OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 75, 3 (2013) 0305-9049
doi: 10.1111/j.1468-0084.2012.00695.x

Earnings Distribution and Labour Supply after a Retirement Earnings Test Reform*

ERIK HERNÆS† and ZHIYANG JIA‡

†*Frisch Centre for Economic Research, Gaustadalleen 21, N-0349 Oslo, Norway (e-mail: erik.hernas@frisch.uio.no)*
‡*Statistics Norway, Box 8131 Dep, N-0033 Oslo, Norway (e-mail: zhiyang.jia@ssb.no)*

Abstract

Norwegian administrative data are used to evaluate the impact of a doubling of the threshold in the retirement earnings test. We find almost no impact on the extensive margin, but a positive effect on the intensive margin. This positive effect is uneven over the earnings distribution, and concentrated on workers around the threshold, increasing with exposure to the reform and leading to a decrease in earnings inequality. Individuals who remain active until retirement age respond more to the reform. Conditional on prereform earnings, we find little evidence that individual characteristics such as working histories influence the responsiveness to the reform.

I. Introduction

One of the measures undertaken by governments around the world to encourage their workforces to delay retirement is to reform or altogether abolish the old age pension earnings test. Among the first to undertake this reform was the United States, which exempted workers aged 70–71 in 1983 and exempted workers above full-retirement age from the Social Security earnings test in 2000.[1] The earnings test in the United Kingdom (the earnings rule), which implied a 100% withdrawal of pension against earnings in an interval corresponding to roughly between 21 and 35 hours week[2] at average earnings, was abolished in 1989. Similar policy reforms have been considered and implemented in many other countries.[3]

*This paper is part of the Frisch Centre project '1307 Strategic research programme on pensions' financed by the Ministry of Labour and Social Inclusion. Data from Statistics Norway have been essential for the project. We thank John Piggott, Oddbjørn Raaum, Knut Røed, Steinar Strøm and two anonymous referees for valuable comments to earlier versions. The usual disclaimer applies.
JEL Classification numbers: J14, H55.

[1]Legislated under the 'Senior Citizens Freedom to Work Act', see Haider and Loughran (2008) for details.

[2]In an interval corresponding to less than 20 weekly hours at average earnings, the test implies a 50% withdrawal, see Disney and Smith (2002) for details.

[3]In a sequence of reforms, Japan has amended its earnings test to be more liberal for those between 60 and 65, while the social security benefit schedule has been revised so that actuarially fair payoffs are received until age 70. In

Considering the potential impact of this particular reform on elderly labour market behaviour, and the theoretical ambiguity of the labour supply response to changes in the earnings test, this type of reform has received remarkably little attention in the literature until recently. To date, most studies have focused on the US reform. The results have varied, as some conclude that there has been little effect on the labour supply (see, e.g. Gruber and Orszag, 2003), although the general thrust of various analyses suggests that earnings test abolition leads to a quite significant increase in the elder labour supply (Friedberg, 2000; Friedberg and Webb, 2006; Engelhardt and Kumar, 2007; Song and Manchester, 2007). In a recent study of US reforms, Haider and Loughran (2008) utilize both survey and administrative data and conclude that the US earnings test in operation until 2000 did significantly reduce the earnings of persons over the full-retirement age, with the main effect felt for those at the lower end of the age bracket.

Similar evidence is also found in other countries. Disney and Smith (2002) confirm the impact of the earnings test in a study based on UK data for the abolishment of the earnings test in 1989. They estimate an increase of 3–4 hours week for men, with approximately half that for women after the reform. For Canada, Baker and Benjamin (1999) find shifts from part-time to full-time work, which they tentatively attribute to fixed costs of work or labour market rigidities. In the end, the conclusion that an earnings test depresses work activity still comes through.

In 2002, the exempt threshold in the Norwegian earnings test, above which 40% of earnings were deducted from the old age pension for persons age 67–69, was doubled. In contrast to the US test, the Norwegian earnings test can be unambiguously viewed as an implicit tax on earnings since there is no actuarial adjustment, i.e. future pension benefits will not be adjusted to ‘compensate’ the immediate reduction due to the earnings test. If individuals are forward-looking and behave in a way consistent with the life cycle framework, an actuarially fair earnings test such as that of the United States would have little or no effect on the labour supply. Hence, the Norwegian reform provides a better opportunity to study the responses of the elderly to changes in financial incentives.

In this paper, we evaluate the impact of the 2002 Norwegian retirement earnings test reform, using 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. In addition, the large number of observations makes it possible to obtain relatively precise estimates. We linked a number of administrative data sets on an individual level, thereby covering the entire Norwegian population. This largely eliminates the need for reconciling data sources as reported by Haider and Loughran (2008), who utilize a variety of data sources.

The Norwegian reform moves the lower kink in the budget curve, but otherwise does not alter the slope. The changes in the Marginal Tax Rate (MTR), as well as the magnitude of the income effect, vary across the earnings distribution. To capture the full picture of the impact of the reform, we allow for an uneven impact over different earnings levels and compare the earnings distribution both before and after the reform. We first look directly at the change of the Complementary Cumulative Distribution Function (CCDF)$\bar{F}(y) = \Pr(Y > y) = 1 - F(y)$ of labour earnings on the full sample by estimating

Australia, the recent abolition of all benefits taxation from age 60 is designed to increase the labour force participation of older workers.

a series of probit models. We find that the impact of the reform is substantial and varies considerably over the earnings distribution, with the strongest responses being found in the interval from the old threshold [50,000 Norwegian Kroner[4] (NOK[4])] to the new threshold (100,000 NOK). We observe a clear drop in the probability of having earnings around the old threshold after the reform, which is accompanied by an increase of a similar magnitude in the probability of having earnings between the prereform and postreform thresholds. These results suggest that individuals who were constrained by the old earnings test increased their earnings after the reform, though no significant change is observed in the right tail of the distribution. For the lower part of the distribution, the results suggest a small but significant increase in the probability of having zero earnings. However, this increase is partially offset by the decrease of the probability of having low earnings ($0 < y < 10,000$ NOK).

To study the effect on the intensive margin in more detail, our focus shifts to the sample of individuals with positive earnings. The results from a quantile regression model confirm the findings on the full sample in identifying a substantial earnings increase of approximately 10,000 NOK in 2002 and 16,000 NOK in 2003 for those with earnings near the old threshold (50,000 NOK). This corresponds to an earnings elasticity with respect to the net-of-tax rate of around 0.29 in 2002 (0.40 in 2003), thus suggesting a much stronger effect from the Norwegian reform than that of the United States, as reported by Song and Manchester (2007). Smaller and negative effects were found in the upper tail, in which the reform gives an income effect, but no incentive effect. Our data also enable us to model the impact of exposure time to the reform in detail, and we can confirm the results of Song and Manchester (2007) that there is a cumulative effect, probably because people need time to plan and adjust.

While following Song and Manchester in modelling the earnings response with a quantile regression, we extend their study in various directions. First, we propose and use a series of probit regressions to directly model the change on the distribution function. This method is easy to interpret and particularly suitable for our study, given our large sample size and the large share of individuals with zero earnings, which makes it difficult to apply a quantile regression method on the full sample.

Second, we look at the impact of the reform in more detail by studying various shape shift measures, which also enables us to shed some light on the impact on the earnings inequality. We find that the reform reduces the right-skewness of the earnings distribution and leads to a decrease in earnings inequality.

Third, our data contain a number of background variables that shows more aspects of the labour market response. Previous studies have primarily focused on the impact on men (Song and Manchester, 2007; Haider and Loughran, 2008), while we include women in our analysis and find that women respond less to the reform than men. We think this is mainly due to lower earnings among women since fewer of the women are around the threshold and constrained by the test. We are also able to examine in detail the impact of the previous labour market status, and our results suggest that the labour market status at age 66 is a strong 'predictor' for the response to the reform. Judging from the location shifts of the

[4]The average exchange rate in 2002 was approximately 7.6 Norwegian Kroner (NOK) per EURO. The average full-time earnings in 2002 were about 320,000 NOK.

earnings distributions, those who are still active in the labour market at age 66, such as the self-employed and wage-earners, react more strongly than those who are already out such as early retirees and disability pensioners. However, when looking at the individuals who are most likely to respond to the reform, i.e. those who were near the prereform threshold, we do not find any substantial differences in response. Therefore, the reason for the variation in response to the earnings test reform with individual characteristics such as gender, and previous labour market history is that they are good indicators of the relative position in the earnings distribution and of being constrained by the earnings test. The individual characteristics have little significant direct effect on the size of responses.

In examining the effect of the earnings test reform, this study extends the knowledge of labour supply responsiveness in older people, and our results suggest that the elderly can and will respond to changes in financial incentives. The significant response to a reform that may appear modest for the age group from 67 to 69 also signals a strong response among younger cohorts to more far-reaching reforms in the Norwegian pension system, which will come into effect from 2011.

This paper is organized as follows: in section II, we give a brief introduction of the Norwegian pension system, whereas in section III, we discuss the potential impact of the earnings test. Section IV presents the data and, section V gives both a brief overview of the elderly labour supply and a descriptive analysis of the impact of the 2002 reform. The difference-in-difference approach and estimation results are discussed in section VI, and section VII looks at the different responses from different subgroups. Section VIII concludes.

II. Institutional setting

The backbone of Norway's retirement provision system is a mandatory, public pay-as-you-go defined benefit plan, the National Insurance System (NIS), which is available from age 67.[5] The system is organized around a unit called the basic amount (G), which is usually adjusted once a year in accordance with the general income level. In 2002, one basic amount (1 G) was about 50,000 NOK.[6] Benefits consist of a basic pension (1 G) and an earnings-related pension. The earnings-related pension is based on the average of the highest 20 years of earnings, with benefits currently set at 42% of earnings between 1 and 6 G and 14% of earnings between 6 and 12 G.[7] The NIS is therefore very progressive and redistributive. Stable earnings at the maximum pension accruing level (12 G), which in 2008 was approximately twice the average of full-time earnings, imply a replacement rate of 33%. In contrast, the replacement level is 60% of stable earnings, which just gives the minimum pension.

In addition to the NIS, there are occupational pensions in both the public and private sectors. The public sector pension is fully integrated with the NIS so that the two combined give a pension that is 66% of the final salary at full accrual, which is 30 years.

[5] An overview of the Norwegian NIS and the proposed pension reform can be found in The Ministry of Labour and Social Inclusion (2006).

[6] In 2008, 1G was approximately 69,000 NOK.

[7] Earnings from before 1 July 1992 and therefore gradually phased out, give percentages of 45 between 1 G and 8 G and 15 between 8 G and 12 G.

In the private sector, the occupational pensions are firm based. In 2001, 85% of all recent retirees received an occupational pension, adding an average of about 30% to their NIS pension. Until 2001, only contributions to occupational pensions which were of the defined benefit type, and that also complied with the specifications in the legislation on these pensions, qualified as a tax deductible cost for the firms. As a consequence, almost all pensions were of this type during our observation period. Private sector occupational pensions of the DB type are like those in the public sector, which are designed to supplement the NIS pension and target a (total) replacement rate defined as the sum of pretax NIS and occupational pension divided by the final salary, with the predominant replacement target being 66% (Pedersen, 2000).

There are also early retirement programmes (AFP) available from age 62 that cover the public sector and about half of the private sector.

Up to 1 January 2008, when it was abolished for persons aged 67, there was an earnings test for persons aged 67–69. During our observation period, the test implied that the NIS pension was reduced by 40% of earnings exceeding a certain threshold, which was 1 G up until 1 January 2002 and 2 G thereafter. The test for occupational pensions is more complex. For workers from the public sector, only earnings from continued public sector employment after age 67, with a wage rate above a special 'retiree' wage rate that is about two-thirds of the average, leads to a reduction in pension income. Earnings from the private sector do not influence the NIS pension or the occupational pension of public sector retirees. Workers with occupational pensions from the private sector can choose whether to receive their occupational pension unabridged or defer it with an actuarial adjustment. Hence, we assume that the earnings test only applies to the NIS pension.

For firms, contributions are taxed like wages. For employees, pensions are taxed under an EET paradigm (contributions and fund earnings are both tax exempt and benefits are taxed under the income tax).[8]

At age 67, all can apply for an old age pension in the standard system (NIS). As they do so, they report expected earnings and the earnings test is applied. Because delaying pension receipt does not increase future benefits (as it would in the United States), it is never optimal to avoid the earnings test by refusing the pension.[9] Also in contrast to the United States, benefits lost to the earnings test in Norway do not increase future benefits. As a result, the Norwegian test can be unambiguously viewed as a tax, and we do not have to assume myopic behaviour as do, e.g. Friedberg and Webb (2006). There is also no concern in the Norwegian case about the standard of living at a very advanced age, which may occur if the removal of the earnings test induces an early take-up at a reduced benefits level of, which is an issue in the United States (Gruber and Orszag, 2003). As described above, the magnitude of the Earnings Test implied tax rate also depends on the impact on occupational pensions.

[8]Tax rates on pension income are lower than tax rates on earnings.

[9]Some may have an incentive to continue working to increase their entitlements, either because they had less than the full number of years of accrual (40 years in the NIS and often 30 in a public or private OP) or to improve the 20 best years in the NIS or the final salary in the OP. This effect is likely to be small and is also not different from that just before the age of 67.

III. The labour supply responses of the earnings test

Standard economic theory depicts the earnings test as a tax on earnings for all of those with earnings above the lower threshold where the earnings test starts to be effective (the exempt amount), though below the level where all NIS pensions are exhausted. Thus, the earnings test creates two kinks in the budget set. The lower kink will be at the exempt amount, whereas the upper kink corresponds to the amount in which all the old age benefits are tested away. Although the lower kink will be the same across the population, the upper kink will be individual since it is dependent on the pension level. For a person with benefits around 125,000 NOK (roughly the average level of old age benefits in 2002), the equivalent marginal tax on earnings before the 2002 reform is 35.8% below 50,000 NOK. For earnings above that amount, it increases to 61.48% until 362,500 NOK and then returns to 35.8%, as shown in Figure 1. During our observation period of 2000–03, there were no structural changes in the tax system, only inflation-based tax bracket adjustments.

The reform in 2002 moved the lower kink in the budget line as illustrated in Figure 1 and reduced the equivalent MTR on earnings between 1 and 2 G from 61.48% to 35.8%. The response to such a reform predicted by standard economic theory will vary across individuals and is dependent on the original position in the earnings distribution. If the elderly can freely adjust their labour supply conditional on wages, the response should be the net result of an increased return to work (incentive effect) and an increased net income (income effect). For those with earnings so high that all the benefits are taxed away, only the income effect applies and we may expect a reduction in the labour supply and earnings. This is the case for only a very few, as can be seen from the empirical overview below in section V. For those with earnings between the two kinks, the income effect and the substitution effect from the reform operate in opposite directions, and the net effect is an empirical question. Some of the individuals with earnings much lower than the lower kink may not be active in the labour market irrespective of the earnings test. For those individuals, a reform of the earnings test will not have any effect. However, as has been argued by many, labour market rigidities may prevent workers from freely adjusting their working hours, particularly for older workers. In this case, we may not be able to make as clear a prediction of the effect of the earnings test reform. On the one hand, labour market

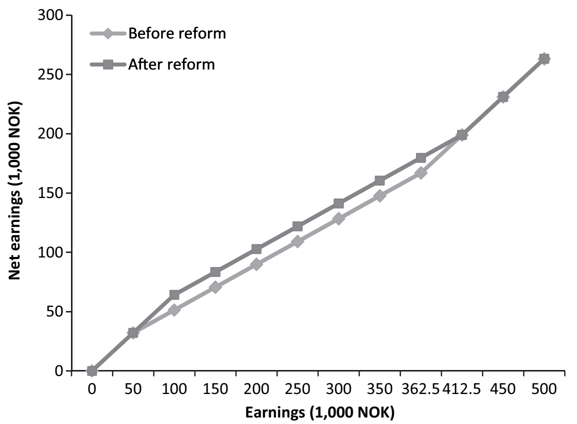

Figure 1. Earnings test and the effective tax rate before and after the 2002 reform

rigidity may prevent or limit the responses even if workers would like to respond, while on the other hand, persons previously below the lower kink due to market rigidities may be induced to seek work, and there may also be a response on the extensive margin in the presence of market rigidities. For example, jobs with a low amount of work hours may not be attractive or available due to the fixed costs facing workers and/or employers. If earnings test reform raises the number of hours that potential workers wish to work, these constraints may no longer bind and we may then also see an effect on the extensive margin (labour force participation).

Additionally, the earnings test reform may also have an effect on younger workers' retirement plans, and they may choose to adjust their labour supply accordingly. However, an evaluation of the long-run impact requires a dynamic life cycle model (see French, 2005) of both the labour supply and consumption behaviour, also taking into account that it may be more difficult to (re)enter the labour market than to stay in it. This is beyond the scope of this analysis, although we look at one aspect of labour market dynamics by conditioning on the labour force status at age 66.

IV. Data

We base our analysis on individual data from merged administrative registers received from Statistics Norway, with permission from the Norwegian Data Inspectorate. We use demographic data files, as well as old age pension registry and tax return records. These files are linked by a unique personal identification number[10] and cover the entire population of Norway.

Demographic files contain both birth and death dates, gender, education and other demographic variables. The old age pension register records the date of the old age pension benefit take-up, the accumulated pension rights and the actual pension paid. Tax return files record summary earnings from various sources, including pension income (the sum of both the NIS and OP pensions), wage income, income from agriculture and other pensions giving income. The sum of all pensions giving income, which essentially includes wages and self-employed earnings and excludes capital income and pensions, is referred to as earnings in the following. In this study, we primarily use data from 1999 to 2003. However, for earnings history, we have also used data reaching as far back as 1967.

As mentioned by Song and Manchester (2007), using administrative data provides both advantages and disadvantages over other data used for studying the effects of the earnings test. Administrative data are not plagued by the self-reporting problems common in survey-based records, which is particularly important in this context since 'accurate earnings data are crucial for analysing responses around the test threshold'. Even though we have data on education, detailed income by sources and information on family characteristics, our data lack some crucial information for labour supply models such as the number of hours of worked and health status.

We limit our sample to all individuals who are between the ages of 67 and 72 from 2000 to 2003, with the summary statistics presented in Table 1.

[10]This number is an encrypted version of the official personal identification number comma and is only used for the internal linking of files at the Frisch Centre.

TABLE 1

Summary statistics

	2000	2001	2002	2003	*Total*
Share of females	0.53	0.53	0.53	0.53	0.53
Years of schooling	9.30	9.42	9.55	9.87	9.53
Financial wealth (in 1,000s of NOK)	339	363	376	404	370
Average pension points (for the NIS)	3.05	3.08	3.11	3.15	3.10
Share of individuals active in labour market*	0.19	0.19	0.19	0.20	0.19
Average earnings (in 1,000s of NOK)	17.3	18.9	19.4	22.1	19.5
Number of observations	205,162	201,562	198,450	197,287	802,461

*Being active in the labour market is defined as having annual earnings over 0.2G (approximately 10,000 NOK), which corresponds to roughly working 1.5 hours week at an average wage rate.

V. Labour market behaviour of the elderly and the 2002 reform

We start our analysis by looking at the pattern of elderly labour market behaviour. The majority of people (about 96%) are registered as receiving old age pension as soon as they reach the eligibility age of 67, and individuals reduce their labour market activities sharply after 67. The average earnings share out of total income drops sharply with age (Figure 2), thereby implying that pension income is the main income source for individuals over 67, with the same pattern holding for both 2000 and 2002.

Figure 3 shows the age-specific transition probabilities from ‘not working’ at age t–1 to ‘working’ at age t for the 1931–04 cohorts from ages 67 through 71. Working is defined as being active in the labour market, i.e. having annual earnings over 0.2G (see footnote in Table 1). As we expected, the probability declines with age. It is quite small, ranging from 1.4% to 2.5%, demonstrating that very few in this age group actually re-enter the labour market. These numbers are very close to what Song and Manchester (2007) found on the US data. In addition, there is no indication of a pattern shift across the various cohorts.

Figure 4 shows the distribution of positive earnings over different ages for the 1932 cohort. Earnings in the year of retirement after retirement, and thus subject to the test, are

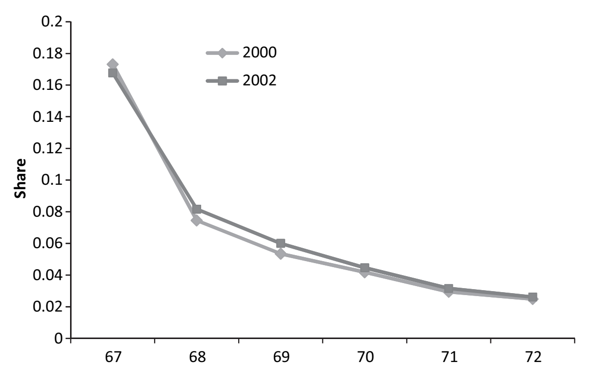

Figure 2. Share of earnings in total income by age and year

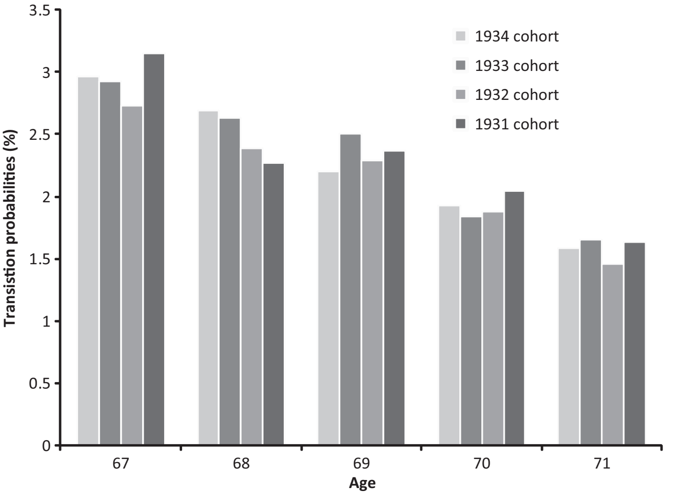

Figure 3. Transition probabilities from not working to working, by age and cohort

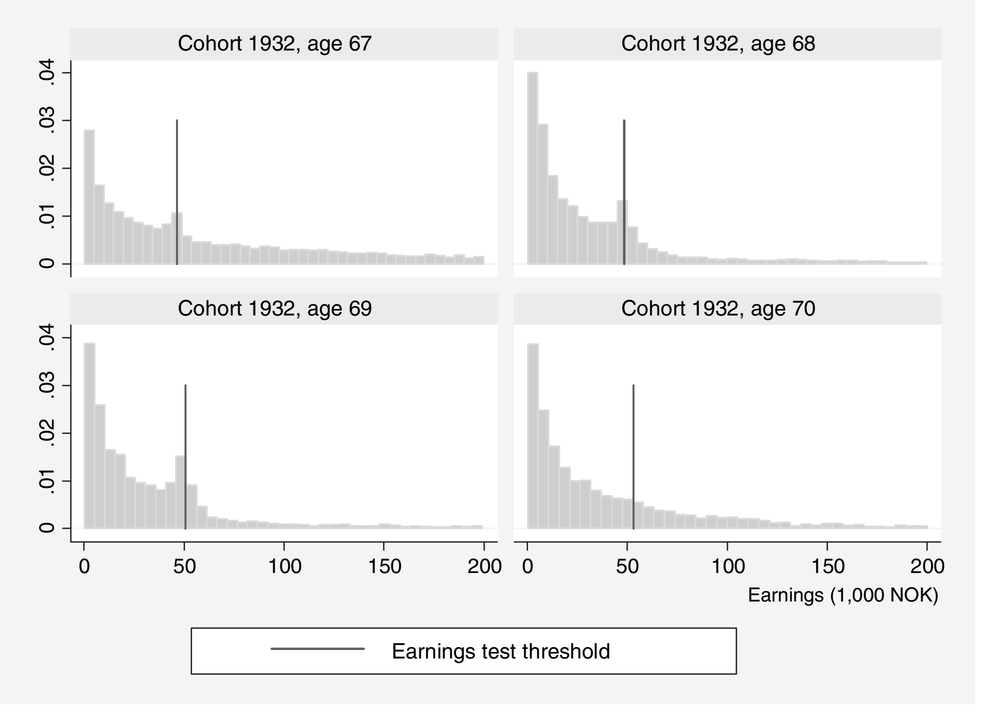

Figure 4. Distribution of earnings of 1932 cohort with positive earnings, by age

imputed using information from previous earnings and retirement take-up date. Around the earnings test threshold, there are peaks which disappear when individuals reach the age of 70. This peak can be explained by the kinks in the budget constraints created by the earnings test. Individuals react to the earnings test by keeping their labour supply down when the earnings test is present. Without the test, we would expect a smooth decline. Similar patterns are found in both the United States (Friedberg, 2000; Song and Manchester, 2007; Haider and Loughran, 2008) and the United Kingdom (Disney and Smith, 2002). This is in clear contrast to findings reported by Saez (2010), where he found no bunching for wage-earners at the kink points generated by the Earned Income Tax Credit in the United States. Saez (2010) proposes a number of reasons for not observing bunching among wage-earners, among them that they may not have the option of adjusting work hours or do not understand the rules. In Norway, earnings test rules are explicitly informed to all the individuals when they apply for the retirement benefits. In addition, the Earnings Test reform and the earned income tax credit are aimed at different types of workers. The labour market environment faced by an experienced old worker and a low income (most likely also low skilled) worker would be certainly quite different. Therefore, we do not find it surprising that we do so clearly observe behavioural responses. However, as discussed in section III, a bunching at the earnings test threshold does not necessarily imply that a retirement earnings test reform, whether it raises the threshold or completely abolishes it, will increase the aggregate labour supply since there is both an income and an incentive effect. We expect that such a reform will increase the labour supply of individuals with earnings around the threshold (the substitution effect dominates). For those with earnings above the threshold, the income effect becomes more important with higher earnings until only the income effect remains for those with earnings so high that all the benefits are taxed away. For this group, the labour supply may be reduced. Even so, there will be very few individuals in this group. For an average pension to be taxed away, the postretirement earnings would have to be approximately 362,500 NOK before the reform and 412,500 NOK after the reform. This is far out in the tail of the retirement earnings distribution, as shown in Figure 4.

The increase in the earnings test threshold from 1 to 2 G from 1 January 2002 gives us a chance to look into how the earnings test affects the elderly labour supply and earnings. Figure 5 shows the distribution of positive earnings for 68 and 69 year olds, both before and after the 2002 reform. The change of the earnings rule threshold had a clear impact on earnings for both age groups. As expected, the clusters around 1G disappear after the reform (in 2002) and are replaced by new clusters at the new threshold 2G. The new peaks are less prominent than the old ones since there are fewer individuals in this part of the earnings distribution and are therefore affected by the earnings test rule after the 2002 reform. The observation of peaks that shift with earnings test thresholds suggests that older workers indeed have some degree of freedom in the labour supply decisions and can respond to changes in incentives.

VI. Regression analysis

Identification strategy

The descriptive empirical evidence indicates that the reform that doubled the exempt amount from 2002 encouraged some individuals to increase their earnings. To evaluate this effect econometrically, we compare the changes in earnings both before and after the reform of two groups of individuals. Group 1 (the treatment group) consists of those who were affected by the earnings rule change, namely those aged 67–69, whereas group 2 (the control group) consists of those who faced no earnings test during the study period,

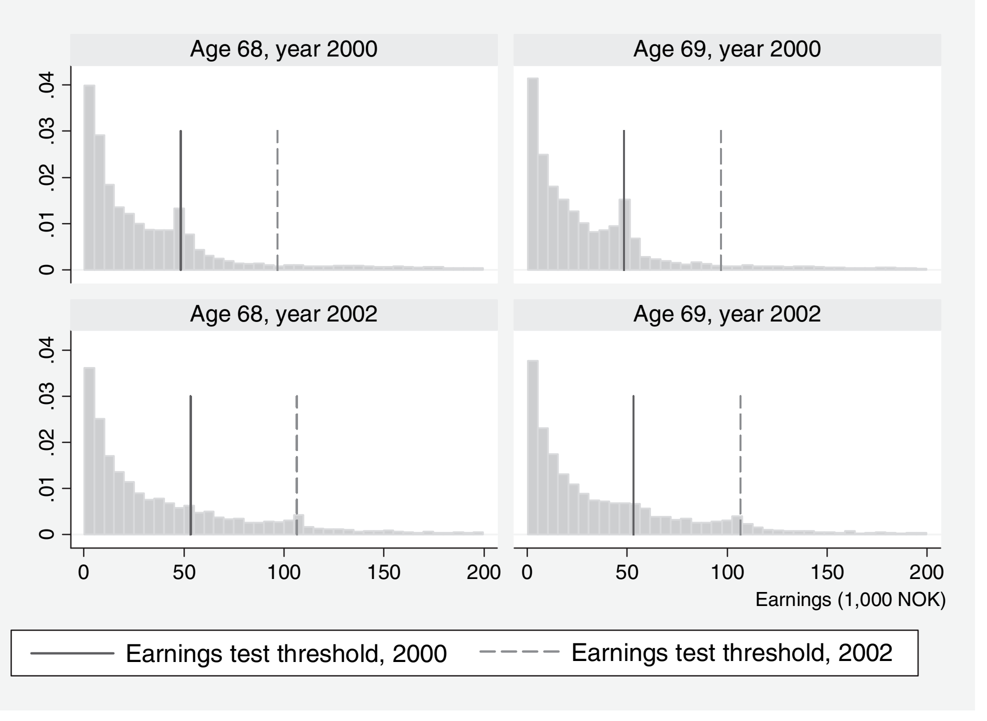

Figure 5. Distribution of earnings before and after 2002 reform, those with positive earnings

namely those aged 70[11] to 72. A similar identification strategy was used by Song and Manchester (2007) and Disney and Smith (2002). The purpose of the control group is to account for important unobserved factors that influence the labour supply behaviour of all such as uncontrolled macro-economic conditions that have a direct impact on labour demand, and changed social norms, e.g. a generally increased acceptance of postretirement employment.

Tables 2 and 3 show the average earnings and participation rate for both the control and treatment groups over the period from 2000 to 2003. As expected, we observe sizable increases in earnings for the treatment group after the reform that are significantly larger than the differences between 2 years before the reform. The differences for the control group are much smaller than those of the treatment group, and they are of a similar magnitude as the differences between the 2 years before reform. For the participation rate, we also observe an increase after the reform for both the control and treatment groups. However, the difference between the control and treatment is not as obvious as for the earnings. Both tables also suggest that there is a sizable variation in the labour supply behaviour of the elderly over different years, which shows the importance of accounting for the yearly effects.

Our identification hinges on the important assumption that there are no shocks other than the retirement earnings test reform that affect the labour supply behaviour of the treatment group differently from the control group. If the model is correctly specified and the

[11] We assume earnings have risen after their 70th birthday, and no test applies, and alternatively act as if this was the case, perhaps to preserve options for further work.

TABLE 2

Average earnings over years and ages (NOK)

	67	68	69	70	71	72
	Treatment group			*Control group*		
2000	*44,019*	*19,983*	*14,153*	11,141	8,535	6,502
2001	*47,253*	*22,825*	*15,847*	12,289	9,059	7,169
2002	**46,978**	**24,892**	**17,602**	12,779	8,473	7,330
2003	**53,415**	**25,494**	**20,139**	(14,422)	9,792	7,246
Difference from 2000 to 2001	3,235	2,842	1,694	1,149	524	667
Difference from 2000 to 2003	9,396	5,511	5,986	3,282	1,257	743

Note: Earnings under an earning test with a threshold of 1 G are marked in italic, and earnings under a 2 G threshold test are marked in bold. Seventy year olds in 2003 are put in parentheses since they were dropped from the analysis (see text).

TABLE 3

Participation rates over years and ages

	67	68	69	70	71	72
	Treatment group (%)			*Control group (%)*		
2000	*34.8*	*23.6*	*17.8*	14.9	11.5	9.90
2001	*35.0*	*24.6*	*18.3*	15.5	12.6	9.68
2002	**34.0**	**25.1**	**19.3**	15.6	12.6	10.4
2003	**36.6**	**24.9**	**19.6**	(16.3)	12.8	10.6
Difference from 2000 to 2001	0.2	1.0	0.5	0.6	1.1	−0.0
Difference from 2000 to 2003	1.8	1.3	1.8	1.4	1.3	0.7

Note: The participation rate is defined as the share of individuals who are active in the labour market, i.e. with earnings > 0.2 G. Participation rates under an earning test with a threshold of 1 G are marked in italics and participation rates under a 2 G threshold test are marked in bold. Seventy year olds in 2003 are put in parentheses since they were dropped from the analysis (see text).

control groups are appropriate, we expect that the estimates for the coefficients associated with the treatment terms prior to the 2002 reform should not be significantly different from zero, which can serve as a simple specification test for our difference-in-difference approach.

There is a potential problem using those born after 1932 as our control group since they were in the treatment age when the reform was introduced in 2002. As Disney and Smith (2002) suggest that, if there are employment dynamics such that labour supply decisions taken at ages 67–69 affect labour supply decisions at ages 70–72, the control group in the difference-in-difference approach will not be appropriate. Given this concern, the sample we used in the regression are the individuals aged 67–72 with positive earnings from 2000 to 2003, except that 70 year olds in 2003 are dropped since they were 69 years old in 2002 when the reform was introduced and are therefore affected.

Modelling the impact of the reform: full earnings distribution

Remember that economic theory suggests that the earnings test reform should have a different impact on different parts of the earnings distribution. The reform not only shifts the location but also changes the shape of the earnings distribution. Simply looking at the location shift by looking at the conditional mean change using a traditional OLS is inappropriate, and at the least, insufficient. To acquire a complete picture, we need to model the distribution shift in more detail.

Quantile regression models have been extensively used in the literature to assess distribution shifts. However, the distribution of earnings is a mixed discrete and continuous distribution with large mass at zero. Given the large sample size and high share of zero values, applying a quantile regression on the full sample turns out to be highly demanding from a numerical standpoint. For this reason, we first apply a less popular but closely related method on the full sample to evaluate the overall effect of the reform. We then use a quantile regression method on the truncated sample with positive earnings to look in more detail at the impact of the reform on the intensive margin.

In the first approach, we look directly at the change of the Complementary Cumulative Distribution Functions (CCDF) $\bar{F}(y\,|\,X)=\Pr(Y>y|X)$before and after the reform. As illustrated in Figure 6, both this method and the quantile regression method can be used to capture the full picture of a distribution change, although from different angles. To describe the distribution shift from distribution F_1 (represented by its CCDF) to distribution F_2 at point A, we can either use the horizontal distance between F_2 and F_1or the vertical distance. The first measure corresponds to a change in earnings (y_2-y_1) at a given quantile p_1, whereas the latter corresponds to a change in probability $F_2^*-F_1^*$ at a given earnings y_1. It is easy to see that the signs of these two measures are the same. In addition, if two distribution functions intersect at A, both measures will be zero. Thus, these two methods will yield the same qualitative conclusion. In theory, we can approximate any distributional statistics of interest once we have estimated the CDF using the CCDF methods or quantiles, which is the inverse of the CDF, using the quantile regression method. In this sense, these two methods are equivalent. Therefore, in an empirical analysis, the choice between these methods is often a matter of convenience. The CCDF method models direct the CDF functions and can easily be used to construct estimates of changes on probability density function. In contrast, the quantile regression is closely related to and has an interpretation similar to the standard OLS, and it is straightforwardly used to generate some shape shift measures using quantile regression methods as shown later in this section.

For any given value y on its support, the CCDF for a random variable Y, $\bar{F}(y\,|\,X)=\Pr(Y>y\,|\,X)$ is a probability, and can be modelled using a parametric binary responses

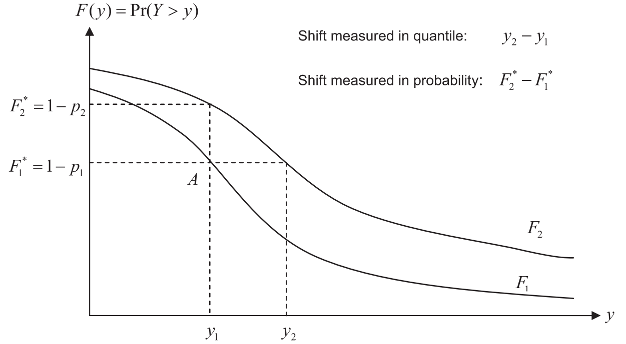

Figure 6. Measurement of distribution shift: two methods

model. To model the entire CCDF, we need to select a series of points which cover the support of Y and estimate one binary response model at each point. We impose no structural connections over these different binary response models, so this method essentially constructs a semi-parametric estimator of the CCDF. This is of particular advantage in our case, in which the distribution of earnings is of a very special shape.

In this paper, we use a series of probit specifications to model the conditional complementary CDF for a number of values of y, which allow a simple application of the difference in a different technique to identify the treatment effect of the reform.

For any given value of $y \geq 0$, we assume that for the individual i at year $t = 2000, \ldots, 2003$:

$$P(y_{it} > y) = \Phi(\alpha + X_{it}\beta + \sum_t \gamma_t D_t + \sum_t \lambda_t D_t \cdot \Delta_{it}), \tag{1}$$

where the dependent variable y_{it} is non-pension earnings, X_{it} denotes individual characteristics, including age and gender dummies, average pension points and financial wealth and years of education. D_t is the time dummy. Δ_{it} equals 1 if individual i belongs to the treatment group at year t. The interaction terms $D_t \cdot \Delta_{it}$ are included to capture the different changes of the treatment group during different years.

To test whether our identification assumption is valid, as suggested in the last section, we first included the full set of interaction terms between treatment group and year dummies ($D_{2000}\Delta_{i,2000}$ is omitted to avoid perfect multicollinearity). If our identification assumption is valid, the 'placebo' treatment dummies λ_{2001} should not be significantly different from 0. We have estimated 41 probit models in which y ranges from 0 to 200,000 NOK in increments of 5,000 NOK. None of the 'placebo treatment' parameter estimates is statistically significant. We then dropped the $D_{2001}\Delta_{i,2001}$ from the specification to obtain a clean specification and re-estimated the model to study the treatment effects for 2002 and 2003. Since it is rather difficult to interpret the parameters of probit models, we present estimates of the marginal effects instead. At a given earnings level$y$, the marginal effect is defined as the difference between the postreform and prereform probability of earning income more than y:$P(Y > y \mid X_{it}, \Delta = 1) - P(Y > y \mid X_{it}, \Delta = 0)$, which is the exact vertical distance between the prereform and postreform CCDFs at y as shown in Figure 6.

Figure 7 shows the estimated marginal effects with 95% confidence envelops over different earnings levels when individual characteristics are set at the sample means. As we expected, the effects of the reform vary considerably over earnings level y, which highlights the importance of studying the distributional effect. The treatment effect follows an inverse U pattern and is concentrated on the interval between the old and the new threshold (50,000 NOK to 100,000 NOK). To see the pattern more clearly, we define four earnings brackets: low earnings $[0, 40, 000)$, at approximately the old threshold [40,000, 55,000), medium earnings [55,000, 105,000) and above the new threshold $[105, 000, \infty)$. Table 4 presents the predicted prereform probabilities of being in those four different earnings bracket, as well as their changes after reform. In the first and last earnings brackets, the changes are small and statistically insignificant. However, we do observe a significant decrease in the probability of having earnings at about the old threshold and an increase of similar magnitude for the middle earnings bracket. This implies that individuals who are constrained by the old test, i.e. those who are holding earnings at about the old threshold level, increase their earn-

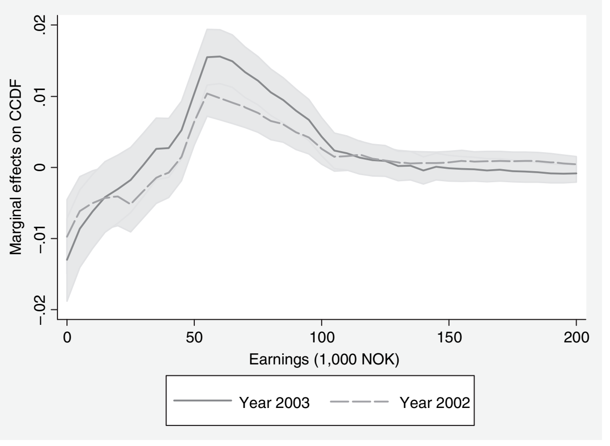

Figure 7. Estimates of marginal effects of the reform: series of probit models

TABLE 4

Probabilities and their changes for selected earnings intervals

		Probability change after reform	
Earnings brackets	*Before reform (%)*	2002 *(%)*	2003 *(%)*
< 40, 000 NOK	90.2	0.07	−0.27
		(0.1)	(0.2)
40,000–55,000 NOK	3.03	−1.10	−1.28
		(0.2)	(0.3)
55,000–105,000 NOK	3.72	0.88	1.31
		(0.2)	(0.3)
> 105, 000 NOK	3.05	0.15	0.23
		(0.1)	(0.2)

Note: Standard deviation in parentheses. Probability change is generated using the marginal effects reported in Figure 7, in which individual characteristics are set to sample averages.

ings after the reform. But once again, the increase is limited by the new threshold, which is consistent with both economic theory and the graphic patterns described in Figure 5.

To our surprise, the results in Figure 7 also suggest the reform may have a negative effect at the lower end of the earnings distribution. We have not yet found a good explanation for this empirical pattern since economic theory predicts that the reform has either no or only some positive participation effects depending on whether there exists market rigidities. This is also in clear contrast to what is found in the literature, in which the typical findings are that the earnings test reform has no, or at best, a weak positive effect on the extensive margin. Further research is needed before we make any conclusions as to whether this is an empirical regularity or only an artefact due to misspecifications or noise in the data. However, the effect on the probability of having zero earnings is partially offset by the decrease in the number of persons with very low earnings, which is also imprecisely estimated.

Despite the fact that the effects of the reform vary strongly across the earnings level, a standard regression on the full sample can still be seen as a simple and intuitive way to simultaneously summarize the intensive and extensive margin effects.

We then run the following regression on the full sample:

TABLE 5

Estimates of the mean treatment effects, NOK

	Estimates	*Standard error*	*P value*
Treatment effect, 2001	1,271	570	0.028
Treatment effect, 2002	1,996	474	0.000
Treatment effect, 2003	3,161	536	0.000

Note: The dependent variable is earnings. Other covariates included in the regressions are a constant, a female dummy, age and year dummies, years of education, average pension points and financial wealth. The standard error is generated using bootstrapping.

$$y_{it} = \alpha + X_{it}\beta + \sum_t \gamma_t D_t + \sum_t \lambda_t D_t \cdot \Delta_{it} + \varepsilon_{it}, \tag{2}$$

where the dependent variable y_{it} is the earnings level, and X_{it}, D_t and Δ_{it} are the same as defined in equation (1), with the results presented in Table 5. Given the fact that many have zero earnings, the error terms are not normally distributed. To be able to perform the standard statistics test, we generated the standard errors of the estimates by bootstrapping.

From Table 5, we see that after the reform the mean of non-pension earnings in the population increased by approximately 2,000 NOK and 3,000 NOK in 2002 and 2003, respectively. The positive intensive margin effect dominates the small negative extensive margin effect, and the total effect of the reform is positive.

One thing we need to note is that the 'placebo' treatment effect in 2001 is positive, with a P value of 0.028. This may raise some questions in terms of the validity of our control groups. Nevertheless, our results from both the probit estimations reported above and the quantile regressions in the next section suggest that the control group is well defined. In addition, the 'placebo' effect is not strongly significant, unlike most of the probit and quantile results, which are significant at a 1% level or even lower. Also, the simple linear regression is obviously misspecified, and only serves as an illustration of the mean treatment effects on the full sample.

Effect on the intensive margin: quantile regression on individuals with positive earnings

The above results suggest a strong and uneven effect on the intensive margin of labour supply behaviour from the 2002 reform, particularly around the old threshold. In the following, we limit our sample to those with positive earnings and focus in more detail on the intensive margin effect. For example, we already know from the results above that some individuals with earnings at about the original threshold increased their earnings. We would like to have a quantitative estimate on the size of the increase, and we would also like to construct distributional statistics/indexes that enable us to summarize and explain the distribution shift.

For this purpose, we have simultaneously estimated the model on 19 equally spaced percentiles (0.05th, ..., 0.95th). For individual i at percentile p in year $t = 2000, \ldots, 2003$:

$$y_{it} = \alpha^p + X_{it}\beta^p + \sum_t \gamma_t^p D_t + \sum_t \lambda_t^p D_t \cdot \Delta_{it} + e_{it}^p, \tag{3}$$

TABLE 6

Estimates of the quantile treatment effects, NOK

	25 *percentile*		*Median*		75 *percentile*	
	Est.	SE	Est.	SE	Est.	SE
Treatment effect, 2001	139	261	775	605	1,492	1,137
Treatment effect, 2002	674	258	3,686	774	10,665	998
Treatment effect, 2003	869	261	6,513	611	16,959	1,011

Note: The dependent variable is earnings. Other covariates included in the quantile regressions are a constant, a female dummy, age and year dummies, years of education, average pension points and financial wealth.

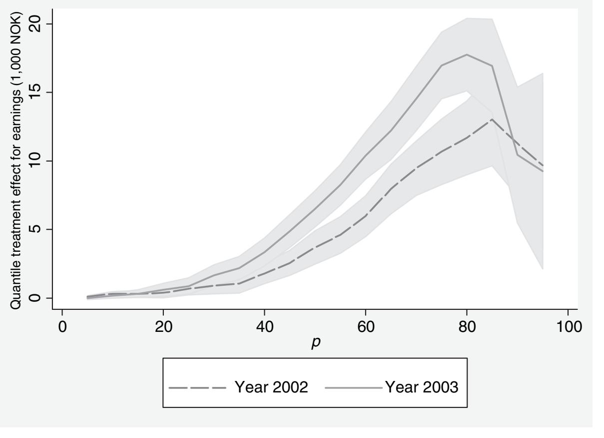

Figure 8. Estimates of quantile treatment effects of the reform

where the dependent variable y_{it} is the earnings level, and X_{it}, D_t and Δ_{it} are the same as defined in equation (1).

Similarly to the probit models, we again estimate two versions of the regressions: a 'full' version with the interaction term $D_{2001}\Delta_{i,2001}$ and a 'clean' version in which this term is dropped. Table 6 reports three selected percentile regression results on earnings, including the 25 percentile, the median (50 percentile) and the 75 percentile for the 'full' version. In Figure 8, we report results from the 'clean' version and plot all the estimated percentiles (in 5% increments) treatment effects, with 95% confidence envelopes to highlight the pattern over the percentiles.

From Table 6, we see that the placebo treatment dummy (λ_{2001}) is not significantly different from zero for any of the 25, 50 or 75 percentiles, while the 'real' treatment dummies both λ_{2002} and λ_{2003} are significant at the 1% level for all three percentiles. After the reform, the median of non-pension earnings in the sample increased by 3,686 NOK and 6,513 NOK in 2002 and 2003, respectively. Figure 8 shows that the treatment dummies for 2002 and 2003 are significant and increasing in magnitude from the 40th conditional percentile upwards, and concentrated on the 60th to 85th percentiles. The 2002 reform raised the earnings threshold from 50,000 NOK, which is slightly lower than the 75th percentile, to 100,000 NOK. The 75th percentile for a typical man (with the individual characteristics value set to the sample average) is approximately 54,000 NOK. At the 75th percentile, earnings increased by 10,665 NOK in 2002 and 16,959 NOK in 2003. The location of the effect therefore corresponds well with what Song and Manchester (2007) found on the US data. They argue that this result indicates the policy change 'has affected the earnings distribution just below the threshold and upwards, as predicted by economic theory'.

As for the magnitude of the effect, remember that for those who are constrained by the earnings test (earnings near the prereform threshold), the equivalent MTR was reduced from 61.5% to 35.8%. Our model predicts about a 20% increase in earnings in 2002 (about 28% in 2003). This corresponds to an earnings elasticity with respect to a net-of-tax rate[12] of approximately 0.29 in 2002 (0.40 in 2003), which is much higher than the elasticity reported in Song and Manchester (2007) for the US data. Their estimates range from 0.09 to 0.24 for those near the earnings test threshold, which we believe is because there is no deferral in the Norwegian earnings test. As we have previously discussed, actuarial adjustments such as those in the United States will dampen the effect of the earnings test for forward-looking individuals, thus reducing the response of a reduction/abolishment of the test.

Briefly, our results suggest that our specification captures a positive effect for the retirement earnings test reform and confirms our findings from the descriptive analysis and the CCDF methods on the full sample above. Old workers can and will respond to economic incentives.

Figure 8 indicates a rather complicated shape shift for the earnings distribution. The location shift can be represented by a change in the median for the conditional income distribution both before and after the reform, i.e. the median treatment effects that have already been discussed above.

To provide a more detailed description of the treatment effect, we chart a graph for the conditional distribution of earnings for a typical man before and after the reform (see box plots in Figure 9). Figure 9 suggests that the reform shifted not only its location but also its spread and skewness of the distributions. To be more precise, we generated the so-called shape-shift quantities (Hao and Naiman, 2007). Since the earnings distributions are not symmetric, the interquantile range (IQR) is a better measure of the spread of a distribution than the commonly used standard deviation. We define the pth IQR as

$$\mathrm{IQR}^p = Q^{1-p} - Q^p \qquad \text{for } 0 < p < 0.5. \tag{4}$$

This measure has the additional advantage that it can easily be generated from the quantile regression results. The scale shift is then defined as the differences between the IQR before and after reform:

$$\mathrm{SCS}^q = \mathrm{IQR}^q_{\text{after reform}} - \mathrm{IQR}^q_{\text{before reform}}. \tag{5}$$

A positive SCS indicates that there is an increase of the spread after the reform and an increase of the earnings gap between the rich and the poor, whereas a negative value indicates the opposite.

[12] Following Saez, Slemrod and Giertz (2010), we define the elasticity as the percentage change in earnings divided by the percentage change in the net of the tax rate.

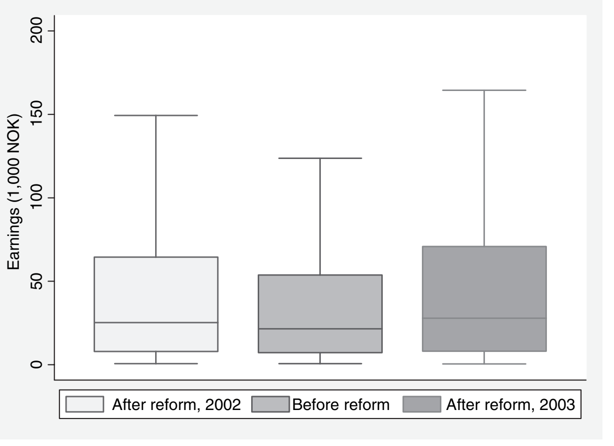

Figure 9. Box plot of the earnings distribution before and after the reform

TABLE 7

Scale and skewness shifts in the earnings distribution from the 2002 reform

	Estimate	*95% confidence interval*	
Scale shifts ($SCS^{0.025}$)			
Year 2002	5,952	−3,632	15,772
Year 2003	6,439	−4,559	16,757
Skewness shifts ($SKS^{0.025}$)			
Year 2002	−0.1476	−0.2020	−0.0848
Year 2003	−0.2444	−0.2956	−0.1864

Similarly, we use a measure of the skewness, which is also quantile based, and is defined as the ratio of the upper spread to the lower spread:

$$\mathrm{QSK}^p = (Q^{1-p} - Q^{0.5})/(Q^{0.5} - Q^p) \qquad \text{for } 0 < p < 0.5 \tag{6}$$

The skewness shift is defined as the ratio of QSKs before and after the reform minus 1.

$$\mathrm{SKS}^q = \mathrm{QSK}^q_{\text{after reform}}/\mathrm{QSK}^q_{\text{before reform}} - 1. \tag{7}$$

A positive value suggests an increase of the right skewness. Table 7 reports the estimated distribution shift parameters over the range of 2.5–97.5% ($p = 0.025$). The confidence interval is derived through bootstrapping with 500 replications, since these shift parameters are based on a combination of several parameter estimates as shown above. As we can see from the top panel of Table 7, there is no indication for both years of a significant change in the scale of the conditional earnings distribution. In other words, the reform does not change the spread of the conditional earnings distribution, it only shifts the distribution upwards.

Still, we do observe significant right skewness shifts for both 2002 and 2003. It seems that the reform has led to a decrease of the right skewness of the conditional earnings distribution. Since the reform induces an almost zero scale shift, a negative right skewness shift can be interpreted as a decrease of income inequality (Hao and Naiman, 2010).

TABLE 8

Estimates of the median treatment effects for different cohorts

	Estimates	*SE*
Treatment dummy, 1933 cohort	1,640	800
Treatment dummy, 1934 cohort	2,720	580
Treatment dummy, 1935 cohort	5,110	540
Treatment dummy, 1936 cohort	12,050	670
N	214,817	
Pseudo R^2	0.043	

Note: The dependent variable is earnings. Other covariates included in the median regressions are a constant, a female dummy, age and year dummies, years of education, average pension points and financial wealth.

Median earnings have increased significantly, whereas earnings at the 2.5 and 97.5 percentiles have not. This fits the theoretical predictions well, since the doubling of the threshold affected medium earners with both an income and a substitution effect. Low earners were not affected, while there was only an income effect among high earners. In other words, the reform not only succeeds in increasing the elderly labour supply (through a positive location shift), but also contributes to reducing the income inequality for the elderly by increasing the earnings of 'middle' individuals.

Another important observation from Table 7 and Figure 8 is that the effects on both the location and shape are stronger for 2003 than for 2002. One possible explanation for this is that different cohorts have a different exposure to the earnings test reform. Individuals born in 1933 were 69 years old in 2002 when the reform was implemented. The window of exposure for them is just 1 year, while for 1935 cohort it is 3 years. Also, the potential benefits of the reform are larger for younger cohorts who are exposed to the system when it is completely phased in and have the higher threshold from the time of retirement age. Additionally, if individuals are forward-looking, they may start planning years before they reach retirement age and make the necessary adjustments to take advantage of the reform. This will also serve to increase the response of those who have more exposure to the reform, namely the younger cohorts.

To investigate this econometrically, we construct cohort-specific treatment dummies. Note that here we only use the year of birth to define the exposure measures to the reform. We could have also included information on the month of birth to construct a finer measure, but that would have increased the number of the treatment dummies considerably, which is not practical.[13] The results are reported in Table 8. As expected, there is a clear trend that the responses increase with the exposure. While the reform increased annual median earnings by 1,640 NOK in the 1933 cohort, the increase was 12,050 in the 1936 cohort. When the reform was enacted, the 1936 cohort was 66, with many still in their preretirement job. They may therefore have had better opportunities to plan for continued work, maybe even at their 'old' workplace.

[13]Conversely, we could restrict the effect to be linear in time and use only one variable that represents the measure of exposure. However, we feel that this restriction is too strong.

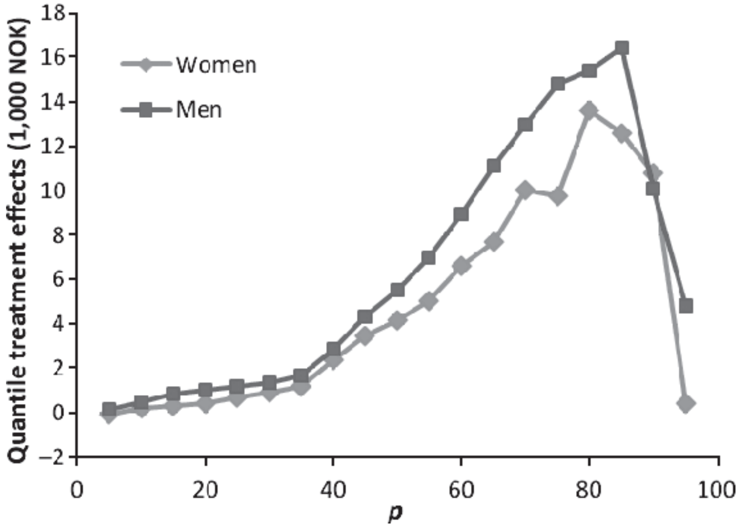

Figure 10. Estimates of quantile treatment effect, by gender

VII. Heterogeneity in the responses to the reform

The quantile regression results presented above clearly show the uneven impact of the reform. In the following, we will look more in detail to see how the effects vary over gender and previous labour market activities.

Gender differences

There is empirical evidence that men and women have shown different retirement behaviour; see, e.g. Dahl, Nilsen and Vaage (2003) and Jia (2005). To investigate the role of gender in the response to the reform, we have run quantile regression models separately for women and men.

The results suggest that women have a lower response than men (Figure 10), which is consistent with the findings in Jia (2005): elderly women typically retire younger, work less and value their leisure more than elderly men. In the sample, women have lower earnings than men, with Figure 11 showing the predicted distribution of earnings for typical men and women,[14] both before and after the reform. We see that the prereform threshold (roughly 50,000 NOK) corresponds to the 75th percentile for women and to around the 65th percentile for men. The corresponding effects of the reform estimated by the quantile regressions are then approximately 9,500 NOK for women and 10,600 NOK for men, respectively, for those who's earning were at the threshold before the reform. The response from women is somewhat lower than that from men, although the difference is not large when controlled for prereform earnings.

Consequently, the reform seems to not be able to contribute to closing the gaps between men and women. Even so, it remains to be seen as to how the effect will differ for younger cohorts with smaller gender differences in labour supply behaviour and earnings history.

[14]A typical man or woman is defined as an individual with the gender-specific sample average characteristics.

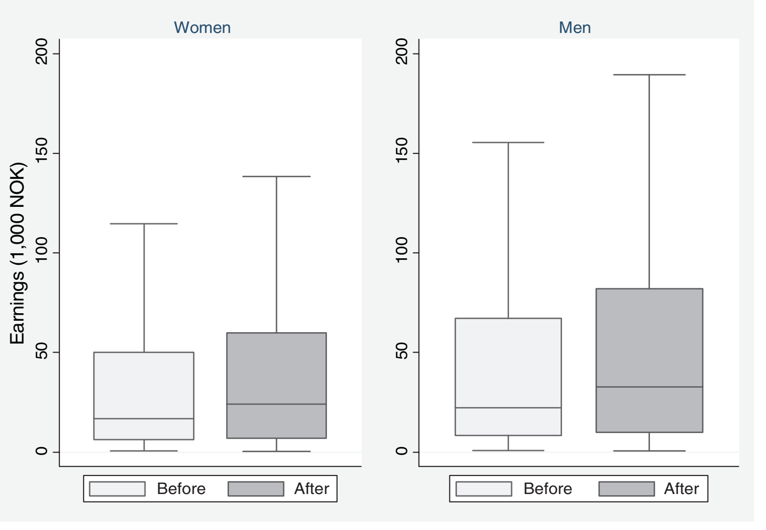

Figure 11. Distribution of earnings before and after reform, by gender

Labour market status at age 66

We mentioned earlier that there is evidence in the literature that labour supply behaviour is positively correlated over time. Friedberg and Webb (2006) also emphasized the importance of controlling for previous work experience when studying the effects of the earnings test. The reason for this is that the preretirement labour market status can provide useful information on important unobserved individual characteristics essential for the postretirement employment decisions such as constraints in the choice set, health status of the individual and unexplained preference differences. It is therefore of interest to divide the sample into subgroups according to their labour market status at 66.

We distinguish between four different groups: self-employed, wage-earners, early retirees through the AFP programme and individuals on disability pension. The model is then estimated separately for the four different groups. As we can see from the upper panel of Table 9, individuals in these groups differ substantially in relation to their education, financial wealth and past experience (represented by the average pension points).

The estimated median treatment effect of the reform (reported in the middle panel in Table 9) indicates that the labour market status at age 66 has a strong effect on the responsiveness to the reform, as measured by the location shift of the conditional earnings distribution before and after the reform. Wage-earners are the most responsive group, followed closely by the self-employed. The responses from the AFP retirees and the disabled are much smaller than from those of the wage-earners and self-employed. The overall effect of the reform for different groups is summarized in Figure 12, in which we present the before and after earnings distribution for a typical member of each group.

Similarly to the discussion on the effect of gender on the impact of the reform, we also look at the response at the threshold level before the reform for the typical members of each group. The distributions of prereform earnings differ substantially across the different groups, and it is of interest to see whether these differences can explain the differences in

TABLE 9

Summary statistics and treatment effects: subgroups on labour market status at 66

	AFP retirees	*Disabled*	*Self-employed*	*Wage-earners*
Years of schooling	10.6	8.9	9.8	10.6
Financial wealth (in 1,000s of NOK)	331	221	780	604
Average pension points	4.04	2.46	3.33	3.77
Share of female	0.45	0.54	0.22	0.50
Median treatment effect (NOK)	2,735	2,132	10,177	13,249
SE	630	563	2,471	1,130
Quantile corresponding to the prereform threshold	80	85	50	50
Estimated effect at this quantile (NOK)	13,861	12,715	10,177	13,249
SE	1,908	1,488	2,471	1,130

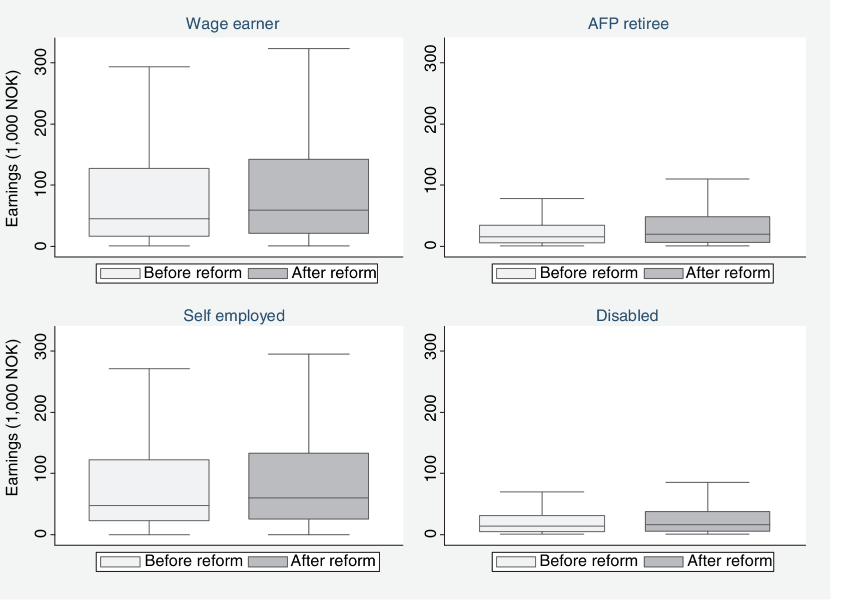

Figure 12. Earnings distribution by work status at 66

the response, with the results summarized in the lower panel of Table 9. Before the earnings test reform, more than half of the wage-earners and self-employed maintain earnings above 50,000 NOK (the prereform threshold), while only around 20% of disabled or AFP early retirees do so (see also Figure 12). However, the impact at around the threshold does not seem to vary much. The apparent effect of the labour market status the year before retirement eligibility on the response to the earnings test reform is caused by its role as an important predictor of potential postretirement earnings. Once the earnings distribution is taken into account, the response does not differ much.

Summing up on heterogeneity and the effect of the earnings test

It is striking that the effect of covariates, both gender and the labour market status at age 66, is sharply reduced or disappears once we take into account the earnings distribution in the subgroups. As is clearly shown in Table 9, the apparent heterogeneity in response between subgroups stems from differences in the earnings distribution. The impact on the part of the earnings distribution around the initial threshold is not very different. What creates the difference between subgroups is that there are very small probabilities that AFP retirees and the disabled would maintain an earning above the threshold before the reform compared with the wage earners and the self-employed. The same is also true for the difference between men and women (see Figure 11). Hence, gender or labour market status at age 66 does not in itself mean much for the response to the earnings test. What matters is the prereform earnings level. This variation in the prereform earnings distribution could be driven by two types of heterogeneity: in the choice set constraints and in the preferences. Individuals who value their leisure time more, or who have a low earnings capacity, tend to leave the labour force early, either through AFP or through disability if that is an option. An increased marginal wage rate generated by the retirement earnings test reform may then not be enough to induce them to return to the labour force. Unfortunately, with the data currently available, we are not able to decompose the heterogeneity by its cause: due to the differences in preferences or options. Nonetheless, it is important to document this pattern.

VIII. Conclusion

The paper studies the effect of a reform of the retirement earnings test on the earnings and labour supply behaviour of the elderly. Using administrative data from Norway, we first analyse the changes in participation and earnings following the increase in the earnings test threshold from 1 to 2 G in 2002. We see clearly from the data that there are earnings ‘humps’ below the thresholds. After the reform, the original ‘humps’ around the old threshold disappeared, while new peaks are observed around the new threshold. This can be seen as an indication that the elderly are willing and able to adjust their labour supply when their financial incentives changes. We then use a difference-in-difference approach to quantify the effect of the reform. We find that the reform has a substantial impact on the earnings distribution of retirees. There is almost no impact on the extensive margin. On the intensive margin, the effect is very uneven across the earnings distribution and strongest at around the level of the threshold. Those who are holding earnings at the old threshold level increased their earnings by around 20–28%, which corresponds to earnings elasticity with respect to a net-of-tax rate of 0.29 to 0.40. The earnings dispersion is reduced by an increase in median earnings without a corresponding increase in earnings in the highest percentiles. This is in accordance with the theoretical predictions since only the income effect applies to those with the highest earnings. There is no clear evidence that individuals also adjust their labour supply behaviour on the extensive margin.

We also find that the effect of the reform on earnings is larger for the younger cohorts. Since the year of birth can be seen as a measure of exposure to the reform, this result suggests that the higher the degree of exposure is, the stronger the response.

Another interesting pattern is that the labour market status at age 66 has a strong effect on the overall responsiveness to the reform since it is a good predictor of the location of the individual on the earnings distribution before the reform. However, for those who are most likely constrained by the earnings test, the size of the responses does not seem to be much related to the labour market status. This indicates that a large part of the differences in responses to the earnings test reform caused by any unobserved heterogeneity, whether in options or in preferences, can be explained by its position in the earnings distribution, which needs to be further explored.

Final Manuscript Received: January 2012

References

Baker, M. and Benjamin, D. (1999). ‘How do retirement tests affect the labour supply of older men?’ *The Journal of Public Economics*, Vol. 71, pp. 27–51.

Dahl, S., Nilsen, O. and Vaage, K. (2003). ‘Gender differences in early retirement behaviour’, *European Sociological Review*, Vol. 19, pp. 179–198.

Disney, R. and Smith, S. (2002). ‘The labour supply effect of the abolition of the earnings rule for older workers in the United Kingdom’, *The Economic Journal*, Vol. 112, pp. C136–152.

Engelhardt, G. V. and Kumar, A. (2007). *The Repeal of the Retirement Earnings Test and the Labor Supply of Older Men*. CRR WP 2007-1, Center for Retirement Research, Boston College.

French, E. (2005). ‘The effects of health, wealth, and wages on labour supply and retirement behaviour’, *Review of Economic Studies*, Vol. 72, pp. 395–427.

Friedberg, L. (2000). ‘The labor supply effects of the Social Security Earnings Test’, *The Review of Economics and Statistics*, Vol. 82, pp. 48–63.

Friedberg, L. and Webb, A. (2006). *Persistence in Labor Supply Effect and the Response to the Social Security Earnings Test*, Center for Retirement Research, Boston College. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1299700. Accessed on 7th February, 2012.

Gruber, J. and Orszag, P. (2003). ‘Does the Social Security earnings test affect labor supply and benefits receipt?’ *National Tax Journal*, Vol. LVI, pp. 755–773.

Haider, S. J. and Loughran, D. S. (2008). ‘The effect of the Social Security earnings test on male labor supply’, *The Journal of Human Resources*, Vol. XLIII, pp. 57–87.

Hao, L. and Naiman, D. Q. (2007). *Quantile Regression. Series: Quantitative Applications in the Social Sciences* Vol. 149, Sage Publications, Inc., Thousand Oaks, CA, USA.

Hao, L. and Naiman, D. Q. (2010). *Assessing Inequality. Series: Quantitative Applications in the Social Sciences*, Vol. 166, Sage Publications, Inc., Thousand Oaks, CA, USA.

Jia, Z. (2005). *Spousal Influence on Early Retirement Behaviour*, Discussion Paper 406, Research Department, Statistics Norway.

Pedersen, A. W. (2000). *The Coverage with Occupational Pensions in Norway*, FAFO-Notat 2000:23.

Saez, E. (2010). ‘Do taxpayers bunch at kink points?’ *American Economic Journal: Economic Policy*, Vol. 2, pp. 180–212.

Saez, E., Slemrod, J. and Giertz, S. (2010). ‘The elasticity of taxable income with respect to marginal tax rates: a critical review, *Journal of Economic Literature* (forthcoming).

Song, J. G. and Manchester, J. (2007). ‘New evidence on earnings and benefits claims following changes in the retirement earnings test in 2000’, *Journal of Public Economics*, Vol. 91, pp. 669–700.

The Ministry of Labour and Social Inclusion. (2006). *Chapter 1: Governmental White Paper No. 5 (2006–2007)*. Available at: http://www.regjeringen.no/en/dep/aid/Topics/pensions.html?id=939. Accessed 7th February 2012.