Salience and Social Security Benefits

Brinch, C.N., E. Hernæs, Z. Jia

Frisch Centre, Gaustadalléen 21, 0349 Oslo, Norway. http://www.frisch.uio.no

Frisch Centre

Salience and Social Security Benefits*

Christian N. Brinch† Erik Hernæs‡ Zhiyang Jia §

June 16, 2015

Abstract

We study the effect of salience in the social security benefit system on labor earnings by exploiting kinks and notches in budget lines introduced by earnings testing and social security accrual mechanisms for 67- to 69-year-old workers in Norway. An earnings test had large effects on labor earnings, while an accrual system discontinuity had no discernible effects. We interpret the difference as likely to be caused by a lack of salience in the accrual incentives: agents are not able or willing to take into account the value of future benefit increases when considering the relevant rewards for working.

Keywords: labor supply, retirement earnings test, social security wealth, difference-in-differences

JEL Classification: J14, H55

*This paper is part of the Frisch Centre project "1175 Working life after the Norwegian pension reform" and Statistics Norway project "227020 Push and pull factors in early retirement", both financed by the Research Council of Norway. We thank Magne Mogstad, Asbjørn Rødseth, Knut Røed, and participants at a seminar at CEPAR, University of New South Wales for valuable comments. While carrying out this research, Brinch has been associated with the centre of Equality, Social Organization, and Performance (ESOP) at the Department of Economics at the University of Oslo, and Statistics Norway, Research Department. ESOP is supported by the Research Council of Norway.

†BI Norwegian School of Business. Corresponding author. E-mail: christian.brinch@bi.no

‡Ragnar Frisch Centre for Economic Research

§Research Department, Statistics Norway

1 Introduction

In most public old-age pension systems, benefit levels are a function of earnings during an individual's working career. Therefore, the taxation of labor earnings, whether in the form of payroll taxes, income taxes or employer contributions to social security systems, partly comprises forced retirement savings for the employee, rather than ordinary taxes. The increase in the expected present value of future benefits caused by increased earnings constitutes a substantial proportion of marginal taxes in many earnings brackets, see, e.g., Feldstein and Liebman (2002). Recent research on the effects of the salience of taxation raises concerns about whether agents take into account non-salient features of tax-benefit systems when adjusting their behavior to such systems, see, e.g., Chetty et al. (2009). In relation to the design of public pension systems, this should be a major concern because rewards for work acting through current net earnings are arguably quite salient compared to rewards for work acting through future benefit increases. Indeed, results from the literature on earnings tests suggest a different valuation of taxes and accrual of social security wealth, as people to some extent reduce their labor supply in response to earnings tests of pensions, even when the withheld pensions are only postponed at actuarially fair rates, see, e.g., Song and Manchester (2007). The potential welfare costs of the current design of retirement pension systems are huge if people respond to taxation by reducing their work efforts, but do not take into account the rewards for work through the increase in their social security wealth.

The contribution of our paper is to address the effects of salience in social security benefit systems, comparing compensation for work through current net income and compensation for work through future pension entitlements transmitted through a set of complex rules. We study two recent policy discontinuities in Norway that removed a kink and a notch in the lifetime budget constraint relevant to labor supply decisions by 67 to 69-year-old workers. In the period 2003-2007, old-age pensions of people aged 67-69 were tested against earnings at a 40 percent rate, without any deferral mechanism, starting from earnings of approximately $20,000. The test generates a kink in the lifetime budget constraint. The test was lifted stepwise from 2008 to 2010. Second, the formula for computing old-age pension benefits depends on the number of years of positive "pension points" generated by earnings in excess of approximately $10,000. Until 2006, all individuals aged 67-69 could increase their annual public old-age pension benefits from age 70 by having annual labor earnings above this threshold, a mechanism that generates

a notch, or a discontinuity, in their lifetime budget constraint. This notch disappeared for the majority of 67 to 69-year-olds still in work in 2007, because of maturation of the system: They already had so many years of earnings above the threshold that another year would not earn them any higher pension. The effects on the incentives to work, measured in terms of the average change in the net present value of earnings as a mechanical consequence of the notch and the kink, were of similar magnitude.

Following Feldstein (1995) and subsequent literature, we study labor supply in terms of individual labor earnings rather than by some measure of working hours. A recent branch of literature has focused on how kinks and notches in individual budget sets due to tax-benefit systems can lead to bunching in the income distribution, see e.g. Saez (2010), Slemrod (2013) and Kleven and Waseem (2013), and uses the bunching observed in the data to back out labor supply elasticities. The main challenge in this approach is the construction of counterfactual earnings distributions, i.e., specifying what earnings distributions would look like in the absence of kinks and notches. In this paper, we exploit the removal of the kink and the notch for subgroups of our population in 2007 and 2008 to construct counterfactual earnings distributions in a difference-in-differences framework, using the full empirical earnings distribution function rather than summary statistics of earnings as the outcome variable.

There are few empirical studies comparing the labor supply effects of compensation in terms of current income and compensation in terms of increased social security wealth. Recent extensive literature on earnings tests, see, e.g., Friedberg (2000); Friedberg and Webb (2006); Song and Manchester (2007); Engelhardt and Kumar (2009) for studies based on US data and, e.g., Disney and Smith (2002) for a similar study for the UK, suggests that these forms of compensation are not equivalent. However, these studies of earnings tests are by their nature not able to separate these two effects. The only study we are aware of that explicitly tries to separate the effects is Reimers and Honig (1996). Using data from the US Longitudinal Retirement History Study in the 1970s, they estimated a hazard model of labor market re-entry for elderly men and women. This study has two important limitations, however. First, it only considers the effect on one particular extensive margin, labor market re-entry. As documented in a number of studies, such as Song and Manchester (2007) and Hernæs and Jia (2013), changes in incentives might have a positive effect on the intensive margin without affecting the extensive margin. Second, in Reimers and Honig (1996), both current pensions and social security wealth enter the analysis as exogenous regressors. However, these variables might plausibly be correlated with unobserved

variables that affect the outcome variables, which might lead to possible bias. To the best of our knowledge, our study is the first to use credible exogenous variation to separate the labor supply effects of compensation in the form of current income and social security wealth.[1]

Studying the earnings distribution at ages 67-69 for the population of workers who were still active in the labor market at age 66, we find marked bunching at $20,000, as predicted by the kink in the budget line introduced by earnings testing. Through the removal of the earnings test for 67-year-olds, the bunching in this group disappears, because people move outwards in the earnings distribution. We do not find any bunching at just above $10,000 or any hole in the earnings distribution just below $10,000, as would be predicted by the notch in the budget line introduced by the benefit accrual mechanism. Our more sophisticated difference-in-differences analyses of the earnings distributions confirm these findings. Our finding contrasts somewhat to studies of the effect of accrual on US data, see, e.g., Krueger and Pischke (1992), Coile and Gruber (2007) and Liebman et al. (2009), although the picture is more mixed than it might seem at first sight. In a study of California teachers, Brown (2013) finds statistically significant, but economically quite small effects of accrual. Our results are also consistent with the standard result from earnings test studies, namely that earnings tests have an impact on labor earnings even when actuarially fair deferral mechanisms are in place.

A number of potential explanations exist for why we might find different results from the two policy discontinuities: Different responses may result from agents having different information about the incentives and the incentive changes. Different responses may result from one of the payoffs being immediate and the other payoff being through an income stream, and from agents discounting future income streams at rates that differ (radically) from market rates. In principle, different responses can also result from different magnitudes of the changes. Our approach is to knock down these alternative explanations one by one, suggesting a strong circumstantial case for the importance of salience in a broad sense. Agents are not able or willing to compute and take into account the value of future benefit increases when considering the relevant rewards for working.

In the literature, the effects of the salience have been studied in the context of taxation, not benefits. Experimental evidence suggests that people respond differently to taxes depending on whether the tax-inclusive prices are posted, even when people know the taxes in question (Chetty

[1] A number of papers, notably Rust and Phelan (1997); Gustman and Steinmeier (2000, 2005); French (2005); Van der Klaauw and Wolpin (2008) use structural approaches to study the effect of social security on retirement behavior.

et al., 2009). The magnitude of the consequences makes the decision studied here qualitatively different. One would expect agents to make more refined decisions and take more time to consider their options in the labor market than when shopping in a grocery store. The response that we find to the earnings test indeed indicates that people take time to evaluate how their labor market decisions affect their current disposable income, rather than just looking at the gross income from work without taking lost benefits into account. However, the effects of current work efforts on future disposable income through social security wealth is a particularly non-salient reward for working. The usual argument in the salience literature is that the individual welfare costs of small optimization errors are of second order (because of the envelope theorem), so that agents should not be expected to optimize perfectly when it is costly to do so. In the current study, agents face first order welfare costs. For example, some agents have earnings just below the threshold that yields a substantial increase in future pension benefits. Such behavior only makes sense if it must also be considered costly to figure out the effects of current behavior on future pensions. The agents have received most of the relevant information, although not all of it is easily interpretable. Unlike Chetty et al. (2009), we do not have an opportunity to discriminate between agents not responding to the incentives because they are not able to figure out the incentives and agents who fail to respond despite actually being able to figure out the incentives.

A recent study by Mastrobuoni (2011) on the stepwise introduction of annual Social Security statements in the US finds that the additional information leads to an increase in knowledge about future pensions, but does not change retirement behavior. His explanation is that workers were either already behaving optimally or the additional information was not sufficient to change behavior. Our results point more in the direction of the second explanation, namely that workers are not able to process the available information because of the complexity of the social security system. In contrast, Liebman and Luttmer (2015) find in an experimental setup that providing individuals with more information about social security may affect both knowledge and behavior. However, the knowledge increase that individuals seem to act upon is of a very general sort: for example, "working more years will lead to higher social security benefits", rather than minute details about the system that require processing to figure out how to act upon them.

The next section describes the institutional framework and the policy discontinuities that we will exploit in the empirical analysis. Section 3 discusses the implications of kinks and notches for earnings distributions. Section 4 describes the data and the empirical strategy, while

Section 5 reviews the main econometric framework we use. Section 6 presents empirical results and section 7 discusses possible mechanisms underlying the empirical findings, while Section 8 concludes.

2 Institutional setting and policy discontinuities

2.1 The NIS old-age pension

The backbone of Norway's retirement provision system is a mandatory defined benefit plan, the National Insurance Scheme (NIS) old-age pension system, which provides basic pension coverage from the age of 67 until death for all individuals with a minimum numbers of years of residence in Norway.[2] It roughly corresponds to the Social Security old-age pension scheme in the US and similar schemes in most OECD countries (see OECD, 2011). The policy discontinuities depend on minute details in the system that are particular to Norway.

The NIS old-age pension system is organized around a unit called the Basic Amount. In 2007, one Basic Amount was equivalent to approximately 70,000 Norwegian kroner, which is roughly a fifth of annual average wage income, or very roughly $10,000 (US dollars). For the remainder of the paper, we will, for the sake of simplicity, refer to amounts in US dollars with this simplified conversion. The amounts are comparable between years during the observation period since the Basic Amount was adjusted annually by the average nominal wage growth.

The NIS pension is the sum of an earnings-independent basic pension and a supplementary pension that depends on individual earnings histories. Starting in 1967, individuals are assigned pension points if their annual pensionable income (labor earnings, temporary benefits, or calculated labor earnings for the self-employed) exceeds $10,000. [3] The supplementary pension depends on the earnings history through two distinct factors: the number of years with positive pension points, N, up to a maximum of 40, and the average number of pension points over the best 20 years (AVP). [4] The supplementary pension, SP, is proportional to both factors:

$$SP \sim \min(N, 40) \cdot AVP. \tag{1}$$

[2]The NIS was reformed in 2011, but only the pre-reform system is relevant to our analysis.

[3]An individual receives one pension point per $10,000 in the bracket $10,000-$60,000 and one pension point per $30,000 in the bracket $60,000-$120,000. No points are awarded for earnings below $10,000 or above $120,000.

[4]The AVP in the Norwegian system plays a similar role to the Average Indexed Monthly Earnings (AIME) in the United States' Social Security system, where AIME is based on the best 35 years. However, the AVP is an average of pension points, which are a nonlinear function of earnings, rather than an average of earnings.

Thus, earning pension points can affect future pensions through two distinct mechanisms: by increasing the number of years with positive pension points and by affecting the AVP. Our focus is on the former mechanism, which introduces a notch in the lifetime budget constraint.

2.2 Policy discontinuities in the NIS system

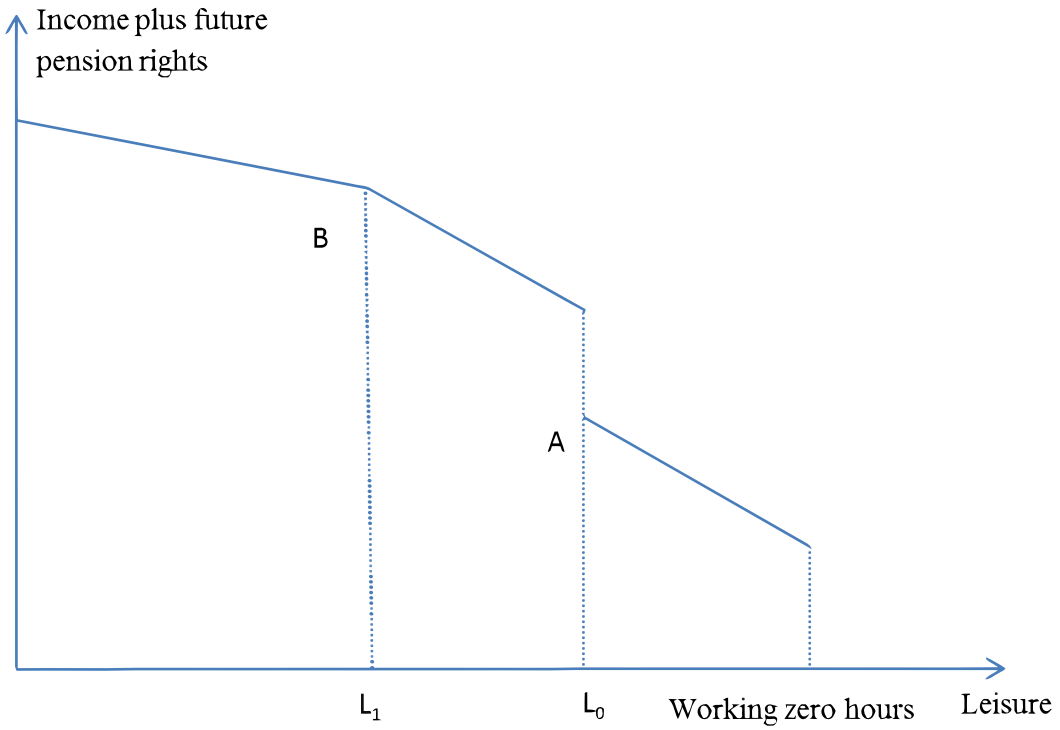
For those with less than 40 years with positive pension points ($N < 40$), one more year with earnings above \$10,000 before the age of 70 will always increase the supplementary pension, by increasing the number of years with positive pension points in the factor in front of AVP in equation (1). The increase in supplementary pensions as a result of moving from just below to just above \$10,000 is equal to 1/40 of the potential supplementary pension. We therefore have a notch in the lifetime budget constraint associated with earning more or less than \$10,000 in each particular year, as illustrated by point A in Figure 1. This notch only exists for those who have not already earned 40 years of pension points. Because the recording of earnings in the system began in 1967, it was not possible before 2007 to have earned 40 years of pension points. Thus, in 2006, all agents considering retirement faced this notch in their budget lines. In 2007, for the first time, those who had worked continuously since 1967 would no longer increase their NIS pensions through the first factor in equation (1) by continuing to work. For them, there is no corresponding notch. Note that this change in the returns to work was not a reform per se, but the consequence of a policy adopted a long time ago, an issue that we return to in Section 6.2.

In the period 2002-2007, an earnings test was applied to the NIS pensions for 67- to 69-year-olds: Labor earnings above \$20,000 led to a reduction in the NIS pension based on a 40 percent rate, without any adjustment of future NIS benefits. Unlike the case in the US system with adjustment via deferral, the Norwegian earnings test can be viewed unambiguously as a tax. The earnings test generates a kink in the individual budget constraints, as illustrated by point B in Figure 1.[5] The NIS earnings test was removed for 67-year-olds in 2008, for 68-year-olds in 2009, and for 69-year-olds in 2010. This reform was motivated by the aim of increasing the elderly labor supply and not by current macroeconomic conditions.[6] The reform was announced just a few months before its introduction, as part of the standard public budget process for 2008.

[5] In principle, there was also a pre-reform kink at a very high earnings level, at the stage when pensions reach 0, though that kink is empirically irrelevant for our analysis.

[6] The international financial crisis did not have a strong impact on the Norwegian labor market.

Figure 1: The kink and notch in the NIS pension system



Note: This graph illustrates the budget constraint before the removal of the notch and the kink in 2007 and 2008. Specifically, working enough hours to generate earnings of $10,000 generates a jump in future benefits (point A), while working enough hours to generate earnings of $20,000 leaves the individual on the part of the budget line where current pension benefits are earnings tested, so that the marginal returns to work are reduced, generating a kink in the budget line (point B).

2.3 Occupational pension and its interaction with the NIS system

In addition to the NIS old-age pension, there are substantial tax-favored occupational pension schemes in both the public and private sectors.[7] The public occupational pension system is a defined benefit system, but still so tightly integrated with the NIS pension system as to almost completely neutralize the policy discontinuities that we study here.

In contrast, the private schemes have no automatic coordination with the NIS. In this age group, the private occupational pensions schemes are dominated by defined-benefit schemes. Even though the schemes usually have certain features in common with the public schemes in targeting a replacement rate of 60-70%, including the NIS pension, they are only coordinated with an industry standard known as "computed NIS" rather than the actual NIS pensions. This standard roughly assumes full coverage from 1967 and after, and every year has the same number

[7]There is also an earlier retirement scheme (AFP) in Norway, which covers roughly half of private sector workers in the relevant period and allows them to retire starting at age 62. For our analysis, this scheme is largely irrelevant.

of pension points as the end year. Thus, the sum of actual pension received from the NIS and the private pension scheme differs from the targeted replacement rate when the actual and computed NIS differ. In particular, unlike the public schemes, the private schemes will not neutralize the policy discontinuities in the NIS old-age pension scheme.

In our main empirical analysis, we therefore only study private sector workers - with and without occupational pension schemes. We carry out corresponding analyses of public sector workers as robustness tests, since any effects found for public sector workers would indicate that our findings for private sector workers were spurious.

3 Labor supply with kinks and notches

The standard neoclassical model for static labor supply has clear predictions of the effect of a kink in the budget line such as that introduced by the earnings test described above. With a convex budget set, those who maximize their utility in the kink point can be characterized by having marginal rates of substitution that can be anywhere between the slopes of the two adjacent line segments. Thus, when aggregating over suitably defined continuous distributions of preferences or wages, there will be a mass point of earnings at the kink point. Allowing for imperfect optimization, there will be bunching around the kink point. For a detailed discussion on bunching in income distribution and kinks in the budget sets, see Saez (2010).

The incentive change relating to the social security wealth accrual mechanism involves a notch in the lifetime budget line, see Slemrod (2013) and Kleven and Waseem (2013), so that the budget set is no longer convex. To find the global optimum, it is necessary to optimize for each individual over each convex subset of the budget set and then compare the local optima. A premium for earning at least $10,000, which we observe, will lead some people who would otherwise earn less than $10,000 to earn $10,000. Assuming leisure is a normal good, individuals who change their behavior as a consequence of the premium will choose to earn exactly $10,000. Allowing for imperfect optimization, we thus expect to see bunching (high earnings density) just above the notch, and a hole (low earnings density) below the level in the earnings distribution associated with the notch. Note that the notch in our empirical analysis does not imply strictly dominated intervals of earnings like the notches in the main analysis of Kleven and Waseem (2013).

In summary, if the elderly can freely adjust their labor supply, the standard economic theory

suggests that there will be bunching in the income distribution above $10,000 and around $20,000 before the changes in incentives. The bunching will disappear as a consequence of the policy discontinuities that remove the kink and notch in the budget line. Removing the earnings test of incomes in excess of $20,000 will lead more individuals to earn more than $20,000. On the other hand, taking away extra social security wealth for having an annual labor income in excess of $10,000 will lead fewer agents to earn more than $10,000.

4 Data and descriptive statistics

We base our analysis on data from merged administrative registers from Statistics Norway covering the full birth cohorts from 1937 to 1941 and containing information on date of birth, gender, education, the date of taking up the old-age pension benefit, accumulated pension rights, and detailed income information observed at an annual level. Our main objective variable, labor earnings, is defined as pensionable income, which essentially includes wages and self-employed earnings and excludes capital income and pensions. As the policy discontinuities affected the 67-69 age group and few old people return to work after they become inactive in the labor market, we study those still working at age 66, operationalized with labor earnings at age 66 of at least $10,000, while not being on the disability pension rolls and not receiving any pension benefits. Table 1 reports the numbers of 66-year-old men who are excluded for different reasons, with some descriptive statistics on earnings and wealth. The analysis population comprises 20 percent of the cohort, since many have exited the labor market through the NIS disability pension scheme or an early retirement program. Because those who are covered by the public occupational pension system will not be affected by or will be affected much less by the incentive changes we study and may differ in important aspects from private sector workers, our main analysis population consists of individuals who work in the private sector or are self-employed. Public sector workers are only used in robustness tests.[8]

[8]We only report results for men for the sake of brevity, but results from a similar analysis for women are reported in a working paper version, see Brinch et al. (2012). The results are largely consistent with what we find for men, but less precise because of a smaller population size.

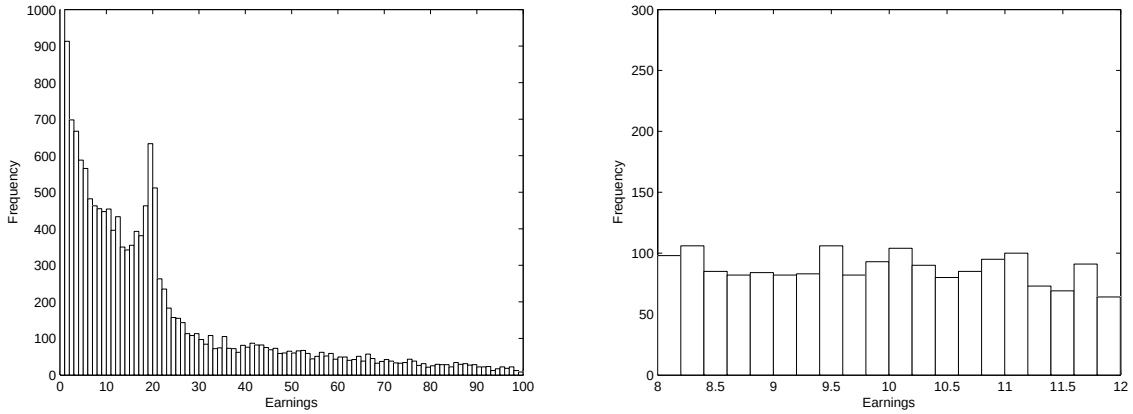
Table 1: Classification of 66-year-old men in Norway, January 1, 2003-2006, by labor market status

	Number	Mean (median) Earnings at age 66	Mean (median) Earnings at age 68	Mean (median) Net wealth
Not active in labor market				
AFP early retirement	18435	6.6 (0.0)	3.2 (0.0)	287 (251)
Disability pension	27425	3.7 (0.0)	2.3 (0.0)	198 (159)
Occupational pension	8956	20.0 (3.3)	10.0 (0.0)	360 (262)
NIS pension	519	15.4 (5.5)	6.5 (0.0)	235 (168)
Some work	259	1.0 (2.3)	16.0 (0.4)	523 (175)
Idle	2316	0.4 (0.0)	2.5 (0.0)	336 (142)
Active in labor market				
Public sector employee	2531	81.5 (72.4)	46.0 (34.3)	383 (316)
Analysis population	10644	66.3 (55.3)	27.1 (14.5)	495 (284)
Full population	71085	18.6 (0.5)	8.9 (0.0)	303 (220)

Notes: The unit of income and wealth is $1,000. AFP early retirement scheme classification does not include those who are only 20 or 40 percent pensioners. Occupational pensions and NIS pensions are defined based on receiving at least $10,000 in pension. Both the groups "some work" and “idle” have less than $10,000 in earnings, while those clarified as “some work” are registered as having an employment contract. Net wealth is the total taxable assets (including housing) in excess of loans.

Figure 2 shows a histogram of the earnings distribution at ages 68-69 for the base population from which we draw our analyses, during the period before the changes in incentives. There is bunching around the $20,000 threshold corresponding to the kink generated by the earnings test, but no bunching around the $10,000 threshold corresponding to the notch generated by the accrual mechanism. The latter observation is reinforced by the right-hand panel in Figure 2, which zooms in on the earnings distribution around $10,000. Based on this preliminary evidence, individuals respond to financial incentives related to current net income, but not to the social security wealth accrual mechanism.

Figure 2: Histograms of earnings distribution for 68- and 69-year-olds, 2003-2006 pooled. The right-hand panel zooms in on earnings in the neighborhood of $10,000.



Note: Histograms of earnings for 68- and 69-year-olds in 2003-2006, for the full population of Norwegian men, conditional on them being active in the labor market at age 66. Earnings measured in $1,000, based on the simplified conversion of one basic amount in the Norwegian national insurance scheme as being equal to $10,000. The leftmost bin in the left-hand panel is truncated and has approximately 5,400 observations.

5 Empirical approach

The descriptive empirical evidence indicates some disparate effects on labor supply from compensation through current net wages and social security wealth accrual. To evaluate these effects more rigorously, we apply a difference-in-differences identification strategy, comparing the change in outcomes for the treatment group in the pre- and post-treatment period with the corresponding change for the control group. The key identifying assumption is then that the mean outcome for the treatment group would, in the absence of treatment, change from the pre- to the post-treatment period in parallel with the change for the control group. Because the two policy discontinuities affect different individuals, we define the treatment and control groups separately for the two analyses.

Because the reforms shift nonlinearities in the budget constraints, we expect an uneven effect over the earnings distribution and analyze the shift in the earnings distribution before and after treatment. A popular approach is to use quantile treatment effects, where the objective is to identify the shift in quantiles of the conditional earnings distribution, see, e.g., Chernozhukov and Hansen (2005) or Athey and Imbens (2006). Our approach is to use an estimator for the effects on the cumulative distribution function of earnings rather than on its inverse, the quantile function. For our analysis, there are two main advantages of working with the distribution function instead

of the quantile function. First, results are easier to interpret. Second, the theoretical predictions from Section 3 are expressed in terms of the earnings distribution function. Our empirical modeling approach is in fact a version of the "distribution regression models" discussed by Chernozhukov et al. (2013). See also Hernæs and Jia (2013) or Havnes and Mogstad (2015) for detailed discussions of this approach and its relationship with standard conditional quantile treatment models.

Denoting the main outcome variable as Y_{it}, we study treatment effects on $\bar{F}(a) = Pr(Y_{it} > a)$ before and after the treatment. To do this, we define a family of derived outcome variables based on the indicator function $Z_{it}^a = 1(Y_{it} > a)$. Now, instead of conducting the difference-in-differences analysis based on Y_{it}, we do so for the derived binary outcome variables Z_{it}^a. In particular, for each choice of a, we apply the following difference-in-differences analysis:

$$E(Z_{it}^a) = P(Y_{it} > a) = \alpha_a + \beta_a X_{it} + \gamma_a \Delta_{it} + \delta_a D_t + \lambda_a(\Delta_{it} \cdot D_t), \tag{2}$$

where X_{it} represents the individual characteristics; $\Delta_{it} = 1$ if individual i belongs to the treatment group in year t, and zero otherwise; and D_t is the dummy variable for the post-treatment period. λ_a captures the treatment effect on the probability that the outcome variable exceeds a. We can then combine the analyses for different a to map out the effect on the entire support of Y_{it}, provided that the identifying assumptions underlying the difference-in-differences approach (most importantly, parallel trends for treatment and control groups in the absence of treatment) are satisfied for all a.

6 Empirical results

6.1 Treatment effects for the earnings test reform

In this section, we present a difference-in-differences analysis of the effects on the income distribution of abolishing the earnings test for 67-year-olds from January 1, 2008. After the repeal of the earnings test for 67-year-olds described in Section 2, earnings testing would again be effective the month after an individual's 68th birthday. Hence, in principle, only those whose 68th birthday was in December 2008 would be exempt from the earnings test during the whole of 2008. However, it was decided in October 2008 that the earnings test for 68-year-olds would also be abolished in 2009, and we observed that the earnings test was not applied to new individuals turning 68 years from October 2008. Hence, for those whose 68th birthday was between October

and December 2008, their entire annual earnings in 2008 were exempt from the earnings test.

Therefore, the treatment group in the earnings test analysis consists of individuals who turned 68 in October-December 2007-2008, with annual earnings in the year they turn 68 as the outcome. Our control group consists of individuals who turned 69 in 2007-2008, with annual earnings in the year they turn 69 as the outcome. The treatment period is 2008, when 69-year-olds were earnings tested, while the 68-year-olds in our treatment group were exempt from the earnings test. Thus, in 2007, both the treatment and control group are subject to the earnings test, while in 2008, only those in the control group are subject to the earnings test. In some of the reported analyses, the data are extended to include pre-treatment data for the years 2004-2006.[9]

Some individuals are in the treatment group in one year and in the control group in the following year. This does not invalidate any point estimates, but it needs to be taken into account when we compute standard errors. We report block-bootstrapped standard errors, with individuals as the unit for bootstrapping.[10]

Table 2 provides descriptive statistics for the treatment and control group among men in the pre- and post-reform periods. Note that the control variables are measured at age 66, while the outcome variables are measured at either 68 (the treatment group) or 69 (the control group). The control and treatment populations look reasonably balanced with respect to the control variables: earnings at 66, the share of self-employed, occupational pension coverage, and education. The outcome variable in the control group, earnings at 69, is lower than the outcome variable in the treatment group, earnings at 68. This is partly because of the reform, as shown below, but also because more of those in the control group have withdrawn from the labor market with zero earnings as they are one year older.

[9]For reasons of power, we do not limit the control group to individuals born between October and December in our baseline specification. This cut-off does not have a large impact on the results. A third alternative would be to limit the control group to individuals born between January and September to avoid the same individuals appearing several times in the data. Such a control group also yields approximately the same results as reported.

[10]Analytical standard errors based on clustering on individuals are similar to what we report.

Table 2: Descriptive statistics of the analysis population for the earning test reform study. Means, with standard deviations for non-binary variables in parentheses, and quantiles where indicated.

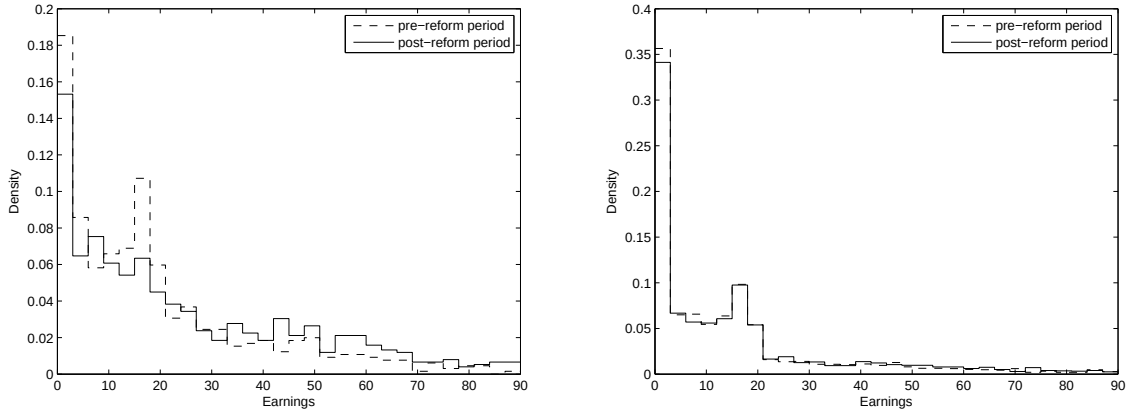
	Treatment group			Control group		
	Before reform		After reform	Before reform		After reform
	2004-2007	2007	2008	2004-2007	2007	2008
Dependent variables						
Earnings (trimmed)	25.120	25.710	33.750	17.880	19.310	20.220
	(32.120)	(31.300)	(36.720)	(29.730)	(30.510)	(30.190)
Earnings > 0	0.866	0.868	0.909	0.696	0.720	0.737
Earnings > 10	0.607	0.640	0.687	0.441	0.480	0.500
Earnings > 20	0.379	0.398	0.509	0.244	0.263	0.286
Earnings > 40	0.198	0.207	0.321	0.134	0.144	0.161
Earnings > 60	0.113	0.112	0.190	0.078	0.086	0.097
1st Quartile	4.320	5.050	7.040	0	0	0
Median	15.700	16.780	20.670	6.630	9.050	10.030
3rd Quartile	30.520	32.210	50.090	19.870	20.360	21.400
Control variables						
Earnings, age 66	61.700	62.640	63.510	61.080	61.220	63.650
	(33.400)	(33.150)	(33.280)	(34.010)	(33.610)	(35.570)
Self-employed	0.302	0.297	0.283	0.300	0.296	0.302
Private occ. pensions	0.341	0.332	0.324	0.359	0.357	0.324
Education in years	11.427	11.593	11.724	11.410	11.533	11.726
	(2.966)	(2.946)	(3.076)	(3.022)	(3.037)	(3.040)
Education > 10 years	0.512	0.525	0.561	0.506	0.520	0.550
Education > 12 years	0.230	0.237	0.258	0.237	0.246	0.261
Education > 14 years	0.122	0.135	0.151	0.126	0.139	0.150
Population size	2393	653	757	9238	2513	2787

Notes: Earnings measured in \$1,000, based on the simplified conversion of one Basic Amount in the Norwegian National Insurance Scheme as being equal to \$10,000. The treatment group is defined as being born in October-December. For the treatment group, the outcome is measured as annual earnings in the year they turn 68. The control group is born throughout the year. The outcome for the control group is measured as annual earnings in the year the workers turned 69. Earnings are trimmed at \$160,000.

Clearly, mean earnings increase from the pre-reform to the post-reform period for the control population, but much more for the treatment population. This particularly reflects a decrease in the proportion earning between \$10,000 and \$20,000 and increasing proportions in the brackets above \$20,000. Figure 3a reports histograms for the immediate pre- and post-reform years for the treatment and control groups. For the control group in the right-hand panel, the pronounced bunching of earnings at \$20,000, where the kink occurs in the budget line, remains both before and after the reform. For the treatment group, the bunching in the pre-reform period almost entirely disappears as the earnings test is removed.

Figure 3: Analysis of earnings test reform.

(a) Histograms of the earnings for the treatment group (left-hand panel), and the control group (right-hand panel). The pre-treatment distributions are from 2007. The post-treatment distributions are from 2008. Earnings measured in $1,000.



(b) The raw difference-in-difference estimates on the distribution function (left-hand panel) and trends in key outcome (right-hand panel) before and contemporaneously with the incentive change (2008). Earnings measured in $1,000.

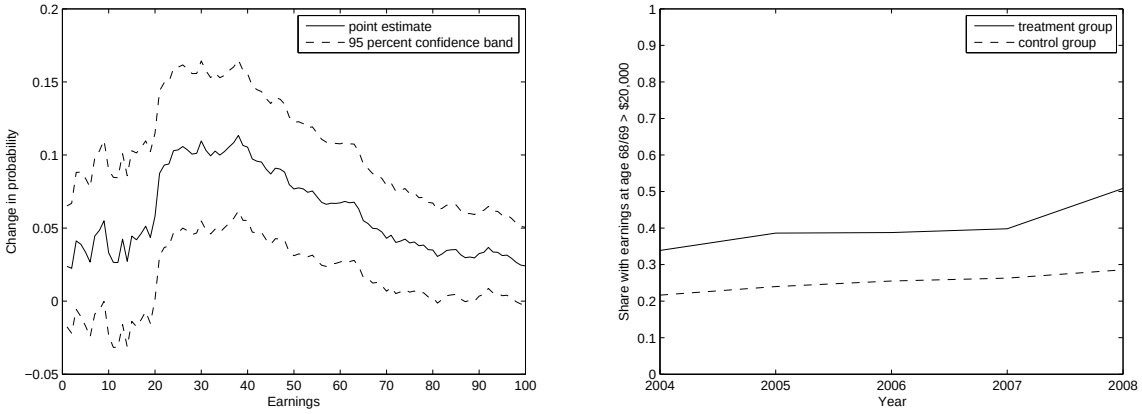


Table 3 reports the results of our difference-in-differences analysis on the earnings distribution. Column 1 shows the estimates without any controls based on data from 2007 and 2008. The first row shows an increase in mean earnings by approximately $7,000. The second row shows the estimated effect on the probability of positive earnings, λ_0 in equation (2), i.e., the extensive margin response, which is positive but not statistically significant. The three next lines show the estimated effects of the earnings test reform on the probability of earning more than $10,000, $20,000, and $40,000, respectively. The effect of the reform kicks in at the $20,000 margin, with an estimated increase of 8.7 percentage points. The similar reported effect on the $40,000 margin shows that this reform effect extends further into the earnings distribution. Corresponding results for all earnings levels are provided in the left-hand panel in Figure 3b,

which clearly shows an effect kicking in at around $20,000 and extending into the upper tail of the earnings distribution.[11]

Column 2 in Table 3 reports the same difference-in-differences effects in a linear regression framework, with controls for a third order polynomial in income at age 66, self-employment at age 66, availability of private occupational pensions, and dummies for highest completed education. The controls hardly change the raw difference-in-differences estimates. Column 3 reports results including the longer pre-reform time series, and including the data from 2004-2006. The inclusion of these additional data gives no substantial changes to the results.

Table 3: Difference-in-differences estimates of the effects of the earnings test reform

Outcome variable	Raw estimates	Estimates with controls	Estimates with controls and ext. pre-reform period	Estimates with controls and separate linear trends
Earnings (trimmed)	7.140	7.760	6.760	7.740
	(1.960)	(1.670)	(1.390)	(1.970)
Earnings > 0	0.024	0.025	0.003	0.020
	(0.021)	(0.021)	(0.017)	(0.025)
Earnings > 10	0.027	0.030	0.024	0.006
	(0.029)	(0.029)	(0.022)	(0.033)
Earnings > 20	0.087	0.092	0.091	0.095
	(0.030)	(0.028)	(0.022)	(0.033)
Earnings > 40	0.097	0.103	0.099	0.102
	(0.025)	(0.023)	(0.019)	(0.027)
Controls		x	x	x
Data from 2004-2006			x	x
Group-specific trends				x

Notes: Earnings measured in $1,000. The controls enter the model as listed in Table 2, except "earnings at age 66", which enter the models as a third order polynomial and "education in years", which is left out, since education is measured in terms of the specified dummy variables. Block-bootstrapped standard errors in parentheses.

The main assumption in the difference-in-differences analysis is the parallel trends assumption. Although control groups and treatment groups are allowed to differ, the changes in the outcomes for the treatment group are assumed to equal the changes for the control group from the pre- to the post-reform period in the absence of a reform. The primary test of this assumption is to check the pre-reform trends. Indeed, this is the primary reason why we included a longer pre-reform period in the analysis. Column 4 reports the difference-in-differences estimates

[11] While it is certainly not difficult to get a feel of the magnitude of the effects from looking at the left-hand panel in Figure 3b, some readers may want to see an elasticity for easy comparison with other analyses. It is not so easy to derive a substitution elasticity of labor supply with respect to net earnings from the estimation results provided here. However, based on Saez (2010), it is possible to estimate an elasticity based on the amount of bunching observed. It is clear that, in our case, such an elasticitywould be between 0.5 and 1. The kink reduced the net-of-tax rate by 40 percent, and the marginal buncher in the sense of Saez (2010) in our analysis decreases his earnings by somewhere in the order of $10,000, from a base of $30,000, as a consequence of the kink. This calculation is only meant to be illustrative. Any exact estimate of the elasticity would require more technical assumptions.

when different linear trends for the treatment and control groups are included as controls in the model. These trends are identified by the pre-reform periods only. The results reported in Column 4 are not much different from those reported in other columns, which suggests that there is not a statistically significant difference in trends for the treatment and control groups. This conclusion is also confirmed by visual inspection of the data. The right-hand panel in Figure 3b shows the trends in the share of individuals with earnings greater than $20,000 and earnings at age 66 over the period from 2004 to 2008. As we can see, there are no indications of different trends between the treatment and control groups before the reform.

We have also run two additional robustness checks. The first is a placebo treatment test, in which we conduct similar difference-in-differences analyses, except that data for 2005 and 2006 are used instead of data for 2007 and 2008. There was no reform in 2006 and, if we found a positive effect, this would point to spurious factors that might also generate the reform effects reported in Table 3. The second robustness test uses the correct reform years, but studies workers who were employed in the public sector at age 66. As discussed in Section 2, these subjects were excluded from the primary analysis population because the incentive changes they faced were different from and much weaker than for the main population in the analysis. We should thus expect to find weaker or no reform effects for the public sector workers. In both robustness checks, we find no significant effect, which supports our model specifications.[12]

Note that, in our analysis sample, those who do not face an earnings test will not face an earnings test the next year either. Our reform effect therefore captures not only the direct effect of the earnings test at age 68 but also an indirect effect: people might continue working to keep their jobs and to take advantage of the absence of an earnings test the next year, and potentially the year after that, if they also guessed right about the policy for 2010.

6.2 Effects of the accrual change

This section presents a difference-in-differences analysis of the consequence of the maturation of the social security wealth accrual mechanisms as set out in Section 2. The treatment group consists of individuals who have had earnings in excess of $10,000 in every year from 1967 up to and including the year they turn 66, and who turn 68 or 69 in 2006 or 2007. The control group consists of individuals without such a complete earnings history who reach 68 or 69 years of age in 2006 or 2007, like the treatment group. The outcome measured is annual earnings in

[12]Results are not tabulated here, but can be found in Table 4 in Brinch et al. (2012).

2006 or 2007. In 2006, all individuals could increase their future NIS pension benefits by earning in excess of $10,000, thus adding one year of pension points, see equation (1), and this is the pre-treatment period in the analysis. In 2007, those with a complete earnings history, at least $10,000 every year from 1967, could no longer increase their pensions through the mechanism of adding more years with earnings in excess of $10,000. The post-treatment period is therefore 2007. In some of the reported analyses, the pre-treatment data are extended to also include the years 2004-2005.

It should be noted that this setup is based on a simplified model, where individuals aged 67-69 do not re-enter the labor market once they become inactive in that market. Allowing for "re-entering" opens the possibility of partial treatment. For example, in the year 2007 (the year of treatment), a 68-year-old belonging to the treatment group could still have the additional incentive for work if she/he were inactive, i.e., earned less than $10,000 at age 67 (in the year 2006). If she/he actually acts upon this incentive and "re-enters" the labor market by earning more than $10,000 in 2007, our estimation results will be biased. Fortunately, as one might expect, such "re-entering" behavior is rather rare. In fact, among those with exactly one year of earnings below $10,000 when aged 67-69, less than 5 percent had this gap at age 67.

Table 4 shows descriptive statistics for the treatment and control groups. The analysis sample is slightly larger than that in the analysis of the earnings test reform. A majority of approximately two-thirds of our sample have earnings in excess of $10,000 every year back to 1967 and belong to the treatment group. The treatment group has higher earnings and more self-employed than the control group at age 66. The control group has more employees with access to private occupational pensions. The residual group of employees without access to private occupational pensions is approximately the same size. In terms of educational level, the treatment and control groups are similar. There are no dramatic changes in the control variables for either the treatment or the control group from the pre-reform to the post-reform period. In terms of the outcome variable, income at age 68 and 69, the control group has a slightly higher mean throughout. The mean earnings are quite a bit higher than the threshold of $10,000. Both the treatment and control group show a slight increase in mean earnings from the pre- to the post-treatment period.

Table 4: Descriptive statistics of the analysis population for the social security wealth accrual change study. Means, with standard deviations for non-binary variables in parentheses, and quantiles where indicated.

	Treatment group			Control group		
Outcome in	2004-2006	2006	2007	2004-2006	2006	2007
Dependent variables						
Earnings (trimmed)	19.400	19.930	21.450	21.300	22.000	23.860
	(30.080)	(29.600)	(30.190)	(32.070)	(31.690)	(33.730)
Earnings > 0	0.754	0.763	0.788	0.771	0.790	0.788
Earnings > 10	0.481	0.502	0.536	0.503	0.526	0.545
Earnings > 20	0.275	0.286	0.313	0.308	0.322	0.338
Earnings > 40	0.147	0.153	0.171	0.164	0.169	0.191
Earnings > 60	0.085	0.092	0.098	0.097	0.103	0.121
1st Quartile	0.190	0.360	0.740	0.460	1.050	1.150
Median	8.820	10.120	12.200	10.140	11.310	12.300
3rd Quartile	21.020	21.590	23.490	24.070	24.980	28.060
Control variables						
Earnings, age 66	64.070	63.570	65.280	53.790	54.200	57.200
	(3.268)	(3.226)	(3.253)	(3.548)	(3.608)	(3.785)
Self-employed	0.352	0.356	0.355	0.186	0.193	0.192
Private occ. pensions	0.285	0.281	0.256	0.514	0.503	0.498
Education in years	11.427	11.515	11.529	11.369	11.584	11.838
	(2.985)	(3.008)	(2.937)	(3.106)	(3.131)	(3.219)
Education > 10 years	0.513	0.520	0.531	0.491	0.513	0.547
Education > 12 years	0.237	0.241	0.236	0.236	0.258	0.288
Education > 14 years	0.122	0.132	0.131	0.133	0.150	0.170
Population size	9497	3253	3474	4261	1530	1826

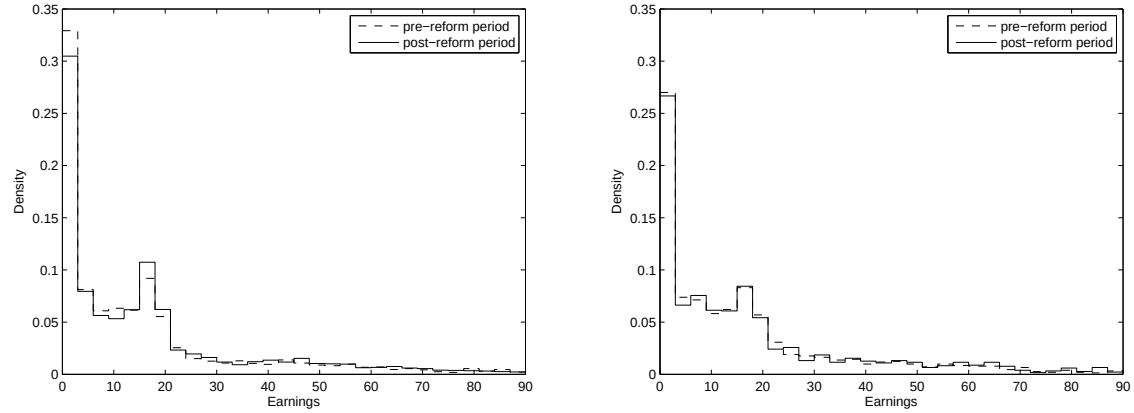
Notes: Earnings measured in \$1,000, based on the simplified conversion of one Basic Amount in the Norwegian National Insurance Scheme as being equal to \$10,000. The treatment group is defined as having an earnings history with at least \$10,000 in annual earnings from 1967 until age 66, while the control group does not have such an earnings history. For both the treatment and the control groups, the outcome is measured as annual earnings in the years they turn 68 and 69. Earnings are trimmed at \$160,000.

Figure 4a reports histograms of earnings for the treatment and control groups in the pre- and post-treatment periods. We see the bunching at \$20,000 associated with the earnings test, which was still in place in both 2006 and 2007. Note that, before the reform, both the control and treatment groups have incentives to maintain earnings of at least \$10,000 because of the existence of the notch in the budget lines, as shown earlier in Figure 1. If individuals respond to these incentives, we should have observed bunching at approximately \$10,000 according to the continuous choice model. However, we do not see any bunching associated with the social security wealth premium for earnings of at least \$10,000. This indicates that, for some reason, the individuals in our sample do not adjust their work efforts as a result of potentially increased future pension entitlements. Thus, it would be very surprising if a positive result turned up in the difference-in-differences analysis. The results from the difference-in-difference analyses are most

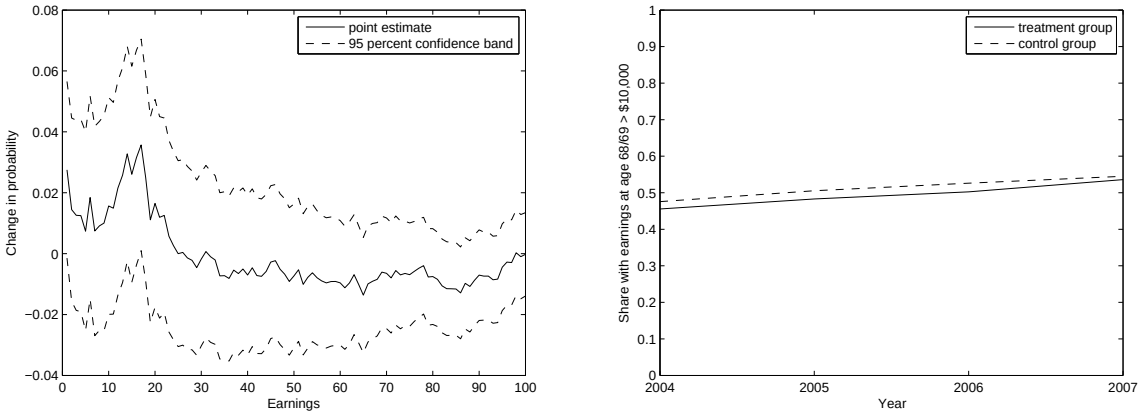
easily seen from the the raw distributional difference-in-differences graphs in the left-hand panel in Figure 4b. If anything, the treatment group shows a slight increase in the number of people earning just above $10,000, which is the opposite of the predicted effect of the incentive change. However, this change is not statistically significant. It should be clear from the confidence bands that we can rule out any strong effect of the incentive change associated with the change in the accrual incentives.

Figure 4: Analysis of social security wealth accrual change.

(a) Histograms of the earnings for the treatment group (left panel), and the control group (right panel). The pre-treatment distributions are from 2006. The post-treatment distributions are from 2007. Earnings measured in $1,000.



(b) The raw difference-in-difference estimates on the distribution function (left-hand panel) and trends in key outcome (right-hand panel) before and contemporaneously with the incentive change (2007). Earnings measured in $1,000.



Detailed results based on different model specifications corresponding to those used for the analysis of the earnings test are reported in Table 5. We do not find any statistically significant effects in any of the analyses, even though the precision of our estimates is quite good. The right-hand panel in Figure 4b shows long-term trends in the shares of the treatment and control

groups that earn in excess of $10,000 and it demonstrates that there are no strong pre-reform trends specific to either the control or treatment groups that could work to cancel out the estimated effects of removing the notch. We also ran the same robustness tests as for the earnings test, using 2004-2005 as the analysis years for the placebo test. Statistically significant treatment effects in the robustness tests would indicate that some assumption is violated, which might hide any effects that should have been evident in the main analysis. However, we do not find any effects of the placebo treatment based on the years 2004-2005 or for the public sector workers.[13] The bottom line is that there is no discernible effect of the change in social security wealth accruals and that this can be stated with some precision.

Table 5: Difference-in-differences estimates of the effects of the social security wealth accrual change

	Raw estimates	Estimates with controls	Estimates with controls and ext. pre-reform period	Estimates with controls and separate linear trends
Earnings (trimmed)	-0.340	0.130	1.170	0.190
	(0.99)	(0.90)	(0.94)	(1.24)
Earnings > 0	0.028	0.024	0.021	0.019
	(0.015)	(0.015)	(0.014)	(0.020)
Earnings > 10	0.015	0.015	0.026	0.012
	(0.017)	(0.017)	(0.016)	(0.023)
Earnings > 20	0.012	0.013	0.023	0.018
	(0.016)	(0.015)	(0.015)	(0.021)
Earnings > 40	-0.005	-0.001	0.011	-0.006
	(0.013)	(0.012)	(0.012)	(0.016)
Controls		x	x	x
Data from 2004-2005			x	x
Group-specific trends				x

Notes: Earnings measured in $1,000. The controls enter the model as listed in Table 2, except "earnings at age 66", which enter the models as a third order polynomial and "education in years", which is left out, since education is measured in terms of the specified dummy variables. Block-bootstrapped standard errors in parentheses.

In contrast to the earnings test reform, which was announced only a few months before implementation, the information about the accrual mechanisms was publicly available as early as 1967, which was when the system was established. Individuals have clearly had an opportunity to adjust their labor supply behavior to take advantage of the implicit rule change in 2007, and this might cause self-selection into the control and treatment groups. From 2007, individuals with "gaps" in their working history before 66 years of age can still receive full old-age pensions after age 70. Thus, it is less costly for persons in the post-treatment cohorts to have gaps in their earnings histories, and workers might self-select themselves into the control group. However, the

[13] Results are not tabulated here, but can be found in Table 8 in Brinch et al. (2012).

gap is only less costly if these workers plan to close that gap by working from the age of 67 to 69. If some individuals move from the treatment to the control group as a consequence of anticipating the policy discontinuity, these individuals plan to work from the age of 67 to 69. We would expect that those who plan to work from the age of 67 to 69 actually act on their plans and have higher earnings when aged 67-69 than the average member of the control groups. Thus, for the the post-reform cohorts, such self-selection should lead to higher earnings in the control group and lower earnings in the treatment group. If such self-selection is prevalent, we will have a negative bias in the estimate of the effect of treatment. Thus, since our predicted effect is negative, our difference-in-differences estimates will be biased away from zero. However, even with this bias, our estimates fail to pick up any significant negative effect.

7 Magnitudes and mechanisms

Our estimates suggest that the kink introduced by earnings testing had a strong effect on labor supply, while the notch introduced by the social security wealth accrual mechanism had no labor supply effects. The two policy discontinuities affected almost identical groups: 67- to 69-year-olds who were working in the private sector at age 66, either as wage earners or self-employed. A subtle difference is that the population of treated only consists of individuals with complete earnings histories in the accrual analysis. Different responses in our analyses could result from different responses to incentive changes between those with and without complete earnings histories. However, auxiliary analyses, not reported here, show no differences in the effects of the earnings test reform between those with and without complete earnings histories. The main differences between the incentive changes is that the earnings test removal affects immediate net earnings, while the accrual discontinuity affects a future benefit stream. Below, we explore various potential explanations of the discrepancy of our results.

Frictions

A first potential explanation of the lack of response to the accrual discontinuity is that labor market frictions and rigidities prevented the elderly from freely adjusting their labor supply. However, the strong effect of the earnings test suggests that the labor market for the elderly is flexible enough to allow adjustments in response to financial incentives, at least around the earnings test threshold of $20,000. It is still conceivable that labor market institutions, such as

working hours regulations and minimum wage regulations, may affect different segments of the labor market differently, so that adjustments around the accrual threshold of $10,000 are quite difficult. However, in a study of an earnings test reform in Norway in 2002, Hernæs and Jia (2013) found that the elderly were able to adjust their earnings around the $10,000 threshold. One could also imagine a different kind of friction, where it is easy to make small adjustments in labor earnings, but harder to make large adjustments. For most individuals, responding to the accrual mechanism would require a large adjustment in labor earnings as it would be necassary to cross the $10,000 threshold, while responding to the kink from the earnings test does not require large changes in earnings. However, our results suggest both that the adjustment in labor earnings in response to the earnings test changes are large and that even those in the neighborhood of the $10,000 threshold do not respond to the notch (see Figure 2, right panel).

The magnitude of changes

A second potential explanation of the discrepancy of the results is that the earnings test reform could result in stronger changes in the incentives for work compared to the accrual mechanism changes. Here, we report our best effort to quantify and compare the magnitude of the effects on the incentive to work. We quantify the incentives to work based on differences between net (after tax) financial rewards for work, for a given earnings distribution, for budget lines with and without the kink or the notch.[14] However, there is a "baseline" problem, that is, the results could depend on whether we base our calculation on the pre- or post-treatment earnings distributions. This is particularly an issue for the earnings test, since the empirical analysis above shows that the earnings test significantly reduces the elderly labor supply. Using the pre-treatment earnings distribution will underestimate the changes in incentives, which are increasing in earnings. We therefore study the mechanical effects of the kink on net income based on the post-treatment earnings distribution for the treatment group in the study, 68-year-olds. If anything, this approach will overestimate the effects of the kink on the rewards for work.[15]

For the accrual mechanism, the choice of which earnings distribution we use for computing the incentive effects does not matter in practice, since the pre-treatment and post-treatment distributions are very similar. To be consistent with the case of the earnings test, we also define

[14] Our incentive measure corresponds to what would be in the numerator of a participation tax rate. The nature of notches rule out comparisons based on the marginal rewards for work.

[15] We get an unbiased estimate of the share of the (relevant) population that is affected by the earnings test, but we get an upper bound on the effect on earnings. In comparison, if we used the pre-treatment earnings distribution, we would estimate a lower bound on both.

the change in the mechanical net (after tax) financial loss due to the accrual changes, based on observed income levels in the absence of the accrual loss (pre-treatment period).[16] In particular, those who choose to earn less than \$10,000 are not affected by the accrual changes and will have zero loss. For those who are affected, the calculation is more complicated than in the earnings test case, since we need to evaluate a permanent change in the future pension payout stream, which depends crucially on how the individuals discount future income streams and take into account mortality risk. We measure the change in the rewards for work by the expected present value (EPV), evaluated at the time of the labor supply decision, t_0, of the loss in future disposable income due to the accrual change:

$$EPV = \sum_{t=70}^{T} P(t|t_0) \frac{\triangle I}{(1+r)^{t-t_0}}, \tag{3}$$

where $P(t|t_0)$ is the probability of surviving to age t conditional on being alive at age t_0. $\triangle I$ is the annual loss in after-tax pension income. The (nominal) discount rate r is set to be equal to the nominal wage growth rate, which is usually higher than the market interest rate in Norway. The survival probabilities $P(t|t_0)$ are taken from the life table of Norwegian males in 2010. As the subgroup still at work at age 66 is typically of better health than the population average, they will likely have higher survival probabilities. Given the higher discount rate and lower survival probability used in the computation, we expect the calculation above to give us a value that is lower than the actuarially fair valuation.

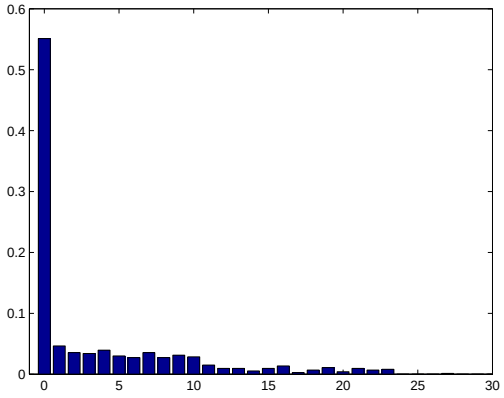
Figure 5 presents distributions of the changes in (the expected present value of) the net incomes as a consequence of the kink and the notch, as described above. While the losses due to the earnings test vary quite a bit from \$0 to \$25,000, the losses due to the accrual change are more concentrated. For the earnings test, around 45 percent of the analysis population was potentially affected since they earned more than \$20,000 in the post-treatment period. On average, they would have lost \$7,600 from the earnings test as a mechanical effect. For the NIS maturation, around half of the treatment population was affected by the change, since they earned more than \$10,000 in 2006, when they earned pensions rights for having such earnings. On average, they would have lost \$6,100 from the loss of pensions accrual. Hence, in terms of the average effect on net income, the two incentive changes are of comparable magnitude, especially taking into account our conservative valuations of social security wealth and the potential overestimation

[16] We have also calculated the change based on the post-treatment income. As we expected, the results are quite similar.

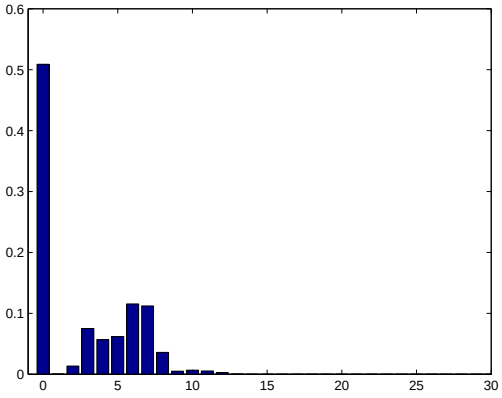
of the change associated with the earnings test kink.

Figure 5: Histograms of distribution of mechanical losses in net earnings and social security wealth from the two incentive changes.

(a) Earnings test



(b) Social security wealth accrual



Note: The loss from the earnings test is measured as the mechanical loss in net earnings based on the earnings distribution in 2008, when no earnings test was in place. The loss from the social security wealth accrual is measured as the mechanical loss in net present value based on the earnings distribution in 2006, when all individuals would accrue social security wealth through labor earnings. Losses measured in \$1,000, based on the simplified conversion of one Basic Amount in the Norwegian National Insurance Scheme as being equal to \$10,000.

Different responses to kinks and notches

A third potential explanation of the lack of response could be that agents respond more strongly to kinks than to notches in the labor market, even if the average mechanical effect of changes is of a similar magnitude, as shown in Figure 5. However, as demonstrated by Kleven and Waseem (2013), it takes only a very small notch to generate discernible bunching in the earnings distribution. The reason for this is quite intuitive. The individual welfare cost of moving from just below to just above the notch, not taking into account the notch, is second order because small departures from an optimal allocation only give second order welfare costs. The welfare gain from moving from just below to just above the notch as a consequence of the notch is first order. For large labor supply elasticities, such as those suggested by our earnings test analysis, we would expect to see effects on the earnings distribution for notches of quite small magnitude.

Rate of time preferences and uncertainty

A fourth potential explanation of the discrepancy of our results is that the agents' subjective valuations of future income streams might be much lower than the expected present values

we calculated earlier, which would reduce the response. A lower valuation could be due to the individuals simply having higher subjective discount rates, or "presence bias", such as hyperbolic discounting.

Note that, with access to a complete credit market, subjective valuations of future incomes that differ from market interest rates would not be an equilibrium outcome: Agents would correct such discrepancies through the credit market. Of course, this argument relies crucially on the assumption that the individuals are not credit constrained. The individuals we study are on average rather well-off, as seen from Table 1, with relatively high above-average incomes before reaching retirement age and, obviously, rather long working careers. Table 6 provides more data on their financial position in 2010 (reliable data on the value of residential properties are not available before this year). We see that our analysis population largely consists of well-off people who to a large extent own their residences and have mortgages of much lower value than their residences. The main reasons are increased property prices and the repayment of mortgages over the years. In addition, Norwegians who are active in the labor market at the end of their 60s have also accumulated wealth in other types of assets. The assets level is substantial, even when assets relating to home ownership are excluded. Thus, there is essentially no reason to expect a large number of our subjects to be rationed in the credit market.

Table 6: Assets and house ownership in 2010 by labor market status at age 66.

	Taxable non-housing assets	Loan	Share of home owners	Net wealth	Share with positive net wealth
Not active in labor market					
AFP early retirement	94 (55)	33 (2)	0.83	287 (251)	0.96
Disability pension	74 (32)	33 (3)	0.66	198 (159)	0.89
Occupational pension	118 (52)	46 (2)	0.57	235 (168)	0.88
NIS pension	389 (65)	78 (4)	0.58	523 (175)	0.87
Idle	207 (37)	43 (0.2)	0.50	336 (142)	0.85
Some work	170 (72)	46 (2)	0.78	360 (262)	0.94
Active in labor market					
Public sector employee	134 (76)	60 (13)	0.85	383 (316)	0.95
Analysis population	312 (90)	64 (6)	0.72	495 (284)	0.93
Full population	137 (52)	41 (3)	0.73	303 (220)	0.92

Notes: Units: \$1,000. Medians reported in parentheses. Non-housing taxable assets include all financial assets (savings, stocks, and funds) in home and foreign countries, real estate ownership (farms, forests) for business purposes, ownership of motor vehicles and leisure boats, etc. The net wealth is the total taxable assets (including housing) in excess of loans.

In the valuation provided above, we do not take into account policy risks that might induce individuals to downwardly adjust their valuations of accumulated pension rights. However, it seems highly unlikely that benefit rules for the elderly individuals we studied (aged 67-69) will

change significantly. For example, individuals born before 1953 were largely unaffected by a comprehensive 2011 Norwegian pension reform, and the Norwegian government is convincingly solvent, with a huge sovereign wealth fund and large revenues from oil extraction. Individual valuations of the benefit stream may also differ from the valuation above, because the notch is a life annuity, which might have an insurance value. Although the NIS system and occupational pensions provide substantial life annuities, it seems unlikely that life annuities should be systematically valued much lower than their expected values.

Information provision

A fifth potential explanation of the discrepancy of our results is that the agents may not be informed about the relevant incentives. However, a few months before reaching their 67th birthday, everybody receives a letter from the local social security administration office with an attached 16-page information folder about relevant pension rules. The information necessary to figure out the financial incentives for working after retirement is included in these folders. Due to the complexity of the accrual rules, the potential accrual is only described in the folder, although it is briefly mentioned in the letter, while the earnings test is accurately described in both the cover letter and the folder. Based on the information that is handed out, one can obtain the relevant details of the pension system, that is, whether benefits will be earnings tested and whether continued working will earn you more pension entitlements.

Lack of salience in social security incentives

Because we find no other reasonable explanation for the discrepancy of our results, we interpret the lack of response to the accrual mechanism as likely to be driven by a lack of willingness or ability to unravel the consequences of the system, or, alternatively, because of a prior belief that the effort of studying how the system works will not lead to substantial net gains through adjustment of individual behavior. In this context, the main difference between the two changes in incentives that we study is that it is easy to understand the earnings test, while it might be more demanding to correctly calculate the value of a gain in terms of increased social security wealth. This difficulty might be caused by the relative complexity of the system that determines future pension entitlements. Since a precise valuation of the gain of being active for one extra year is not available due to the non-salience of the pension rules, individuals may have relied on some crude approximations that greatly underestimated the value. In other words, we conclude

that the most likely explanation of the lack of response to the social security wealth accrual is lack of salience in the incentives implicit in the social security wealth accrual mechanisms.

Note that we use the concept of salience in a slightly broader sense than Chetty et al. (2009). In their experimental set-up, they can distinguish between the mechanisms whereby individuals are unable to calculate the relevant prices and whereby individuals do not respond to the relevant prices even though are able to calculate them. They pinpoint the mechanism to the fact that agents do not respond to prices even when they are able to calculate them correctly. In our non-experimental retrospective setup, there is no way we can separate these mechanisms. In an experimental setup, we would have set up several randomized groups, with the reference group receiving the same information as in our analysis. We could then have used several treatment groups to try to pinpoint where the differences appear between the subjective and objective valuations of the pension accrual. As an example, we could have had one group receiving extra information on the potential pension levels with and without one year of extra earnings in excess of $10,000, and another group that, in addition, received information about the expected present value of the differences between these benefit streams.

8 Conclusion

Using our difference-in-differences approach on the earnings distribution, we find statistically significant and economically substantial effects of an earnings test reform for 67-year-old men. A visual inspection of the earnings distributions before and after the reform reaffirms this finding, with bunching in the earnings distribution at the kink point introduced by earnings testing disappearing as the earnings test is removed. Thus, the earnings test has a sizable impact on those who are still active in the labor market at age 66. Several robustness tests bolster our results, as we are not able to find results in placebo setups where agents are not affected by the reform.

In contrast, we do not find any effect of the changes in the incentives in terms of social security wealth accrual. Before the incentive change, there appears to be no bunching in the earnings distribution above the threshold that provides a substantial boost in the value of future pension benefits. This observation alone suggests that there might be no change in the earnings distribution when this threshold and boost disappears. Indeed, our statistically insignificant point estimates of the effects even show the wrong sign. These null results are particularly

interesting in light of the strong responses to the earnings test reform for the same groups, which show that the lack of response is not simply a consequence of labor market frictions for the elderly.

The two changes in incentives are of the same order of magnitude. The main differences between the incentive changes is that one affects an immediate sum while the other affects a future income stream and that, while both work through the social security old-age pension system, the former might be considered much simpler, more transparent and thus more salient than the latter. The results show that only the former incentive has a strong effect on behavior. The groups affected by the two changes in incentives are also just about identical: 67- to 69-year-olds who were working in the private sector at age 66, either as wage earners or self-employed. We would imagine that the group we study would be close to the perfect group for actually finding labor supply effects of social security wealth accrual. The benefits streams are quite close in time to the decisions and the group that we study, Norwegians active in the labor market at age 66, are reasonably wealthy and healthy.

We interpret lack of salience in the social security wealth accrual as being the most likely reason for the disparate effects on elderly labor supply behavior that we have found. Although potential pensioners are informed about the detailed rules, they were not able to make use of this information to allocate optimally, most likely as a result of the complexity of these rules. Our results are relevant beyond the Norwegian context, given the similarity in social security old-age pension systems across many OECD countries. While the Norwegian pension accrual rules in place at the time of our study were undeniably complex, so are the corresponding rules in other countries. It would be rash, however, to conclude that social security wealth accrual never has an effect on labor supply. Our interpretation points more in the direction that it cannot be taken for granted that rewards for work in the form of accrual have the same effect on the labor supply as more direct rewards for work. Our analysis is a particularly stark case where we would expect to find effects of accrual, but find none.

Arguably, the most difficult aspect for potential retirees is the evaluation of a not clearly specified future income stream for an unknown number of years under not yet decided tax systems. This aspect is intrinsic to the evaluation of pension entitlements. Our results indicate that taxing current income, while at the same time offering incentives to work in the form of future pension entitlements - which is a major feature of most public old-age pension systems - might lead to deadweight loss because of the non-salience of social security wealth accrual

incentives. A potential remedy that would not require large-scale changes to pension systems would be to improve the salience of the accrual mechanisms. However, one would then need to know more about what actually works. Chan and Stevens (2008) combine self-reported data on pension incentives with employer-reported and administrative data. They show that the "perceived" incentives are powerful predictors of behavior, even after controlling for "actual" incentives. This supports the case for further studies on how to improve perceptions.

References

Athey, Susan, and Guido W. Imbens. 2006. Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, 74, no. 2: 431–497.

Brinch, Christian N., Erik Hernæs, and Zhiyang Jia. 2012. Labor supply on the eve of retirement. Disparate effects of immediate and postponed rewards to working. Discussion Papers no. 698, Research Department, Statistics Norway.

Brown, Kristine M. 2013. The link between pensions and retirement timing: Lessons from California teachers. *Journal of Public Economics*, 98: 1–14.

Chan, Sewin, and Ann Huff Stevens. 2008. What You Don't Know Can't Help You: Pension Knowledge and Retirement Decision Making. *The Review of Economics and Statistics*, 90, no. 2: 253–266.

Chernozhukov, Victor, Ivn Fernndez-Val, and Blaise Melly. 2013. Inference on Counterfactual Distributions. *Econometrica*, 81, no. 6: 2205–2268.

Chernozhukov, Victor, and Christian Hansen. 2005. An IV model of quantile treatment effects. *Econometrica*, 73, no. 1: 245–261.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. Salience and Taxation: Theory and Evidence. *American Economic Review*, 99, no. 4: 1145–1177.

Coile, Courtney, and Jonathan Gruber. 2007. Future Social Security Entitlements and the Retirement Decision. *The Review of Economics and Statistics*, 89, no. 2: 234–246.

Disney, Richard, and Sarah Smith. 2002. The labour supply effect of the abolition of the earnings rule for older workers in the United Kingdom. *The Economic Journal*, 112, no. 478: C136–C152.

Engelhardt, Gary V., and Anil Kumar. 2009. The repeal of the retirement earnings test and the labor supply of older men. *Journal of Pension Economics and Finance*, 8, no. 04: 429–450.

Feldstein, Martin. 1995. The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act. *Journal of Political Economy*, 103, no. 3: 551–572.

Feldstein, Martin, and Jeffrey B. Liebman. 2002. Social security. In *Handbook of Public Economics, Vol 4*. eds. by Alan J. Auerbach, and Martin Feldstein. Amsterdam: Elsevier.

French, Eric. 2005. The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour. *The Review of Economic Studies*, 72, no. 2: 395–427.

Friedberg, Leora. 2000. The labor supply effects of the social security earnings test. *Review of Economics and Statistics*, 82, no. 1: 48–63.

Friedberg, Leora, and Anthony Webb. 2006. Persistence in labor supply and the response to the social security earnings test. Center for Retirement Research at Boston College Working Paper no. 27.

Gustman, Alan L., and Thomas L. Steinmeier. 2000. Retirement in Dual-Career Families: A Structural Model. *Journal of Labor Economics*, 18, no. 3: 503–545.

Gustman, Alan L., and Thomas L. Steinmeier. 2005. The Social Security early entitlement age in a structural model of retirement and wealth. *Journal of Public Economics*, 89, no. 2-3: 441–463.

Havnes, Tarjei, and Magne Mogstad. 2015. Is universal child care leveling the playing field? *Journal of Public Economics*, 127: 100 – 114.

Hernæs, Erik., and Zhiyang Jia. 2013. Earnings Distribution and Labour Supply after a Retirement Earnings Test Reform. *Oxford Bulletin of Economics and Statistics*, 75: 410–434.

Van der Klaauw, Wilbert, and Kenneth I Wolpin. 2008. Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics*, 145, no. 1: 21–42.

Kleven, Henrik J, and Mazhar Waseem. 2013. Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *The Quarterly Journal of Economics*, 128, no. 2: 669–723.

Krueger, Alan B., and Jorn-Steffen Pischke. 1992. The Effect of Social Security on Labor Supply: A Cohort Analysis of the Notch Generation. *Journal of Labor Economics*, 10, no. 4: 412–37.

Liebman, Jeffrey B., and Erzo F. P. Luttmer. 2015. Would People Behave Differently If They Better Understood Social Security? Evidence from a Field Experiment. *American Economic Journal: Economic Policy*, 7, no. 1: 275–299.

Liebman, Jeffrey B., Erzo F. P. Luttmer, and David G. Seif. 2009. Labor supply responses to

marginal Social Security benefits: Evidence from discontinuities. *Journal of Public Economics*, 93, no. 11-12: 1208–1223.

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, no. 7: 913–925.

OECD. 2011. *Pensions at a Glance 2011: Retirement-income Systems in OECD and G20 Countries*. OECD Pensions at a Glance. Paris: OECD Publishing.

Reimers, Cordelia, and Marjorie Honig. 1996. Responses to Social Security by Men and Women: Myopic and Far-Sighted Behavior. *Journal of Human Resources*, 31, no. 2: 359–382.

Rust, John, and Christopher Phelan. 1997. How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets. *Econometrica*, 65, no. 4: 781–831.

Saez, Emmanuel. 2010. Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, 2, no. 3: 180–212.

Slemrod, Joel. 2013. Buenas notches: lines and notches in tax system design. *eJournal of Tax Research*, 11, no. 3: 259–283.

Song, Jae G., and Joyce Manchester. 2007. New evidence on earnings and benefit claims following changes in the retirement earnings test in 2000. *Journal of Public Economics*, 91, no. 3-4: 669–700.