Discussion Papers
Statistics Norway
Research department

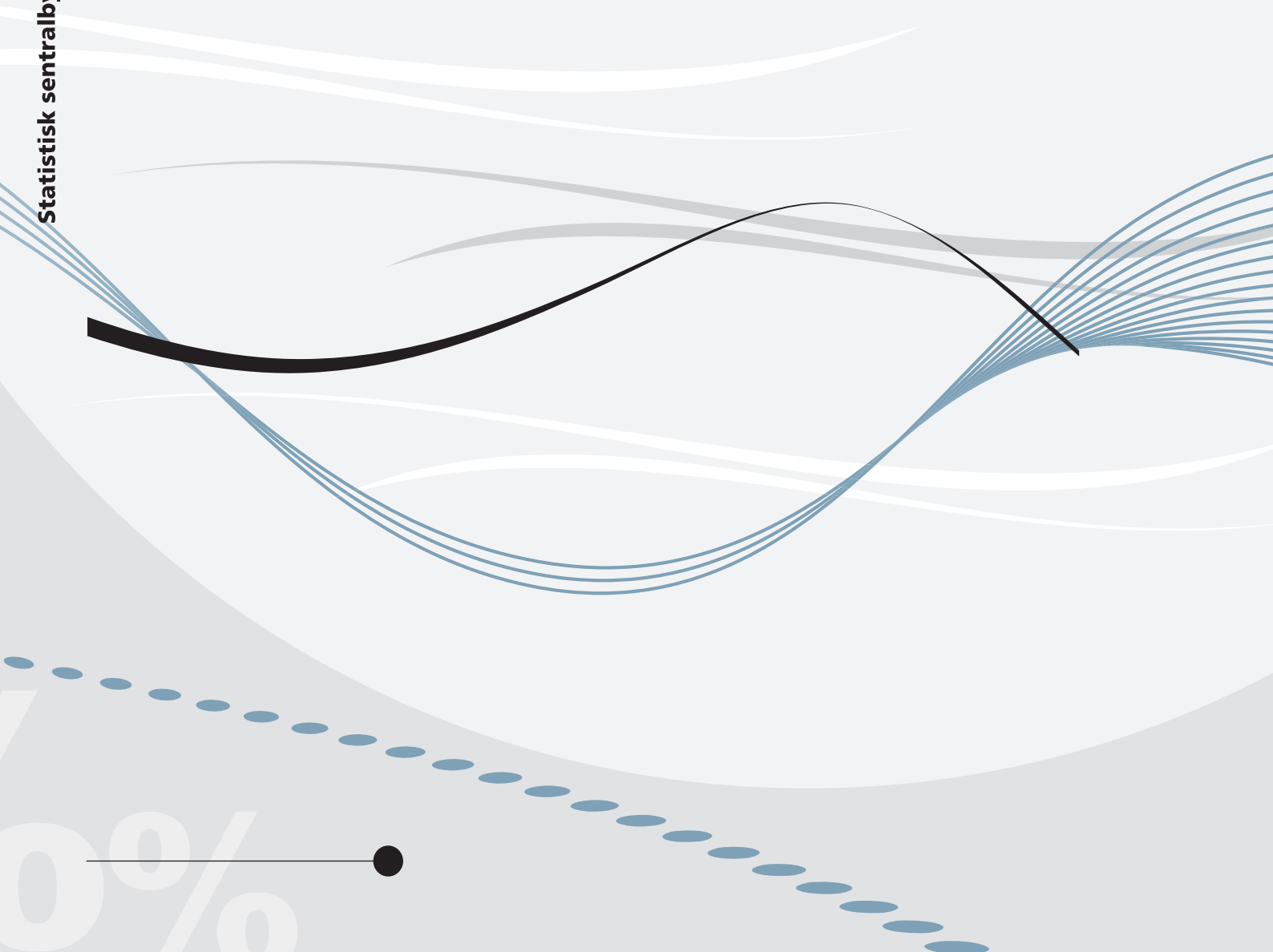
No. 698
July 2012

Statistics Norway
Statistisk sentralbyrå

Christian N. Brinch, Erik Hernæs and Zhiyang Jia

Labor supply on the eve of retirement

Disparate effects of immediate and postponed rewards to working



Discussion Papers No. 698, July 2012
Statistics Norway, Research Department

Christian N. Brinch, Erik Hernæs and Zhiyang Jia

Labor supply on the eve of retirement

Disparate effects of immediate and postponed rewards to working

Abstract:
We study two recent changes in incentives to work facing 67-69 year old workers in Norway: an earnings test reform which increases current earnings from work, and a pension system maturation which removes pension accrual from work. Within a difference-in-differences framework, we exploit these changes to investigate the effects of economic incentives. We find the earnings test reform has large effects, while the pension system maturation has no significant effects. The findings confirm that 67-69 year olds can adjust their work efforts to economic incentives, but do so only to thoses related to current income and not to future pensions.

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

JEL classification: J14, H55.

Acknowledgements: We acknowledge financial support from the Ministry of Labour (Norway). We thank John Dagsvik, Asbjørn Rødseth and Knut Røed for valuable comments to an earlier version. The usual disclaimer applies.

Address: Christian N. Brinch, Statistics Norway, Research Department. E-mail: cnb@ssb.no

Erik Hernæs, The Ragnar Frisch Centre for Economic Research. Email erik.hernas@frisch.uio.no.

Zhiyang Jia, Statistics Norway, Research Department. E-mail: jia@ssb.no

Discussion Papers	comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:
Statistics Norway
Telephone: +47 62 88 55 00
E-mail: Salg-abonnement@ssb.no

ISSN 0809-733X
Print: Statistics Norway

Sammendrag

To begivenheter i Folketrygden gjorde det mulig det mulig å studere hvordan eldre arbeidstakere tilpasser arbeidsinntekten etter økonomiske insentiver. Den første var at avkortingen av folketrygd mot arbeidsinntekt for personer mellom 67 og 70 år gradvis ble opphevet fra 2008 til 2010. Den andre begivenheten var at det i 2007 for første gang siden innføringen av folketrygden i 1967, ble mulig å ha full opptjening (40 år) uten å ha opptjenning (inntekt over 1 G) helt fram til fylte 70 år. Begge deler endret nettoinntekten av arbeid og den samlete virkningen var av samme størrelsesorden. Den fjernete avkortingen påvirket imidlertid nettoinntekten samme år, mens bortfallet av mulig videre opptjening av pensjon påvirket nivået på hele strømmen av framtidige pensjonsutbetalinger. Vi finner klar tilpasning til avkorting, men ingen tilpasning til opptjening av framtidig pensjon. Dette kan bety at framtidige ytelser er mindre verdsatt enn årets inntekt, eller det kan bety at avkortingen var bedre forstått enn opptjening av framtidig pensjon. I informasjonen som ble sendt til framtidige pensjonister ble alle reglene beskrevet i en vedlagt brosjyre, mens avkortingen i tillegg ble klart beskrevet i oversendelsesbrevet

1 Introduction

This paper addresses a key issue in the literature on retirement and pension schemes: the absolute and relative importance of current income and future pension entitlements for labor supply behavior in the crucial years around common retirement ages. We separate and compare these two effects by taking advantage of two changes in the incentive structure in the Norwegian pension system. The first is an earnings test reform, which implies a pure change in effective tax rate on labor income after retirement. The other change is caused by the maturation of the Norwegian National Insurance Scheme (NIS) old age pension rules, which involves a pure change in the incentives to work in terms of social security wealth. The two incentive changes are large and of comparable magnitude.

Knowledges on the absolute and relative importance of current income and future entitlements are essential for understanding and predicting behavioral responses when pension schemes are structured so that postponing retirement is compensated by an actuarially fair adjustment of future pension rights. If future pension rights, properly discounted, are considered as inferior to income today, an earnings test with actuarially fair adjustment will in fact behave like a tax, reducing labor supply in the crucial years when most people retire from the labor market. When pensions are actuarially adjusted according to timing of take-up, people will then tend to take up pension as soon as possible. Although there is no obvious labor supply distortion in this case, the actualral adjustment becomes superflous since few will postpone their pension take-up. Studies of behavioral responses to future entitlements are typically based on the assumption that individuals are forward looking with a single positive discount parameter. This discount factor is often used to construct the social security wealth, which is considered as one of the most important incentive factors of the retirement and is often used to form expected future utilities in a classical dynamic analysis. However, empirical evidences on time preference have documented the inadequacy of such modeling approach, see for example Frederick et al. (2002). In addition, it is not clear how long planning horizon people have when they plan for retirement, since risk and market imperfections can lead to retirement behavior that appears myopic. A recent strand of the labor supply literature also focuses on the importance of context when studying the effects of incentives, see e.g. Saez (2010) or Chetty et al. (2011). In the two cases of incentive change that we study, the labor supply responses may differ not only because of different timing of the rewards to continued working, but also because the NIS maturation may be less transparent than the earnings test.

In the literature, studies of the effect of incentives on labor supply at retirement age often focus on earnings tests of pensions against labor earnings. If deferred pensions are compensated in the form of social security wealth, the labor supply effects will be crucially dependent on how potential retirees evaluate current and future income. If future entitlements are totally disregarded, such tests contribute to very high implicit tax rates on labor income and even moderate labor supply elasticities may give rise to large welfare costs. If evaluation implies discounting with the same factor as in the actuarial adjustment, this is no source of distortion. To date, most empirical studies of the impact of earnings test have focused on the pension system in the US, where the reduction in pension income due to the earnings test is compensated by increased future pensions. Some studies conclude that there is little effect on the labor supply, see for example, Gruber and Orszag (2003), but more recent analyses suggest that earnings test has significant effects on labor supply among the elderly. Both structural and reduced form approaches have found such results (Friedberg, 2000; Friedberg and Webb, 2006; Song and Manchester, 2007; Engelhardt and Kumar, 2009). Similar evidence is found also in other countries, for example, in UK (Disney and Smith, 2002), Canada (Baker and Benjamin, 1999) and Norway (Hernæs and Jia, 2012). That even earnings tests with actuarially neutral deferral mechanisms seem to matter, indicates that social security wealth is valued less than what is implied by the discount factor in use.

In contrast to many earnings test systems in other OECD countries, the Norwegian earnings test has no actuarial adjustment and can be seen as an implicit tax on earnings.[1] Thus, the Norwegian earnings test reform provides a unique opportunity to study labor supply effects of changes in current net earnings. From January 1st, 2008, the earnings test was abolished for 67-year olds while initially retained for 68-year olds, providing clean difference-in-differences identification of a change in effective marginal tax rates from about 70 to about 45 percent (only taking into account income taxes and earnings testing of pensions, not employer paid payroll taxes and value added taxes). We are not aware of studies in the literature of correspondingly large magnitude changes in the tax-benefit system for persons born in one year while the rest of the system is kept in place for persons only one year older. The current paper thus adds to the literature on labor supply and taxation in general through a well-identified quasi-experimental setup. We find large changes, with an elasticity in the order of 0.6, implying that the reform was most likely self-financing when all taxes are taken into account.

[1] As far as we know, all earnings test reforms studied in the literature, except for the systems in Canada and Norway, involved some deferral mechanism.

There is a large literature on how future social security and pension entitlements will affect retirement decisions, with a focus on the direct effect of the level of benefits on retirement behavior, rather than the effect of the rewards of postponing retirement through the benefits which is what we study. The consensus is that the level of future benefits has only modest effects on retirement behavior. See for example, Rust and Phelan (1997), Stock and Wise (1990) and Samwick (1998). In a recent analysis, Coile and Gruber (2007) used data from the US Health and Retirement Study to construct both accrual and forward looking incentive variables. They found that the level of pension wealth has a significant positive effect on retirement. Unfortunately, these results can not be directly used to form comparisons between incentives with immediate payoffs and payoffs in terms of future income streams.

Reimers and Honig (1996) is probably the first and only study on the relative importance of the current and the future benefits on the elderly labor supply responses. Using data from the US Longitudinal Retirement History Study in 1970s, they estimated a hazard model of labor market re-entry for both elderly men and women. They found strong gender differences. Women take account of their social security wealth, but not their current benefits, while men respond only to current benefits not to the change to future benefits. Their work, however, had two important limitations. First, they only consider 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 (2012), change in incentives may have a positive effect on the intensive margin, without any effect on the extensive margin. Secondly, both current pensions and social security wealth enter the analysis as (assumed) exogenous regressors in Reimers and Honig (1996). However, these variables may plausibly be correlated with unobserved variables affecting outcome variables, which may lead to possible bias.

Our work remedies these two limitations. First, as an outcome variable we study the full labor earnings distribution for those eligible for pensions, thus taking into account both adjustments at the intensive and the extensive margin. Secondly, we address the issue of the potential endogeneity of individual earnings and social security wealth by using a quasi-experimental design. In this design, we do not exploit the individual specific differences in pensions or social security wealth, but only differences that are generated through changes in the incentives for large groups and the interaction of these changes with birth cohorts. In addition, while the data used by Reimers and Honig (1996) date back to 1970s, we use recent Norwegian data based on administrative records. Such data are not plagued by self reporting errors on earnings, which

are of particular importance in this type of study. We linked a number of administrative data set on an individual level, which cover the whole Norwegian population - although our study focuses on the subgroup who have still not left the labor market by age 66.

Both the events we study have sizable effects on the incentives to work. The earnings test reform affects current income and had a large effect on the labor supply and earnings for men, with a smaller effect, although not statistically significant at the five percent level, for women. The maturation of NIS affected future pension entitlements only and had no significant effects, not even statistically insignificant effects pointing in the right direction. Therefore, people around the retirement age seem to respond to incentives affecting current income, but not to incentives affecting future pension when they decide how much to work. It is clear from the study of the earnings test reform that the elderly are indeed able to adjust their labor supply. Thus, the lack of response to the change in future pension entitlements cannot be due to restrictions in the labor market. Our conjecture is it is due to the corresponding incentives are not sufficiently well understood since the related pension rules are quite complicated.

The paper is organized as follows. We explain briefly the relevant institutional features of the retirement system including the incentive changes exploited in the analysis in Section 2. We then move on to discuss the theoretical aspect of the possible impact of the changes in Section 3. In Section 4, we discuss the data used. Section 5 and 6 present respectively our empirical strategy and empirical results. Section 7 concludes.

2 Institutional framework

2.1 Retirement in Norway

The backbone of Norway's retirement provision system is a mandatory defined benefit plan, the NIS old age pension system. In addition, there are occupational pensions both in the public and in the private sector. The mandatory retirement age in Norway is 70 years. This means that when you reach 70 years, your employer may force you to quit your job without any other reason than old age. However, starting from the month after a persons 67th birthday, one can receive the NIS old age pensions. In reality, the majority of people start collecting pension benefits as soon as they reach the eligibility age.

There is a voluntary early retirement program (AFP), available from age 62, covering the public sector and about half of the private sector. Under this scheme, a large proportion of

workers can decide to retire at the age of 62 instead of the ordinary 67, without (in most cases) receiving lower pensions than they would, had they worked to the age of 67. In addition, almost half of the total population retires through the disability pensions scheme before they reach 67 (see Table 1 and 9).

In the following we explain in detail the Norwegian pension systems, and describe the NIS old age pension system, the public occupational pensions scheme, private occupational pensions schemes, and how the pensions from the different schemes depend on each other and on labor earnings. A major pension reform was implemented in 2011, but the description here covers the previous system which was in operation during our observation period.

2.2 The national insurance scheme old age pensions

The NIS provides basic pension coverage for all individuals with a minimum amount of years of residence in Norway. The system is organized around a unit called the basic amount (G). In 2007, one basic amount (1G) is around 70 000 NOK[2]. We will use G as the currency in the following, noting that it is comparable between years over the observation period as it was adjusted annually by the average nominal wage growth.

For each year from 1967 onwards, individuals are assigned pension points if their pensionable income (basically, labor earnings, temporary benefits or calculated labor earnings for the self-employed) exceeds 1G[3]. Pension points are calculated based on total labour earnings in excess of 1 G. Although one can retire at 67 years, it is possible to earn pension points also from age 67 to age 70. However, these points are included only in a person's pension calculation from age 70.

The pension is the sum of the basic pension (1 G for singles, 0.75 G for married) and a supplementary pension that depends on the pension point history. The supplementary pension depends on the average labor income in the best twenty years and the number of years with positive pension points. To obtain a full pension, 40 years with positive pensions points are required. For individuals with less than 40 years' pension points, the supplementary pension is reduced proportionally. Since the pension point calculation started in 1967, it was not possible to acquire 40 years of points until 2006. Thus, even with a full earnings history from 1967, it is still necessary for individuals born before 1940 to maintain at least 1G pensionable income in

[2]approximately 12 000 USD.

[3]roughly equals to a fifth of annual average wage income.

one or more years during the period when they are between 67 and 69 years old to receive full supplementary pension from age 70.

Individuals can combine pension income with labor earnings. Persons above age 70 receive the NIS pension regardless of whether they receive any labor earnings. For 67-69 year olds, however, a retirement earnings test was applied to the pensions. The system during 2002-2007 was as follows: any labor earnings in excess of 2G leads to a reduction in the NIS pension based on a 40 percent rate. Note that there is no adjustment of future NIS benefits due to the earnings test, so that unlike the case in the US with adjustment via deferral, the Norwegian test can unambiguously be viewed as a tax.

In practice, since the NIS pension can be obtained the month after the 67th birthday, the earnings test of the pension was based on average monthly income after retirement. For example, a person born 1st May 1938, can receive old age pension from June 2005 (the month after his/her 67th birthday) and receive pension for 7 months in 2005. In this case, earnings up to (7/12)·2G during these 7 months are allowed before the earnings test applies. Similarly, when she is 70 years old in 2008, earnings up to (5/12)·2G before June are exempt, as are all earnings from June, 2008.

2.3 Occupational pension schemes

The occupational pension schemes differ greatly between the public and the private sector in Norway.

Full pension coverage in public pension system requires thirty years of employment in the public sector in a full or part time job, above approximately 40% position. The pension from the public schemes is based on the finale wage. The combined pension from NIS and the public scheme is 66 percent of the final salary as working. The public scheme pays out *literally whatever is not covered by the NIS*. Persons fully covered by public pension schemes are thus not affected by changes in the NIS for most cases.

Private occupational pension schemes in Norway are far more heterogeneous than the public schemes. Currently, there exist tax-favored schemes for both defined-benefits and defined-contribution style pension schemes, and a minimum level of occupational pension schemes are mandatory. The dominant type of scheme offering substantial occupational pensions to individuals in our data set is defined benefits schemes similar to the public schemes: Around 30 years of service are required for full pension rights and 60-70 percent of the final wage as final

pension level taking into account NIS pensions.

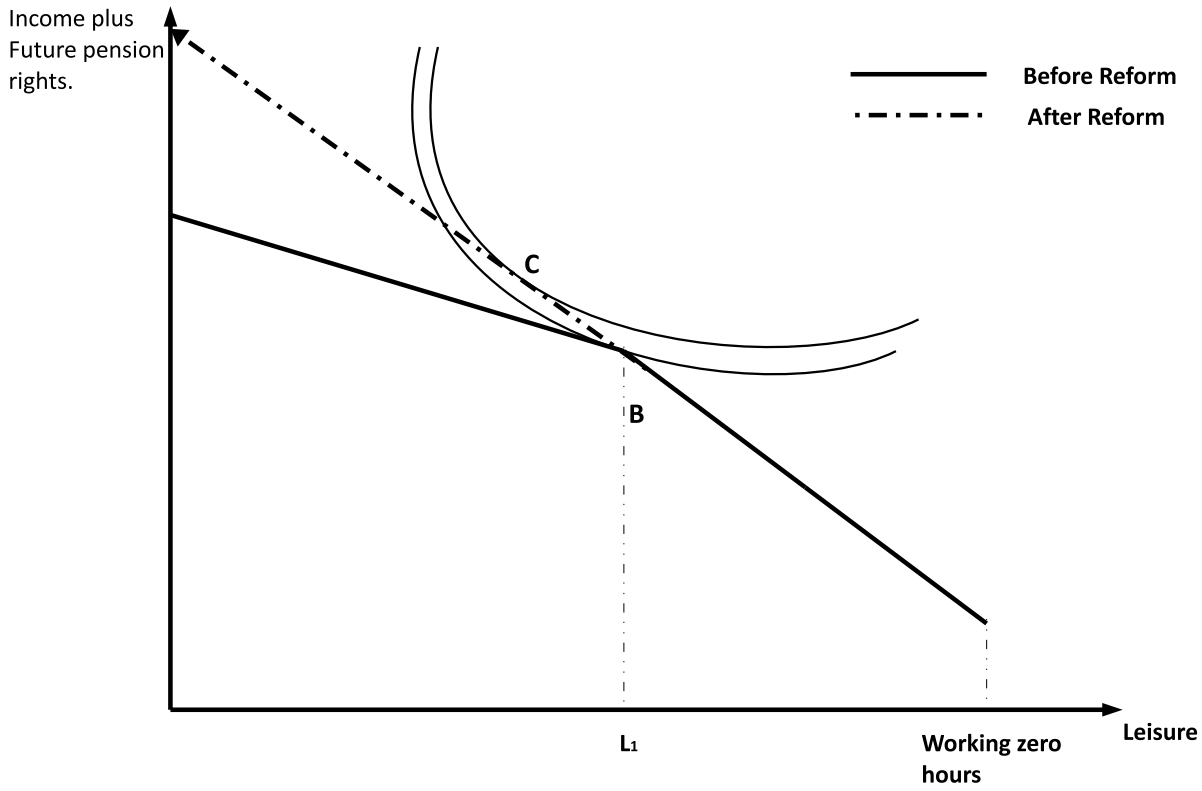
A major difference between public and private defined benefits occupational pension schemes is the way the systems are coordinated with the NIS system. The public scheme interacts fully with the NIS scheme - neutralizing effects of changes and reforms in NIS. The private scheme is only coordinated with an industry standard called "computed NIS", which roughly assumes full coverage from 1967 and upwards, every year with the same number of pension points as the end year. Thus, the actual pension one receives from NIS and the private pension scheme will differ from the targeted replacement rate because actual and computed NIS will differ. In particular, unlike the public schemes, the private schemes will not neutralize the additional NIS pension received from working when 67-70 years old. Similarly, the private schemes will not compensate retirees for the loss of NIS pensions due to incomplete income histories. Since both reforms we are going to consider in this paper change only the NIS pension, persons covered by the public occupational pension schemes are not affected - or at least affected to a much smaller extent. We will focus only on those individuals who are not entitled in public occupational pensions.

2.4 The changes in incentives studied here

The two events studied in this paper are changes in NIS earnings test rules and in the interaction between birth cohort and NIS rules.

The earnings test in NIS was abolished for 67 year olds in 2008, for 68 year olds in 2009 and for 69 year olds in 2010. Our data cover the abolishment for 67 year old and this is the event we consider in this analysis. The test implied that the NIS pension was reduced by 40 per cent of earnings exceeding 2 G and the abolishment in 2008 gives considerably higher pecuniary incentives to continue working at age 67 for persons covered only by NIS or by NIS and a private sector occupational pension. The shift in the labor supply budget line associated with the reform is illustrated in Figure 1. The kink (change in the slope) of the net earnings as function of hours worked was removed as part of the reform. There is also in principle a kink where all the pension have been earning tested away, though that kink is empirically irrelevant and not shown in the figure. The predicted effects of such a change is discussed in Section 3. Workers with occupational pensions from the private sector can in general choose whether to receive their occupational pension unabridged or defer it with an actuarial adjustment. Hence, we assume that the earnings test for private sector workers with occupational pension applies

Figure 1: Illustration of work incentive effects of the Earnings Test reform



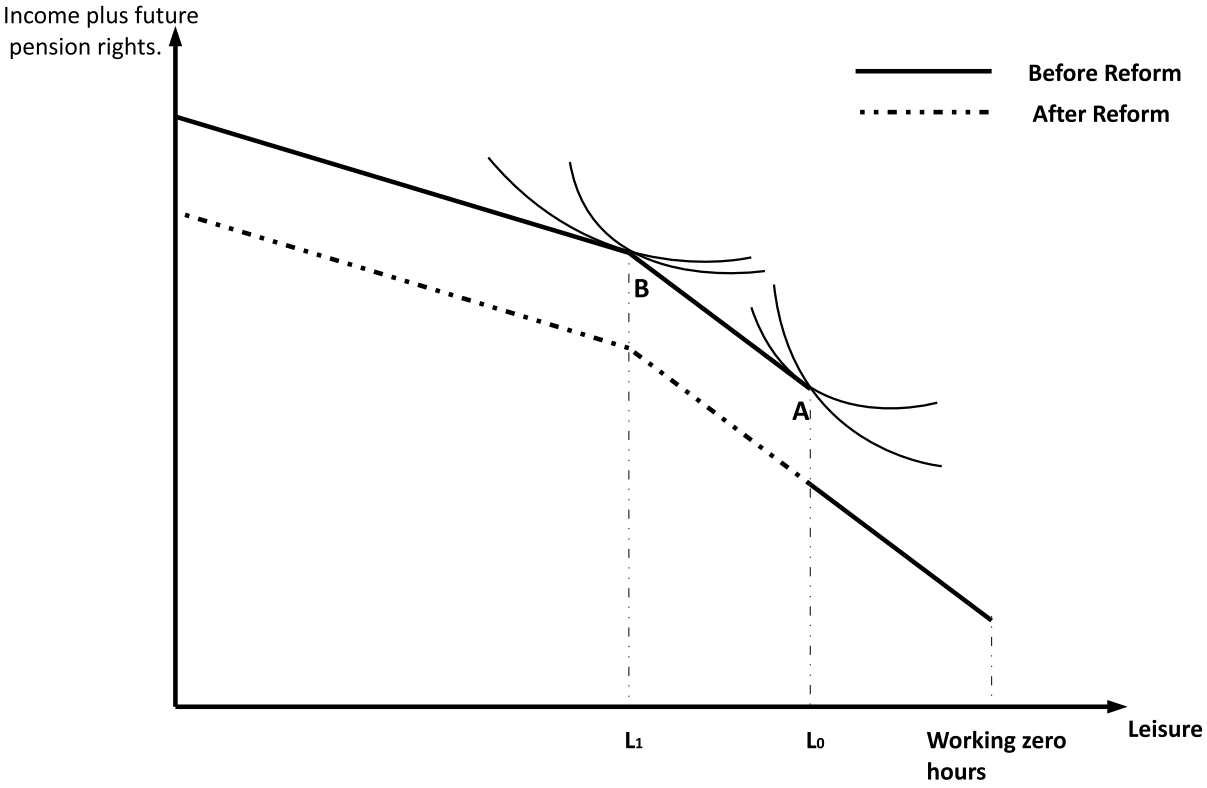
only to the NIS part of their pension.

The second change in incentives is not a proper reform - it is rather an effect of maturing of the old NIS phasing-in rules. Since recording of earnings in the system began in 1967, it was not possible before 2007 to have earned 40 years of pension points. Hence in 2006, all retirees between 67 and 69 (1937-1939 cohorts) had an incentive to continue working (and earn at least 1 G) to increase their social security wealth. The premium for earning at least 1 G gives a notch (discontinuity) in the labor supply budget line, as illustrated by Figure 2, which also includes the earnings test that was still in place. In 2007, for the first time, those who had worked continuously since in 1967, would no longer increase their NIS pensions by continuing working after pension take-up.

The first reform works through the immediate gains and the second reform through expected gains in the future. These two changes in the rules studied are both of large and comparable magnitude. Consider a person who had a complete working history from 1967 to age 66, with 3 G of supplementary pensions, who continues to work and earns 5 G at 67.[4] The abolishment

[4]The income level in the example is chosen to approximately equalize the effect of the incentive changes. Most individuals have earnings lower than 5 G at age 67 and the abolishment of the earnings test will have somewhat

Figure 2: Illustration of work incentive effects of the pension system maturing



of the earnings test will increase current income (before tax, after earnings test) by 1 G. After maturing of NIS, for the same individual a potential 2.5% increase in annual supplementary pension from one extra year of earnings disappeared from 2006 to 2007. Note that the amounts are substantial in present value terms. The annual gain is around 0.075 G. With 14 years of expected pensions after age 70[5], the total is then 1.05 G. This computation requires that a person discounts future nominal income as the same rate as the rate of increase in G - the same as the increase in the nominal wage rate. This is a relatively high discount rate, compared to what one could expect (after tax) in the financial market.

3 Labour supply with kinks and notches in budget lines

In the following empirical analysis, we will measure the effect of changes in the tax-benefit system on labour supply directly on the empirical cumulative income distributions. We will first

lower effect on their current before-tax, after earnings test, income.

[5]In 2007, the remaining life expectancy for a 70 years old Norwegian is 13.7 years and 16.5 years for male and female respectively.

give a very brief introduction to the standard static neoclassical labor supply model, focusing on the prediction the model gives rise to under changes in budget constraints corresponding to those we will measure the effect of. We will then proceed beyond the neoclassical model, deriving predictions from a discrete choice model with extraordinarily weak assumptions - opening up the possibility for labor supply responses on the extensive margin. Both the neoclassical model and the more general discrete choice model give predictions about the effect of the changes in budget constraints we study on the earnings distribution.

The standard model for static labor supply builds on a continuous choice model, where individuals have preferences for leisure and consumption and can freely choose any combination of them within the budget constraints. Without strong assumptions on preferences, the model does not give rise to strong predictions about labor supply curves even in the most basic case, because increasing wage implies increasing purchasing power, giving an income effect (an individual can afford more leisure) that may or may not offset the substitution effect (leisure becomes more expensive as the wage rises).

The first issue to discuss here is the predictions of the model of the effect on introducing a kink in the budget line as shown in Figure 1. In the presence of such a kink, individuals who maximize their utility will allocate themselves to the kink in the budget line for a range of wages. Essentially, this means that the earnings as a function of hourly wages will be constant over an interval of wages. In the same interval, the elasticity of the individual labour supply curve will be -1, as hours of work are adjusted in response to hourly wages such that their product is constant. The key prediction, when aggregating over individuals with suitably defined continuous distributions over wage rates and preferences, is that there will be a spike in the earnings distribution at the kink, because a range of combinations of wage rates and preferences will lead to just that allocation. Allowing for imperfect optimization, there will be bunching around the kink points. See for example, Saez (2010) for a detailed discussion on bunching in income distribution and kinks in the budget sets.

The incentive change involving social security wealth is less elegant to study with the neoclassical model, involving a "notch" in the "lifetime" budget line. Essentially, it is necessary to optimize for each individual over each convex subset of the budget set and then compare the local optima to find the global optimum. A premium for earning at least 1 G, which we observe, will lead some people who would otherwise earn less than 1 G to earn just above 1 G. The prediction is then that they would earn just above 1 G - unless their indifference curves are

non-convex. So we expect to see some bunching just above the notch.

The earnings test reform in 2008 effectively removes the kink in the budget line, while the maturation of the NIS pension rules removes the "notch". The responses to such reforms predicted by standard economic theory are complicated and vary across individuals . See Friedberg (2000), Hernæs and Jia (2012) for a discussion for the case of earnings test reform. It is also far from obvious that the most important labor supply responses are on the intensive margin. This holds in particular for the labor supply responses that we study here, which we expect to be closely related to the retirement decision, which we usually (maybe falsely) think about as a binary choice.

We therefore also derive predictions from a generic discrete choice model, where an individual chooses among a finite or countable number of alternatives. In such a model we can study the effect of changes in the tax-benefit system on the distribution of earnings - as long as we assume that unobserved factors that help to explain which alternative is the best are independent of the tax-benefit system.

Formally, an individual may choose among $n \in \mathbb{N}$ job alternatives, potentially including full and partial withdrawals from the labor market. Each job alternative is associated with a gross earnings Y_i, working hours h_i and other, possibly unobserved, job characteristics, ξ_i. The utility associated with job alternative $i \in (1, \dots n)$ is $U(C_i, h_i, \xi_i)$, where C_i denote the disposable income from job alternative i, which is a function of gross earnings $C_i = f(Y_i)$, The individual chooses the job that gives highest utility, or the alternative with lowest index in case of a tie. Assume that utility is strictly increasing in disposable income. Let $Y^*(f)$ denote the gross earnings of the preferred alternative under tax-benefit system f.

Note that the model discussed here is more general than the continuous choice model, in the sense that it is possible to approximate the continuous choice model arbitrarily closely with a special case of the discrete choice model: For any individual, define $U(C_i, h_i, \xi_i) = v(C_i, h_i)$, where v is a standard utility function over consumption and leisure, and assume that the individual can choose among a countable number of different alternatives such that the set of alternatives is dense in the admissible choice set given by the continuous choice model. Hence any result that holds for the generic discrete choice model will also hold for the standard neoclassical labour supply model.

The following proposition considers the labor supply behavior of one agent in terms of earnings, as a response to changes in the tax-benefit system. Let the indicator function $1()$ take

the value 1 if the argument is true, 0 if the argument is false.

Proposition 1 *Consider two tax-benefit systems f_1 and f_2 such that $f_2(Y) > f_1(Y)$ for $Y \geq \tilde{Y}$ and $f_2(Y) \leq f_1(Y)$ for $Y < \tilde{Y}$.*
(i) $1(Y^*(f_2) \geq \tilde{Y}) \geq 1(Y^*(f_1) \geq \tilde{Y})$.
(ii) If $f_2(Y) = f_1(Y)$ for $Y < \tilde{Y}$, then $1(x_0 \leq Y^*(f_2) \leq x_1) \leq 1(x_0 \leq Y^*(f_1) \leq x_1)$ for any $x_0, x_1 \in (0, \tilde{Y})$ with $x_0 \leq x_1$.

Proof.

Point (i) holds trivially if $1(Y^*(f_1) \geq \tilde{Y}) = 0$. If $1(Y^*(f_1) \geq \tilde{Y}) = 1$, then $1(Y^*(f_2) \geq \tilde{Y}) = 1$, because if $U(f_1(Y_j), h_j, \xi_j) = \max(U(f_1(Y_1), h_1, \xi_1), \ldots, U(f_1(Y_n), h_n, \xi_n))$ and $Y_j \geq \tilde{Y}$, then $U(f_2(Y_j), h_j, \xi_j) > U(f_2(Y_k), h_k, \xi_k)$ for any k such that $Y_k < \tilde{Y}$.

Point (ii) follows because if $U(f_1(Y_j), h_j, \xi_j) \leq \max(U(f_1(Y_1), h_1, \xi_1), \ldots, U(f_1(Y_n), h_n, \xi_n))$ with $Y_j < \tilde{Y}$, then $U(f_2(Y_j), h_j, \xi_j) \leq \max(U(f_2(Y_1), h_1, \xi_1), \ldots, U(f_2(Y_n), h_n, \xi_n))$. As a consequence, either $Y^*(f_2) > \tilde{Y}$ or $Y^*(f_2) = Y^*(f_1)$.

■

As a consequence of the change in the tax-benefit system, the utilities of alternatives will change as well. Informally, point (i) holds because no-one switches from an alternative where utility increases to an alternative where utility stays unchanged or even decreases. Point (ii) holds because no-one switches to an alternative with unchanged utility in a situation where utility is not decreased for any alternative. The reason for writing up the result in the somewhat awkward fashion with indicator functions is that these aggregate easily to distribution functions.

Consider a situation where individuals are heterogeneous - in any respect. Typically not all heterogeneities can be accounted by observable characteristics. Even among observable identical individuals, their observed income, namely the income in their preferred alternatives, can differ and should be treated as random. In turn, we need to study the effect of tax benefit system changes on the distribution of observed income. This leads us to the following:

Proposition 2 *Consider two tax-benefit systems f_1 and f_2 such that $f_2(Y) > f_1(Y)$ for $Y \geq \tilde{Y}$ and $f_2(Y) \leq f_1(Y)$ for $Y < \tilde{Y}$.*
(i) $Pr(Y^*(f_2) \geq \tilde{Y}) \geq Pr(Y^*(f_1) \geq \tilde{Y})$.
(ii) If $f_2(Y) = f_1(Y)$ for $Y < \tilde{Y}$, then $Pr(x_0 \leq Y^*(f_2) \leq x_1) \leq Pr(x_0 \leq Y^*(f_1) \leq x_1)$ for any $x_0, x_1 \in (0, \tilde{Y})$ with $x_0 \leq x_1$.

Proof.

Based on a population of N heterogeneous individuals, each of them following the model leading to Proposition 1 with separate response functions to the tax benefit system $Y_1^*(f), \ldots Y_n^*(f)$. Define $Y^*(f)$ as a random draw from this population. Now $Pr(Y^*(f) \leq a) = \frac{1}{N}\sum_{i=1}^{N} 1(Y_i^*(f) \leq a)$ - so Proposition 1 aggregates directly to Proposition 2.

■

Similar to Proposition 2, it is straightforward to aggregate Proposition 1 based on models where individual choice is considered random (e.g random utility models), because Proposition 1 holds conditional on the realization of random variables. Since the results hold for any realization of the random variables, they aggregate up and hold for individual distribution functions as well as population distribution functions.

Both the earnings test reform and the NIS maturation are changes in tax-benefits systems that satisfy points (i) and (ii) in Proposition 1 and 2. The following is a summary of the predictions of the discrete choice model and the additional predictions generated by the neoclassical continuous choice labour supply model (with optimization errors):

1. The discrete choice model predicts that abolishing the earnings test of incomes in excess of 2G will lead more individuals to earn more than 2G. The density in the empirical income distribution below 2G will not be higher after the abolition. The continuous choice model in addition predict that the density in the income distribution below 2G will be unchanged.

2. The discrete choice model predicts that taking away extra social security wealth for having an annual labor income in excess of 1G will lead fewer agents to earn more than 1G. The continuous choice model in addition predicts that these agents that change behavior would otherwise earn just above 1G.

3. The continuous choice models imply that there will be "bunchings" in the income distribution at 1 G and 2 G before the changes in incentives and that these bunchings will disappear as a consequence of the incentive changes. Importantly, the discrete choice model also opens up a possible interpretation of small responses to incentives that is implicitly assumed away in the continuous choice model. It is possible that the response to incentives is weak because those affected have few alternatives to choose from in the labor market. In other words: a change in incentives enables us identify the absolute effect

of incentives. If this effect is small, this may be (i) because agents do not care very much about pecuniary incentives or (ii) because few agents actually have the opportunity to adjust: in terms of our simple model, their choice sets are limited. One of the attractions of our strategy in this paper is that we study two different changes in incentives. With a large effect from one of the changes, we can informally rule out the interpretation that agents do not respond because of constraints in their choice set.

4 Data

We base our analysis on administrative data, which are merged administrative registers from Statistics Norway up to 2008. We use demographic data files, old age pension registry and tax return records. Our data cover the full Norwegian population in the birth cohorts 1937 to 1941. We have information on date of birth, gender, education, the date of the old age pension benefit taking up, accumulated pension rights as well as detailed income information. 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 Hernæs and Jia (2012) show there is quite strong heterogeneity in the responses to a previous earnings test reform in 2002, driven by labor market status just before the eligibility of the NIS pension (at age 66), which is a strong "predictor" of the response to the reform. We observe very small responses from those who are already out (most often retired with AFP or on disability pension). In the present study we characterize individuals based on their welfare program and labor market participation, pension incomes and labor earnings in the year they become 66 years old. Table 1 shows that most individuals are already pensioners at this stage, primarily through the disability pension scheme and the AFP early retirement scheme.

Since there is very few return to work, we study only the minority of the population that is still working prior to receiving old age pension rights, 20 per cent of the male and 12 per cent of the female cohort. For this group, we are able to characterize their incentives precisely.

We analyze the labor supply decision separately for men and women to facilitate the interpretation of results. In the main text, we focus on the results for men, mainly for reasons for precision. The corresponding analysis for women is given in appendix A. The results for women are not qualitatively different from the effects for men, though point estimates of the effects are smaller and standard errors are larger, due to a smaller sample size.

Table 1: Average annual income at age 68 by labor market status at age 66. Classification of 66-year olds men by January 1st 2003-2006 .

	Number	Fraction of full population	Mean earnings at age 68	Fraction of total earnings
Men				
Not active in labor market				
AFP early retirement	18435	0.26	0.32	0.09
Disability pension	27425	0.39	0.23	0.10
Occupational pension	519	0.01	0.65	0.01
NIS pension	259	0.00	1.60	0.01
Idle	2316	0.03	0.25	0.01
Some work	8956	0.13	1.00	0.14
Active in labor market				
Public sector employee	2531	0.04	4.60	0.18
Analysis population	10644	0.15	2.71	0.46
Full population	71085	1.00	0.89	1.00

Notes: AFP early pension scheme does not include those who are only 20 or 40 percent pensioners. Occupational pensions and NIS pensions are defined based on receiving at least 1 G in pensions. "Idle" defined based on no income while "some work" means less than 1 G in income.

We focus on the behavior of the group of persons who are still active in labor market at age 66. As we have discussed earlier, the reform will have different impact on persons under different occupational pension schemes. Those who are covered by the public occupational pension system will not be affected by either change, so our analysis population consists of individuals who work in private sector or are self-employed.

5 Empirical approach

The backbone of our empirical approach is a difference-in-differences identification strategy, see e.g Angrist and Pischke (2009); Imbens and Wooldridge (2009). With both of the changes in incentives we study, we separate data into treatment and control groups based on whether individuals are affected by the changes or not. We then study outcomes for the treatment and control groups before and after treatment. The difference-in-differences estimator compares 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. Proper definition of the treatment and control groups is essential to obtain sensible results. Since the two changes

in incentives we study affect different individuals, we need to define the treatment and control groups separately for these two reforms, which will be described in detail in the respective sections presenting the analysis of each change.

Since nonlinearlity in budget constraints induces bunchings in the income distributions and the reforms we study shift such budget constraints, we expect an uneven effect over the earnings distribution, as suggested by Proposition 2. To take account of this, we analyze the reform effects by constructing difference-in-differences estimators for the full cumulative distribution function (cdf) of the earnings distribution. A popular approach to studying distributional treatment effects is to use the "quantile treatment effects" where the objective is to identify the shift of quantiles of the conditional earnings distribution before and after treatment, see e.g. Imbens (2004), Chernozhukov and Hansen (2005) or Athey and Imbens (2006). While we do supply some quantile treatment effects below, our main focus is on distributional treatment effects, where we study the reform effects on the cumulative distribution function rather than on its inverse (which gives the quantile treatment effects). There are two reasons for working directly with the cdf rather than its inverse, the quantile functions. First, results are easier to interpret. Secondly, our theoretical predictions from Section 3 are in terms of the cdf of the earnings distribution, not the quantile function.

In practice, distributional treatment effects are completely straightforward to calculate. If the main outcome variable is Y_{it}, we define a family of derived outcome variables based on the indicator function $Y_{ita} = 1(Y_{it} > a)$. Now, instead of doing the difference-in-differences analysis based on Y_{it} as the outcome variable, we do the corresponding analysis for the derived outcome variables Y_{ita}. For each choice of a, the analysis gives the 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 cdf, provided that the identifying assumptions underlying the difference-in-differences approach (essentially, parallel trends for treatment and control groups in the absence of treatment) are satisfied for all a. See Havnes and Mogstad (2010) or Hernæs and Jia (2012) for a detailed discussion of this approach and its relationship with standard conditional quantile treatment models. This strategy is closely related to the unconditional quantile treatment effects, see Firpo et al. (2009).

Within this approach, it is straightforward to control for covariates using the linear probability model. In addition to the standard difference-in-differences identification strategy, we use a supplementary identification strategy as a robustness check. This identification strategy

is based on controlling for separate linear time trends for treatment and control groups. These time trends are identified by using more than one pre-treatment period. Both approaches are straightforwardly admitted into ordinary least squares frameworks.

We also provide some quantile treatment effects - as these are useful as a supplement for interpreting the main results. Without control variables, the quantile treatment effect are completely straightforward (e.g. just a difference-in-differences based on medians rather than means). With controls, quantile treatment effects are more complicated. The quantile treatment effects we report for models with controls are based on the same linear probability regressions as our main results. Based on these regressions, we calculate post-treatment distributions for the control and treatment groups based on standard prediction from the linear regression models with the pre-treatment covariate compositions. The quantile treatment effects are then found by comparing the relevant quantiles in the actual and predicted outcome distributions for the treatment and control groups. The method collapses to standard quantile treatment effects if there are no covariates. Since in our applications in the following, the outcome distributions for the control groups are almost identical in the pre- and post-reform periods, the quantile treatment effects we report are also almost identical to effects that would be found based on the changes-in-changes model (Athey and Imbens, 2006).

The general approach of studying labour supply through the earnings distribution also circumvents the need for data on working hours, which are not of good quality in the data sets available. We use the distributions of annual earnings during a whole calendar year for individuals reaching ages 68 and 69 years in the relevant year. We do not use the annual incomes for those reaching age 67, since it is difficult to identify earnings before and after retirement eligibility for this cohort.

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 1st 2008. In principle, only those with 67th birthdays in December 2007 would be exempt from the earnings test during the whole of 2008, as the earnings test initially was lifted only for 67 year olds and were set to become effective again from the month after their 68th birthday. However, it was decided in

October 2008 that the earnings test also for 68 year olds would be abolished from 2009. It then turned out, that in practice, the whole annual earnings in 2008 was exempt from the earnings test for those who became 68 during October to December 2008. Therefore, the treatment group was defined as those becoming 68 in October-December 2004-2008, with treatment taking place in 2008.

As control group we chose in a similar way, persons becoming 69 years in 2004-2008, who were subject to the earnings test throughout the period. Thus, before treatment (in 2007 and earlier), both the treatment and control group are subjected to the earnings test, while after treatment (in 2008), only those in the control group are subjected to the the earnings test. Limiting also the control group to persons born in October-December leads to generally larger point estimates but much less precise results (not reported). Due to small population sizes, we use both a simple pre-treatment period (2007) and a longer pre-treatment period covering 4 years (2004-2007).

Table 2 gives descriptive statistics for the treatment and control group among men in the pre- and post-reform periods for the outcome variables and control variables measured at age 66. For the control variables, the control and treatment populations look reasonably balanced. The income at age 66 increases somewhat from the pre- to the post-reform period, but at least as much for the control group as for the treatment group. The share of self-employed and the share of employees in firms with private occupational pension schemes are reasonably constant within both groups. The length of highest completed education also increases from the pre- to the post-reform period, but again to at least the same extent for the control group as for the treatment group. There is, however, quite a difference in the amount of income earned. This primarily reflects that more of those in the control group have withdrawn from labor marketwith zero earnings, since they are one year older than those in the treatment group.

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 - they may stay on working to keep their job to take advantage of the lack of an earnings test the next year (and potentially the year after that, if they guessed right about the policy for 2010) as well.

Clearly, mean incomes in the years of analysis are increasing from the pre-reform to the post-reform population for both treatment and control populations, though the increase is much stronger for the treatment population. This particularly reflects a decrease in the proportion of

Table 2: Descriptive statistics, population for analysis of means testing reform, men. Summary statistics for outcomes at age 68 (treatment group) and age 69 (control group) and control variables (measured at age 66), 2004-2007, 2007 and 2008. Means, standard deviations for non-binary variables in parentheses, quantiles when indicated.

	Treatment group			Control group		
	Before reform		After reform	Before reform		After reform
	2004-2007	2007	2008	2004-2007	2007	2008
Dependent variables						
Income (trimmed)	2.512	2.571	3.375	1.788	1.931	2.022
	(3.212)	(3.130)	(3.672)	(2.973)	(3.051)	(3.019)
Income > 0	0.866	0.868	0.909	0.696	0.720	0.737
Income > 1G	0.607	0.640	0.687	0.441	0.480	0.500
Income > 2G	0.379	0.398	0.509	0.244	0.263	0.286
Income > 4G	0.198	0.207	0.321	0.134	0.144	0.161
Income > 6G	0.113	0.112	0.190	0.078	0.086	0.097
1st Quartile	0.432	0.505	0.704	0.000	0.000	0.000
Median	1.570	1.678	2.067	0.663	0.905	1.003
3rd Quartile	3.052	3.221	5.009	1.987	2.036	2.140
Control variables						
Income, age 66	6.170	6.264	6.351	6.108	6.122	6.365
	(3.340)	(3.315)	(3.328)	(3.401)	(3.361)	(3.557)
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: Treatment group defined as being born in October-December. For the treatment group the outcome is measured as annual income in the year they become 68 years. The control group is born throughout the year. The outcome for the control group is measured as annual income in the year they become 69 years. Some individuals are in the treatment group for one year and the control group another year. This is taken into account when computing standard error.

Table 3: Difference-in-differences estimates of the effects of abolishing mean-testing

	Raw estimates	Estimates with controls	Estimates with controls and ext. pre-reform period	Estimates with controls and separate linear trends
trimmed mean	0.714	0.776	0.676	0.774
	(0.196)	(0.167)	(0.139)	(0.197)
Prob(Y>0)	0.024	0.025	0.003	0.020
	(0.021)	(0.021)	(0.017)	(0.025)
Prob(Y>1G)	0.027	0.030	0.024	0.006
	(0.029)	(0.029)	(0.022)	(0.033)
Prob(Y>2G)	0.087	0.092	0.091	0.095
	(0.030)	(0.028)	(0.022)	(0.033)
Prob(Y>4G)	0.097	0.103	0.099	0.102
	(0.025)	(0.023)	(0.019)	(0.027)
1st Quartile	0.208	0.173	0.213	0.028
	(0.105)	(0.102)	(0.084)	(0.080)
Median	0.295	0.318	0.180	0.215
	(0.174)	(0.165)	(0.147)	(0.218)
3rd Quartile	1.621	1.585	1.710	1.698
	(0.348)	(0.316)	(0.256)	(0.361)
Controls		x	x	x
Data from 2004-2006			x	x
Group-specific trends				x

Notes: Controls are listed in Table 2 as they appear in the model - except Earnings at age 66 which enters as a third order polynomial.

earning between 1G and 2G and an increase in the proportion of earnings above 2G, above 4G etc. Figure 3a gives kernel density estimates for the immediate pre- and post-reform years for the treatment and control groups. For the control group in the right panel, clearly, there is almost no change in the earnings distribution, with a pronounced bunching of earnings at 2G, where the kink is in the budget set due to the means-testing. For the treatment group, the pronounced bunching around 2G in the pre-reform period almost entirely disappears as the reform of the earnings test is introduced and the kink in the budget set disappears.

Difference-in-differences estimates[6] on the earnings distribution are reported in Table 3. The first column gives the raw difference-in-differences estimates based on the short pre-reform

[6]Note that some persons who are in the control group in the post-treatment period are also in the treatment group the year earlier. Since their incomes as 68 year olds and 69 year olds are most likely correlated, this invalidates the standard formulas for the standard errors of difference-in-difference estimates. Therefore, all standard errors in our analysis are based on bootstrap simulations, designed to capture the correlation between the within-person-correlations. (Essentially, the persons who figure in several groups are bootstrapped separately and included in both groups.)

time series (2007 only). The mean of income increases by about 0.7G. This confirms a positive effect of the earnings test reform. The effects on the full income distribution are given in Figure 3b with both distributional and quantile treatment effects reported. As we expected, the increase is uneven over the earnings distribution. Almost none of the increase in earnings is explained by changes on labor market participation, i.e. there is no significant change on the share of individuals who works (working is defined as having positive earnings). This is consistent with findings by Song and Manchester (2007) and Hernæs and Jia (2012) where they also find that the earning test reforms in US and Norway have very small or almost no effect on the extensive margin. The main effects on the income distribution are concentrated around the earnings test threshold 2G. After the reform, the percentage of individuals with earnings above 2G increases by 8.7 percentage points, which is about a quarter of the pre-reform level. A similar change carries up quite far into the income distribution.

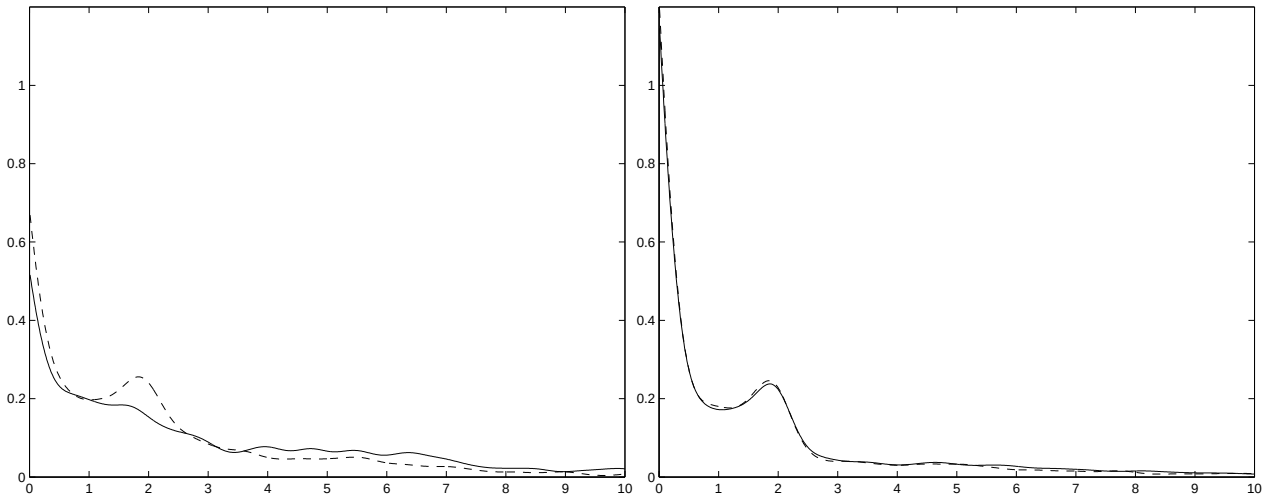
In addition to the distributional treatment effect which measures the changes in cdf directly, we also present the corresponding unconditional quantile treatment effects. The unconditional quantile effects follow an inverse U pattern with a long left tail and peak around 70th percentile. The reform has significant effects only in the upper half of the income distribution. In Table 3 this translates into statistically insignificant changes at the first quartile and the median, with large changes at the upper quartile, which is around 1.6 G and amounts to an increase of about 50 percent. In the more extreme upper tail of the income distribution we find smaller and insignificant results, which may be due to the fact that the reform gives a strong income effect for individuals with high earnings before the reform.

Column 2 in Table 3 reports the same difference-in-differences effects in a linear regression / linear probability 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. As is clear, the controls hardly change the raw difference-in-differences numbers. Column 3 reports results with the longer pre-reform time series (2004-2007). Again, there are no substantial changes to the results as a consequence of this change, although we gain quite a bit in precision due to increased sample size.

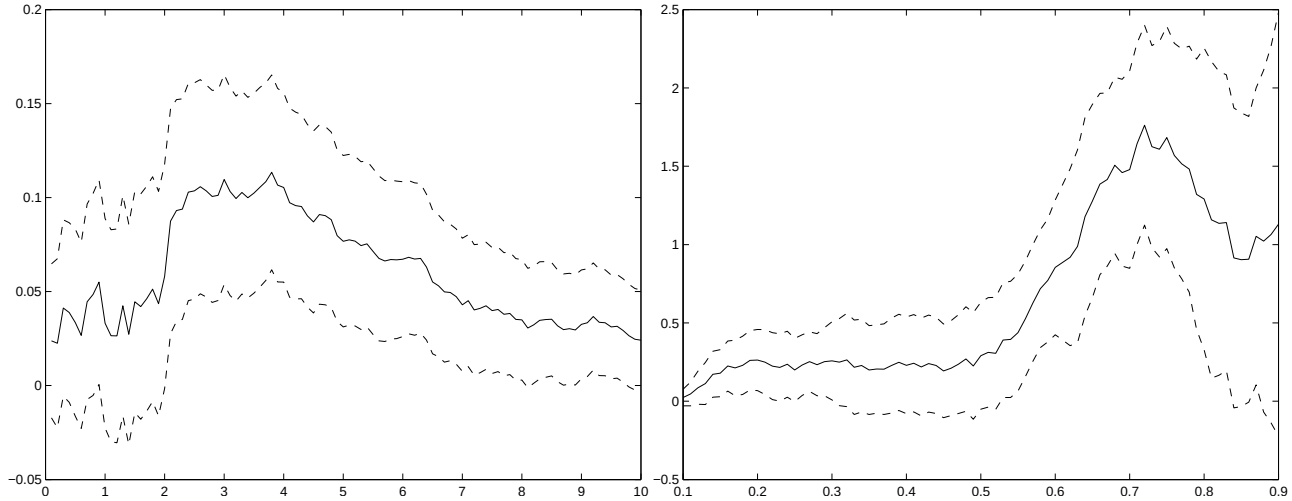
The main assumption of difference-in-differences analysis is the parallel trends assumption. Although control groups and treatment groups are allowed to differ, the changes in the outcomes for the treatment group are assumed to equal the changes for the control group from the pre- to the post-reform period in the absence of a reform. The main test of this assumption is to

Figure 3: Analysis of earnings test reform.

(a) Kernel density estimates of the income for the treatment group, left panel, and the control group, right panel. Pre-treatment distributions are stippled. Post-treatment distributions are solid lines.



(b) Raw difference-in-difference estimates on the distribution function (left panel) and quantile function of incomes. Stippled lines indicate 95 percent confidence bands.



(c) Trends in key outcome 1(Y>2G) (left panel) and income at age 66 (pre-treatment value), before and contemporaneously with the reform.

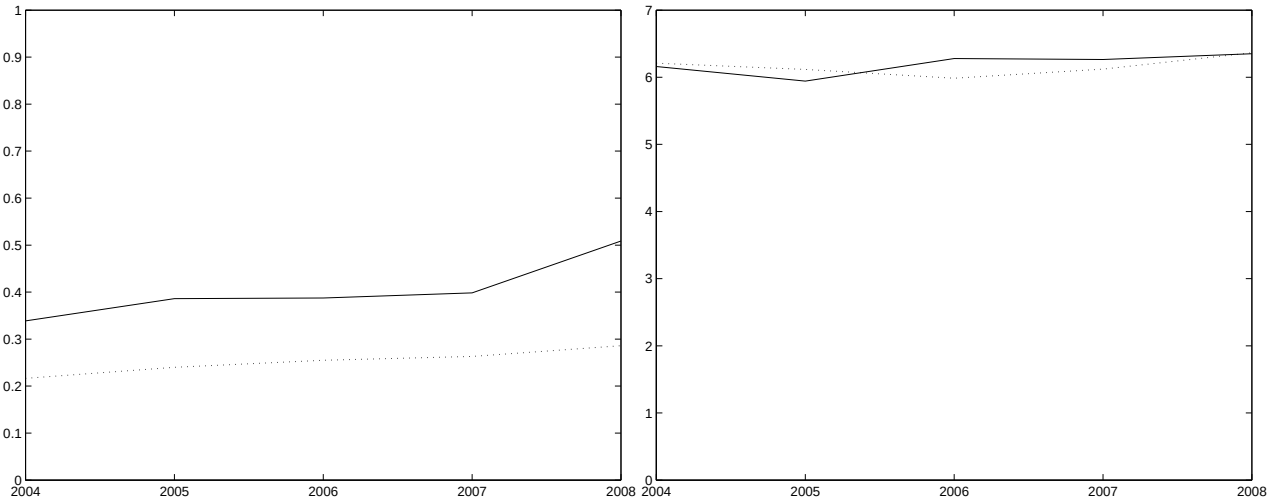


Table 4: Robustness test, Earnings test reform: men

	Test 1: Placebo treatment	Test 2: Public sector workers
trimmed mean	-0.043	-0.071
	(0.238)	(0.298)
Prob(Y>0)	-0.015	0.011
	(0.025)	(0.041)
Prob(Y>2G)	0.017	0.033
	(0.034)	(0.055)
Prob(Y>2G)	-0.014	-0.002
	(0.033)	(0.057)
Prob(Y>4G)	-0.001	-0.007
	(0.028)	(0.050)

check the pre-reform trends. Indeed, this is one of the main reason why we included a longer pre-reform period in the analysis. Column 4 reports difference-in-differences estimates when different linear trends for the treatment and control groups are included as controls in the model. (These 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. Figure 3c shows the trends in share of individuals with earnings more than 2G and earnings just before retirement (earnings at 66) over the period 2004 to 2008. As we can see there is no indication of different trends between the treatment and control groups.

In Table 4, we report results from two additional robustness checks. In column 1, we report results from a placebo treatment test. The estimates correspond to the raw difference-in-differences estimates in Table 3, except that data for 2005 and 2006 are used instead of data for 2007 and 2008. Since there was no reform in 2006, we should not find an effect of the reform. If we found a positive effect, this would point to spurious factors that might also generate the reform effects reported in Table 3. As reported, we find no effect from the placebo reform, which support our model specifications.

Robustness test 2 of Table 4 continues in the same fashion by using the correct reform years, but studying workers who were employed in the public sector at age 66. As discussed in Section 2, these were excluded from the main analysis population because the incentive changes they faced was different from and much weaker than for the main population of analysis. We should thus expect to find weaker or no reform effects for the public sector workers - and as reported there are no such effects.

Table 5: Tax revenue changes due to the earnings test reform

Labor earnings before reform	2G	3G	4G
Increased pension payment (I)	0.00	0.40G	0.80G
Increased labor income	0.60G	1.62G	1.02G
Tax on increased pension (II)	0.00	0.14G	0.28G
Tax on increased labor income(III)	0.27G	0.72G	0.45G
Payroll tax on increased labor income (IV)	0.06G	0.16G	0.10G
Revenue effect (II+III+IV-I)	0.33G	0.62G	0.03G

Note: The increased labor income is based on the raw estimates in Table 3. For a given pre-reform labor earnings, we first calculated the corresponding quantile. The increased labor income is then simply the estimated unconditional quantile treatment effect. Marginal tax rate on labor income is set as 45% and payroll tax rate as 10%.

There is thus no indication that the difference-in-difference effects of the reform are spurious. The estimates suggest an increase of almost 50% in labor income for a individual with pre-reform labor income around 3G (which is corresponding to the third quantile). Note that the effective marginal tax rate changed from 70% to 45% after the earnings test reform. The response corresponds to an earnings elasticity with respect to a net-of-tax rate of approximate 0.6. This is somewhat larger than Hernæs and Jia (2012) found (0.3-0.4) for an earlier Norwegian earnings test reform where the earnings test threshold was changed from 1G to 2G. The main reason for the difference, we suspect, is due to the fact that we have limited our sample only to those who clearly have an incentive to respond, whereas the results in Hernæs and Jia (2012) are based on the whole population, including also those who will not respond, for example the public sector retirees.

An interesting question, if a bit off-topic, is how much government tax revenue and pension layouts are affected by such a reform. We provide a brief back-of-envelope-calculation to illustrate the order of magnitude of such revenue effects. After the earnings test reform, the government will pay out higher pensions, however the extra cost may be offset by increased income taxes on the pensions and increased labor earnings from the retirees, as well as increased payroll tax collected from employer. In Table 5, we list calculations for three different levels of pre-reform labor earnings. As we can see, there are positive tax revenue gains for all three cases. The earnings test presumably reflected a perceived trade-off between deadweight loss due to distorted incentives and a positive effect on public finances. As it does not seem that the earnings test had a positive effect on public finances, the abolishement of the earnings test must be considered a successful reform. The caveat is that the reform may have had some negative revenue effects through public sector workers since there does not seem to have been much

response for such workers.

6.2 Effects of the NIS maturation

This section describes the analysis of the change in incentives to continue to work the year one reaches 68 or 69 years as a consequence of maturation of the NIS rules. The exposition of the analysis is structured the same way as the analysis of the earnings test reform above.

As described in Section 2, the income history dependent part of the NIS old age pension depends crucially on the number of years with earnings in excess of 1G. More than 40 years does however not give any payoff. Since the first year of pension generating incomes was 1967, up to 2006, individuals who are younger than 70 will always have social security wealth increased by around 2.5% as long as they maintain earnings above 1G. In contrast, 2007 is the first year when it was possible to earn in excess of 1G without increasing the accumulated social security wealth. This maturation of the NIS old age pension rules changes the working incentives for persons aged from 67-69 who had a full history of earnings in excess of 1G from 1967 up to age 66. For example, an individual born in 1937 (aged 69 in 2006) with full history will increase social security wealth through working in all three years after reaching age 67. An individual with similar earnings history born in 1938 (aged 69 in 2007) will only increase social security wealth by working in two out of the three years between 67 and 69. However, if this individual does not have a complete working history, no matter which cohort she was born (1937 or 1938), she can increase her social security wealth by working all three years.

We take advantage of this change, and define the treatment group and control group according to the working history from 1967 up to age 66. The treatment group consists of all individuals in the year they turn 68 or 69 years, who had full earnings history up to age 66, while similar individuals, but without full earnings history constitute the control group. The main analysis sample thus includes 68 years old and 69 year olds in 2006 and 2007, with treatment taking place in 2007.

When they were at age 66, all those in treatmetn group in 2007 were in a situation where it is possible to have a gap in the working history from age 67-69 without any loss on social security wealth. In theory, those individuals can decide not to work at any one year (or two years for the 68 year olds) during age 67 to 69 as responses to the rule change. However, among those who have a one-year gap in working history during the ages 67-69, around 83% will had it at age 69, 19% at age 68 and only 4% at age 67. So we assume that if the rule change will

Table 6: Descriptive statistics, population for analysis of NIS maturation, men. Summary statistics for outcomes at age 68 and 69 pooled and control variables (measured at age 66), 2004-2006, 2006 and 2007. Means, standard deviations for non-binary variables in parentheses, quantiles when indicated.

	Treatment group			Control group		
Outcome in	2004-2006	2006	2007	2004-2006	2006	2007
Dependent variables						
Income (trimmed)	1.940	1.993	2.145	2.130	2.200	2.386
	(3.008)	(2.960)	(3.019)	(3.207)	(3.169)	(3.373)
Income > 0	0.754	0.763	0.788	0.771	0.790	0.788
Income > 1G	0.481	0.502	0.536	0.503	0.526	0.545
Income > 2G	0.275	0.286	0.313	0.308	0.322	0.338
Income > 4G	0.147	0.153	0.171	0.164	0.169	0.191
Income > 6G	0.085	0.092	0.098	0.097	0.103	0.121
1st Quartile	0.019	0.036	0.074	0.046	0.105	0.115
Median	0.882	1.012	1.220	1.014	1.131	1.230
3rd Quartile	2.102	2.159	2.349	2.407	2.498	2.806
Control variables						
Income, age 66	6.407	6.357	6.528	5.379	5.420	5.720
	(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

induce some individuals who would otherwise work and earn more than 1G to adjust their labor supply behavior, they will shave off one year from their working career at the end (and not by relaxing at 67 and then working again at age 68 or relaxing at age 68 and working again at age 69.) Then we can measure the treatment effect with the earnings at age 68 and 69.

Table 6 shows descriptive statistics for the treatment and control group for the men. As is clear, the populations are a bit bigger than for the analysis of the earnings test reform. In this analysis we study 68-year and 69-year olds throughout the year. A majority of about two-thirds is in the treatment group, that is, with earnings in excess of 1G every year back to 1967.

The treatment group have higher earnings and more self-employed than the control group at age 66. The control group have more employees with access to private occupational pensions. The residual group of employees without access to private occupational pensions are of about

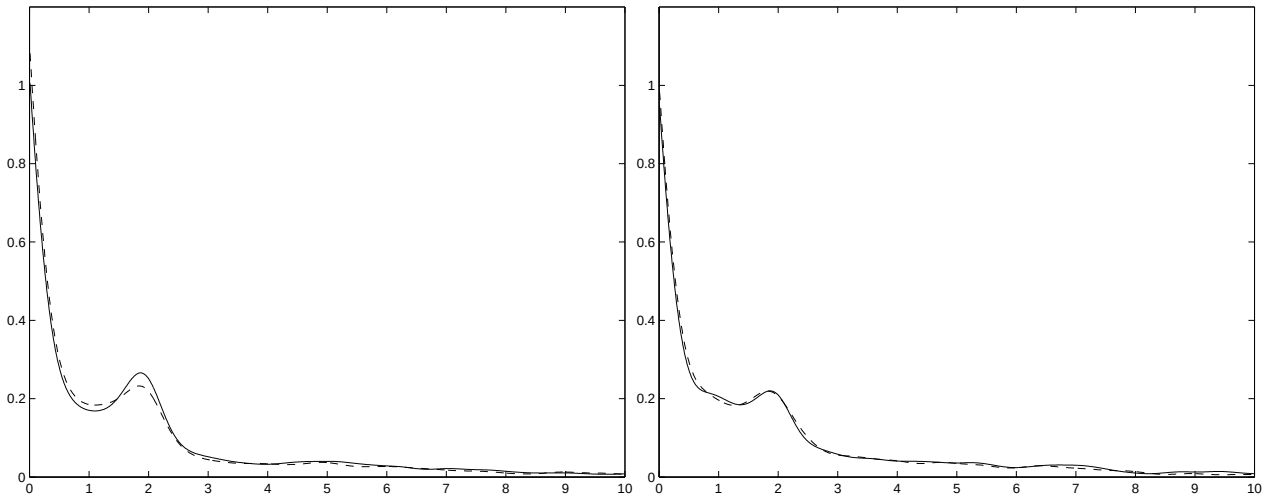
the same size. In terms of educational level, the treatment and control groups are similar. There are no dramatic changes in the control variables for either the treatment or the control group from the pre-reform to the post-reform period. In terms of the outcome variable, income at age 68 and 69, the control group has a slightly higher mean throughout. The mean income is quite a bit higher than the critical threshold at 1G. Both the treatment and control group get a slight increase in mean incomes moving from the pre- to the post-treatment period.

Figure 4a reports kernel density estimates of the earnings for the treatment and control groups in the pre- and post treatment period. We see the bunching at 2 G associated with the earnings test (which was still in place in both 2006 and 2007). Note that before reform (in 2006), both the control and treatment group have incentives to maintain earnings at least 1G, due to the existence of the "notch" in the "lifetime" budget lines, as shown earlier in Figure 2. If they respond to these incentive, we should have observed a bunching around 1G as predicted by theory. However, we do not see any bunching associated with the social security wealth premium for earning at least 1 G. This fact indicates that inviduals in our sample, for some reason, do not adjust their work efforts with respect to potentially increased future pension entitlements. Therefore, we expect that the maturation will not lead to any strong changes, which is confirmed by the difference-in-differences analysis, as shown by the raw distributional and quantile based difference-in-differences graphs in Figure 4b. If anything, the treatment group has a slight increase in the number of people earning just above 1 G, which is the opposite of the predicted effect of the incentive change, but 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 maturation of the National Insurance Scheme. Figure 4c demonstrates that there is no strong pre-reform trends specific to either control or treatment group that could work to cancel out effects of the reform. Details of the difference-in-differences analysis from different specifications are given in Table 7. The bottom line is that there is no discernible effect of the reform and that this can be stated with some precision.

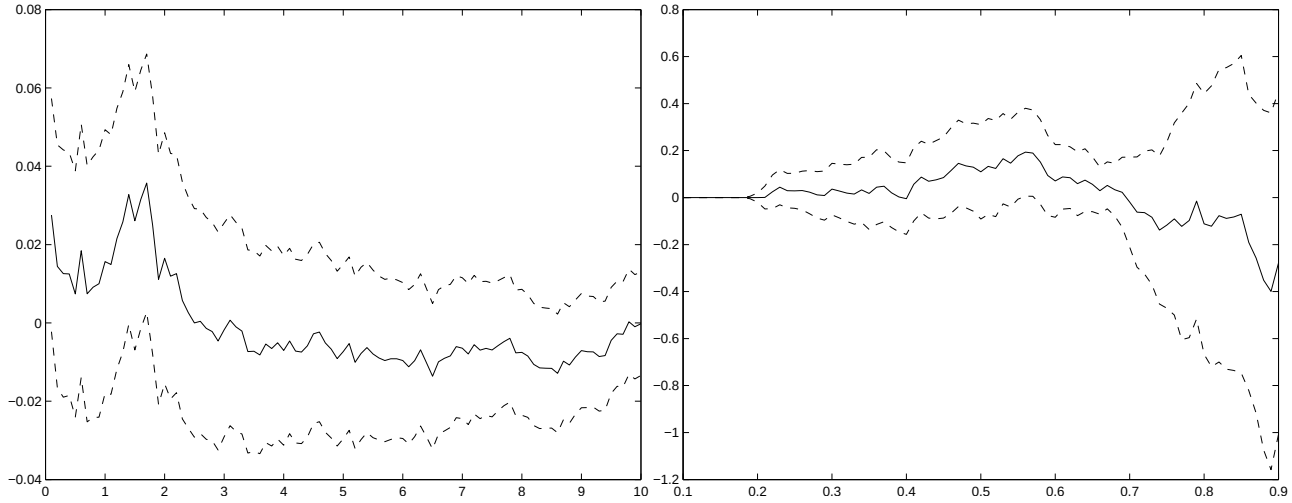
Similarly to the case of the earnings test above, we have also run robustness tests (Table 8). Statistically significant treatment effects in the robustness tests would indicate that something is wrong with the assumptions, e.g. the parallel trends assumptions - which might hide any effects that should have been there in the main analysis. However, we do not find any effects for the placebo treatment or the public sector workers. Note that we cannot attribute this unresponsiveness as a result of labor market restrictions, since our analysis on earnings test above

Figure 4: Analysis of National insurance scheme maturation. Men.

(a) Kernel density estimates of the income for the treatment group, left panel, and the control group, right panel. Pre-treatment (2006) distributions are stippled. Post-treatment (2007) distributions are solid lines.



(b) Raw difference-in-difference estimates on the distribution function (left panel) and quantile function of incomes. Stippled lines indicate 95 percent confidence bands.



(c) Trends in key outcome 1(Y>1G) (left panel) and income at age 66 (pre-treatment value), before and contemporaneously with the reform.

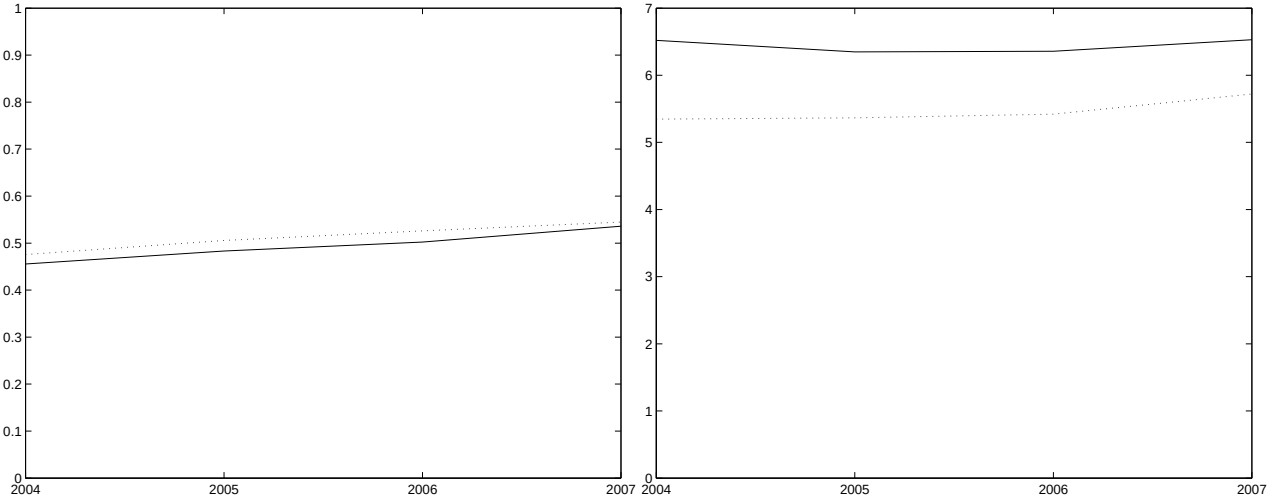


Table 7: Difference-in-differences estimates of the effects of NIS maturation, men

	Raw estimates	Estimates with controls	Estimates with controls and ext. pre-reform period	Estimates with controls and separate linear trends
trimmed mean	-0.034	0.013	0.117	0.019
	(0.099)	(0.090)	(0.094)	(0.124)
1(Y>0)	0.028	0.024	0.021	0.019
	(0.015)	(0.015)	(0.014)	(0.020)
1(Y>1G)	0.015	0.015	0.026	0.012
	(0.017)	(0.017)	(0.016)	(0.023)
1(Y>2G)	0.012	0.013	0.023	0.018
	(0.016)	(0.015)	(0.015)	(0.021)
1(Y>4G)	-0.005	-0.001	0.011	-0.006
	(0.013)	(0.012)	(0.012)	(0.016)
1st Quartile	-0.009	0.000	0.000	0.000
	(0.026)	(0.023)	(0.003)	(0.004)
Median	0.139	0.133	0.194	0.070
	(0.102)	(0.094)	(0.081)	(0.108)
3rd Quartile	-0.134	-0.059	0.089	0.003
	(0.171)	(0.159)	(0.153)	(0.222)
Controls		x	x	x
Data from 2004-2005			x	x
Group-specific trends				x

Notes: Controls are listed in Table 6 as they appear in the model - except Earnings at age 66 which enters as a third order polynomial.

already showed that the elderly workers can adjust their work efforts when economic incentives of working changes. In other words, our results seem to indicate that economic incentives in terms of increased future pensions are at least not as attractive as immediate gains.

In contrast to the earnings test reform, which was announced only few months before implementation, the information on maturation of the NIS rules were publicly available as early as when the NIS was established in 1967. Individuals have the opportunity to adjust their labor supply behavior to take advantage of the implicit rule change in 2007 and this may cause self-selection into control and treatment groups. From 2007, individuals with "gaps" in their working history before 66, can still receive full old age pension from age 70. Essentially, the only reason why the change we study should affect whether people have complete earnings histories or not is that it is less costly for persons in the post-treatment period to have gaps in their earnings history, because they plan to close that gap by working during ages 67-69 - and only for that reason. Under the assumption that these individuals who plan to work during ages

Table 8: Robustness test, NIS maturation: men

	Test 1: Placebo treatment	Test 2: Public sector workers
trimmed mean	-0.087	-0.227
	(0.113)	(0.283)
Prob(Y>0)	0.015	0.005
	(0.017)	(0.028)
Prob(Y>2G)	-0.002	-0.012
	(0.019)	(0.034)
Prob(Y>2G)	-0.016	0.003
	(0.017)	(0.035)
Prob(Y>4G)	-0.001	0.000
	(0.014)	(0.035)

67-69 work and earn more than the average in the treatment and control groups in the same years, such misallocation to control and treatment groups in the post-reform period should lead to higher earnings in the control group and lower earnings in the treatment group - hence a downward bias in the estimated effect of the reform.

Since the expected effect of the treatment is negative, i.e., there are less individuals working in response to lower compensation, our difference-in-differences estimate will be biased away from zero. However, even with this bias, our estimates fails to pick up any significant effect.

7 Concluding discussion

We find statistically significant and economically substantial effects of the earnings tests reform for 67-year old men, using our difference-in-differences approach on the earnings distribution. 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 abolished. Hence the earnings test has a sizable impact on labour supply of the relevant target group that we study here - 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 should not be affected by the reform. Smaller sample sizes make it more difficult to reach strong conclusions for women, though point estimates point towards similar effects as for men.

In contrast, we do not find any effect of the maturation of the National Insurance Scheme that only affects incentives through the accrual of entitlements to future pensions . There seems to be no bunching in the earnings distribution around or above the threshold that gives a sub-

stantial boost in the value of future pensions and no change in the earnings distribution when this threshold (and boost) disappears. Indeed our statistically insignificant point estimates even show the wrong sign. Again, small sample sizes make it hard to reach sharp conclusions for women, but results are largely confirming the results for men. These null results are particularly interesting in the light of the strong responses to the earnings test reform for the same groups, which show that our subjects are not constrained in the labour market.

The two changes in incentives can roughly be interpreted as being of the same order of magnitude for large parts of the sample, even though the sizes of the two changes in incentives are not directly comparable. The effect of lifting the earnings test depends on the level of potential earnings. The effect of the maturing depends on the level of pension entitlements and applies to the remaining life time stream of pension. The differences between them is that one is an immediate sum and the other a future income stream, and that while both work through the NIS pension system, the former may be considered simpler or more transparent that the latter. The results show that only the former has a strong effect on behavior. The groups affected by the two changes in incentives are also just about identical: 67-69 year old persons who were working in the private sector at age 66, either as wage earners or self-employed. There are essentially two potential explanations for different responses to incentives. The first explanation is that 67-year olds discount future incomes heavily. The second explanation is that the potential pensioners do not understand the complicated pension rules and fail to acknowledge the potential gains from one extra year of working.

While we have no data to bolster this claim, it does not seem reasonable that the main explanation for the different responses is heavy discounting of future incomes. The individuals we study are on average rather well-off, with above average incomes and, obviously, rather long working careers. Hence these are largely well-off people, usually owning their own residences and with mortgages that are of much lower values than these residences - due to both increased property prices and down payments over the years. Hence, there is essentially no reason to expect a large number of our subjects to be rationed in the credit market. Valuation of current and future consumption that differs from market interest rates should then primarily be corrected through activity in the credit market rather than labour market behavior.

Thus, we are left with the likely explanation that individuals do not respond to the changes in incentives in terms of future pension entitlements because these incentives are intransparent or at least not sufficiently well understood by the subjects. It is worthwhile dwelling on a

few points of this likely explanation. A few months before reaching her 67th birthday, the subject receives a letter with an attached information folder on relevant pension rules from her local Social Security Administration offices. The necessary information for figuring out the incentives is included in these folders. Based on the information that is handed out, one can easily obtain the relevant details of the pension system (will incomes be earnings tested, will you earn more pension entitlements from continued working and work out quite accurate estimates of gain or loss from continuing working. Since the information is available, the lack of response must be interpreted as a lack of willingness or ability to unravel the consequences of the system or alternatively, a prior belief that the effort of studying how the system works will not lead to substantial net gains through adjustment of own behavior. In this context, the main difference between the two changes in incentives that we study is that it is very easy to understand the earnings test, while it may be more demanding to calculate "correctly" the value of a gain in terms of increased pension entitlement, due to the relative complexity of the system determining future pension entitlements. Furthermore, while the folder contains all the detailed information, the earnings are also clearly described in the cover letter, which we guess was more widely read than the folder.

We believe that these insights may have some validity also with respect to other pension systems than the Norwegian. 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 as well. The arguably hardest part for potential retirees is the evaluation of the value of a future income stream in an unknown number of years under not yet decided tax systems. This part is intrinsic in the evaluation of pension entitlements. Our results may indicate that taxing current income while at the same time offering incentives in terms of future pension entitlements - which is a major part of most public old age pension systems - may indeed lead to what we can frame as a "deadweight loss due to intransparent incentives".

In a pension reform current being implemented in Norway, pensions in the NIS are actuarially adjusted and delinked from earnings. Employers are also required to try to accommodate requests for part time work. The idea is to improve work incentives and give flexibility to adjust earnings, pensions and consumptions. If people have as bleak a view of future pension incomes as suggested by our results, they can be expected to take out pension as early as possible - transforming the intransparent currency of future pension entitlements into hard cash. Indeed, this seems to have happened. However, the new pension system does not require people to quit

working, or even quit working full time, to take up pensions. So our results give the dual prediction of a large degree of early pension take-up while at the same time inducing more work during those years when pension is an option, particularly on the intensive margin.

References

Angrist, J. and J. S. Pischke (2009) *Mostly harmless econometrics: an empiricist's companion*: Princeton University Press.

Athey, S. and G. W. Imbens (2006) "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, pp. 431–497.

Baker, M. and D. Benjamin (1999) "How do retirement tests affect the labour supply of older men?," *Journal of Public Economics*, Vol. 71, No. 1, pp. 27–52.

Chernozhukov, V. and C. Hansen (2005) "An IV model of quantile treatment effects," *Econometrica*, Vol. 73, No. 1, pp. 245–261.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011) "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *The Quarterly Journal of Economics*, Vol. 126, No. 2, p. 749.

Coile, Courtney and Jonathan Gruber (2007) "Future Social Security Entitlements and the Retirement Decision," *The Review of Economics and Statistics*, Vol. 89, No. 2, pp. 234–246, 03.

Disney, R. and S. Smith (2002) "The labour supply effect of the abolition of the earnings rule for older workers in the United Kingdom," *The Economic Journal*, Vol. 112, No. 478, pp. C136–C152.

Engelhardt, G.V. and A. Kumar (2009) "The repeal of the retirement earnings test and the labor supply of older men," *Journal of Pension Economics and Finance*, Vol. 8, No. 04, pp. 429–450.

Firpo, S., N.M. Fortin, and T. Lemieux (2009) "Unconditional quantile regressions," *Econometrica*, Vol. 77, No. 3, pp. 953–973.

Frederick, Shane, George Loewenstein, and Ted "O'Donoghue (2002) "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, Vol. 40, pp. 351–401.

Friedberg, L. (2000) "The labor supply effects of the social security earnings test," *Review of Economics and Statistics*, Vol. 82, No. 1, pp. 48–63.

Friedberg, L. and A. 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. CRR WP*, Vol. 2006, p. 27.

Gruber, J. and P. Orszag (2003) "Does the Social Security earnings test affect labor supply and benefit receipt?" *National Tax Journal*, Vol. 45, pp. 1–22.

Havnes, T. and M. Mogstad (2010) "Is Universal Child Care Leveling the Playing Field? Evidence from Non-Linear Difference-in-Differences," *IZA Discussion Papers*.

Hernæs, E. and Z. Jia (2012) "Earnings Distribution and Labour Supply after a Retirement Earnings Test Reform," *Oxford Bulletin of Economics and Statistics*, Vol. forthcoming.

Imbens, G.W. (2004) "Nonparametric estimation of average treatment effects under exogeneity: A review," *Review of Economics and Statistics*, Vol. 86, No. 1, pp. 4–29.

Imbens, G. W. and J. M. Wooldridge (2009) "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, Vol. 47, No. 1, pp. 5–86.

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

Rust, John and Christopher Phelan (1997) "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, Vol. 65, No. 4, pp. 781–832, July.

Saez, E. (2010) "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, Vol. 2, No. 3, pp. 180–212.

Samwick, Andrew A. (1998) "New evidence on pensions, social security, and the timing of retirement," *Journal of Public Economics*, Vol. 70, No. 2, pp. 207–236, November.

Song, J.G. and J. Manchester (2007) "New evidence on earnings and benefit claims following changes in the retirement earnings test in 2000," *Journal of Public Economics*, Vol. 91, No. 3-4, pp. 669–700.

Stock, James H and David A Wise (1990) "Pensions, the Option Value of Work, and Retirement," *Econometrica*, Vol. 58, No. 5, pp. 1151–80, September.

A Corresponding analysis for women

Earnings test reform for women

The population size is substantially smaller for women than for men, both because a larger share of men are still working at age 66 and because a larger share of the women work in the public sector(Table 9). Table 10 reports pre- and post-reform descriptive statistics for the treatment and control group among women. Note that there are some strikingly large changes in background characteristics from the pre-reform to the post-reform cohorts. These changes are however roughly similar for the treatment and control groups. Looking at the income data, it is clear that the women have lower incomes than comparable men. The increase in incomes that is contemporaneous with the reform for the treatment group is about 0.3 G or an increase of about 20 percent. This reflects an increase in women earning more than 1G and in particular more than 2G. Also the control group face an increase in income of about 0.1 G, which primarily reflects an change where some people in the pre-reform period earned between 0 and 1 G now earn between 1 G and 2 G. Most of the action is for the treatment group in the upper half of the income distribution, although there is some action at the median as well. The top panel in Figure 5 provides kernel density estimates of the income distribution for the treatment group and the control group. It is clear that there is some bunching around the kink point of 2 G for the control group and for the treatment group in the pre-reform period. The bunching becomes less marked, but does not entirely go away with the reform, which may or may not be coincidence. The treatment group is the post-reform period consist of about 400 women, where about half earn less than 1.1 G anyway, so there is sparse information to estimate the income densities.

Table 11 shows the results of difference-in-difference analysis as well as the robustness test results. The means and standard deviations in outcome variables and potential control variables

for the treatment group and the control group in 2007 and 2008. As suggested by the numbers in the above paragraph, the raw difference-in-difference on the mean income is about 0.2 G, which is however not statistically significant. The number is not negligibly small though, suggesting a 15 percent increase in mean incomes. The result is of the same magnitude and statistically significant at the 5 percent level when the larger pre-reform sample is used (column 3) a long with controlling for background variables. The change reflects an increase in the probability of having more than 2G and 4G in income, corresponding to changes in the upper half of the income distribution. Figure 5 shows that most of the action for the women goes on above the 70th percentile, corresponding to about 2 G of income. As is seen from the Table, some of the distributional and quantile treatment effects are statistically significant even with the short pre-reform period in Table 10, and indeed the strongest results are for the probability of earning more then 2 G - the key prediction from our labor supply model. The results for women are not robust to inclusion of control and treatment group specific time trends. Our interpretation is that this primarily reflects a lack of precision - no statistically different trends are estimated. We also see this in Figure 5. For the key distributional outcome parameter 1(I>2G), there are no clear pre-reform trends for either the treatment or control group. Still, the precision of the difference in difference parameter for this outcome disappears when we allow for group specific trends.

The placebo and public sector robust analysis go through, with the exception of the curious result that the placebo reform seems to affect the income distribution in the very top (Table 12).

It seems fair to conclude that the earnings test reform affected women as well as men, though our smaller population sizes for women mean that we need to be somewhat more cautious in the conclusion. Like the results for men, the reform seems only to affect the part of the income distribution that would anyway earn almost 2 G, as predicted by the continuous choice model. Allowing for the fact that this is a smaller share of the women than the men, the elasticity for women may be considered to be of almost the same magnitude as that for men.

Effects of the NIS maturation for women

Table 13 shows the descriptive statistics for the analysis sample for women. Compared with men, there are far few individuals belongs to the treatment group. Most of women have gaps during their working history from 1967. The difference in difference results are presented in Table 14 and illustrated in Figure 6. Due to the smaller sample size for the treatment group, the

estimates are less precise as the case for men. But main results remain as the same. Similar to the analysis for men, we check the model setup using two robustness tests (Table 15). Again, the robustness tests go through.

B Supplement tables and figures

Table 9: Average annual income at age 68 by labor market status at age 66. Classification of 66-year olds women by January 1st 2003-2006 .

	Number	Fraction of full population	Mean earnings at age 68	Fraction of total earnings
Women				
Not active in labor market				
AFP early retirement	14238	0.19	0.21	0.10
Disability pension	35048	0.47	0.12	0.13
Occupational pension	3607	0.05	0.26	0.03
NIS pension	657	0.01	0.27	0.01
Idle	8185	0.11	0.44	0.12
Some work	4135	0.05	0.52	0.07
Active in labor market				
Public sector employee	3147	0.04	2.21	0.22
Analysis population	6295	0.08	1.56	0.32
Full population	75312	1.00	0.41	1.0

Notes: AFP early pension scheme does not include those who are only 20 or 40 percent pensioners. Occupational pensions and NIS pensions are defined based on receiving at least 1 G in pensions. "Idle" defined based on no income while "some work" means less than 1 G in income.

Table 10: Descriptive statistics, population for analysis of means testing reform, women. Summary statistics for outcomes at age 68 (treatment group) and age 69 (control group) and control variables (measured at age 66), 2004-2007, 2007 and 2008. Means, standard deviations for non-binary variables in parentheses, quantiles when indicated.

	Treatment group			Control group		
	Before reform		After reform	Before reform		After reform
	2004-2007	2007	2008	2004-2007	2007	2008
Dependent variables						
Income (trimmed)	1.572	1.625	1.937	1.036	1.051	1.152
	(1.978)	(1.974)	(2.205)	(1.787)	(1.797)	(1.786)
Income > 0	0.868	0.870	0.850	0.610	0.623	0.638
Income > 1G	0.480	0.489	0.539	0.333	0.339	0.375
Income > 2G	0.262	0.264	0.333	0.164	0.171	0.176
Income > 4G	0.102	0.117	0.169	0.066	0.067	0.076
Income > 6G	0.039	0.042	0.062	0.023	0.021	0.022
1st Quartile	0.254	0.255	0.205	0.000	0.000	0.000
Median	0.902	0.959	1.160	0.214	0.220	0.395
3rd Quartile	2.032	2.075	2.866	1.516	1.590	1.736
Control variables						
Income, age 66	4.101	4.097	4.279	3.997	4.000	4.178
	(2.293)	(2.350)	(2.150)	(2.176)	(2.180)	(2.342)
Self-employed	0.225	0.196	0.231	0.225	0.214	0.221
Private occ. pensions	0.203	0.191	0.174	0.210	0.224	0.198
Education in years	10.533	10.627	11.023	10.430	10.554	10.745
	(2.289)	(2.215)	(2.438)	(2.288)	(2.348)	(2.320)
Education > 10 years	0.281	0.303	0.329	0.262	0.284	0.316
Education > 12 years	0.140	0.147	0.204	0.133	0.140	0.167
Education > 14 years	0.048	0.046	0.083	0.050	0.063	0.064
Population size	1496	409	432	5608	1595	1622

Table 11: Difference-in-differences estimates of the effects of abolishing mean-testing. Women.

	Raw estimates	Estimates with controls	Estimates with controls and ext. pre-reform period	Estimates with controls and separate linear trends
trimmed mean	0.210	0.170	0.234	0.100
	(0.155)	(0.138)	(0.108)	(0.160)
1(Y>0)	-0.036	-0.039	-0.051	-0.063
	(0.030)	(0.029)	(0.023)	(0.034)
1(Y>1G)	0.015	0.012	0.014	0.023
	(0.039)	(0.037)	(0.029)	(0.046)
1(Y>2G)	0.064	0.055	0.054	0.057
	(0.034)	(0.032)	(0.025)	(0.037)
1(Y>4G)	0.042	0.036	0.053	0.032
	(0.025)	(0.023)	(0.019)	(0.027)
1st Quartile	-0.065	-0.084	-0.080	-0.113
	(0.058)	(0.053)	(0.041)	(0.050)
Median	0.043	0.025	0.069	0.087
	(0.156)	(0.142)	(0.120)	(0.247)
3rd Quartile	0.646	0.535	0.529	0.326
	(0.337)	(0.298)	(0.248)	(0.326)
Controls		x	x	x
Data from 2004-2006			x	x
Group-specific trends				x

Table 12: Robustness test for Earnings Test reform, women

	Test 1: Placebo treatment	Test 2: Public sector workers
trimmed mean	0.275	-0.071
	(0.178)	(0.298)
Prob(Y>0)	0.018	0.011
	(0.033)	(0.041)
Prob(Y>2G)	-0.001	0.033
	(0.045)	(0.055)
Prob(Y>2G)	0.003	-0.002
	(0.038)	(0.057)
Prob(Y>4G)	0.028	-0.007
	(0.026)	(0.050)

Table 13: Descriptive statistics, population for analysis of NIS maturation, women. Summary statistics for outcomes at age 68 and 69 pooled and control variables (measured at age 66), 2004-2006, 2006 and 2007. Means, standard deviations for non-binary variables in parentheses, quantiles when indicated.

	Treatment group			Control group		
	2004-2006	2006	2007	2004-2006	2006	2007
Dependent variables						
Income (trimmed)	1.628	1.575	1.647	1.147	1.143	1.212
	(2.605)	(2.473)	(2.704)	(1.747)	(1.721)	(1.717)
Income > 0	0.696	0.690	0.632	0.708	0.701	0.728
Income > 1G	0.408	0.405	0.395	0.375	0.375	0.403
Income > 2G	0.251	0.256	0.256	0.188	0.191	0.201
Income > 4G	0.137	0.129	0.153	0.070	0.071	0.074
Income > 6G	0.064	0.052	0.074	0.023	0.023	0.021
1st Quartile	0.000	0.000	0.000	0.000	0.000	0.000
Median	0.520	0.542	0.409	0.443	0.434	0.563
3rd Quartile	2.009	2.013	2.023	1.722	1.743	1.815
Control variables						
Income, age 66	5.391	5.321	5.353	3.826	3.827	3.927
	(2.530)	(2.455)	(2.480)	(2.065)	(2.072)	(2.183)
Self-employed	0.338	0.325	0.313	0.210	0.210	0.205
Private occ. pensions	0.226	0.250	0.237	0.206	0.217	0.208
Education in years	11.245	11.200	11.398	10.321	10.426	10.554
	(2.696)	(2.707)	(2.801)	(2.199)	(2.252)	(2.251)
Education > 10 years	0.389	0.376	0.409	0.244	0.258	0.286
Education > 12 years	0.235	0.239	0.262	0.118	0.127	0.140
Education > 14 years	0.086	0.083	0.112	0.044	0.051	0.057
Population size	941	348	367	7370	2606	2850

Table 14: Difference-in-differences estimates of the effects of NIS maturation, women

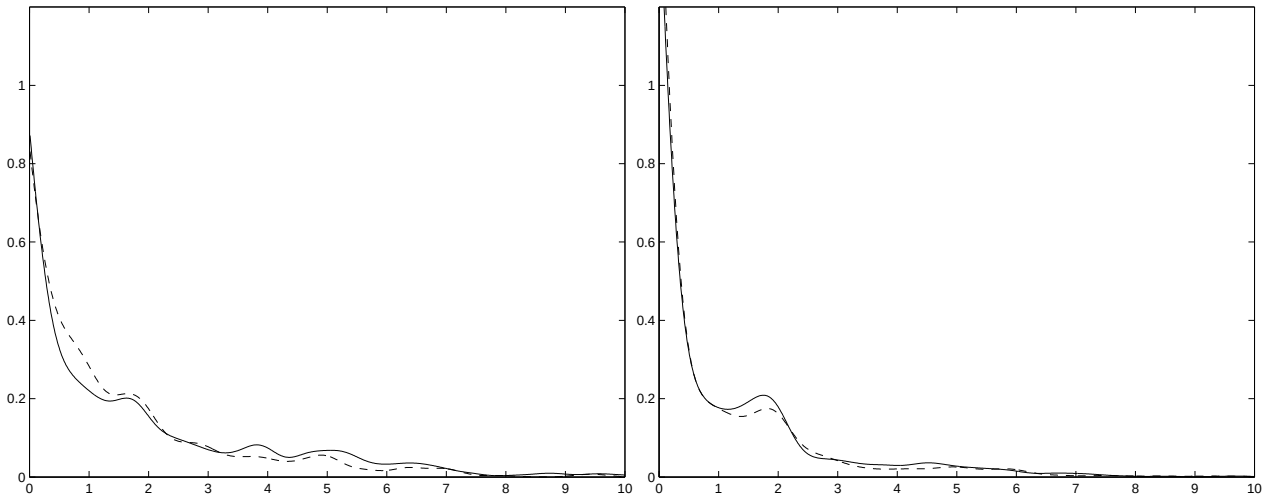
	Raw estimates	Estimates with controls	Estimates with controls and ext. pre-reform period	Estimates with controls and separate linear trends
trimmed mean	0.004	0.003	-0.017	-0.026
	(0.148)	(0.129)	(0.127)	(0.178)
1(Y>0)	-0.084	-0.084	-0.080	-0.077
	(0.033)	(0.032)	(0.029)	(0.042)
1(Y>1G)	-0.038	-0.038	-0.037	-0.042
	(0.031)	(0.030)	(0.028)	(0.042)
1(Y>2G)	-0.010	-0.010	-0.005	-0.009
	(0.029)	(0.027)	(0.025)	(0.038)
1(Y>4G)	0.020	0.019	0.014	0.028
	(0.021)	(0.019)	(0.019)	(0.027)
1st Quartile	-0.021	-0.021	-0.020	-0.022
	(0.008)	(0.008)	(0.006)	(0.009)
Median	-0.259	-0.254	-0.202	-0.333
	(0.104)	(0.099)	(0.098)	(0.173)
3rd Quartile	-0.078	-0.075	-0.055	-0.075
	(0.198)	(0.184)	(0.127)	(0.185)
Controls		x	x	x
Data from 2004-2005			x	x
Group-specific trends				x

Table 15: Robustness test, NIS maturation: women

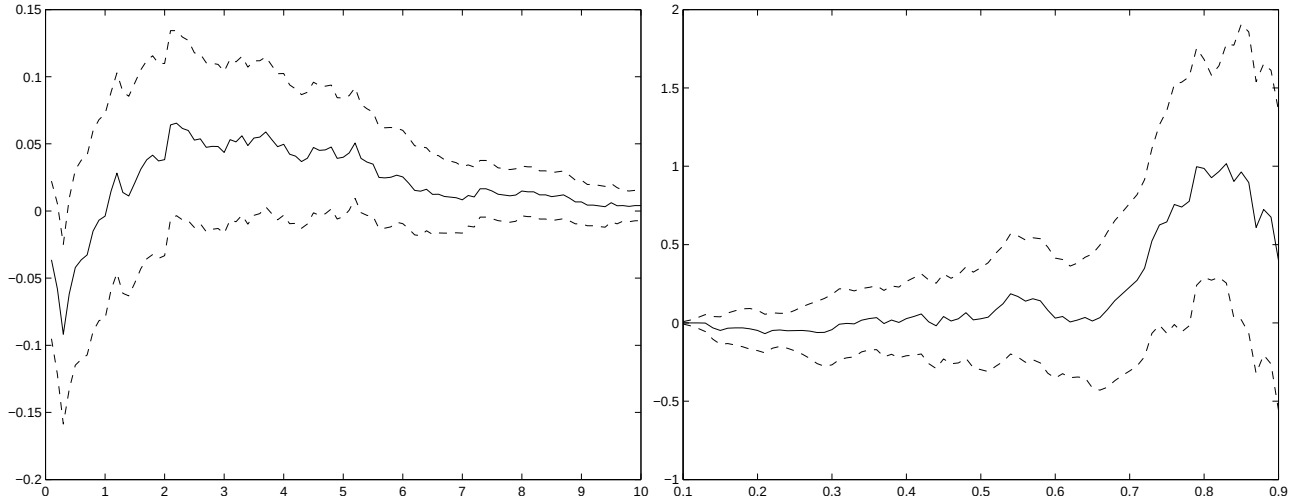
	Test 1: Placebo treatment	Test 2: Public sector workers
trimmed mean	0.165	-0.227
	(0.163)	(0.277)
Prob(Y>0)	-0.005	0.005
	(0.035)	(0.029)
Prob(Y>2G)	0.023	-0.012
	(0.034)	(0.034)
Prob(Y>2G)	0.003	0.003
	(0.029)	(0.035)
Prob(Y>4G)	-0.006	0.000
	(0.023)	(0.034)

Figure 5: Analysis of earnings test reform. Women.

(a) Kernel density estimates of the income for the treatment group, left panel, and the control group, right panel. Pre-treatment distributions are stippled. Post-treatment distributions are solid lines.



(b) Raw difference-in-difference estimates on the distribution function (left panel) and quantile function of incomes. Stippled lines indicate 95 percent confidence bands.



(c) Trends in key outcome 1(Y>2G) (left panel) and income at age 66 (pre-treatment value), before and contemporaneously with the reform.

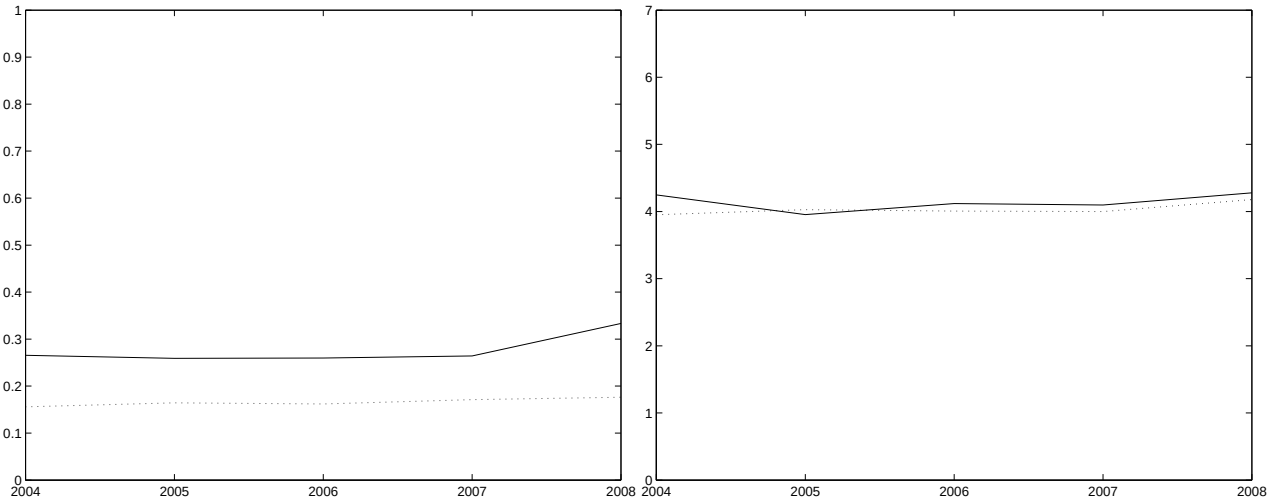
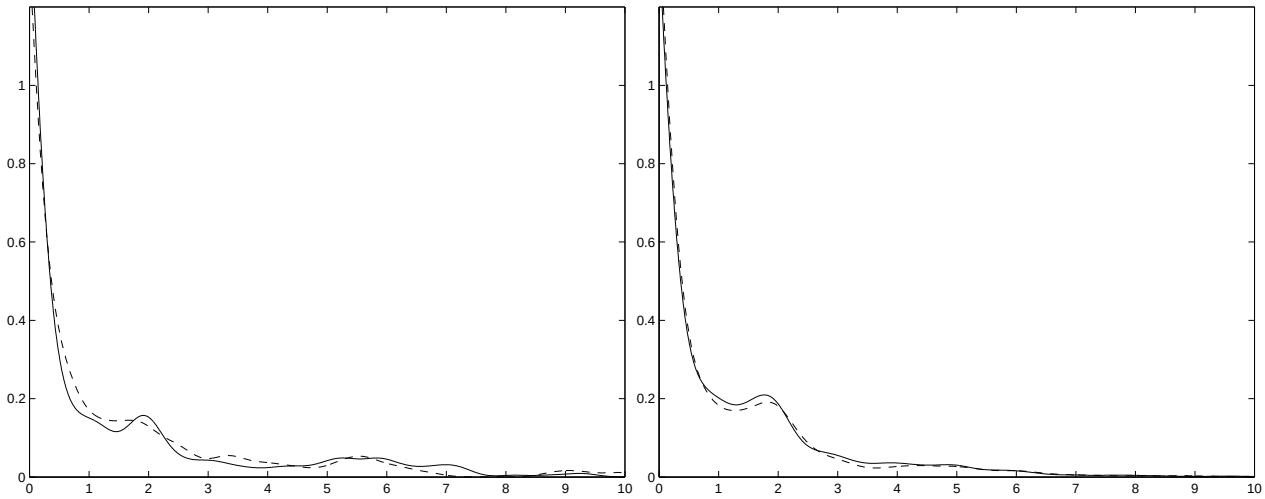
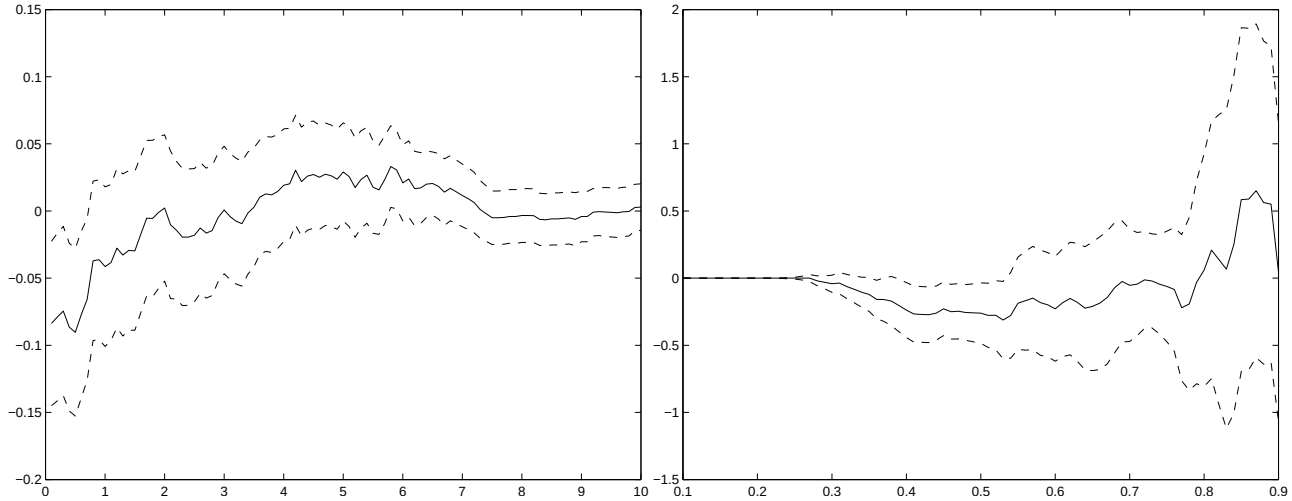


Figure 6: Analysis of National insurance scheme maturation. Women.

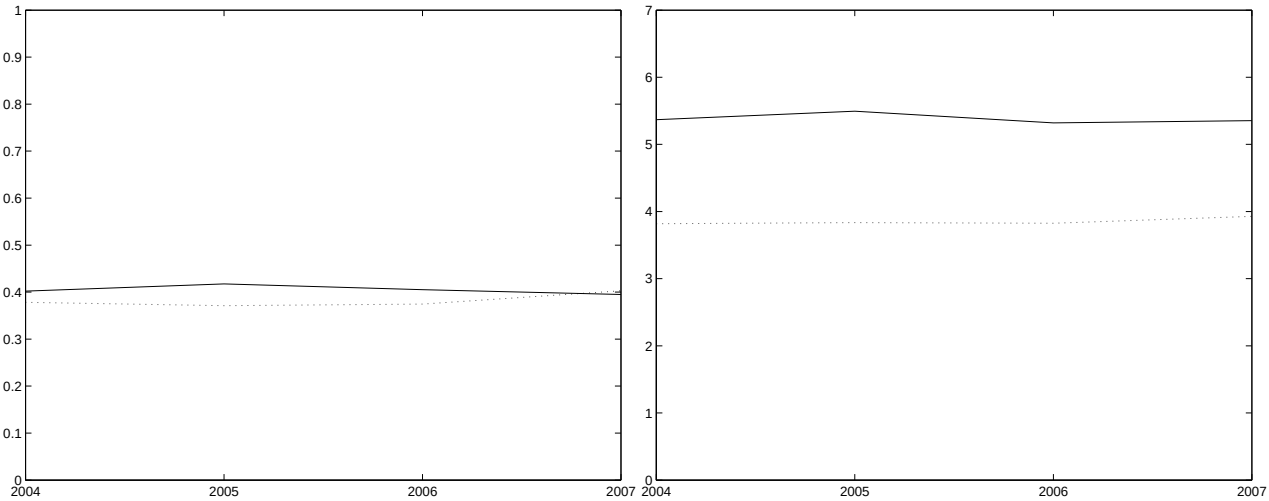
(a) Kernel density estimates of the income for the treatment group, left panel, and the control group, right panel. Pre-treatment (2006) distributions are stippled. Post-treatment (2007) distributions are solid lines.



(b) Raw difference-in-difference estimates on the distribution function (left panel) and quantile function of incomes. Stippled lines indicate 95 percent confidence bands.



(c) Trends in key outcome 1(Y>1G) (left panel) and income at age 66 (pre-treatment value), before and contemporaneously with the reform.



B

Returadresse:
Statistisk sentralbyrå
NO-2225 Kongsvinger

Statistics Norway

Oslo:
PO Box 8131 Dept
NO-0033 Oslo
Telephone: + 47 21 09 00 00
Telefax: + 47 21 09 00 40

Kongsvinger:
NO-2225 Kongsvinger
Telephone: + 47 62 88 50 00
Telefax: + 47 62 88 50 30

E-mail: ssb@ssb.no
Internet: www.ssb.no

ISSN 0809-733X

Statistisk sentralbyrå
Statistics Norway