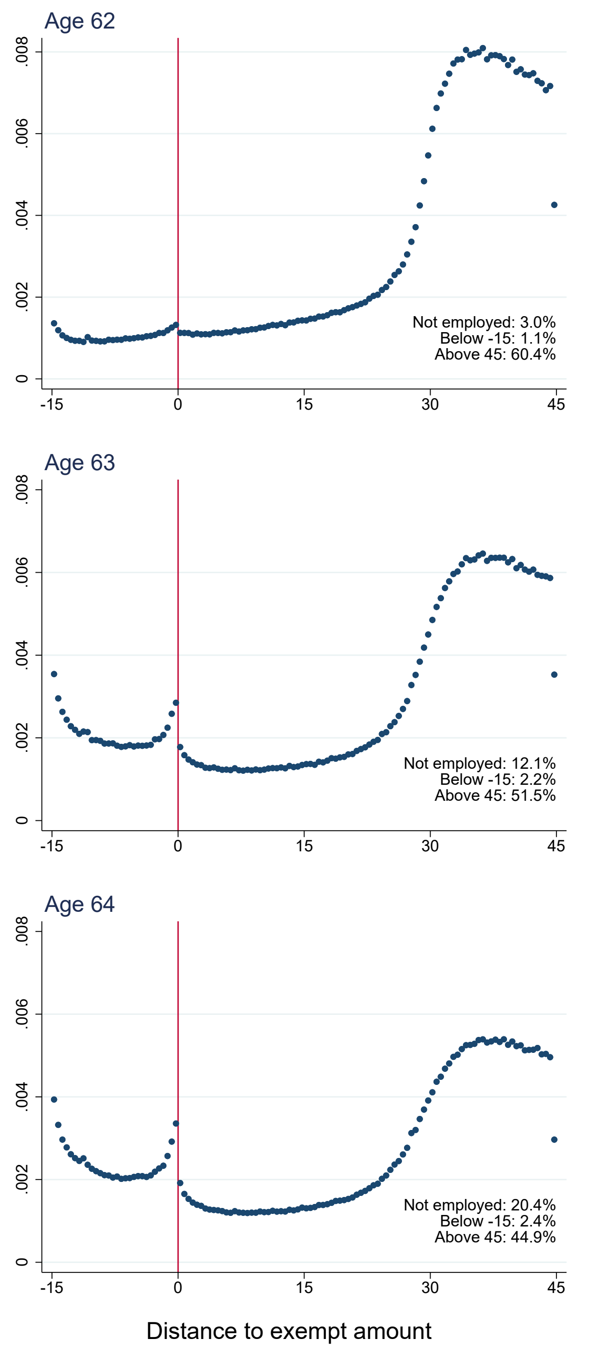
Notes: See notes from prior page. “Strong Loss Aversion" means the effective budget set slopes downward. “Diminishing Sensitivity" is loss aversion with diminishing sensitivity to gains and losses. “Spotlighting" means the kink is misperceived as a notch.

Figure 2: Distribution of earnings relative to exempt amount, by age



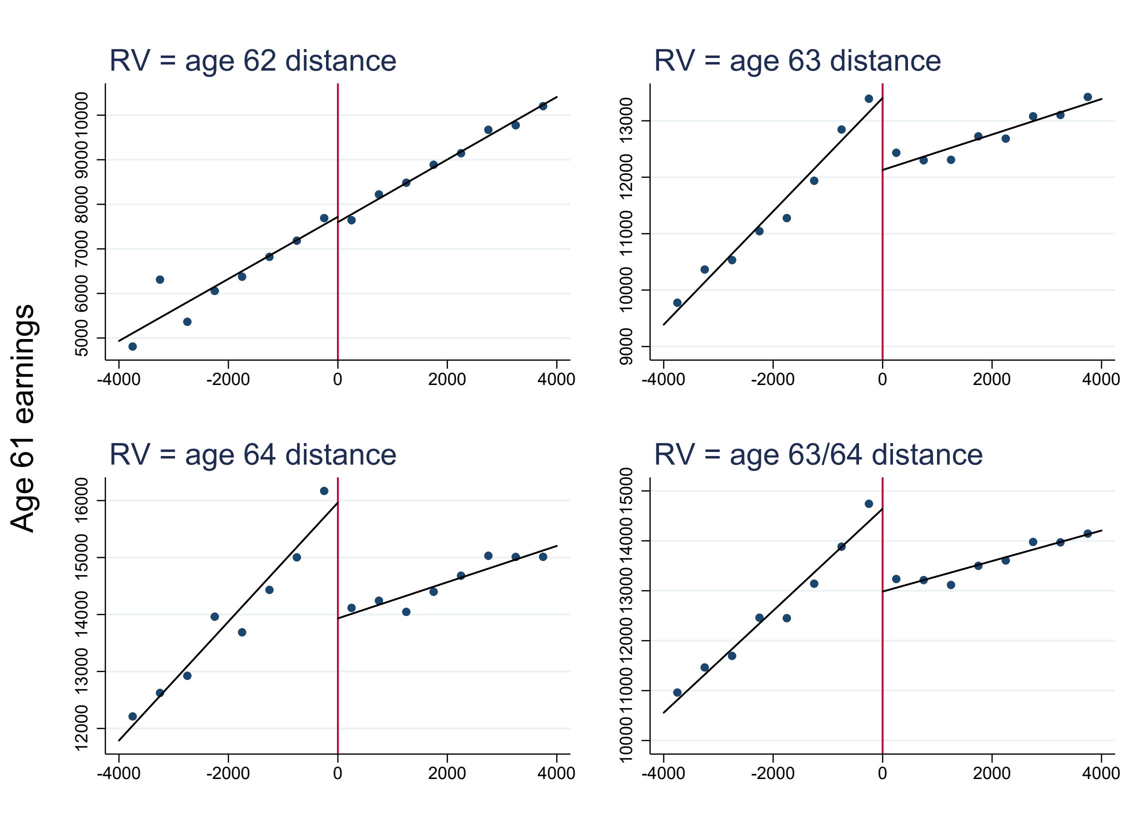
Notes: Figure plots the number of observations in each bin of earnings (relative to the exempt amount), by age. Sample is anyone born 1939-1954 with a Social Security number and in the tax system in 1999-2018, and no age 61 self-employment income. The gray dots show the number of observations at age 61. The smooth black line is a degree 5 fit estimated using all data from the indicated age except in the range (-4000, 4000).

Figure 3: Earnings distribution at ages 62-64 among people with high age 61 earnings



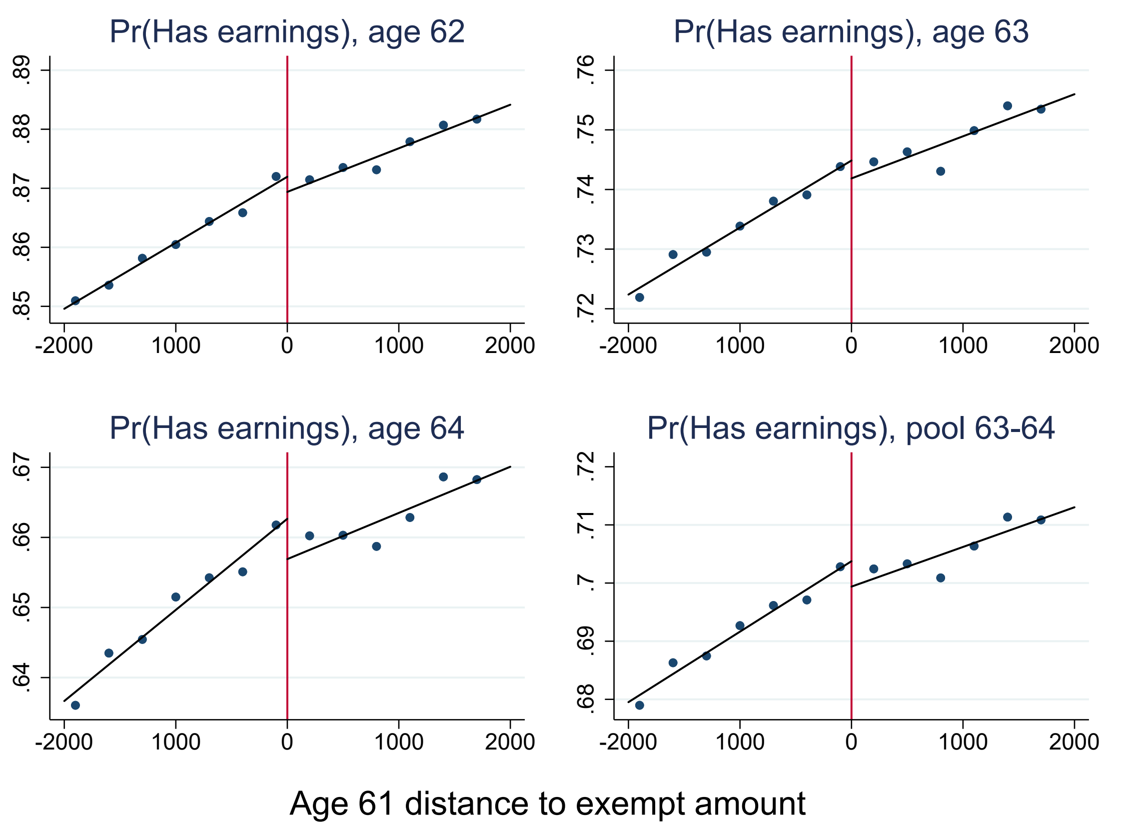
Notes: Figure plots the earnings distribution, at the indicated age, among people whose age 61 earnings was at least $30,000 greater than the exempt amount. Figure also reports the fraction of people with zero earnings or earnings outside the range. Sample is anyone born 1939-1954 with a Social Security number, and no age 61 self-employment income.

Figure 4: Left bunchers have higher earnings than right bunchers at age 61



Notes: Figure plots the mean age 61 earnings (relative to exempt amount) in each bin of earnings (relative to the exempt amount) number of observations in each cell of earnings, by age. Sample is anyone born 1939-1954 with a Social Security number, no age 61 self-employment income, and positive age 61 income.

Figure 5: Discontinuity in probability of having earnings



Notes: Figure plots the probability of having non-zero earnings at the indicated ages, given age 61 earnings relative to exempt amount. Sample is anyone born 1939-1954 with a Social Security number, and no age 61 self-employment income.

Table 1: Person-level summary statistics

Sample	Full	Main
Female	0.498	0.511
Year of birth	1947.4	1947.4
Claim age	61.9	61.9
Retirement age	60.0	60.0
Ever has SE earnings	0.141	0.061
N people	58,451,246	53,473,042

Notes: Full sample is anyone born 1939-1954 with a Social Security number, and at least one year of positive earnings age 60-64. Main sample conditions on no self-employment income at age 61. Claim age is the youngest age at which we observe non-zero OASI or DI income. Retire age is the youngest age at which we observe a disbursement from a tax-advantaged retirement account.

Table 2: Summary statistics by age, main sample

Age	60	61	62	63	64
Earnings (Medicare covered)					
Mean (if pos)	57200	56300	54500	52900	51800
25th percentile (if pos)	19700	18600	16000	13300	12300
50th percentile (if pos)	39500	38400	36200	34000	32400
75th percentile (if pos)	68500	67400	65300	63900	62800
Pr(any)	0.537	0.509	0.476	0.426	0.387
Exempt amount					
Mean (if pos)	15800	15900	16000	16200	16200
25th percentile (if pos)	15500	15700	15700	15700	15800
50th percentile (if pos)	15800	15800	16000	16300	16300
75th percentile (if pos)	16300	16300	16400	16400	16600
DI income					
Mean (if pos)	16000	16200	16000	16100	16100
25th percentile (if pos)	10400	10600	10300	10500	10700
50th percentile (if pos)	14700	14900	14900	15100	15200
75th percentile (if pos)	20500	20800	20900	21000	21100
Pr(any)	0.103	0.110	0.120	0.124	0.124
OASI income					
Mean (if pos)	8300	11300	6900	12300	12900
25th percentile (if pos)	3800	6600	2700	7800	8300
50th percentile (if pos)	7300	11000	5600	11400	12100
75th percentile (if pos)	11500	15700	9900	17300	18100
Pr(any)	0.029	0.037	0.254	0.353	0.392
N obs	53,473,042	53,473,042	53,473,042	53,473,042	53,473,042

Notes: Sample is anyone born 1939-1954 with a Social Security number, at least one year of positive earnings age 60-64, and no self-employment income at age 61. All dollar amounts are 2018 real and rounded to \$100 to avoid disclosing individual taxpayer incomes.

Table 3: Excess mass around the exempt amount, by age, main sample

Counterfactual	Smooth quintic		Age 61	
	Left	Right	Left	Right
	(1)	(2)	(3)	(4)
Age 60	0.001	0.001	-0.004	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Age 62	0.018	0.001	0.022	-0.003
	(0.001)	(0.001)	(0.005)	(0.005)
Age 63	0.043	-0.001	0.052	-0.020
	(0.001)	(0.001)	(0.015)	(0.013)
Age 64	0.045	-0.003	0.058	-0.024
	(0.002)	(0.001)	(0.015)	(0.016)

Notes: Sample is anyone born 1939-1954 with a Social Security number, at least one year of positive earnings age 60-64, and no self-employment income at age 61. Excess probabilities in a given bin are calculated as the actual bin probability less the counterfactual probability. Left and right excesses are the total excess amounts between [-4000, 0) and [0, 4000). We consider two counterfactuals: the fit from a smooth quintic (estimated using data from (-12000, 12000) but excluding the range (-4000, 4000)), and the age 61 earnings distribution. Bootstrapped standard errors (resampling residuals) in parentheses..

Table 4: RD estimates of discontinuity in age 61 earnings (relative to exempt amount) given age 62-64 earnings (relative to exempt amount

Age for running variable	62	62	62	63	63	63	64	64	64
Width for “at exempt amount""	Omit	50	100	Omit	50	100	Omit	50	100
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Above exempt amount	90.7	85.6	87.4	-871.6	-924.7	-902.4	-1307.5	-1383.6	-1356.2
	(272.5)	(270.7)	(271.2)	(118.3)	(118.5)	(118.5)	(122.3)	(123.4)	(122.6)
At exempt amount		373.5	362.8		2029.4	1347.7		2979.9	2129.7
		(172.7)	(185.6)		(132.2)	(113.1)		(240.5)	(170.9)
BW	2847.7	2847.7	2847.7	3160.0	3160.0	3160.0	3172.4	3172.4	3172.4
N	1,948,933	1,948,933	1,948,933	2,065,678	2,065,678	2,065,678	1,916,671	1,916,671	1,916,671

Notes: Sample is anyone born 1944-1951 with a Social Security number and in the tax system in 1999-2018, and no age 61 self-employment income, and positive age 61 income. Table reports estimates from a linear RD regression of age 61 earnings on distance to exempt amount at the indicated age, plus an indicator for “at exempt amount", defined as earnings within the indicated range of the exempt amount. Robust standard errors in parentheses. We use the CCT bandwidth.

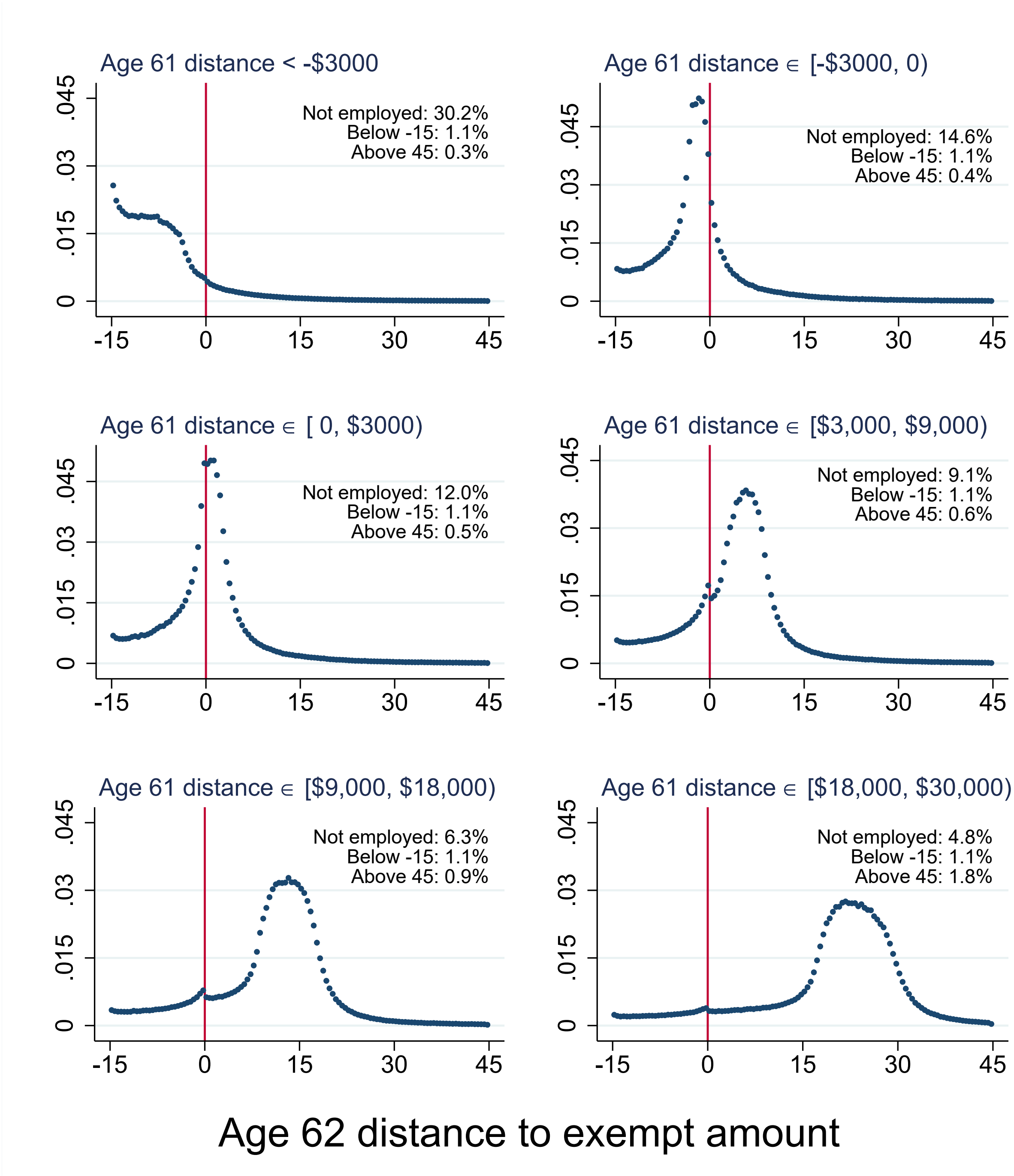
Table 5: RD estimates of discontinuity in employment at ages 62-64, given age 61 distance

Outcome age	62 (1)	63 (2)	64 (3)	Pool 63-64 (4)
Discontinuity	-0.0023	-0.0037	-0.0044	-0.0048
	(0.0012)	(0.0014)	(0.0016)	(0.0015)
BW	1,867	2,232	2,171	1,988
N	1,224,577	1,458,505	1,419,539	2,604,784

Notes: Sample is anyone born 1944-1951 with a Social Security number and in the tax system in 1999-2018, and no age 61 self-employment income. Table reports estimates from a linear RD regression of an indicator for positive earnings at the indicated outcome age on age 61 distance to the exempt amount. Robust standard errors in parentheses. We use the CCT bandwidth.

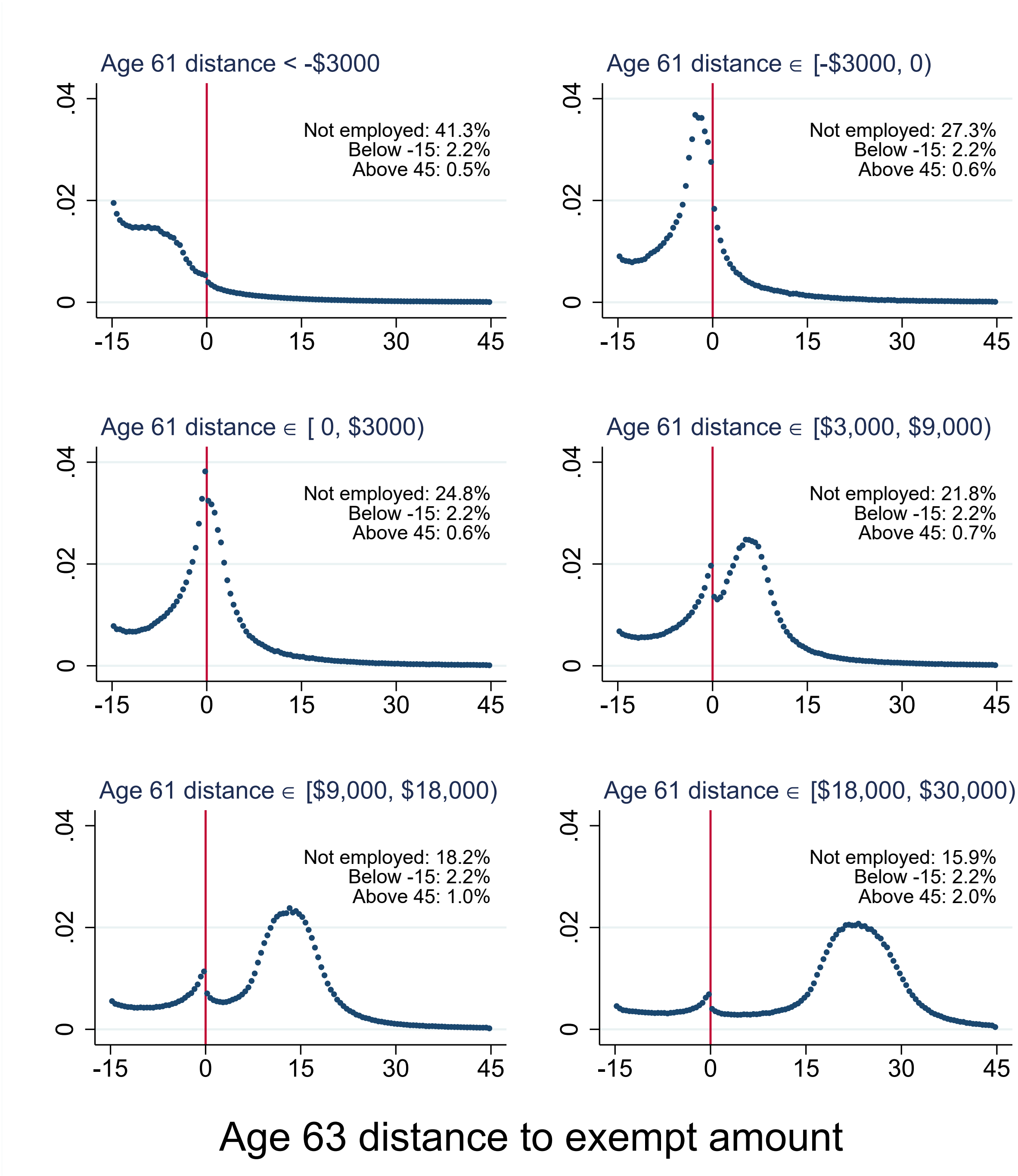
A Appendix Exhibits

Figure A.1: Even people with high age 61 earnings left bunch (age 62, sort of left bunching)



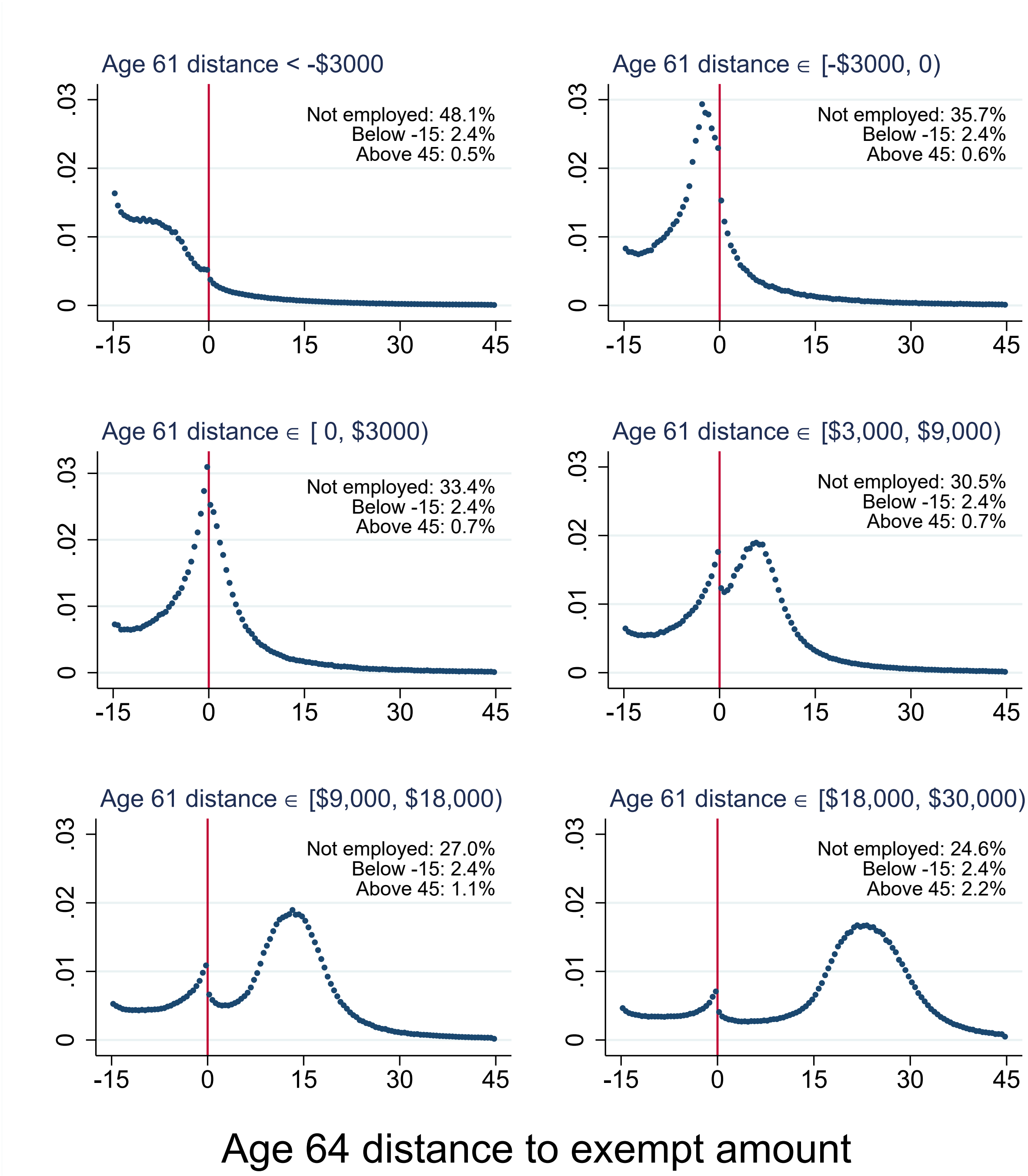
Notes: Figure plots the earnings distribution, at age 62, conditional on age 61 distance to exempt amount. The plotted figure (plus the reported amount) reflect the full set of people with age 61 distance in the indicated range. Sample is anyone born 1939-1954 with a Social Security number and in the tax system in 1999-2018, and no age 61 self-employment income.

Figure A.2: Even people with high age 61 earnings left bunch (age 63)



Notes: Figure plots the earnings distribution, at age 63, conditional on age 61 distance to exempt amount. The plotted figure (plus the reported amount) reflect the full set of people with age 61 distance in the indicated range. Sample is anyone born 1939-1954 with a Social Security number and in the tax system in 1999-2018, and no age 61 self-employment income.

Figure A.3: Even people with high age 61 earnings left bunch (age 64)



Notes: Figure plots the earnings distribution, at age 64, conditional on age 61 distance to exempt amount. The plotted figure (plus the reported amount) reflect the full set of people with age 61 distance in the indicated range. Sample is anyone born 1939-1954 with a Social Security number and in the tax system in 1999-2018, and no age 61 self-employment income.

B Modeling Extensive margin responses

We show that the combination of a (real or perceived) notch in the budget set, along with intensive margin adjustment frictions, results in a downward discontinuity in the extensive margin, i.e. a discontinuous drop in employment. A (real or perceived) kink, however, produces only a kink in the extensive margin.[11]

Throughout, we make use of a potential outcomes framework (Rubin, 1974). We index two potential states of the world by $j \in \{0,1\}$. In state 0 individuals face a linear tax τ_0, *i.e.* $T_0(z) = \tau_0 z$, where $T_j(\cdot)$ denotes their tax liability in state j and z denotes their pre-tax earnings. Alternatively, in the case of a kink, in state 1 the tax schedule exhibits a change in slope at z^*:

$$T_1(z) = \begin{cases} \tau_0 z, & \text{if } z < z^* \\ \tau_0 z^* + \tau_1 (z - z^*) & \text{if } z \geq z^* \end{cases} \tag{3}$$

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

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

$$c_{nj} = z_{nj} - T_j(z_{nj}) \tag{4}$$

[11] This material previously appeared in the working paper Gelber et al. (2017). The material on extensive margin responses to kinks was published in Gelber et al. (2021). The material on extensive margin responses to notches has not been previously published.

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

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

B.1 Intensive Margin Responses

Given this setup, we briefly review the intensive margin effect of a kink or notch. As shown in Saez (2010), a kinked budget set leads to a discontinuity in the earnings density at the kink due to intensive margin responses. Assuming a smooth distribution of ability n, a range of individuals who would earn between z^* and $z^* + \Delta z^*$ in state 0 will respond in state 1 by reducing earnings to the kink at z^*. This is referred to as "bunching" at the kink. The reduction in earnings Δz^* of the "marginal buncher"—*i.e.* the buncher who earns the most, $z^* + \Delta z^*$, in state 0—can be related to the size of the change in the marginal tax rate at the kink, $d\tau \equiv \tau_1 - \tau_0$, in order to estimate an intensive margin elasticity (Saez, 2010).

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

B.2 Extensive Margin Responses

In addition to the intensive margin response, individuals may also respond at the extensive margin. In the model of a kink briefly, preferences and budget sets are convex, which restricts a small tax change only to affect the choice between zero and infinitesimally small earnings supply (*e.g.* Kleven and Kreiner (2005)). To capture the realistic pattern of potential entry to or exit from non-trivial levels of earnings, we introduce a fixed cost of employment (Cogan, 1980; Eissa, Kleven and Kreiner, 2008). Utility conditional on working is now given by:

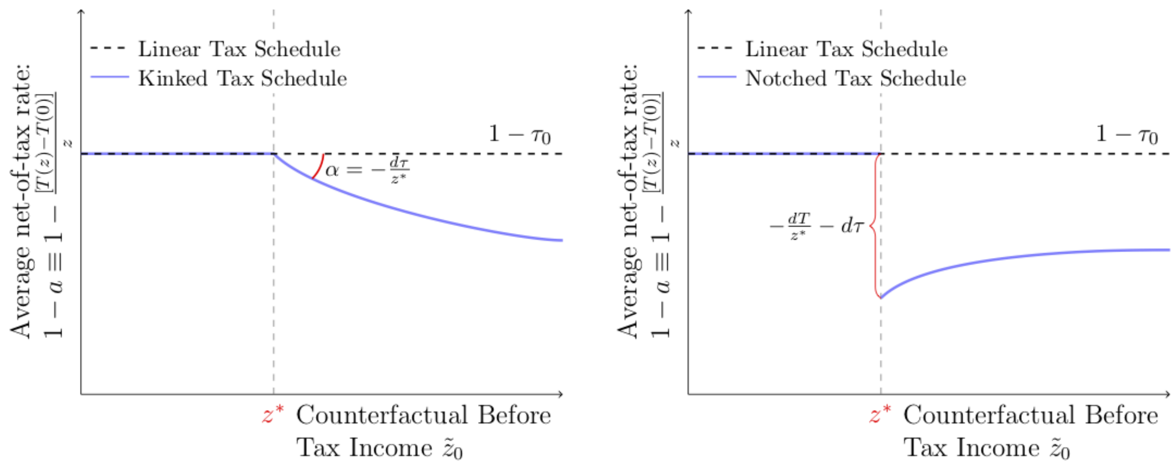
$$u\left(c_{nj}, z_{nj}; n\right) = v\left(c_{nj}, z_{nj}; n\right) - q_{nj} \cdot 1\left\{z_{nj} > 0\right\} \tag{5}$$

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

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

[14]Writing the fixed cost q as separable from v, as in (5) above, simplifies the exposition. Without loss of generality, this model is equivalent to a model in which these are not separable *per se*, and instead we express utility simply as $v(c, z; n)$.

Figure B.1: Extensive Margin Incentives



(a) Kinked Tax Schedule

(b) Notched Tax Schedule

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

the form of a fixed cost, as such frictions could be considered part of the fixed cost of working. We often refer to the fixed cost of employment as the fixed cost of working.

We pay special attention to whether or not the individual locates at a corner, *i.e.* $z_{nj} = 0$. Let $\tilde{z}_{nj}$ denote the optimal level of earnings in state j conditional on working. This is chosen by maximizing $u(c,z;n)$ subject to (4). The individual works in state j if:

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

Our key behavioral response of interest is the extensive margin response to the presence of a kink or notch. Here we define an individual's type by their optimal interior earnings conditional on working in state 0, *i.e.* $\tilde{z}_{n0}$. (We will often refer to $\tilde{z}_{n0}$ as "counterfactual" or "desired" earnings.) An isomorphism exists between this earnings amount and ability n, and for empirical purposes using an earnings amount is natural to implement. The probability of working in state j conditional on type $\tilde{z}_{n0}$ is:

Letting c_n^0 be consumption when not working, we can then posit a discontinuity in $v(c,z;n)$ at the boundary of the support of z that reflects the fixed cost. Thus, we can define a fixed cost q_n as: $q_n \equiv \lim_{z->0^+}\left[v\left(c_n^0, 0; n\right) - v\left(c_n^0, z, n\right)\right]$.

$$\Pr\left(z_{nj} > 0 \middle| \tilde{z}_{n0}\right) = \Pr\left(q_{nj} \leq v\left(\tilde{z}_{nj} - T_j\left(\tilde{z}_{nj}\right), \tilde{z}_{nj}; n\right) - v^0 \middle| \tilde{z}_{n0}\right) = G\left(\overline{q}_{nj} \middle| n\right) \tag{7}$$

where:

$$\overline{q}_{nj} \equiv v\left(\tilde{z}_{nj} - T_j\left(\tilde{z}_{nj}\right), \tilde{z}_{nj}; n\right) - v^0 \tag{8}$$

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

B.3 Modeling Incentives with a Kink or Notch

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

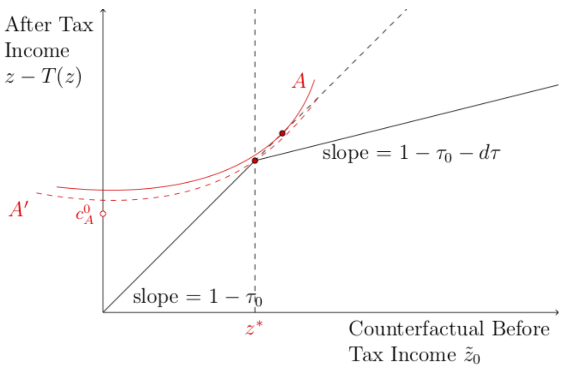
In Figure B.2 we illustrate the indifference curves governing the extensive margin decision under the alternative tax schedules. We model the fixed cost of working visually by allowing agents to choose a level of earnings along the prevailing tax schedule, or to earn zero and receive a level of consumption of c^0.[16]

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

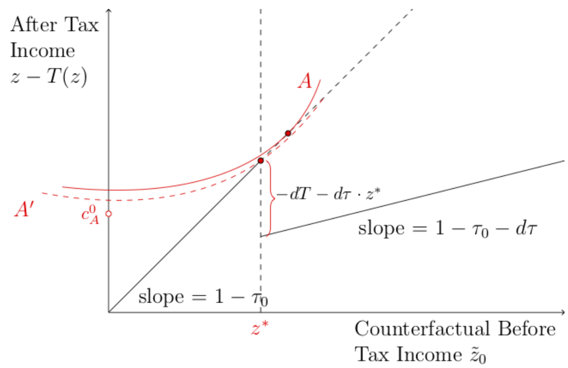
[15]If extensive margin responses were instead only driven by the value function $v(\cdot)$, we might generate unexpected predictions or restrictions on the employment function. For example, if $G(\cdot)$ were homogeneous in our model, we would necessarily require that labor force participation be upward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

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

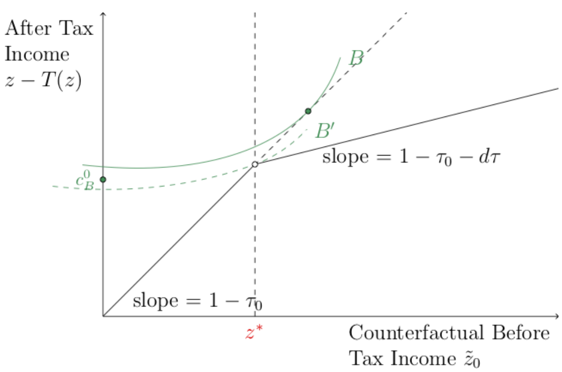
Figure B.2: Intensive and Extensive Margin Response to Kink or Notch



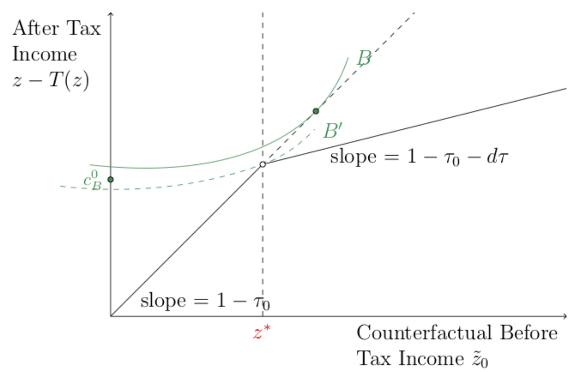
(a) Kink, No Exit



(b) Kink, Exit



(c) Notch, No Exit



(d) Notch, Exit

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

B.4 Employment Probability with a Kink and Unconstrained Intensive Margin Responses

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

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

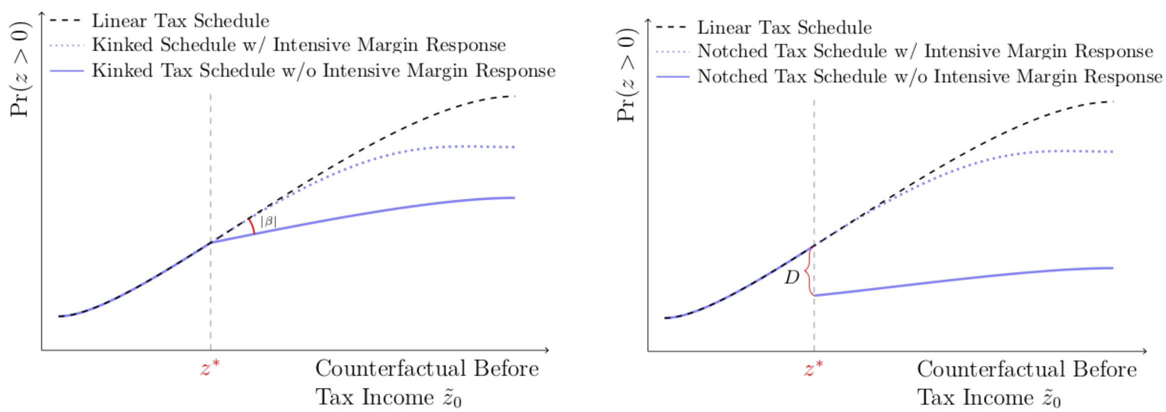
$$\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d\Pr(z_{n1} > 0 | \tilde{z}_{n0})}{d\tilde{z}_{n0}} = \lim_{\tilde{z}_{n0} \to z^{*-}} \frac{d\Pr(z_{n1} > 0 | \tilde{z}_{n0})}{d\tilde{z}_{n0}}. \tag{9}$$

Figure B.3A illustrates this result. The x-axis measures counterfactual earnings in state 0 conditional on having positive earnings, *i.e.* $\tilde{z}_{n0}$. The y-axis plots an illustrative employment rate. The dashed line represents a presumed smooth relationship between the employment rate in state 0 under a linear tax schedule, *i.e.* $\Pr(z_{n0} > 0 | \tilde{z}_{n0})$, and earnings conditional on having positive earnings, *i.e.* $\tilde{z}_{n0}$. The dotted line plots the employment rate in state 1 under a kinked tax schedule, $\Pr(z_{n1} > 0 | \tilde{z}_{n0})$, while the x-axis continues to plot $\tilde{z}_{n0}$. In this case, we assume that individuals are unrestricted in their earnings choices. We see that the employment function is unchanged at counterfactual earnings levels below z^*, as the tax schedule remains the same in the two states. Above z^* we see a gradual decrease in the probability of positive earnings in state 1 relative to state 0, due to the decrease in the ANTR (Figure B.1A). Nonetheless, the kink in the ANTR does not translate into a kink in the employment rate. Intuitively, the ability to adjust on the intensive margin smooths the first-order changes in the slope of the ANTR at z^*.

B.5 Employment Probability with a Kink and Constrained Intensive Margin Responses

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

Figure B.3: Extensive Margin Responses by State 0 Counterfactual Earnings



(a) Kinked Tax Schedule

(b) Notched Tax Schedule

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

of adjustment that would prevent adjustment to the kink for those in this region (*e.g.* Dickens and Lundberg (1985); Chetty et al. (2011*b*); Gelber, Jones and Sacks (2020)). Modeling and estimating frictions that could give rise to such restrictions is the focus of other work (Gelber, Jones and Sacks, 2020). Here we do not take a stand on what specific process gives rise to such restrictions, as the existence of such restrictions is sufficient to generate our results. In Appendix C, we prove the following result:

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

$$\lim_{\tilde{z}_{n0}\to z^{*+}} \frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} - \lim_{\tilde{z}_{n0}\to z^{*-}} \frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} = -d\tau\cdot\lambda_{n^*}\cdot g\left(\overline{q}_{n^*1}\middle| n^*\right) \tag{10}$$

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

Returning to Figure B.3A, the solid line depicts the relationship between counterfactual earnings in state 0 and the probability of positive earnings when a kink is present in state 1 and $\tilde{z}_{n1} \equiv \tilde{z}_{n0}$. The slope of the employment rate now discontinuously changes at z^*, where the ANTR also changes slope. We have closed one of the channels through which individuals respond to the increase in tax liability, and thus the slope of the employment rate decreases discontinuously at z^*. Equation (10) has an intuitive interpretation: the kink in the employment rate is proportional to $d\tau$, the size of the kink in the tax schedule; λ_{n^*}, the marginal utility of after-tax income; and $g\left(\overline{q}_{n^*1}\middle| n^*\right)$, the density of workers who are on the margin of entering employment in state 0. These parameters apply to the individual earning z^* in state 0. Because the kink in the employment rate we model is only detectable in the presence of frictions in intensive margin adjustment (whether due to fixed costs or constraints on earnings), our method can therefore also provide an incidental test of whether intensive margin frictions exist.

B.6 Extensive Margin Response to a Notch

We have shown that kinked budget set, with intensive margin frictions, produces a kink in the employment rate but not a discontinuity. We now show that the combination of notch and intensive margin frictions is sufficient and necessary for a downward discontinuity. As in the case of a kink, we first explore the setting in which individuals are free to adjust earnings on the intensive margin. We then contrast this with a setting with frictions on the intensive margin, as documented in Kleven and Waseem (2013). In Appendix C we prove the following:

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

$$\lim_{\tilde{z}_{n0}\to z^{*+}} \Pr(z_{n1}>0|\tilde{z}_{n0}) = \lim_{\tilde{z}_{n0}\to z^{*-}} \Pr(z_{n1}>0|\tilde{z}_{n0}). \tag{11}$$

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

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

This is illustrated in the dotted line in Figure B.3B. Its slope is continuous at z^*.

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

$$\lim_{\tilde{z}_{n0} \to z^{*+}} \Pr(z_{n1} > 0 | \tilde{z}_{n0}) - \lim_{\tilde{z}_{n0} \to z^{*-}} \Pr(z_{n1} > 0 | \tilde{z}_{n0}) = G\left(\bar{q}^{+}_{n^*1} \middle| n^*\right) - G\left(\bar{q}_{n^*0} \middle| n^*\right) \tag{13}$$

where the critical earnings levels are defined as

$$\begin{aligned}
\bar{q}_{n^*0} &\equiv v\left(z^* - T_0\left(z^*\right), z^* n^*\right) - v^0 \\
\bar{q}^{+}_{n^*1} &\equiv v\left(z^* - T^{+}_1\left(z^*\right), z^*; n^*\right) \\
&= v\left(z^* - T_0\left(z^*\right) - dT - d\tau \cdot z^*, z^*; n^*\right) \\
T^{+}_1\left(z^*\right) &\equiv \lim_{z \to z^{*+}} T_1(z).
\end{aligned}$$

Moreover, using a first-order approximation for the $G(\cdot)$ and $v(\cdot)$, we have:

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

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

This is illustrated in the solid line in Figure B.3B. This line shows a discontinuity in the level of employment at z^*. In the case of intensive margin frictions, the employment probability also features a discontinuity in slope, due to the discontinuity in the ANTR shown in Appendix Figure B.1b. The magnitude of this change in slope depends on the functional form of the utility function and the distribution of fixed costs. Even in the case of a pure notch, *i.e.* $dT > 0$ but $d\tau = 0$, with intensive margin frictions we will still have a change both the slope and level of the employment probability.

C Proofs

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

$$\frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} = g\left(\overline{q}_{n1}\middle|n\right)\frac{d\overline{q}_{n,1}}{d\tilde{z}_{n0}} + \frac{\partial G\left(\overline{q}_{n1}\middle|n\right)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} \tag{15}$$

Focusing on the first term in the expression for $d\Pr(z_{n1}>0|\tilde{z}_{n0})/d\tilde{z}_{n0}$ in (15), we have:

$$\begin{aligned}\frac{d\overline{q}_{n,1}}{d\tilde{z}_{n0}} &= \frac{\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)}{\partial\tilde{z}_{n1}}\frac{d\tilde{z}_{n1}}{d\tilde{z}_{n0}} + \frac{\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} \\ &= \frac{\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}}\end{aligned} \tag{16}$$

When agents are unrestricted in their intensive margin earnings choice, we can set the first term on the right side of (16) to zero. For those with $\tilde{z}_{n0}<z^*$ or $\tilde{z}_{n0}>z^*+\Delta z^*$, $\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)/\partial\tilde{z}_{n1}=0$ due to the envelope theorem.[17] For those with $z^*\leq\tilde{z}_{n0}\leq z^*+\Delta z^*$, we have $d\tilde{z}_{n1}/d\tilde{z}_{n0}=0$, since $\tilde{z}_{n1}=z^*$ for everyone in this set—*i.e.* these agents bunch at z^*. Substituting for $d\overline{q}_{n1}/dz_{n0}$ in (15) using (16), we have:

$$\frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} = g\left(\overline{q}_{n1}\middle|n\right)\frac{\partial v(\tilde{z}_{n1}-T_1(\tilde{z}_{n1}),\tilde{z}_{n1};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} + \frac{\partial G\left(\overline{q}_{n1}\middle|n\right)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} \tag{17}$$

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

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

$$\lim_{\tilde{z}_{n0}\to z^{*+}}\frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} = \lim_{\tilde{z}_{n0}\to z^{*-}}\frac{d\Pr(z_{n1}>0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} \tag{18}$$

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

Proposition 2: If $\tilde{z}_{n1}\equiv\tilde{z}_{n0}$, the general expression for $d\Pr(z_{n1}>0|\tilde{z}_{n0})/d\tilde{z}_{n0}$ from (15) still holds. However, we now have a slightly different expression for the critical level of fixed costs, which is now evaluated at $\tilde{z}_{n0}$, implying $\overline{q}_{n1}\equiv v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)-v^0$. Accordingly, we have a different expression for $d\overline{q}_{n1}/d\tilde{z}_{n0}$ relative to (16). Since $\tilde{z}_{n1}=\tilde{z}_{n0}$ for everyone, we have:

$$\frac{d\overline{q}_{n1}}{d\tilde{z}_{n0}} = \frac{\partial v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)}{\partial\tilde{z}_{n0}} + \frac{\partial v(\tilde{z}_{n0}-T_1(\tilde{z}_{n0}),\tilde{z}_{n0};n)}{\partial n}\frac{dn}{d\tilde{z}_{n0}} \tag{19}$$

where the key difference is that $\partial v/\partial z$ and $\partial v/\partial n$ are evaluated at $\tilde{z}_{n0}$ instead of $\tilde{z}_{n1}$. For those with

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

$\tilde{z}_{n0} < z^*$, since $T_1(\cdot) = T_0(\cdot)$ and $\tilde{z}_{n1} = \tilde{z}_{n0}$, it is still the case that $\partial v(\tilde{z}_{n0} - T_1(\tilde{z}_{n0}), \tilde{z}_{n0}; n)/\partial \tilde{z}_{n0} = 0$ due to the envelope theorem. However, the first term in (19) for those with $\tilde{z}_{n0} > z^*$ is now:

$$\begin{aligned} \frac{\partial v(\tilde{z}_{n0} - T_1(\tilde{z}_{n0}), \tilde{z}_{n0}; n)}{\partial \tilde{z}_{n0}} &= (1-\tau_1) v_c + v_z \\ &= \lambda_n \left[(1-\tau_1) + \frac{v_z}{v_c}\right] \end{aligned} \tag{20}$$

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

Thus, we now have:

$$\frac{d\Pr(z_{n1} > 0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} = \begin{cases} g(\overline{q}_{n1}|n) \frac{\partial v(\tilde{z}_{n0} - T_1(\tilde{z}_{n0}), \tilde{z}_{n0}; n)}{\partial n} \frac{dn}{d\tilde{z}_{n0}} + \frac{\partial G(\overline{q}_{n1}|n)}{\partial n} \frac{dn}{d\tilde{z}_{n0}}, & \text{if } \tilde{z}_{n0} < z^* \\ g(\overline{q}_{n1}|n) \left[\lambda_n \left[(1-\tau_1) + \frac{v_z}{v_c}\right] + \frac{\partial v(\tilde{z}_{n0} - T_1(\tilde{z}_{n0}), \tilde{z}_{n0}; n)}{\partial n} \frac{dn}{d\tilde{z}_{n0}}\right] + \frac{\partial G(\overline{q}_{n1}|n)}{\partial n} \frac{dn}{d\tilde{z}_{n0}}, & \text{if } \tilde{z}_{n0} \geq z^* \end{cases} \tag{21}$$

Note also that:

$$\lim_{\tilde{z}_{n0} \to z^{*+}} \frac{v_z(\tilde{z}_{n0} - T_1(\tilde{z}_{n0}), \tilde{z}_{n0}; n)}{v_c(\tilde{z}_{n0} - T_1(\tilde{z}_{n0}), \tilde{z}_{n0}; n)} = -(1-\tau_0) \tag{22}$$

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

$$\begin{aligned} \lim_{\tilde{z}_{n0} \to z^{*+}} \frac{d\Pr(z_{n1} > 0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} - \lim_{\tilde{z}_{n0} \to z^{*-}} \frac{d\Pr(z_{n1} > 0|\tilde{z}_{n0})}{d\tilde{z}_{n0}} &= \lim_{\tilde{z}_{n0} \to z^{*+}} g(\overline{q}_{n1}|n) \cdot \lambda_n \left[(1-\tau_1) + \frac{v_z}{v_c}\right] \\ &= g(\overline{q}_{n^*1}|n^*) \cdot \lambda_{n^*} [(1-\tau_1) - (1-\tau_0)] \\ &= -d\tau \cdot \lambda_{n^*} \cdot g(\overline{q}_{n^*1}|n^*) \end{aligned} \tag{23}$$

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

Proposition 3: To illustrate the first part of the result, note that $\tilde{z}_{n,1} = z^*$ for everyone with $\tilde{z}_{n,0} \in [z^*, z^* + \triangle z^N]$, i.e. the set of bunchers. Thus, as we cross z^*, there is no discontinuous change in after-tax income or optimal earnings, conditional on working, and thus no discontinuous change in the employment probability. To show the second result, our proof proceeds almost step-for-step as in the case of a kink. The key step of that proof is that the following term from (16) continues to vanish:

$$\frac{\partial v(\tilde{z}_{n1} - T_1(\tilde{z}_{n1}), \tilde{z}_{n1}; n)}{\partial \tilde{z}_{n1}} \frac{d\tilde{z}_{n1}}{d\tilde{z}_{n0}} = 0 \tag{24}$$

For those with $\tilde{z}_{n0} < z^*$ or $\tilde{z}_{n0} > z^* + \Delta z^N$, $\partial v(\tilde{z}_{n1} - T_1(\tilde{z}_{n1}), \tilde{z}_{n1}; n)/\partial \tilde{z}_{n1} = 0$ due to the envelope theorem.[18] For those with $z^* \leq \tilde{z}_{n0} \leq z^* + \Delta z^N$, we have $d\tilde{z}_{n1}/d\tilde{z}_{n0} = 0$, since $\tilde{z}_{n1} = z^*$ for everyone in this set—*i.e.* these agents bunch at z^*. The result then follows as in the case of a kink.

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

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

$$\begin{aligned} G\left(\bar{q}_{n^*1} \middle| n^*\right) - G\left(\bar{q}_{n^*0} \middle| n^*\right) &\approx g\left(\bar{q}_{n^*,0} \middle| n^*\right)\left[\bar{q}_{n^*1} - \bar{q}_{n^*0}\right] \\ &= g\left(\bar{q}_{n^*,0} \middle| n^*\right)\left[v\left(z^* - T_1^+\left(z^*\right), z^*; n^*\right) - v\left(z^* - T_0\left(z^*\right), z^*; n^*\right)\right] \end{aligned} \tag{25}$$

$$\begin{aligned} v\left(z^* - T_1^+\left(z^*\right), z^*; n^*\right) - v\left(z^* - T_0\left(z^*\right), z^*; n^*\right) &\approx \frac{\partial v}{\partial c}\left[z^* - T_1^+\left(z^*\right) - z^* + T_0\left(z^*\right)\right] \\ &= -\lambda_{n^*}\left[dT + d\tau \cdot z^*\right] \end{aligned} \tag{26}$$