American Economic Review 2021, 111(4): 1126–1165
https://doi.org/10.1257/aer.20191136

Reference Points for Retirement Behavior: Evidence from German Pension Discontinuities[†]

By Arthur Seibold*

This paper studies the large concentration of retirement behavior around statutory retirement ages, a puzzling stylized fact. To investigate this fact, I estimate bunching responses to 644 pension benefit discontinuities, using administrative data on the universe of German retirees. Financial incentives alone cannot explain retirement patterns, but there is a large direct effect of statutory retirement ages. I argue that the framing of statutory ages as reference points for retirement provides a plausible explanation. Simulations based on a model with reference dependence highlight that shifting statutory ages via pension reforms is an effective policy to influence retirement behavior. (*JEL* D91, H55, J26, J32)

For many countries, population aging poses looming questions over the fiscal sustainability of public pension systems. The average OECD country already spends 18 percent of total public expenditure on pensions, and the old-age dependency ratio is predicted to almost double by 2050 (OECD 2019). Extending individuals' working lives is an important margin of adjustment to these demographic trends. In standard economic models, retirement behavior can be influenced by appropriate financial incentives. In this paper, I present evidence that how retirement incentive schedules are perceived by workers matters, and this can have larger effects on retirement behavior than the incentives themselves.

In particular, I analyze the role of saliently featured age thresholds that I term *statutory retirement ages*. These are used by pension systems to frame retirement rules and they usually include an early retirement age and a full or normal retirement age. Panel A of Figure 1 shows that job exits of German workers are strongly concentrated around statutory retirement ages. There are sharp spikes in the distribution

* Department of Economics, University of Mannheim (email: seibold@uni-mannheim.de). Stefano DellaVigna was the coeditor for this article. I thank advisors Henrik Kleven and Camille Landais for their continued guidance and support. I am grateful to Michel Azulai, Youssef Benzarti, David Card, Erik Eyster, Xavier Jaravel, Dana Kassem, Felix Koenig, Erzo F. P. Luttmer, Panos Mavrokonstantis, Guy Michaels, Joana Naritomi, Jan Nimczik, Steve Pischke, Daniel Reck, Dominik Sachs, Emmanuel Saez, Johannes Spinnewijn, Stefan Staubli, three anonymous referees, and numerous seminar and conference participants for very helpful comments and discussions. I gratefully acknowledge the support of Tatjana Mika, Michael Stegmann, and their colleagues at the Research Data Centre of the German State Pension Fund, as well as financial support from LSE/STICERD, from the Daimler and Benz Foundation, and from the German Research Foundation (DFG) through CRC TR 224.
[†]Go to https://doi.org/10.1257/aer.20191136 to visit the article page for additional materials and author disclosure statement.

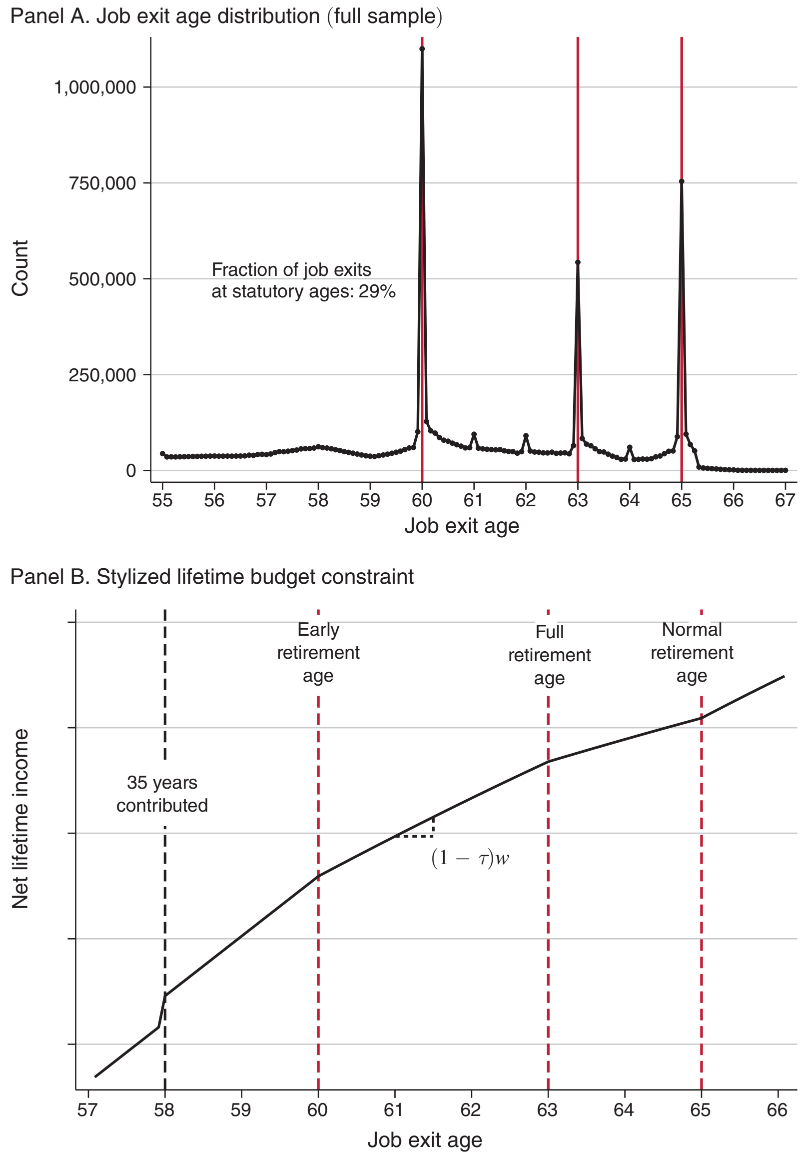


FIGURE 1. JOB EXIT AGE DISTRIBUTION AND LIFETIME BUDGET CONSTRAINT

Notes: Panel A shows the pooled distribution of retirement (job exit) ages for the full individual sample, i.e., for all workers born between 1933 and 1949. The connected dots show the count of job exits within monthly bins. Vertical red lines indicate the main locations of statutory retirement ages throughout the sample period. *Fraction of job exits at statutory ages* refers to the fraction of job exits at ages 55 to 67 that occur exactly in the month when the worker reaches a statutory retirement age. Panel B shows a stylized lifetime budget constraint for a worker who faces an early retirement age of 60, a full retirement age of 63, and an normal retirement age of 65, and who becomes eligible for a more generous pathway into retirement requiring 35 years of contributions at age 58. The slope of the budget constraint is the implicit net wage $(1-\tau)w$ (see Section IB). The stylized shape of the budget constraint corresponds to incentives faced by the average worker: On average, workers face a 22 percent reduction in the implicit net wage (i.e., a 22 percent kink size) at age 60, a 28 percent reduction at age 63, and a 32 percent increase in the implicit net wage at age 65.

at ages 60, 63, and 65, the main locations of statutory ages.[1] In total, 29 percent of job exits at age 55 and above occur precisely in the month when the worker reaches a statutory age. These retirement spikes are not only large, but also puzzling from the point of view of standard labor supply models. To preview this, consider the stylized lifetime budget constraint in panel B of Figure 1. Most workers face a reduction in the marginal return to work, i.e., an incentive to stop working, at ages 60 and 63, but a disincentive to retire at age 65. Nevertheless, large bunching occurs at all three ages.

I investigate this stylized fact and make three main contributions. First, I provide new, large-scale reduced-form evidence, building on Mastrobuoni (2009) and especially Behaghel and Blau (2012) who document bunching at the full retirement age in smaller samples using US survey data. I estimate bunching responses to more than 600 benefit discontinuities in the German public pension system, using administrative data on the universe of German retirees. I find that financial incentives alone cannot explain retirement patterns: on average, responses to statutory retirement ages are seven times larger than to pure financial incentives. These results suggest a first-order impact of nonstandard behavior on the retirement age distribution. Second, based on additional evidence, I argue that a parsimonious model with reference dependence fits the empirical patterns well. Third, counterfactual simulations suggest that shifting statutory ages is an effective policy tool to influence retirement behavior and such reforms can generate a positive fiscal impact.

As the empirical setting, the German public pension system provides several advantages. To begin with, there is rich variation in statutory retirement ages and financial incentives: there are six pathways into retirement entailing different statutory ages and benefit schedules, and a series of pension reforms provide additional cohort-based variation at the monthly level. This creates 644 discontinuities in pension benefits over the sample period, corresponding to kinks and notches in lifetime budget constraints. Discontinuities vary in the size of the local financial incentive, ranging from sizable incentives for retirement to disincentives. Moreover, some discontinuities, namely statutory retirement ages, are framed as reference points for retirement, while others are pure financial incentives. Statutory ages are linked to notions such as a "normal" time to retire, and a "full" level of pension benefits. Taken together, this independent variation allows me to disentangle responses to underlying financial incentives and the direct effect of presenting a threshold as a statutory age.

Another advantage of the setting is that high-quality administrative data are available to exploit this fine-grained variation. The analysis is based on a novel dataset provided by the German State Pension Fund, covering the universe of workers who retired between 1992 and 2014. The main sample contains 8.6 million individuals. The data include a rich set of worker characteristics related to earnings careers and pension eligibility, based on which monthly job exits and individual lifetime budget constraints can be calculated.

I divide the analysis in the paper into three parts. The first part of the paper uses bunching methods to estimate retirement responses to the 644 benefit

[1] Note that different statutory ages apply to workers depending on their birth cohort and characteristics such as gender and contribution histories.

discontinuities. I establish two main results. First, financial incentives alone fail to explain retirement patterns. There are large responses to statutory ages even if there is a close to zero or negative financial incentive to retire at the discontinuity. Second, presenting a threshold as a statutory retirement age directly affects retirement behavior. At all types of statutory ages and irrespectively of kink sizes, large additional bunching occurs compared to pure financial incentive discontinuities.

These results emerge from two complementary approaches. In the first approach, I focus on some cases of specific discontinuities that lend themselves to natural comparison. For instance, workers respond more strongly to a full retirement age kink than to a pure financial incentive kink of similar size occurring at the same retirement age. In the second approach, I use the full set of discontinuities to generalize the results. Expressed in terms of observed elasticities of the retirement age with respect to the net-of-tax rate, the average response to statutory ages is seven times larger than to pure financial incentives. I also propose a reduced-form strategy to formalize the joint estimation of responses to statutory ages and financial incentives, combining the large number of bunching estimates in a regression. The identification assumption is that responses to different types of discontinuities are driven by the same underlying parameters. The estimated direct effect of statutory ages is large and significant, and the "true" net-of-tax elasticity of around 0.05 is modest. Results are robust to controlling for heterogeneity in age, income, education, and other observable characteristics of workers facing different types of discontinuities.

The second part of the paper explores mechanisms behind the reduced-form effect of statutory ages. I begin by showing evidence from two reforms, suggesting that the effect is indeed due to the government setting statutory ages, and that the framing of statutory ages can affect retirement behavior. The first reform increases the full retirement age for women. Large bunching moves in lockstep with the statutory age while it is increased by one month for each month of birth over a five-year period. In addition, I exploit a second reform where the frequency of information letters sent to workers is substantially increased. After the reform, more workers retire at the normal retirement age around which explanations in letters are framed. Moreover, I discuss potential alternative mechanisms. On the one hand, firm responses do not seem to drive much of the results. For instance, self-employed workers and those in small firms below the employment protection threshold also bunch strongly at statutory ages. On the other hand, liquidity constraints could explain at least part of the response to the early retirement age, although this remains hard to verify directly in the absence of data on assets. Hence, the behavioral interpretation in the remainder of the paper focuses on full or normal retirement ages.

The third part of the paper turns to an interpretation of the empirical findings in a simple model of retirement with reference dependence. The reference point is given by a salient threshold in the form of a full or normal retirement age, for instance because workers perceive it as a normal time to retire. Reference dependence is modeled as a change in marginal disutility from continuing work at the reference point. Incorporating this standard formulation into a bunching framework yields predictions consistent with the empirical patterns, namely sharp bunching at statutory ages irrespectively of financial incentives.

Reference dependence may not be the only possible behavioral explanation for bunching at statutory ages, but I argue that it provides a parsimonious model which fits the data well. In particular, the shape of the empirical retirement age distribution around statutory ages is consistent with an asymmetric density shift as predicted by reference dependence with loss aversion. On the other hand, the empirical density does not exhibit missing mass in the neighborhood of statutory ages, as would be predicted by alternative models where individuals derive a fixed utility premium from retiring at this age. Such a discrete utility gain is one way to represent alternative behavioral mechanisms where individuals perceive retiring exactly at statutory ages as implicit advice by the government, or as a social norm. Moreover, if workers follow a suggestion by the government, one may expect that responses to statutory ages are concentrated among less financially sophisticated workers who find it difficult to make optimal retirement decisions. However, I do not find a negative relationship between bunching at statutory ages and proxies for financial literacy.

Based on the model, the magnitude of observed bunching can be directly related to parameters governing the strength of reference dependence, and these parameters can be straightforwardly recovered via structural bunching estimation. The estimation exploits the same variation in statutory ages and financial incentives across discontinuities used in the reduced-form analysis. Estimated local utility kinks at full and normal retirement ages are large and significant, with magnitudes equivalent to variation in the implicit tax rate of at least 51 percent.

Finally, counterfactual simulations highlight an important policy implication: reforms shifting statutory ages are effective in influencing retirement behavior and can generate a positive fiscal impact, which would be more difficult to achieve via financial incentives. First, I simulate an increase in the normal retirement age from 65 to 66. This leads to an increase in average actual retirement ages by 3 months. The second simulated reform provides stronger financial incentives for late retirement in the form of a "delayed retirement credit." In order to match the effect on retirement behavior from the first scenario, financial rewards would have to be almost doubled from their current level. Although both policies have the same effect on average retirement ages, the fiscal impact is very different: a back-of-the-envelope calculation suggests that the normal retirement age increase entails a long-term annual fiscal gain of €1.1 billion, whereas the financial rewards leads to a net fiscal loss of €1.0 billion. The difference in fiscal effects arises because workers pay contributions for longer in both scenarios, but in contrast to the second scenario, shifting statutory ages can induce workers to retire later without having to increase pension benefits at older retirement ages.

This paper contributes to three strands of literature. First, I contribute to the empirical literature on retirement behavior. A number of studies estimate the effects of pension reforms involving statutory retirement ages, but evidence on the *direct* effect of statutory ages is scarce. For instance, Staubli and Zweimüller (2013) and Manoli and Weber (2018) find sizable effects of an early retirement age increase using Austrian administrative data. Importantly, this type of reform simultaneously changes statutory ages and the financial incentives linked to them, such that the total reform effect is a mixture of the two.[2] Most closely related to this paper, Mastrobuoni (2009) documents sizable responses to a change in the full retirement

[2]Similar recent studies estimating the total effects of pension reforms involving statutory ages and financial incentives include Lalive and Staubli (2015); Cribb, Emmerson, and Tetlow (2016); and Fetter and Lockwood

age using US survey data, and Behaghel and Blau (2012) argue that loss aversion is a potential explanation for benefit claiming spikes.[3] In this paper, I leverage a unique setting combining rich, independent variation in statutory ages and financial incentives and full-population administrative data over two decades. I obtain compelling and precise estimates of the large direct effect of statutory retirement ages on job exit behavior, and provide new evidence on behavioral mechanisms. To my knowledge, this paper is the first to jointly quantify reference point effects and standard elasticities, which allows me to simulate and explicitly compare the effects of statutory age reforms to pure financial incentives.

Second, I contribute to a growing literature on the role of reference points in field settings. In particular, Allen et al. (2017) and Rees-Jones (2018) investigate bunching at reference points among marathon runners and income tax filers, respectively. Building on these approaches, I take the use of bunching methods further and estimate underlying reference dependence parameters by exploiting variation in financial incentives and statutory retirement ages across multiple discontinuities.[4] The bunching approach is complementary to full structural approaches such as DellaVigna et al. (2017) and Thakral and Tô (forthcoming). In addition, the results in this paper highlight the empirical relevance of salient thresholds as reference points,[5] and contribute to a broader literature on the importance of individuals' perception of incentives set by policy (e.g., Duflo et al. 2006).

Third, this paper builds on and contributes to the literature on bunching methods (Saez 2010, Chetty et al. 2011, Kleven 2016). Studies such as Kleven and Waseem (2013); Bastani and Selin (2014); and Gelber, Jones, and Sacks (2020) emphasize the importance of contextual factors in determining responses, mostly focusing on optimization frictions. Estimating bunching at many discontinuities, this paper shows that reference point effects can magnify bunching responses and highlights that bunching methods can be used to estimate related preference parameters.

The remainder of this paper is organized as follows. Section I outlines context and data, Section II describes the empirical methodology, Section III presents reduced-form evidence, Section IV discusses mechanisms behind the statutory age effect, Section V develops the conceptual framework, Section VI presents the estimation and counterfactuals, and finally, Section VII concludes.

(2018). Moreover, Brown (2013) and Manoli and Weber (2016) analyze retirement responses to pure financial incentives.

[3] In addition, some studies present survey and experimental evidence in favor of framing effects or reference dependence in intended retirement behavior, including Brown, Kapteyn, and Mitchell (2016); Merkle, Schreiber, and Weber (2017); and Shoven, Slavov, and Wise (2017).

[4] Existing bunching approaches including Allen et al. (2017) and Rees-Jones (2018) are unable to recover underlying reference dependence parameters, as suitable variation to estimate the curvature of the cost of effort function is not available in those settings (see DellaVigna 2018).

[5] This contributes to an ongoing debate about the relevance of salient/backward-looking reference points versus forward-looking reference points (see O'Donoghue and Sprenger 2018).

I. Context and Data

A. *The German Public Pension System*

Germany has a pay-as-you-go pension system that covers the vast majority of workers in the country (86 percent of the labor force in 2014). Enrollment is mandatory for private-sector employees, but most self-employed workers and civil servants are exempt. Contributions are levied as a payroll tax on gross earnings. Benefits are defined according to a pension formula based on a worker's lifetime contribution history.[6] Hence, pensions are roughly proportional to lifetime income and there is relatively little redistribution. The average net replacement rate is just over 50 percent (OECD 2019). Public pensions are the main source of income for most recipients.[7] Moreover, there is an earnings test for pension recipients where earnings above €450 per month lead to reductions in benefit payments. Only 2.5 percent of workers in the data have any income from employment while receiving a pension, making retirement an absorbing state for most.

The system features three types of statutory retirement ages. First, the *early retirement age* (*ERA*) is the earliest age from which a pension can be claimed. Second, the *full retirement age* (*FRA*) is the age from which workers can claim their *full pension*. Third, the *normal retirement age* (*NRA*) is the age from which workers can get more than their full pension.[8]

Discontinuous Benefit Rules: The key advantage of the empirical setting is that there are more than 600 pension discontinuities. Three types of discontinuous pension benefit rules are at their source. First, marginal pension adjustment changes at statutory retirement ages. A full monthly benefit level is defined at the FRA, and there are permanent benefit reductions for workers claiming before the FRA as well as permanent benefit increases for claiming after the NRA. The benefit adjustment function follows a kinked schedule, with a penalty of 0.3 percent for each month of retirement before the FRA, no adjustment between the FRA and the NRA, and a reward of 0.5 percent per month after the NRA.[9]

Second, workers become eligible for discontinuously higher pensions at some contribution thresholds, where they qualify for more generous pathways into retirement. Pathways are summarized in Table 1. The regular pathway requires just 5 years of contributions, but pensions can only be claimed from the NRA. At 15 and 35 years of contributions, workers become eligible for pathways with ERAs between 60 and 63, and FRAs between 60 and 65. Thus, they can receive a pension for more years

[6] Online Appendix Section B provides additional details on benefit calculation and other aspects of the institutional setting.

[7] In a 2003 survey, 11 percent of retirees reported to receive any income from employer pension schemes and only 1 percent had a private pension. Among retirees with any employer or private pension, the average income from that source corresponds to 34 percent and 23 percent of their public pension, respectively (Heien, Kortmann, and Schatz 2005).

[8] The distinction between the full and normal retirement age is somewhat peculiar to the German pension system. Historically, FRAs were introduced in some pathways to allow certain workers to claim a full pension before the NRA. However, late claiming rewards are only available after the NRA for all workers.

[9] In contrast to the United States, there is no discontinuity in the availability of public health insurance at statutory ages.

Table 1—Pathways into Retirement

Pathway	Required contributions	Other requirements	Statutory retirement ages (cohort 1941) Early	Full	Normal	Share of Sample (percent)
Regular	5 years	—	65	65	65	5
Long-term insured	35 years	—	63	65	65	19
Women	15 years 10 years full	Female	60	61	65	32
Unemployed/part-time	15 years 8 years full	Unemployed or in part-time work before retirement	60	64	65	20
Invalidity	35 years	Disability status	60	60	65	12
Disability	5 years 3 years full	Stricter disability status	—			11

Notes: The table presents an overview of pathways into retirement. For each pathway, statutory retirement ages are shown for a worker born in January 1941. Note that statutory ages vary over the sample period as shown in online Appendix Figure A2. The disability pathway does not have any statutory ages. For the unemployed/part-time pathway, unemployment for at least 1 year or old-age part-time work for at least 2 years after age 58 is required. For the invalidity pathway, an officially recognized disability of a certain degree is required; the disability pathway entails a stricter disability requirement, such that the worker is not able to work more than 3 hours a day in any job. *full* contribution years exclude periods where contributions are paid voluntarily. The last column shows the share of workers in each pathway in the full individual sample.

and/or the benefit level is higher at any given age due to more favorable adjustment, which implies a discontinuous increase in pension wealth. Some pathways have additional requirements including gender, disability, and periods of unemployment. Finally, the third type of discontinuous pension rule occurs in a pathway without statutory retirement ages where pensions can be claimed at any age. The disability pathway has a low contribution requirement of 5 years, but a relatively strict disability requirement. In this pathway, benefits are increased by 0.3 percent per month for retiring between 60 and 63, with no further adjustment when claiming before 60 or after 63.[10]

Framing of Statutory Retirement Ages: Statutory ages are one source of pension discontinuities, but the way they are presented to workers differs fundamentally from other, "pure" financial incentives. Figure 2 provides an example of the framing of statutory ages from a leaflet designed to inform workers about a future pension reform that increases the NRA to 67. First, statutory ages are saliently featured as normal retirement dates. The title *Retirement at 67* refers to the post-reform NRA. In fact, this title is a commonly used name for the reform in the media and public discourse. Using a hypothetical worker (Maria F.), readers are then told that if they want to retire "as early as possible" they can retire at the ERA, but if they wish a full pension, they should retire at the FRA. Furthermore, workers are warned of losses if they retire before the FRA ("the penalty will remain for her entire retirement").

The example illustrates how statutory ages are framed as reference points. By invoking notions such as a "normal" retirement age, statutory ages are presented as reference ages, and "early" and "late" retirement is defined relative to them.

[10] Moreover, contribution points are credited in the disability pathway as if the individual had continued working until age 60, making benefits less dependent on their contribution history.

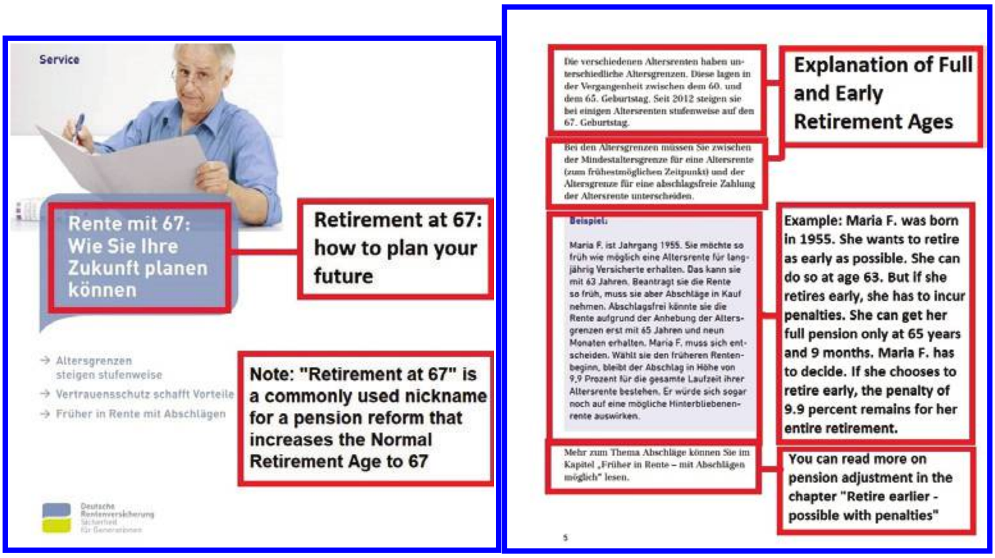


Figure 2. Framing

Notes: The figure shows excerpts from an information leaflet about a future pension reform. Explanation of the main points is provided in the red boxes on the right. The example in the right panel refers to early and full retirement ages in the long-term insured pathway. See online Appendix Figure A1 for full brochure, including similar examples from the other pathways. Note that some of the pension rules in the leaflet can differ from those described in Section I, as the leaflet describes a planned reform.

Source: Deutsche Rentenversicherung (2017)

Moreover, pension adjustment is framed as a loss (penalty) or gain (reward) relative to a "full" reference level linked to a statutory age.[11] Such framing of retirement ages has been shown to affect reported retirement plans in experimental settings (e.g., Brown, Kapteyn, and Mitchell 2013; Merkle, Schreiber, and Weber 2017).

B. *Lifetime Budget Constraint Discontinuities*

In order to see how the pension system affects incentives for the timing of retirement, the net present value of a worker i's net lifetime income can be written as a function of her retirement (job exit) age R_i:

$$NPV_i(R_i) = \sum_{t=0}^{R_i-1} \delta^t w_{it}(1-\tilde{\tau}_{it}) + \sum_{t=\max(R_i,ERA)}^{T_i} \delta^t B_i(R_i). \tag{1}$$

The worker earns a gross wage w from starting age 0 to the period before retirement, which is subject to payroll tax $\tilde{\tau}$. Pension benefits B depend on R both via contributions paid until retirement and via pension adjustment. Benefits can be claimed from the job exit age if the worker has already reached her ERA, and from the ERA

[11] More generally, statutory ages play a crucial role in the way pensions and retirement are presented to workers. For instance, pension reforms tend to be presented as changes to statutory ages rather than changes to benefit levels they might effectively entail.

otherwise, and are paid until time of death T, which is assumed to be known for simplicity. Finally, all payments are discounted at factor δ.

The slope of the budget constraint, that is the marginal gain in lifetime consumption possibilities C from delaying retirement by one period, defines the implicit net wage $w^{net} = dC/dR$. Expressing the consumption gain as a fraction of the gross wage, the *implicit net-of-tax rate* is $1 - \tau = w^{net}/w$. Delaying retirement generally affects consumption in three ways. First, the worker gains an additional period of wage earnings. Second, she sees a permanent change in her benefit eligibility dB/dR. In the German case dB/dR is always strictly positive, since later retirement implies both more favorable pension adjustment and a larger sum of contribution points. Third, if she is already eligible to claim benefits, there is an opportunity cost of work in terms of forgoing one period of benefits.

Panel B of Figure 1 shows a stylized version of the lifetime budget constraint. The discontinuous benefit rules described in Section IA introduce discontinuities into the budget constraint.

Kinks at Statutory Retirement Ages: There are kinks at all statutory ages, but their sign and magnitudes differ. Kinks at the ERA and the FRA are convex, i.e., the marginal net-of-tax rate is reduced. Moreover, there is a non-convex kink, i.e., an increase in the marginal return to work, at the NRA.[12] The kinks at the FRA and NRA are a direct consequence of discontinuous pension adjustment, where marginal adjustment decreases from 3.6 percent per year to 0 at the FRA and increases from 0 to 6 percent per year at the NRA. The kink at the ERA arises due to a combination of pension adjustment and an additional opportunity cost of working, since workers start forgoing benefits from the ERA onward.[13]

Pure Financial Incentive Discontinuities: Contribution Notches and Disability Kinks: There are two sources of *pure financial incentive discontinuities*. First, contribution requirements of different pathways create budget constraint discontinuities in the form of notches, i.e., jumps in the average net-of-tax rate. In panel B of Figure 1, for instance, the worker reaches 35 years of contributions when working until age 58, where he becomes eligible for the long-term insured pathway with higher implied pension wealth. Similarly, workers face notches in all main pathways where eligibility requires 5, 15, or 35 years of contributions.[14] Note that the age location of these notches is worker-specific since it depends on the individual career starting age. As a second source of pure financial incentive discontinuities, the kinks in the benefit schedule of the disability pathway imply budget constraint

[12] An exception is the regular pathway where pensions can only be claimed from the NRA, in which case there is a convex kink at the NRA.

[13] The ERA kink could be smoothed out by actuarially fair adjustment of pensions. However, the adjustment of 3.6 percent annually is less than actuarially fair (see Börsch-Supan and Wilke 2004).

[14] The notches at 5 years of contributions are not used in the analysis because the data on workers with less than 5 years of contributions are incomplete.

kinks, where the marginal net-of-tax rate changes due to changes in marginal pension adjustment.

C. *644 Discontinuities*

Two sources of variation generate more than 600 budget constraint discontinuities.[15] First, the six pathways described in Table 1 vary in statutory ages and contribution requirements. Second, a series of cohort-based pension reforms have been enacted since the early 1990s. Online Appendix Figure A2 shows the evolution of ERAs and FRAs for birth cohorts 1933 to 1949. In addition to cross-sectional variation, statutory ages were changed in different pathways at different times. For instance, the women's FRA was gradually increased from 60 to 65 for cohorts 1940 to 1944, such that each monthly birth cohort faces a one-month change in the FRA. Similar gradual changes to the FRA or ERA were also implemented in the other pathways.

In total, this yields 386 budget constraint kinks linked to statutory ages. Contribution notches and disability kinks amount to 258 pure financial incentive discontinuities. Combining variation across pathways, cohorts, and age groups yields a total of 180 contribution notches. Including a gradual introduction period, there are 78 disability pension kinks. To illustrate the variation, online Appendix Figure A3 provides some examples of lifetime budget constraints, where workers face different statutory ages and pure financial incentive discontinuities depending on their month of birth, gender, contribution history, and disability status.

Panel A of Table 3 summarizes the 644 budget constraint discontinuities. At statutory retirement ages, there is strong heterogeneity in underlying kink sizes. At ERAs and FRAs, the average kink size is between 0.22 and 0.25, i.e., the net-of-tax rate decreases between 22 percent and 25 percent at the threshold. On the other hand, NRAs feature sizable non-convex kinks of average size -0.50. The average change in the net-of-tax rate is 0.32 at pure financial incentive kinks, and 0.44 at contribution notches. The figure at notches is obtained from the approximation as a kink for the marginal buncher following Kleven and Waseem (2013). A further advantage of the setting is that there is substantial variation in kink sizes across discontinuities of a given type. For instance, the standard deviation of kink sizes across statutory ages is 0.39. There is also some within-group variation in effective kink sizes due to different individual earnings histories, but the within-group standard deviations are relatively small. The average retirement age at which statutory ages are located is 62.5, while pure financial incentive discontinuities occur at an average age of 60.4.

D. *Data*

The analysis is based on a novel administrative dataset covering the universe of retirees who claim a public pension between 1992 and 2014 provided by the German State Pension Fund (FDZ-RV 2015). The sample is limited to workers in the six main

[15] See online Appendix Section D.2 for the a complete list of all discontinuities used. This paper refers to both kinks and notches as budget constraint discontinuities. More precisely, kinks are discontinuities in the marginal net-of-tax rate, whereas notches are discontinuities in the average net-of-tax rate.

pathways who claim a pension for the first time between ages 55 and 67, have earned at least 5 contribution points from at least 5 years of contributions, and do not continue work after retirement. Moreover, East Germans retiring in 1995 and earlier are excluded since their pensions were calculated under a particular set of post-reunification rules. The analysis focuses on birth cohorts 1933 and 1949, for whom the relevant part of the retirement age distribution is fully observed. After applying those restrictions, the *individual sample* contains around 8.6 million observations.

The data include all variables necessary for the pension fund to determine a worker's pension eligibility as well as a number of socioeconomic characteristics. Monthly benefit claims and last contributions can be directly observed. The month of job exit can be inferred from the time of the last contribution for most of the sample. For those workers where the last contribution does not coincide with employment, the time of job exit is imputed using additional information on the insurance status in the last three years before retirement.[16] Lifetime earnings and average annual earnings are backed out using information on contribution periods and contribution points,[17] and a pension benefit simulator is built to calculate each individual's benefit eligibility across possible retirement ages. Lifetime budget constraints are simulated as a version of equation (1) with a 3 percent discount rate and heterogeneous life expectancies by year of birth and gender. In order to account for the fact that observed take-up of pathways may reflect workers' choices, pathways are assigned in terms of eligibility as far as possible.

In addition, survey data from the German Socioeconomic Panel are used for part of the analysis (SOEP 2015). SOEP is an unbalanced panel of around 1.4 million individual-year observations spanning the period 1984 to 2013. It contains a wide range of socioeconomic variables including labor market outcomes. Variables of interest are collapsed at the three-digit occupation level and merged with the main data where occupations can be observed from 2000 onward. This sample is referred to as the *occupation-matched sample*.

As explained in Section IC, pension discontinuities differ across pathways and cohorts. In practice, workers can be grouped by pathway and year of birth to capture this variation. Workers born during reform periods where policy varies at the monthly level are grouped by pathway and month of birth instead. The sample split yields 375 groups each of whom faces a distinct set of statutory ages and lifetime budget constraint discontinuities. When analyzing contribution notches, groups by pathway and year of birth are further divided into those retiring at ages 55 to 60 and 60 to 65 in order to capture variation of notch sizes with retirement age. For the analysis across discontinuities, bunching observations are collected in the *bunching sample*, where each of the 644 observations represents a discontinuity faced by a particular group of workers, and control variables are added at the group level. Table 2 shows summary statistics for the individual sample in column 1, for the occupation-matched sample in column 2 and for the bunching sample in column 3. The average job exit age is around 61, and the time between the first and last

[16] The imputation is mostly relevant for job exits before the ERA, and affects relatively few job exits at the different types of discontinuities. See online Appendix Section C for further details of the data and key variables.

[17] Contribution points are generally proportional to gross earnings. The only caveat is top-coding of earnings above the contributions cap.

Table 2—Summary Statistics

	Individual sample (1)	Occupation-matched sample (2)	Bunching sample (3)
Job exit age	60.88	61.89	61.11
	(2.80)	(2.67)	(1.54)
Benefit claiming age	62.05	62.80	62.39
	(2.33)	(2.12)	(1.41)
Career length	43.60	44.19	43.70
	(6.53)	(6.93)	(2.67)
Contribution points	37.03	39.02	37.07
	(17.24)	(18.07)	(11.37)
Net lifetime income	1,689,142	1,745,749	1,678,875
	(655,797)	(677,296)	(432,089)
Female	0.46	0.45	0.45
	(0.50)	(0.50)	(0.43)
Married	0.76	0.76	0.77
	(0.42)	(0.43)	(0.06)
Education (years)	10.61	10.74	10.68
	(1.59)	(1.79)	(0.30)
Sick leave (years)	0.07	0.06	0.07
	(0.25)	(0.21)	(0.04)
East Germany	0.18	0.20	0.18
	(0.38)	(0.40)	(0.09)
Small firm		0.27	
		(0.18)	
Large firm		0.44	
		(0.18)	
Tenure		8.95	
		(2.80)	
Unlimited contract		0.83	
		(0.09)	
Observations (individuals)	8,557,797	3,954,968	
Observations (discontinuities)			644

Notes: The table presents summary statistics for the samples used. The individual and occupation-matched samples are at the worker level, while the bunching sample is at the discontinuity level. Job exit and benefit claiming ages are in years. Career length is time between first and last contribution. Contribution points are collected from pension contributions, where one point corresponds to earning the population average gross income for one year. Net lifetime income is defined as in equation (1) and calculated in terms of net present value at age 65. *East Germany* is a dummy for residence in East Germany. *Small firm* and *Large firm* are indicators for firms with less than 20 employees and more than 200 employees, respectively. Firm size, tenure, and fraction in unlimited contract are at the occupation level. Standard deviations in parentheses.

contribution is around 44 years. Just below one-half of the sample are female and three-quarters are married. These and other key observables are relatively balanced across the different samples.

II. Empirical Methodology

A. *Basic Bunching Method*

The first step of the empirical analysis is to measure retirement responses at each discontinuity. The bunching method developed by Saez (2010) and Chetty et al.

(2011), which can be applied to the retirement age distribution,[18] provides a way of detecting such responses. The bunching mass B at an age threshold $\hat{R}$ is the observed local spike above a counterfactual retirement age density $h_0(\hat{R})$, which can be obtained by fitting a polynomial to the observed density excluding the threshold. The excess mass $b = B/h_0(\hat{R})$ is computed as the bunching mass relative to the counterfactual.

Assuming that the density would have been smooth in the absence of the threshold,[19] bunching can be interpreted in terms of a local retirement response. A standard approach focused on responses to financial incentives then computes an elasticity by relating the excess mass to the kink size $\Delta\tau/(1-\tau)$, defined as the local percentage change in the implicit net-of-tax rate. The elasticity of the retirement age with respect to the net-of-tax rate can be calculated as

$$\hat{\varepsilon} = \frac{b/\hat{R}}{\Delta\tau/(1-\tau)}. \tag{2}$$

The formula is based on the result that the excess mass is directly related to the labor supply response of the marginal bunching individual (Saez 2010), here $b \approx \Delta R$. Elasticities computed according to (2) are referred to as *observed elasticities* for the remainder of the paper.

B. *Estimation Using Multiple Bunching Observations*

The observed elasticity $\hat{\varepsilon}$ corresponds to a structural labor supply elasticity in a frictionless model where workers only respond to financial incentives. In this case, bunching is only a function of the elasticity and a vector of observables x related to the threshold, including the counterfactual density and the kink size. Following the notation of Kleven (2016), $B = B(\varepsilon, x)$, and ε can be estimated from bunching at a single discontinuity as above. However, bunching responses can serve to identify additional parameters. Writing bunching at threshold i as $B_i = B(\varepsilon, \omega, x_i)$, where ω is a vector of k additional parameters, identification requires observing $n \geq k+1$ bunching moments. If $n = k+1$, the implied system of n equations has an exact solution given the set of observed bunching moments. In this paper, bunching is observed at many discontinuities, such that $n > k+1$ and parameters can be estimated across "bunching observations" B_i.

Specifically, I aim at estimating the direct effect of statutory ages on bunching, which is later interpreted in terms of reference dependence. Denoting D_i an indicator for the presence of a statutory age at threshold i,

$$B_i = B(\varepsilon, \omega(D_i), x_i). \tag{3}$$

[18] See, e.g., Brown (2013) and Manoli and Weber (2016) for previous work on retirement bunching.

[19] The empirical implementation allows for round-number effects in addition. See online Appendix Section D.1 for details of the bunching estimation in practice.

Hence, statutory ages directly affect bunching via ω. Parameters can be identified when bunching is observed at sufficiently many thresholds that vary in D_i and x_i under the following assumption.

ASSUMPTION A: $E(\varepsilon_i | D_i) = \varepsilon$. *That is, structural elasticities do not vary systematically between statutory retirement ages and pure financial incentive discontinuities.*

Intuitively, the assumption rules out that stronger responses to financial incentives are falsely interpreted as a direct effect of statutory ages. Note that the assumption is concerned with underlying structural elasticities, which differ from observed elasticities estimated according to (2) in the presence of statutory age effects. In fact, equations (2) and (3) imply differences in observed elasticities across types of discontinuities as a corollary. An observed elasticity at a statutory age overestimates the true elasticity if some of the bunching occurs due to non-financial factors.[20] It is also important to note that the bunching approach generally allows for heterogeneity in underlying elasticities (and other parameters). In this case, bunching identifies an average retirement response, and local average parameter values at the threshold (Kleven 2016).

Within-Group Estimation: For part of the analysis, parameters can be estimated within groups indexed by g:

$$B_{ig} = B\left(\varepsilon_g, \omega_g(D_{ig}), x_{ig}\right). \tag{4}$$

This requires observing bunching both at statutory ages and pure financial incentive discontinuities for the same group of workers g. Restricting the analysis to groups facing both types of discontinuities allows for identification under a weaker assumption.

ASSUMPTION B: $E(\varepsilon_{ig} | D_{ig}) = \varepsilon_g$. *That is, a given group of workers g exhibits the same structural elasticity at statutory retirement ages and pure financial incentive discontinuities.*

Hence, elasticities can vary across groups in unrestricted ways, but a given group of workers are required to respond to all financial incentives in the same manner. I return to discussing the empirical validity of Assumptions A and B in Sections IIIB and IIIC.

Optimization Frictions: Evidence from previous work indicates that optimization frictions seem to play a relatively minor role for the timing of retirement (e.g., Manoli and Weber 2018) and extensive-margin responses more generally

[20] Existing studies estimating additional parameters from bunching focus mostly on optimization frictions, such as a fraction of workers unable to adjust or a fixed cost of adjustment (e.g., Chetty et al. 2011; Kleven and Waseem 2013; Gelber, Jones, and Sacks 2020). In a situation with optimization frictions, the observed elasticity underestimates the true elasticity.

(Chetty 2012). These findings are also mirrored by the sharp retirement responses documented in this paper. However, it is not necessary to assume that there are no frictions for the purpose of the above analysis. Denoting a vector of friction parameters by ϕ, if $B_i = B(\varepsilon, \omega(D_i), \phi, x_i)$, the additional assumption necessary to identify statutory age effects is that frictions do not vary systematically with D_i. In other words, if frictions attenuate responses to different discontinuities in the same way, the relative magnitude of the effects of interest can still be identified.[21]

III. Reduced-Form Evidence

A. *Basic Bunching Analysis*

Bunching at Specific Discontinuities: Some Cases.—I begin by presenting some cases of bunching at specific discontinuities that lend themselves to two natural comparisons between statutory retirement ages and pure financial incentives.

Statutory Retirement Age versus Contribution Notch within Group: First, panels A1 and A2 of Figure 3 show that the same group of workers respond more strongly to a discontinuity linked to a statutory age than to pure financial incentives. Panel A1 plots the job exit age distribution of women born in 1945 and 1946 around their ERA of 60. The average kink size is 0.07, implying a 7 percent reduction in the implicit net-of-tax rate at the threshold. There is large excess mass of 12.2 and the observed retirement age elasticity calculated according to equation (2) is 1.46. Panel A2 shows the distribution of years of contributions of women in the same birth cohorts around the threshold of 15 years required for the women's pathway. The jump in the average tax rate is 0.7pp, and this notch corresponds to an approximate kink size of 0.28. The average job exit age around the notch is 60.4. There is sharp bunching at 15 years and some missing mass to the left. However, the excess mass of 1.36 is significantly less than that at the ERA in panel A1 where workers face a smaller kink. The observed elasticity of 0.04 is much smaller than that of the same group at the ERA.

Statutory Retirement Age versus Pure Financial Incentive Kink: For the second comparison, panels B1 and B2 show bunching at two similar kinks, with and without a statutory retirement age. Panel B1 shows bunching around the FRA at 63 for cohorts 1944 to 1946 in the invalidity pathway. The kink size is 0.33 and the excess mass is estimated at 10.4, which implies an observed elasticity of 0.20. Panel B2 shows the distribution of job exit ages for workers born between 1938 and 1946 in the disability pathway. They face a pure financial incentive kink of size 0.29 at age 63. Consequently, workers in panels B1 and B2 face similar kinks at the same age, but the threshold is not presented as a full retirement age in the disability pathway. In contrast to the large excess mass at the FRA, bunching is hardly visible and the

[21] For instance, this would be given if there was a constant share of non-optimizers, leading to a proportional attenuation of bunching as in Kleven and Waseem (2013).

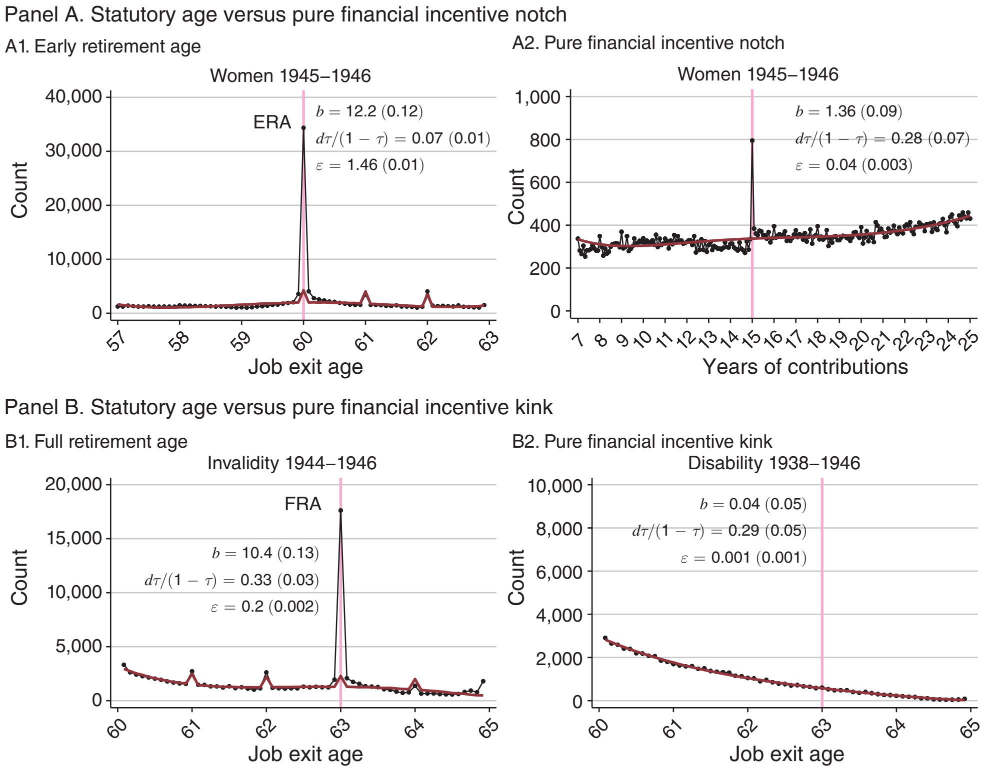


Figure 3. Bunching at Specific Discontinuities

Notes: The figure shows bunching at selected discontinuities. Panel titles indicate the type of discontinuity and panel subtitles indicate pathways and birth cohorts used. In panels A1, B1, and B2, the connected black dots show counts of job exit ages in monthly bins among the respective group of workers. In panel A2, the black dots show counts of years of contributions instead. In all panels, the red line shows the counterfactual distribution estimated as a seventh-order polynomial, including round-age dummies in panels A1 and B1. Vertical red lines indicate the location of the discontinuity. The variable b is the excess mass, $d\tau/(1-\tau)$ is the change in the implicit net-of-tax rate at the discontinuity (kink size), and ε is the observed elasticity of the retirement age with respect to the implicit net-of-tax rate. In panel A2, the legend displays the notch size, i.e., the jump in the average tax rate, and the average job exit age at the notch in addition. For the excess mass and observed elasticity, bootstrapped standard errors are in parentheses. For the kink size, notch size, and average job exit age, standard deviations are in parentheses.

excess mass is only 0.04 at the disability kink. Consequently, the observed elasticity of 0.001 is far below the estimate at the FRA.

Bunching across All 644 Discontinuities.—Panel B of Table 3 summarizes bunching responses across all 644 discontinuities in the data. In column 1, the average excess mass of 19.8 across the 386 kinks linked to statutory ages is very large. Columns 2 to 4 show that this is driven by large responses to all three types of statutory ages, with the largest excess mass at NRAs. Attributing all bunching to the change in the implicit net-of-tax rate implies an average observed elasticity of 0.49. Again, elasticities are large across all types of statutory ages.[22] Next, columns

[22]Non-convex NRA kinks are not included in the elasticity estimation since bunching in response to those would imply a negative observed elasticity.

TABLE 3—BUNCHING RESPONSES ACROSS 644 DISCONTINUITIES

	Statutory retirement age plus financial incentives				Pure financial incentives		
	All (1)	Early (2)	Full (3)	Normal (4)	All (5)	Kinks (6)	Notches (7)
Panel A. Summarizing discontinuities							
Kink size $\Delta\tau/(1-\tau)$							
Mean	0.04	0.22	0.25	−0.50	0.42	0.32	0.44
Standard deviation across discontinuities	0.39	0.14	0.23	0.30	0.16	0.12	0.15
Standard deviation within discontinuity	0.05	0.02	0.04	0.12	0.08	0.05	0.08
Mean retirement age	62.5	61.1	62.7	65.0	60.4	60.5	60.4
Panel B. Bunching responses							
Excess mass b	19.8	14.1	21.6	32.7	3.81	0.10	4.31
	(0.79)	(0.98)	(0.86)	(1.77)	(0.28)	(0.04)	(0.34)
Observed elasticity $\hat{\varepsilon}$	0.49	0.56	0.44	1.01	0.07	0.01	0.08
	(0.02)	(0.03)	(0.02)	(0.14)	(0.01)	(0.002)	(0.01)
Observations (discontinuities)	386	117	257	93	258	78	180

Notes: Panel A of the table summarizes discontinuities in the bunching sample by type. Kink size is the percentage reduction in the net-of-tax rate at the discontinuity. *Standard deviation across discontinuities* is standard deviations of kink size across discontinuities of a given type. *Standard deviation within discontinuity* is standard deviation of kink size within a group of workers facing the same discontinuity. *Mean retirement age* is the average retirement age at which a type of discontinuity is located. Panel B shows bunching responses by type of discontinuity. Excess mass and observed elasticities are computed as described in Section II. Note that the number of discontinuities in columns 2 to 4 are larger than the total in column 1 because some kinks are linked to more than one type of statutory age. Observed elasticities are only calculated at convex kinks, that is excluding non-convex NRA kinks. All statistics are weighted by group size. In panel B, standard errors are in parentheses.

5 to 7 report bunching responses to the 258 pure financial incentive discontinuities. The average excess mass is 3.81, and the average observed elasticity is 0.07. Among pure financial incentives, the elasticity is 0.08 at notches and 0.01 at kinks.[23]

The difference in observed elasticities suggests that, conditional on kink size, the response to statutory ages is about seven times larger than that to pure financial incentives. This is even more marked than the difference in raw excess mass, reflecting that kink sizes are larger at pure financial incentives on average. The observed elasticity at statutory ages is also an order of magnitude above previous estimates from pure financial incentives of around 0.01 to 0.04 (Brown 2013, Manoli and Weber 2016). Moreover, a first indication that bunching at statutory ages seems to occur independently of financial incentives is given by the large excess mass at non-convex NRA kinks, where there is a disincentive to bunch.

To further investigate the extent to which differences in bunching are driven by differences in financial incentives, Figure 4 shows binned scatterplots of the excess mass at a discontinuity against kink size. Two main insights emerge from the figure. First, financial incentives alone cannot explain the bunching patterns.

[23] The larger observed elasticities at notches could be driven by several factors. First, kinks occur in the disability pathway where workers may display a lower true elasticity than in other pathways. Second, observed elasticities measured at notches represent an upper bound: Kleven and Waseem (2013) point out that the approximation of the notch as a kink for the marginal buncher in order to compute a reduced-form elasticity underestimates the size of the discontinuity, since everyone between the marginal buncher and the notch faces a larger change in the marginal tax rate. Third, additional months of contributions could come from some non-work periods, such that workers may have additional margins of adjustment to bunch at contribution notches.

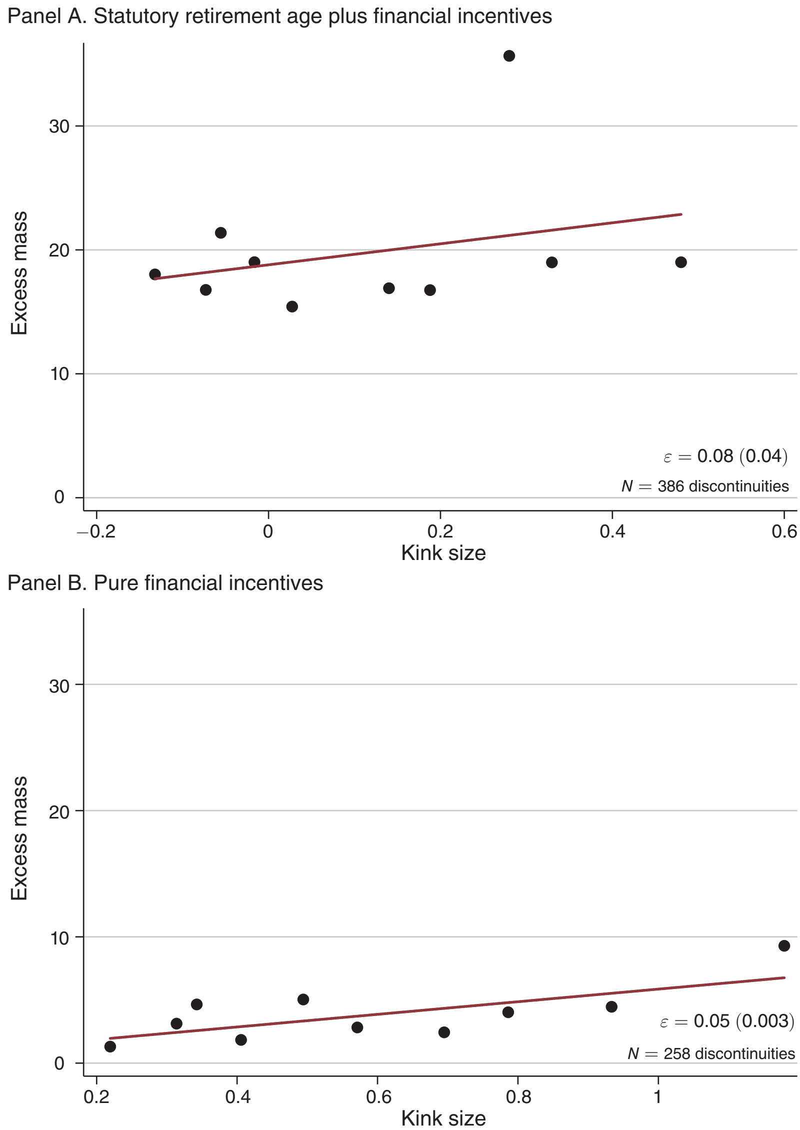


Figure 4. Bunching by Size of Financial Incentive

Notes: The figure shows binned scatterplots of the retirement response (excess mass) versus the underlying financial incentive (kink size) at a discontinuity, separately for statutory retirement ages (panel A) and pure financial incentive discontinuities (panel B). In panel A, the type of statutory ages (early, full, or normal retirement age) is controlled for. Each panel also includes the coefficient from a discontinuity-level regression of normalized excess mass $b/\hat{R}$ on kink size, which can be interpreted as a difference-in-bunching elasticity, with bootstrapped standard error in parentheses. Online Appendix Figure A4 shows additional plots separately by statutory age types.

In panel A, there is large excess mass at statutory ages across all kink sizes, even when there is a zero or negative incentive to retire. The second insight is that whether a discontinuity is presented as a statutory age matters directly for bunching. There are much larger responses at statutory ages in panel A than at pure financial incentives in panel B for any given kink size. Even the largest pure financial incentives induce less bunching than statutory ages. Note that this does not imply that there are no responses to financial incentives. Both panels A and B show a modest, but significantly positive relationship between excess mass and underlying kink size. The estimated slopes correspond to difference-in-bunching elasticities of 0.08 and 0.05, respectively.

B. *Reduced-Form Estimation*

The analysis so far suggests a large amount of additional bunching at statutory retirement ages. In order to quantify the importance of this "statutory age effect," I employ the following regression specification:

$$\frac{b_i}{\hat{R}_i} = \varepsilon \frac{\Delta \tau_i}{1-\tau_i} + \sum_s \beta^s D_i^s + Z_i' \gamma + \nu_i, \tag{5}$$

where an observation i corresponds to a discontinuity in the bunching sample. The variable D_i^s is an indicator for a statutory age of type $s \in \{ERA, FRA, NRA\}$ linked to discontinuity i, and the coefficients β^s measure the reduced-form effect of the respective statutory age type.[24] Finally, Z_i is a vector of control variables, and ν_i is an error term.

Equation (5) may be a natural reduced-form specification, but it can be also be interpreted as a simple, linear version of the bunching equation (3), where the parameter vector ω consists of a set of linear regression coefficients on the dummies D_i^s. The empirical setting provides many more bunching observations than parameters in the equation, which has two advantages. First, additional regressors can be included, allowing to control for group-level characteristics and fixed effects in a flexible way. Second, rather than finding an exact solution, the equation can be estimated via OLS, combining the information from all available bunching moments. Intuitively, statutory age effects are identified from the difference in bunching *between* statutory ages and pure financial incentive discontinuities, while the elasticity is identified from variation in kink size *within* each type of discontinuity. Standard errors are obtained via bootstrap by re-sampling bunching observations.

The key identification assumption for this specification is Assumption A. In practice, including control variables and fixed effects somewhat weakens the required assumption, such that elasticities should be independent of D_i conditional on these. Direct empirical support for Assumption A is lent by the results from Figure 4. The estimated slopes in panels A and B suggest that within type of discontinuity, the elasticity with respect to financial incentives is similar at statutory ages and pure financial incentive discontinuities.

Table 4 reports results from regressions based on equation (5). To begin with, column 1 shows results from a basic specification without controls. This yields

[24] In this section, I estimate the reduced-form effects of all types of statutory ages. Later, the structural estimation focuses on FRAs and NRAs, as these are arguably the more clear-cut cases to put forward a behavioral explanation.

Table 4—Reduced-Form Estimation

	Excess mass $b/\hat{R}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Kink size $\Delta\tau/(1-\tau)$	0.036	0.035	0.025	0.034	0.050	0.077	0.050	0.086
	(0.004)	(0.004)	(0.018)	(0.034)	(0.042)	(0.008)	(0.011)	(0.012)
Statutory age at kink:								
Early retirement age	0.066	0.046	0.039	0.049	0.051	0.060	0.072	0.082
	(0.007)	(0.006)	(0.018)	(0.032)	(0.041)	(0.004)	(0.010)	(0.010)
Full retirement age	0.071	0.059	0.071	0.079	0.072	0.069	0.075	0.099
	(0.010)	(0.009)	(0.017)	(0.034)	(0.039)	(0.008)	(0.009)	(0.015)
Normal retirement age	0.162	0.190	0.201	0.215	0.218	0.244	0.235	0.242
	(0.017)	(0.018)	(0.034)	(0.068)	(0.080)	(0.014)	(0.022)	(0.016)
Observations (discontinuities)	644	644	644	644	583	627	627	627
R^2	0.67	0.71	0.86	0.87	0.84	0.91	0.82	0.94
Statutory age interactions	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Worker controls	No	No	Yes	Yes	Yes	No	No	No
Pathway fixed effects, year-of-birth fixed effects	No	No	Yes	Yes	Yes	No	No	No
Pathway × year-of-birth fixed effects	No	No	No	Yes	Yes	No	No	No
Occupation-level controls	No	No	No	No	Yes	No	No	No
Heterogeneous coefficients								
by pathway	No	No	No	No	No	Yes	No	Yes
by year of birth	No	No	No	No	No	No	Yes	Yes
by pathway × year of birth	No	No	No	No	No	No	No	Yes

Notes: The table shows results from discontinuity-level regressions of normalized excess mass $b/\hat{R}$ on kink size as well as dummies for the presence of statutory retirement ages, using the bunching sample. Columns 1 to 5 report results from the baseline estimation according to equation (5). Statutory age interactions are interactions between dummies for each statutory age type. Worker controls include dummies for female, married, and East Germany, education years, lifetime earnings, last income before retirement, career length, sick leave years, parental leave years, and retirement age at the discontinuity. Occupation-level controls include firm size index, unionization rate, tenure in the firm, fraction in unlimited contracts, active union member rate, fraction receiving severance pay, and fraction of involuntary job exits. The number of observations is smaller in column 5 because occupation-level controls are only available for the occupation-matched sample, and discontinuities corresponding to too few individual observations are dropped. Columns 6 to 8 report weighted averages of heterogeneous coefficients estimated according to equation (6), where column 6 defines groups by pathway, 7 defines groups by year of birth, and 8 by pathway × year of birth. Groups with insufficient variation have to be dropped in columns 6 to 8, such that the number of observations is slightly smaller than in columns 1 to 4. All regressions are weighted by group size and bootstrapped standard errors in parentheses.

large and significant statutory age effects and an elasticity of 0.04. Next, column 2 adds interactions between different statutory age types in order to account for the fact that more than one type is present at some discontinuities. Column 3 adds a set of characteristics including income, education, gender, marital status, and retirement age at the discontinuity as controls, as well as pathway and year-of-birth fixed effects. Column 4 adds the largest set of group fixed effects, controlling for pathway times year-of-birth fixed effects. Finally, column 5 controls for occupation-level characteristics including firm size and unionization rates. With a coefficient of 0.16 to 0.22, the NRA has the largest reduced-form effect on bunching, while the FRA effect is 0.06 to 0.08 and the ERA effect is 0.04 to 0.07. In spite of the varying set of controls and fixed effects, the point estimates remain stable across specifications, although the ERA effect becomes insignificant due to larger standard errors in columns 4 and 5. Estimates of the elasticity range between 0.03 and 0.05, but they are only significant in columns 1 and 2.

C. *Heterogeneity*

A potential concern with comparing responses between statutory ages and pure financial incentives is that workers facing different discontinuities might differ in some relevant characteristics. For instance, pure financial incentive discontinuities tend to occur at somewhat younger retirement ages, and many apply to the disability pathway. This may in turn be correlated with elasticities, violating identification Assumption A. The main estimation is robust to controlling for a range of observables and fixed effects, but in this section I present additional evidence that the effect of statutory ages is not confounded by such differences.

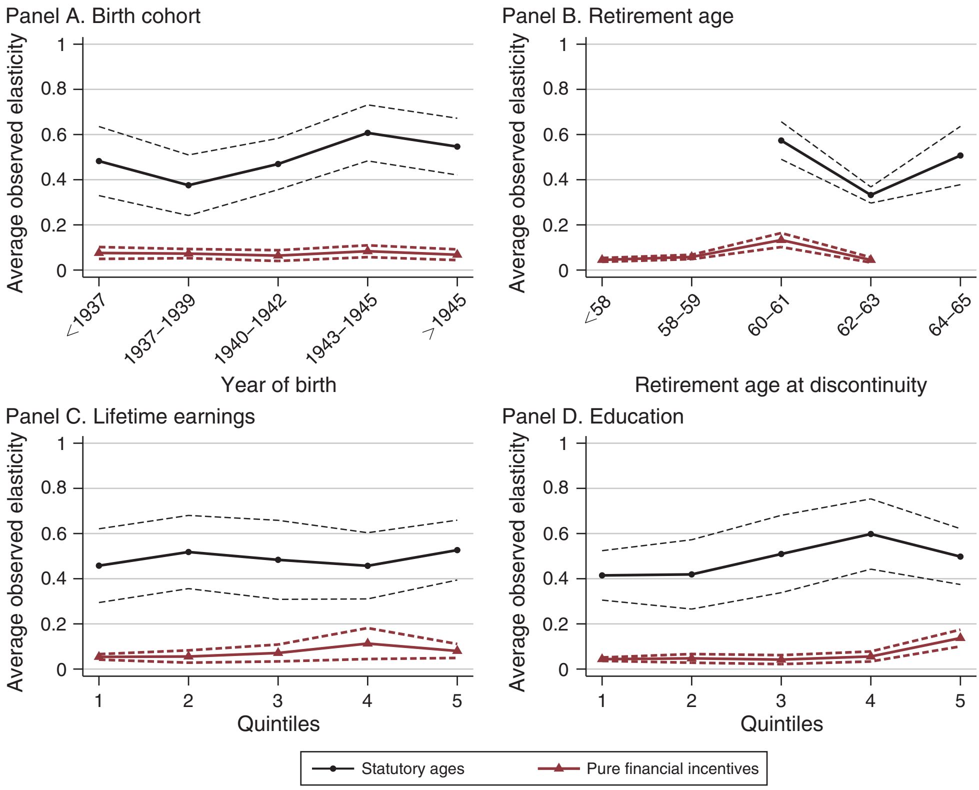
Can Heterogeneity Explain Differences in Bunching Responses? Figure 5 shows average observed bunching elasticities at statutory ages and pure financial incentive discontinuities by a range of observables. First, panels A and B sort bunching observations by birth cohort and the retirement age at the discontinuity, respectively. The remaining panels of the figure sort bunching observations by quintiles of individual characteristics, including lifetime earnings (panel C), years of education (panel D), and health status proxied by the negative of sick leave periods (panel E), as well as occupation-level characteristics including firm size (panel F), unionization (panel G), and tenure in the firm (panel H). In order to obtain partial correlations, each characteristic is first residualized via a regression on a set of basic variables including the other characteristics in the figure. Bunching observations are then sorted into quintiles by the residual from this regression.[25] There are large and significant differences in observed elasticities at statutory ages versus financial incentives among all birth cohorts, across the available range of retirement ages, and in each quintile of each characteristic. Hence, the strongly differential responses across types of discontinuities do not seem to be driven by differences across workers along the lines of age or other characteristics. Online Appendix Figure A5 shows that this conclusion is robust to using different sets of controls in the residualization regression or using raw characteristics, although the slope of bunching responses within each type of discontinuity can be sensitive to these choices.[26]

Estimation with Heterogeneous Parameters: As discussed above, a concern for identification arises if parameters are heterogeneous across workers facing different types of discontinuities. A second approach to address this is to allow directly for heterogeneous parameters in the following specification:

$$\frac{b_{ig}}{\hat{R}_{ig}} = \varepsilon_g \frac{\Delta \tau_{ig}}{1 - \tau_{ig}} + \sum_s \beta_g^s D_{ig}^s + \nu_{ig}, \tag{6}$$

[25] Rees-Jones (2018) uses a similar method in order to capture heterogeneity in bunching by tax reductions conditional on individual characteristics.

[26] In addition, online Appendix Table A1 reports results from a Oaxaca-Blinder decomposition, suggesting that the joint explanatory power of observable characteristics is limited. The decomposition attributes differences in excess mass between pure financial incentive discontinuities and statutory retirement ages to a component explained by differences in observables and an unexplained component. Eighty-four percent of the additional bunching at statutory ages cannot be explained by differences in observable characteristics. Financial incentives account for around 4 percent of observed differences, while worker and firm variables including those discussed here explain 15 percent and −3 percent, respectively.



(*Continued*)

Figure 5. Heterogeneity in Bunching Responses

where g indexes groups. An arguably natural definition of groups is to allow for heterogeneity at the level where benefit schedules and statutory ages are determined, namely pathway and year of birth. This strategy corresponds to a linear version of the within-group bunching equation (4). The specification requires identification Assumption B, which is weaker than Assumption A. Assumption B states that the same group of workers exhibits the same elasticity at different types of discontinuities, while true elasticities can vary arbitrarily across groups, allowing for a lower elasticity in the disability pathway, for instance.

Columns 6 to 8 of Table 4 report results from estimating equation (6) with varying group definitions. Note that the table reports weighted averages of coefficients, while selected pathway- and cohort-specific estimates are shown in online Appendix Table A2. First, column 6 estimates a specification with pathway-specific coefficients, and column 7 repeats the exercise with groups defined by birth cohorts. Column 8 reports estimates with groups defined by pathway and birth cohort. In the spirit of the comparison presented in Figure 3, this specification estimates elasticities and statutory age effects within narrowly defined groups such as women born in 1945. Overall, results remain very similar to the baseline estimation. In all specifications, statutory age effects are highly significant and increase slightly to between 0.06 and 0.24 compared to columns 1 to 5. The estimated elasticity is between 0.05

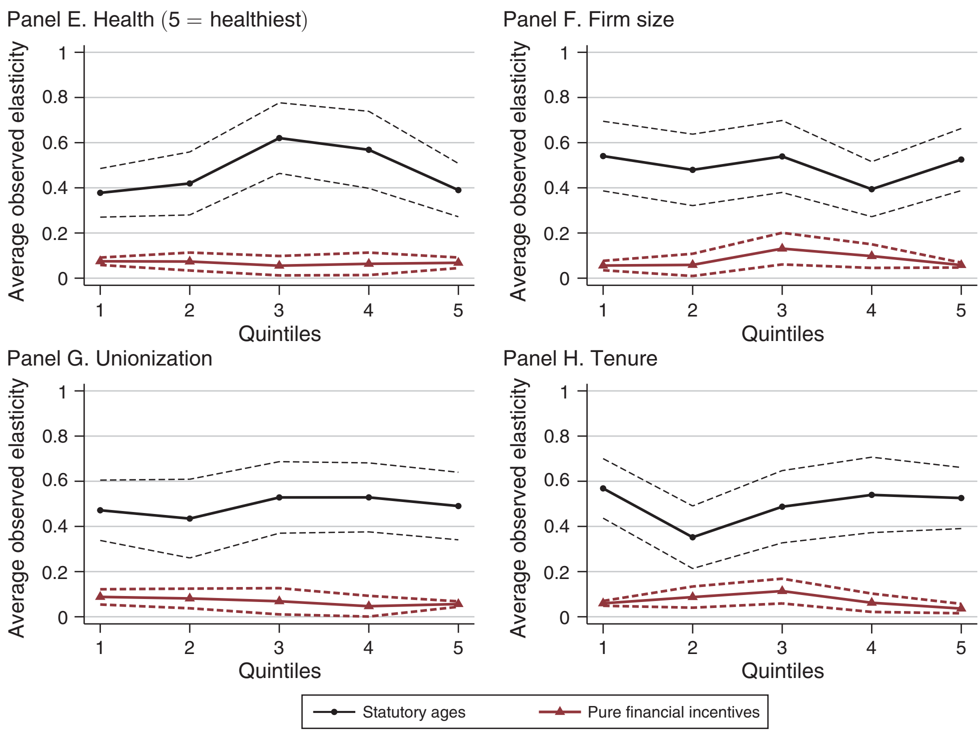


Figure 5. Heterogeneity in Bunching Responses (*Continued*)

Notes: The figure shows average observed bunching elasticities by birth cohort, the retirement age at the discontinuity, and worker and firm-related characteristics, namely lifetime earnings, years of education, health status (5 = healthiest), a firm size index computed from discrete size categories, unionization rate, and tenure. In order to obtain the partial effect of each characteristic in panels C to H, it is first regressed on a set of other basic variables including lifetime earnings, education, health, gender, marital status, parental leave, firm size, unionization, tenure, and year-of-birth fixed effects. Bunching observations are then sorted by quintiles of the residual from this regression. See online Appendix Figure A5 for analogous graphs using raw characteristics and alternative controls in the residualization regressions. In all panels, black dots indicate bunching at statutory ages, and red triangles indicate bunching at pure financial incentive discontinuities. The dashed lines around the point estimates mark 95 percent confidence intervals based on bootstrapped standard errors.

and 0.09. In particular, the fact that estimated statutory age effects change little suggests that differences in elasticities across pathways do not seem to introduce much bias in the baseline results.

IV. Mechanisms

In this section, I discuss potential mechanisms behind the reduced-form effect of statutory retirement ages.

A. *Can the Government Effectively Change Statutory Retirement Ages?*

First, I show that workers' retirement decisions react directly to a change in statutory ages, suggesting that the government can effectively set and change statutory ages. To this avail, I exploit variation due to cohort-based reforms (see online Appendix

Figure A2). One prominent reform enacted over the sample period is the increase in the FRA in the women's pathway from age 60 to 65 for birth cohorts 1940 to 1945. As it is implemented gradually, the reform creates fine-grained variation where each monthly birth cohort faces an additional one-month increase in the FRA.

Figure 6 shows the effect of the FRA increase on retirement behavior. Panel A displays the average job exit age in the women's pathway by month of birth around the reform. Among the pre-reform cohorts 1935 to 1939, the average job exit age is around 61 and exhibits no clear trend, besides some seasonal fluctuations. Starting with January 1940, there is a remarkably linear upward trend in job exit ages while the FRA is gradually increased. For the post-reform cohorts, the average job exit age is just below 63 and again remains stable. A before-after estimate indicates an effect of the reform on job exit ages of 1.70 years, corresponding to a 4.1 months increase in actual retirement ages per one-year increase in the FRA.[27]

Panel B shows job exit age distributions of the last pre-reform birth cohort 1939, the first post-reform cohort 1945, as well as selected monthly cohorts during the transition period. The graphs suggest that the increase in the average job exit age is driven by a shift in the distribution from the pre-reform FRA to the post-reform FRA. Before the reform, there is a large job exit age spike at age 60 and a relatively small spike at 65. After the reform, a large spike at 65 emerges. Since the women's ERA remains at age 60 after the reform, there is still a smaller spike at this age. In addition, job exit age distributions among selected transition cohorts are shown, namely June 1940, February 1941, April 1942, and July 1943, whose FRA is 60 and 6 months, 61 and 2 months, 62 and 4 months, and 63 and 7 months, respectively. For each cohort, there is large bunching precisely in the month of the FRA, even though the policy changes at a high frequency and FRAs are located at non-round ages. Online Appendix Figure A6 shows the complete set of distributions for the 60 monthly birth cohorts during the transition period. Across all cohorts, the spike in retirement moves in lockstep with the monthly FRA change.

B. *The Effect of Framing*

Does the framing of statutory ages affect retirement behavior? This is difficult to test directly, as the framing is ubiquitous and to my knowledge the way statutory ages are presented per se has not changed over the last decades. To obtain suggestive evidence, I exploit a reform affecting the intensity of framing instead, where the German State Pension Fund increased the frequency of information letters sent to workers. Before June 2002, workers received a letter only once in their lifetime, when they turned 55. Under the new regime phased in between June 2002 and December 2003, letters are sent annually to all workers (see Dolls et al. 2018). The stated goal of the reform was to better inform workers about benefit and retirement rules. Online Appendix Figure A7 shows an example of a letter. Letters provide detailed, personalized information on the worker's contributions so far, pension benefit calculation, and some guidance on making intertemporal decisions. Projected benefit amounts at different retirement ages are also shown. However, letters emphasize statutory ages

[27] Manoli and Weber (2018) use a regression kink design to analyze an ERA increase in Austria and find effects of similar magnitude on average job exit ages.

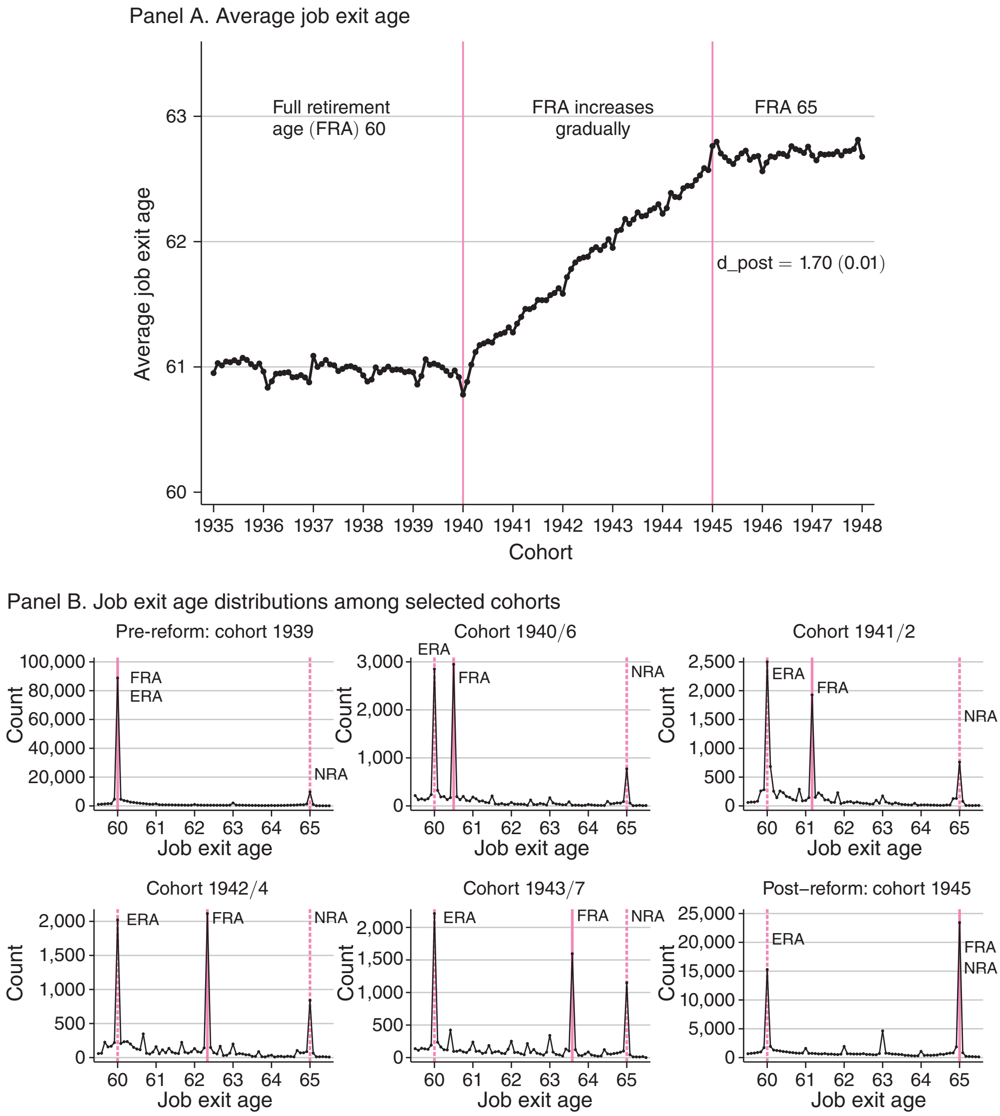


FIGURE 6. THE EFFECT OF INCREASING THE FULL RETIREMENT AGE

Notes: The figure shows the effect of a reform that increases the full retirement age (FRA) in the women's pathway. For birth cohorts 1939 and older, the FRA is 60 and from cohort 1945 onward the FRA is 65. For the 60 monthly birth cohorts born between 1940 and 1944, the FRA increases by one month for each month of birth. Panel A displays the average job exit age among workers in the women's pathway retiring at age 60 and above. The graph also includes the coefficient from an individual-level before-after regression, see online Appendix Table A3 for details. Panel B shows selected job exit age distributions throughout the reform. The first and last graph are for the last pre-reform cohort 1939 and the first post-reform cohort 1945, respectively. The remaining graphs show distributions among selected monthly cohorts during the transition period where the FRA increases on a monthly basis. In each graph, the connected dots show the count of job exits within monthly bins. The solid vertical red line indicates the location of the FRA, and dashed vertical red lines indicate other statutory retirement ages. Online Appendix Figure A6 shows the full set of monthly job exit age distributions during the transition period.

as reference dates, in particular the NRA. For instance, the first paragraph shows the exact date when the individual will reach the NRA. Moreover, two out of three benefit scenarios in the letter use the NRA as the hypothetical retirement date.

Panel A of online Appendix Figure A8 shows the fraction of workers bunching at different types of discontinuities by calendar quarter around the reform. First, there is no visible change in the response to pure financial incentives, in spite of the goal of providing better information. On the contrary, the probability of bunching at statutory ages increases, and this is driven by a significant increase in the probability of bunching at the NRA. The before-after coefficient shown in the figure indicates a 3pp increase at the NRA. In addition, the reform creates variation in the number of letters across birth cohorts (see online Appendix Section B.4 for details). In panel B of the figure, there is a gradual increase in the probability of retiring at the NRA for cohorts 1941 onwards as the number of letters received in the years before the NRA increases. However, there is no clear effect for earlier cohorts who receive just one letter before the NRA. This lack of immediate effect may be related to receiving the letter only in one year just before the NRA, leaving relatively little time to change retirement plans. Overall, the probability of bunching at the NRA increases by 2pp among the post-reform cohorts.

These findings are consistent with an effect of direct communication by the pension administration that emphasizes statutory ages, but they remain somewhat suggestive. Mastrobuoni (2011) uses a similar strategy exploiting the introduction of the US social security statement. In line with my results, the study finds no effect on the responsiveness to financial incentives. However, he also finds no substantial effect on the probability of bunching at statutory retirement ages.[28] Finally, it is worth noting that the estimated effect is small and bunching is already remarkably large before the reform when workers receive only one letter in their lifetime. This may point at the importance of the broader framing discussed in Section IA.

C. *Alternative Mechanisms*

The Role of Firms.—Laying off workers at statutory ages is sometimes cited as a way for firms to avoid costs of firing older workers. In the German labor market, mandatory retirement is possible at the NRA, but not at the ERA or FRA. Recent evidence by Rabaté (2019) suggests that similar mandatory retirement rules can only explain 12 percent of bunching at the NRA in France. Similarly, firm responses do not seem to be the main driver of statutory age retirements in the German setting. In particular, I investigate bunching at statutory ages among two subgroups where firm incentives play no role or a smaller role. First, although limited, there are a number of self-employed individuals enrolled in the public pension system.[29] Second, small firms with less than 10 employees are exempt from employment protection rules, so there should be little need for employers to lay off older workers specifically at statutory ages. Online Appendix Figure A9 shows job exit age distributions among the full occupation-matched sample (panel A), self-employed workers enrolled in the public

[28] This difference could be driven by two factors. First, the content of the US statements differs somewhat from the German letters. In fact, Mastrobuoni (2011) interprets them as an information treatment rather than a framing treatment. While still centered around the NRA, the US statements show expected social security benefits at a range of possible retirement ages between 62 and 70. The German letters, on the other hand, only show expected benefits from retiring at the NRA, or immediately. Second, the small positive estimates from this paper might be within the confidence intervals of Mastrobuoni (2011), which are not shown.

[29] Self-employed individuals can be enrolled in the public pension system for two reasons. First, a small set of self-employed occupations are mandated to participate. Second, self-employed workers can enroll voluntarily.

pension system (panel B), and the 20 occupations most frequently in small firms (panel C). There are sharp spikes among the self-employed at the main statutory ages and the fraction bunching of 28 percent is only 3pp less than in panel A. Hence, most bunching at statutory ages seems to persist in the absence of firm responses. Moreover, although most contracts are not subject to employment protection rules, there are also sharp spikes at statutory ages in panel C and the fraction bunching is 30 percent.

More generally, variables related to firms' incentives do not seem to explain much of the bunching at statutory ages. Firing frictions may be more severe for larger firms, in more unionized sectors, and for workers with longer tenure and unlimited contracts, for instance. Moreover, in a tighter labor market it may be more valuable to keep older workers beyond statutory ages. In Figure 5, there is large additional bunching at statutory ages compared to pure financial incentives in all quintiles of firm size, unionization, and tenure, which are observed at the occupation level. In addition, online Appendix Table A5 shows results from individual-level regressions of the probability of bunching at statutory ages on these characteristics, as well as the fraction of workers in unlimited contracts and a measure of labor market tightness at the state-year level. The probability of bunching increases with firm size, but somewhat surprisingly decreases in unionization, tenure, and unlimited contracts and increases in labor market tightness, while the coefficients are modest in magnitude.

Statutory Retirement Ages as Implicit Advice?—The results so far suggest a behavioral explanation for bunching at statutory ages, but reference dependence is not the only possible mechanism. A leading alternative behavioral mechanism may be that individuals interpret statutory ages as an implicit suggestion or advice by the government and those who find it difficult to make optimal retirement decisions follow this suggestion. A natural implication of this mechanism would be that responses to statutory ages are concentrated among less financially sophisticated workers. In the data, there is no clear negative relationship between bunching at statutory ages and available proxies for financial literacy. In Figure 5, workers at all education and income levels respond strongly to statutory ages. If anything, higher education and income seem to be associated with slightly larger responses to statutory ages, while there is also an increase in the response to pure financial incentives in the highest education quintile. These correlations are somewhat sensitive to the choice of control variables, however (see online Appendix Figure A5). Online Appendix Table A5 performs a similar correlation test at the individual level. In column 1, workers retiring at statutory ages have higher education and are more likely to be in an economically trained occupation. They also have higher lifetime income and higher last earnings before retirement, providing no indication that higher financial literacy diminishes bunching at statutory ages. Column 2 shows that the results are robust to limiting the sample to retirements no more than one year away from statutory ages. These correlations are similar to results by Behaghel and Blau (2012) who find that benefit claiming at the FRA is positively related to some survey measures of cognitive ability and wealth.[30]

[30]Moreover, the ratio of pension wealth to annual earnings, a proxy for the relative importance of public pensions for the worker, is positively related to the probability of bunching at statutory ages in online Appendix

Liquidity Constraints.—Since pension benefits can only be claimed from the ERA onward, liquidity constraints may provide a potential reason to retire at the ERA. Liquidity-constrained workers may not be able to smooth lifetime consumption throughout the gap between job exit and ERA to the desired extent. Recent evidence by Goda et al. (2018) suggests that liquidity constraints are not the main driver of ERA retirements in the United States. In addition, online Appendix Table A5 shows no indication of workers retiring at statutory ages being liquidity constrained, as they have both higher lifetime incomes and higher last earnings before retirement. However, the importance of this mechanism remains hard to check directly in the absence of data on assets and I cannot fully exclude that liquidity constraints explain part of the response to the ERA, besides serving as a potential reference point among earlier retirees. Hence, the conceptual interpretation and structural estimation in the remainder of the paper focuses on the FRA and NRA.

V. Retirement Bunching and Reference Points in a Simple Model

In this section, I incorporate reference dependence into a simple model of retirement decisions. It is arguably natural that workers perceive a salient benchmark presented by government policy as a normal time to retire as a reference point, in particular given that retirement is a one-off decision where other potential reference points such as previous outcomes or a status quo are not available. Moreover, I show in Section VC that the retirement density around statutory ages is consistent with the reference-dependent model, but more difficult to reconcile with alternative behavioral models. As discussed above, the remaining analysis focuses on the FRA and NRA, as they are not confounded by liquidity constraints.

A. *Basic Setup and Bunching at a Budget Constraint Kink*

Consider a simple static model of retirement decisions where workers maximize lifetime utility $U = u(C) - v(R,n)$.[31] Here, C is lifetime consumption, R is the worker's retirement age relative to a career starting age normalized to 0, and n is a parameter capturing earnings ability at old age. Utility is increasing and concave in consumption and disutility from lifetime labor supply is strictly convex such that $u'(C) > 0$, $u''(C) \leq 0$, $v_R > 0$, and $v_{RR} > 0$. Moreover, low ability increases disutility from postponing retirement such that $v_{Rn} < 0$. The lifetime budget constraint expresses consumption C as a function of R as in equation (1). The slope of the budget constraint divided by the gross wage again defines the implicit net-of-tax rate $1 - \tau$.

Table A5. Thus, higher stakes do not seem to diminish responses, which speaks against inattention explaining retirements at statutory ages.

[31] The static model corresponds to the "lifetime budget constraint" model of retirement suggested by Burtless (1986). Similar static models are used in recent applications such as Brown (2013) and Manoli and Weber (2018) in order to quantify bunching at local discontinuities. Online Appendix Section E provides an outlook on the relationship with dynamic models and discusses how a number of extensions, including parameter heterogeneity and income effects, can be incorporated into the analysis.

Consider first the case of a linear budget constraint $C = w(1-\tau)R$, and assume, as is standard in the bunching literature, that utility is quasi-linear in consumption and iso-elastic in labor supply such that

$$U = w(1-\tau)R - \frac{n}{1+\frac{1}{\varepsilon}}\left(\frac{R}{n}\right)^{1+\frac{1}{\varepsilon}},$$

where ε is the elasticity of the retirement age with respect to the implicit net-of-tax rate. Workers' utility maximization yields

$$R = n\left[w(1-\tau)\right]^{\varepsilon}.$$

If the distribution of ability $F(n)$ is smooth, this implies a smooth distribution of retirement ages with density $h_0(R)$.

Bunching at a Budget Constraint Kink: Suppose now that there is a kink in the lifetime budget constraint such that the marginal implicit tax rate increases by $\Delta\tau$ at some retirement age threshold $\hat{R}$. Analogously to standard bunching models, the framework predicts bunching at the kink. Online Appendix Figure A10 illustrates the effect of the kink following Saez (2010) and Kleven (2016), and the full derivation is shown in online Appendix Section F.1. The amount of bunching can be linked to the retirement response of a marginal bunching individual whose indifference curve is tangent to the initial budget set at R^* and to the upper part of the new budget set at $\hat{R}$. Bunching is characterized by

$$\frac{R^*}{\hat{R}} = \left(\frac{1-\tau}{1-\tau-\Delta\tau}\right)^{\varepsilon} \tag{7}$$

or, if $\Delta\tau$ is small,

$$\frac{b}{\hat{R}} \approx \varepsilon\frac{\Delta\tau}{1-\tau}, \tag{8}$$

where $b = B/h_0(\hat{R})$ is the excess mass. This corresponds to the Saez (2010) bunching formula and implies equation (2), which is used to calculate observed elasticities in the reduced-form estimation.

B. *Bunching at a Reference Point*

Next, reference dependence can be incorporated into this standard bunching framework. Reference dependence captures the notion that workers evaluate their retirement age relative to a threshold $\hat{R}$. In particular, I consider a fixed, exogenous reference point set by policy in the form of a full or normal retirement age. Preferences of a reference-dependent agent are

$$U = u(C) - v(R,n) - \mathbf{1}(R \geq \hat{R}) \cdot \tilde{\lambda}(R-\hat{R}). \tag{9}$$

The last term in equation (9) introduces a discontinuity in individual marginal disutility from continuing work at $\hat{R}$. Marginal disutility from increasing labor supply beyond the reference point $\hat{R}$ is greater than marginal disutility from approaching

$\hat{R}$ from the left, where the parameter $\tilde{\lambda} > 0$ captures this kink in the utility function. This is consistent with an interpretation where workers perceive postponing retirement as a loss relative to a normal time to retire. Similar formulations of reference dependence with loss aversion are commonly used in the literature (e.g., DellaVigna et al. 2017 and Rees-Jones 2018). Note that the model represents a reduced-form way of incorporating loss aversion in two dimensions. First, gain utility is abstracted from for simplicity. Second, the parameter λ captures reference dependence directly in terms of the retirement age, rather than also considering reference dependence in consumption.[32] As I show below, this entails the advantage that bunching responses are analogous to those at a budget constraint kink, and underlying parameters can be identified straightforwardly via structural bunching estimation.

Figure 7 illustrates bunching responses to the reference point in a budget set diagram and a density diagram. Initially, indifference curves are smooth and an individual with ability $\hat{n}$ is located at $\hat{R}$, while n^* is located at R^*. When the reference point is introduced, indifference curves rotate counter-clockwise above $\hat{R}$ and now exhibit a kink at $\hat{R}$. The individual whose indifference curve was initially tangent to the budget line at R^* is now tangent at $\hat{R}$. This individual is the marginal buncher: All workers initially located between $\hat{R}$ and R^* bunch at the reference point, while all individuals initially to the left of the reference point leave their retirement age unchanged and all individuals initially to the right of R^* stay above the reference point. Like a budget constraint kink, the reference point does not produce a hole in the density of retirement ages, since workers initially above R^* also retire earlier, causing a leftward shift in the density above $\hat{R}$ that fills the hole. The bunching mass B is given by

$$B = \int_{\hat{R}}^{R^*} h_0(R)\, dR \approx h_0(\hat{R})(R^* - \hat{R}),$$

where $h_0(R)$ is the counterfactual density and the approximate equality holds if $h_0(R)$ is constant on $[\hat{R}, R^*]$. The two tangency conditions for the marginal buncher imply $R^* = n^*[w(1-\tau)]^{\varepsilon}$ and $\hat{R} = n^*[w(1-\tau) - \tilde{\lambda}]^{\varepsilon}$. Hence,

$$\frac{R^*}{\hat{R}} = \left(\frac{1-\tau}{1-\tau-\lambda}\right)^{\varepsilon}, \tag{10}$$

where $\lambda = \tilde{\lambda}/w$ expresses the reference dependence parameter relative to the gross wage w. Equation (10) implies that a kink in disutility from work has a bunching effect equivalent to a budget constraint kink. Workers respond *as if* there was a local change in the implicit net-of-rax rate of size λ. This result has two important implications. First, a natural interpretation of the magnitude of λ arises, as it can be scaled equivalently to kink size, a standard measure used in the bunching literature. Second, λ can be estimated based on bunching observed at a reference point, but one also needs to know or estimate the elasticity ε for this purpose. Intuitively, ε plays a

[32] A previous version of this paper considered reference dependence in terms of both the retirement age and consumption. Reference dependence in consumption could be motivated by the gain-loss framing of benefit levels around the FRA, for instance. While both models predict bunching at a reference point, the present formulation is better suited in terms of transparent identification of the key parameters and is well in line with the retirement distribution around statutory ages (see Section VC).

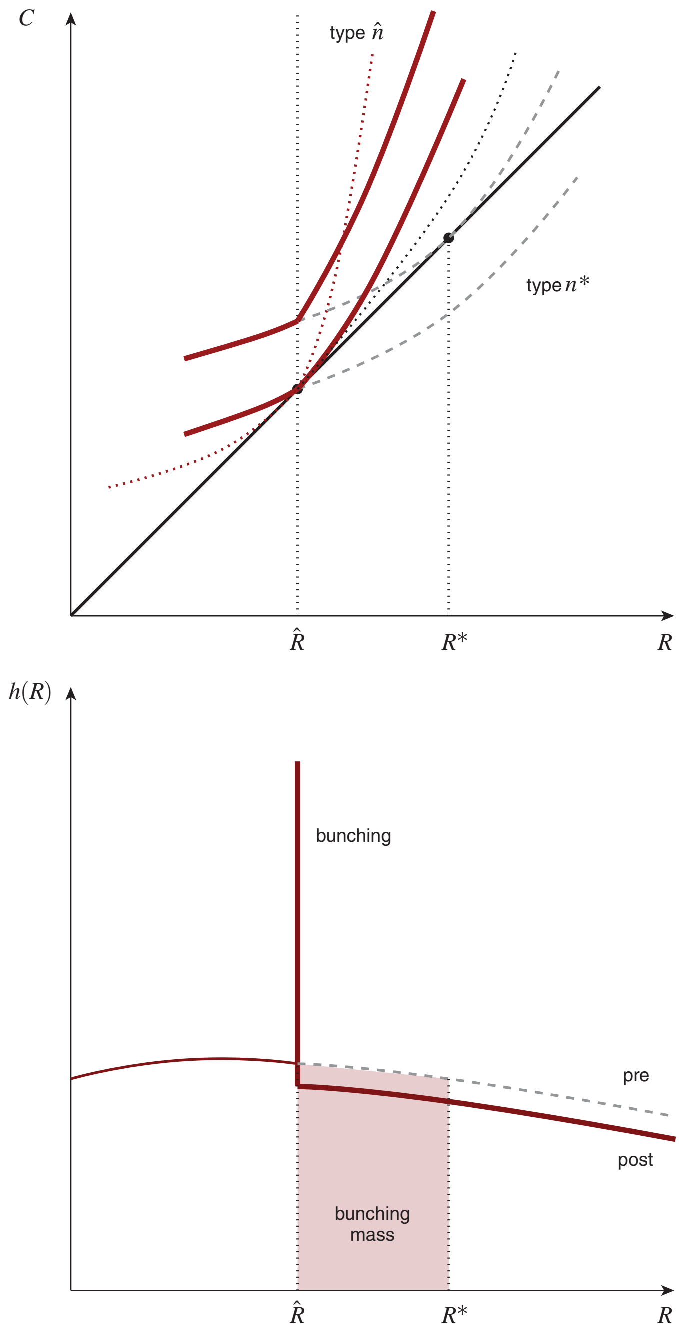


Figure 7. Retirement Bunching at a Reference Point

Notes: The figure shows predicted bunching responses to a retirement age reference point in an indifference curve diagram (top panel) and density diagram (bottom panel). In the top panel, the dashed gray curves are the initial (pre) indifference curves of the marginal buncher with ability n^*, whereas the solid red curves are her indifference curves after introducing the reference point (post). The dotted curves are indifference curves pre- (gray) and post-reference point (red) of an individual with ability $\hat{n}$ who retires at $\hat{R}$ before and after the change. The marginal buncher is tangent at R^* in the absence of the reference point, and tangent at $\hat{R}$ with the reference point. In the bottom panel, the solid red line denotes the post-reference point density, whereas the dashed gray line denotes the initial density. The red shaded area is the initial location of the mass of workers bunching in response to the reference point.

role for the amount of bunching for a given λ as it governs the utility cost to workers of adjusting their retirement age towards the reference point.

Combining Financial Incentives and Reference Points: At a statutory retirement age, a potential reference point coincides with a change in financial incentives. Thus, in order to compute total bunching, an initial situation without any discontinuity needs to be compared to a situation with the budget set kink and reference point. Online Appendix Figure A11 illustrates bunching responses to such a combined threshold. One can identify a marginal buncher whose original indifference curve is tangent to the original budget set at R^* and whose kinked indifference curve is tangent to the upper part of the kinked budget set at $\hat{R}$. Again, all individuals initially located between $\hat{R}$ and R^* bunch at the threshold and the total bunching mass is $B \approx h_0(\hat{R})(R^* - \hat{R})$.

The two tangency conditions for the marginal buncher imply $R^* = n^*\left[w(1-\tau)\right]^{\varepsilon}$ and $\hat{R} = n^*\left[w(1-\tau-\Delta\tau-\lambda)\right]^{\varepsilon}$. Hence, the excess mass $b = B/h_0(\hat{R})$ is

$$\frac{b}{\hat{R}} = \left(\frac{1-\tau}{1-\tau-\Delta\tau-\lambda}\right)^{\varepsilon} - 1. \tag{11}$$

Thus, if a retirement age reference point is at the same location as a budget constraint kink $\Delta\tau$, the additional bunching effect due to the reference point is as if the size of the kink increases by λ.

C. *Further Predictions and Alternative Behavioral Models*

Consistent with the empirical patterns, the model predicts sharp bunching at statutory retirement ages. However, reference dependence with loss aversion also makes further, distinctive predictions about the shape of the retirement age distribution. In particular, the framework implies a shift of the distribution on one side of the reference point: in Figure 7, the rotation of indifference curves causes a leftward shift of all retirement ages above $\hat{R}$. Workers initially located between $\hat{R}$ and R^* bunch, while workers initially above R^* retire earlier but remain above $\hat{R}$. Hence, there is no hole or missing mass, but an asymmetry of the distribution around $\hat{R}$, where the density is shifted to the left (downwards) over the range above the reference point.

These additional predictions can help assess the plausibility of the reference-dependent model against alternative behavioral models. Online Appendix Section F.2 sets out a model with a fixed utility premium that individuals derive from retiring at a statutory age. This can be interpreted as a reduced-form representation of alternative mechanisms where individuals view retiring exactly at statutory ages as implicit advice by the government (see Section IVC), or as a social norm. The model also predicts sharp bunching, but a different shape of the density around statutory ages. As there is a fixed utility cost of deviating from the statutory age in either direction, individuals whose counterfactual retirement age would be sufficiently close to the threshold bunch, but there are no retirement responses further away from the threshold. Hence, the alternative model predicts missing mass on both sides and no further shift in the density.

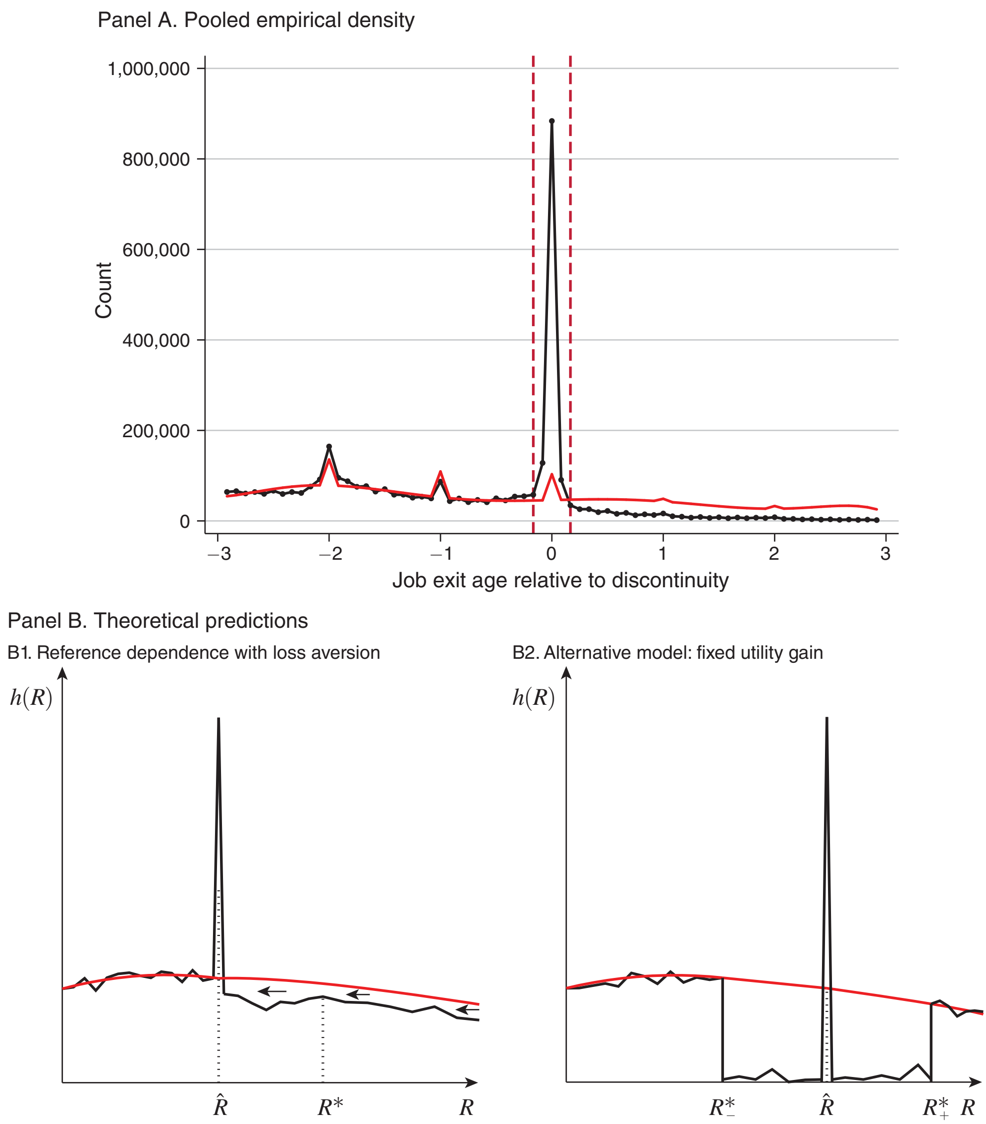


Figure 8. Empirical Density versus Model Predictions

Notes: Panel A of the figure shows the pooled empirical density around all full and normal retirement ages, with the age at the discontinuity normalized to zero. The black connected dots show the count of job exits within monthly bins. The red line shows a counterfactual distribution estimated as a seventh-order polynomial including round-age dummies. Vertical dashed lines indicate the bunching region excluded from the counterfactual estimation. To the right of the bunching region, the counterfactual is corrected for a leftward density shift as predicted by the reference dependence model. Panel B shows stylized density graphs, illustrating the predicted shape of the density around statutory ages under the reference dependence model (panel B1) and an alternative behavioral model with a fixed utility gain from retiring at statutory ages as described in online Appendix F.2 (panel B2). Panels B1 and B2 correspond to the lower panels in Figure 7 and online Appendix Figure F1, respectively, adapted to the shape of the empirical density. The black line shows a stylized observed distribution under each model emerging from a counterfactual given by the red line.

Figure 8 shows the empirical density pooled around all FRAs and NRAs (panel A), as well as stylized graphs of the density predicted by the reference-dependent model (panel B1), and the alternative model with a fixed utility gain (panel B2). The empirical distribution exhibits the usual sharp bunching and the density to the right of the threshold is visibly lower than on the left. This is consistent with a leftward shift of the density above statutory ages. The red curve in the graph shows a counterfactual density fitted via a polynomial with an additional correction on the right such that the bunching mass equals the density shift. The counterfactual fits well, illustrating that a leftward shift as predicted in panel B1 is a plausible source of observed bunching. On the contrary, the empirical density is not well in line with the alternative model, as there is no visual indication of missing mass in the neighborhood of the threshold. If anything, the density seems to locally increase toward the threshold on both sides.[33] Finally, online Appendix Figure A12 shows that the empirical density exhibits similar patterns when considering FRAs and NRAs separately.

VI. Structural Bunching Estimation and Counterfactuals

A. *Structural Bunching Estimation*

The previous section establishes a straightforward link between bunching at a reference point and the parameters governing the strength of reference dependence. Equations (7) and (11) imply that bunching observed at different discontinuities provides sufficient statistics to estimate these parameters. In particular, the variation in the presence of statutory ages and in kink sizes used for the reduced-form estimation can be exploited to identify ε and λ. Similarly to the reduced-form part, the estimation can be implemented at the discontinuity level, without having to estimate a full model of retirement decisions at the individual level. The availability of independent variation in statutory ages and financial incentives presents a crucial advantage, as existing bunching approaches to reference dependence are unable to estimate λ or similar parameters in the absence of an estimate of the cost of adjusting the relevant behavioral margin.[34]

Bunching at discontinuity i can be written as

$$\frac{b_i}{\hat{R}_i} = \left(\frac{1-\tau_i}{1-\tau_i-\Delta\tau_i-\Lambda(D_i)}\right)^{\varepsilon} - 1 + \xi_i, \tag{12}$$

where D_i is a vector of indicators for statutory ages, $\Lambda(D_i)$ denotes reference point effects as a function of statutory ages, and ξ_i is an error term. Reference point effects are then specified as a simple linear combination of the different types of statutory ages:

$$\Lambda(D_i) = \sum_s \lambda^s D_i^s,$$

[33] Similarly, the empirical density is not well in line with alternative functional forms of reference dependence, including a utility notch (a discontinuity in the level of utility) and diminishing sensitivity (a discontinuity in the second derivative of utility). Both of these formulations would induce sharp bunching at the reference point together with missing mass on one side (see the simulations in Allen et al. 2017 and Rees-Jones 2018).

[34] See DellaVigna (2018). For instance, in Rees-Jones (2018) the cost of effort to change one's tax liability is not known. Similarly, in Allen et al. (2017), the cost of running effort is unknown.

where λ^s is a parameter governing reference point effects of statutory age type s. Thus, the degree of reference dependence is allowed to vary with the type of statutory age as in the reduced-form estimation.

Online Appendix Table A6 reports results from a corresponding nonlinear least squares estimation at the discontinuity level. The estimation focuses on the FRA and NRA, and all discontinuities linked to an ERA are excluded. The baseline specification estimates λ^s by type of statutory age, and also includes interaction effects between types of statutory ages in order to account for the fact that they can coincide. The estimated λ^s parameters are positive and highly significant, with magnitudes of 0.23 at the FRA and 0.38 at the NRA. The estimates imply that marginal disutility from working an additional period changes by 23 percent to 38 percent of a worker's gross wage at the respective statutory age. In addition, the parameters can be scaled in terms of budget constraint kink equivalents. The estimated magnitude of reference dependence corresponds to a 51 percent kink at the FRA and 120 percent at the NRA. The very large estimate at the NRA is due to large observed bunching in spite of non-convex kinks. Finally, the elasticity of 0.05 is precisely estimated and similar to the reduced-form results. Online Appendix Table A7 shows that the parameter estimates are robust to a range of alternative specifications of reference point effects, including direct estimation of kink size equivalents, estimation without interaction effects, and separate estimation by type s.

B. *Counterfactual Simulations*

Finally, I simulate the effects of counterfactual policy scenarios. I focus on two policies often considered as options for pension reform. The first reform increases the NRA, as a number of countries including the United States and Germany are doing or planning to do. In the simulation, the NRA is raised from 65 to 66, without providing additional financial incentives to postpone retirement. In order to focus on reference point effects, the simulated reform does not entail a benefit cut across the board below the NRA.[35] The second reform increases financial rewards for late retirement similar to the US "Delayed Retirement Credit," while the NRA remains at 65. Simulating and comparing the effects of both scenarios is possible as the structural bunching estimation yields joint estimates of the parameters governing the responses to reference points and to financial incentives.

Table 5 summarizes the effects of both scenarios simulated for birth cohort 1946.[36] As a result of the one-year NRA increase, actual job exit ages increase by 3 months on average, and the increase among individuals who retire at 65 and above is 10 months. Online Appendix Figure A13 shows the simulated effect on the job exit age distribution. There is un-bunching of the spike at age 65, and the density above 65 increases. A new, large job exit spike emerges at the post-reform NRA of 66. A key implicit assumption behind the simulation is that the NRA shifted to the

[35] Reforms that increase the NRA in practice often feature such a benefit cut across the board, which entails a sizable positive mechanical fiscal effect. In the reform simulated here, only a relatively small mechanical effect arises due to late retirement rewards being paid from the new NRA onwards. See online Appendix Section G for further details of the policy simulations.

[36] I focus on workers born in 1946 as this is the last birth cohort not subject to a planned gradual increase in the NRA by 2031, such that the actual NRA is still 65 for these workers.

Table 5—Policy Counterfactuals

	Actual	Counterfactuals	
	(1)	(2)	(3)
Policy		Normal retirement age increase from 65 to 66	Increase in rewards for late retirement from 6% to 11.4%
Average job exit age (65 and above)	65.0	65.9	65.9
change (months)		+10.1	+10.1
Average job exit age (60 and above)	62.8	63.1	63.1
change (months)		+3.1	+3.2
Excess mass at NRA	31.3	28.4	12.6
change		−2.9	−18.7
Net fiscal effect (NPV for one cohort)		+€1,048m	−€420m
Contributions collected		+€425m	+€425m
Benefits paid out		−€623m	+€845m

Notes: The table shows results from a simulation of two counterfactual policies: an increase in the NRA (column 2) and an increase in financial rewards for late retirement (column 3). The effects of both policies are simulated for birth cohort 1946. Fiscal effects are calculated in terms of net present value at age 65 for this birth cohort. Excess mass figures are weighted by group size. See online Appendix Figure A13 for graphs of the simulated retirement age distribution under both scenarios.

new location is perceived by workers as a reference point similarly to the previous NRA. While this may not be true for arbitrarily large changes to the NRA, support that the assumption holds over some range of ages is provided by the evidence from Section IVA, where bunching moves in lockstep with the reform. It is also reassuring that the change in average retirement ages in the simulation is similar to estimated reform effects in Section IVA.

In the second scenario, the increase in late retirement rewards is calibrated to match the effect on the average retirement age in the first scenario. In order to yield the same effect, financial rewards would have to be almost doubled from currently 6 percent per year to 11.4 percent. Providing stronger financial incentives for late retirement leads to a drop in the excess mass at the NRA by more than one-half, and the former bunchers disperse along the density above age 65. Hence, both types of policies could achieve an increase in average actual retirement ages. However, the estimated fiscal impact of the two scenarios is very different. The NRA increase has a positive net fiscal effect of +€1,048 million in net present value terms for birth cohort 1946. This is due to the additional contributions paid by workers postponing retirement, combined with the shorter duration for which they receive pension benefits. On the contrary, the net fiscal effect of increased financial rewards is negative at −€420 million. Workers also contribute longer in this scenario, but this is more than offset by the large increase in pension benefits at older retirement ages necessary to induce workers to postpone retirement. Under some additional assumptions, a back-of-the-envelope calculation yields similar figures for the annual fiscal impact on the pension system (see online Appendix Section G.2). Assuming that a series of identical cohorts retire until reforms are fully phased in, the two scenarios generate a long-term annual fiscal impact of around +€1.1 billion and −€1.0 billion, respectively.

These results further highlight that statutory retirement ages are an effective policy tool for governments to influence retirement decisions. Such reforms can improve the fiscal balance of the pension system, as they can lead to an increase in actual retirement ages without requiring high financial rewards for postponing retirement.[37] On the other hand, the simulations show that positive fiscal effects are more difficult or impossible to achieve using pure financial incentives such as a delayed retirement credit.

VII. Conclusion

In this paper, I provide new, comprehensive evidence that the way retirement incentives are presented to workers has large effects. The results highlight the direct role of statutory ages, at which almost one third of job exits occur. Reference dependence provides a behavioral explanation for this phenomenon. In comparison, responses to pure financial incentives emphasized by standard models of retirement are modest.

There are implications for the design of pension systems and reform options. Having established their direct impact on behavior, statutory retirement ages themselves can be viewed as policy tools. Policy simulations suggest that shifting statutory retirement ages can be an effective way to increase actual retirement ages with a positive fiscal impact. Thus, such reforms can help adapt pension systems to demographic change.

Two limitations of the analysis are worth pointing out. First, this paper is agnostic about the welfare consequences of policies that set or manipulate statutory retirement ages. Such an evaluation would require a normative stance on the extent to which reference point effects enter welfare calculations. In addition, statutory age reforms likely have some distributional implications. Second, I do not study the formation of individual reference points around government policies. In particular, an open question may be how far governments can push statutory retirement ages beyond their established, historical range such that individuals still perceive them as "credible" reference points. Similarly, some caution might be warranted when completely decoupling statutory ages from financial incentives to which they were historically linked. Exploring these questions could be promising avenues for future research.

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): 1657–72.

Bastani, Spencer, and Håkan Selin. 2014. "Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule." *Journal of Public Economics* 109: 36–49.

Behaghel, Luc, and David M. Blau. 2012. "Framing Social Security Reform: Behavioral Responses to Changes in the Full Retirement Age." *American Economic Journal: Economic Policy* 4 (4): 41–67.

Börsch-Supan, Axel, and Christina B. Wilke. 2004. "The German Public Pension System: How It Was, How It Will Be." NBER Working Paper 10525.

[37] It is important to note that the magnitude of fiscal gains from increasing the NRA depends on the characteristics of the pension system. For instance, providing steep marginal pension increases for late retirement at the same time would dampen the fiscal effect. At the extreme, if the reform featured full actuarially fair benefit adjustment between the old and the new NRA, there would be no behavioral fiscal effects of increasing retirement ages per se but only mechanical fiscal effects.

Brown, Jeffrey R., Arie Kapteyn, and Olivia S. Mitchell. 2016. "Framing and Claiming: How Information-Framing Affects Expected Social Security Claiming Behavior." *Journal of Risk and Insurance* 83 (1): 139–62.

Brown, Kristine M. 2013. "The Link between Pensions and Retirement Timing: Lessons from California Teachers." *Journal of Public Economics* 98: 1–14.

Burtless, Gary. 1986. "Social Security, Unanticipated Benefit Increases, and the Timing of Retirement." *Review of Economic Studies* 53 (5): 781–805.

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

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

Cribb, Jonathan, Carl Emmerson, and Gemma Tetlow. 2016. "Signals Matter? Large Retirement Responses to Limited Financial Incentives." *Labour Economics* 42: 203–12.

DellaVigna, Stefano. 2018. "Structural Behavioral Economics." In *Handbook of Behavioral Economics: Foundations and Applications 1*, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 613–723. Amsterdam: Elsevier.

DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F. Schmieder. 2017. "Reference-Dependent Job Search: Evidence from Hungary." *Quarterly Journal of Economics* 132 (4): 1969–2018.

Deutsche Rentenversicherung. 2017. "Rente mit 67: Wie Sie Ihre Zukunft Planen Können." Information Brochure.

Dolls, Mathias, Philipp Doerrenberg, Andreas Peichl, and Holger Stichnoth. 2018. "Do Retirement Savings Increase in Response to Information about Retirement and Expected Pensions?" *Journal of Public Economics* 158: 168–79.

Duflo, Esther, William Gale, Jeffrey Liebman, Peter Orszag, and Emmanuel Saez. 2006. "Saving Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H&R Block." *Quarterly Journal of Economics* 121 (4): 1311–46.

Fetter, Daniel K., and Lee M. Lockwood. 2018. "Government Old-Age Support and Labor Supply: Evidence from the Old Age Assistance Program." *American Economic Review* 108 (8): 2174–211.

Forschungsdatenzentrum der Rentenversicherung (FDZ-RV). 2015. "Versichertenrentenzugang 1992–2014." Research Data Center of the German State Pension Fund (accessed August 14, 2020).

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.

Goda, Gopi Shah, Shanthi Ramnath, John B. Shoven, and Sita Nataraj Slavov. 2018. "The Financial Feasibility of Delaying Social Security: Evidence from Administrative Tax Data." *Journal of Pension Economics and Finance* 17 (4): 419–36.

Heien, Thorsten, Klaus Kortmann, and Christof Schatz. 2005. *Altersvorsorge in Deutschland 2005*. Report by TNS Infratest Sozialforschung.

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

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

Lalive, Rafael, and Stefan Staubli. 2015. "How Does Raising Women's Full Retirement Age Affect Labor Supply, Income, and Mortality?" NBER Retirement Research Center Paper NB14-09.

Manoli, Day, and Andrea Weber. 2016. "Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions." *American Economic Journal: Economic Policy* 8 (4): 160–82.

Manoli, Dayanand, and Andrea Weber. 2018. "The Effects of the Early Retirement Age on Retirement Decisions." Unpublished.

Mastrobuoni, Giovanni. 2009. "Labor Supply Effects of the Recent Social Security Benefit Cuts: Empirical Estimates Using Cohort Discontinuities." *Journal of Public Economics* 93 (11–12): 1224–33.

Mastrobuoni, Giovanni. 2011. "The Role of Information for Retirement Behavior: Evidence Based on the Stepwise Introduction of the Social Security Statement." *Journal of Public Economics* 95 (7–8): 913–25.

Merkle, Christoph, Philipp Schreiber, and Martin Weber. 2017. "Framing and retirement age: The Gap between Willingness-to-Accept and Willingness-to-Pay." *Economic Policy* 32 (92): 757–809.

O'Donoghue, Ted, and Charles Sprenger. 2018. "Reference-Dependent Preferences." In *Handbook of Behavioral Economics: Foundations and Applications 1*, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 1–77. Amsterdam: Elsevier.

Organisation for Economic Co-operation and Development (**OECD**)**.** 2019. *Pensions at a Glance 2019: OECD and G20 Indicators*. Unpublished.

Rabaté, Simon. 2019. "Can I Stay or Should I Go? Mandatory Retirement and the Labor-Force Participation of Older Workers." *Journal of Public Economics* 180: 104078.

Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies* 85 (2): 1251–78.

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

Seibold, Arthur. 2021. "Replication Data for: Reference Points for Retirement Behavior: Evidence from German Pension Discontinuities." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E120903V1.

Shoven, John B., Sita N. Slavov, and David A. Wise. 2017. "Social Security Claiming Decisions: Survey Evidence." NBER Working Paper 23729.

Socio-Economic Panel (**SOEP**)**.** 2015. "SOEP Data for Years 1984–2013, version 30i." Research Data Center SOEP. doi:10.5684/soep.v30i.

Staubli, Stefan, and Josef Zweimüller. 2013. "Does Raising the Early Retirement Age Increase Employment of Older Workers?" *Journal of Public Economics* 108: 17–32.

Thakral, Neil, and Linh T. Tô. Forthcoming. "Daily Labor Supply and Adaptive Reference Points." *American Economic Review.*