IFAU
Institute for Evaluation of Labour Market and Education Policy

Studies on social insurance, income taxation and labor supply

Lisa Laun

DISSERTATION SERIES 2012:1

Presented at the Department of Economics, Stockholm University

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P O Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

This doctoral dissertation was defended for the degree of Doctor in Philosophy at the Department of Economics, Stockholm University, November 9, 2012. The third essay has previously been published by IFAU as Working paper 2012:2 and the fourth essay has previously been published by IFAU as Working paper 2012:18.

ISSN 1651-4149

Acknowledgements

First of all, I thank my supervisor, Mårten Palme, whose door has always been open to me. He has provided me with endless support and encouragement, and has taken the time to critically assess my research in a way that has substantially benefited this thesis. Through him, I have gotten a better sense of what research is all about. Without hesitation, he has also included me in several interesting research projects and introduced me to researchers around the world. I am very grateful for his generosity.

I am also indebted to my co-advisor, Per Johansson. I have benefited from numerous discussions with him about economics and econometrics, and his excitement over our collaborations has served as great motivation. I am also thankful to my other co-authors, Ingemar Svensson for sharing his expertise on the Swedish Social Insurance system, Peter Skogman Thoursie for important methodology discussions, and Benjamin Friedrich, Costas Meghir and Luigi Pistaferri for providing an international perspective to my work.

I thank everyone at the Department of Economics. Ingela Arvidsson and Anita Karlsson provided great administrative support. Many fellow graduate students made this experience more fun. Charlotta Boström, Margherita Bottero, Maria Cheung, Susan Niknami, Martin Nybom, Martin Olsson, André Romahn, Marieke Schnabel, Eric Sjöberg and Linnéa Wickström Östervall deserve a special mention. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

I would not have applied to the graduate program without the encouragement of Matz Dahlberg. I thank Laura Hartman for her support from the very beginning, inviting me to interesting collaborations and introducing me to the Institute for Evaluation of Labour Market and Education Policy (IFAU). I am grateful for the opportunity to spend time at IFAU and for the openness, support and profound excitement about labor economics I have experienced there.

Last but not least, I am fortunate to have met my husband, Tobias Laun, in the graduate program. Without him, this journey would have been much less enjoyable. I thank him for fruitful research discussions and for careful guidance. He is a constant source of advice and the joy of my life.

Table of Contents

Chapter 1

Introduction

Increased longevity and lower birth rates have changed the demographic composition in many developed countries. The age-dependency ratio in Sweden, i.e., the fraction of the population aged 65 or older to the population aged 15 to 64, is projected to increase from 28 to 40 percent from 2010 to 2040. An aging population puts pressure on public finances. The retirement age has not adjusted to increased life expectancy and the number of years in retirement has risen. At the same time, the expectations on welfare services have increased. A challenge for modern welfare states in order to sustain the current level of welfare services in presence of the demographic changes is to enhance labor supply at all possible margins.

Two important instruments for policy makers to affect labor supply in old ages are financial incentives and the design of the social insurance system. The first three papers in this thesis regard the performance of social insurance programs, in particular the sickness and the disability insurance program, through which the early withdrawal from the labor force in Sweden primarily takes place. The fourth paper studies the impact of financial incentives for delayed retirement implemented in the income tax system, in the form of age-targeted tax credits. The last paper concerns a slightly different aspect of the social insurance system, namely how income risk is shared between the individual, the firm and the social insurance system. The paper proposes a strategy for studying the amount and characteristics of income risk, in particular the part that can be attributed to the employer.

Paper 1: Disability Insurance, Population Health, and Employment in Sweden (with Mårten Palme and Ingemar Svensson)

This paper is part of the International Social Security Project at the National Bureau of Economic Research and is published in David A. Wise (2012). The

paper describes the development of population health and disability insurance utilization for older workers in Sweden during the last half century and analyzes the relation between the two. The historic perspective gives a background to the questions analyzed in the remaining studies of this thesis.

We pose three main research questions. The first question is whether the development of disability insurance recipiency over the past decades can be explained by changes in the health status of the population. We find some support for this hypothesis, in the sense that the demographic groups with the least advantageous health development are the same groups with the least advantageous development of disability insurance recipiency. In particular, we find a more adverse development of the health of women compared to men, and the health of younger compared to older individuals. The same pattern can be found in the development of disability insurance recipiency. Overall, however, changes in the underlying health of the population do not seem to be able to explain the large fluctuations in the inflow to the disability insurance program. The relation between population health and disability insurance recipiency is least apparent for the oldest age group, aged 60 to 64.

The second question is whether the changes in disability insurance recipiency can be explained by changes in the eligibility rules in the disability insurance program. We focus on the introduction and abolishment of two sets of eligibility rules that only affected the oldest age group, aged 60 to 64. The first is the special eligibility rules for older workers, implying an exempt from rehabilitation and retraining, lowered requirements for the medical assessment of working capacity and a possibility to consider functional limitations due to normal aging for eligibility to disability benefits. The second is the pure labor market reasons for older workers, making them eligible for disability benefits if still unemployed when reaching the time limit for unemployment benefits. The introduction of the special eligibility rules in 1970 seems to have had a large impact on disability insurance recipiency, whereas the introduction of pure labor market reasons in 1972 seems to have induced only a small additional increase in program entry. The abolishment of the pure labor market reasons in 1991 seems to have reduced disability insurance recipiency at older ages. The abolishment of the special eligibility rules for older workers in 1997 seems to have merely had a temporary impact.

The final question is to what extent the changes in eligibility rules for older workers affected employment and labor force participation. The answer is ambiguous. The 1970 reform seems to have opened the disability insurance program as a much more frequent exit route from employment than before. In fact, the disability insurance program became almost the only pathway out of the labor force before the normal retirement age in the decades following the reform. The results are complicated to analyze for the 1991 and 1997 reforms.

The 1991 reform coincided with a sharp drop in the employment rate, caused by a labor demand shock from a severe recession. The 1997 reform did not seem to lead to a significant increase in employment, but rather an increased utilization of the unemployment and sickness insurances.

Paper 2: Does Privatization of Vocational Rehabilitation Improve Labor Market Opportunities? Evidence from a Field Experiment in Sweden (with Peter Skogman Thoursie)

This paper analyzes if privately provided vocational rehabilitation can improve labor market opportunities of individuals on long-term sick leave, compared to rehabilitation provided by the public. The motivation for private provision of public services is to induce the incentives for innovation and cost reduction, but the risk is that the incentives to engage in pure cost reduction are too strong, which could deter the quality of the services.[1] The merit of outsourcing public services is ultimately an empirical question, and existing studies suggest that the result depends on the nature of the service provided.

We use a field experiment performed by the Employment Service and the Social Insurance Agency in Sweden from 2008 to 2010, in which over 4,000 participants were randomly offered private and public rehabilitation. The participants had been receiving sickness benefits for at least two years or were receiving temporary disability benefits, and were either unemployed or unable to return to their previous workplace for health reasons. The private providers competed for contracts through public tenders, competed for clients through consumer choice and were rewarded based on results in terms of acquiring employment. The public provider faced no financial incentives but was subject to a quantitative performance goal when operating on a regular basis. Rehabilitation could be provided for up to one year.

Our main finding is that there was no difference in the success of acquiring employment following rehabilitation between private and public rehabilitation providers up to two years after the vocational rehabilitation was initiated. There was a substantial transition to employment following rehabilitation for participants at both types of providers but the transition rates were the same across the two groups. The development over time by month since randomization was also remarkably similar. Rough calculations indicate that the cost of private and public rehabilitation was the same. This suggests that there are no large efficiency gains from privatizing vocational rehabilitation.

[1] See, e.g., Hart, Schleifer and Vishny (1997).

Paper 3: Screening Stringency in the Disability Insurance Program (with Per Johansson and Tobias Laun)

This paper proposes a strategy for assessing how the inflow to the disability insurance program has been governed over time. While the first paper in this thesis analyzes the association between population health and disability insurance recipiency, this study provides a more formal framework to account for changes in health when studying program inflow. We analyze the ex-ante health of individuals entering the program by using the ex-post mortality from all years in which the individuals are observed in the future, compared to individuals not entering the program in the same year.

Our strategy captures the development of screening stringency, due to changes in formal eligibility criteria as well as in implementation. The latter aspect has been difficult to assess in previous studies. There may also be other factors affecting the relative health of new disability beneficiaries compared to non-beneficiaries. Program generosity is a different policy instrument for governing program inflow. The impact of labor market factors may also be important, and our measure can be taken as to capture the ability of the disability insurance program to withstand the influence of outside forces.

To structure the analysis, we provide a theoretical model of the application decision to the disability insurance program. We focus on screening stringency as a key determinant of program inflow. A potential indicator of screening stringency is the denial rate. Since Parsons (1991), however, it is well known that initial eligibility determination is an important self-screening mechanism. We show that the composition of the applicant pool changes with screening stringency, and the effect of stringency on the denial rate is therefore undetermined. Going further, we show that screening stringency has implications for the health of new disability beneficiaries. We derive a relationship showing that the relative health of new beneficiaries compared to non-beneficiaries improves when stringency is reduced.

We use duration analysis to estimate the mortality ratio of new disability beneficiaries to non-beneficiaries over time. The differential development of health across cohorts is accounted for by estimating cohort-specific baseline mortality hazards. The estimation strategy can be used to assess how program inflow has been governed in general as well as across population groups. We apply the strategy to the disability insurance program in Sweden between 1985 and 2008. Sweden is an interesting application since the inflow to the disability insurance program has fluctuated considerably over time, in a fashion that cannot always be related to changes in formal program rules.

We find that changes in the mortality ratio of new disability beneficiaries compared to non-beneficiaries correspond well to the fluctuations in the dis-

ability benefit entry rate over time. This suggests that the variation in program inflow cannot be accounted for by changes in the underlying health of the population. Changes to formal eligibility criteria are indeed reflected in the mortality ratio. The removal of eligibility for pure labor market reasons for workers aged 60–64 in 1991 can be seen in a substantial and lasting increase in the relative mortality ratio of older compared to younger beneficiaries. Also the removal of the special eligibility rules for older workers and eligibility for labor market and medical reasons combined in 1997 is reflected in the relative health of new beneficiaries across age groups, although this reform seems to have had more of a transitory impact. During the period under study, program inflow was more strictly governed in Stockholm compared to other regions, but the relative mortality ratio across regions has converged after the possibilities of taking labor market reasons into account were removed in the 1990s.

We also find fluctuations in the mortality ratio of new beneficiaries compared to non-beneficiaries during periods when no formal program changes were being pursued. The large increase in program entry during the early 2000s is related to a decrease in the relative mortality of new beneficiaries, and the large inflow of women compared to men during this period corresponds to a relatively lower mortality ratio of female compared to male beneficiaries. Perhaps most strikingly, the rapid fall in program entry since 2005 is reflected in a substantial increase in the mortality ratio of new beneficiaries compared to non-beneficiaries, despite the lack of formal program changes. The turning point in 2005 coincides with the re-organization of the regional Social Insurance Agency offices into one central authority.

Paper 4: The Effect of Age-Targeted Tax Credits on Retirement Behavior

A key remedy for the fiscal pressure of an aging population, facing many developed countries, is to delay retirement. The main approach to encourage a delayed labor force exit has been through social security reform. The financial incentives for early retirement inherent in many social security systems, as shown in Gruber and Wise (1999), have been removed in many countries and age limits have been increased. An alternative approach, that has not yet been widely used, is to adjust labor tax rates close to the retirement age. If the labor supply elasticity near retirement is large, as suggested by French (2005), the cost of reduced tax rates might be offset by the revenue from an increased tax base. If the elasticity is small, reduced labor tax rates for older workers will merely work as a transfer to individuals who would have continued to work regardless.

This paper analyzes the effect on retirement behavior of two age-targeted

tax credits that were simultaneously introduced in Sweden in 2007: an earned income tax credit and a payroll tax credit. The identification strategy exploits that the tax credits applied to workers aged 65 or above at the beginning of the tax year. Therefore, individuals who turn 65 just before or just after the new year face different labor tax rates at essentially the same age. Public pension benefits, on the other hand, are determined by the age at retirement and other differences between the groups can be controlled for using a difference-in-differences approach, to the extent that they do not change discontinuously at the time of the reform.

The results suggest that the age-targeted tax credits increased employment in the year following the 65th birthday by 1.5 percentage points, earnings as a share of previous earnings by 1.8 percentage points and the number of remunerated months by 0.133 among individuals with annual earnings above the 2007 tax liability threshold at least one year three to five years earlier. The results are shown to be robust to various specification tests. An analysis of heterogeneous responses indicates that the effects are primarily driven by men and are particularly large for self-employed. In an attempt to study the relative importance of the two types of tax credits the variation in the size of the credits by income and year is being used, but the analysis gives no conclusive result.

From a policy perspective, the changes in retirement behavior induced by the reform appear to be moderate. The public finance costs of the tax credits are large and rough calculations suggest that the benefits from the behavioral changes amount to only a small share of reform costs. Interestingly, for the behavioral changes to fully finance the reform would require employment elasticities of very large magnitudes, compared to what has typically been found in the labor supply literature. Although the tax credits, in terms of foregone tax revenues, showed to be a costly way to increase labor supply at older ages, however, changes in retirement behavior may have long-run benefits that are not accounted for in the study and that may interact with other possible future policy changes.

The paper makes two main contributions to the previous literature. First, although there is a large literature on the effect of social security systems on retirement behavior and a growing literature on labor market responses to different earned income tax credit and payroll tax initiatives, it is, to the author's knowledge, the first paper to study the effect of income tax policy changes on retirement behavior. As argued in Blundell and MaCurdy (1999), the participation decision is likely the most responsive margin of labor supply. This motivates reforms directly aimed at increasing participation among certain types of workers with potentially elastic labor supply; an often targeted group has been low income families. The study analyzes how targeted income tax reforms affect another group with a potential labor supply reserve, namely workers at the

margin of retirement. Second, the identification strategy, which exploits differences in the organization of the income tax and the social security systems with respect to cohort versus age based incentives, has not been previously used in the empirical public finance literature.

Paper 5: Wage Dynamics and Firm-Level Shocks (with Benjamin Friedrich, Costas Meghir and Luigi Pistaferri)

Finally, the fifth paper proposes a framework for introducing the firm into empirical models of the dynamic income process. The firm side has typically been neglected in the literature studying the nature of the income process, but the recent availability of matched employer-employee data gives rise to major new opportunities in this direction. In a competitive labor market, workers only bare the risk of shocks to their productivity, but in the presence of search frictions, there are multiple sources of risk distinct from workers' productivity shocks. A potential source of risk is shocks at the firm level, but we know little about the extent to which firm-level productivity shocks are transmitted to wages.

First, most equilibrium search models that have been estimated on empirical data assume no productivity shocks. Second, models that have been estimated on matched employer-employee data, without a specific economic structure, have focused on sorting and firm-worker heterogeneity rather than the dynamics of shocks. A recent paper by Guiso, Pistaferri and Schivardi (2005) has indeed measured the impact of firm-level shocks on wages using matched employer-employee data. However, their approach is limited by the fact that they ignore job to job mobility and transitions between employment and unemployment. Such transitions may well hide the impact of shocks on wages because a worker may quit instead of suffering too large a pay cut.

In the study, we propose a model of the income process that can be used to analyze whether firm-level productivity shocks are transmitted to wages. The key innovation is that we account both for job to job transitions and for transitions between employment and unemployment, to capture the role of job mobility and labor force participation in hiding the impact of shocks. We allow for a rich stochastic structure of the income process, both at the individual and the firm-worker match level. We also present a strategy for estimation of the model and identification of the key parameters. The paper is the opening investigation into an agenda that will lead to richer structural models taking the firm side more seriously an allowing for a rich dynamic stochastic structure of the income process.

References

Blundell, R., and T. MaCurdy (1999), "Labor Supply: A Review of Alternative Approaches," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics Volume 3*, Elsevier Science B.V., pages 1559–1695.

French, E. (2005), "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies*, vol. 72, pages 395–427.

Gruber, J., and D. A. Wise (1999), *Social Security and Retirement Around the World*, Chicago: The University of Chicago Press.

Guiso, L., L. Pistaferri, and F. Schivardi, (2005). "Insurance within the Firm," *Journal of Political Economy*, vol. 113, pages 1054–1087.

Hart O., A. Schleifer, and R. Vishny (1997), "The Proper Scope of Government: Theory and an Application to Prisons," *Quarterly Journal of Economics*, vol. 112, pages 1127–1161.

Parsons, D. (1991), "Self-Screening in Targeted Public Transfer Programs," *Journal of Political Economy*, vol. 99, pages 859–876.

Wise, D. A. (2012), *Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms*, Chicago: University of Chicago Press.

Chapter 2

Disability Insurance, Population Health, and Employment in Sweden*

Lisa Laun Mårten Palme Ingemar Svensson

Abstract

This paper describes the development of population health and disability insurance utilization for older workers in Sweden and analyzes the relation between the two. We use three different measures of population health: (1) the mortality rate (measured between 1950 and 2009); (2) the prevalence of different types of health deficiencies obtained from Statistics Sweden's Survey on Living Conditions (ULF, 1975–2005); (3) the utilization of health care from the inpatient register (1968–2008). We also study the development of the relative health between disability insurance recipients and non-recipients. Finally, we study the effect of the introduction of less strict eligibility criteria for older workers in 1970 and 1972 as well as the subsequent abolishment of these rules in 1991 and 1997, respectively.

Keywords: Disability insurance, Population health
JEL-codes: H51, H55, I18, J26

*This paper is a part of the National Bureau of Economic Research project *International Social Security* and is published in David A. Wise (ed.) (2012), *Social Security Programs and Retirement Around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms*, Chicago: The University of Chicago Press. We are grateful to Agneta Kruse and Peter Skogman Thoursie as well as participants on seminars at the ISS meeting for comments on previous drafts of the paper. We gratefully acknowledge financial support from the Bank of Sweden Tercentenary Foundation, the Swedish Council for Working Life and the Jan Wallander and Tom Hedelius Foundation.

1. Introduction

Compared to other industrialized countries, a large share of the Swedish population receives support from the disability insurance program (see Gruber and Wise, 2010). In 2009, 20 percent of the males and more than 30 percent of the females in the age group 60–64 received disability benefits. The disability insurance is the most common pathway out of the labor force for those who exit before the normal retirement age. In 2009, the expenditures from this program corresponded to 1.8 percent of GDP.

Despite the extensive usage of the disability insurance, Sweden has a comparatively high employment rate among older workers. About 70 percent of the population aged 55–64 were employed in 2005, as compared to about 50 percent in Germany, 40 percent in Italy and 60 percent in the United States (see Gruber and Wise, 2010). As in most other European countries, however, there has been a dramatic fall in the employment rate of older men in recent decades. For example, the employment rate of males aged 60–64 has decreased from above 80 percent in the early 1960s to slightly above 60 percent today. This development has caused concern in view of future financial burdens of an aging population.

An explanation for the comparatively high employment rate among older workers combined with high disability insurance recipiency is that Sweden does not have a generous early retirement program. Many European countries introduced such programs in the 1970s and 1980s and the large cross-country differences in employment rates among older workers emerged during this period. For some time, however, the disability insurance program in Sweden developed towards an early retirement scheme. From only awarding disability benefits for health reasons in the 1960s, less strict eligibility criteria, especially for older workers, were introduced in the 1970s. These rules were abolished in the 1990s, and since 1997 an impaired work capacity for health reasons is again the sole eligibility criteria for disability benefits.

In this paper, we study to what extent the evolution of disability insurance utilization can be explained by changes in the population health status and by changes in eligibility rules, respectively. We focus on the age group 45–64, which is the most important for the utilization of the disability insurance.[1] We pose three main research questions. First, is there a relationship between disability insurance utilization and the development of population health status in recent decades? Second, did the changes in eligibility rules for older workers affect disability insurance utilization? Third, did the changes in eligibility rules

[1] Disability insurance recipiency in younger ages has increased over time, which is a source of concern. In this paper, however, we limit our focus to the utilization of the disability insurance in older ages.

for older workers affect labor market outcomes such as employment and labor-force participation, or where they "crowded out" by the utilization of other income security programs?

Wadensjö (1996) and Hedström (1987) have previously analyzed the effect of changes in eligibility rules of the disability insurance program in Sweden, in particular the introduction of eligibility rules for labor market reasons targeted at older workers in the early 1970s. Karlström, Palme and Svensson (2008) studied the abolishment of the special eligibility rules for older workers in 1997. In this study, we extend the previous literature by considering the full 40 year history of eligibility changes starting in the early 1970s and by relating it to different labor market outcomes. In addition, we put together a comparatively wide set of population health measures and relate the development of these measures to the development of disability insurance utilization in different demographic groups.

Although we strive to have a broad scope for the empirical analysis, we leave out several plausible explanations for the fluctuations in the utilization of the disability insurance. Previous studies have analyzed the effect of economic incentives on the disability insurance in Sweden (e.g. Kruse and Söderström, 1989; Skogman Thoursie, 1999; and Palme and Svensson, 1999 and 2004). Changes in social norms regarding the utilization of the sickness insurance program have been studied by Lindbeck, Palme and Persson (2009), and should be a plausible explanation also for changes in the utilization of the disability insurance. The implementation of rules may also be affected by administrative policies within the social insurance system. Finally, changes in the demand for labor with disabilities have not yet been properly studied on Swedish data, but it is an interesting topic for further research.

The rest of the paper is organized as follows. Section 2 gives a brief history of the development of the disability insurance program in Sweden. Section 3 describes the development of various population health measures over time. Section 4 describes the development of the utilization of disability insurance and the development of labor market outcomes. Section 5 studies the relation between population health and disability insurance utilization. Section 6 studies the relation between the eligibility reforms for older workers, disability insurance utilization and various labor market outcomes. Section 7 concludes.

2. Historical Overview of the Disability Insurance in Sweden

The disability insurance (DI) is one of Sweden's most important income security programs. Its main objective is to replace foregone earnings for workers below the retirement age with a permanently impaired working ability for health reasons. The related sickness insurance replaces foregone earnings due

to a temporarily impaired working ability for health reasons. Disability benefits can be granted part time or full time, depending on the extent of the work impairment.

Sweden's disability insurance has a comparatively long history. The first public pension system covering all citizens, including an invalidity pension, was implemented already in 1914. The recent history of Sweden's disability insurance, which we analyze in this paper, started when a public income related supplementary pension scheme (ATP) was introduced, following a referendum in 1957. The new scheme came into place in 1960 and the first payments were made in 1963, but since the program was phased in, it did not reach its full maturity until the beginning of the 1990s. The pension benefit under this scheme replaced 60 percent of the average of an individual's 15 best years of earnings up to a social security ceiling. The benefit was linearly reduced if the worker contributed less than 30 years to the scheme and it was financed through payroll taxes.

The new income related pension also included disability insurance. The size of the individual benefit was calculated in the same way as the old-age pension benefit, with the actual earnings history replaced by an assumed earnings profile. Eligibility for disability benefits was initially based on health. Disability benefits were awarded by the local Social Insurance Agency after a physical examination by a medical doctor. If the health status prevented the worker from doing his or her regular job, but not one that suited the worker's general qualifications, the worker was required to go through a retraining program.

Eligibility rules for disability benefits changed on several occasions after the new disability insurance was first introduced. Table 2.1 summarizes the main eras in this history. The first major reform took place in 1970 and had two main components. First, special eligibility rules were introduced for workers aged between 63 and the normal retirement age, the age of 67 at that time. These rules implied that (a) no rehabilitation or retraining for a new occupation was required if the worker's health status did not permit his or her regular work; (b) the medical requirements for assessing inability to work were substantially lower for this age group; and (c) also functional limitations due to normal aging could be considered for eligibility for DI. Second, unemployment was made an additional criterion for DI eligibility in all age groups. Long-term unemployed workers with functional limitations were made eligible for DI after having been unemployed for 1–2 years.

The next reform towards more generous eligibility rules for DI took place in 1972, when pure labor market reasons for older workers were introduced. These rules implied that workers aged between 63 and the normal retirement age could become eligible for DI if they were still unemployed when reaching

Period	Medical reasons	Labor market and medical reasons combined	Special eligibility rules for older workers	Pure labor market reasons for older workers
–1962	Yes	Very small	No	No
1963–1970(June)	Yes	Some	No	No
1970(July)–1972(June)	Yes	Yes	Yes, aged 63–66	No
1972(July)–1974(June)	Yes	Yes	Yes, aged 63–66	Yes, aged 63–66
1974(July)–1976(June)	Yes	Yes	Yes, aged 63–66	Yes, aged 60–66
1976(July)–1991(Sept)	Yes	Yes	Yes, aged 60–64	Yes, aged 60–64
1991(Oct)–1996	Yes	Yes	Yes, aged 60–64	No
1997–	Yes	Very small	No	No

Table 2.1: Changes in eligibility rules for the disability insurance

the time limit in the unemployment insurance, even without any health limitations. In 1974, the age limit for pure labor market reasons was lowered from age 63 to age 60, and in 1976 the age limit for the special eligibility rules for older workers, introduced in 1970, was lowered from age 63 to age 60. The latter change was partly made as a consequence of the decrease in the normal retirement age from age 67 to 65 in 1976.

Two changes led to higher replacement levels in the disability insurance during the 1970s and 1980s. First, the maturity of the supplementary pension scheme (ATP) led to higher replacement levels in general. Second, the introduction of a "special supplement" in 1969 led to improvements for low income pensioners. It was reduced on a one-to-one basis against income from the supplementary pension (ATP). The special supplement applied to all types of pension and was gradually increased from 1969 to 1981. In 1977, the amount of the special supplement was doubled for DI pensioners only, and the subsequent development of the supplement for DI beneficiaries followed the gradual increase for regular pensioner but at twice as high a level.

The policy toward more liberal DI eligibility rules was reversed in the 1990s. The eligibility for DI for pure labor market reasons for older workers, introduced in 1972, was abolished in 1991. In 1991 and 1992, a new legislation was also enacted with the purpose of reducing sickness absence, which required employers to improve the work environment and take responsibility for the rehabilitation of employees. The Social Insurance Agencies were made responsible for the coordination of rehabilitation among the employer, the public health care system, labor market authorities, the local government and the individual. In 1997, also the favorable eligibility rules for older workers and the eligibility for DI for labor market reasons and medical reasons combined, introduced in 1970, were abolished. Since then, an impaired work capacity for health reasons has been the only eligibility criterion for disability insurance.

On 1 January 2003, the disability insurance was shifted from the public pension system to the public sickness insurance system, following a major pension reform. Benefits were renamed and the calculation of benefits changed, but the assessment of eligibility remained the same. Benefits were calculated as 64 percent of the assumed income, i.e., the income of the best three of the last five to eight years, depending on age, up to a social security ceiling. Individuals not qualifying for the income related insurance received a guarantee benefit.

From January 2005 onwards a re-assessment of the working capacity for granted individuals should be made every third year. Also in 2005, the organization of the Social Insurance Agency changed, when the 21 regional offices were integrated into one central authority. In 2008, the eligibility for disability benefits was substantially tightened. For all cases granted after 1 July 2008, working capacity had to be permanently reduced in relation to the entire labor market in order to qualify for benefits.

3. The Development of Population Health

Changes in population health would be a natural determinant of the evolution of disability insurance recipiency over time. To describe the development of population health in Sweden over the last decades, we use three main groups of health measures: the mortality rate, self-reported health measures from the Survey on Living Conditions and, finally, the utilization of inpatient care from the National Patient Register. For our purposes, each of the three measures of population health has its obvious advantages and disadvantages.

The main advantage of mortality as a health measure is that it is objective and accurately measured through the population censuses, which makes it easily comparable across countries and time. A disadvantage is that mortality is more of an outcome measure, in part reflecting technological advances within the health care system rather than the average health status of the population. It might also target a slightly inappropriate population for our purposes, since it measures the health of the marginal survivors rather than the marginal workers. Diseases that cause death and diseases that reduce working capacity may be different. Finally, a higher survival rate implies that individuals who would previously have died now survive, although possibly with a bad health status. This could change the composition of the population and in turn affect the average health status negatively although mortality decreases.

The advantage of the self-reported health measures is that they are better targeted towards the aspects of health that are relevant for the ability to remain in the labor force. The main disadvantage relates to the fact that they are subjective. Fluctuations over time might capture changes in the interpretation

of the questions and the general view of health rather than changes in actual health. Another disadvantage is that the self-reported health measures may be state dependent. For example, since an impaired work capacity is an eligibility criterion for disability benefits, individuals may be more likely to report an impaired work capacity as a result of receiving benefits.

The advantage of the utilization of inpatient care as a health measure is that it is accurately measured, since it is obtained from registers, and closely related to the health status of the worker. It has, however, the disadvantage of the self-reported measures of being sensitive to changes in the general view of health. Furthermore, the utilization of inpatient care might be influenced by public health care spending, working procedures at the hospitals and the division of labor between outpatient and inpatient care.

3.1. Mortality

The mortality rate is defined as the number of deaths during a year divided by the average size of the population in a particular age group. We present the development of mortality in Sweden in three different ways. Figure 2.1 shows the development from 1960 to 2009 of the age at which men and women faced the same mortality rate as that which the 60 and 65 year olds faced in 1960. Figure 2.2 shows the mortality rate by age in 1960 and 2005 for men and women, respectively. Finally, Figure 2.3 shows the development of the mortality rate for men and women, respectively, at the age of 55, 60 and 65 from 1950 to 2009.

These figures reveal two interesting results. First, there has been a marked decrease in mortality for both men and women during the period under study. Figure 2.1 shows that the age of equal mortality as that which the 60 year olds faced in 1960 increased by 8.4 years for men and 7.1 years for women until 2009, and that the age of equal mortality as that which the 65 year olds faced in 1960 increased by about 7.5 years for both genders until 2009. Figure 2.2 shows that the age at which the mortality rate passes five percent increased by 5 years for men and 7 years for women between 1960 and 2005. Finally, Figure 2.3 shows that the mortality rate has halved from 1950 to 2009 for men and women in all age groups.

The second result is that the large decrease in mortality happened much earlier for women. Figure 2.3 shows that the mortality rate for 65 year old women decreased from about 2 percent in 1950 to about 1 percent in 1980. Since then, it has decreased by less than 0.25 percentage points. Also for women at younger ages, the mortality rate decreased most rapidly before 1980. For men, on the other hand, the main improvement occurred after 1980. Between 1980 and 2009, the mortality rate for 65 year old men decreased by

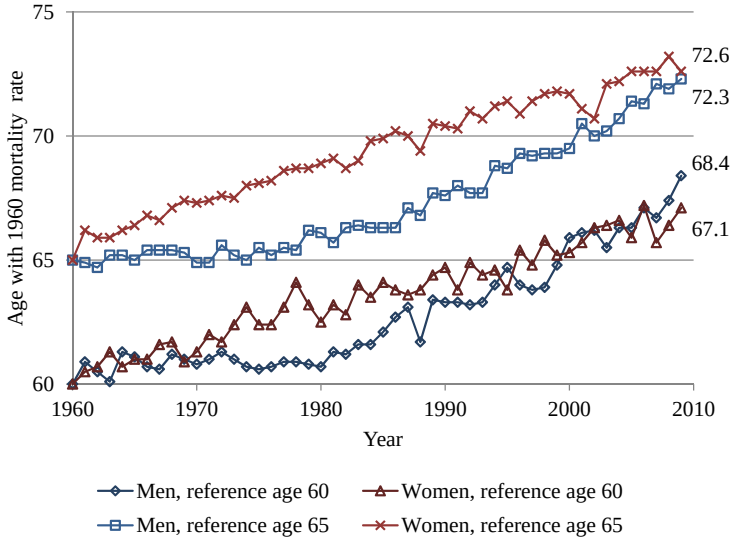


Figure 2.1: Ages of equal mortality probability, 1960–2009
Source: Statistics Sweden

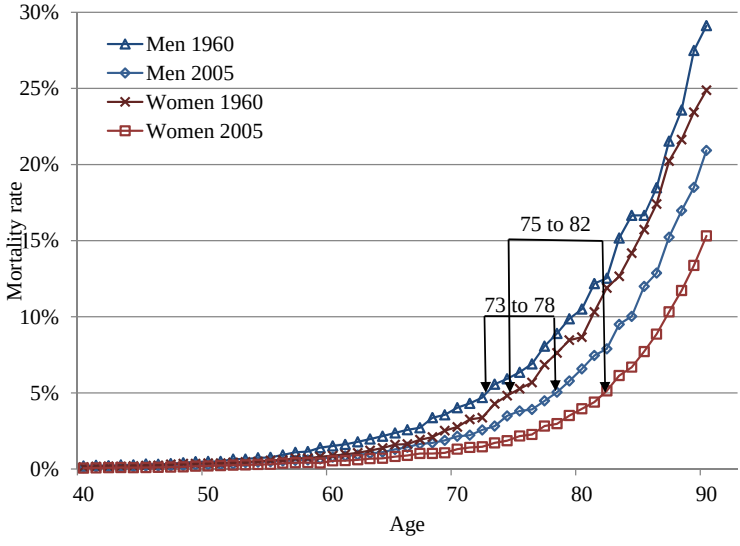
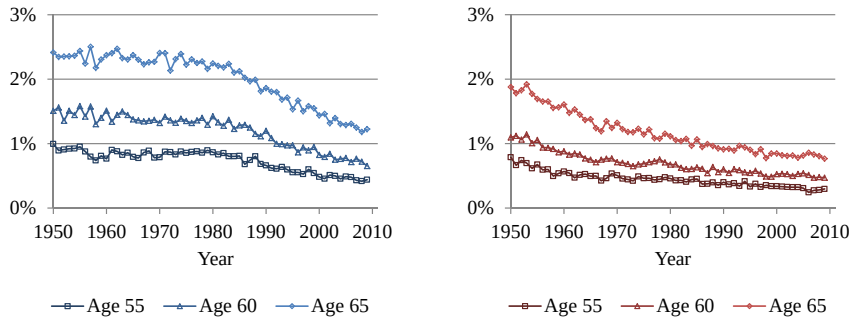


Figure 2.2: Mortality rates by age, 1960 and 2005
Source: Statistics Sweden



(a) Men (b) Women

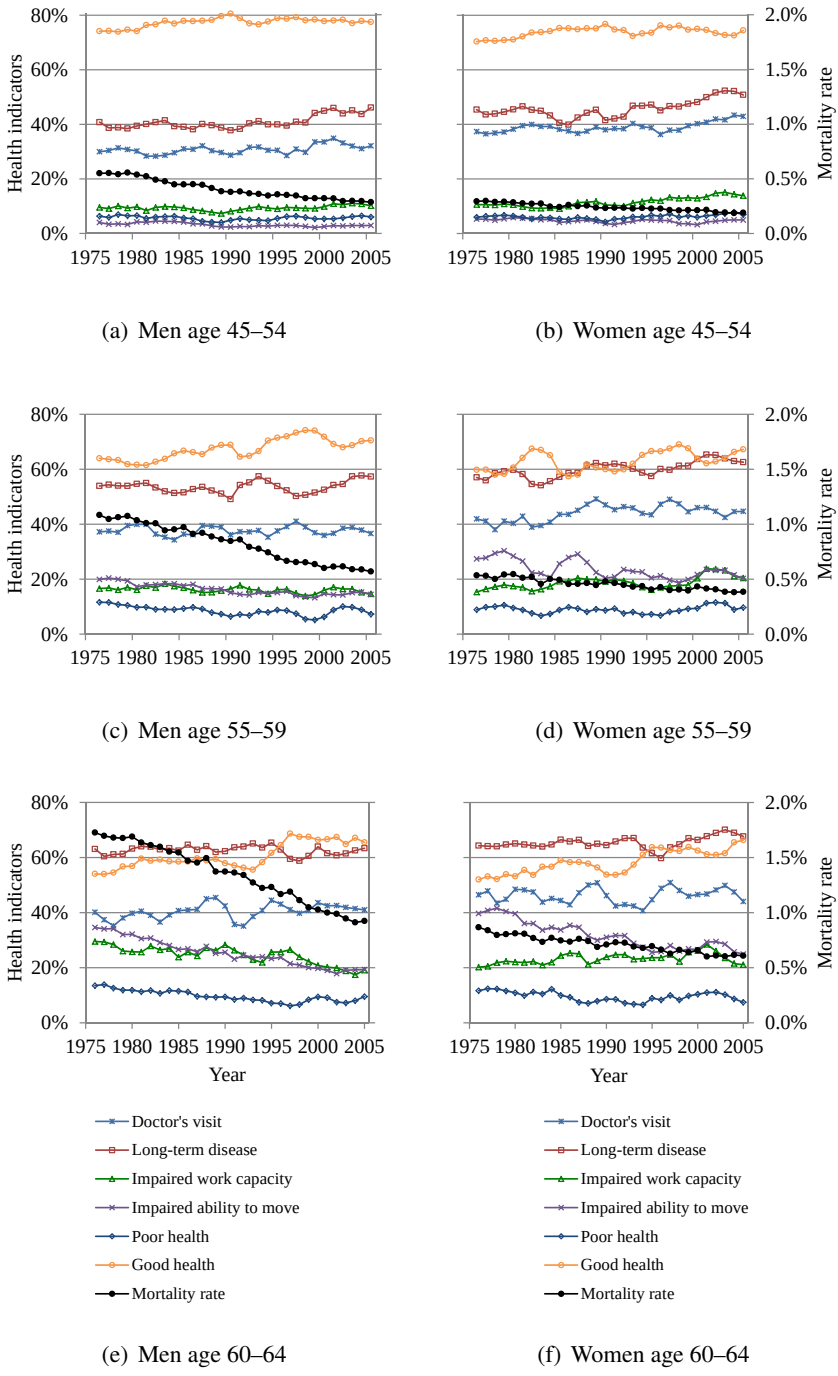
Figure 2.3: Mortality rates by year, 1950–2009
Source: Statistics Sweden

1 percentage point to just over 1 percent. Also for younger men, mortality decreased most rapidly during this period. The same pattern is visible in the series of equal mortality probability in Figure 2.1. The gap between men and women broadened until the mid 1980s, and thereafter narrowed substantially. The development for the youngest males is so steep after 1980 that it even surpasses the development for women in the mid 2000s.

3.2. Self-Reported Health

Self-reported information about the health of the Swedish population is collected by Statistics Sweden through the Survey on Living Conditions (ULF). It is a yearly survey of a random sample of about 7,500 individuals aged 18–64 that has been produced since 1975. The survey contains a large set of questions about health in general as well as about particular diseases. We present the results from the survey for men and women in the age groups 45–54, 55–59 and 60–64. All series show the share of the population in the age group with a certain condition. To reduce the problem of large stochastic errors due to small sample sizes within each demographic group, we present three-year moving averages. We also focus on the long-run development of the series rather than fluctuations in single years.

Figure 2.4 presents the development of a set of general health indicators from the survey. The indicator "Doctor's visit" shows the share of the population who visited a doctor within the last three months. The indicator "Long-term disease" shows the share of the population with at least one disease in a list of diagnoses and the indicator "Impaired work capacity" shows the share of the population reporting that the long-term disease causes an impaired work



(a) Men age 45–54

(b) Women age 45–54

(c) Men age 55–59

(d) Women age 55–59

(e) Men age 60–64

(f) Women age 60–64

Figure 2.4: Self-reported health indicators and mortality, 1976–2005
Source: Statistics Sweden

capacity. The indicator "Impaired ability to move" shows the share of the population who are not able to run 100 meter. Finally, the indicators "Poor health" and "Good health" show the self-assessed health status, based on a question where the individual evaluates his or her general health on a particular scale.[2] The mortality rate in each demographic group is included as a comparison.

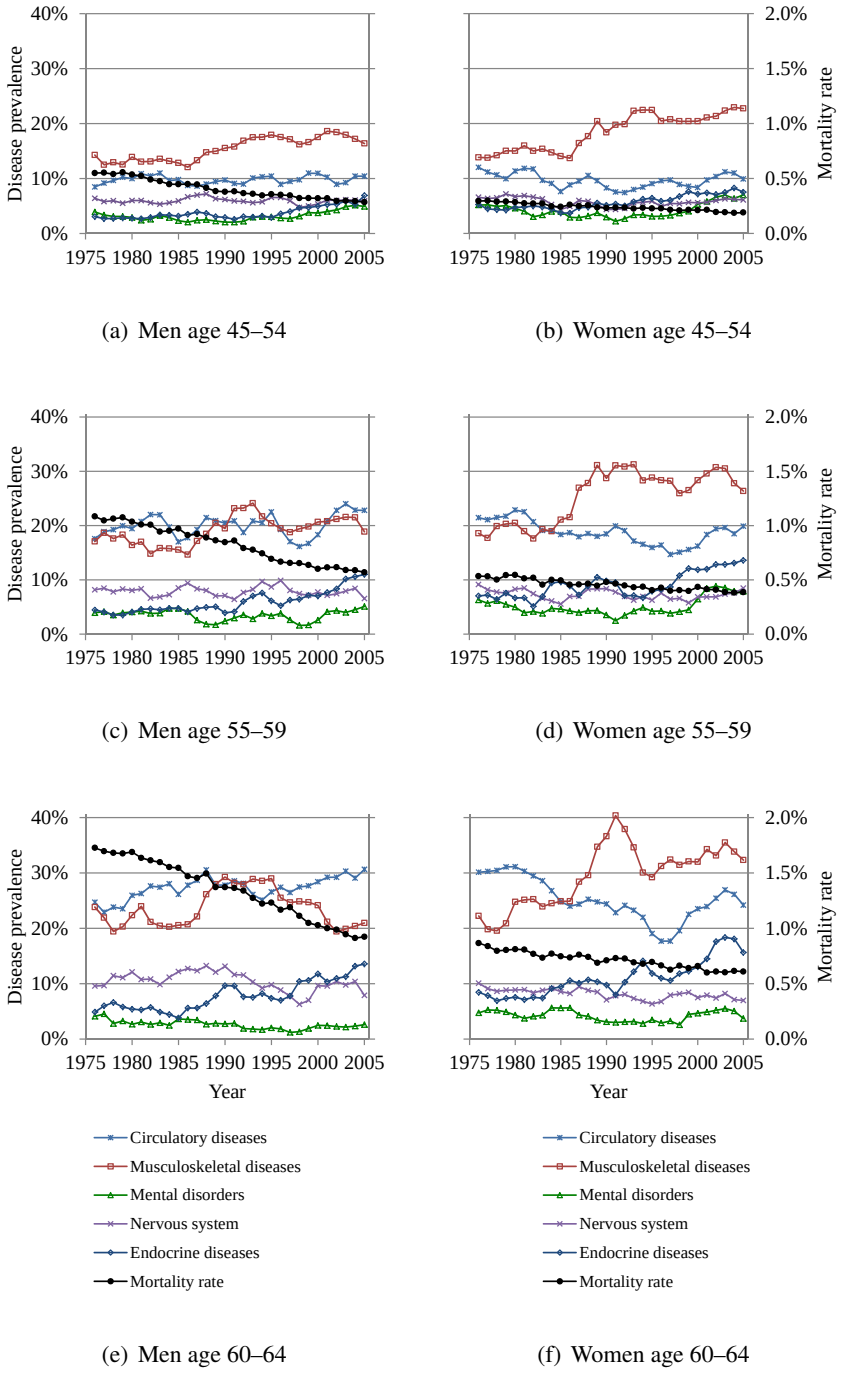
As opposed to the development of mortality, Figure 2.4 shows no overall trend for the development of the self-reported health indicators over time for neither men nor women. Two of the indicators, the share of the population with a long-term disease and the share of the population that recently visited a doctor, show an adverse or invariant development over time in all demographic groups. Both of these indicators have increased for men and women aged 45–54 and women aged 55–59 and have remained stable for men aged 55–59 and men and women aged 60–64. In general, these two indicators seem to have developed more adversely for women than for men and for younger than for older age groups.

Other self-reported health indicators have developed in opposite directions in different demographic groups. The share of the population with an impaired work capacity has remained stable over time for men aged 45–54 and 55–59, but has decreased by a third for men aged 60–64 since 1976. For women, there is instead an upward trend, in particular for the 45–54 and 55–59 year olds. Also women aged 60–64 show a slight upward trend, implying that the share of women with an impaired work capacity has not developed in the same promising way as the share of men with an impaired work capacity in this age group.

The remaining health indicators show an invariant or advantageous development over time. The health indicator with the most favorable development is the share of the population with an impaired ability to move, which has decreased in all demographic groups. The reduction has been particularly large for men and women aged 60–64, where the share has almost halved over the period. This is the only indicator improving over time for all female age groups.

The share of the population in poor health remained stable for men and women aged 45–54 and women aged 55–59, and decreased slightly for men aged 55–59 and men and women aged 60–64. This broad pattern is supported by the development of the share of the population in good health, which has remained stable for men and women in the age group 45–54 and has improved for the two older age groups. In the age group 60–64, the share of the population in good health has increased from about 55 to 65 percent for both men and women.

[2] See Statistics Sweden (2009) for more information.



(a) Men age 45–54

(b) Women age 45–54

(c) Men age 55–59

(d) Women age 55–59

(e) Men age 60–64

(f) Women age 60–64

Figure 2.5: Self-reported disease prevalence and mortality, 1976–2005
Source: Statistics Sweden

In addition to the general health indicators, the Survey on Living Conditions contains information about self-reported disease prevalence for a number of diagnoses. Figure 2.5 shows the development over time for the prevalence of diseases for five diagnosis groups that are of particular importance for the disability insurance. These include circulatory diseases, musculoskeletal diseases, mental disorders, diseases in the nervous system and endocrine diseases. The development of the mortality rate in each demographic group is included as a comparison.

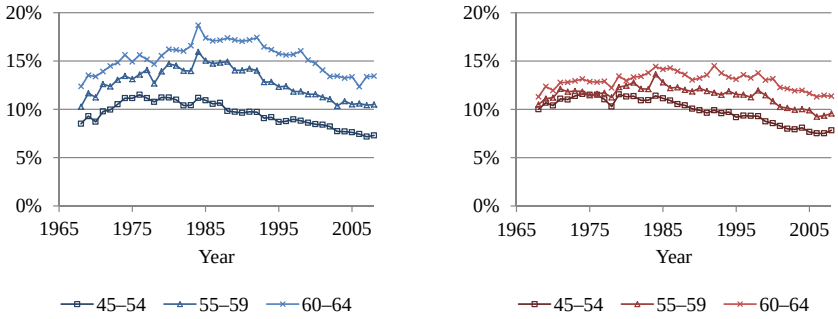
The development of self-reported disease prevalence does not reveal any unambiguous trend. Different diagnosis groups follow different patterns over time. The prevalence of musculoskeletal diseases has increased in all demographic groups, except men aged 60–64, since the mid 1980s. The increase has been particularly striking for women. Also the prevalence of endocrine diseases has increased in all demographic groups since the mid 1980s. This is likely to reflect that problems related to obesity, such as diabetes, have become more common. The prevalence of mental disorders has remained stable for most of the period, but has increased since 1995 in the two youngest age groups. The prevalence of nervous diseases has remained stable for the two youngest age groups and has decreased slightly in the age group 60–64. Although fluctuations in the prevalence of circulatory diseases have been large, there are no clear patterns in the long run.

3.3. Inpatient Care

The utilization of inpatient care is registered in The National Patient Register at The National Board of Health and Welfare. The register contains all overnight hospital visits in Sweden from 1987 and onwards. For a selection of counties, however, the register contains information from as far back as 1968. To get a longer perspective, we use information about inpatient care for four of Sweden's twenty-one counties from 1968 to 2008. The presented series show the share of the population in the four counties taken together that experienced at least one overnight hospital visit during the year.[3]

Figure 2.6 shows the development of inpatient care from 1968 to 2008 for men and women in the three age groups. In the two oldest age groups, the utilization of inpatient care has been higher for men than for women throughout the period. In the youngest age group, utilization has been very similar across genders, although slightly higher for women. As expected, the utilization of inpatient care increases by age. Over time, the series show a slightly

[3]The selected counties are Dalarna, Gävleborg, Uppsala and Jämtland. The development of inpatient care in these counties taken together follows that of the entire Sweden from 1987 onwards.



(a) Men (b) Women

Figure 2.6: Share of population receiving inpatient care during the year, 1968–2008
Source: National Board of Health and Welfare

increasing trend in the utilization of inpatient care up to the mid 1980s, followed by a substantial decrease to levels similar to, or even lower than, those in the late 1960s. The decline was much more pronounced for men than for women in the two oldest age groups, which has lead to a convergence in the utilization of inpatient care across genders. This suggests an improvement in the health of men relative to women in these age groups since the mid 1980s, which is consistent with the results from the self-reported health measures and the development of the mortality rates reported previously.

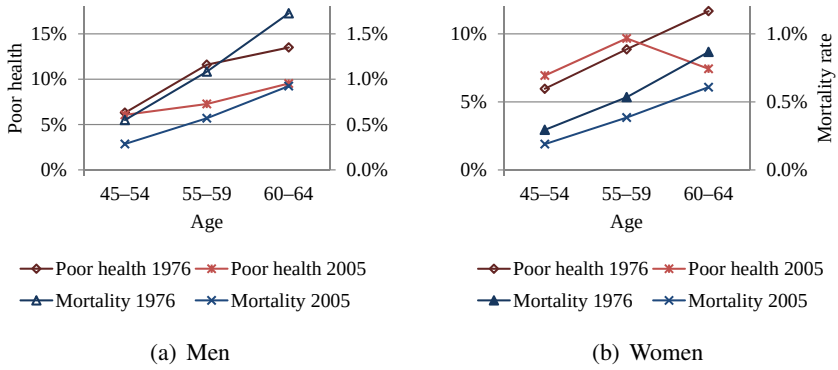
3.4. Conclusions about the Development of Population Health

The three main measures of population health give an ambiguous picture of the development of the general health status in Sweden over the past decades. The development of the mortality rate suggests a marked health improvement over time for both men and women. The development of the self-reported health indicators does, however, not confirm this result. Only men aged 60–64 show an invariant or positive development for all self-reported health indicators over time. Also the development of the utilization of inpatient care does not indicate a clear-cut health improvement over time.

Two main conclusions emerge from our analysis. First, health seems to have developed more adversely for females than for males since the mid 1970s. Female mortality decreased most rapidly up until the 1970s, while male mortality decreased substantially from 1980 onwards. For the self-reported health measures, we only have information from 1976 onwards. Since then, these measures show a less advantageous development of the health of females than

the health of males. The development of the utilization of inpatient care confirms this pattern. Since the early 1980s, the utilization of inpatient care in the two oldest age groups has decreased more for men than for women. The development in the youngest age group, aged 45–54, however, has been similar for men and women.

The second main conclusion is that the health of younger age groups seems to have deteriorated compared to older ones. The self-reported health indicators suggested a worsened health status over time for both men and women in the youngest age group, aged 45–54, whereas the health status of the oldest age group, aged 60–64, improved. This is supported by the development of the mortality rate, where the decline was steeper at age 65 than at ages 55 and 60. It is not apparent, however, in the development of inpatient care utilization.



(a) Men (b) Women

Figure 2.7: Share of population in poor health and the mortality rate, 1976 and 2005
Source: Statistics Sweden

The two main conclusions are highlighted in Figure 2.7, which shows the share of individuals with a self-reported poor health and the mortality rate in the three age groups in 1976 and 2005. Figure 2.7(a) shows that the health of men in the youngest age group, aged 45–54, has hardly improved over the 30 year period, while the health of men in ages 55–59 and 60–64 has improved substantially. For women, the gains in mortality are much smaller than for men over the period, and the share of women with a self-reported poor health has even increased for the younger age groups, aged 45–54 and 55–59. For older women, aged 60–64, on the other hand, the share in poor health has decreased substantially over the period.

4. Disability Insurance Utilization and the Relation to Labor Market Outcomes

4.1. The Development of Disability Insurance Recipiency

Figure 2.8 shows the prevalence of disability insurance recipiency by the end of the year from 1962 to 2009 for men and women, respectively.[4] The prevalence is defined as the share of the population in an age group that receives full or partial disability benefits in a given year. Figure 2.8(a) reveals a clear upward trend of disability insurance prevalence for men in all age groups until the early 1990s. The prevalence for men in the oldest age group, aged 60–64, increased from around 10 percent in 1962 to above 35 percent in 1995. The increase from about 5 percent in 1962 to about 18 percent in 1995 for men in the age group 55–59 is also notable. After the mid 1990s, there is a clear trend break for men in the oldest age group, with a decrease in the disability insurance prevalence from above 35 percent to about 20 percent. A smaller decrease can be seen in the age group 55–59. The prevalence of disability insurance recipiency for men aged 45–54, however, continued to increase until the late 2000s.

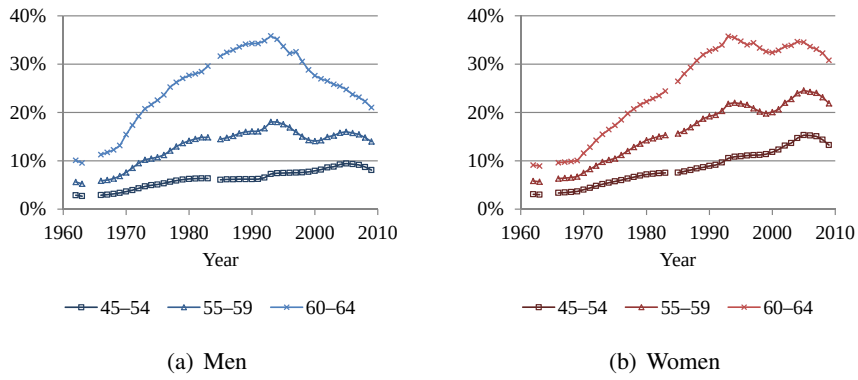


Figure 2.8: Disability insurance prevalence by gender and age group, 1962–2009
Source: Swedish Social Insurance Agency

Figure 2.8(b) shows a similar increase in the prevalence of disability insurance recipiency for women until the early 1990s. The increase was most rapid for women aged 60–64, for whom the prevalence of disability insurance recipiency rose to the same level as that for men. For the two younger age

[4]The prevalence of DI recipiency is measured in January from 1963 to 1984 and in December from 1985 onwards. In our analysis, we let the January figures from 1963 to 1984 represent DI recipiency at the end of the previous year. For example, DI recipiency in January 1963 is presented as DI recipiency in 1962.

groups, however, DI recipiency increased to even higher levels than for men of the same age. A similar trend break as that for men can be seen also for women in the early 1990s, but the development since then has been far less favorable than that for men. The prevalence of DI recipiency for women in the oldest age group has remained above 30 percent, and the prevalence in the younger age groups continued to rise until the mid or late 2000s. Compared to the early 1960s, the level of DI recipiency has more than tripled for all women.

Figure 2.9 shows the development of the incidence of disability insurance recipiency from 1971 to 2009. The incidence is defined as the share of the population at risk that starts to receive full or partial disability benefits in a given year, whereas the population at risk is defined as all individuals in the age group that are not already receiving disability benefits. Due to limited data availability, we redefined the youngest age group to age 50–54 rather than 45–54 as in the previous section.

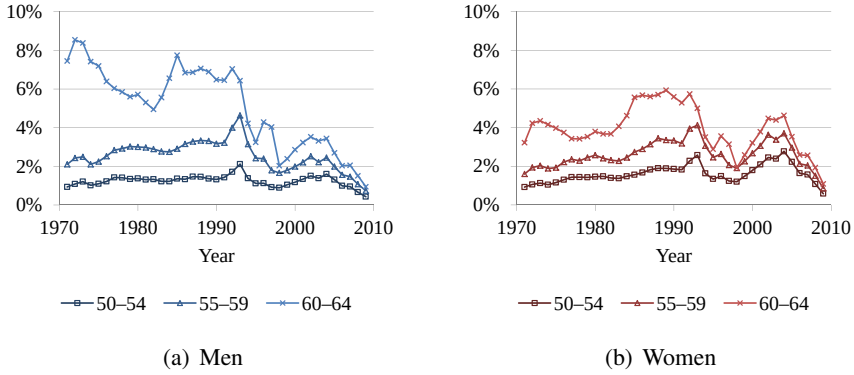
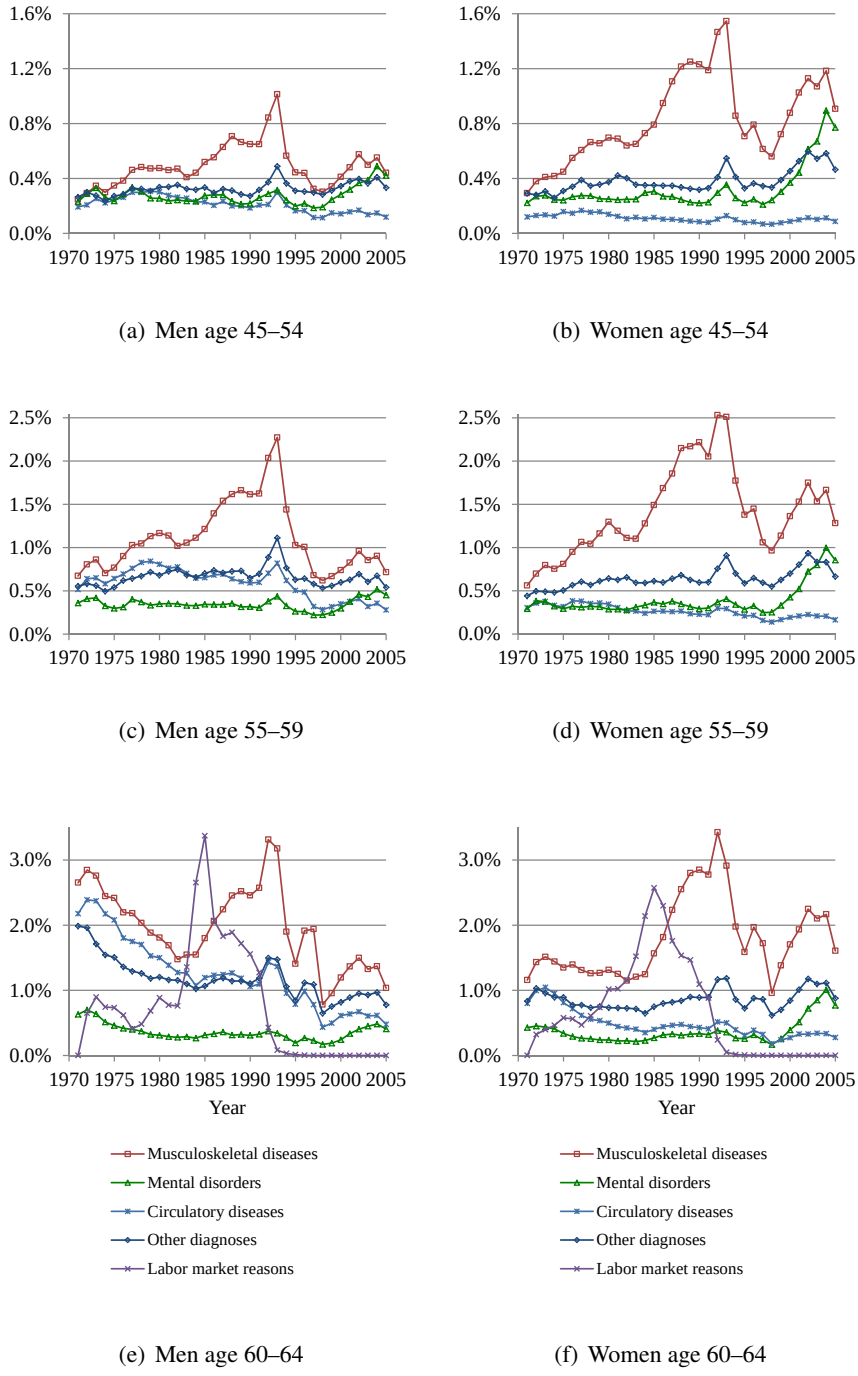


Figure 2.9: Disability insurance incidence by gender and age group, 1971–2009
Source: Swedish Social Insurance Agency

Figure 2.9 shows that the incidence of DI recipiency in the two youngest age groups increased in a similar manner for men and women until the early 1990s. In the oldest age group, however, the development is somewhat different across genders. The incidence for men aged 60–64 started at a historically high level in the beginning of the 1970s and decreased until the early 1980s, while the incidence for women aged 60–64 was relatively stable until the early 1980s. Thereafter, the incidence of DI recipiency in the oldest age group increased for both men and women until the early 1990s, although the level for men was still higher than that for women. From the early 1990s onwards, the incidence of DI recipiency has developed similarly across demographic groups. The incidence decreased markedly in all age groups during the 1990s. In the early 2000s, the incidence once more increased, in particular for women,



(a) Men age 45–54

(b) Women age 45–54

(c) Men age 55–59

(d) Women age 55–59

(e) Men age 60–64

(f) Women age 60–64

Figure 2.10: Disability insurance incidence by diagnosis, 1971–2005
Source: Swedish Social Insurance Agency

but then declined to historically low levels in all groups until 2009.

The recent decline in the incidence of disability insurance recipiency is remarkably large. Since the upward trend was broken in 2005, the incidence has steadily declined to very low levels in all age groups. As described in Section 2, new regulation came into place in July 2008 introducing stricter eligibility criteria for disability benefits. It is apparent, however, that the decline begun well before then. The decline in incidence can also be seen in the development of the prevalence of disability insurance recipiency in Figure 2.8, although the levels are still high. If the incidence remains at these low levels, we are likely to see a steep decline in the prevalence of DI recipiency over the coming years.

Figure 2.10 shows the development of disability insurance incidence by diagnosis. We present the three most common groups of diagnoses for disability benefits award: circulatory diseases, musculoskeletal diseases and mental disorders. We also include a category for all other diagnoses. For the group aged 60–64 we add a category for labor market reasons for the part of the period when this was a sufficient criterion for DI eligibility in this age group, as described in Section 2.

The figure shows that musculoskeletal diseases have been the most common diagnosis for disability benefits award in all demographic groups throughout the period. The large increases in the incidence of DI recipiency until the early 1990s and the subsequent sharp declines were primarily attributed to musculoskeletal diagnoses. The figure also reveals that mental disorders have become increasingly important in recent years in all demographic groups, in particular for the youngest age group and for women. Circulatory diagnoses have lost importance over time in all demographic groups. Even though musculoskeletal diseases and mental disorders explain a large share of the increase in incidence during the early 2000s, also the series capturing other diagnoses show a similar increase.

For the oldest age group, the pure labor market reasons, introduced in 1972, did not become important until the 1980s. In the mid 1980s, labor market reasons were the most common reason for granting disability benefits to 60–64 year olds. Thereafter, the importance of labor market reasons declined substantially while the importance of musculoskeletal diagnoses increased.

4.2. The Development of Labor Market Outcomes

To describe the development of some central labor market outcomes we use data from the Labor Force Surveys, collected by Statistics Sweden. We present the development from 1963 to 2007 for men and women in the age groups 45–54, 55–59 and 60–64. Figure 2.11 shows the development of the employment rate. Figure 2.11(a) reveals a similar pattern across age groups for the devel-

opment of male employment over time. Male employment decreased from the early 1960s until the early 1990s, and then experienced a pronounced dip during the recession in Sweden in the early 1990s. From the late 1990s onwards, male employment again increased. The pattern is most pronounced for the age group 60–64, where the employment rate decreased from above 80 percent in 1963 to around 50 percent in the mid 1990s, and then increased to about 65 percent in 2009. Changes in employment in the age group 45–54 are much smaller. Employment in this age group remained above 90 percent until the economic crisis in the early 1990s, and then shifted down to a level just below 90 percent.

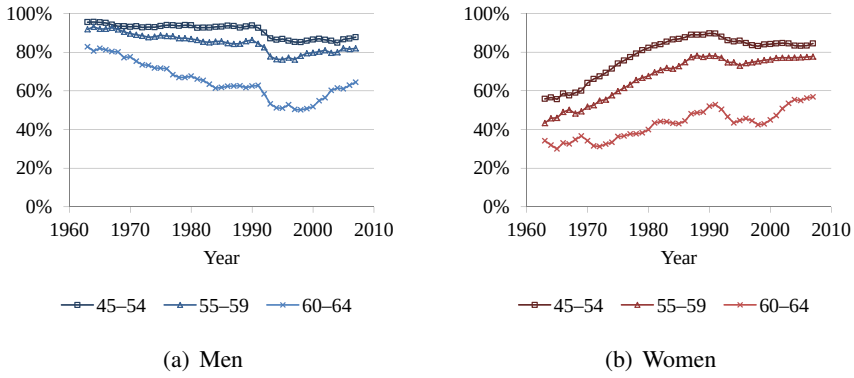
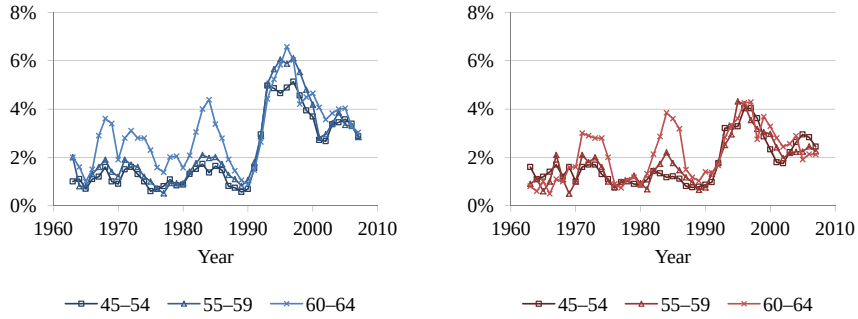


Figure 2.11: Employment rates by gender and age group, 1963–2007
Source: Swedish Labor Force Survey, Statistics Sweden

Figure 2.11(b) shows a different development of the female employment rate. Until the 1990s, employment increased substantially for women in all age groups. The increase was most rapid for the two youngest age groups, where employment increased from 55 to 90 percent in the age group 45–54 and from 40 to 80 percent in the age group 55–59. By the time of the economic crisis in the beginning of the 1990s, the stable increase in employment ceased and was followed by a slight decrease in all age groups. In recent years, however, there has been a marked increase in the employment rate of women aged 60–64. Between 2000 and 2007, employment increased by 15 percentage points to about 58 percent in this group. Also the female employment rate in the age groups 45–54 and 55–59 has increased since the late 1990s.

Figure 2.12 shows the development of the unemployment rate. As opposed to the employment rate, the unemployment rate has developed in a similar manner for men and women. There are two notable properties of the development. First, unemployment in the oldest age group seems to have been more sensitive to business cycle movements than other age groups before 1990. Af-

ter 1990, the series for different age groups correspond remarkably well. Second, the unemployment rate seems to have increased to a permanently higher level in all age groups after the recession in the beginning of the 1990s.



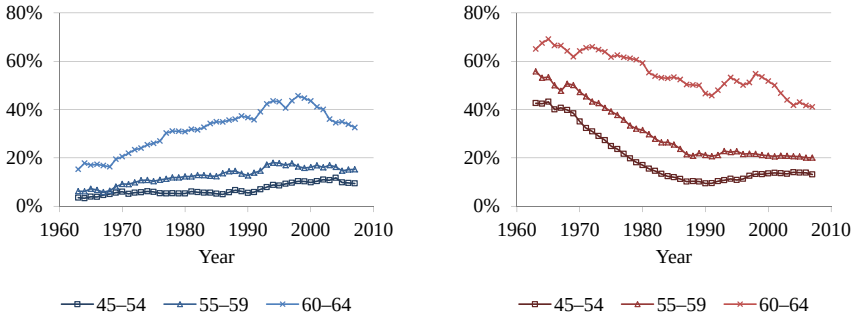
(a) Men (b) Women

Figure 2.12: Unemployment rates by gender and age group, 1963–2007
Source: Swedish Labor Force Survey, Statistics Sweden

Figure 2.13 shows the share of the population not participating in the labor force. Figure 2.13(a) reveals an increasing trend in the share of the male population out of the labor force in all age groups. The steepest increase occurred for the oldest age group, aged 60–64, from about 15 percent in 1963 to about 45 percent in 2000. Thereafter, however, the trend reversed and non-labor force participation for men aged 60–64 decreased to a level just above 30 percent in 2009. The increase in non-labor force participation in the two younger age groups has been much more modest and the series have stabilized rather than declined in the 2000s.

Figure 2.13(b) reveals a very different development of the non-labor force participation rate for women compared to men. There has been a decreasing trend in all age groups and the decrease has been most pronounced in the two younger groups. Non-labor force participation decreased from 45 to 10 percent for the age group 45–54 and from 55 to 20 percent for the age group 55–59 between 1963 and 1990. Thereafter, the level has remained stable for the age group 55–59 and increased slightly for the age group 45–54. The non-labor force participation rate for the oldest age group, aged 60–64, decreased less rapidly, from 65 to 40 percent over the full period.

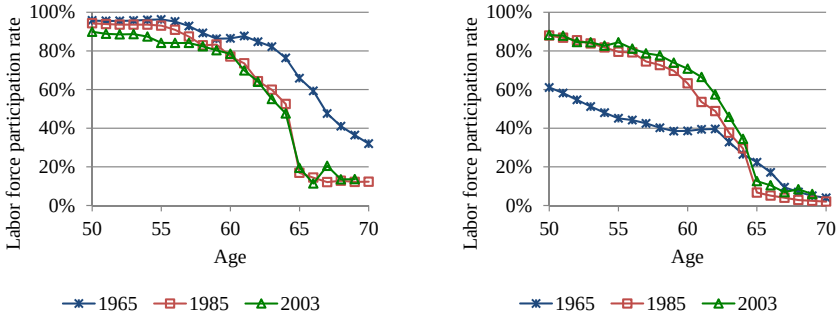
Figure 2.14 shows the labor force participation rate in one-year age groups in 1965, 1985 and 2003 for men and women, respectively. Figure 2.14(a) reveals a marked decrease in the labor force participation of men above the age of 58 between 1965 and 1985, but no visible change between 1985 and 2003. In the ages below 58, however, there was a small but visible decrease in the



(a) Men (b) Women

Figure 2.13: Non-labor force participation rates by gender and age group, 1963–2007
Source: Swedish Labor Force Survey, Statistics Sweden

labor force participation both between 1965 and 1985 and between 1985 and 2003. Figure 2.14(b) reveals a large increase in female labor force participation at ages below 65 between 1965 and 1985, and a smaller increase between 1985 and 2003. Beyond the age of 65, there was a decrease in labor force participation between 1965 and 1985 that can be explained by a decrease in the normal retirement age from 67 to 65 in 1976.

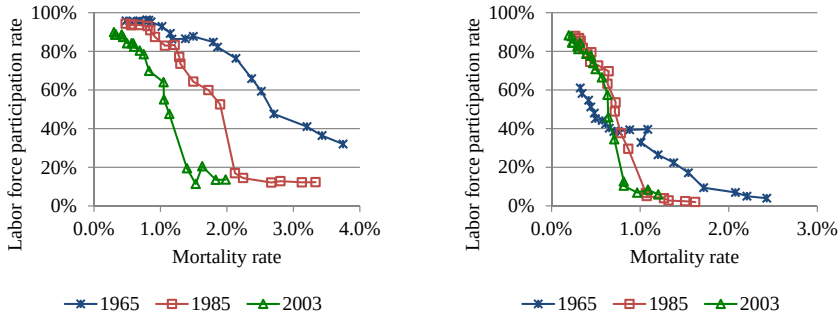


(a) Men (b) Women

Figure 2.14: Labor force participation by age and year
Source: Own calculations based on data from Statistics Sweden

In Section 3 we showed that the mortality rate has fallen quite dramatically in all age groups in recent decades. An increased life expectancy can affect the time an individual spends in the labor force, the time an individual spends as a retiree, or both. To report the changes in labor force participation over time

without correcting for changes in mortality, as we have done so far, implicitly implies that all gains in life expectancy are taken out in time as a retiree. The other extreme would be that the time as a retiree is held constant and the increased life expectancy only affects the time spent in the labor force. One way of investigating this alternative is to calculate labor force participation rates at constant mortality risks.



(a) Men (b) Women

Figure 2.15: Labor force participation by mortality rate and year
Source: Own calculations based on data from Statistics Sweden

The results from this exercise are shown in Figure 2.15, which shows the labor force participation rate at a given mortality rate. As was apparent in Figure 2.14, labor force participation for men decreased even when not taking the decrease in mortality into account. Figure 2.15 hence reveals an even larger decrease in labor force participation rates for men. In contrast to the results in Figure 2.14, however, there is also a large decrease between 1985 and 2003, a period during which the mortality rate of men decreased significantly. Interestingly, the increased labor force participation for females between 1965 and 1985 that was shown in Figure 2.14 reverses in Figure 2.15, since mortality improves more than labor force participation increases. Only for very low mortality rates labor force participation still increased between 1965 and 1985. For women at higher mortality rates, labor force participation decreased substantially between 1965 and 1985, and continued to decrease until 2003.

4.3. Pathways to Retirement

Figure 2.16 shows the development of the share of the population receiving disability benefits, the share of non-employed and the share not participating in the labor force from 1963 to 2009. Figure 2.16(c) and 2.16(e) show a large increase in non-employment and non-labor force participation of elderly men

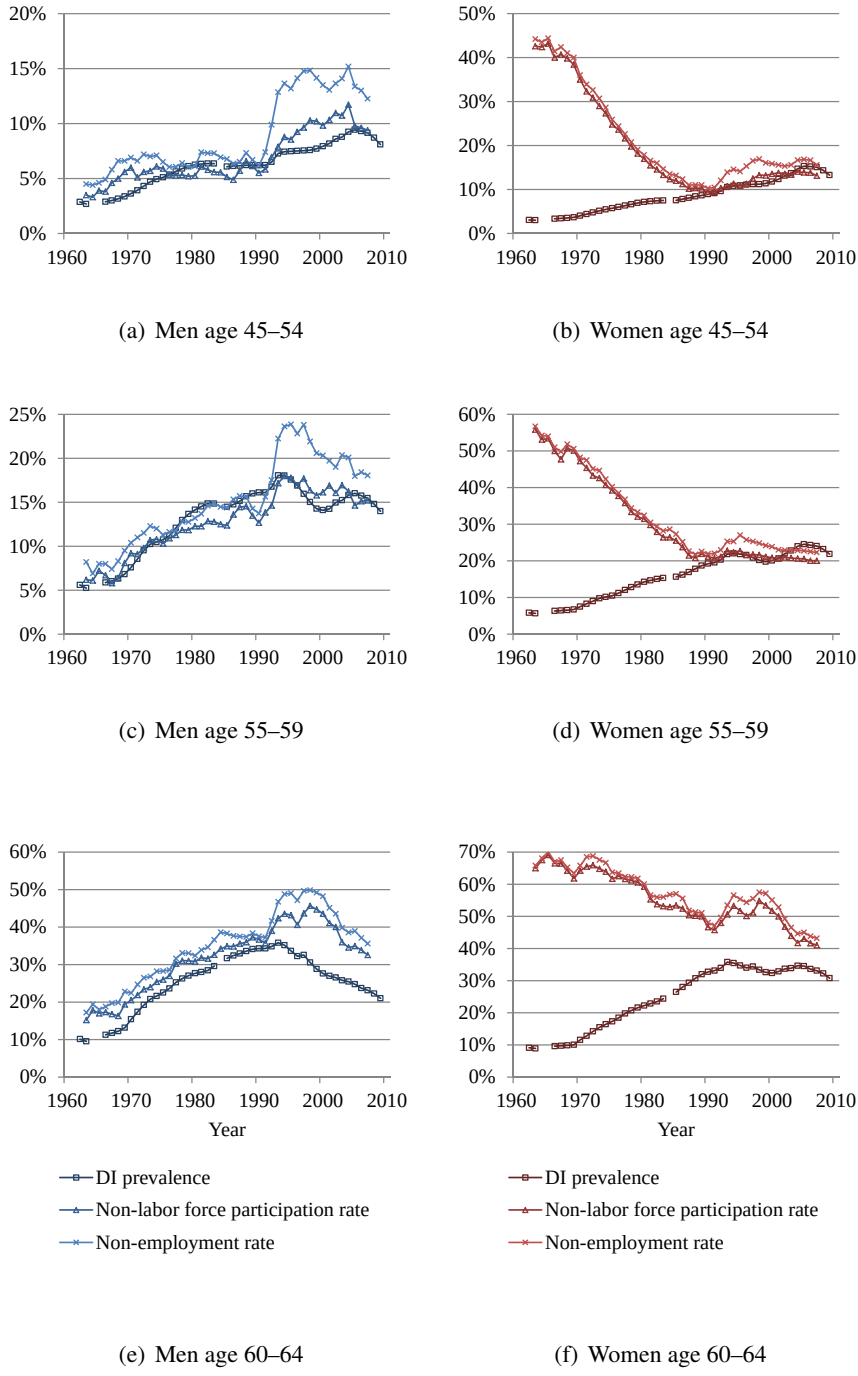


Figure 2.16: DI prevalence, non-labor force participation and non-employment
Sources: Swedish Social Insurance Agency and Statistics Sweden

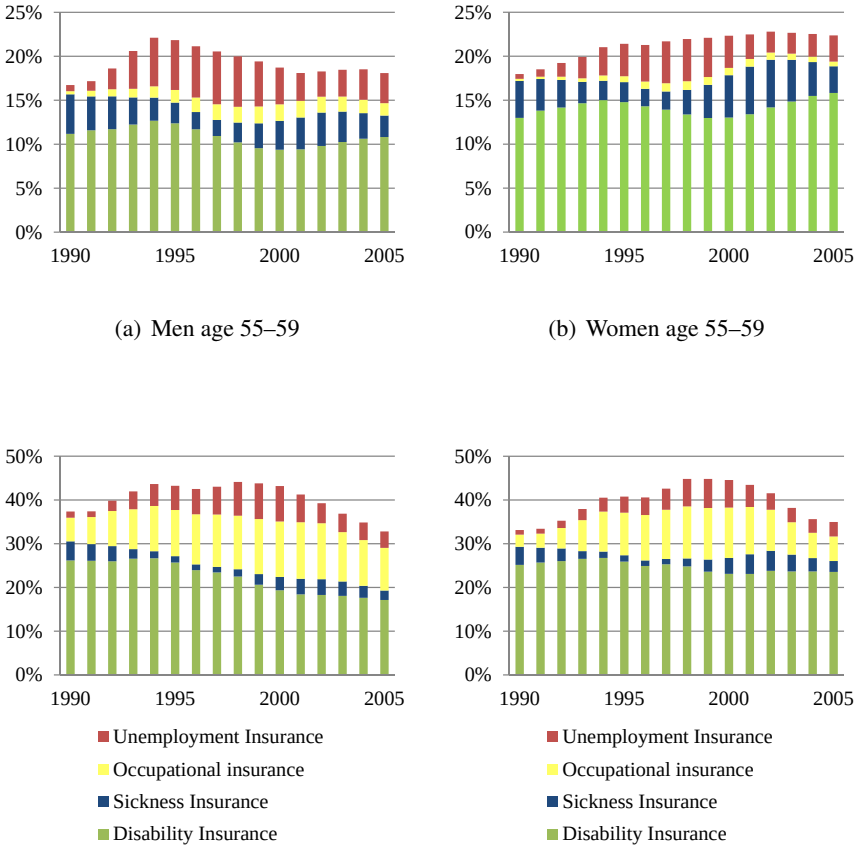
between 1970 and 1988, along with a corresponding increase in disability insurance recipiency. In the age group 55–59, non-employment increased by 5.2 percent between 1970 and 1988, non-labor force participation by 5.3 percent and disability insurance recipiency by 8.1 percent. For men aged 60–64 the corresponding figures were 15.0, 15.5 and 15.2. The disability insurance hence seems to have been the dominating pathway to retirement for older males until 1990.

For men aged 45–54, the same close correspondence between non-employment and non-labor force participation on the one hand and disability insurance recipiency on the other is not apparent. Non-employment and non-labor force participation increased in the end of the 1960s and did not increase again until 1990, while disability insurance recipiency was gradually increasing.

The right-hand panels in Figure 2.16 show a very different pattern for females until 1990. The large gap between disability insurance recipiency and the non-labor force participation rate consists of the diminishing fraction of homemakers. The fact that the gap closes earlier for younger age groups tells us that this development is primarily a cohort effect. Since the close link between non-employment, non-labor force participation and disability insurance recipiency for men was broken in 1990, the development for women has been more similar to that for men. In all age and gender groups, a gap emerged between disability insurance recipiency and non-employment from 1990 onwards. In the oldest age group, this gap primarily consisted of increased non-labor force participation that was not due to increased disability insurance recipiency.

To study the background to this development, we use the annual income statistics from tax returns. Figure 2.17 shows the share of men and women aged 55–59 and 60–64 with one of the four main income security programs in Sweden as the main income source. These include the disability insurance, the sickness insurance, the unemployment insurance and occupational insurance. An individual is classified as receiving his or her main income from a certain program if the benefits from the program account for 50 percent or more of the total yearly income.

The upper panels in Figure 2.17 show that the gap between disability insurance recipiency and non-employment in the age group 55–59 has been mainly accounted for by the unemployment insurance. Also the sickness insurance has been a main income source for a substantial share of the population in this age group, and the relative importance of the sickness and the unemployment insurance has shifted over time. The lower panels in Figure 2.17 show a different pattern for the age group 60–64. A large share of the gap between disability insurance recipiency and non-labor force participation in this age group after 1990 has been accounted for by benefits from occupational insurance schemes.



(a) Men age 55–59

(b) Women age 55–59

(c) Men age 60–64

(d) Women age 60–64

Figure 2.17: Main income source, 1990–2005
Source: Annual income statistics, Statistics Sweden

4.4. Conclusions about the Disability Insurance and Labor Market Outcomes

The results in this section showed some general patterns. The development of disability insurance recipiency and labor market outcomes can be divided into two main eras; before and after 1990. From the early 1960s until 1990, the share of disability insurance recipients increased in all demographic groups. For males, the increase was closely accompanied by an equal increase in non-labor force participation and non-employment. For females, the correspondence between disability insurance recipiency and labor market outcomes was weak, but the gap was closing over time as female labor force participation increased.

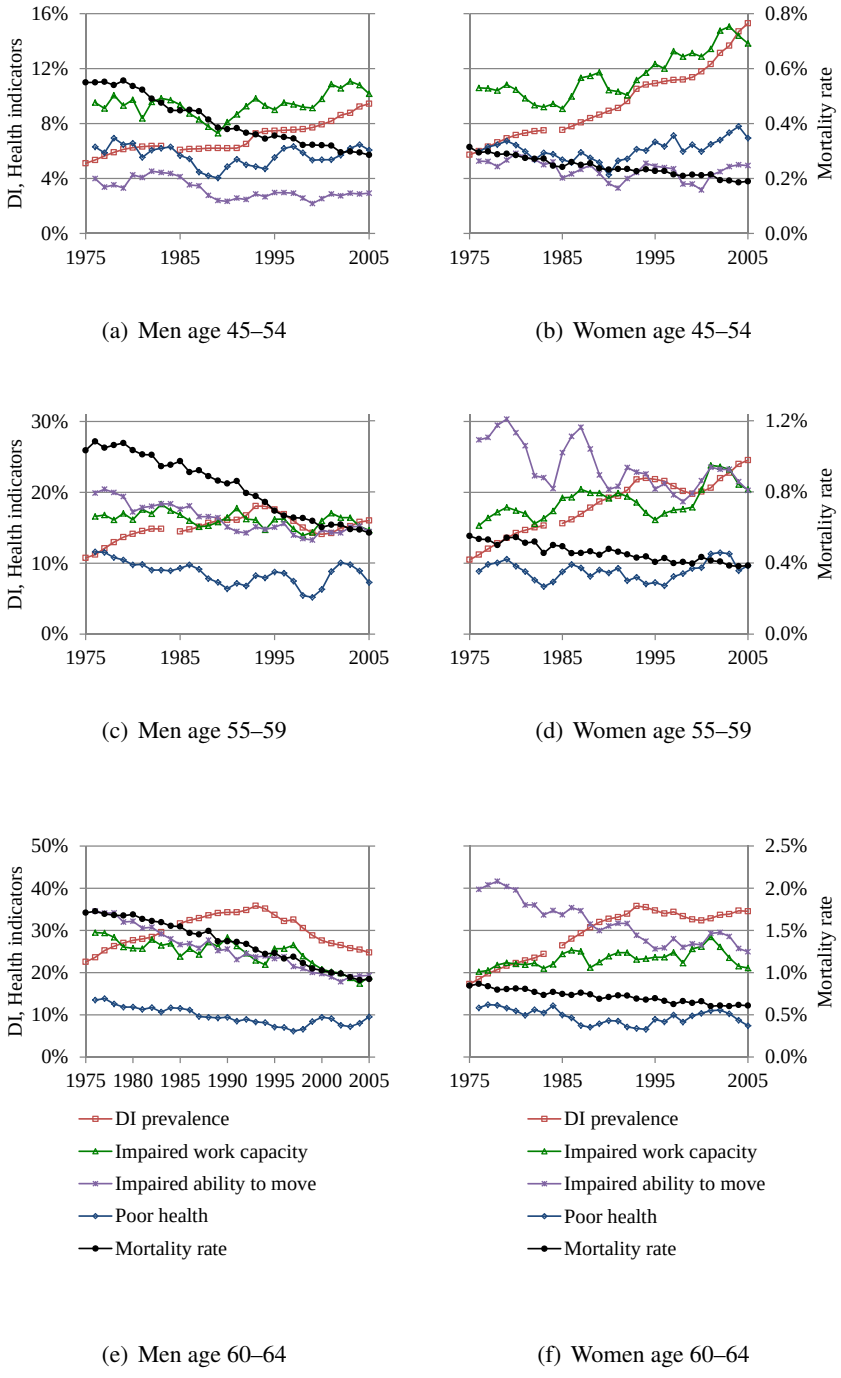
After 1990, the development of disability insurance recipiency was different across demographic groups. While disability insurance recipiency decreased for men aged 55–59 and 60–64, it continued to increase for men and women aged 45–54 and women aged 55–59, and only stabilized for women aged 60–64. From the early 1990s, a gap emerged between disability insurance recipiency and non-employment. A closer study showed that the disability insurance program was losing importance as a pathway to permanent exit from the labor force. In the age group 55–59, the unemployment and sickness insurance programs became more important while in the age group 60–64, the role of occupational insurances increased substantially after 1990.

Finally, we should note the sharp decrease in the incidence of disability insurance utilization in recent years. The incidence reached a historically low level of below 1 percent of the risk population in all demographic groups in 2009. If these exceptionally low levels of incidence continue, it will lead to a sharp decrease in the prevalence of disability insurance utilization in the coming years.

5. Population Health and Disability Insurance

5.1. Disability Insurance Prevalence and Population Health

Figure 2.18 presents the development of the prevalence of disability insurance recipiency along with the mortality rate and the share of the population with a self-reported impaired work capacity, the share of the population with an impaired ability to move and the share of the population with a self-assessed poor health from the Survey on Living Conditions. The most apparent result from this figure is the lack of correlation between the mortality rate and the prevalence of DI recipiency in all of the demographic groups. The increase in the prevalence of DI recipiency for all groups until the mid 1990s instead coincided with decreasing mortality rates.



(a) Men age 45–54

(b) Women age 45–54

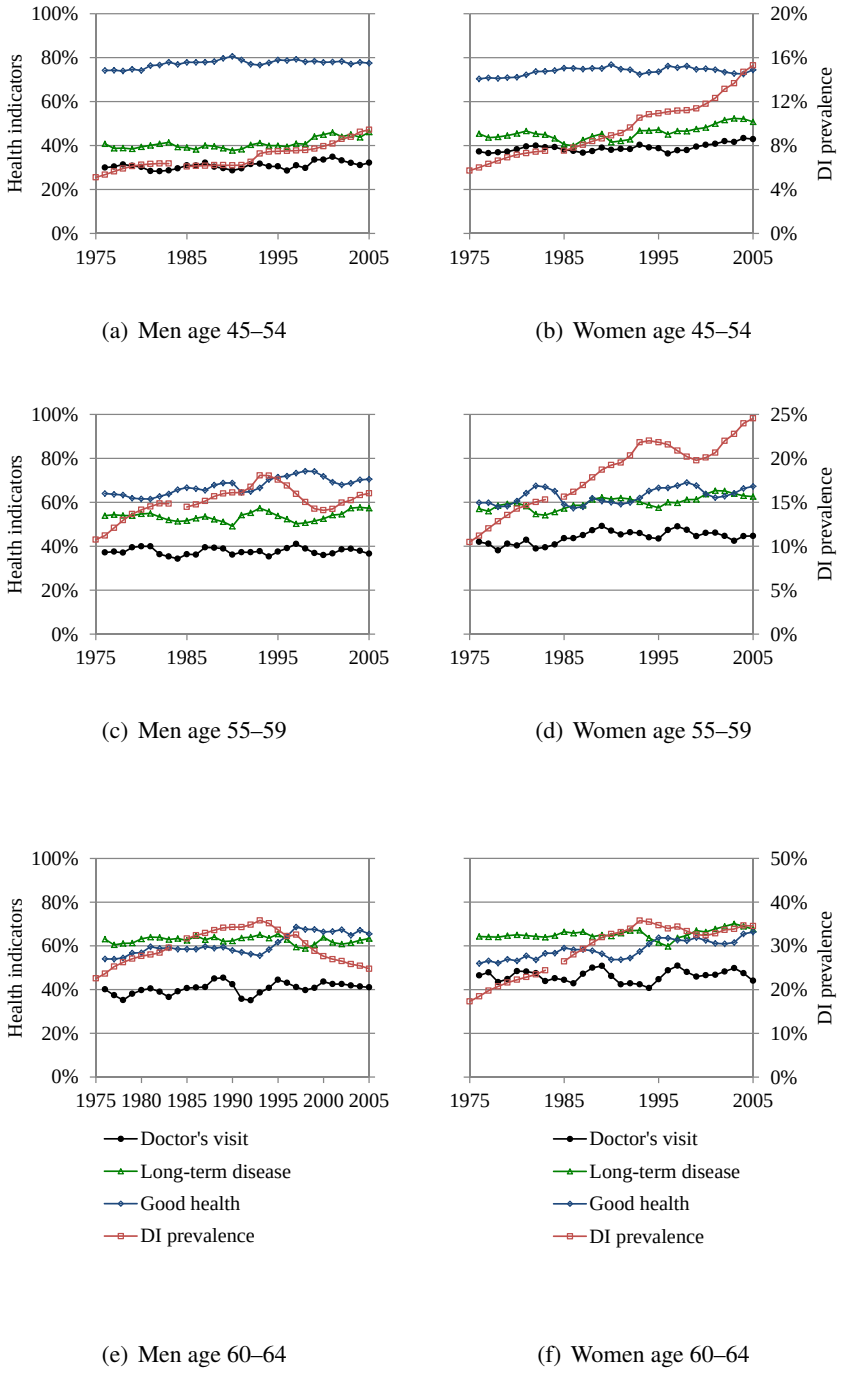
(c) Men age 55–59

(d) Women age 55–59

(e) Men age 60–64

(f) Women age 60–64

Figure 2.18: Disability insurance prevalence, mortality and health indicators
Sources: Swedish Social Insurance Agency and Statistics Sweden



(a) Men age 45–54

(b) Women age 45–54

(c) Men age 55–59

(d) Women age 55–59

(e) Men age 60–64

(f) Women age 60–64

Figure 2.19: Disability insurance prevalence, mortality and health indicators
Source: Swedish Social Insurance Agency and Statistics Sweden

Turning to the self-reported health indicators in Figure 2.18, there is a resemblance between the development of impaired work capacity and disability insurance prevalence. The two series shared a similar pattern from the mid 1980s onwards for the two younger age groups, and from the early 1990s onwards for the oldest age group. Both the prevalence of DI recipiency and the share of the population with an impaired work capacity increased more rapidly for women than for men in the two youngest age groups. In addition, the sharp drop in DI recipiency among men aged 60–64 since the mid 1990s has coincided with a decrease in the share of the population with an impaired work capacity. The same correspondence is not present between the development of the share of the population in poor health and disability insurance recipiency or the share of the population with an impaired ability to move and disability insurance recipiency.

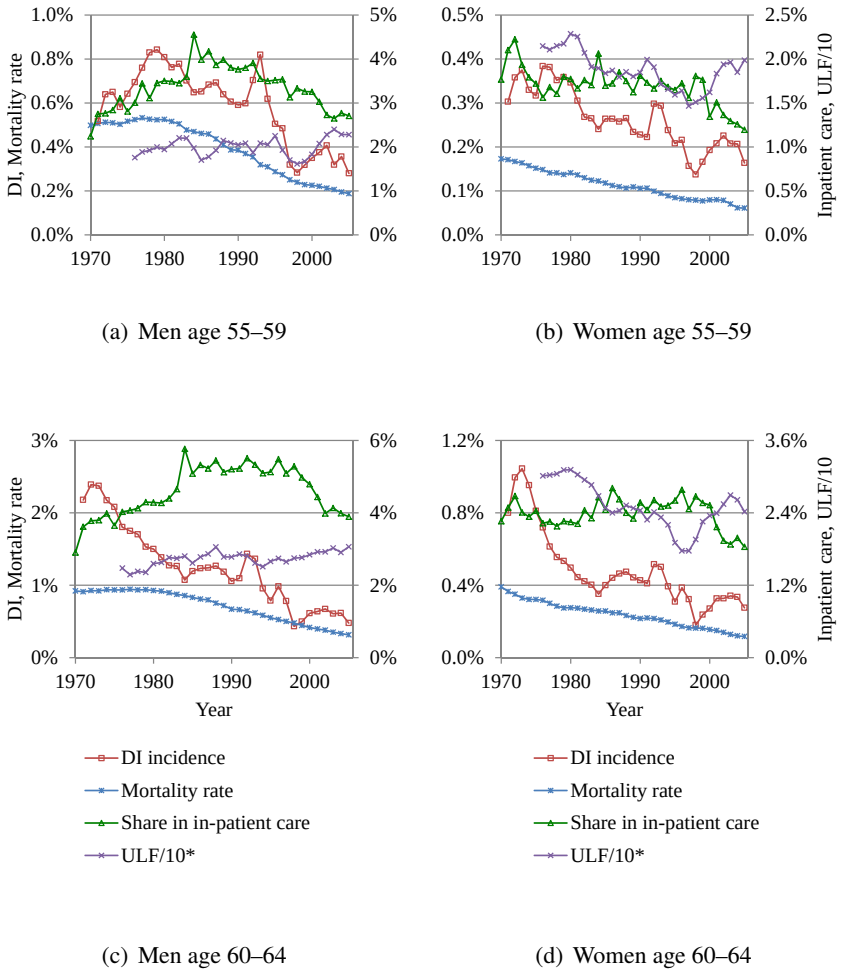
Figure 2.19 presents the development of disability insurance prevalence along with three additional health indicators: the share of the population who visited a doctor during the last three months, the share of the population with a long-term disease and the share of the population with a self-assessed good health. From the development of these indicators it is not possible to reject that disability insurance recipiency and population health are unrelated. The indicators did, however, develop more adversely for women and for younger age groups, which was also the case for disability insurance recipiency.

5.2. Disability Insurance Incidence and Population Health

To further explore the relation between population health and disability insurance recipiency, Figure 2.20 through 2.22 show the development of three diagnosis-specific health measures: (i) the share of the population receiving inpatient care for a specific diagnosis, (ii) the corresponding self-reported prevalence of a specific disease from the Survey on Living Conditions (ULF) and (iii) the diagnosis-specific mortality rate; along with diagnosis-specific incidence of DI recipiency. The figures present this information for the three most common diagnoses for DI eligibility: circulatory diseases, musculoskeletal diseases and mental disorders. We show the development from 1971 to 2005 for men and women in the age groups 55–59 and 60–64.

Figure 2.20 shows the development for circulatory diagnoses. There has been a decreasing importance of circulatory diagnoses as a ground for disability insurance recipiency in all demographic groups over time. This decrease has been shared with the diagnosis-specific mortality rate. Mortality decreased later for men than for women, and for the younger age group, aged 55–59, the decrease in DI incidence showed a similar pattern. For the older age group, however, the decrease in DI incidence occurred simultaneously across demo-

graphic groups, although mortality decreased later for men.



(a) Men age 55–59

(b) Women age 55–59

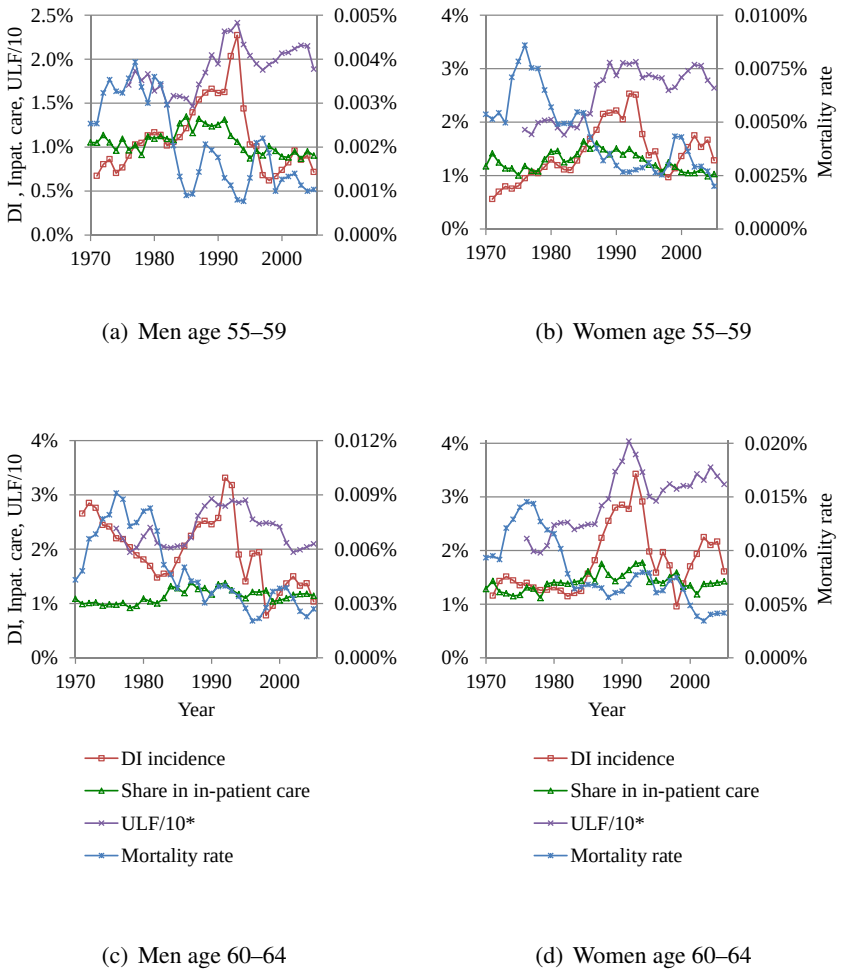
(c) Men age 60–64

(d) Women age 60–64

Figure 2.20: Circulatory diagnoses by gender and age group, 1971–2005
*ULF/10=Share of population reporting the particular disease in the Survey on Living Conditions (ULF), divided by ten

The decreasing trend is not as apparent for the utilization of inpatient care or the self-reported prevalence of circulatory diagnoses. For the older age group, aged 60–64, the development of the utilization of inpatient care showed no resemblance with the development of DI incidence. For the younger age group, aged 55–59, however, there was a similarity between the utilization of inpatient care and DI incidence from the mid 1980s onwards. For men, the development of the self-reported prevalence of circulatory diseases showed no

correspondence with DI incidence. For women, however, there was a resemblance between the two series.



(a) Men age 55–59

(b) Women age 55–59

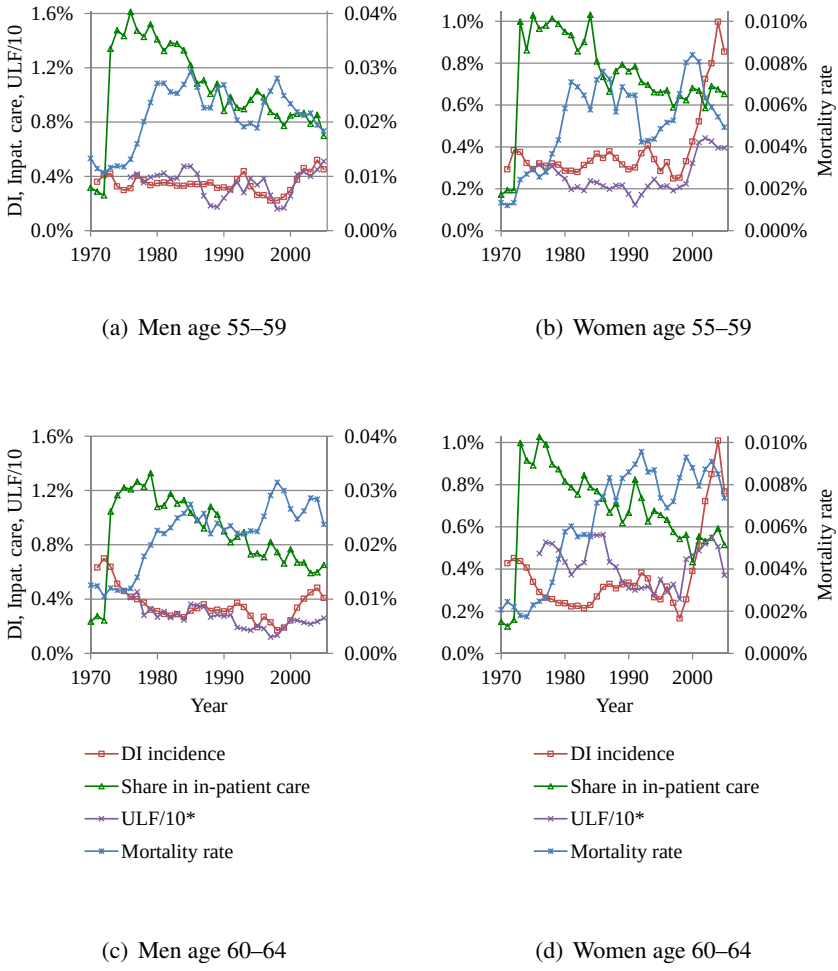
(c) Men age 60–64

(d) Women age 60–64

Figure 2.21: Musculoskeletal diagnoses by gender and age group, 1971–2005
*ULF/10=Share of population reporting the particular disease in the Survey on Living Conditions (ULF), divided by ten

Figure 2.21 shows the development for musculoskeletal diagnoses. There was a clear peak in the importance of musculoskeletal diagnoses for DI incidence from the mid 1980s until the mid 1990s in all demographic groups. Interestingly, this peak is also visible in the self-reported prevalence of musculoskeletal diseases. A smaller upturn in DI incidence can be seen in the early 2000s. Also this pattern is visible in the self-reported disease prevalence, ex-

cept for men aged 60–64. The development of inpatient care utilization due to musculoskeletal diagnoses has been relatively stable over time, although a slight increase is visible as DI incidence increases in the late 1980s. The mortality rate in musculoskeletal diagnoses has been highly volatile but has followed a decreasing trend that did not correspond to the development of the diagnosis-specific DI incidence.



(a) Men age 55–59

(b) Women age 55–59

(c) Men age 60–64

(d) Women age 60–64

Figure 2.22: Mental disorders by gender and age group, 1971–2005
*ULF/10=Share of population reporting the particular disease in the Survey on Living Conditions (ULF), divided by ten

Finally, Figure 2.22 shows the development for mental disorders. The importance of mental disorders for DI incidence has been invariant or falling in all

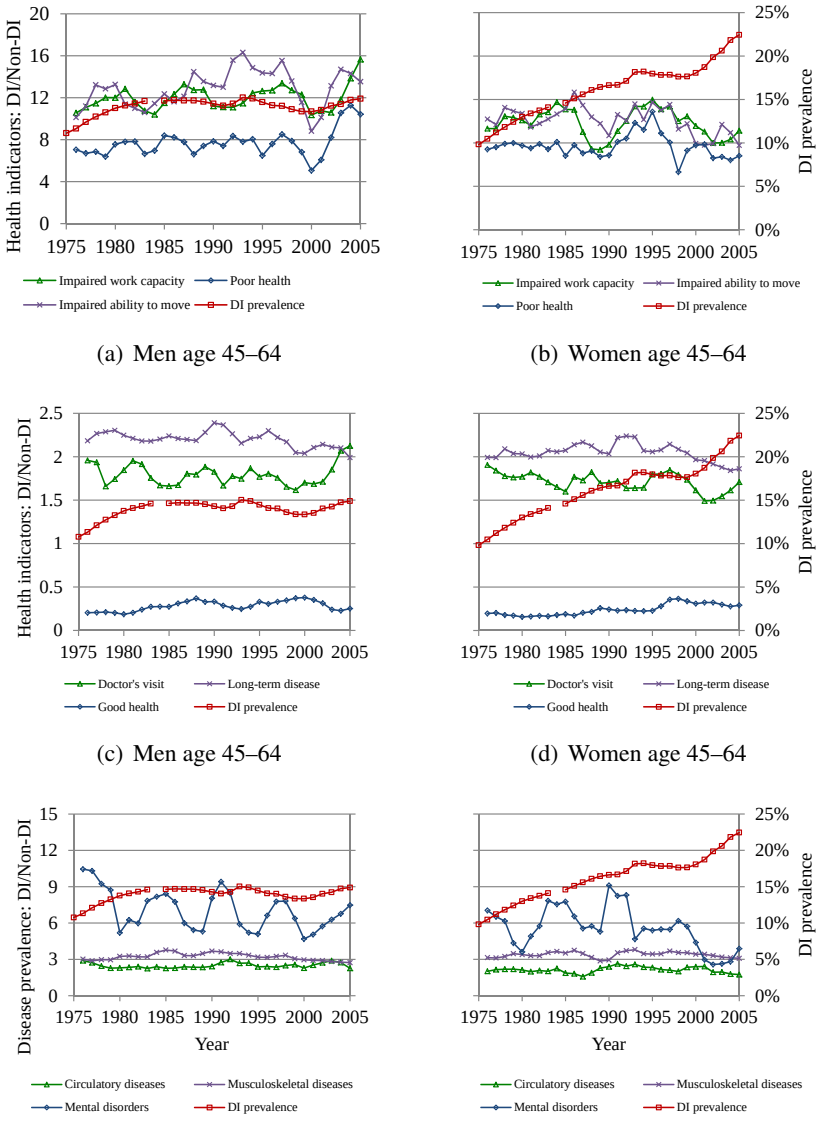
demographic groups until the late 1990s. This was followed by an enormous increase in the importance of mental disorders for granting disability benefits to women and a slight upturn for men. This increase is well reflected in the self-reported disease prevalence, except for men aged 60–64. Also the stable or falling trend in DI incidence before the 2000s corresponded to the development of self-reported disease prevalence, except for women aged 60–64.

The utilization of inpatient care due to mental disorders shows no correspondence with diagnosis-specific DI incidence. Inpatient care utilization increased dramatically in the early 1970s and was then consistently falling over time. Such a consistent development might be due to changing working procedures in the health care system with this type of patients rather than an underlying trend in health. The mortality rate also shows no correspondence with the development of DI incidence. It should be noted, however, that the mortality rates in musculoskeletal diseases and mental disorders are extremely low and hence very dependent on diagnosing patterns.

5.3. The Relative Health of Disability Insurance Recipients to that of Non-Recipients

An alternative way of examining the role of health for the development of disability insurance recipiency is to compare the health of disability insurance recipients with the health of non-recipients. With a fixed health threshold for DI recipiency, we would expect the health of recipients relative to non-recipients to remain constant even if the prevalence of DI recipiency changes. If changes in the prevalence of DI recipiency are instead induced by for example economic incentives, less stringent health requirements for eligibility or a change in demand for workers with health-induced work limitations, we would expect the health of DI recipients relative to the health of non-recipients to improve as recipiency increases.

We divided the respondents in the Survey on Living Conditions into disability insurance recipients and non-recipients. Figure 2.23 shows the development of the relative health of DI recipients relative to non-recipients for nine self-reported health indicators from the survey along with the prevalence of DI recipiency. For each indicator, the relative measure shows the prevalence of a particular condition in the DI population as a fraction of the prevalence in the non-DI population. For example, the upper left-hand panel in Figure 2.23 shows that in 1976, a ten times larger proportion of those receiving disability benefits reported an impaired work capacity as compared to non-DI recipients. Due to sample size restrictions, the results are presented for the entire age group 45–64 year olds and the presented series are three-year moving averages. The left-hand panels show the development for men and the right-hand



(a) Men age 45–64

(b) Women age 45–64

(c) Men age 45–64

(d) Women age 45–64

(e) Men age 45–64

(f) Women age 45–64

Figure 2.23: Relative health of DI recipients compared to non-DI recipients, 1975–2005

Source: Swedish Social Insurance Agency and Statistics Sweden

panels for women.

Figure 2.23 reveals much volatility, but no obvious trend, for the health of DI recipients relative to non-recipients before 1995. For men, most indicators remain constant also after 1995, although the share of men in poor health, the share with an impaired ability to move and the share who recently visited a doctor increase in the 2000s. This suggests, if anything, a worsened health of male DI recipients relative to non-recipients in recent years. For women, there is an opposite trend as that for men from 1995 to 2005. During this period, disability insurance recipiency for women increased substantially. The health indicators in Figure 2.23(b), showing the prevalence of impaired work capacity, impaired ability to move and poor health, show a downward trend since 1995. Also the prevalence of long-term disease in Figure 2.23(d) has been falling, while the share of women with a self-reported good health has increased, for DI recipients relative to non-recipients. The share of women who visited a doctor, reported in Figure 2.23(f), fell at least initially as DI caseloads took off. This suggests a relative improvement in the health of female DI recipients as compared to non-recipients between 1995 and 2005.

5.4. Conclusions about Health and the Disability Insurance

The analysis in this section showed that there is no relation between mortality and disability insurance recipiency in general in any age and gender group. This result does not, however, preclude that there is an underlying relation between population health and DI recipiency. As discussed in Section 3, mortality might be a poor measure of the aspects of population health that are important for DI recipiency.

For the self-reported health indicators, there are some results indicating that population health indeed is an important determinant of disability insurance utilization. In particular, the share of the population with an impaired work capacity developed in a similar manner as the prevalence of DI recipiency. The health in younger age groups has declined compared to older groups, and the health of women has declined compared to men. This pattern is consistent with the development of disability insurance recipiency. From the analysis of the diagnosis-specific health indicators, we also saw a corresponding pattern between disability insurance incidence and the development of self-reported diagnosis-specific diseases.

The relation between health and disability insurance seems to be strongest for the younger age groups. For the age group 60–64, a potential relation between health and DI recipiency can be seen only from 1990 onwards. For men in this age group, the drop in DI recipiency during the last decade coincided with a drop in the share of the population with an impaired work capacity. For

women in this age group, however, the development of disability insurance recipiency during the last decade seem to be less health related than for men. As DI recipiency increased, the health of disability insurance recipients relative to non-recipients improved. This implies that relatively healthier women than before started to receive disability benefits.

6. Changes in Disability Insurance Eligibility

The changes in the design of the disability insurance program in Sweden were described in Section 2 of this paper. Two major reforms in the history of the program were directed towards older workers only. First, special eligibility rules for older workers were introduced in 1970 and abolished in 1997. Second, a possibility of granting disability benefits based on pure labor market reasons for older workers was introduced in 1972 and abolished in 1991. The age limits were initially set to age 63–66, but was changed to 60–64 in 1974 for pure labor market reasons and in 1976 for the special eligibility rules.

The fact that the changes in eligibility affected a limited group only makes the implementation of these rules favorable from an evaluation point of view. It enables us to analyze the effect of changes in eligibility in a demographically defined group and use the younger age group aged 55–59 as an unaffected control group. In this section, we analyze the introduction of the special eligibility rules in 1970 and the pure labor market reasons in 1972 and the subsequent abolitions in 1997 and 1991. We analyze the impact of the eligibility reforms on disability insurance recipiency and study to what extent the reforms also affected labor market outcomes.

As described in Section 2, also another set of eligibility rules were in effect between 1970 and 1997. These rules affected all workers, and implied that long-term unemployed with functional limitations were made eligible for disability benefits after having been unemployed for 1–2 years. Since these rules affected all workers, there is no control group to use in order to distinguish the effect of the rules from general time trends. When studying the effect of the special eligibility rules that were in effect during the same period, we implicitly assume that the unemployment as an additional criterion for disability insurance eligibility affected the age groups 55–59 and 60–64 equally. If the unemployment criterion in fact was more important in the older age group than in the younger, the effect of these rules will be subsumed in the effect of the special eligibility rules for older workers.

As was also described in Section 2, eligibility for disability insurance recipiency was recently changed again. Since 1 July 2008, an individual's working capacity has to be permanently reduced in relation to the entire labor market in order to qualify for disability benefits. Since this change affected all

disability insurance applicants simultaneously, we cannot evaluate the impact in a control group setup. Figure 2.9 showed a substantial decrease in the incidence of disability insurance recipiency during the last years, and the decrease is particularly steep since 2008.

6.1. Program Eligibility and Disability Insurance Recipiency

The upper panels in Figure 2.24 show the development of the prevalence of disability insurance recipiency for men and women, respectively. Vertical lines mark the introduction of the special eligibility rules for elderly workers in 1970 and the abolition of these rules in 1997, as well as the introduction of labor market reasons in 1972 and the subsequent abolition in 1991. The lower panels in Figure 2.24 show the differences in disability insurance recipiency between the group aged 60–64 and the younger group aged 55–59. Figure 2.25 presents similar panels for the incidence of disability insurance recipiency, i.e., the admitted disability insurance recipients as a share of the risk population in each age group.

The upper panels in Figure 2.24 show a clear increase in the growth rate of DI recipiency after the 1970 reform. The increase in the growth rate is particularly large for the oldest age group. The lower panels show that the difference in DI prevalence between age groups 60–64 and 55–59 is fairly constant before 1970, at least for women, but increases rapidly after 1970. This indicates that there was an effect of the special eligibility rules for older workers that were introduced in 1970. Unfortunately, we do not have data on the incidence of disability insurance recipiency in these age groups before 1971. The number of entrants into DI in all ages, however, almost doubled from around 23 000 in 1968 to around 44 000 in 1970.

The next reform is the introduction of pure labor market reasons for older workers in 1972. It is not possible to perceive any effect of this reform on the growth rate of DI prevalence in Figure 2.24. From Figure 2.10(e) and 2.10(f), however, we know that pure labor market reasons accounted for an entry rate of around 1 percent of the risk population from its introduction until 1983. The lower panels in Figure 2.25 show the difference in the incidence of disability insurance recipiency between the age group 60–64 and 55–59. The difference is slightly larger in 1972 and 1973 than in 1971, which might indicate a small immediate effect of the 1972 reform on total DI entry rates.

As discussed in Section 4, there was a clear trend break in disability insurance recipiency in the older age groups in the early 1990s. This coincided with the abolishment of pure labor market reasons in 1991 and the rehabilitation reform in 1992. It also coincided, however, with a deep recession in Sweden. The upper panels of Figure 2.24 show this trend break in the prevalence of dis-

ability insurance recipiency in age groups 55–59 and 60–64. The increase in 1992 and 1993 is due to the fact that a large number of recipients of sickness benefits were transferred to the disability insurance program as a consequence of a rehabilitation reform affecting the work of the Social Insurance Agencies.

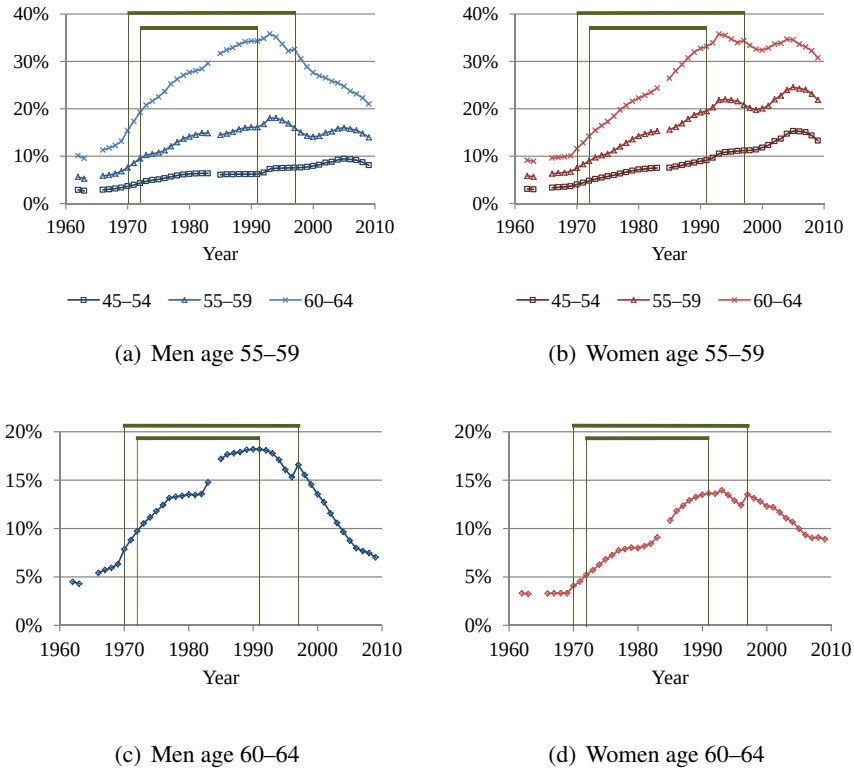
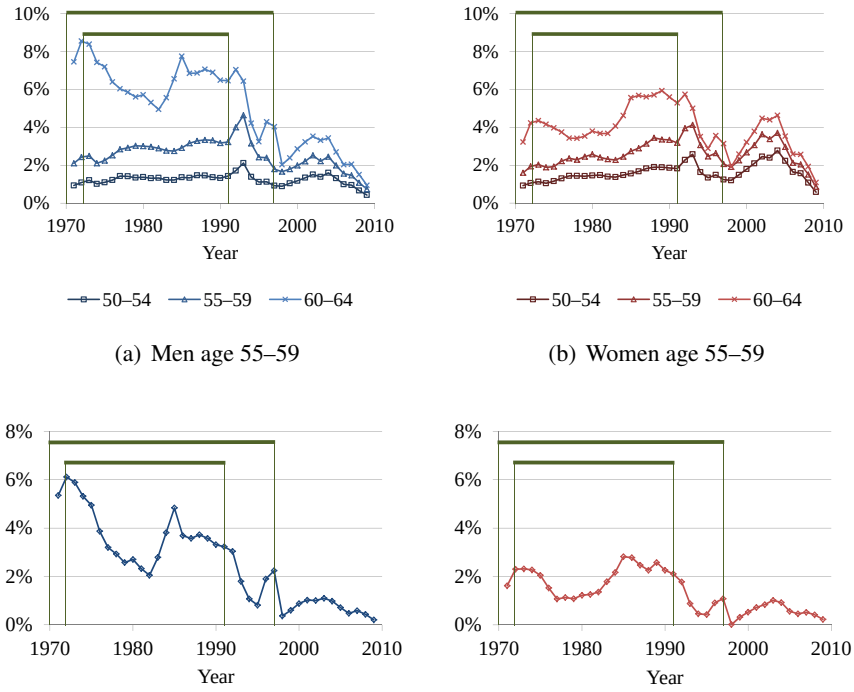


Figure 2.24: Prevalence of disability insurance utilization and the timing of DI reforms
Source: Swedish Social Insurance Agency

The lower panels in Figure 2.25 show that the difference in disability insurance entry rates between age groups 60–64 and 55–59 was substantially lower after the 1991 reform than before. The abolishment of the pure labor market reasons for aged 60–64 in 1991 hence seems to have had an effect on disability insurance recipiency in the affected age group. The effect was larger for men than for women. Entry rates into disability insurance were higher for men before the reform, but of the same magnitude as for women after the reform.

The abolishment of special eligibility rules for older workers in 1997 has been thoroughly analyzed in Karlström, Palme and Svensson (2008). According to their analysis, there is no significant effect on entry rates into the disabil-

ity insurance. There is, however, a significant anticipation effect – an increase in entry rates into DI just before the reform – corresponding to almost 2 percent of the labor force in ages 60–64. Since the new eligibility rules were announced long before they were implemented, workers who believed they would pass the pre-reform eligibility rules, but not the post-reform ones, could apply under the pre-reform regime. This effect is seen in Figure 2.17(c) and 2.17(d) from the increase in the difference in DI entry rates during 1996 and 1997.



(a) Men age 55–59

(b) Women age 55–59

(c) Men age 60–64

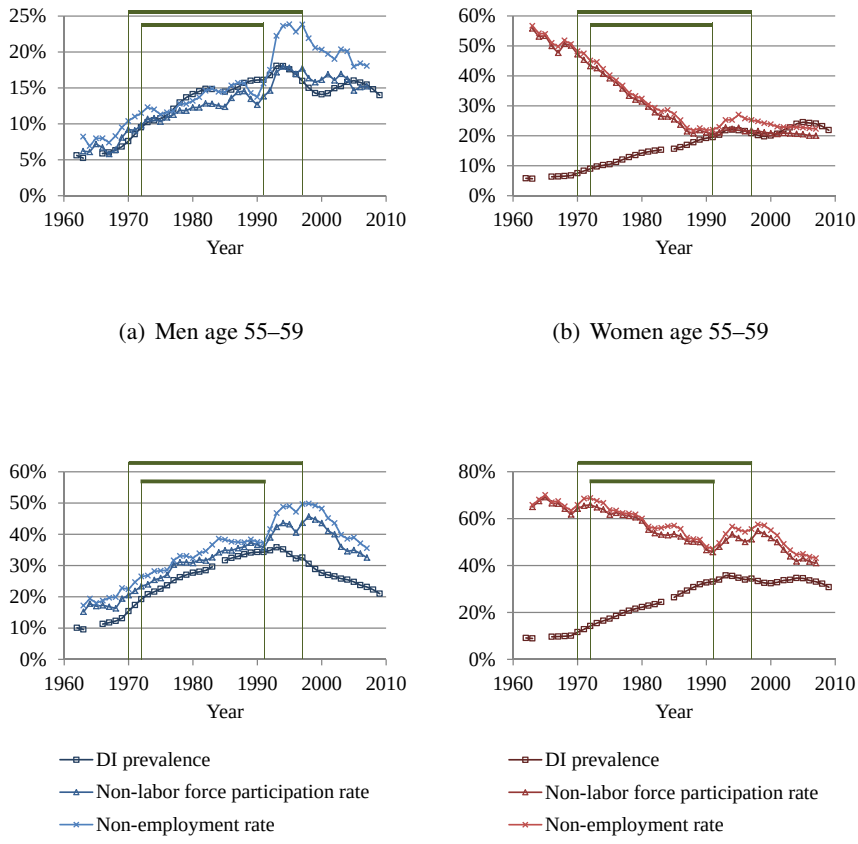
(d) Women age 60–64

Figure 2.25: Incidence of disability insurance utilization and the timing of DI reforms
Source: Swedish Social Insurance Agency

6.2. Program Eligibility and Labor Market Outcomes

The eligibility reforms for older workers seem to have had an effect on the utilization of the disability insurance. An extended question is to what extent these effects were translated into effects on employment and labor force participation rates. Figure 2.26 shows the development of disability insurance

prevalence, non-labor force participation and non-employment for men and women aged 55–59 and 60–64. Figure 2.27 shows the difference in non-labor force participation and non-employment rates between the age groups 60–64 and 55–59, along with the corresponding difference for the prevalence and incidence of disability insurance recipiency. The reforms under study are marked with vertical lines.



(a) Men age 55–59

(b) Women age 55–59

(c) Men age 60–64

(d) Women age 60–64

Figure 2.26: DI prevalence, non-labor force participation and non-employment
Source: Swedish Social Insurance Agency and Statistics Sweden

The left-hand panels in Figure 2.26 show that the change in the prevalence of disability insurance recipiency after the reform in 1970 was indeed translated into a correspondingly large increase in non-employment and non-

labor force participation for the male population in both age groups. Figure 2.27 shows that the differences in non-labor force participation and non-employment between the age groups 60–64 and 55–59 increased in the same manner as disability insurance recipiency during the 1970s. The pure labor market reasons, introduced in 1972, were not extensively used until the mid 1980s. When they were used, however, we do see an increase in non-employment and non-labor force participation that suggests a continuously close relationship between the prevalence of disability insurance recipiency and labor market outcomes also in the 1980s. For the 1970s and 1980s, the utilization of the disability insurance program hence seems to have translated into effects on non-employment and non-labor force participation rates.

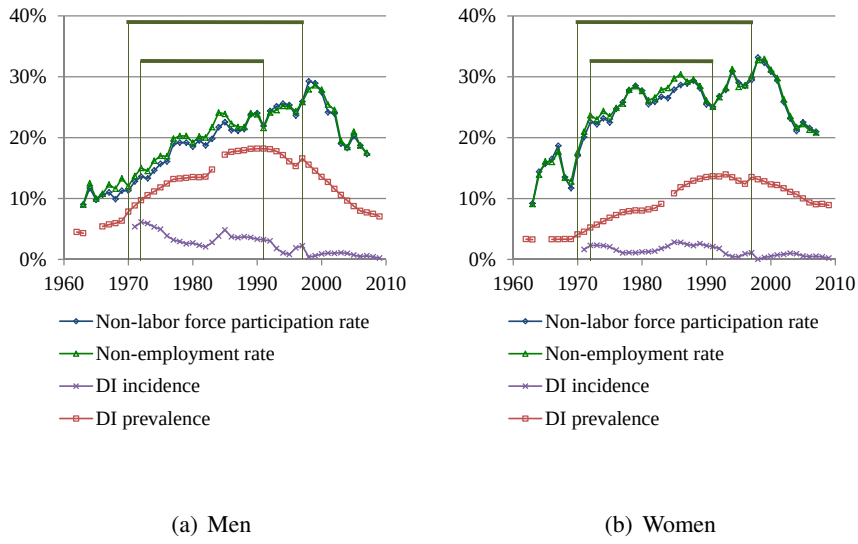


Figure 2.27: Differences between age groups 60–64 and 55–59
Source: Swedish Social Insurance Agency and Statistics Sweden

The 1991 reform seems to have had a very different effect. While there was a rapid decrease in disability insurance recipiency, both non-employment and non-labor force participation increased in the age group 60–64. In the age group 55–59, the decrease in disability insurance recipiency coincided with an increase in non-employment, but not in non-labor force participation. The background to this result might be the deep economic recession in the early 1990s which resulted in a sharp decrease in labor demand. As concluded in Section 4, disability benefits were replaced by income from unemployment benefits, occupational pension and sickness benefits.

Contrary to the 1991 reform, Figure 2.27 suggests that the abolition of the special eligibility rules for workers aged 60–64 in 1997 was followed by increased employment and labor force participation. A detailed analysis of the effects of this reform on employment and on the utilization of the sickness and unemployment insurance programs is found in Karlström, Palme and Svensson (2008). They did not, however, find a significant effect of the reform on employment, but did find an effect on both entry and persistence in the unemployment and sickness insurance programs (not considered in Figures 2.26 and 2.27). Their conclusion is that the other income security programs worked like communicating vessels that crowded out the employment effect of the stricter eligibility rules enacted in the 1997 reform. Looking closer at Figure 2.27, the decrease in non-employment and non-labor force participation did not come until a few years after this reform. It is therefore difficult to attribute the drop to the reform itself.

6.3. Conclusions about Program Eligibility, Disability Insurance and Labor Market Outcomes

Our analysis of the changes in the eligibility rules shows that the introduction of special eligibility rules for older workers in 1970 seems to have had an effect on the utilization of the disability insurance and that the effect translated into effects on labor force participation and employment. We did not find support for an additional effect of the introduction of pure labor market reasons for older workers in 1972 on disability insurance recipiency. The pure labor market reasons were not being used extensively until the early and mid 1980s. At that time, however, there is also an increase in non-employment and non-labor force participation.

For the 1990 and 1997 reforms, the analysis shows that the marked change in utilization of the disability insurance was "crowded out" by changes in the utilization of other income security programs. However, the long term relative increase in employment and labor force participation of the age group 60–64 among both males and females after 1997 suggests that there was an effect of the 1997 reform in prolonging the time before the permanent exit from the labor market of older workers.

7. Overall Conclusions

In this paper, we posed three main research questions. The first question was whether the development of disability insurance recipiency over the past decades can be explained by changes in the health status of the population. We found some support for this hypothesis. We focused on the development

for men and women in the age groups 45–54, 55–59 and 60–64. The analysis showed that the demographic groups with the least advantageous health development were the same groups with the least advantageous development in disability insurance recipiency. In particular, we found a more adverse development of the health of women compared to men, and the health of younger compared to older. The same pattern can be found in the development of disability insurance recipiency. The relation between population health and disability insurance recipiency was least apparent for the oldest age group, aged 60–64.

The second question was whether the changes in disability insurance recipiency can be explained by changes in the eligibility rules in the disability insurance program. We focused on the introduction and abolishment of two sets of eligibility rules that affected the oldest age group, aged 60–64, only. The first were the special eligibility rules for older workers, implying an exempt from rehabilitation and retraining, lowered requirements for the medical assessment of working capacity and a possibility to consider functional limitations due to normal aging for eligibility to disability benefits. The second was the introduction of pure labor market reasons for older workers, making them eligible for disability benefits if they were still unemployed when reaching the time limit for unemployment benefits. The special eligibility rules for older workers were in effect between 1970 and 1997, while the pure labor market reasons were in effect between 1972 and 1991.

For some of the changes in eligibility for older workers, we found evidence of an effect on disability insurance recipiency. The introduction of the special eligibility rules in 1970 seems to have had a large impact on disability insurance recipiency. The introduction of pure labor market reasons in 1972, however, seems to have induced only a small additional increase in the entry rates into the disability insurance program. The abolishment of the pure labor market reasons in 1991 seems to have had a larger effect on disability insurance recipiency. The abolishment of the special eligibility rules for older workers in 1997 did affect disability insurance recipiency but the effect on employment was crowded out by an increased utilization of the sickness and unemployment insurances. In the long run, however, the difference in non-employment rates between the age groups 60–64 and 55–59 has been decreasing after the reform, which might suggest that the eligibility changes in the disability insurance in 1997 eventually spilled over on employment.

The final question was to what extent the changes in eligibility rules for older workers affected employment and labor force participation. The answer is ambiguous. For the 1970 reform, this seems to be true in the sense that the reform opened the disability insurance program as a much more frequent exit route from employment than before. In fact, the disability insurance program

became almost the only pathway out of the labor force before the normal retirement age in the decades following the reform. The results are complicated to analyze for the 1991 and 1997 reforms. The 1991 reform coincided with a sharp drop in the employment rate, caused by a labor demand shock from a severe recession. The conclusion for the 1997 reform is that it did not lead to a significant increase in employment. The effect on the disability insurance utilization was crowded out by an increase in the unemployment rate and increased utilization of the sickness insurance. However, several years after the reform, we have seen significant improvements in employment rates among older workers. It is an open question to what extent the new regime within the disability insurance after 1997 contributed to this development.

There are several significant changes in the usage of disability benefits that cannot be directly related to either changes in health or reforms of the rules of the program. Throughout the many graphs shown in this paper we have seen that trends tend to continue, without visible changes in eligibility rules or population health. Possible explanations are (1) changes in the demand for labor with health impairments; (2) formation of norms on eligibility to disability insurance in the social security administration and in the society in general; (3) administrative policies within the social insurance system; or (4) changes in economic incentives the disability insurance program primarily attributed to maturation of the supplementary pension program (the ATP system). The relative importance of these factors is a subject for further research in this area.

References

Gruber, J., and D. A. Wise (2010), *Social Security and Retirement Around the World: The Relationship to Youth Employment*, Chicago: The University of Chicago Press.

Hedström, P. (1987), "Disability Pension: Welfare or Misfortune?" *International Journal of Sociology*, vol. 87, pages 1261–1284.

Karlström, A., M. Palme, and I. Svensson (2008), "The Employment Effect of Stricter Rules for Eligibility to DI: Evidence from a Natural Experiment in Sweden," *Journal of Public Economics*, vol. 92, pages 2071–2082.

Kruse, A., and L. Söderström (1989), "Early Retirement in Sweden," in W. Schmäl (ed.), *Redefining the Process of Retirement*, Berlin: Springer.

Lindbeck, A., M. Palme, and M. Persson (2009), "Social Interaction and Sickness Absence," Working Paper 2009:4, Department of Economics, Stockholm University.

Palme, M., and I. Svensson (1999), "Social Security, Occupational Pensions and Retirement in Sweden," in J. Gruber and D. Wise (eds.), *Social Security and Retirement Around the World*, Chicago: The University of Chicago Press.

Palme, M., and I. Svensson (2004), "Income Security Programs and Retirement in Sweden," in J. Gruber and D. Wise (eds.), *Social Security and Retirement Around the World: Micro-estimates*, Chicago: The University of Chicago Press.

Statistics Sweden (2009), "Undersökning om levnadsförhållanden 2009", available at
http://www.scb.se/Statistik/LE/LE0101/_dokument/LE0101_DO_2009.pdf.

Skogman Thoursie, P. (1999), *Disability and Work in Sweden*, Ph.D. thesis No. 39, Stockholm: Swedish Institute for Social Research.

Wadensjö, E. (1996), "Early Exit from the Swedish Labour Market," in E. Wadensjö (ed.), *The Nordic Labour Markets in the 1990s*, Amsterdam: North-Holland.

Wise, D. A. (ed.) (2012), *Social Security Programs and Retirement Around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms*, Chicago: The University of Chicago Press.

Chapter 3

Does Privatization of Vocational Rehabilitation Improve Labor Market Opportunities? Evidence from a Field Experiment in Sweden*

Lisa Laun Peter Skogman Thoursie

Abstract

This paper analyzes if privatization of vocational rehabilitation can improve labor market opportunities for individuals on long-term sickness absence. We use a field experiment performed by the Employment Service and the Social Insurance Agency in Sweden during 2008 to 2010, in which over 4,000 participants were randomly offered private and public rehabilitation. We find no differences in employment rates following rehabilitation between individuals who received rehabilitation by private and public providers. Also the average cost of rehabilitation was essentially equal for the two types of providers. This suggests that there are no large efficiency gains from privatizing vocational rehabilitation.

Keywords: Vocational rehabilitation, Privatization, Field experiment
JEL-codes: J14, J68, L33

*We would like to thank the Swedish Social Insurance Agency and the Employment Service for their engagement in this project. We are especially grateful to Tomas Jeppsson at the Employment Service and Eva-Charlott Jansson at the Swedish Social Insurance Agency. We also thank Staffan Brantingson, Andrzej Dudziuk, Oscar Dunge and Caroline Kalerud with all help to provide the data. We are also indebted to Laura Hartman for her cooperation in the initial phase of this project. We are grateful for comments from Erica Lindahl, Eva Mörk, Mårten Palme, Olof Åslund and seminar participants at the Department of Economics at Uppsala University, the Institute for Social Research at Stockholm University, the Department of Economics at Linnaeus University, the Workshop on Absenteeism and Social Insurance, the Workshop of the Centre for Research in Active Labour Market Policy Effects, the 24th Conference of the European Association of Labour Economists and the 3rd National Conference of Swedish Economists. Lisa Laun gratefully acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation.

1. Introduction

There has been an intense debate about whether the private sector can provide a variety of public services more efficiently than the government. The controversy has concerned services ranging from education, job placement services and health care to transportation and garbage collection. One motivation for privatization is to induce incentives for innovation and cost reduction (Grossman and Hart, 1986; Hart and Moore, 1990; Hart et al., 1997). As pointed out by Hart et al. (1997), however, incentives to engage in pure cost reduction may be too strong, which could deter the quality of the services. This is particularly problematic when quality is imperfectly observable.

The merit of outsourcing public services is ultimately an empirical question, but the empirical evidence is too scarce for drawing a general conclusion of whether privatization improves efficiency. Existing studies suggest that the result depends on the nature of the service provided. In a recent review, Andersson and Jordahl (2011) argue that outsourcing works relatively well for public services with little contractibility problems, such as garbage collection, but appears more problematic for credence goods, such as residential youth care. The authors also conclude that the lack of exogenous variation in outsourcing remains a major weakness in the empirical literature.

In this paper, we study if privately provided vocational rehabilitation can improve labor market opportunities of individuals on long-term sick leave, compared to rehabilitation provided by the public. The understanding of how to motivate individuals with a long history of health-related absence to return to work is limited (Autor and Duggan, 2006) and the efficiency of vocational rehabilitation per se is an unresolved question (see Johansson et al., 2011). The Ticket to Work program in the US offers disability beneficiaries increased opportunities to obtain return-to-work services but the participation rate has been low (Stapleton et al., 2008). The Pathways to Work program, offering support for incapacity benefit claimants in the UK, however, seems to have increased the return to work (Adam et al., 2008). Since vocational rehabilitation is a complex service and little is known about the relative efficiency of different types of rehabilitation measures, the potential for innovation might be large. This makes it particularly interesting to study what private providers can achieve.

Our estimation strategy relies on a field experiment in which 4,090 participants were randomly assigned to private and public rehabilitation. The experiment was conducted during 2008–2010 in four Swedish regions with different local labor markets, which increases the external validity of the results. The participants had been receiving sickness benefits for at least two years or were receiving temporary disability benefits, and were either unem-

ployed or unable to return to their previous workplace for health reasons. The private providers competed for contracts through public tenders, competed for clients through consumer choice and were rewarded based on results in terms of acquiring employment. The public provider faced no financial incentives but was subject to a quantitative performance goal when operating on a regular basis. Rehabilitation could be provided for up to one year.

The research design avoids the problem of endogenous selection into private and public vocational rehabilitation that would be the case without the random assignment. Individuals who were randomized to private rehabilitation could demand rehabilitation by the public, but not the other way around, which implies one-sided non-compliance. We estimate the causal effect of privately provided vocational rehabilitation using an instrumental variable approach in which the initial assignment is the instrument for private rehabilitation. We follow the participants up to two years after randomization and compare the probability of acquiring employment in the second year, when rehabilitation should be completed.

The main contribution of this paper is to provide direct evidence on the relative performance of the private and the public sector in providing vocational rehabilitation services to individuals on long-term sickness absence. It also contributes to the empirical literature on the effectiveness of privatization in general.[1] Given the extent to which public services are being outsourced, there is remarkably little research on how private actors perform. In particular, there is little experimental evidence. Bennmarker et al. (2009) analyze an experiment with privately provided job placement services in Sweden and find no differences in the relative performance of private and public providers in terms of acquiring employment. Behaghel et al. (2012) use an experiment with privately provided job placement services in France and find that the public provider acquired employment to a substantially larger extent than the private providers. The target group for vocational rehabilitation is in general further from the labor market than the target group for job placement services and the vocational rehabilitation services are typically more comprehensive and provided for a longer period than job placement services. In that sense, we study the virtues of privatization of a more complex public service compared to the above-mentioned studies.

Our main finding is that there is no difference in the success of acquiring employment following rehabilitation between private and public rehabilitation providers. There was a substantial transition to employment following rehabilitation for both types of providers but the transition rates were the same across the two groups. The development over time by month since randomization was

[1] See, e.g., Dewenter and Malatesta (2001), Duggan (2004) and Aizer et al. (2007).

also remarkably similar. There are no indications of one type of provider performing better than the other up to two years following randomization. Rough calculations suggest that also the public cost of private and public rehabilitation was the same. This suggests that there are no large efficiency gains from privatizing vocational rehabilitation.

One difference we found is that private and public providers made use of different types of employment subsidies. Whereas private providers were more successful in providing employment with a new start job subsidy, which is based on the length of absence from the labor market, the public provider more often provided employment with a wage subsidy for disabled workers, which is based on the degree of reduced work capacity. A likely reason for this is that the assessment of work capacity for work subsidies was carried out by the Employment Service and that this type of subsidy therefore might have been more accessible to the public provider. Up to two years following randomization, however, we find no difference in employment duration for the two types of subsidized employments. A study by the Swedish Agency for Public Management (2011) also shows that the average subsidy payment per participant in general is similar across the two types of subsidies. Therefore, this difference has no financial implications for the short-run effects estimated in this paper. Future studies may show if the type of subsidized employment affects labor market opportunities in the long run.

The remainder of the paper is organized as follows. Section 2 describes the institutional framework of vocational rehabilitation and Section 3 presents detailed information about the experiment. Section 4 discusses the theoretical aspects and provides a conceptual framework of treatment. Section 5 describes the data while Section 6 discusses the empirical strategy. Section 7 presents the results of private compared to public rehabilitation in terms of acquiring employment or education and Section 8 discusses the implications of the results.

2. Background

The two main income support programs in Sweden for individuals who are unable to work for health reasons are the sickness insurance and the disability insurance programs. The sickness insurance program provides income replacement for individuals with a temporarily reduced working capacity whereas the disability insurance program replaces foregone earnings for individuals with a lasting impairment. Before 1 July 2008 there was no time limit for receiving sickness benefits and disability benefits could be temporary or permanent, which implied substantial overlap between the two programs. Since then, sickness benefits can be received for up to two and a half years and temporary

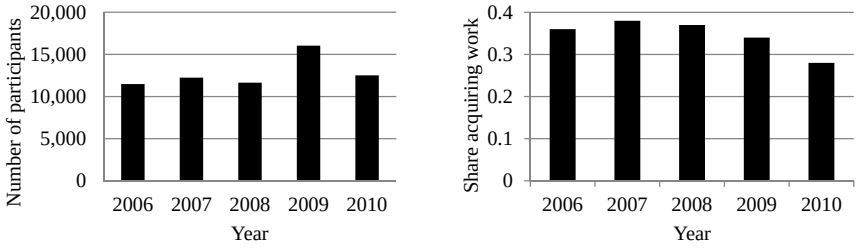
disability benefits are no longer granted. Medical conditions are typically assessed by a physician but eligibility for benefits is determined by the Social Insurance Agency. Sickness and disability benefits can be granted full-time or to 25, 50, or 75 percent. The replacement rate is 80 percent of foregone earnings in the sickness insurance and 64 percent of foregone earnings in the disability insurance, up to a cap. Most workers in Sweden also receive supplementary compensation from collectively bargained insurances with varying degrees of compensation.

The Social Insurance Agency is responsible for the rehabilitation of individuals in the sickness and the disability insurance programs. Rehabilitation can be of medical, social or vocational character. The Social Insurance Agency assesses the need for rehabilitation and coordinates the rehabilitation process, but does not provide the rehabilitation services. The health care system provides medical rehabilitation, the social services provide social rehabilitation and the employer provides vocational rehabilitation for individuals who are able to return to their previous workplace. The Employment Service provides vocational rehabilitation for individuals who are unemployed or unable to return to their workplace because of the health impairment. The cooperation between the Social Insurance Agency and the Employment Service regarding rehabilitation of the individuals with no employment to return to is formalized within "the vocational rehabilitation program".

The vocational rehabilitation program started in 2003 and was implemented across the country by 2005. The motivation was a large increase in recipients of sickness and disability benefits during the late 1990s and early 2000s and the fact that only 8 percent of the long-term sick participated in a rehabilitation activity (The Social Insurance Agency and the Employment Service, 2011). The purpose of the vocational rehabilitation program is to provide support during the process from health related absence to work. The target group is individuals on sickness or disability benefits who are unemployed or unable to return to their previous workplace because of the impairment and are assessed by the Social Insurance Agency to be in need of vocational rehabilitation. The most common diagnoses are mental disorders and musculoskeletal diseases. Whereas the Social Insurance Agency provides administrative and financial support, the Employment Service provides the rehabilitation activities, which can last for up to one year. Typical activities are counseling, job training and job search assistance.

The vocational rehabilitation program is subject to two explicit goals. The first goal is that 15,000 individuals should begin vocational rehabilitation each year. This figure can be related to the total number of individuals receiving sickness and disability benefits, which was about 550,000 in 2010. The second goal is that 40 percent of the participants should be in work or education

one year after entering the program. This includes unsubsidized or subsidized employments, regular education and occupational training programs. Figure 3.1 shows the fulfillment of the two goals during 2006–2010. The first goal of recruiting 15,000 participants yearly was only achieved in 2009, when more than 16,000 individuals entered the program. In the other years, the number of participants entering the program was around 12,000 per year. The second goal of acquiring work or education for 40 percent of the participants has never been achieved, but the exit rate to work or education exceeded 35 percent during 2005–2008.



(a) Number of new participants per year (b) Share acquiring work or education

Figure 3.1: Number of new participants in the vocational rehabilitation program and the share of participants acquiring work or education, 2006–2010

This type of quantitative goal might have selection effects such that certain types of individuals are selected for rehabilitation. This relates to the literature on cream skimming, defined as case workers selecting eligible clients into a program who would have done well without participation, rather than persons with the greatest expected gain from participating (see Anderson et al., 1993; Barnow, 1992; Heckman et al., 1997 and 2002). Cream skimming is not an issue in this study since participants were randomly selected to private and public providers.

3. The Experiment

In 2008, the Swedish government assigned the Social Insurance Agency and the Employment Service to perform an experiment with private provision of vocational rehabilitation as an alternative to the rehabilitation provided by the Employment Service. The purpose was to promote innovation and individualization of vocational rehabilitation services. The experiment was performed during 2008–2010 in four Swedish regions: Stockholm, Gothenburg, Dalarna

and Västerbotten. The target group was individuals who had been collecting sickness benefits for more than two years or were collecting temporary disability benefits, were either unemployed or unable to return to their previous workplace for health reasons, and were assessed to be in need of vocational rehabilitation.[2] The private providers competed for contracts through public tenders, competed for clients through consumer choice and were rewarded based on results in terms of acquiring employment.

The Employment Service contracted the private providers through public tenders. The contestants were assessed based on a set of known criteria regarding the services offered, the degree of individualization, the labor market relevance, the methodology, the degree of innovation and the qualification of the personnel. In each region, the highest ranked providers were contracted. The compensation scheme was determined in advance. Since the government explicitly wanted to encourage the participation of non-profit actors from the social economy,[3] and these actors were presumed to be more credit constrained than for-profit actors, non-profit actors faced a slightly more favorable compensation scheme. In this paper, we cannot compare the performance of for-profit and non-profit private providers, however, since participants randomized to private rehabilitation chose among the providers.

A provider could receive a total of 60,000 SEK per participant (6,700 €), paid in three steps based on performance in terms of acquiring employment:

1. Assignment fee: for-profit 45 %, non-profit 55 %
2. Acquiring employment: for-profit 25 %, non-profit 20 %
3. Retaining employment for six months: for-profit 30 %, non-profit 25 %

The assignment fee was granted after two weeks of rehabilitation. Unsubsidized and subsidized employment entitled to full compensation whereas self-employment entitled to full compensation if the business was running without support from the Employment Service after six months. Regular education and occupational training reduced compensation in steps 2 and 3 to 25 percent in total, granted after three months.[4] Rehabilitation could be provided for up to

[2]This is a subset of the target group for the vocational rehabilitation program in general, which also includes individuals on sickness absence for a shorter period than two years and individuals collecting permanent disability benefits.

[3]The social economy refers to organized activities which primarily aim at serving the community, are being built on democratic values, and are organizationally independent of the public sector (The Swedish Government, 1998). These activities are run mainly by associations, cooperatives or foundations and the main driving force is the benefit of the public or the members, and not profit.

[4]Non-profit actors were entitled to compensation for employment in the own establishment while for-profit actors were compensated for employment in the own establishment in step 2 only if the participant had acquired an employment outside the own establishment in step 3.

12 months and should be provided on a full-time basis.

There are two main types of subsidized employments. The first is a "wage subsidy for disabled workers", which is granted for individuals with a documented reduced work capacity. The subsidy compensates the employer for the lost productivity and the size is determined by the degree of impairment. The assessment of work capacity is conducted by the Employment Service. The second is a "new start job subsidy", which is granted for individuals with a certain period of absence from the labor market due to unemployment, sickness absence, immigration or imprisonment. The subsidy is paid during a period equal to the length of absence from the labor market. The size of the subsidy is twice the payroll tax, at a normal rate of 31.42 percent in 2010, for individuals aged 26–65 and one times the payroll tax for individuals aged 20–26.[5] According to the Swedish Agency for Public Management (2011), the average subsidy payment per participant is essentially equal across the two types of subsidies.

One concern in the experiment was that the public provider would have easier access to wage subsidies for disabled workers, since the assessment of work capacity was conducted by the Employment Service. Therefore, the Employment Service was explicitly instructed not to discriminate private providers who asked for such an assessment. Still, the public caseworkers might be more experienced handling this type of subsidy. The new start job subsidy, on the other hand, was a relatively new type of subsidy. Since a new start job could be organized without involvement of the Employment Service, there is reason to expect that new start job subsidies were more attractive to use for the private providers.

The public tenders were completed in August 2008 for for-profit actors and in January 2009 for non-profit actors, with a total of 1,770 rehabilitation slots contracted. There were more competitors than contracts in all regions. In Stockholm, for example, a total of 30 private providers competed and 11 received a contract. There were 9 providers contracted in Gothenburg, 6 in Dalarna and 3 in Västerbotten.

The Social Insurance Agency recruited the participants to the experiment through two different pathways. First, caseworkers continuously identified potential participants from the case files at the Social Insurance Agency. Second, the Social Insurance Agency sent out close to 24,400 information letters to all

[5]The payroll tax rate is lower for individuals aged 26 or below, amounting to 15.49 percent in 2010. Individuals aged 27–54 must have been absent for at least 12 of the last 15 months. For individuals aged 20–26 or 55–65, absence for at least 6 of the last 9 months is enough for qualification. The maximum period is 12 months for individuals aged 20–26, 5 years for individuals aged 26–55 and 10 years for individuals aged 55–65. Individuals aged 20–26 who qualify for a new start job subsidy due to sickness absence receive an additional subsidy equal to the normal payroll tax rate and can be granted the subsidy for up to 5 years.

individuals who had been collecting sickness benefits for more than two years or were collecting temporary disability benefits, in which they offered active rehabilitation services in cooperation with the Employment Service. The purpose was to attract individuals who were not identified by the caseworkers but were motivated to participate in rehabilitation. Individuals responding to the letter were recruited to the experiment if they were assessed to be in need of vocational rehabilitation.

A total of 4,090 individuals were recruited to the experiment from June 2008 to August 2009, of which 3,587 entered through the "ordinary pathway" and 503 through the "information pathway". Once the participants had been recruited, they were randomly offered private and public rehabilitation. Figure 3.2 shows the number of individuals randomized to private and public rehabilitation each month during the recruitment period. The increased probability of being randomized to private rehabilitation from March to May 2009 was due to a tilting of the randomization because providers waited for being assigned participants.

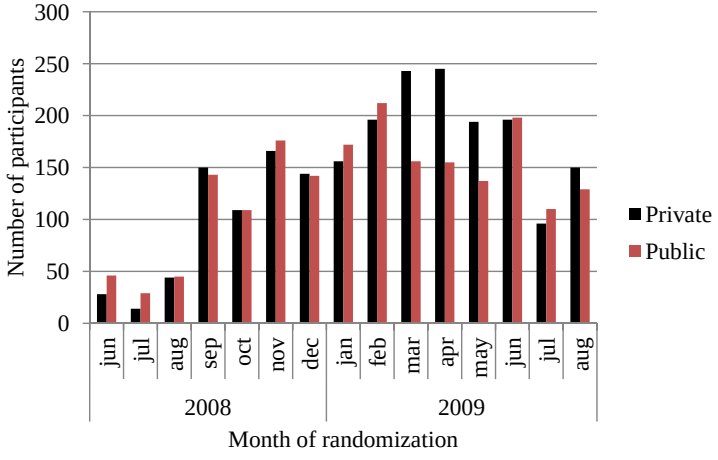


Figure 3.2: Number of participants randomized to private and public rehabilitation, by month of randomization

Individuals randomized to private rehabilitation were asked by the Social Insurance Agency caseworkers to choose among the private providers in the neighborhood based on information leaflets produced by the providers. The caseworkers were explicitly instructed not to influence the choice of provider. Individuals could not be forced to choose a private provider, however. If they denied privately provided rehabilitation, they were directed to the regular voca-

tional rehabilitation program at the Employment Service. Individuals randomized to public rehabilitation were directed to the Employment Service without having the option of private rehabilitation. We therefore have one-sided noncompliance. Table 3.1 presents the number of participants randomized to private and public rehabilitation by column and the number of participants who complied with the assignment by row. Out of the 4,090 individuals who were recruited to the experiment, 2,131 were randomized to private rehabilitation and 1,959 were randomized to public rehabilitation. Among the individuals who were randomized to private rehabilitation, 1,730 complied with the assignment while 401 denied private rehabilitation. This implies a compliance rate of above 81 percent.

		Randomization outcome (Z)		
		Private rehabilitation	Public rehabilitation	Total
Compliance (D)	Private rehabilitation	1,730	0	1,730
	Public rehabilitation	401	1,959	2,360
	Total	2,131	1,959	4,090

Table 3.1: Number of individuals randomized to private and public rehabilitation and the compliance with the assignment

A study by Malmö University (2010) examines the differences in initial resources and the types of rehabilitation services provided across actors, based on interviews with providers and participants in the experiment. The study finds small differences in the academic competence and background of the employees, but larger differences in the organizational experience with the target group. Although many private provider employees had prior experience with the target group, the organization as a whole did not. Another difference was that the private providers did not have access to the clients' case files, and thus had less information about the client history. Private providers also spent more time on guidance, charting of individual needs and job search assistance, which might be due to the lack of information about the clients. The public provider was more focused on job training.

Important for the comparison of private and public rehabilitation is to also take costs into account. We can only perform rough calculations of the cost of rehabilitation by private and public providers. The cost of public rehabilitation is calculated based on the 2007 yearly report for the vocational rehabilitation program, and includes costs for administration and the provision of rehabilitation services for the Employment Service (The Social Insurance Agency and the Employment Service, 2007). The cost of private rehabilitation includes procurement costs and realized payments to the private providers in accordance with the compensation scheme. Our calculations suggest a cost per

client of around SEK 24,000 (2,700 €) for both private and public providers. The similar cost of rehabilitation for the two types of providers motivates the comparison of labor market outcomes of the participants following rehabilitation.

4. Theoretical Framework

The total effect of privatization estimated in this paper consists of several different components, which can broadly be divided into two groups: initial resources and incentives. Differences in initial resources and incentives across providers can affect the way rehabilitation is carried out, which may in turn affect the labor market opportunities of the participants. To formalize the discussion about the components of the treatment effect, we formulate a rehabilitation production function of the form

$$Pr(Emp_i) = r(R(Private), I(Private), \mu_i, W), \tag{3.1}$$

where R are initial resources, I are incentives, $Private$ indicates organization type, μ_i represents individual human resources such as ability, motivation, education and experience and W are external factors such as local labor market conditions. Due to the random assignment of participants into private and public rehabilitation, the individual's human resources μ_i and the external factors W are independent of organization type. The function r describes how the input factors create different types of rehabilitation, which in turn affect the probability that the participant acquires employment, $Pr(Emp)$.

In this simple model, the total effect of privatization is given by

$$\frac{dPr(Emp_i)}{dPrivate} = \frac{\partial r}{\partial R}\frac{\partial R}{\partial Private} + \frac{\partial r}{\partial I}\frac{\partial I}{\partial Private}, \tag{3.2}$$

where the first term represents the effect of initial resources and the second term represents the effect of incentives. In terms of initial resources, one might claim that the private providers were disadvantaged. We know from Section 3 that they had less organizational experience with the target group and less client information. If these differences in initial resources ($\partial R/\partial Private < 0$) are important for providing efficient rehabilitation ($\partial r/\partial R > 0$), the first term in equation (3.2) is negative.

In terms of incentives, there are several differences between private and public providers. The first incentives component is the private ownership. The basis is the property-rights theory of Grossman and Hart (1986) and Hart and Moore (1990), which was applied to public service contracting by Hart, Schleifer and Vishny (1997). Assuming contractual incompleteness, the residual control right that comes from the ownership of an asset will increase the

bargaining power in the renegotiation of the contract. This increases the incentives for innovation and cost reduction, since the owner will secure a larger share of the surplus from the investments. The implication from the model is that private provision of a public service will reduce costs but have an ambiguous effect on quality. The risk is that the private provider reduces costs in a way that deteriorates non-contractible quality.

The second incentives component is competition. The experiment introduced competition for both public and private providers, which might have affected quality also of publicly provided rehabilitation (for contributions that point out that more competition may stimulate innovations, see for instance, Aghion et al., 2001 and 2005). The private providers, however, faced additional competition since they competed for contracts to participate in the experiment and competed for clients through consumer choice.

The third incentives component, which is closely related to the notion of competition, is reputation building. In settings with repeated interactions, private providers might have large gains from establishing a credible reputation for high quality (Francois and Vlassopoulus, 2008). The additional competition facing the private providers reinforces the importance of reputation building. The performance during the experiment is important for the prospects in future public tenders, and is likely to affect the probability of a permanent shift from public to private provision of vocational rehabilitation. If private providers put an extra amount of effort during the experimental period in order to build up a good reputation, we would overestimate the effect of privatization. Reputation building is also important for attracting new clients during the experiment.

The fourth incentives component is the compensation scheme. The private providers were compensated based on performance. The assignment fee gave an incentive to attract clients, the payment when acquiring employment encouraged the provision of efficient rehabilitation measures, and the compensation when retaining employment for six months motivated the achievement of high quality matches between workers and firms. The public provider had no financial incentives for acquiring employment, but was subject to a quantitative goal of acquiring employment for 40 percent of the participants (for effects of payment schemes, see for example Barnow, 1992; Heckman et al., 2002).

The fifth incentives component relates to the literature on pro-social motivation, suggesting that public sector employees might be more motivated than private sector employees, which would increase the quality of the provided service (see e.g., Besley and Ghatak, 2005; Akerlof and Kranton, 2005). Extrinsic incentives, such as monetary rewards, might also harm the intrinsic motivation of mission-oriented workers at the private providers (see Francois and Vlas-

sopoulos, 2008 for a review).

Taken together, the theory about incentives suggests that private providers faced stronger incentives for providing efficient rehabilitation in terms of differences in ownership, the degree of competition, the importance of reputation building and the terms of compensation. The exception is the theory on how employees in different types of organizations may be differently motivated. If stronger incentives for providing efficient rehabilitation ($\partial I/\partial Private > 0$) translate into increased rehabilitation efficiency ($\partial r/\partial I > 0$), the second term in equation (3.2) is positive. In total, however, a positive effect of incentives could be counteracted by a negative effect of initial resources for private compared to public providers. The experimental estimate obtained in this paper will capture the total effect of privatization and cannot disentangle the impact of the separate components.

5. Data

It is crucial for the experimental design that the randomization of individuals into private and public rehabilitation was carried out properly. Administrative records from the Social Insurance Agency contain demographic information about the participants. Table 3.2 presents the average characteristics of the two groups at the time of randomization along with their differences and the normalized difference. There are no significant differences between individuals randomized to private and public rehabilitation, which indicates that the division of the two groups was indeed random. This is also confirmed if we consider the scale-free normalized difference in means, which is reported in column 4 of Table 3.2.[6]

We create a set of outcome variables by combining daily records of the collection of sickness and disability benefits from the Social Insurance Agency with daily registrations of unemployment and program participation from the Employment Service, from June 2007 until August 2011.[7] Based on these records, we categorize the participants into eight mutually exclusive states on each day since randomization. The first state is to be registered as unemployed at the Employment Service, without participating in an organized activity. The second state is to participate in rehabilitation, including the initial charting pe-

[6]The scale-free difference in means is calculated as $(\mu_1 - \mu_0)/\sqrt{(\sigma_1^2 + \sigma_0^2}$. Imbens and Wooldridge (2009) recommend reporting this difference since it does not systematically increase with the sample size which is the case when relying on the t-statistic. As a rule of thumb, a normalized difference exceeding 0.25 is likely to lead to sensitive results. As shown in Table 3.2 all normalized differences are substantially below 0.25.

[7]The records at the Employment Service regard the registration status and not the collection of unemployment benefits.

Variable	Private	Public	Difference	Normalized difference
Birth year	1966.40	1966.63	–0.233	0.016
	(0.215)	(0.229)	(0.314)	
Woman	0.654	0.666	–0.012	0.019
	(0.010)	(0.011)	(0.015)	
Immigrant	0.225	0.236	–0.012	0.019
	(0.009)	(0.010)	(0.013)	
Employed	0.166	0.172	–0.006	0.012
	(0.008)	(0.009)	(0.012)	
Information pathway	0.130	0.116	0.014	0.029
	(0.007)	(0.007)	(0.010)	
Part-time benefits	0.097	0.084	0.012	0.031
	(0.006)	(0.006)	(0.009)	
Type of benefits				
Temporary disability benefits	0.696	0.687	0.009	0.014
	(0.010)	(0.010)	(0.014)	
Sickness benefits	0.304	0.313	–0.009	0.014
	(0.010)	(0.010)	(0.014)	
Diagnosis				
Mental disorder	0.355	0.364	–0.009	0.013
	(0.010)	(0.011)	(0.015)	
Musculoskeletal disease	0.146	0.136	0.010	0.021
	(0.008)	(0.008)	(0.011)	
Other/Unknown/Combination	0.499	0.500	–0.001	0.002
	(0.011)	(0.011)	(0.016)	
Education				
Less than high school	0.230	0.226	0.005	0.008
	(0.009)	(0.009)	(0.013)	
High school	0.552	0.567	–0.014	0.020
	(0.011)	(0.011)	(0.016)	
College	0.217	0.208	0.009	0.016
	(0.009)	(0.009)	(0.013)	
Region				
Stockholm	0.481	0.490	–0.009	0.012
	(0.011)	(0.011)	(0.016)	
Gothenburg	0.263	0.265	–0.002	0.003
	(0.010)	(0.010)	(0.014)	
Västerbotten	0.145	0.138	0.007	0.014
	(0.008)	(0.008)	(0.011)	
Dalarna	0.110	0.107	0.003	0.007
	(0.007)	(0.007)	(0.010)	
Number of observations	2,131	1,959		

Table 3.2: Pre-program summary statistics by randomization status

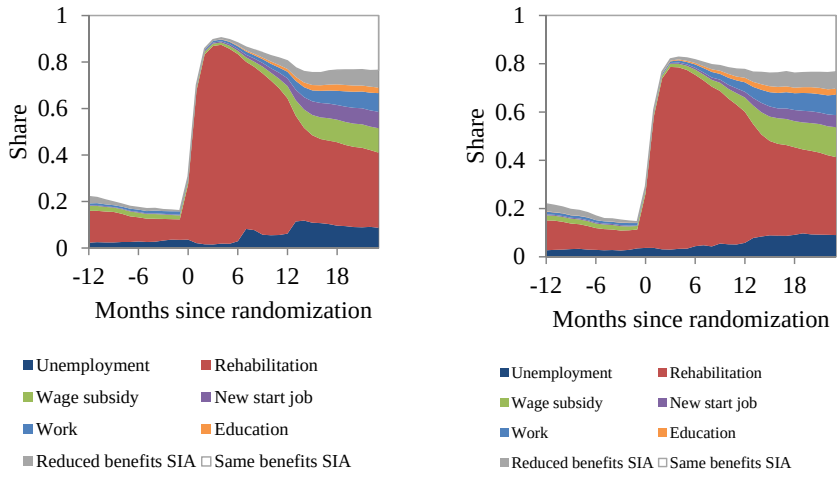
riod as well as active rehabilitation services and labor market programs. The third state is to be employed with a wage subsidy for disabled workers and the fourth state is to be employed with a new start job subsidy, both of which were discussed in Section 3. The fifth state is to be in unsubsidized employment and the sixth state is to participate in regular education or occupational training. The seventh state is to be unregistered at the Employment Service for other reasons than work or education but having decreased the degree of benefits collection from the Social Insurance Agency compared to the day of randomization. The eighth state is to be unregistered at the Employment Service and receive benefits from the Social Insurance Agency to the same extent as on the day of randomization.

The explicit goal for the private as well as the public providers was to get the participants into any type of employment or education. A positive outcome of rehabilitation is therefore defined as being in subsidized employment with a wage subsidy for disabled workers or a new start job subsidy, being in unsubsidized employment or being in regular education or occupational training, i.e., the third through the sixth state. These four states of activity will constitute our main outcome measure, which we will refer to simply as 'employment'.

We follow the activities of the participants from one year before randomization until two years after randomization. Figure 3.3 shows the average share of individuals in each type of activity per month since randomization, for individuals randomized to private and public rehabilitation. Some individuals were registered at the Employment Service, in particular in rehabilitation, even before participating in the experiment, but the share was equal across the two types of provider. Immediately after randomization, the share of individuals participating in rehabilitation increased sharply in both groups. Twelve months later, the share in rehabilitation decreases again and other types of activity gain importance. In particular, a substantial fraction of individuals transfer to employment, i.e., unsubsidized employment, education and subsidized employments. This indicates that rehabilitation in general was successful in making participants return to the labor market. The difference in outcomes between individuals randomized to private and public rehabilitation will be formally analyzed in the results section.

6. Empirical Strategy

The randomized field experiment allows for a simple estimation strategy to study the causal effect of private rehabilitation. The main outcome variable is "employment", which is defined as being in unsubsidized or subsidized employment, regular education or occupational training. In the first part of the analysis, the outcome is the average employment rate during the second year



(a) Private rehabilitation (b) Public rehabilitation

Figure 3.3: Types of activity per month since randomization

since randomization, i.e., when rehabilitation is supposed to have ended. In the second part of the analysis, we study the development of labor market outcomes over time per month since randomization.

Our interest is to estimate the effect on employment of receiving vocational rehabilitation by private providers instead of public providers. However, around 20 percent of the individuals randomized to private providers chose to receive rehabilitation by the public. These individuals might systematically differ from those who chose to comply with the random assignment with respect to individual characteristics that can affect labor market outcomes. If this is the case, the comparison of average differences in outcomes between individuals who received rehabilitation by private and public providers is flawed.

Still, we begin by studying the results of such an analysis by estimating the following OLS model:

$$Y_{di} = \delta Private_i + \sum_{m=12}^{23} \lambda_m Month_m + \beta' X_i + \varepsilon_{di}, \tag{3.3}$$

where Y_{di} indicates if individual i has acquired employment at day $d = 366, \ldots,$ 730 after randomization. The set of month dummies $Month_m$ is included to capture the average effect per month since randomization. The matrix X contains the pre-treatment individual covariates presented in Table 3.2 and a set of

indicator variables for the month of randomization as presented in Figure 3.2. The indicator variable *Private* takes the value 1 if the individual received rehabilitation by a private provider. The coefficient of interest is δ which measures the average outcome difference during the second year after randomization between those who received rehabilitation by private and public providers. Since the participants are observed every day, standard errors are clustered at the individual level.

If receiving private rehabilitation is endogenous to individual characteristics that affect labor market outcomes, the estimation of equation (3.3) yields a biased effect of private rehabilitation. Comparing differences in outcomes between individuals who were randomized to private and public providers would, on the other hand, reflect the causal effect of being offered the opportunity to receive rehabilitation by private providers. This effect is of direct policy interest since private rehabilitation would likely be voluntary if implemented in a large scale and non-compliance would be inevitable. We estimate this intention-to-treat effect by the following reduced form model:

$$Y_{di} = \pi Z_i + \sum_{m=12}^{23} \lambda_m Month_m + \beta' X_i + \varepsilon_{di}, \tag{3.4}$$

where Z is an indicator variable for being randomized to private rehabilitation and the parameter π is the estimated intention-to-treat effect.

We can also estimate the causal effect of actually receiving private rehabilitation by using the initial random assignment as an instrument for private rehabilitation. Since individuals randomized to public providers could not choose private rehabilitation we have one-sided non-compliance. The instrumental variable estimate can therefore be interpreted as an average treatment-on-the-treated effect rather than a local average treatment effect, which would have been the case if we had two-sided non-compliance (Imbens and Angrist, 1994). We estimate the causal effect of private rehabilitation by the reduced form model represented by equation (3.4) and the following first stage equation:

$$Private_i = \theta Z_i + \sum_{m=12}^{23} \lambda_m Month_m + \beta' X_i + \varepsilon_{di}, \tag{3.5}$$

The instrumental variable estimate is obtained by dividing π in the second stage regression by θ from the reduced form regression. The first stage regression, represented by equation (3.5), is interesting in itself since it reveals which individual characteristics affect the probability of choosing public rehabilitation even if randomized to private rehabilitation. Comparing the flawed OLS estimate from equation (3.3) and the IV-estimate also highlights the importance of utilizing the experimental variation for drawing conclusions about privately provided rehabilitation.

In order to capture the development over time, the second part of the analysis estimates the causal effect of private rehabilitation per month since randomization. In addition to the outcome in terms of employment, we also estimate separate effects for the different types of activities included in the employment variable, namely unsubsidized work, education, new start job and wage subsidy for disabled workers. This analysis is based on the following equation:

$$Y_{kdi} = \sum_{m=0}^{23} \delta_{km} Month_m \times Private_i + \sum_{m=0}^{23} \lambda_{km} Month_m + \beta' X_i + \varepsilon_{kdi}, \qquad (3.6)$$

where Y_{kdi} indicates if individual i has achieved outcome k at day $d = 1, ..., 730$ after the day of randomization. The variables of interest are the interactions between the indicator for month since randomization $Month_m$ and the variable indicating whether the individual received rehabilitation by private providers, $Private$. The coefficients δ_{km} measure the differences in outcomes between individuals receiving private and public rehabilitation at month $m = 0, ..., 23$ after the day of randomization. Since $Private$ is potentially endogenous we instrument all interactions between $Private$ and the month dummies with the corresponding interactions between month dummies and the initial random assignment.

7. Results

Table 3.3 presents the main results of the effect of private rehabilitation. Column 1 presents the OLS estimate obtained from equation (3.3), which suggests that the probability of acquiring employment was 3.1 percentage points lower for individuals who received rehabilitation by a private provider. As discussed in Section 6, however, this is not a causal effect since receiving private rehabilitation is not randomly assigned. Column 2 presents the causal effect of being offered private rehabilitation, obtained from the reduced form equation (3.4). The estimates show that there is no significant difference in acquiring employment during the second year since randomization between individuals randomized to private and public rehabilitation. Thus, the causal effect of being offered the opportunity to receive rehabilitation by private providers, i.e., the intention-to-treat effect, is zero. Since the statistical significance level of the treatment effect from this reduced form equation is equal to the significance level of the IV-estimate, also the causal effect of actually receiving rehabilitation by private providers must be zero. This is confirmed by the insignificant IV-estimate reported in Column 4 of Table 3.3.

The results from the first stage regression, presented in Column 3 of Table 3.3, show what characterizes the individuals who complied with the assignment to a private provider. Since the majority of the participants complied

Variables	(1) OLS	(2) Reduced form	(3) First stage	(4) IV
Private rehabilitation	-0.031***			-0.008
	(0.011)			(0.014)
Randomization to private		-0.007	0.813***	
		(0.011)	(0.008)	
Birth year	0.003***	0.003***	-0.000	0.003***
	(0.001)	(0.001)	(0.000)	(0.001)
Woman	-0.012	-0.013	0.045***	-0.013
	(0.012)	(0.012)	(0.010)	(0.012)
Immigrant	-0.058***	-0.058***	0.011	-0.058***
	(0.013)	(0.013)	(0.011)	(0.013)
Employed	0.028*	0.028*	0.028**	0.028*
	(0.016)	(0.016)	(0.011)	(0.016)
Information pathway	-0.078***	-0.080***	0.060***	-0.079***
	(0.017)	(0.017)	(0.012)	(0.017)
Part-time benefits	0.146***	0.146***	-0.022	0.146***
	(0.022)	(0.022)	(0.016)	(0.022)
Sickness benefits	0.043***	0.043***	0.011	0.043***
	(0.013)	(0.013)	(0.010)	(0.013)
Musculoskeletal disease	0.017	0.016	-0.000	0.016
	(0.019)	(0.019)	(0.014)	(0.019)
Other diagnosis	-0.005	-0.005	-0.001	-0.005
	(0.013)	(0.013)	(0.010)	(0.013)
High school	0.024*	0.023*	0.032***	0.023*
	(0.014)	(0.014)	(0.012)	(0.014)
College	0.029*	0.028	0.035**	0.028
	(0.017)	(0.017)	(0.014)	(0.017)
Gothenburg	0.005	0.003	0.049***	0.004
	(0.014)	(0.014)	(0.010)	(0.014)
Västerbotten	-0.035**	-0.034**	-0.022	-0.034**
	(0.017)	(0.017)	(0.015)	(0.017)
Dalarna	-0.030	-0.031	0.014	-0.031
	(0.019)	(0.019)	(0.015)	(0.019)
Constant	-5.739***	-5.770***	0.374	-5.766***
	(1.148)	(1.149)	(0.946)	(1.149)
Observations	1,486,472	1,486,472	1,486,472	1,486,472
R-squared	0.043	0.042	0.687	0.042

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the individual level in parentheses.

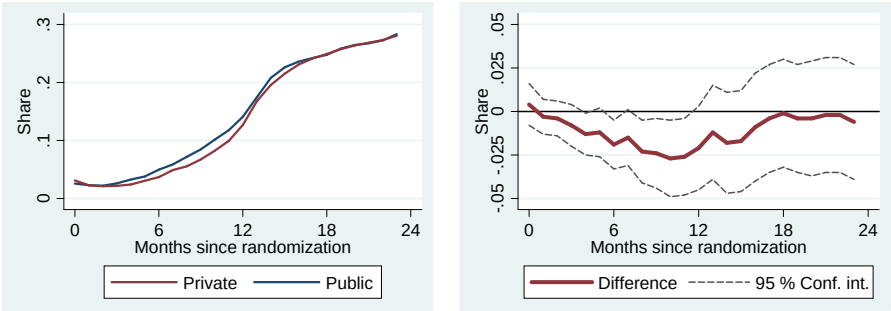
Table 3.3: Employment effects of private rehabilitation

with the randomization it is not surprising that the instrument is highly significant. We also see that certain types of individual characteristics are related to the propensity of accepting private rehabilitation. Based on the estimated coefficients we see that women, individuals with an underlying employment and individuals who entered the experiment through the information pathway were more likely to comply with the random assignment. Having a higher education or being in the Gothenburg region also increased the probability of accepting private rehabilitation.

It is not straightforward to relate the results from the first stage regression to the result obtained by the biased OLS estimate of privatization. The OLS estimate yielded a negative effect of privatization despite the fact that, for example, relatively more highly educated individuals chose to receive rehabilitation by private providers. However, there are no a priori reasons to expect that individuals with certain types of characteristics will benefit more from rehabilitation than others. One explanation pointed out by the Social Insurance Agency (2010) is that a main reason for denying rehabilitation by private providers was that the individual already had a plan for rehabilitation at the Employment Service. If these individuals were closer to achieving employment from the start this might explain why the OLS estimate is negative while the IV estimate is zero.

Next, we consider the monthly development of the difference in outcomes between private and public rehabilitation. We begin by analyzing total employment, including unsubsidized or subsidized employment, regular education and occupational training. Results are reported in Figure 3.4. Figure 3.4(a) shows the monthly development of raw averages of employment, separately for individuals randomized to private and public rehabilitation. Figure 3.4(b) shows the monthly differences in employment, based on the IV-estimates from equation (3.6), including 95 percent confidence intervals of the estimated effects. As shown in the figure, the employment rate of individuals receiving public rehabilitation was slightly higher during the first year, when rehabilitation was supposed to take place. This implies that private rehabilitation on average lasted for a longer time period. Most important, however, is that one year after randomization, when the individuals were supposed to have finished their rehabilitation, there is no significant difference between private and public rehabilitation. By the end of the observation period, the employment rate of individuals receiving private and public rehabilitation is essentially identical.

Figure 3.5 presents the effects of private rehabilitation for the separate types of activity included in the total employment measure. Figure 3.5(a) and 3.5(b) present the result for acquiring unsubsidized employment. From the seventh month after randomization significantly fewer of the participants in private rehabilitation acquired unsubsidized employment, but the difference



(a) Employment rates

(b) Estimated difference, employment

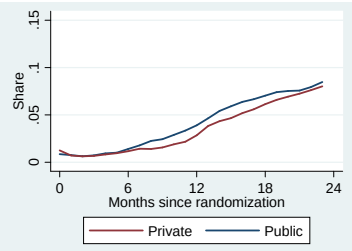
Figure 3.4: Employment rates and IV-estimates of the difference in employment between individuals randomized to private or public rehabilitation, per month since randomization

disappears after the twelfth month. This might also be explained by private rehabilitation lasting for a longer time period. Figure 3.5(c) and 3.5(d) show that there is no significant effect of private rehabilitation on the probability of participating in education.

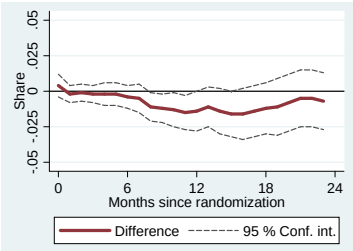
Figure 3.5(e) and 3.5(f) present the results for employment with a new start job subsidy. From about twelve months after randomization, there was a large and significant difference between private and public rehabilitation in the share of individuals having a new start job. During this period, the probability of being employed with a new start job subsidy was almost two percentage points higher for participants in private rehabilitation. In terms of percent, this is a large effect since the fraction of individuals rehabilitated by public providers who received a new start job is five percent. Figure 3.5(g) and 3.5(h) show the results for employment with a wage subsidy for disabled workers. The public provider was more successful in getting participants into employments with a wage subsidy than private providers. The difference almost fully counteracts the larger extent to which the private providers managed to acquire employment with a new start job subsidy, although insignificant.[8] The differences in the types of employment subsidy private and public providers used are consistent with the hypothesis that wage subsidies for disabled workers were more easily accessible for the public provider while new start job subsidies were easier to obtain for private providers, as discussed in Section 3.

Based on the results from the above analysis we conclude that there is no

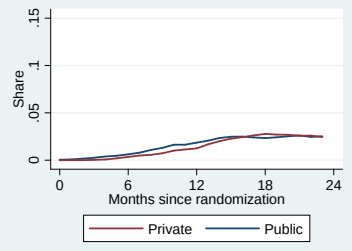
[8] We have also estimated heterogeneous effects with respect to the available individual characteristics but find no significant results. We also find no differences in the probability of being unregistered at the Employment Service while having reduced the benefits from the Social Insurance Agency, which could potentially be seen as a positive outcome.



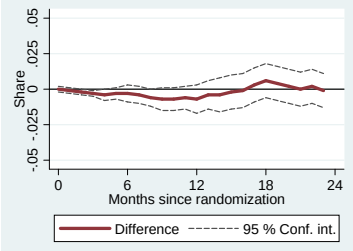
(a) Unsubsidized employment



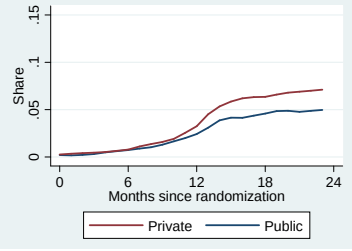
(b) Estimated difference, unsubsidized employment



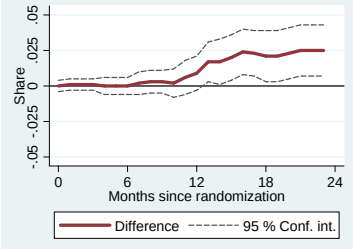
(c) Education



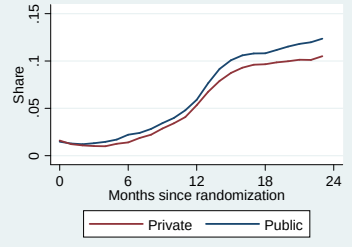
(d) Estimated difference, education



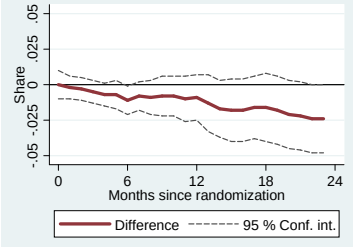
(e) New start job subsidy



(f) Estimated difference, new start job



(g) Wage subsidy for disabled workers



(h) Estimated difference, wage subsidy

Figure 3.5: Raw averages in outcomes and IV estimates of the effect of private rehabilitation, per month since randomization

effect on employment of private compared to public rehabilitation. The estimated coefficient is insignificant and very close to zero. Studying the development over time by month since randomization, we show that the employment rates for the two types of providers converge to exactly the same level following rehabilitation. When looking at separate types of activities we find that private providers managed to acquire employment with a new start job subsidy for their clients to a larger extent than the public provider. The public provider, on the other hand, seems to have acquired employment with a wage subsidy for disabled workers to a larger extent. This might be due to wage subsidies being a more established type of subsidy at the Employment Service whereas new start job subsidies were more accessible for private providers.

A final question is whether private and public providers had access to different employer networks which may have resulted in differences in match quality between the worker and the firm. One might worry that jobs provided by one type of provider lasted for a longer period. If a small number of individuals received stable jobs in one type of organization type but a large number of individuals received jobs for a short period in the other type of organization, such differences are not detected in the estimated treatment effects. To address this issue we have estimated survival rates for the first employment spell obtained, separately for individuals randomized to public and private providers. Results show very small and insignificant differences in survival rates.

8. Conclusion

The merit of privatization is an unresolved question and the empirical evidence is limited. A randomized field experiment gave us the unusual opportunity to assess the effect of privately provided vocational rehabilitation on individual labor market outcomes. The results show that there is no difference in the probability of acquiring employment following rehabilitation for individuals receiving private or public rehabilitation. The estimated effect is insignificant and very close to zero. Although a substantial share of the participants returned to work, the exit rate to employment followed a remarkably similar development over time for private and public providers. Rough calculations show that also the average cost of rehabilitation was essentially equal for the two types of actors. This suggests that there are no large efficiency gains from privatizing vocational rehabilitation.

The difference we find regards the types of employment subsidies used between private and public providers. Private providers acquired employment with a new start job subsidy, for which the subsidy is based on a certain period of absence from the labor market, to a larger extent. Public providers, on the other hand, more often acquired employment with a wage subsidy for disabled

workers, which is based on the degree of reduced work capacity. This difference could be expected since the assessment of work capacity was performed by the Employment Service and wage subsidies therefore might have been easier accessible for the public provider. Since the average size of the subsidy is similar across the two types of subsidies and the duration of employment is the same during the follow-up period in this paper, however, this difference has no financial implications in the short run. Future research will show if the type of subsidy affects labor market prospects in the long run.

How can the absence of differences between private and public providers be interpreted? Although a large share of participants from both private and public providers returned to employment, this paper cannot evaluate whether vocational rehabilitation was effective in general. One might argue that the lack of differences between private and public providers speaks in favor of private providers, since they had less initial experience with the target group and no access to the clients' case files. On the other hand, because of the importance of reputation building and the prospects of vocational rehabilitation being outsourced on a permanent basis, the private providers might have put down additional effort during the experimental period than what would be possible in the long run. This would counteract the potential positive effect of increased experience and would lead to an overestimation of the effect of privatization.

It is important to emphasize that a complete welfare analysis of privatization should include the potential welfare gain from getting the opportunity to choose among rehabilitation providers. The study by Malmö University (2010) shows that the opportunity to choose among the private providers was appreciated, although some participants found the choice to be difficult. An analysis of the value of choice is, however, beyond the scope of this paper. Given that the experiment with privatization of vocational rehabilitation was motivated by efficiency considerations rather than the value of consumer choice, however, we think that the focus on relative performance of private and public providers still is motivated.

References

Adam, S., A. Bozio, C. Emmerson, D. Greenberg, and G. Knight (2008), "A Cost-Benefit Analysis of Pathways to Work for New and Repeat Incapacity Benefit Claimants," Department for Work and Pension, Report No. 498.

Aghion, P., C. Harris, P. Harris, and J. Vickers (2001), "Competition, Imitation, and Growth with Step-by-Step Innovation," *Review of Economic Studies*, vol. 68, pages 467–492.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005), "Competition and Innovation: an Inverted-U Relationship," *Quarterly Journal of Economics*, vol. 120(2), pages 701–728.

Aizer, A., J. Currie, and E. Moretti (2007), "Does Managed Care Hurt Health? Evidence from Medicaid Mothers," *Review of Economics and Statistics*, vol. 89(3), pages 385–399.

Akerlof, G., and R. Kranton (2005), "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, vol. 19(1), pages 1167–201.

Anderson, K. H., R. V. Burkhauser, and J. E. Raymond (1993), "The Effect of Creaming on Placement Rates Under the Job Training Partnership Act," *Industrial and Labor Relations Review*, vol. 46(4), pages 613–24.

Andersson, F., and H. Jordahl (2011), "Outsourcing Public Services: Ownership, Competition, Quality and Contracting," IFN Working Paper No. 874, Research Institute of Industrial Economics.

Autor, D. H., and M. G. Duggan (2006), "Growth in Social Security Disability Rolls: A Fiscal Crisis Unfolding," *Journal of Economic Perspectives*, vol. 20(3), pages 71–96.

Barnow, B. (1992), "The Effects of Performance Standards on State and Local Programs," in C. Manski and I. Garfinkel (eds.), *Evaluating Welfare and Training Programs*, Cambridge: Harvard University Press, pages 277–309.

Behaghel, L., B. Crépon, and M. Gurgand (2012), "Private and Public Provision of Counseling to Job-Seekers: Evidence from a Large Controlled Experiment," IZA Discussion Paper No. 6518, Institute for the Study of Labor.

Bennmarker, H., E. Grönqvist, and B. Öckert (2009), "Effects of Outsourcing Employment Services: Evidence from a Randomized Experiment," IFAU

Working Paper 2009:23, Institute for Evaluation of Labour Market and Education Policy.

Besley, T., and M. Ghatak (2005), "Competition and Incentives with Motivated Agents," *American Economic Review*, vol. 95(3), pages 616–636.

Dewenter, K. L., and P. H. Malatesta (2001), "State-Owned and Privately Owned Firms: An Empirical Analysis of Profitability, Leverage, and Labor Intensity," *American Economic Review*, vol. 91(1), pages 320–334.

Duggan, M. (2004), "Does Contracting Out Increase the Efficiency of Government Programs? Evidence from Medicaid HMOs," *Journal of Public Economics*, vol. 88(12), pages 2549–2572.

Francois, P., and M. Vlassopoulos (2008), "Pro-Social Motivation and the Social Services," *CESifo Economic Studies*, vol. 54(1), pages 22–54.

Grossman, S., and O. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, vol. 94, pages 691–719.

Hart O., and J. Moore (1990), "Property Rights and the Nature of the Firm," *Journal of Political Economy*, vol. 98, pages 1119–1158.

Hart O., A. Schleifer, and R. Vishny (1997), "The Proper Scope of Government: Theory and an Application to Prisons," *Quarterly Journal of Economics*, vol. 112, pages 1127–1161.

Heckman, J., C. Heinrich, and J. Smith (1997), "Assessing the Performance of Performance Standards in Public Bureaucracies," *American Economic Review Papers and Proceedings*, vol. 87(2), pages 389–395.

Heckman, J., C. Heinrich, and J. Smith (2002), "The Performance of Performance Standards," *Journal of Human Resources*, vol. 37(4), pages 778–811.

Imbens, G., and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, vol. 62(2), pages 467–475.

Imbens, G., and J. Wooldridge (2009), "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, vol. 47, pages 5–86.

Johansson, P., E. Aydin, S. Bergendorff, N. Granqvist, M. Josephson, and I. Sohlberg (2011), "Arbetslivsinriktad rehabilitering," Underlagsrapport nr. 7 till den parlamentariska socialförsäkringsutredningen (S 2010:4).

Malmö University (2010), "Samma på olika vis. En jämförande utvärdering av försöksverksamheten med alternativa insatser för långtidssjukskrivna," Malmö högskola, enheten för kompetensutveckling och utvärdering.

The Social Insurance Agency and the Employment Service (2007), "Återrapportering om insatser för att fler sjukskrivna ska återfå arbetsförmåga och få ett arbete", Report No. 29251–2007.

The Social Insurance Agency and the Employment Service (2011), "Samverkan mellan Arbetsförmedlingen och Försäkringskassan under 2000-talet," Underlagsrapport nr. 2 till den parlamentariska socialförsäkringsutredningen (S 2010:04).

The Social Insurance Agency (2010), "Utvärdering av handläggningsprocessen inför deltagande i aktiv insats. En försöksverksamhet med alternativa insatser för långtidssjukskrivna," Social Insurance Report 2010:12.

Stapleton, D., G. Livermore, C. Thornton, B. O'Day, R. Weathers, K. Harrison, S. O'Neil, E. S. Martin, D. Wittenburg, and D. Wright (2008), "Ticket to Work at the Crossroads: A Solid Foundation with an Uncertain Future," Mathematica Policy Research Report No. 8977.

The Swedish Agency for Public Management (2011), "Delredovisning av uppdraget om subventionerade anställningar m.m. – Statistiskt sifferunderlag," Report No. 2011/181–5.

The Swedish Government (1998), "Social ekonomi i EU-landet Sverige – tradition och förnyelse i samma begrepp," DS 1998:48.

Chapter 4

Screening Stringency in the Disability Insurance Program*

Per Johansson Lisa Laun Tobias Laun

Abstract

In this paper, we propose a strategy for assessing how the inflow to the disability insurance program has been governed over time. We analyze the ex-ante health of individuals entering the program by using the ex-post mortality from all years in which the individuals are observed in the future, compared to individuals not entering the program in the same year. To structure the analysis, we also provide a theoretical model of the application decision to the disability insurance program, which gives a relation between screening stringency and the relative health of new beneficiaries to non-beneficiaries. Applying the empirical strategy to Sweden, we find large variation in the relative health of new beneficiaries compared to non-beneficiaries over time. Some of the fluctuations correspond well to formal changes to screening stringency. However, we also find large variation in health during periods when no changes to formal eligibility criteria have been pursued.

Keywords: Disability insurance, Screening stringency, Cox proportional hazard model
JEL-codes: C41, I18, J14

*We are thankful for comments from Marcus Eliason, John Ham, Karl-Oskar Lindgren, Eva Mörk and Mårten Palme. We also thank seminar participants at the Institute for Evaluation of Labour Market and Education Policy and the Department of Economics, Uppsala University. Lisa Laun and Tobias Laun gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Foundation.

1. Introduction

The disability insurance program has become one of the largest income maintenance programs in modern welfare states.[1] Autor (2011) shows that, in the U.S., the share of 25–64 year olds receiving benefits from the Social Security Disability Insurance (SSDI) increased from 2.3 to 4.6 percent between 1989 and 2009, and the SSDI Trust Fund is projected to be exhausted between 2015 and 2018. Sweden has also experienced large growth in disability benefits recipiency. Between 1985 and 2008, the share of 30–64 year olds receiving disability benefits increased from 8 to 12 percent. The variation in program inflow over this period was large, but the reasons for the fluctuations are not fully understood.

Policy makers have two main instruments for governing the inflow to the disability insurance program. Program *generosity* affects the attractiveness of the program and can be modified through changes in the compensation level or the cap on benefits. Program *availability* affects the likelihood of being awarded benefits and depends on screening stringency, which includes formal eligibility criteria as well as the implementation of program rules. While eligibility criteria are easily modified at the political level, the control of the implementation process is likely to be imperfect. Caseworker discretion and internal processes at the Social Insurance Agency are important in the eligibility determination process but are difficult for the insurance provider to monitor. Because of variation in implementation over time, the control of program inflow may change also during periods when no reforms at the political level are being pursued.

A key question is whether fluctuations in the inflow to the disability insurance program can be explained by changes in the underlying health of the population, or whether they are due to changes in the control of program inflow. In this paper, we propose a strategy for assessing how the inflow to the disability insurance program has been governed over time, by studying the development of the relative health of new beneficiaries compared to non-beneficiaries. We analyze the ex-ante health of individuals entering the disability insurance program by using the ex-post mortality from all years in which the individuals are observed in the future, compared to individuals not entering the disability insurance program in the same year. This approach allows us to control for changes in the underlying health when estimating fluctuations in program entry.

The association between population health and disability insurance recipiency has previously been assessed in Wise (2012). This paper provides a more

[1] See, e.g., Wise (2012) for a detailed description of disability insurance programs in several developed countries.

formal framework for analyzing how program entry has been governed, after accounting for health changes. Our strategy captures the development of screening stringency due to changes in formal eligibility criteria as well as in implementation. The latter aspect has been difficult to assess in previous studies. Also other factors may affect the relative health of new disability beneficiaries compared to non-beneficiaries. As discussed above, program generosity is a different policy instrument for governing program inflow. However, there are no changes in the benefit level over the studied period. The attractiveness of outside options may also be important for the application decision. The impact of labor market factors will be included in the estimates, and our measure can be taken as to capture the ability of the disability insurance program to withstand the influence of outside forces.

To structure the analysis, we first provide a stylized theoretical model of the application decision to the disability insurance program. We focus on screening stringency as a key determinant of program inflow. A potential indicator of screening stringency is the denial rate. Since Parsons (1991), however, it is well known that initial eligibility determination is an important self-screening mechanism. His results suggest that a 10 percent increase in the initial denial rate induced a 4 percent decrease in application rates after 2 years. Halpern and Hausman (1986) show that the probability of acceptance has a significant effect on the probability of applying for disability benefits. The denial rate is therefore an imperfect measure for studying the development of screening stringency over time. Also in our model we show that the composition of the applicant pool changes with screening stringency, and the effect of stringency on the denial rate is therefore undetermined. Going further, we show that screening stringency has implications for the relative health of new beneficiaries to non-beneficiaries. We derive a relationship showing that the relative health of new beneficiaries to non-beneficiaries improves with a reduction in stringency.

We then outline the empirical strategy for assessing how the inflow to the disability insurance program has been governed over time. We analyze the ex-ante health of individuals entering the disability insurance program by using the ex-post mortality from all years in which the individuals are observed in the future, compared to individuals not entering the disability insurance program in the same year. We analyze the ex-ante health of individuals entering the disability insurance program by using the ex-post mortality from all years in which the individuals are observed in the future, compared to individuals not entering the disability insurance program in the same year. We do this using duration analysis, in which the differential development of health across cohorts is accounted for by estimating cohort-specific baseline mortality hazards. The strategy can be used to assess how program inflow has been governed in

general as well as across population groups. We apply the proposed strategy to the disability insurance program in Sweden between 1985 and 2008. Sweden is an interesting application since the inflow to the disability insurance program has fluctuated considerably over time, in a fashion that cannot always be related to changes in formal program rules. We study the development of the relative health of new beneficiaries compared to non-beneficiaries at the time of formal program changes, but are also able to study the development during periods when no formal program changes have been pursued.

We find that changes in the mortality ratio of new disability beneficiaries compared to non-beneficiaries correspond well to the fluctuations in the disability benefit entry rate over time. This suggests that the variation in program inflow cannot be accounted for by changes in the underlying health of the population. Changes to formal eligibility criteria are indeed reflected in the mortality ratio. The removal of eligibility for pure labor market reasons for workers aged 60–64 in 1991 can be seen in a substantial and lasting increase in the relative mortality ratio of older compared to younger beneficiaries. Also the removal of the special elderly rules and eligibility for labor market and medical reasons combined in 1997 is reflected in the relative health of new beneficiaries across age groups, although this reform seems to have had more of a transitory impact. During the period under study, program inflow was more strictly governed in Stockholm compared to other regions, but the relative mortality ratios across regions have converged after the possibilities of taking labor market reasons into account were removed in the 1990s.

We also find fluctuations in the mortality ratio of new beneficiaries compared to non-beneficiaries during periods of no formal program changes. The large increase in program entry during the early 2000s is related to a decrease in the relative mortality of new beneficiaries, and the large inflow of women compared to men during this period corresponds to a relatively lower mortality ratio of female compared to male beneficiaries. Perhaps most strikingly, the rapid fall in program entry since 2005 is reflected in a substantial increase in the mortality ratio of new beneficiaries compared to non-beneficiaries, although no formal changes to the program were being pursued. The turning point in 2005 coincides with the re-organization of the regional Social Insurance Agency offices into one central authority.

Screening stringency has been recognized as one of the most important determinants of the growth of the disability insurance program.[2] Previous studies have mainly analyzed the effects of known changes to screening stringency, affecting the eligibility criteria or the implementation of program rules, and have to a large extent concluded that such changes are important for pro-

[2]See, e.g., Bound and Burkhauser (1999) and Duggan and Imberman (2009).

gram growth. To our knowledge this is the first paper that studies changes in screening stringency over time. Several studies have analyzed the effects of changes to formal eligibility criteria. Autor and Duggan (2003) use the 1984 liberalization of the disability determination process in the U.S. together with an unforeseen increase in the effective replacement rate and declined demand for less skilled workers, and find a substantial impact of these forces on labor force participation of high school dropouts. Karlström, Palme and Svensson (2008) study the removal of looser eligibility criteria for workers aged 60 to 64 in Sweden in 1997, and find no effect on employment but a spillover effect to other social insurance programs. Staubli (2011) analyzes the effect of tightened eligibility criteria for men above a certain age in Austria in 1996. He finds a substantial decline in disability enrollment and an increase in employment, but also spillover effects into the unemployment and sickness insurance programs.

Other studies have exploited variation in the strictness of the eligibility screening process. Gruber and Kubik (1997) use the disability funding crisis in the U.S. in the late 1970s, which induced a sharp increase in initial denial rates. Exploiting variation across U.S. states, they find that stricter screening significantly reduced labor force non-participation among older men. De Jong, Lindeboom and van der Klaauw (2011) use a large-scale experiment in the Netherlands in which caseworkers in certain regions were instructed to increase the screening of disability benefit applications. They find a reduction in long-term sickness absence and disability insurance applications from stricter screening, without negative spillover effects on the utilization of the unemployment insurance.

The paper proceeds as follows. Section 2 describes the disability insurance program in Sweden. Section 3 presents the theoretical model of the application decision to the disability insurance program and a numerical simulation of the model. Section 4 outlines the empirical strategy and Section 5 describes the data. Section 6 presents the results and Section 7 concludes the paper.

2. The Swedish Disability Insurance Program

The Swedish disability insurance program replaces foregone earnings due to a lasting reduced working capacity for workers aged 19–64. The basis in the eligibility determination process is a clear relationship between medical causes and the reduction in working capacity in order to qualify for disability benefits. During certain time periods and for certain groups of workers consideration could also be given to the individual's labor market situation. From 1972 to October 1991, individuals aged 60 to 64 could be granted disability benefits for pure labor market reasons, without a health assessment, if they were still un-

employed when reaching the time limit in the unemployment insurance. From 1970 to January 1997, special rules for workers aged 60 to 64 further implied less strict health assessments, no test of employability and lower requirements for changing occupation or area of residence in order to find work. During the same time period, workers of all ages could be granted disability benefits for labor market reasons and health reasons combined, if they suffered from a reduced working capacity for medical reasons and had been unemployed for 1–2 years.

Since 1997, individual job opportunities should not be taken into account in the eligibility determination for disability benefits. In 2003, the disability insurance program was moved from the public pension system to the social insurance system as part of a public pension reform, but the formal eligibility criteria were unchanged. In 2005, the 21 regional offices of the Social Insurance Agency were integrated into one central authority. In the same year, the reassessment of eligibility for disability beneficiaries became stricter. In July 2008, the eligibility criteria to the program were substantially tightened.

Disability benefits can be granted part-time or full-time, depending on the extent of the work impairment. The compensation rate is 64 percent of the assumed income, up to a ceiling. The level of benefits has remained fairly constant over time. Before 2003, benefits were calculated according to the formula that applied to old-age pension. The average level of compensation was, however, similar before and after 2003. Disability benefits can be supplemented with payments from occupational insurances, covering the majority of Swedish workers.[3] The total compensation rate, including occupational insurances, is about 80 percent for workers in the private and local government sector and about 85 percent for workers in the central government sector, for earnings below the ceiling. Other social insurance programs include the sickness insurance, which replaces foregone earnings due to a temporarily reduced working capacity, and the unemployment insurance. Taking occupational insurances into account, disability benefits compare well with the level of sickness and unemployment benefits. The re-assessment of working capacity is also least strict in the disability insurance program, making it a desirable route for individuals who want to stop working permanently.

The development of program participation in the disability insurance is presented in Figure 4.1. The solid line, plotted against the left axis, shows the share of individuals aged 30–64 receiving disability benefits. The dashed line, plotted against the right axis, shows the disability benefit entry rate among

[3] Sjögren Lindquist and Wadensjö (2007) estimate that 96 percent of Swedish workers are covered by a collective agreement allowing for occupational insurance, and that almost all of these workers fulfill the criteria for receiving occupational insurance. Between 60 and 80 percent of workers claim occupational insurance when receiving disability benefits.

the individuals not already receiving benefits. The vertical dashed lines mark the reforms to formal eligibility criteria in 1991 and 1997, discussed above. Between 1985 and 2008, the share of disability beneficiaries increased from just over 8 percent to about 12 percent, but the fluctuations in the entry rate were large. The rate of entry was high during the late 1980s and early 1990s, declined during the mid 1990s and increased rapidly during the late 1990s and the early 2000s. Since the mid 2000s, there has been a remarkable drop in the disability benefit entry rate, which has even resulted in a decline in the participation rate in recent years.

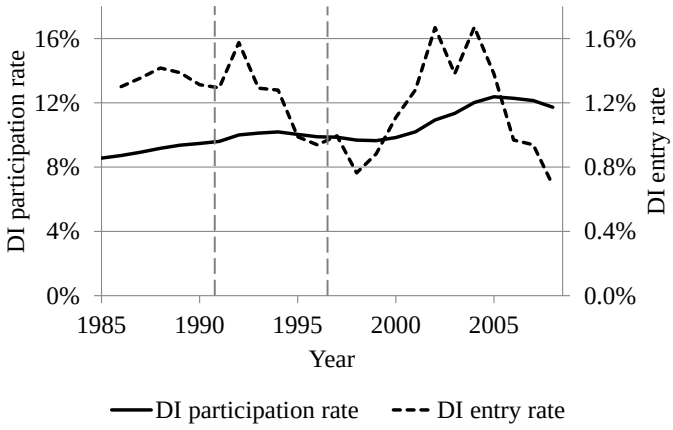
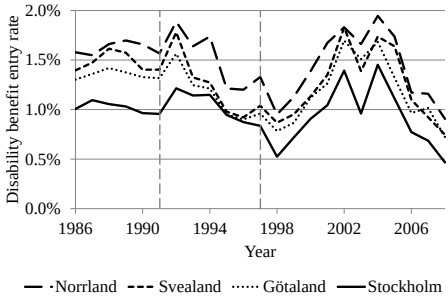


Figure 4.1: The share of individuals receiving disability benefits and the disability benefit entry rate in ages 30–64 in Sweden, 1985–2008

The reasons for the fluctuations in the disability benefit entry rate over time are not fully understood. The decline in the entry rate in the early 1990s coincided with the removal of the pure labor market reasons for workers aged 60 to 64 in 1991, but the adaption did not appear immediately. There were also fluctuations in the entry rate around the reform in 1997, when the special rules for workers aged 60 to 64 and the eligibility rules for health and labor market reasons combined were removed. The large increase during the early 2000s and the subsequent decline since the mid 2000s cannot be attributed to any formal program changes. The integration of the 21 regional offices of the Social Insurance Agency into one central authority in 2005 might have affected the internal processes within the agency and thereby the entry rate. It is apparent that the large drop in program entry appeared well before the tightening of the formal eligibility criteria in 2008.

(a) The disability benefit entry rate in the age groups 30–59 and 60–64

(b) The disability benefit entry rate among men and women in ages 30–64



(c) The disability benefit entry rate in ages 30–64 by region

Figure 4.2: The inflow to the disability insurance program in Sweden for different population groups, 1986–2008

The development of program entry for different groups of the population, presented in Figure 4.2, can shed further light on program growth. Several of the changes to formal eligibility criteria concerned the age group 60 to 64. Figure 4.2(a) shows the disability benefit entry rate in the age groups 30 to 59 and 60 to 64. Following the removal of the pure labor market reasons for workers aged 60 to 64 in October 1991, the entry rate in this age group fell from between 6 and 7 percent between 1986 and 1992 to just above 2 percent in 1998. The fluctuations in the entry rate around the removal of the special elderly rules in 1997 also appeared exclusively for the older age group. Since 1998, the pattern of the entry rates has been similar across the two age groups, increasing during the late 1990s and early 2000s, and decreasing since the mid 2000s. In the mid 2000s, the entry rate among 30 to 59 year olds was higher than ever before.

Figure 4.2(b) shows the disability benefit entry rate from 1986 to 2008, separately for men and women. Women have been more likely to be awarded disability benefits throughout the period. Between 1986 and 1998, the development of entry rates of women and men was very similar. The increase in awards between 1998 and 2005, however, was much larger for women. The subsequent decline in program entry since 2005 has again led to a convergence in entry rates across gender.

The development of program entry has also varied across Swedish regions. We have divided Sweden into four large regions: the northern part of Sweden, Norrland; the middle part of Sweden, Svealand (excluding Stockholm); the southern part of Sweden, Götaland; and the greater Stockholm area. Figure 4.2(c) presents the development of the disability benefit entry rates across these four regions. The entry rate has been lowest in the Stockholm region and highest in Norrland throughout the period, whereas Svealand and Götaland have had intermediate entry rates.

3. The Model

In this section we present a stylized model of an individual's application decision to the disability insurance program. We show that the application decision, and with that, the composition of the applicant pool, are endogenous to screening stringency. In a numerical simulation we analyze the relationship between screening stringency and the relative health of new beneficiaries to non-beneficiaries. We further introduce mortality into the model and examine how screening stringency affects the relative mortality of new beneficiaries to non-beneficiaries.

3.1. Environment

In the model, an individual decides whether to apply for disability benefits or whether to continue working. If the individual applies and the individual's health is lower than the disability benefit threshold, the individual receives disability benefits. If the application is rejected, the individual is assumed to return to employment. There is an application cost which is increasing in the individual's wage.

This model is close to the self-screening setup of Parsons (1991). In his environment, an individual decides whether or not to apply for disability benefits. The acceptance probability is higher for a truly disabled individual than it is for an able one. The cost of applying consists of foregone wage earnings since the individual is not allowed to work while the application is being decided.

Let m_{it} be the disability benefit threshold that individual i faces at calendar time t. This threshold can be expressed as

$$m_{it} = m_t + \mu_{it},$$

where m_t is the average disability benefit threshold at time t and μ_{it} is an idiosyncratic error term, with zero mean, a cumulative distribution function $F(\cdot)$ and a probability density function $f(\cdot)$. The average disability benefit threshold, m_t, is known by the individual.

When the individual applies for disability benefits, the caseworker can observe the individual's health, h_{it}. If the health is below the disability benefit threshold the individual faces, m_{it}, then the individual is awarded disability benefits. Hence,

$$D_{it}^{AC} = \begin{cases} 1 & \text{if} \quad h_{it} - m_t < \mu_{it} \\ 0 & \text{if} \quad h_{it} - m_t \geq \mu_{it}, \end{cases}$$

where D_{it}^{AC} is an indicator of individual i's application for disability benefits at time t being accepted.

The expected value from applying for disability benefits is given by

$$V^{AP}(h_{it}, m_t, w_{it}) = p(h_{it}, m_t)V^{AC}(h_{it}, w_{it}) + [1 - p(h_{it}, m_t)]\, V^{RJ}(h_{it}, w_{it}) - C(w_{it})$$

where $V^{AC}(\cdot)$ is the value of the application for disability benefits being accepted, $V^{RJ}(\cdot)$ is the value of the application being rejected, $C(\cdot)$ is the application cost and

$$p(h_{it}, m_t) = \Pr(h_{it} - m_t < \mu_{it}) = 1 - F(h_{it} - m_t),$$

is the probability of the application being accepted. It follows that

$$p_h(h_{it}, m_t) \equiv \frac{\partial p(h_{it}, m_t)}{\partial h_{it}} = -f(h_{it} - m_t) < 0$$

and

$$p_m(h_{it}, m_t) \equiv \frac{\partial p(h_{it}, m_t)}{\partial m_t} = f(h_{it} - m_t) > 0.$$

That is, the probability of obtaining disability benefits is decreasing in health and increasing in the average disability benefit threshold.

The value of an accepted application for disability benefits, V^{AC}, is a function of the individual's health, h_{it}, and the wage, w_{it}. The individual's consumption, when receiving disability benefits, is given by δw_{it}, where δ is the compensation rate. The other source of utility is the individual's health, h_{it}. The value of receiving disability benefits is therefore increasing in all arguments.

The value of a rejected application for disability benefits, V^{RJ}, is a function of the health, h_{it}, and the wage, w_{it}. Health generates utility and V^{RJ} is therefore increasing in h_{it}. We assume that the individual returns to employment when being denied disability benefits. The value of being denied disability benefits is therefore increasing in the wage.

The application cost is increasing in the wage. This reflects the fact that in order to apply for disability benefits, for example in the U.S., the applicant has to leave his or her current employment and forego wage earnings. In Sweden, many individuals are on sick-leave when applying for disability benefits, which is more costly for high income earners. Hence,

$$C_w(w_{it}) \equiv \frac{\partial C(w_{it})}{\partial w_{it}} > 0.$$

3.2. Solution

We can now determine when it is optimal for the individual to apply for disability benefits. The individual applies for disability benefits if the difference between the value of applying and the value of working is positive, i.e., if

$$\Delta(h_{it}, m_t, w_{it}) \equiv V^{AP}(h_{it}, m_t, w_{it}) - V^{WO}(h_{it}, w_{it}) > 0,$$

where the value of working is increasing in both health and wage.

We assume that in the worst health state it is optimal for the individual to apply for disability benefits while in the best health state it is optimal to work. This assumption is not very restrictive and can be motivated by a disutility of working which is decreasing in health. In other words, a healthy individual has a low disutility of working as well as a low probability of receiving disability benefits. This makes working more attractive than applying for disability benefits. An individual in bad health, on the other hand, faces a large disutility of working while at the same time having a high probability of applying

successfully, which makes applying for disability benefits more attractive than working. Defining $\underline{h}$ as the lowest and $\overline{h}$ as the highest possible health state, we can express these assumptions as

$$\Delta(\underline{h}, m_t, w_{it}) > 0 \quad \text{and} \quad \Delta(\overline{h}, m_t, w_{it}) < 0. \tag{4.1}$$

In order to show that there exists only one level of health for which the individual is indifferent between applying for disability benefits and working, we are interested in how the difference between the value of applying and the value of working changes with health

$$\begin{aligned}\frac{\partial \Delta(h_{it}, m_t, w_{it})}{\partial h_{it}} =& p_h(h_{it}, m_t)\left[V^{AC}(h_{it}, w_{it}) - V^{RJ}(h_{it}, w_{it})\right] \\ &+ p(h_{it}, m_t)\left[V_h^{AC}(h_{it}, w_{it}) - V_h^{RJ}(h_{it}, w_{it})\right] \\ &+ V_h^{RJ}(h_{it}, w_{it}) - V_h^{WO}(h_{it}, w_{it}).\end{aligned}$$

In order to determine the sign of this derivative, we need to make two assumptions. The first one states that the value of an accepted application for disability benefits is larger than the value of a rejected application,

$$V^{AC}(h_{it}, w_{it}) > V^{RJ}(h_{it}, w_{it}). \tag{4.2}$$

The other assumption concerns the marginal utility of health. We assume that

$$V_h^{AC}(h_{it}, w_{it}) \leq V_h^{RJ}(h_{it}, w_{it}) \leq V_h^{WO}(h_{it}, w_{it}). \tag{4.3}$$

In other words, a marginal increase in health is least valuable for an individual receiving disability benefits and it is most valuable for a working individual.

These two assumptions together with the fact that $p_h(\cdot) < 0$ allow us to determine that the difference between the value of applying and the value of working is decreasing in health, i.e.,

$$\frac{\partial \Delta(h_{it}, m_t, w_{it})}{\partial h_{it}} < 0.$$

Assumptions (4.1)–(4.3) guarantee the existence of a health level below which it is optimal to apply for disability benefits and above which it is optimal to continue working. Let us define this health cut-off level as $h^{CO}(m_t, w_{it})$, where $\Delta(h^{CO}(m_t, w_{it}), m_t, w_{it}) = 0$. The application decision can then be formalized as follows

$$D_{it}^{AP} = \begin{cases} 1 & \text{if} \quad h_{it} < h^{CO}(m_t, w_{it}) \\ 0 & \text{if} \quad h_{it} \geq h^{CO}(m_t, w_{it}), \end{cases}$$

where D_{it}^{AP} is an indicator of individual i applying for disability benefits at time t.

In order to analyze how the composition of the applicant pool changes when stringency changes, we need to know how this cut-off level changes with the average disability benefit threshold, m_t. To determine this we compute how a change in m_t affects the difference between the value of applying and the value of working. We get that

$$\frac{\partial \Delta(h_{it}, m_t, w_{it})}{\partial m_t} = p_m(h_{it}, m_t) \left[V^{AC}(h_{it}, w_{it}) - V^{RJ}(h_{it}, w_{it}) \right] > 0,$$

because of $p_m(\cdot) > 0$ and assumption (4.2). In other words, the difference between the value of applying and the value of working is increasing in m_t.

As illustrated in Figure 4.3, a reduction in stringency, i.e., an increase from m_t to m_t', pushes the difference between the value of applying for disability benefits and the value of working upwards, from $\Delta(h_{it}, m_t)$ to $\Delta(h_{it}, m_t')$. This implies that the health cut-off level increases from $h^{CO}(m_t)$ to $h^{CO}(m_t')$. We therefore know that

$$\frac{\partial h^{CO}(m_t, w_{it})}{\partial m_t} > 0. \tag{4.4}$$

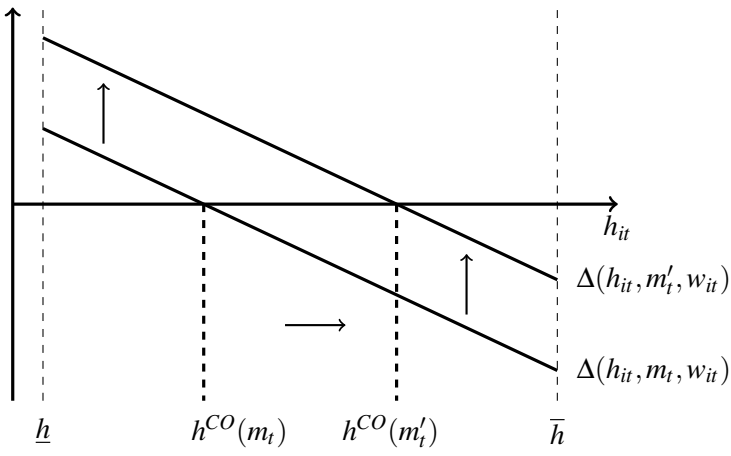


Figure 4.3: Change of the health cut-off as stringency decreases

As can be seen in Figure 4.3, the pool of disability benefit applicants increases from all individuals whose health falls in the interval $[\underline{h}, h^{CO}(m_t)]$ to all individuals with objective health in $[\underline{h}, h^{CO}(m_t')]$. In other words, when stringency decreases, it becomes more attractive to apply for disability benefits and individuals who were previously too healthy to apply find it now worthwhile

doing so. The average health among applicants therefore increases when stringency decreases. This change resembles the results of Parsons (1991) who shows that the composition of disability benefit applicants depends, among other things, on the stringency of the screening process.

3.3. Numerical Simulation

For the empirical strategy we propose in the next section, we are interested in what happens to the health of disability benefit recipients, i.e., to applicants who are granted disability benefits, relative to the health of non-recipients. In the empirical analysis we use mortality as a measure of health. We therefore introduce mortality into the model and examine how screening stringency affects the relative mortality of new beneficiaries to non-beneficiaries. To analyze this relationship, we simulate the model numerically and make the following assumptions on functional forms

$$\begin{aligned}
V^{AC}(h_{it}, w_{it}) &= \delta w_{it} + h_{it}, \\
V^{RJ}(h_{it}, w_{it}) &= \alpha w_{it} + h_{it} - \phi(1 + \overline{h} - h_{it}), \\
V^{WO}(h_{it}, w_{it}) &= w_{it} + h_{it} - \phi(1 + \overline{h} - h_{it}).
\end{aligned}$$

In other words, utility is linear in consumption as well as in health. When working, the individual incurs disutility of labor which is decreasing in health. An individual on disability benefits receives the compensation rate times the wage as a benefit. When being denied disability benefits, the individual returns to employment and earns the wage. However, there can be a wage penalty, $\alpha \leq 1$, associated with having applied for disability benefits. Applying for disability benefits might signal low productivity to the employer. Also, there could be stigma associated with application to the disability insurance program.

The cost of applying for disability benefits is assumed to be linear in the wage

$$C(w_{it}) = \psi w_{it}.$$

The idiosyncratic error term, μ_{it}, is assumed to be logistically distributed and therefore

$$p(h_{it}, m_t) = \frac{1}{1 + \exp(\gamma(h_{it} - m_t))}.$$

The parameter γ is set so that the application of an individual in the worst health is accepted with a probability of 95% at the lowest stringency level.

The health of individual i at age a is given by

$$h_{ia} = 1 - \exp(-\kappa(100 - a) + \varepsilon_i) \quad \text{for} \quad a = 30, \ldots, 100,$$

where

$$\varepsilon_i \sim N(0, \sigma^2).$$

An individual in this model survives from one period to the next as long as her health is above a threshold, i.e., as long as $h_{ia} \geq q_1$. However, there is also a small probability of dying, $q_2(a)$, which is independent of health and increasing in age. All parameters associated with the health process are calibrated to replicate the survival probabilities that we see in the data. Figure 4.4 plots the age-dependent probability of dying in the upper panel. In the lower panel we compare the probability of surviving from one period to the next, conditional on living to this period, in our model with the corresponding numbers in the data.

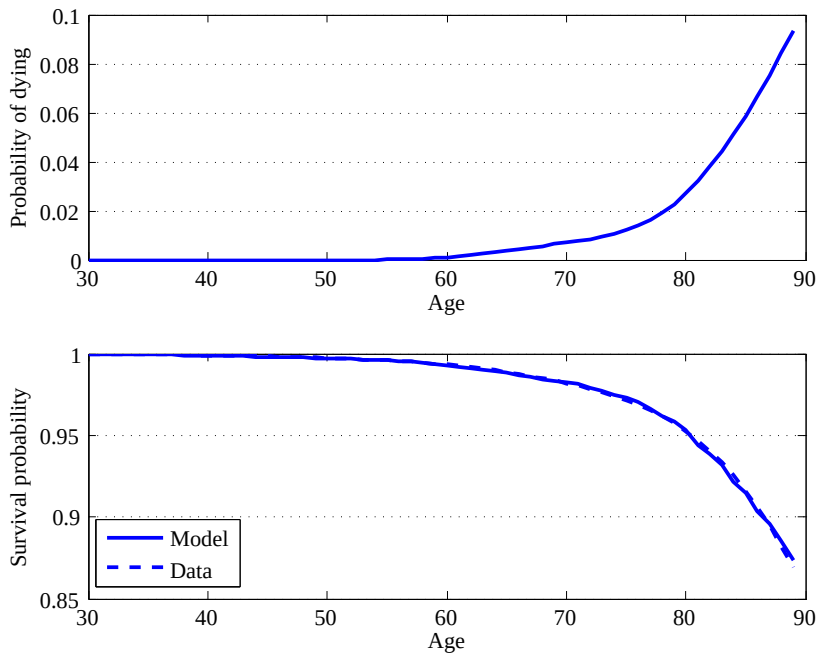


Figure 4.4: Probability of dying and conditional survival probabilities

All parameter values used to simulate the model are presented in Table 4.1. For the purposes of this analysis we keep the wage constant over time and identical for all individuals, i.e., $w_{it} = w$. The parameters are chosen such that assumptions (4.1)–(4.3) are fulfilled.

Each period $n = 1000$ individuals aged 30 enter the model. They live at most to the age of 100. We draw a health shock, ε_i, for each individual as well as an error term, μ_{it}, for each individual in each period. We then simulate the model for each level of stringency $m = 0, 0.05, \ldots, 1$.

Parameter	Value	Description
α	0.95	Wage penalty
δ	0.8	Disability compensation rate
$\overline{h}$	1	Perfect health
$\underline{h}$	0	Worst health
κ	0.02	Health parameter
ϕ	1.25	Disutility of labor multiplier
ψ	1	Application cost multiplier
q_1	0.05	Survival cut-off
σ	0.5	Standard deviation of health shock
w	1	Wage

Table 4.1: Parameter values

In Figure 4.5 we show how the health cut-off and the average health of individuals applying for disability benefits as well as of individuals accepted to the disability insurance program vary with the stringency level m.

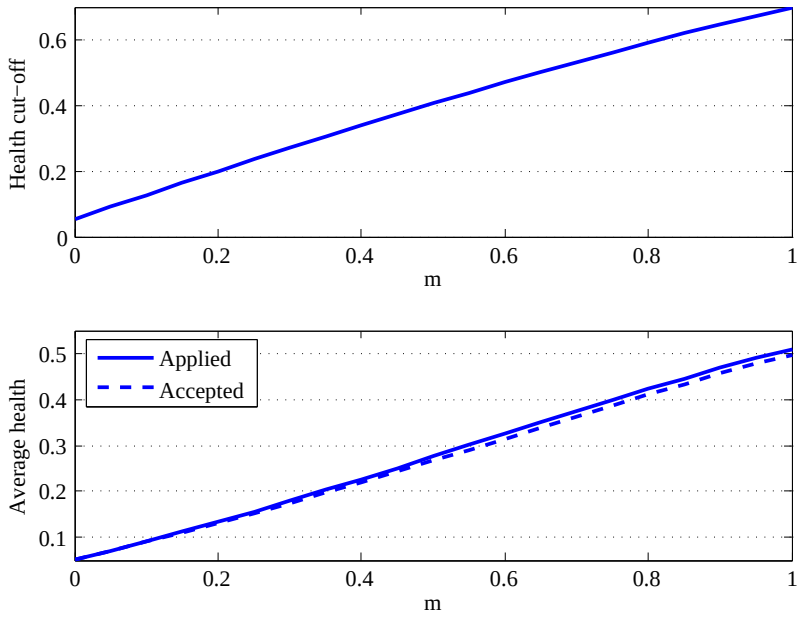


Figure 4.5: Simulation of the health cut-off and the average health for different levels of stringency

Recall that an increase in m corresponds to a decrease in stringency. In the upper panel of Figure 4.5 we see the increase in the health cut-off described in condition (4.4). As stringency decreases, the health level below which it is optimal to apply to disability benefits increases. As a consequence we see that the average health among disability benefits applicants and recipients increases

with lower stringency.

In Figure 4.6 we compare average health and mortality of disability benefits applicants to non-applicants and of recipients to non-recipients. Here, mortality is measured by using longevity in the form of the difference between the maximum number of years alive and the actual number of years alive.[4]

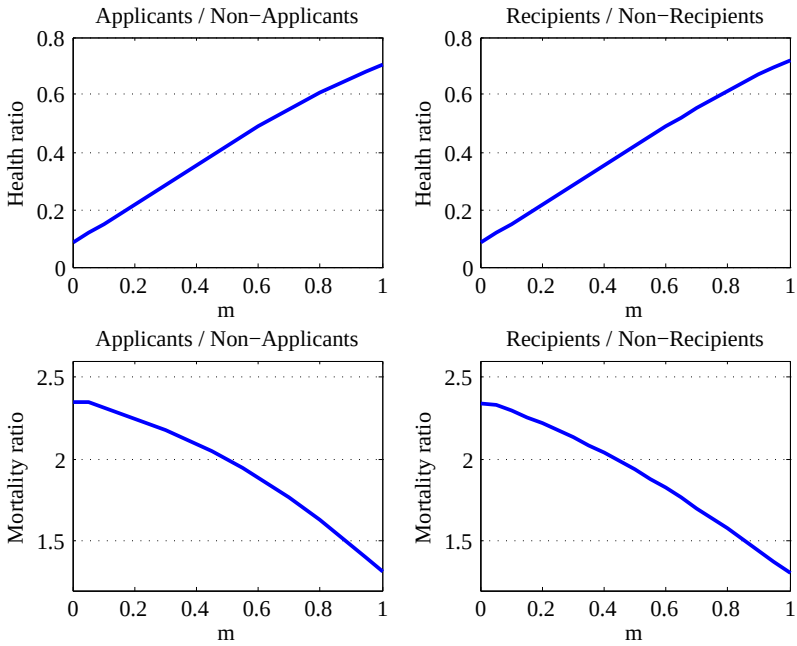


Figure 4.6: Comparison of the health and mortality of disability benefit applicants to non-applicants and of disability benefit recipients to non-recipients

In the upper left panel of Figure 4.6 we see the average objective health of disability insurance applicants relative to the average objective health of non-applicants. We see that, as stringency decreases, the disability insurance applicants become relatively healthier. In the upper right panel we see that the same holds for disability benefits recipients relative to non-recipients.

The two lower graphs in Figure 4.6 show the average mortality of disability insurance applicants relative to non-applicants on the left side and disability insurance recipients relative to non-recipients on the right side. We can see that mortality of applicants and recipients relative to non-applicants and non-recipients decreases when stringency decreases.

[4]Recall that agents enter the model at age 30 and die with certainty at age 100. The maximum number of years alive is therefore 71.

3.4. Alternative Explanations

In the above description, we focused on the relation between screening stringency and the health of disability beneficiaries. The result that we derived, that the average health of disability insurance recipients relative to non-recipients declines with screening stringency, can be obtained also by changing other parameters of the model. Program generosity is a different instrument that policy makers can use to change the attractiveness of the disability insurance program. The compensation rate is represented by the parameter δ in the model, and decreasing the benefits produces the same result as increasing screening stringency. Also factors related to the attractiveness of outside options may be important. In the model, the wage is constant and equal for all individuals, and there is no risk of unemployment. An increased value of working relative to collecting benefits, due to, for example, increased wages or improved employment prospects, could also lead to a decline in the relative health of disability beneficiaries compared to non-beneficiaries. The implications of this for the results in this paper are further discussed below.

4. Empirical Strategy

In the theoretical model, we derived that the health of new disability applicants improves with reduced stringency. This is an intuitive result. As stringency decreases, healthier individuals are induced to apply for benefits. The same result may appear, however, because of other changes affecting the attractiveness of the disability insurance program, such as increased benefits generosity or a decline in labor market opportunities. In this paper, we analyze how the inflow to the disability insurance program has been governed over time, which includes all of these forces. With respect to screening stringency, a strength of the strategy is that it captures whether changes to formal eligibility criteria impact the eligibility determination process, and whether there is variation in the implementation of program rules during periods when no formal program changes have been pursued. Regarding benefit generosity, we know that the compensation rate has not substantially changed during the period under study. With respect to labor market opportunities, our approach can be argued to capture the ability of the disability insurance program to withstand the influence of outside forces.

To study how the inflow to the disability insurance program has been governed over time, we are interested in the relative health of new disability beneficiaries to non-beneficiaries. However, health is difficult to measure. Self-rated health measures are subjective and likely to suffer from measurement error. Furthermore, self-rated health tends to be endogenous to labor supply choices,

in the sense that individuals receiving disability benefits are more prone to report an impaired work capacity.[5] An objective measure of health that is accurately measured through population censuses and is difficult to fake is mortality. The immediate mortality of new disability beneficiaries is certainly low. However, we can analyze the ex-ante health of individuals entering the disability insurance program by using the ex-post mortality from all years in which the individual is observed in the future, compared to individuals not entering the disability insurance program in the same year. In Figure 4.6 we see that the relative mortality of new beneficiaries to non-beneficiaries is decreasing as stringency decreases.

The strategy is employed using duration analysis. First, we make an assumption about the distribution of longevity. We could assume a specific form of the underlying distribution of time to death and perform the analysis within a full maximum likelihood framework. Instead we assume that, under constant control of program inflow, the mortality of new disability beneficiaries at each age is proportionally related to the mortality of non-beneficiaries. For a number of known longevity distributions this assumption is the same as assuming that there is a constant difference in the mean log longevity between the two groups. This will, e.g., be the case when longevity follows the Weibull and the exponential distribution. Thus the proportionality assumption allows us to estimate the mortality ratio without completely specifying the longevity distribution. The mortality hazard of new disability beneficiaries, $\lambda(a)$, is then given by

$$\lambda(a) = g(a)\exp(\beta),$$

where $g(a)$ is the baseline hazard at age a and $\exp(\beta)$ is the mortality ratio of new disability beneficiaries compared to non-beneficiaries. We have that $\exp(\beta) > 1$ if the mortality of disability insurance beneficiaries is larger than the general mortality at each age.

The mortality hazard is likely to change across cohorts due to, for example, the expansion of education, changes in nutrition or advances within the health care system. This can be accounted for within the duration analysis framework by stratifying the baseline mortality hazard by cohort. Another reason to stratify the baseline hazard by cohort is to solve the problem of length-bias sampling, that comes from the fact that we observe different cohorts during different ages. We do this by estimating the Cox proportional hazard model

$$\lambda_{ic}(a) = g_c(a)\exp(\beta_t D_{it}(a)), \quad i = 1,...,n, \quad t = 1,....T, \tag{4.5}$$

where $\lambda_{ic}(a)$ is the mortality hazard of individual i in cohort c at age a and $g_c(a)$ is the baseline hazard of cohort c. $D_{it}(a)$ is a step function that takes

[5] See, e.g., Burkhauser, Daly, Houtenville and Nargis (2002).

the value one after individual i begins receiving disability benefits at age a in year t, else zero. Hence, t indicates the year in which the individual enters the disability insurance program. For an individual who is not receiving disability benefits during the studied time period, $D_{it}(a)$ is zero for all ages. Variation in the estimated coefficients β_t over time indicates that there were changes in how the inflow to the disability insurance program was governed.

Identification of β_t comes from the within cohort comparison of mortality hazards at age a for those who are awarded disability benefits at calendar time t compared to those who have *not yet* obtained disability benefits. The individuals in the comparison group may, however, begin receiving disability benefits later if eligible then. Since individuals are no longer eligible for disability benefits above age 64, we censor the observations above this age.[6]

We can also use the model to compare how program inflow has been governed across different groups of the population. To study the development across age groups, we estimate the model

$$\lambda_{ic}(a) = g_c(a)\exp(\beta_t D_{it}(a) + \delta_t^a D_{it}(a) I(a_i = a)), \quad i = 1,...,n, \quad t = 1,....T, \tag{4.6}$$

where the coefficients $\exp(\delta_t^a)$ are the mortality ratios of new beneficiaries to non-beneficiaries in age group a at time t compared to the mortality ratio in other age groups. To analyze the development across other types of populations, p, such as gender or region of residence, we assume a proportional health at each age a in the different populations. We then have that $g_c(a, p = 2) = g_c(a, p = 1)\exp(\delta_0)$, which gives the proportional hazard model

$$\begin{aligned}\lambda_{ic}(a) &= g_c(a)\exp(\delta_0 I(p_i = 2) + \beta_t D_{it}(a) + \delta_t^p D_{it}(a) I(p_i = 2)), \\ &i = 1,...,n, \quad t = 1,....T,\end{aligned} \tag{4.7}$$

where the coefficients $\exp(\delta_t^p)$ are the mortality ratios of new disability beneficiaries to non-beneficiaries in population group 2 at time t compared the mortality ratio in population group 1. The difference in mortality in general across the groups is captured by the parameter δ_0.

One potential criticism against the empirical strategy is that mortality is a bad measure of health at the individual level. Mortality can be argued to capture a different margin of health than work capacity. Furthermore, the development of mortality over time can be argued to capture advances within

[6]The reason is that we do not know if an individual aged, say, 66 would still have remained in the control group. Had the eligible age been 66 years, his or her health could have been such that he/she instead would have received disability benefits if applying. Keeping individuals in the control group after the age of 64 could lead to an attenuation of the estimated screening effect. However, when we include data above age 64 in the analysis, the patterns of the results are similar.

the health care system rather than the general health of the population. The strategy proposed here, however, requires only that there *is* a relation between health and subsequent mortality in the future at the group level. Mortality does not need to take place immediately after admittance to the disability insurance program, but at some point in the future. In addition, the strategy employs a within cohort comparison of mortality, and thus allows for advances in the health care system.

A second concern is that admittance to the disability insurance program may have a direct effect on health. In this paper, such an effect is not what we are interested in. Since our interest is how program inflow has been governed over time, this is a problem only if there are heterogeneous causal effects of disability insurance recipiency on the health of individuals. We regard this as an extreme case, but if so, this will also be part of our estimates.

5. Data

We use data covering all individuals in Sweden aged 30–64 from 1985 to 2008. The data originates from administrative records, collected and maintained by Statistics Sweden. For each individual, we observe birth year, gender and county of residence. We also have access to detailed spell data on the collection of disability benefits, provided by the Swedish Social Insurance Agency. For each year, we record if the individual receives disability benefits and whether it is the entry year to the program. Most individuals remain in the disability insurance program until retirement, once they enter. For individuals with multiple benefit spells, the first year of benefit collection is regarded as the entry year to the program.[7] Since the data is based on disability benefit payments, there might be a slight delay of the first payment compared to the decision to award benefits. Finally, we add information about mortality, provided by the National Board of Health and Welfare. Mortality is measured until 2010 or until the year the individual turns 64.

The sample used for estimation consists of individuals aged 30–64 between 1985 and 2008 who are not already receiving disability benefits during the first year in which they are observed. This includes individuals who are observed from the beginning of the sample period, in 1985, and are not already receiving disability benefits at that time, as well as individuals entering the sample by turning 30 during 1986–2008 and not already receiving disability benefits at that age. The sampling strategy implies that different cohorts will be observed for different periods of time and at different ages. This problem of length bias sampling is solved by stratifying the baseline hazard by cohort in

[7]This simplification leads to only 2.4 percent of the registrations of program participation not corresponding to the data.

the Cox regression model. In total, the sample consists of about 100 million observations.

The empirical strategy is used to study changes in how the inflow to the disability insurance program has been governed over time in the full population and across groups defined by age, gender and region of residence. As explained in Section 2, we divide Sweden into four large regions: the greater Stockholm area, the northern region Norrland, the middle region Svealand (excluding Stockholm) and the southern region Götaland. Region is determined by the county of residence, defined as the county in which the individual resided when first awarded disability benefits. For individuals never awarded benefits, the county of residence is defined as the county in which the individual lived for the most years during the period under study.

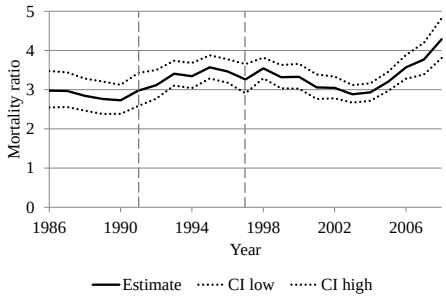
6. Results

This section presents the results from the empirical analysis. We first provide the estimation results of how the inflow to the disability insurance program has been governed over time, in general and for population groups defined by age, gender and region. Thereafter, we discuss the interpretation of the results. In particular, we are interested in the relation to the changes to formal eligibility criteria in 1991 and 1997, discussed in Section 2. Furthermore, we are interested in how the inflow to the program has been governed during periods when no reforms to formal program rules have been pursued.

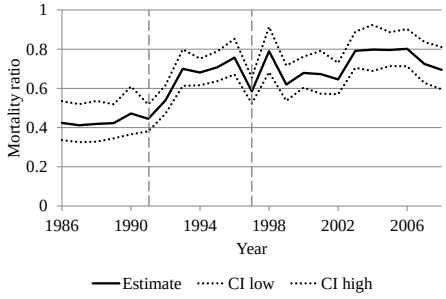
6.1. Estimation Results

Figures 4.7 and 4.8 present the estimation results of how the inflow to the disability insurance program in Sweden has been governed between 1986 and 2008. The solid lines plots the estimated mortality hazard ratios and the dotted lines plot the 95 percent confidence interval, with standard errors estimated by clustering by birth cohort. The vertical dashed lines mark the reforms to formal eligibility criteria that were discussed in Section 2. The first vertical line indicates the removal of eligibility for pure labor market reasons for workers aged 60 to 64 in 1991. The second vertical line indicates the removal of the special eligibility criteria for workers aged 60 to 64, and the removal of eligibility for labor market and health reasons combined for workers of all ages, in 1997.

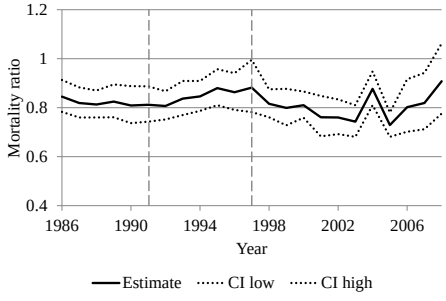
Figure 4.7(a) presents the main estimation results from estimating equation (4.5). The figure shows the estimated mortality ratio of new disability beneficiaries compared to non-beneficiaries from 1986 to 2008. The mortality ratio develops in a smooth fashion over time and reveals several interesting results.



(a) Estimated mortality hazard ratio of new disability beneficiaries compared to non-beneficiaries in ages 30–64



(b) Estimated mortality hazard ratio of new disability beneficiaries to non-beneficiaries in age 60–64 compared to age 30–59



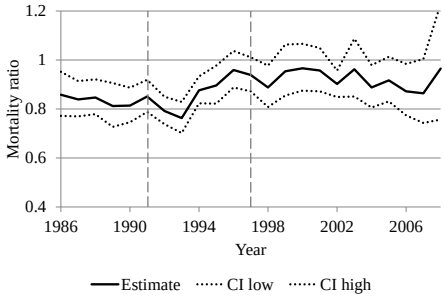
(c) Estimated mortality hazard ratio of new disability beneficiaries to non-beneficiaries for women compared to men in ages 30–64

Figure 4.7: Estimation results, 1986–2008

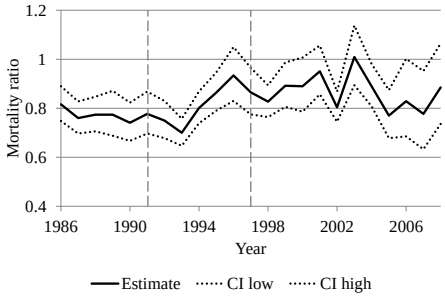
The overall picture is that substantial variation in program inflow remains after accounting for changes in the underlying health of the population. During the 1980s the mortality hazard of new beneficiaries in the disability insurance program was comparatively low, less than three times as large as that of non-beneficiaries. In the early 1990s, the relative mortality of new beneficiaries compared to non-beneficiaries increased to a higher level, and remained rather constant at the new level during the mid 1990s. Between 1998 and 2004, the mortality ratio steadily declined to a level comparable to that in the 1980s. From 2004, however, the mortality ratio rapidly increased to a higher level than ever before during the period under study. The mortality hazard of new disability beneficiaries in 2008 was more than four times higher than that of non-beneficiaries, suggesting that the access to the program was very restricted compared to before during the studied period.

Figure 4.7(b) presents the estimated mortality hazard ratio of new disability beneficiaries compared to non-beneficiaries in the age group 60–64 compared to the age group 30–59, as specified by equation (4.6). The relative mortality ratio is below one throughout the period, which suggests that old beneficiaries are in better health than young beneficiaries, compared to non-beneficiaries of the same age. The relative health changes substantially during the period, however, with the relative mortality ratio of older compared to younger beneficiaries ranging from 0.4 to 0.8. During the late 1980s, the relative mortality ratio of new beneficiaries compared to non-beneficiaries in the two age groups remained more or less constant at a low level. During the early 1990s, however, the mortality of older compared to younger beneficiaries increased rapidly. Thereafter, the mortality ratio of older compared to younger beneficiaries has fluctuated at the higher level throughout the studied period.

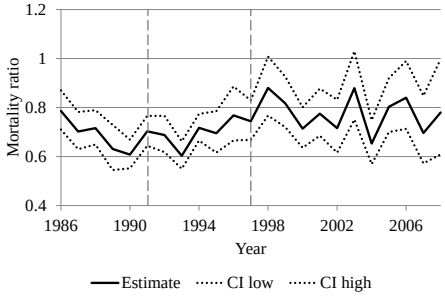
To study how program inflow has been governed across gender, we use the model presented by equation (4.7) and let $p_i = 2$ indicate being a woman. Figure 4.7(c) presents the estimation results, showing the estimated mortality hazard ratio of new beneficiaries compared to non-beneficiaries of women relative to men. The relative mortality ratio is below one throughout the period, suggesting that the health of new female beneficiaries has typically been better than that of male beneficiaries, compared to non-beneficiaries in the same gender group. Recall that we control for the relatively lower mortality of women compared to men in the population by estimating the parameter δ_0. During the late 1980s and the early 1990s, the relative mortality of new female compared to male beneficiaries was constant, while in the mid 1990s, the mortality ratios across gender slowly converged. Between 1997 and 2003, the mortality hazard ratio of female compared to male beneficiaries decreased to lower levels than ever before, suggesting a greater access to the program for women compared to men. The sudden spike in the relative mortality in 2004 is dif-



(a) Götaland



(b) Svealand



(c) Norrland

Figure 4.8: Estimated mortality hazard ratio of new disability beneficiaries to non-beneficiaries in (a) Götaland, (b) Svealand and (c) Norrland compared to the Stockholm region, ages 30–64, 1986–2008

ficult to explain but appears to be due to the transitory reduction in program entry during that particular year affecting women more than men. During the end of the period, from 2005 onwards, the relative mortality ratio increased to similar levels as in 1997. In 2008 there is, for the first time during the period under study, no statistically significant difference in the mortality ratio of new disability beneficiaries compared to non-beneficiaries across gender.

To study the development of the relative mortality of new disability beneficiaries to non-beneficiaries across Swedish regions over time, we estimate equation (4.7), with four different groups instead of two. We use Stockholm as the benchmark and simultaneously estimate the mortality ratio of new beneficiaries compared to non-beneficiaries in Götaland, Svealand (excluding Stockholm) and Norrland compared to Stockholm by including main effects and interaction terms for each region except for Stockholm. Figure 4.8 shows the estimated mortality hazard ratios of new beneficiaries compared to non-beneficiaries in each region, compared to the mortality ratio in the Stockholm region, during 1986–2008. The estimated mortality hazard ratios of Götaland, Svealand and Norrland compared to Stockholm are more volatile than the previous results, but still, an interesting pattern emerges. During the late 1980s and the early 1990s, the relative mortality of new beneficiaries compared to non-beneficiaries in the three other regions was significantly lower than in Stockholm. After 1993, the mortality of new beneficiaries converged towards the Stockholm level, in particular for Götaland and Svealand, and has fluctuated at a higher level throughout the period under study.

6.2. Interpretation

We can compare the estimation results of the mortality ratio of new beneficiaries to non-beneficiaries, presented in Figures 4.7 and 4.8, with the development of inflow to the disability insurance program, presented in Figures 4.1 and 4.2. The relatively stable entry rates in the end of the 1980s are reflected in a fairly constant mortality ratio in general as well as across groups defined by age, gender and region. During this period, the entry rate was high and the health requirements were low. Workers aged 60 to 64 could be eligible for disability benefits for pure labor market reasons, which appears to have substantially impacted how program inflow was governed. Older workers were awarded benefits almost eight times as much as younger workers during the late 1980s, and the mortality ratio of older compared to younger new disability beneficiaries was at the lowest level during the period under study.

In October 1991, the eligibility to disability benefits for pure labor market reasons for workers aged 60 to 64 was removed. This was followed by a substantial drop in program entry in general, as depicted in Figure 4.1. Fig-

ure 4.2 shows that the decrease was mainly driven by the age group 60–64, that women and men were similarly affected, and that the drop in entry was apparent in all four regions of Sweden. The drop in program entry is well reflected in a changed mortality ratio of new beneficiaries compared to non-beneficiaries, depicted in Figures 4.7 and 4.8. In particular, the mortality ratio of older compared to younger beneficiaries, shown in Figure 4.7(b), increased dramatically following the reform, which also affected the general mortality ratio, presented in Figure 4.7(a). The transitory spike in entry in 1992 is not reflected in a change in how program inflow was governed. Anecdotal evidence suggests that a shortage of caseworkers led to a queue of cases during the preceding years, and that the spike in 1992 is due to a shifting of assessments across years. This would explain that the mortality ratio is unaffected.

Interestingly, the removal of pure labor market reasons for older workers in 1991 was followed by a convergence in the relative mortality ratio of new beneficiaries compared to non-beneficiaries across Swedish regions, presented in Figure 4.8. During the late 1980s and the early 1990s, the mortality ratio of new beneficiaries compared to non-beneficiaries in Götaland, Svealand and Norrland was significantly lower than in Stockholm. These regions also had a higher unemployment rate and worse labor market prospects on average than the Stockholm region. After 1993, the mortality ratio converged towards the Stockholm level, in particular for Götaland and Svealand, and has since remained at a higher level. This suggests that the removed possibilities of taking labor market reasons into account affected eligibility determination and reduced the differences in the health of new beneficiaries across regions. Taken together, the removal of the eligibility criteria for older workers for pure labor market reasons in 1991 appears to have significantly impacted the government of program inflow. The change appeared rather smoothly over several years, however, reaching a new level of relative health in the mid 1990s. This suggests that it takes some time for formal program changes to impact the implementation process.

The next reform to formal eligibility criteria took place in 1997, when the looser eligibility criteria for workers aged 60 to 64 and the eligibility for labor market and health reasons combined were removed. Interestingly, this reform does not seem to have had a lasting impact on the relative health of new disability beneficiaries compared to non-beneficiaries. In Figure 4.2(a) there is substantial volatility in program entry for workers aged 60 to 64 around the reform. This transitory variation is well reflected in the relative mortality ratio of older compared to younger new disability beneficiaries, presented in Figure 4.7(b). In 1997, there is a substantial drop in the mortality ratio of older compared to younger beneficiaries. This captures the anticipation effect of the reform that was found in Karlström, Palme and Svensson (2008). Applications

under the old regime could be filed until 31 December 1996, which means that many of the award decisions were taken in 1997. The significant drop in the relative mortality ratio of new beneficiaries compared to non-beneficiaries across age groups in 1997 was followed by a significant spike in 1998, when the special rules had been removed. After that, the relative mortality ratio returned to the pre-reform level.

After the transitory responses to the 1997 reform followed a rapid increase in program entry into the disability insurance between 1998 and 2004, as shown in Figure 4.1. Figure 4.2 shows that the increase appeared both for younger and older workers and in all regions. The parallel development of program entry for women and men was broken, as program entry increased more rapidly for women than for men. The large increase in program entry during this period is reflected in a declining mortality ratio of new beneficiaries compared to non-beneficiaries, as shown in Figure 4.7(a). This may be due to informal changes to the screening of applicants during the early 2000s. Figure 4.7(b) shows that the relative mortality ratio of new beneficiaries compared to non-beneficiaries across age groups was fairly stable during several years. The relative mortality ratio for women compared to men, as shown in Figure 4.7(c), decreased, however, suggesting that the large increase in disability benefit awards for women during this period may be explained by a looser screening of new female compared to male disability beneficiaries.

Finally, there has been a remarkable decline in the entry rate to the disability insurance program since 2004, as shown in Figure 4.1. Figure 4.2 shows that the decline appeared across the board, for both age groups, both genders and in all regions. The rapid decrease in disability benefits awards since 2004 is reflected in a marked increase in the mortality ratio of new beneficiaries compared to non-beneficiaries, as depicted in Figure 4.7(a). Although no formal changes to the program were pursued until 2008, the mortality ratio increased already from 2005 to levels not seen before during the studied period. The increased mortality ratio may be due to an informal change in the eligibility assessment of new applicants that preceded the formal changes. The turning point coincides with the re-organization of the Social Insurance Agency into one central authority in 2005. Interestingly, the increased mortality ratio also implied a convergence in the relative mortality ratio of new beneficiaries compared to non-beneficiaries across gender, as shown in Figure 4.7(c).

To sum up, the results in the empirical analysis show that the estimated changes in the mortality ratio of new disability beneficiaries compared to non-beneficiaries correspond well to the development of the disability benefit entry rate, overall as well as within groups defined by age and gender. Changes in formal eligibility criteria are captured in the estimation results, and we also find that fluctuations in the entry rate during periods of no formal program changes

are reflected in a differential health of new program entrants compared to non-entrants. Overall, our results suggest that changes in how program inflow has been governed over time are important for the fluctuations in the entry rate to the disability insurance program.

7. Conclusion

Changes in the screening stringency of applications to the disability insurance program are potentially important for explaining program growth. Screening stringency is, however, inherently difficult to observe since it depends on the implementation of program rules as well as formal eligibility criteria. In this paper, we present a theoretical model of the application decision to the disability insurance program. We further propose an empirical strategy for capturing how the inflow to the disability insurance program has been governed over time. As an application, we study the inflow to the disability insurance program in Sweden between 1986 and 2008.

In the theoretical model, an individual decides whether to apply for disability benefits or whether to continue working. The application cost is increasing in the individual's wage. The probability of a successful application is increasing in the average disability benefit threshold, which is known to the individual, and decreasing in the individual's health. We show that it is optimal for an individual to apply for benefits if the individual's health is below a certain cut-off, which is increasing in the average disability benefit threshold. When screening stringency is reduced, i.e., when the average threshold increases, the health cut-off increases and the average health among applicants improves. We further show through numerical simulation that the health of disability benefits recipients relative to non-recipients improves when stringency is reduced.

We propose an empirical strategy for assessing how the inflow to the disability insurance program has been governed over time. We use a Cox proportional hazard model in which we compare the mortality of new disability beneficiaries to non-beneficiaries, taking age into account and allowing for cohort-specific mortality hazards. Applying the strategy to Sweden, we find that the estimated changes in the mortality ratio of new beneficiaries compared to non-beneficiaries correspond well to the development of the disability benefit entry rate, overall as well as within groups defined by age and gender. Changes in formal eligibility criteria are well reflected in the mortality ratio. We also find that fluctuations in the entry rate during periods of no formal program changes are reflected in changes in the relative health of new disability beneficiaries. Overall, our results suggest that substantial variation in program inflow remains after accounting for changes in the underlying health of the population.

References

Autor, D. (2011), "The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences, and Policy Options," NBER Working Paper No. 17697, National Bureau of Economic Research.

Autor, D., and M. Duggan (2003), "The Rise in Disability Rolls and the Decline in Unemployment," *Quarterly Journal Economics*, vol. 118, pages 157–206.

Bound, J., and R. V. Burkhauser (1999), "Economic Analysis of Transfer Programs Targeted on People With Disabilities," *Handbook of Labor Economics Volume 3*, Elsevier Science B. V., pages 3417–3528.

Burkhauser, R. V., C. Daly, A. J. Houtenville, and N. Nargis (2002), "Self-Reported Work-Limitation Data: What They Can and Cannot Tell Us," *Demography*, vol. 39(3), pages 541–555.

De Jong, P., M. Lindeboom, and B. van der Klaauw (2011), "Screening Disability Insurance Applications," *Journal of European Economic Association*, vol. 9, pages 106–129.

Duggan, M., and S. Imberman (2009), "Why Are the Disability Rolls Skyrocketing? The Contribution of Population Characteristics, Economic Conditions, and Program Generosity," in D. Cutler, and D. A. Wise, *Health at Older Ages: The Causes and Consequences of Declining Disability among the Elderly*, Chicago: The University of Chicago Press.

Gruber, J., and J. Kubik (1997), "Disability Insurance Rejection Rates and the Labor Supply of Older Workers," *Journal of Public Economics*, vol. 64, pages 1–23.

Halpern, J., and J. Hausman (1986), "Choice under Uncertainty: A Model of Applications for the Social Security Disability Insurance Program," *Journal of Public Economics*, vol. 31, pages 131–161.

Karlström, A., M. Palme, and I. Svensson (2008), "The Employment Effect of Stricter Rules for Eligibility for DI: Evidence from a Natural Experiment in Sweden," *Journal of Public Economics*, vol. 92, pages 2071–2082.

Parsons, D. (1991), "Self-Screening in Targeted Public Transfer Programs," *Journal of Political Economy*, vol. 99, pages 859–876.

Sjögren Lindquist, G., and E. Wadensjö (2007), "Ett svårlagt pussel - kompletterande ersättningar vid inkomstbortfall," Report for the Expert Group on Economic Studies No. 2007:1.

Staubli, S. (2011), "The Impact of Stricter Criteria for Disability Insurance on Labor Force Participation," *Journal of Public Economics*, vol. 95, pages 1223–1235.

Wise, D. A. (2012), *Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms*, Chicago: The University of Chicago Press.

Chapter 5

The Effect of Age-Targeted Tax Credits on Retirement Behavior*

Lisa Laun

Abstract

This paper analyzes the effect of two age-targeted policy initiatives to delay retirement that were simultaneously implemented in Sweden in 2007: an earned income tax credit and a payroll tax credit. Both policies were targeted at workers aged 65 or above at the beginning of the tax year. The paper exploits that the special rules for elderly were governed by the year of birth while the social security system is governed by age at retirement, i.e., the day of birth, in analyzing the effect of the new policies. The results suggest that the age-targeted tax credits increased employment in the year following the 65th birthday by 1.5 percentage points among individuals with annual earnings above the 2007 tax liability threshold three to five years earlier. An analysis of fiscal implications indicates, however, that the increase in employment was not large enough to offset the implied decrease in tax revenues.

Keywords: Labor supply, Retirement, Earned income tax credit, Payroll taxes
JEL-codes: H24, J14, J18, J21

*I am grateful for valuable comments from Johan Egebark, Peter Fredriksson, Eva Mörk, Mårten Palme and Håkan Selin as well as seminar participants at the Institute for Evaluation of Labour Market and Education Policy, the Department of Economics at Stockholm University, the 2nd Workshop in Empirical Public Policy at Uppsala University and the 2nd National Conference of Swedish Economists. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

1. Introduction

A key remedy for the fiscal pressure of an aging population, facing many developed countries, is to delay retirement. The main approach to encourage a delayed labor force exit has been through social security reform. The financial incentives for early retirement inherent in many social security systems, as shown in Gruber and Wise (1999), have been removed in many countries and age limits have been increased. An alternative approach, that has not yet been widely used, is to adjust labor tax rates close to the retirement age. If the labor supply elasticity near retirement is large, as suggested by French (2005), the cost of reduced tax rates might be offset by the revenue from an increased tax base. If the elasticity is small, reduced labor tax rates for older workers will merely work as a transfer to individuals who would have continued to work regardless.

In 2007, the Swedish government introduced two different labor tax credits for workers aged 65 or above at the beginning of the tax year, with the purpose to promote work at older ages. The first was an earned income tax credit introduced for all workers, that was substantially larger for workers above age 65. The additional tax credit for a worker at the 25th percentile of the earnings distribution, aged 65 at the beginning of 2007, amounted to about 9 percent of net earnings. The second was a payroll tax credit that reduced the payroll tax rate by about 16 percentage points. This paper studies if the age-targeted labor tax credits for workers above age 65 affected retirement behavior. The identification strategy exploits that individuals who turn 65 just before or just after the new year face different labor tax rates at essentially the same age. Public pension benefits, on the other hand, are determined by the age at retirement and other differences between the groups can be controlled for using a difference-in-differences approach, to the extent that they do not change discontinuously at the time of the reform. The sample consists of individuals who turned 65 from November to February between 2001 and 2009.

The results suggest that the age-targeted tax credits increased employment in the year following the 65th birthday by 1.5 percentage points, increased earnings as a share of previous earnings by 1.8 percentage points and increased the number of remunerated months by 0.133 among individuals with annual earnings above the 2007 tax liability threshold three to five years earlier. The results are shown to be robust to various specification tests. An analysis of heterogeneous responses indicates that the effects are primarily driven by men and are particularly large for self-employed. In an attempt to study the relative importance of the two types of tax credits the variation in the size of the credits by income and year is being used, but the analysis gives no conclusive result. From a policy perspective, the changes in retirement behavior induced

by the reform appear to be moderate. The public finance costs of the tax credits are large, and rough calculations suggest that the benefits from the behavioral changes are not large enough to account for reform costs.

This study relates to three branches of the previous literature. The first branch is the impact of financial incentives on the retirement decision, which has mainly focused on the effects of social security provisions.[1] The second is the impact of earned income tax credits on labor supply. The focus has been on the Earned Income Tax Credit in the U.S. and the Working Families Tax Credit in the U.K., mainly targeted towards low income families.[2] The Swedish earned income tax credit has primarily been assessed in simulation studies in which workers above age 65 have been excluded.[3] Edmark, Liang, Mörk and Selin (2012) use a quasi-experimental approach and conclude that the impact of the Swedish earned income tax credit for workers below age 65 is difficult to evaluate. The third branch of the previous literature is the effect of payroll taxation on employment.[4]

The paper makes two main contributions to the previous literature. First, although there is a large literature on the effect of social security systems on retirement behavior and a growing literature on labor market responses to different earned income tax credit and payroll tax initiatives, it is, to the author's knowledge, the first paper to study the effect of income tax policy changes on retirement behavior. As argued in Blundell and MaCurdy (1999), the participation decision is likely the most responsive margin of labor supply. This motivates reforms directly aimed at increasing participation among certain types of workers with potentially elastic labor supply; an often targeted group has been low income families. This paper studies how targeted income tax reforms affect another group with a potential labor supply reserve, namely workers at the margin of retirement. Second, the identification strategy, which exploits differences in the organization of the income tax and the social security systems with respect to cohort versus age based incentives, has not been previously used in the empirical public finance literature.

The study also contributes to the intense public policy debate in Europe and elsewhere on how to increase labor supply at older ages. Although the challenge of aging populations is not new, the question has gained importance in the wake of the financial crisis. The results in the paper suggest that social security reform is not the only way to enhance the incentives for a delayed labor force exit. Another approach is to consider the incentives in the income

[1] See Lumsdaine and Mitchell (1999) for a review.

[2] See Eissa and Hoynes (2006) and Blundell and Hoynes (2004) for reviews.

[3] See, e.g., Aaberge and Flood (2008), Ericson and Flood (2009) and Sacklén (2009).

[4] See, e.g., Gruber (1997), Bohm and Lind (1993), and Bennmarker, Mellander and Öckert (2009).

tax system, that affect the gains from working directly. From a public finance perspective, changing the rules in the social security system may be less costly than introducing financial incentives for work through tax credits. However, policies aimed at encouraging individuals who are healthy enough to continue working are likely easier to promote than forcing regulations for delayed retirement.

The results may also be informative of the effectiveness of the Swedish earned income tax credit in general, given that no quasi-experimental study has been successful in evaluating the reform. This paper studies the impact for workers who are already in the labor market, who have the right to stay at their job until age 67 and who are likely to have an elastic extensive margin labor supply. If there would be no effects for this group, it is unlikely that the earned income tax credit for workers below age 65 would get unemployed workers, who are likely to face much larger labor demand restrictions, into employment.

The paper is organized as follows. Section 2 describes the structure of the age-targeted tax credits and retirement institutions. Section 3 discusses the theoretical predictions. Section 4 outlines the empirical strategy and Section 5 describes the data. Section 6 presents the empirical results and Section 7 concludes the paper.

2. Institutional Setting

2.1. The Age-Targeted Labor Tax Credits

The individual is the tax-paying unit in the Swedish income tax system. Individual income is subject to a personal income tax that consists of a proportional local government tax and a progressive central government tax. The taxable income is the sum of labor and transfer income reduced by a standard deduction, that is phased in at low incomes and phased out at high incomes, and a deduction for certain costs of acquiring income. The local government tax rate varies across Sweden's 290 municipalities, ranging from 26.5 to 34.41 percent with an average rate of 31.55 percent in 2007. The central government tax schedule has two thresholds, with a tax rate of 20 percent above the first and 25 percent above the second threshold.[5]

The first labor tax credit studied in this paper is an earned income tax credit that reduced the personal income tax on labor income only. It was introduced on 1 January 2007 for workers of all ages, with the purpose of increasing

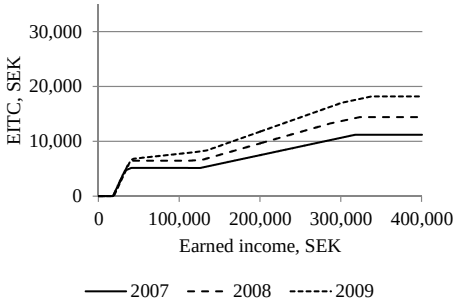
[5]In 2007, the standard deduction ranged from SEK 17,100 to 31,100 ($ 2,400 to 4,400). The first threshold for the central government tax was SEK 316,700 ($ 45,200) and the second threshold was SEK 476,700 ($ 68,100). Throughout the paper, I use an exchange rate of 7 SEK for 1 USD.

the returns from working relative to collecting public transfers. Motivated by the particular importance of encouraging older workers to remain in the labor force, the tax credit was substantially larger for workers aged 65 or above at the beginning of the tax year. Apart from age, the size of the tax credit was a function of the earned income, the standard deduction and the local government tax rate. It was a non-refundable credit that could not reduce the local government tax liability below zero and was deducted automatically on the monthly paycheck. The earned income tax credit was expanded in 2008, 2009 and 2010. Due to data availability, this paper focuses on the period 2007–2009.

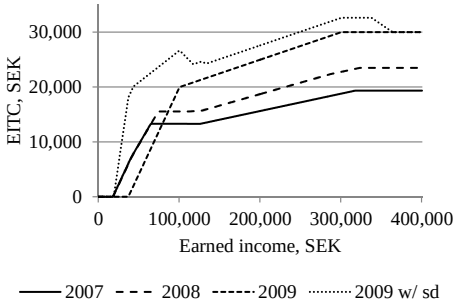
Figure 5.1 presents the structure of the earned income tax credit during 2007–2009, assuming the average local tax rate and no taxable transfers.[6] Figures 5.1(a) and 5.1(b) show the earned income tax credit as a function of earned income for workers who are below or above age 65 at the beginning of the tax year. Unlike most earned income tax credits in other countries, there was no phase-out region of the credit. In 2007 and 2008, the shape of the tax credit schedule was the same for workers below and above age 65, but the larger initial phase-in region made the tax credit substantially more generous for the older age group. In 2009, an additional standard deduction for workers above age 65 was introduced. The earned income tax credit for older workers was therefore made independent of the standard deduction and the shape of the tax credit changed. Since the additional standard deduction applied to transfer income as well as labor income it did not explicitly encourage work, and the analysis in this paper can therefore be argued to capture the difference that is due to the earned income tax credit per se.

Figure 5.1(c) shows the difference in tax amounts for workers above compared to below age 65, with and without the additional standard deduction in 2009. The additional tax credit for workers above age 65 increased slightly from 2007 to 2008 and increased substantially from 2008 to 2009, except in the interval where the additional standard deduction limited the tax liability of older workers. Between the tax liability thresholds for workers below and above age 65 in 2009, the tax credit is zero for older workers and positive for younger workers, which explains the negative difference in this interval. For an employed worker at the 25th percentile of the earnings distribution who was aged 65 at the beginning of the tax year, the additional earned income tax credit amounted to 9.1 percent of net earnings in 2007, 9.6 percent in 2008

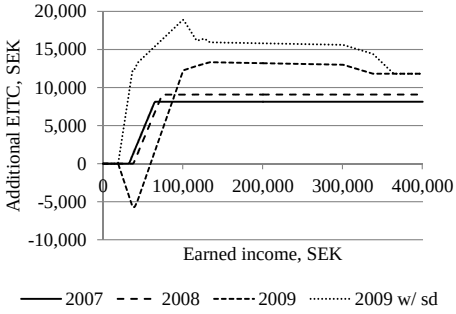
[6]The EITC depends on taxable transfers through the standard deduction. The assumption of no taxable transfer is strong for individuals above age 65 of which more than 90 percent collect public pension. Larger taxable transfers will, however, only slightly increase the difference in credit amounts for workers below and above age 65 presented in Figure 5.1(c).



(a) EITC below age 65



(b) EITC above age 65



(c) Additional EITC above age 65

Figure 5.1: The earned income tax credit as a function of earned income 2007–2009, below age 65 in (a), above age 65 in (b), additional tax credit above age 65 in (c), with and without additional standard deduction in 2009

and 12 percent in 2009.[7] The earned income tax credit for older workers thus substantially increased the gains from continued work.

Labor income is also subject to a proportional payroll tax levied on all wages paid out by employers. The payroll tax includes a general wage tax and specific contributions to various social insurance programs. The benefits to which an individual is entitled increase with the level of income up to a cap that varies across programs. For income below the cap, the payroll tax can hence partly be seen as an insurance premium. Eligibility for disability and unemployment benefits ceases on the 65th birthday, however, and access to sickness benefits is restricted after that age. The payroll tax rate has therefore traditionally been slightly lower for individuals aged 65 or above at the beginning of tax year. The lower rate has been cohort-specific because of the introduction of the new pension system, described in the next section.

Figure 5.2: The payroll tax rate by age at the beginning of the tax year 2001–2009

The second labor tax credit studied in this paper is a payroll tax credit that further reduced the tax rate for workers aged 65 or above at the beginning of the tax year. Like the earned income tax credit, it was introduced on 1 January 2007. Figure 5.2 presents the payroll tax rate by age at the beginning of the tax

[7]Employment is defined as having annual earnings above one income base amount (see Section 5 for details and data). Earnings at the 25th percentile of the earnings distribution among workers aged 65 at the beginning of the year was SEK 89,200 in 2007, 95,100 in 2008 and 102,800 in 2009.

year during 2001–2009. The normal payroll tax rate in 2007 was 32.42 percent for regular employers and 30.71 percent for self-employed. The payroll tax rate for workers above age 65 was slightly lower even before 2007 but was reduced from 26.37 percent in 2006 to 10.21 percent in 2007. Since then, it only includes pension contributions. The payroll tax credit thus reduced the payroll tax rate for older workers by 16.16 percentage points. Also this reform was motivated by the importance of promoting work among older workers. While the earned income tax credit was aimed at stimulating labor supply, the purpose of the payroll tax credit was to stimulate labor demand. It could, for example, compensate for productivity declines or workplace accommodations.

2.2. Retirement Institutions

In Sweden, the decision to retire from the labor force is separate from public pension collection, in the sense that there are no restrictions on the size of labor earnings when collecting income-related public pension. The marginal tax rate increases with total income, however, and guaranteed pension is means-tested. The mandatory retirement age, i.e., the age at which an employer can ask an employee to leave an employment, is determined in collective agreements but the minimum level is regulated under the Employment Protection Act. In 2001, the minimum mandatory retirement age in Sweden increased from 65 to 67. Legally, the retirement decision at age 65 among employed workers is therefore in the hands of the individual.

The Swedish public pension system was reformed during the late 1990s. The minimum retirement age under the old system was 61 and the normal retirement age was 65. For each month before age 65 that the individual collected old-age pension, benefits were reduced by 0.5 percent, and for each month after age 65 that the individual postponed collecting pension, benefits were increased by 0.7 percent. Under the new pension system, the minimum retirement age is also 61 but there is no normal retirement age. Income-related pension benefits are calculated by adjusting the notional account balance at the time of retirement by an annuity divisor that is based on life expectancy and a real rate of return during the expected life of the annuity. Guaranteed pension can be collected from the 65th birthday. Although there is no normal retirement age, most people start collecting public pension when turning 65. In 2007, about 20 percent of individuals collected public pension in the year they turned 64 while more than 90 percent collected public pension in the year they turned 65.

The old public pension system was a defined benefit scheme consisting of a flat-rate basic pension and an income-related supplementary pension based on the best 15 out of 30 years of earnings. The new system is a combination

of notional defined contributions on a pay-as-you-go basis and a smaller financial defined contribution scheme. Individuals with small or no pension claims receive a guaranteed pension. The main factor determining pension benefits is the age at retirement, through the actuarial adjustments described above, but there is a cohort-based element through the phase-in of the new pension scheme. The 1938 cohort was the first to receive pension from the new scheme, with 4/20 of their pension benefits from the new scheme and 16/20 from the old scheme. Each successive cohort receives an additional 1/20 from the new scheme and 1/20 less from the old scheme. Individuals born in 1954 or later are completely in the new scheme. The phase-in of the new public pension system is thus very slow and the size of the incentives change is the same for each cohort. The exception is the 1938 cohort aged 65 at the beginning of 2003, i.e., well before the age-targeted tax credits were introduced.

3. Theoretical Framework

To discuss the theoretical predictions of the age-targeted labor tax credits, I distinguish between the *gross* wage w paid by the employer, the *taxable* wage z received by the worker and the *net* wage c consumed by the worker. According to economic theory, the incidence of taxation does not depend on which side of the market the tax is levied, but will fall mostly on the group for which the response to price changes is least elastic. Figure 5.3 presents the potential effects of the tax credits in a simple model of supply and demand in the labor market. The earned income tax credit increases the net wage of workers and shifts the labor supply curve from S_0 to S_1, which reduces the taxable wage from z_0 to z_1 and increases employment from E_0 to E_1. The payroll tax credit lowers labor costs and shifts the labor demand curve from D_0 to D_1, which increases the taxable wage from z_0 to z_2 and increases employment from E_0 to E_1. Combined, the two tax credits shift both the labor supply and labor demand curve to the new equilibrium given by E_2 and z_0, resulting in increased employment but with an ambiguous effect on taxable wages. The magnitude of the employment effect and the sign and magnitude of the wage effect depend on the relative elasticities of labor supply and labor demand, i.e., the relative slopes of the two curves.[8]

In a market with perfect competition, the gross wage equals the marginal product of the worker. Labor demand is hence infinitely elastic, which implies

[8]Summers (1989) and Gruber (1997) point out that if payroll tax revenues are used to finance programs which benefit workers, the tax-benefit linkage will shift labor supply outwards as well, since the tax is buying workers benefits. Workers above age 65 were not eligible to the programs the payroll tax financed before the reform, however, so there was no tax-benefit linkage.

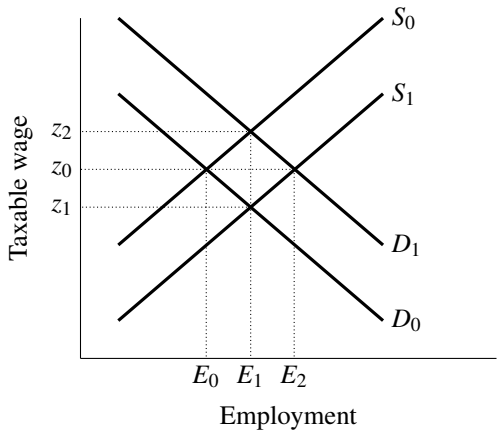


Figure 5.3: The incidence of taxation

that the full tax incidence falls on the worker. In the standard static labor supply model, the individual maximizes the utility from consumption and leisure subject to a budget constraint. Following Lehmann, Marical and Rioux (2011), I denote one minus the labor income tax rate the net-of-income-tax rate and one minus the payroll tax rate the net-of-payroll-tax rate. On a linear part of the payroll tax schedule, $z = \tau^P w + R^P$, where τ^P is the marginal net-of-payroll-tax rate and R^P is the virtual taxable income. Likewise, on a linear part of the income tax schedule, $c = \tau^I z + R^I$, where τ^I is the marginal net-of-income-tax rate and R^I is the virtual net income. Combining these expressions gives the individual budget constraint if the full tax incidence falls on the worker:

$$c = \tau^I \tau^P w + \tau^I R^P + R^I. \tag{5.1}$$

The average net-of payroll-tax rate is given by $\rho^P = z/w = \tau^P + R^P/w$ and the average net-of-income-tax rate is given by $\rho^I = c/z = \tau^I + R^I/z$. In this model, labor supply depends on the global marginal net-of-tax rate $\tau = \tau^I \tau^P$, the global virtual income $R = \tau^I R^P + R^I$ and the global average net-of-tax rate $\rho = c/w = \rho^P \rho^I$.

There are several reasons for why the above description may not hold. Labor market frictions resulting in imperfect competition could limit the adjustment of taxable wages. Wage negotiations typically take place with a lag and taxable wages may be rigid due to, for example, union bargaining. If so, it does matter to which side of the market the tax is levied. The payroll tax credit would accrue to the employer and potentially affect employment through increased demand for labor, although empirical findings suggest that the employ-

ment effects of payroll tax reductions may be small.[9] Only the earned income tax credit would accrue to the worker.

A point highlighted by Liebman (1998) and Chetty and Saez (2009) is that the salience of taxation may be important for the responsiveness to tax changes. Riksrevisionen (2009) shows that only 40 percent of respondents were aware of the earned income tax credit in 2009. Since retirement regards the decision to leave the labor force and the earned income tax credit is deducted automatically on the monthly paycheck, however, older workers may adjust their behavior to the tax credit without explicit knowledge of its existence or structure. Furthermore, there is evidence that indirect taxes are less salient than direct taxes, which implies that the payroll tax credit may have a smaller impact than the earned income tax credit.[10]

It should also be noted that workers above age 65 are not eligible for the social programs financed by payroll taxes. In that sense, the payroll tax reform could be motivated by "fairness" rather than by promoting work. For these reasons, it is worth to also consider the case in which the statutory incidence determines the economic incidence of taxation. If only the earned income tax credit accrues to the worker, the individual budget constraint would be given by:

$$c = \tau^I z + R^I, \tag{5.2}$$

where the taxable wage z is unaffected by the reform.

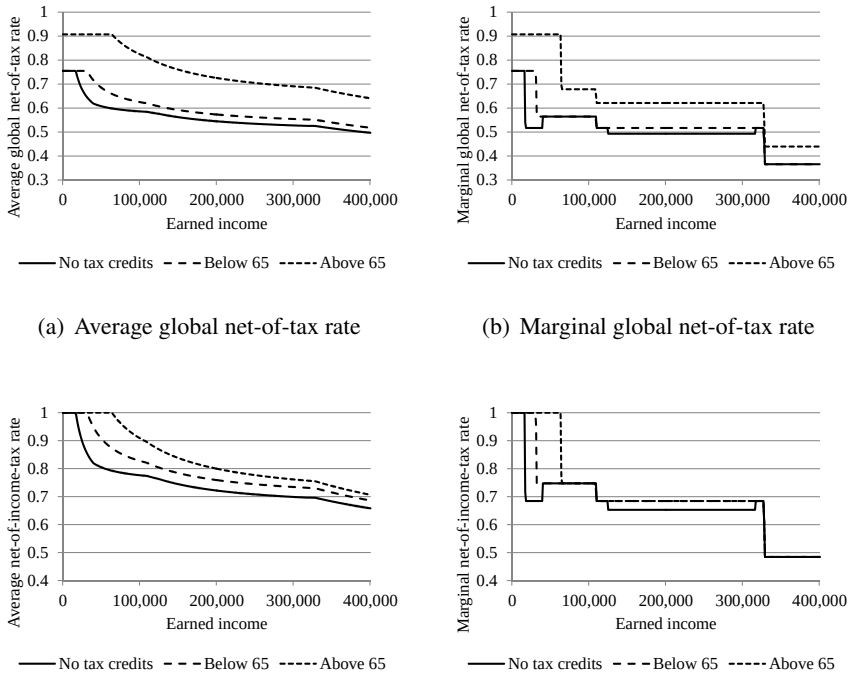
Figure 5.4 shows how the age-targeted tax credits altered the average and marginal net-of-tax rates of individuals below and above age 65 in 2007. The upper panels present the global net-of-tax rate, taking both the payroll tax credit and the earned income tax credits into account, as in equation (5.1). The lower panels present the net-of-income-tax rate, only taking the earned income tax credit into account, as in equation (5.2). At the extensive margin, where the individual decides whether to work or not, the predictions of the age-targeted tax credits are clear. Since the average net-of-tax rates in Figure 5.4(a) and 5.4(c) are higher for workers above compared to below age 65 throughout the earnings distribution, the tax credits unambiguously increased the relative incentives for employment among older compared to younger workers.

At the intensive margin, the effects of the tax credits on labor supply are ambiguous. A reduction in the marginal tax rate affects labor supply through a positive substitution effect, due to the higher relative price of leisure, and a negative income effect, due to the increase in the net wage, given that leisure

[9] See, e.g., Bennmarker, Mellander and Öckert (2009) and Huttunen, Pirttilä and Uusitalo (2010).

[10] See, e.g., the discussion in Pirttilä and Selin (2011).

is a normal good. The substitution effect typically dominates the income effect. In addition to the price effect, there is an additional income effect from the reduction in the average tax rate, affecting labor supply in a negative direction. In the upper panels of Figure 5.4, including both the earned income tax credit and the payroll tax credit, the positive substitution effect from the higher marginal net-of-tax rate and the negative income effect from the higher average net-of-tax rate for workers above age 65 implies that the effect on labor supply at the intensive margin is undetermined. In the lower panels, where only the earned income tax credit is accounted for, the marginal net-of-tax rate is the same beyond the phase-in region of the credit whereas the average net-of-tax rate is higher for workers above age 65. In this region, the substitution effect is the same but the income effect is larger for older compared to younger workers, resulting in unambiguously smaller incentives for work in the older age group.



(a) Average global net-of-tax rate

(b) Marginal global net-of-tax rate

(c) Average net-of-income-tax rate

(d) Marginal net-of-income-tax rate

Figure 5.4: The average and marginal global net-of-tax rate and net-of-income-tax rates in 2007, without tax credits and with tax credits above and below age 65

4. Empirical Strategy

The aim of the empirical analysis is to examine to what extent the two labor tax credits for workers above age 65 introduced in 2007 affected labor market outcomes in the targeted group. Since the earned income tax credit and the payroll tax credit were introduced simultaneously and applied to the same group of workers, the main analysis studies the joint impact of the two credits. To isolate the effect of the policy changes from other confounding factors that may affect retirement, I exploit a feature inherent in the design of the tax policies. Eligibility for the age-targeted tax credits was determined by the age at the beginning of the tax year, which created a discontinuity in labor income tax rates that depended on the date of the 65th birthday.[11] Individuals who turned 65 just before the new year were eligible for the tax credits from 1 January while individuals who turned 65 just after the new year were eligible for the tax credits one year later. I use this discontinuity in labor taxation to identify the effect of the age-targeted tax credits on labor market outcomes in the year immediately following the 65th birthday.

Individuals who turn 65 on different sides of the new year may differ in terms of labor market outcomes even in absence of the age-targeted tax credits. The treatment group is slightly older than the control group, which reduces labor force attachment at a given point in time. As discussed in Section 2.2, the new public pension system is also slowly phased in by an equal amount for each successive cohort. Furthermore, the cutoff for school start is defined by calendar year in Sweden. Fredriksson and Öckert (2009) show that individuals born early in the year, who start school at an older age, perform better in terms of long-run labor market outcomes than individuals born late in the year. To the extent that relative labor market outcomes of individuals who turn 65 on different sides of the new year are constant across cohorts, however, such differences can be controlled for using a difference-in-differences approach.

I estimate a model which compares the difference in labor market outcomes of individuals who turn 65 on different sides of the new year before and after the implementation of the tax credits:

$$y_{it} = \alpha + \gamma Reform_{it} + \beta' X_{it} + \lambda_a + \lambda_t + \varepsilon_{it}, \tag{5.3}$$

where y_{it} is the labor market outcome of individual i in year t, X_{it} is a vector of individual characteristics, λ_a is a set of indicator variables for the individual's age in months at the beginning of the tax year, λ_t is a set of indicator variables for the year in which the outcome is measured and ε_{it} is the error term. The variable $Reform_{it}$ is an indicator variable that takes the value one if the

[11] The Swedish tax year follows the calendar year.

individual is aged 65 or above at the beginning of the tax year and the year is 2007 or later, when the age-targeted tax credits were in place. The coefficient γ captures the treatment effect of the age-targeted tax credits. The fact that the control group will receive the tax credits one year later may impact their labor market outcomes during the year of analysis as well, for example by remaining employed until becoming eligible. Such behavior would attenuate the estimated effects of the tax credits. The treatment effect can thus be interpreted as the effect of receiving the age-targeted tax credits one year earlier.

The identifying assumption is that the relative labor market outcomes of the treatment and the control group would have remained constant in absence of the reform. To assess this assumption, I perform a t-test for the presence of a linear time trend in the pre-reform period. I also estimate the models:

$$y_{it} = \alpha + \sum_{t=s}^{S} \gamma_t \delta_{treat} \times \lambda_t + \beta' X_{it} + \lambda_a + \lambda_t + \varepsilon_{it}, \tag{5.4}$$

$$y_{it} = \alpha + \sum_{a=m}^{M} \gamma_a \delta_{post} \times \lambda_a + \beta' X_{it} + \lambda_a + \lambda_t + \varepsilon_{it}, \tag{5.5}$$

where the second term in equation (5.4) are treatment$\times$year interactions that take the value one if the individual belongs to the treatment group and the year is equal to t, and the second term in equation (5.5) are post$\times$age interactions that take the value one if the age in months at the beginning of the tax year is a and the year is 2007 or later. In these models, there should be no effect of the reform in the years before 2007 and for ages below 65. The effect should appear where eligibility for the tax credits truly begins.

One threat to identification is that other reforms have affected the same cohorts as the age-targeted tax credits. The large compulsory school reform analyzed by, e.g., Meghir and Palme (2005) primarily affected cohorts born in the end of the 1940s onwards. The youngest cohort included in this study were born in 1944, and Holmlund (2007) shows that less than 5 percent of this cohort was affected by the compulsory school reform. Also the public pension system was reformed during the period under study. The first cohort in the new pension system was aged 65 at the beginning of 2003, however, several years before the introduction of the tax credits.

If the error terms of individuals in different groups are correlated, standard errors could be biased. Common shocks that create dependence of individuals within a group could appear for a number of reasons. It might be difficult to imagine local labor market shocks that affect individuals only a few months apart in age differently, but cohort based reforms earlier in life could have lasting differential impacts on the treatment and control groups. I handle this issue by aggregating the analysis to the group level using the two-step approach

suggested by Donald and Lang (2007).[12] In the first step, I construct covariate adjusted group-year effects, μ_{gt}, by estimating:

$$y_{igt} = \mu_{gt} + \beta' X_{it} + \lambda_a + v_{igt}, \tag{5.6}$$

where y_{igt} is the outcome variable, X_{it} are individual characteristics, λ_a are indicator variables for the age in months at the beginning of the tax year and v_{igt} is the error term. In the second step, the estimated group year effects, $\hat{\mu}_{gt}$, are regressed on the variables that vary only at the group and year level using the equation:

$$\hat{\mu}_{gt} = \alpha + \gamma Reform_{gt} + \lambda_t + u_{gt}, \tag{5.7}$$

where $Reform_{gt}$ is the reform indicator and λ_t are year dummies.

5. Data

I use register-based data compiled and maintained by Statistics Sweden. The Longitudinal Database on Education, Income and Employment (LOUISE) provides demographic and socioeconomic information, the Income and Tax Register (IoT) provides individual income tax records, and the Register-Based Labor Market Statistics (RAMS) provides records of employment spells. The estimation sample is a cross-section of individuals who turned 65 during a certain window around the new year between 2001 and 2009.[13] Historical records are used to construct individual baseline characteristics. Since the tax credits were introduced in 2007 and the last observation year is 2009, events that occurred at least three years earlier are exogenous to the policies for all years. To limit transitory fluctuations, I use the three to five years before the year of analysis as the "baseline years" for constructing covariates. These are the years when individuals with birthday before the new year turned 61–63 and individuals with birthday after the new year turned 60–62. The data cover the period 1996 to 2009.

I define previous earnings as the maximum annual earnings during the baseline years, which can be thought of as "potential earnings" if the individual would continue working after age 65.[14] To exclude individuals who exited the labor force early, the estimation sample is limited to individuals with previous

[12] A similar strategy is to cluster the standard errors at the group level, but since this method requires a large number of clusters, it is not feasible here.

[13] The data cover the entire Swedish population aged 16–64 before 2001 and aged 16–74 during 2001–2009. Because of the age restriction before 2001, the first year of analysis is 2001.

[14] I use maximum rather than average earnings to limit the impact of low annual earnings due to individuals exiting the labor force.

earnings above the 2007 tax liability threshold.[15] This sample restriction is assessed below. Education level is determined by the maximum level during the period. A variable of previous sickness indicates that the individual received sickness or disability benefits from the Social Insurance Agency during the baseline years.[16] Using household identifiers, I identify the spouse in the baseline years and record whether the spouse was younger, older or if there was no spouse.[17] Self-employment status, immigrant status and the municipality and county of residence is determined by the most common record during the baseline years.

The year and month of birth are used to define eligibility for the age-targeted tax credits. The day of birth is not available in the data. The analysis requires that the window of birth months around the new year is determined. The most restrictive approach is to compare individuals turning 65 in December to individuals turning 65 in January, who are only one month apart in age but face different labor tax rates. Expanding the window is a trade-off between increasing the sample size, which improves efficiency, and increasing the age difference between the treatment and control group, which reduces comparability.

The first three columns in Table 5.1 present the raw differences in baseline characteristics between the treatment and the control group for samples of individuals who turned 65 within one to three months around the new year during 2001–2009. The treatment group is slightly less educated than the control group, which can be expected since the treatment group belongs to an older cohort and education has expanded over time. There are also fewer immigrants in the treatment group, which might be explained by anecdotal evidence that immigrants with an unknown birthday are recorded on 1 January.[18] Individuals in the treatment group are also more likely to have an older spouse. Since they are born late in the year, while individuals in the control group are born early, the treatment group is more likely to be younger than their spouse when marrying someone in the same birth cohort. The differences in previous sickness, self-employment status and previous earnings are small and becomes insignificant on a five percent level when the window is narrowed to 1 or 2 months around the new year.

The difference-in-differences strategy allows the treatment and control groups to be different, as long as the characteristics do not change at the same

[15] The 2007 tax liability threshold was SEK 17,100 ($ 2,400).

[16] Since the employer period is two weeks this includes individuals absent for more than two weeks.

[17] Some of the individuals recorded as having no spouse may, however, be co-habiting.

[18] Since the recorded birthday is used for tax purposes, this should not be a problem. Furthermore, the share of immigrants is low.

time as the reform was implemented. The last three columns in Table 5.1 report whether the individual characteristics of the treatment group change discontinuously at the time of the reform. The table presents the results from the model:

$$Reform_{it} = \alpha + \beta' X_{it} + \lambda_a + \lambda_t + \varepsilon_{it}, \tag{5.8}$$

which captures how the reform indicator $Reform_{it}$ depends on individual characteristics X_{it} when controlling for age in months and time effects, as in the main analysis. An F-test tests the joint restriction of all coefficients on the variables in X_{it} being zero.

	Raw difference treatment–controls			Estimation of reform indicator on covariates		
Variables	Dec-Jan	Nov-Feb	Oct-Mar	Dec-Jan	Nov-Feb	Oct-Mar
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.003	-0.003	-0.001	-0.001	0.001	0.001
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
High School	-0.001	-0.002	-0.004**	-0.002	0.000	0.001
	(0.003)	0.002	(0.002)	(0.002)	(0.001)	(0.001)
College	-0.013***	-0.016***	-0.016***	-0.007***	-0.002	-0.001
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Immigrant	-0.006***	-0.005***	-0.003**	0.003	0.003*	0.003**
	(0.002)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)
Previously sick	0.000	-0.004*	-0.004**	0.000	-0.001	-0.001
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Older spouse	0.006***	0.009***	0.007***	0.002	0.000	0.001
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Younger spouse	-0.009***	-0.011***	-0.008***	0.002	0.002	0.002
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Self-employed	0.002	0.002	0.003***	-0.001	-0.002	-0.000
	0.002	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)
Previous earnings	-2,270*	-1,566*	-2,761***	0.000	0.000	0.000
	(1,166)	(899)	(713)	(0.001)	(0.000)	(0.000)
County dummies				Yes	Yes	Yes
Year dummies				Yes	Yes	Yes
Age dummies				Yes	Yes	Yes
Observations				92,893	181,184	281,944
R-squared				0.617	0.614	0.607
F-test				1.465	0.994	1.097
Prob > F				0.0508	0.474	0.329

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Treatment and control groups consist of individuals turning 65 in Oct–Dec and Jan–Mar, respectively, 2001–2009, with previous earnings above 17,100 (2007 SEK). Excluded education cat.: less than high school, excluded marital status: no spouse.

Table 5.1: Raw differences between the treatment and control groups and estimation of the reform indicator on covariates for different sample selections

For the sample of individuals with birthday in December and January, the F-statistic suggests that I can reject the joint restriction of all coefficients being zero on a five percent significance level. For such a limited sample, the share

of college educated decreases discontinuously in the treatment group at the time of the reform. A smaller share of college educated in the treatment group is likely to attenuate any effects, since educated workers are more likely to remain employed. The difference is small and disappears when expanding the sample. For the sample of individuals born between November and February and between October and March, the F-statistics suggest that I cannot reject the joint restriction of all coefficients on the variables in X_{it} being zero, which is reassuring. To keep the treatment and control groups as similar as possible, I use the sample of individuals with birthday from November to February in the estimation, i.e., a window of two months around the new year. Summary statistics for this sample are presented in Table A-1 in the Appendix.

In terms of outcome variables, there is no information about hours or wages for workers above age 64 in the Swedish register data. There is, however, information about annual taxable labor earnings, which is the sum of labor earnings from formal employment and self-employment net of certain costs for acquiring the income. There are also records of the number of remunerated months reported by employers. Given the theoretical predictions, extensive margin responses are the most interesting. Theory predicts an unambiguously positive employment effect of the two tax credits. The first outcome variable is employment, defined as having annual taxable labor earnings above one income base amount (SEK 45,900 or about $\$6,600$ in 2007).[19] Since individuals in the control group turn 65 in January or February during the year in which the outcomes are measured, a positive cutoff helps capturing employment past age 65. In a sensitivity analysis, other cutoffs for employment are being used, as well as extensive margin responses based on the number of remunerated months.

To make use of the data at hand, I also construct three outcome variables capturing responses at the extensive and intensive margins combined. Although the tax credits are predicted to increase labor supply at the extensive margin, negative income effects may decrease labor supply at the intensive margin, implying an undetermined total effect. The second outcome variable is the annual taxable labor earnings, including zero records. The third outcome variable is the annual taxable labor earnings as a share of previous earnings, defined above. The difference between the second and third outcome variable is merely an issue about functional form. Whereas the second outcome variable captures the effect on earnings linearly, the third outcome variable captures the effect on the extent of work in relation to the "potential" extent of work. This may be closer to capturing the effect on hours worked. Finally, the fourth outcome variable is the number of remunerated months. This outcome

[19]The income base amount is determined by the government each year and is used for calculations in the public pension system. It closely follows the development of income in the economy.

variable is a valuable complement since it makes use of another source of data, originating from employer records.

Table 5.2 shows the averages of the outcome variables in the treatment and control groups between 2001 and 2006, before the tax credits were in place. The outcome variables are measured on a tax year basis. Since the treatment group is slightly older than the control group during the year in which the outcome is measured, the treatment group has a lower employment rate, lower labor earnings, earns a lower share of previous earnings and has fewer remunerated months than the control group.

	Treatment Group (1)	Control Group (2)	Difference (3)
Employment	0.226	0.295	-0.070***
	(0.002)	(0.002)	(0.003)
Taxable labor earnings	45,183	54,900	-9,717***
	(512)	(474)	(697)
Share of previous earnings	0.184	0.222	-0.038***
	(0.002)	(0.002)	(0.003)
Remunerated months	3.844	4.318	-0.475***
	(0.024)	(0.023)	(0.033)
Observations	51,019	55,565	

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Treatment and control groups consist of individuals turning 65 in Nov–Dec and Jan–Feb, respectively, 2001–2009, with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount.

Table 5.2: Summary statistics, outcome variables

As explained above, the sample is limited to individuals with previous earnings above the 2007 tax liability threshold. This affects the external validity of the results, since the conclusions drawn from the empirical analysis can only be applied to this particular group. As is shown below, however, more than two thirds of the individuals are included in the estimation sample, and the results thus apply to a large share of the population. The motivation for the sample restriction is that individuals with very low earnings in all years three to five years before turning 65 are unlikely to be affected by financial incentives for continued work after age 65. They may be homemakers or have exited the labor force early due to, for example, health reasons. To impact the labor supply of these workers, reforms to systems that are more relevant for early retirement, such as the disability insurance or occupational insurance schemes, are likely to be more effective.

To make sure that the sample restriction does not drive any results, I study if the probability of being included in the estimation sample is continuous at

the cutoff. I estimate the following model on the full population:

$$Reform_{it} = \alpha + \beta Included_{it} + \lambda_a + \lambda_t + \varepsilon_{it}, \tag{5.9}$$

which captures how the reform indicator depends on the probability of being included in the sample. The results, presented in the first column of Table A-2 in the Appendix, show that the probability of being included in the estimation sample does not change discontinuously at the cutoff, which is reassuring. The second column of Table A-2 shows the effect of covariates on the probability of being included in the sample. The value of the constant implies that on average 71 percent of the cohort are included. Women, immigrants and individuals who were previously sick are less likely to be included in the sample, whereas educated workers, self-employed and workers with a spouse are more likely to be included. I also estimate the main model specified in equation (5.3) for the excluded sample. The results, presented in Table A-3 in the Appendix, show that the reform effect is insignificant and close to zero, which suggests that any findings for the estimation sample will not be counteracted by a differential effect for the sample of individuals excluded from the analysis because of their labor market history.

6. Results

This section presents the results from the empirical analysis. First, I present the main estimation results of the impact of the age-targeted labor tax credits on labor market outcomes, along with a variety of robustness checks. Thereafter, I study heterogeneous effects for different groups of the population and try to distinguish between potential mechanisms. Next, I make an attempt to assess the relative importance of the two types of tax credits. Finally, I translate the estimation results into elasticities and perform rough calculations of public finance implications, given the behavioral responses found in the analysis.

6.1. Main Results

Table 5.3 presents the main results from estimating equation (5.3) for the sample of individuals who turned 65 from November to February during 2001–2009 and had annual earnings above the 2007 tax liability threshold three to five years earlier. Each column presents the results from a separate regression for each outcome variable. Column (1) shows that the age-targeted tax credits increased the probability of employment during the year immediately following the 65th birthday by 1.5 percentage points among individuals with previous earnings above 17,100 (2007 SEK). Relating this to the average employment rate in the sample, presented at the end of the table, implies an employment

effect of 4.9 percent. The estimated effect on taxable labor earnings in column (2) is positive but not significantly different from zero. For taxable labor earnings as a share of previous earnings, presented in column (3), however, there is a positive and significant effect of 1.8 percentage points or about 7.6 percent. That the impact on earnings as a share of previous earnings is significant, while the effect on labor earnings is not, might be due to the latter variable being too volatile for capturing labor market responses. Finally, the results for the number of remunerated months in column (4) suggest a significant increase of 0.133 months, or 2.9 percent, due to the age-targeted tax credits.

Variables	Employment (1)	Taxable labor earnings (2)	Share of previous earnings (3)	Remunerated months (4)
Reform	0.015***	1,518	0.018***	0.133***
	(0.004)	(1,335)	(0.004)	(0.051)
Female	-0.069***	-27,134***	-0.035***	-0.439***
	(0.002)	(718)	(0.003)	(0.029)
High School	0.027***	8,910***	0.022***	0.396***
	(0.002)	(538)	(0.002)	(0.029)
College	0.138***	56,949***	0.099***	1.693***
	(0.003)	(948)	(0.003)	(0.034)
Immigrant	-0.003	2,815**	0.017***	-0.246***
	(0.003)	(1,188)	(0.004)	(0.040)
Self-employed	0.114***	14,563***	0.206***	2.843***
	(0.005)	(1,495)	(0.007)	(0.055)
Previously sick	-0.129***	-34,974***	-0.090***	-1.458***
	(0.002)	(529)	(0.002)	(0.026)
Older spouse	-0.063***	-12,465***	-0.058***	-0.582***
	(0.003)	(722)	(0.003)	(0.032)
Younger spouse	-0.004	1,358*	-0.017***	0.066**
	(0.003)	(812)	(0.003)	(0.032)
Constant	0.282***	62,577***	0.207***	3.920***
	(0.005)	(1,409)	(0.006)	(0.062)
County dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Age dummies	Yes	Yes	Yes	Yes
R-squared	0.086	0.096	0.055	0.077
Observations	181,184	181,184	181,184	181,184
p-val parallel trends test	0.327	0.192	0.639	0.919
Mean of dep. variable	0.306	64,382	0.242	4.541
Effect in percent	0.049	0.024	0.076	0.029

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Nov–Feb 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount.

Table 5.3: The effect of the age-targeted tax credits on labor market outcomes

As discussed in Section 4, the underlying assumption in the difference-in-differences approach is that the relative labor market outcomes of the treatment and control groups would have remained constant in absence of the reform.

The third to last row of Table 5.3 presents the p-value from the t-test of the presence of a linear time trend in the pre-reform period, suggesting that I cannot reject the parallel trends assumption. Table 5.4 presents the estimation results for the model with unrestricted treatment×year interactions, specified by equation (5.4). The data cover the period 2001–2009 and the excluded years for the interaction terms are 2001–2003. The results show no significant effects of the tax credits in the pre-reform period, on a five percent significance level, which also supports the parallel trends assumption. The significant effects show up from 2007, when the age-targeted tax credits were implemented, and can mainly be seen in 2008 and 2009. The late response could be due to an increased awareness of the tax credits over time or by the expansion of the earned income tax credit.

Table 5.5 presents the estimation results for the model with unrestricted post×age interactions, specified by equation (5.5). The sample includes individuals who turned 65 from October to March during 2001–2009 and the excluded category for the interaction term is turning 65 in March. The effect on employment clearly appears for those turning 65 in December and earlier, where eligibility for the age-targeted tax credits begins. There is also a shift around the new year for earnings as a share of previous earnings and the number of remunerated months, although not all coefficients are significant. The estimates for taxable labor earnings are volatile, which might explain why there were no significant effects for this variable in the main analysis. For the other outcome variables, the results support the assumption of parallel trends and suggest that it is in fact the effect of the tax credits that is being captured in the analysis.

The results from the two-step approach suggested by Donald and Lang (2007), given by equations (5.6) and (5.7), are presented in Table A-4 in the Appendix. All effects are still significant on a ten percent level, which is reassuring. Table A-5 in the Appendix shows that the results do not depend on the inclusion of control variables. Finally, to make sure that the results are robust to other sample selections than individuals who turned 65 from November to February, Table A-6 in the Appendix presents the estimated coefficient on the reform variable for samples of individuals with birthday within 1–3 months around the new year. The pattern of the results is the same also for other sample selections.

The estimation results are not sensitive to the cutoff value for employment. Table A-7 in the Appendix presents the results for different earnings cutoffs and shows that the results are similar for earnings above 0 to 1.5 income base amounts, which is a reasonable range for an employment cutoff. For earnings above 2 income base amounts, about 90,000 SEK in 2007 ($ 12,900), there is no significant effect of the tax credits and the pattern is the same also for higher

Variables	Employment (1)	Taxable labor earnings (2)	Share of previous earnings (3)	Remunerated months (4)
Treatment × Year 2004	0.005	-403	-0.001	0.003
	(0.007)	(2,046)	(0.007)	(0.090)
Treatment × Year 2005	0.010	780	0.009	0.035
	(0.007)	(1,802)	(0.007)	(0.089)
Treatment × Year 2006	-0.008	-1,011	-0.003	0.034
	(0.008)	(1,998)	(0.008)	(0.091)
Treatment × Year 2007	0.007	2,434	0.018***	0.043
	(0.007)	(2,373)	(0.007)	(0.086)
Treatment × Year 2008	0.018***	-191	0.020***	0.191**
	(0.007)	(2,102)	(0.008)	(0.083)
Treatment × Year 2009	0.023***	2,046	0.020***	0.189**
	(0.007)	(2,084)	(0.007)	(0.081)
R-squared	0.366	0.259	0.258	0.448
Observations	181,184	181,184	181,184	181,184

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Nov–Feb 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount. Excluded interactions: 2001–2003. Includes the controls in Table 5.3.

Table 5.4: The effect of the age-targeted tax credits on labor market outcomes, estimation with treatment×year interactions

Variables	Employment (1)	Taxable labor earnings (2)	Share of previous earnings (3)	Remunerated months (4)
Post × 65 in February	0.004	-984	-0.004	0.034
	(0.006)	(1,805)	(0.006)	(0.067)
Post × 65 in January	0.007	-3,372*	-0.007	-0.057
	(0.006)	(1,788)	(0.005)	(0.067)
Post × 65 in December	0.017***	-645	0.010*	0.083
	(0.006)	(1,979)	(0.006)	(0.068)
Post × 65 in November	0.023***	-720	0.016**	0.161**
	(0.006)	(1,858)	(0.006)	(0.070)
Post × 65 in October	0.012**	-3,421*	-0.000	0.103
	(0.006)	(1,797)	(0.006)	(0.068)
R-squared	0.384	0.267	0.270	0.453
Observations	281,944	281,944	281,944	281,944

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Oct–Mar 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount. Excluded interaction: 65 in March. Includes the controls in Table 5.3.

Table 5.5: The effect of the age-targeted tax credits on labor market outcomes, estimation with post×age interactions

cutoff values. This is in accordance with the fact that there was no significant effect on taxable earnings in the main estimation. Table A-8 presents the results for the probability of working more than a certain number of months. The effect is largest for the probability of working more than zero months, and is the same, around 1 percentage point, for the probability of working more than 3, 6 or 9 months. Taken together, the results suggest that the age-targeted tax credits affected the probability of being employed at earnings levels that imply less than full time work.

To sum up, the results suggest that the age-targeted labor tax credits increased employment in the year following the 65th birthday by 1.5 percentage points or about 5 percent, increased earnings as a share of previous earnings by 1.9 percentage point or almost 8 percent and increased the number of remunerated months by 0.137 or about 3 percent for individuals with annual earnings above the 2007 tax liability threshold three to five years earlier. The specification tests support the identifying assumption of parallel trends in the outcome variables, and the results are shown to be robust to a variety of robustness checks.

6.2. Heterogeneous Effects

To better understand the origin of the positive effects found in the main analysis, this section provides an analysis of heterogeneous responses to the tax credits. Table 5.6 presents the results from separate estimations of equation (5.3) for different groups of the population. The first column presents the number of observations in each group along with the F-statistic from testing whether individual characteristics change discontinuously at the time of the reform within the group, as specified by equation (5.8). The test is performed for the covariates in X_{it} excluding the variable that defines the group, and for none of the groups can the joint restriction of all coefficients being zero be rejected, which is reassuring.

Labor supply studies of prime-aged workers typically find that responses are larger among women than men. Male labor force participation is in general high and most men work full time, which gives little room for behavioral responses. Interestingly, panel A shows that the effects of the age-targeted tax credits on labor market outcomes can be attributed entirely to the male subsample. The results suggest that the tax credits increased the probability of employment by 2.4 percentage points, earnings as a share of previous earnings by 3 percentage points and the number of remunerated months by 0.191 for men, while none of the estimates are significantly different from zero for women. For workers at the margin of retirement, men thus seem to be more responsive than women to changes in financial incentives.

Group	Obs Prob>F (1)	Employ-ment (2)	Taxable labor earnings (3)	Share of previous earnings (4)	Remune-rated months (5)
A. Gender					
Men	93,048	0.024***	2,962	0.030***	0.190***
	0.878	(0.006)	(2,274)	(0.007)	(0.073)
Women	88,136	0.005	-368	0.006	0.066
	0.166	(0.006)	(1,299)	(0.005)	(0.072)
B. Education					
Less than high school	61,498	0.006	-114	0.016**	0.100
	0.373	(0.007)	(1,588)	(0.007)	(0.088)
High school	72,492	0.016**	1,196	0.014**	0.079
	0.366	(0.007)	(1,653)	(0.007)	(0.081)
College	47,194	0.021**	4,903	0.025**	0.192*
	0.884	(0.009)	(3,782)	(0.010)	(0.103)
C. Health					
Previously sick	71,986	0.008	-97	0.015**	0.122
	0.403	(0.006)	(1,347)	(0.006)	(0.078)
Not previously sick	109,198	0.020***	2,561	0.021***	0.142**
	0.303	(0.006)	(2,009)	(0.006)	(0.068)
D. Age of spouse					
Older spouse	52,969	0.010	3,723*	0.013*	0.030
	0.246	(0.007)	(2,072)	(0.007)	(0.092)
Younger spouse	68,252	0.028***	2,360	0.033***	0.227***
	0.723	(0.007)	(2,557)	(0.007)	(0.085)
No spouse	59,963	0.006	-882	0.008	0.117
	0.339	(0.007)	(2,100)	(0.008)	(0.089)
E. Type of employment					
Regular employee	169,050	0.011**	711	0.012***	0.140***
	0.176	(0.004)	(1,355)	(0.004)	(0.053)
Self-employed	12,134	0.066***	12,499*	0.100***	0.064
	0.683	(0.018)	(6,446)	(0.027)	(0.212)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Nov–Feb 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount. Separate estimations for different population groups. Includes the controls in Table 5.3.

Table 5.6: Heterogeneous effects of the age-targeted tax credits on labor market outcomes

Panel B indicates that the effects increase with education, which could be due to institutional factors or preferences for work. A recent report to the Swedish government (SOU 2012:28) shows that low income earners who qualify for guaranteed pension at age 65 have little to gain from working past this age, since pension benefits are unaffected by delayed retirement.[20] In this group, women are overrepresented. If individuals have adverse financial incentives for work to begin with, the age-targeted tax credits are unlikely to impact their retirement decision. In addition, since eligibility for unemployment benefits ceases at age 65, the only option for unemployed workers above this age is to retire, and the risk of unemployment is higher among low-educated workers. Occupational agreements may also be more binding for this type of workers, affecting the flexibility for continued work, and social conventions about retirement ages may be particularly strong in certain workplaces. In terms of preferences, individuals who dislike their job or have physically demanding jobs that are hard to pursue at older ages are likely to be less responsive to reforms, which is also linked to education.

Relatedly, individuals in bad health incur a larger disutility of labor and may be less responsive to financial incentives for delayed retirement. Panel C divides the sample according to previous sickness history, where sickness is defined as having collected sickness or disability benefits from the Social Insurance Agency at some point during the baseline years, three to five years earlier. The responses to the age-targeted tax credits are indeed concentrated among the healthier individuals. Several studies have also documented a strong correlation in the timing of retirement between spouses, suggesting an additional value of joint retirement.[21] Panel D separates the sample by the age of the spouse during the baseline years, which is a proxy for the labor market status of the spouse that is exogenous to the reform. Interestingly, the responses to the reform can be almost completely attributed to workers with a younger spouse.

Self-employed are particularly autonomous in their labor supply decision and are little constrained by institutional factors. Panel E presents separate results for self-employed and regular employees, where self-employment is defined based on the main type of employment during the baseline years. The results are striking, suggesting an effect on employment for self-employed of 6.8 percentage points, compared to 1.1 percentage point for regular employees. The effect on earnings as a share of previous earnings is 10 percentage points for self-employed compared to 1.2 percentage points for regular employees.

[20] Individuals who also qualify for housing benefits or other means-tested transfers have even smaller incentives for continued work, since benefits are reduced against income.

[21] See, e.g., Hurd (1990), Baker (2002) and Schirle (2008). Zweimüller et al. (1996) find that husbands are sensitive to changes in their wives incentives but not vice versa.

Even the coefficient on labor earnings is large and significant for this group of workers, while there is no significant effect on the number of remunerated months. The reported taxable labor earnings of self-employed may, however, be a result of tax planning rather than reflecting labor supply.

The heterogeneity analysis showed that the responses to the age-targeted tax credits were largest among men, high educated, workers in good health, workers with a younger spouse and self-employed. Since men are healthier, more likely to have a younger spouse and more likely to be self-employed, the question is if these results are due to men in general responding more to the age-targeted tax credits, or if men respond more because their characteristics affect the responsiveness. To shed further light on potential mechanisms, Tables 5.7 and 5.8 present the results from the heterogeneity analysis separately for men and women. The number of observations in the first column of the two tables indicates that some groups are very small. Individuals in these groups may of course be different in many respects. The F-statistics in the first column suggest that we cannot reject the joint restriction of all coefficients on the covariates being zero, however, except for women who had not been previously sick. For all other groups, the characteristics did not change discontinuously at the time of the reform.

The results by education group, presented in panel A of Table 5.7, suggest that the largest responses for men appeared among workers with high school education. The standard errors are large, however, and the coefficients are not significantly different from each other across education groups. For women, presented in panel A of Table 5.8, there are indications of a positive response only among the college educated. Taken together, education may play a role for the responsiveness to the reform. Panel B in the two tables show that no differences remain across group with different histories of sickness absence once the sample is divided by gender, which suggests that health is not an important determinant of the responsiveness to the tax credits.

Panel C in Table 5.7 shows that, for men, the magnitude of the response cannot be significantly separated across ages of the spouse, although men with no spouse tend to respond the least. Among the few women with a younger spouse, presented in panel C of Table 5.8, there are positive and significant effects of the tax credits on employment and the number of remunerated months, which indicates that the value of joint retirement might be important for women. This is of course a small and selected group, however, and the results should be interpreted with some caution.

Finally, panel D in Table 5.8 shows that self-employed women responded to the age-targeted tax credits, at least in terms of employment. Panel D in Table 5.7 shows that self-employed men responded more than regular employees, but there are substantial effects also for regularly employed men. The

Group	Obs Prob>F (1)	Employ-ment (2)	Taxable labor earnings (3)	Share of previous earnings (4)	Remune-rated months (5)
A. Education					
Less than high school	34,426	0.011	1,499	0.029***	0.142
	0.664	(0.010)	(2,451)	(0.010)	(0.120)
High school	36,343	0.033***	2,105	0.027***	0.224*
	0.609	(0.010)	(2,849)	(0.010)	(0.117)
College	22,279	0.021	7,897	0.035**	0.114
	0.983	(0.013)	(7,104)	(0.018)	(0.152)
B. Health					
Previously sick	31,898	0.025**	2,872	0.031***	0.253**
	0.709	(0.010)	(2,385)	(0.011)	(0.121)
Not previously sick	61,150	0.025***	3,097	0.030***	0.161*
	0.730	(0.008)	(3,190)	(0.009)	(0.091)
C. Age of spouse					
Older spouse	11,645	0.053***	17,159**	0.060***	0.303
	0.505	(0.017)	(7,129)	(0.021)	(0.205)
Younger spouse	54,997	0.025***	2,081	0.033***	0.181*
	0.897	(0.008)	(3,072)	(0.009)	(0.096)
No spouse	26,406	0.012	-1,168	0.011	0.185
	0.218	(0.011)	(3,706)	(0.014)	(0.135)
D. Type of employment					
Regular employee	84,560	0.021***	1,659	0.021***	0.186**
	0.644	(0.006)	(2,349)	(0.007)	(0.076)
Self-employed	8,488	0.055**	16,329*	0.121***	0.274
	0.348	(0.022)	(8,727)	(0.033)	(0.252)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Men turning 65 Nov–Feb 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount. Includes the controls in Table 5.3.

Table 5.7: Heterogeneous effects of the age-targeted tax credits on labor market outcomes, men

Group	Obs Prob>F (1)	Employ-ment (2)	Taxable labor earnings (3)	Share of previous earnings (4)	Remune-rated months (5)
A. Education					
Less than high school	27,072	-0.001	-2,418	-0.001	0.029
	0.542	(0.010)	(1,767)	(0.009)	(0.129)
High school	36,149	-0.001	363	0.002	-0.071
	0.787	(0.009)	(1,634)	(0.008)	(0.111)
College	24,915	0.022*	2,059	0.015	0.264*
	0.903	(0.012)	(3,288)	(0.012)	(0.139)
B. Health					
Previously sick	40,088	-0.004	-2,245	0.004	0.033
	0.355	(0.008)	(1,500)	(0.007)	(0.102)
Not previously sick	48,048	0.013	1,196	0.007	0.096
	0.012	(0.008)	(2,016)	(0.008)	(0.101)
C. Age of spouse					
Older spouse	41,324	-0.002	-548	-0.000	-0.049
	0.175	(0.008)	(1,654)	(0.007)	(0.103)
Younger spouse	13,255	0.035**	893	0.027*	0.420**
	0.243	(0.015)	(3,451)	(0.015)	(0.185)
No spouse	33,557	0.002	-716	0.004	0.051
	0.766	(0.010)	(2,358)	(0.009)	(0.118)
D. Type of employment					
Regular employee	84,490	0.001	-558	0.003	0.087
	0.126	(0.006)	(1,319)	(0.005)	(0.073)
Self-employed	3,646	0.096***	4,901	0.059	-0.353
	0.709	(0.032)	(7,272)	(0.043)	(0.391)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Women turning 65 Nov–Feb 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount. Includes the controls in Table 5.3.

Table 5.8: Heterogeneous effects of the age-targeted tax credits on labor market outcomes, women

fact that low educated and women respond the least to the tax credits, while high educated and self-employed respond the most, is in accordance with the importance of institutions discussed above.

6.3. Separating the Effect of the Two Tax Credits

The results presented thus far show the joint effect of the additional earned income tax credit and the payroll tax credit. In this section, I make an attempt to assess the relative contribution of the two types of tax credits by exploiting that the size of the earned income tax credit varies by income and year, whereas the size of the payroll tax credit is constant across years and equal for all workers.

I use the main estimation strategy from the previous sections, but replace the reform dummy with two new variables capturing the size of the earned income tax credit and the payroll tax credit, respectively:

$$y_{it} = \alpha + \gamma_1 \Delta \log(\tau_{it}^I) + \gamma_2 \Delta \log(\tau_{it}^P) + \beta' X_{it} + \lambda_a + \lambda_t + \varepsilon_{it}, \tag{5.10}$$

where τ_{it}^I is the net-of-income-tax rate on taxable earnings and τ_{it}^P is the net-of-payroll-tax rate on gross earnings. The variable $\Delta \log(\tau_{it}^I)$ is the difference in the natural logarithm of the net-of-income-tax rate that is due to the additional earned income tax credit for workers above age 65. It is constructed using the individual's previous earnings, i.e., the maximum annual earnings three to five years earlier, together with the current year tax rules and can thus be interpreted as the change in potential earnings due to the earned income tax credit. It is positive for workers above age 65 during 2007–2009, and varies across individuals depending on the size of previous earnings and year. The variable $\Delta \log(\tau_{it}^P)$ is the difference in the natural logarithm of the net-of-payroll-tax rate that is due to the payroll tax credit for workers above age 65. It is positive and equal across individuals in the treatment group from 2007.

The difference in the natural logarithm of the net-of-tax rate approximately gives the percentage increase in net earnings, divided by one hundred. The specification of the two variables allows for an interpretation of the coefficients as the level change in the outcome variable from a one hundred percent increase in the net-of-income-tax rate and the net-of-payroll-tax rate, respectively. The mean of $\Delta \log(\tau_{it}^I)$ in the treatment group 2007–2009 is 0.056, which implies that the additional earned income tax credit increased potential net earnings by approximately 5.6 percent on average. The variable $\Delta \log(\tau_{it}^P)$ is 0.137 in the treatment group 2007–2009, which implies that the payroll tax credit increased taxable earnings by approximately 13.7 percent, assuming that the full tax incidence fell on the worker. Combined, the tax credits increased potential net earnings, or the global net-of-tax rate, by approximately 19.3 percent on average. The model captures the differential effects of the two tax

credits under the assumption that there is a proportional relationship between the size of the tax reduction and the change in behavior. Comparing γ_1 and γ_2 measures which of the two tax credits was most correlated with the behavioral changes found in the main analysis.

In addition to the control variables included in the main analysis, I add flexible controls for previous earnings. I rank all individuals in the sample according to their previous earnings and construct indicator variables for each percentile of the earnings distribution. High income earners are more likely to work after age 65, and the size of the earned income tax credit is negatively correlated with income. By controlling for previous earnings, the estimates capture the relative behavior of individuals with the same level of previous earnings in the treatment and the control group, who only differ in terms of labor tax rates.

Table 5.9 presents the results for the three outcome variables for which there was a significant effect in the main analysis. The first three columns show the results for the change in the natural logarithm of the global net-of-tax rate, i.e., the sum of $\Delta\log(\tau_{it}^{I})$ and $\Delta\log(\tau_{it}^{P})$. The last three columns show the results for the separate effects of the two variables. The estimates for the global net-of-tax rate in columns (1) to (3) are positive and significant for all outcome variables and the size of the coefficients are in the same ballpark as those in the main analysis. Using the average change in the global net-of-tax rate of 19.3 percent gives an average effect on employment and earnings as a share of previous earnings of 1.7 percentage points and on the number of remunerated months of about 0.1. This compares well to the estimates reported in Table 5.3. Separately including the two tax credit variables in columns (4) to (6) shows that the effect on employment is primarily attributed to the earned income tax credit variable, but that the opposite holds for earnings as a share of previous earnings, although only significant at a ten percent level, and the number of remunerated months.

The analysis of the relative importance of the two tax credits thus gives contradicting results for different outcome variables. The results also tend to be sensitive to the inclusion and the specification of control variables. The analysis does therefore not provide any conclusive evidence about the relative contribution of the two types of tax credits. The analysis is, however, challenging since the two tax credits were introduced at the same time and applied to the same group of individuals. The estimation strategy is demanding in terms of the variation required to separate the two effects, and the variation in the size of the tax credits by income and year may not be sufficient for generating robust estimates. Edmark, Liang, Mörk and Selin (2012) use a similar approach to evaluate the Swedish earned income tax credit for workers below age 65, exploiting the variation in local tax rates across municipalities. The authors

Variables	Employ-ment (1)	Share of previous earnings (2)	Remune-rated months (3)	Employ-ment (4)	Share of previous earnings (5)	Remune-rated months (6)
EITC + Payroll tax	0.089***	0.090***	0.532**			
	(0.021)	(0.026)	(0.261)			
EITC				0.256***	-0.028	-1.227
				(0.067)	(0.176)	(0.931)
Payroll tax				-0.001	0.153*	1.487***
				(0.045)	(0.081)	(0.567)
Constant	0.134***	0.213***	2.784***	0.134***	0.213***	2.779***
	(0.010)	(0.013)	(0.132)	(0.010)	(0.013)	(0.132)
R-squared	0.144	0.061	0.101	0.144	0.061	0.101
Observations	181,184	181,184	181,184	181,184	181,184	181,184

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Nov–Feb 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount. Includes the controls in Table 5.3, and indicator variables of the percentile of previous earnings.

Table 5.9: Separate effects of the earned income tax credit and the payroll tax credit on labor market outcomes

find a positive effect of the earned income tax credit but also a positive placebo effect, which may be due to the failure of the method to properly account for underlying trends in labor market outcomes. The problems encountered in the above analysis may well be similar to theirs.

6.4. Policy Analysis

The main analysis established a positive joint effect of the age-targeted tax credits, and the heterogeneity analysis indicated that the impact was primarily driven by men. Unfortunately, the attempt to separate the effect of the two tax credits gave no conclusive result. Still, for the reasons outlined in Section 3, we might expect the earned income tax credit to have a larger impact than the payroll tax credit. This section evaluates the size of the effects found in the above analysis. First, to facilitate comparison with other labor supply studies, I translate the estimated effects into elasticities. Second, I analyze the public finance implications of the reform. Both analyses are performed under the two scenarios that both tax credits contributed equally to the behavioral changes, on the one hand, or that the changes can be entirely attributed to the earned income tax credit, on the other hand. The first scenario is the most appropriate since there is no evidence of a differential impact, but the second scenario is an interesting alternative hypothesis. Furthermore, to give an indication of the fiscal implications in the hypothetical scenario that the reform effect for women would be as large as that for men, the public finance analysis is performed only for men as well as for the full sample.

6.4.1. Implied Elasticities

Table 5.10 translates the estimated reform effects from the main analysis into elasticities, which can be interpreted as the percentage change in the outcome variable from a one percent increase in net labor income, due to the age targeted tax credits. The elasticities are presented for the full sample in panel A and for men only in panel B. The first row in each panel assumes that both tax credits were equally important for the observed effect. The second row assumes that the entire effect was due to the earned income tax credit, and presents the elasticity only with respect to the earned income tax credit. The elasticities are calculated as the percentage change in the outcome variable, from the last row of Table 5.3, divided by the percentage change in net labor earnings due to the different tax credits. The change in net labor earnings due to the earned income tax credit is calculated from the average change in the logarithm of the net-of-income-tax rate due to the earned income tax credit for workers above age 65 in the treatment group during 2007–2009, which is 7.11 percent.[22] The change in net labor earnings due to the payroll tax credit is calculated as the change in the natural logarithm of the net-of-payroll-tax-rate from the 2007 reform, which is 13.68 percent. The total change is calculated as one minus the product of the net-of-income-tax rate and the net-of-payroll-tax rate, which is 19.82 percent.[23]

	Employment (1)	Share of previous earnings (2)	Remunerated months (3)
A. Full sample			
Elasticity, effect from EITC and payroll tax credit	0.248	0.383	0.148
Elasticity, effect only from EITC	0.691	1.068	0.412
B. Men			
Elasticity, effect from EITC and payroll tax credit	0.349	0.565	0.195
Elasticity, effect only from EITC	1.006	1.631	0.563

Table 5.10: Implied elasticities

The first row in panel A of Table 5.10 suggests an elasticity of employment of 0.25, an elasticity of earnings as a share of previous earnings of 0.38 and an elasticity of the number of remunerated months of 0.15, under the assumption that both tax credits contributed equally to the observed effects. Assuming

[22]This is based on *realized* earnings whereas the values in Section 6.3 were based on *potential* earnings.

[23]For men only, the change due to the earned income tax credit is 6.77 percent and the total change is 19.52 percent.

that the entire effect was due to the earned income tax credit, in the second row, gives an elasticity of employment of 0.69, an elasticity of earnings as a share of previous earnings of 1.07 and an elasticity of the number of remunerated months of 0.41. The elasticities for men, reported in panel B, are larger, ranging from 0.2 to 0.57 when accounting for both tax credits and from 0.56 to 1.63 when only accounting for the earned income tax credit. The elasticities for women are not reported since there were no significant effects of the tax credits for this group in the main analysis. They should thus be close to zero.

The estimated elasticities can be compared to results in the previous literature. Hotz and Scholz (2003) summarize the literature on the effects of expansions in the U.S. earned income tax credit on labor force behavior. They find elasticities of labor force participation with respect to net income of 0.69 to 1.69 among single women with children, which was the group primarily affected by the EITC, and regard elasticities of this size as large. Immervoll, Kleven, Kreiner and Saez (2007) review the evidence on participation elasticities in the labor supply literature, and find elasticities of 0.5 to 1 for certain subgroups of the population, such as married women and low income earners, and elasticities close to zero for prime-aged males. In their modeling, the authors use an average participation elasticity of 0.2, ranging from 0.4 in the lowest income decile to 0 in the highest income decile. In simulation studies of the Swedish earned income tax credit for workers below age 65, Sacklén (2009) obtains a relatively low participation elasticity of 0.1, whereas Aaberge and Flood (2008) obtain a participation elasticity of 0.3 for single mothers.

When only considering the earned income tax credit, the implied elasticity of employment with respect to net income in this study is high and compares well to the participation elasticities for women and low income earners in the previous literature. When considering the total impact of the earned income tax credit and the payroll tax reduction, however, the elasticities are small, at least if we believe that labor force participation should be more elastic among older workers than in the population in general. The elasticities do, however, compare with the participation elasticities for single women in Sweden from the simulation model in Aaberge and Flood (2008).

6.4.2. Public Finance Implications

The size of the estimated effects can also be evaluated by analyzing the public finance implications of the reform. As discussed in Palme and Svensson (2007), the overall change in public sector finances can be decomposed into a "mechanical" and a "behavioral" component. The mechanical component is the change in public finances if individuals would not change their retirement behavior, whereas the behavioral component is the change in public fi-

nances from the changes in retirement behavior due to the reform. Table 5.11 presents the results from such a decomposition. The first column shows the results for all individuals in the treatment group after the reform, i.e., individuals who turned 65 in November and December during 2007–2009 with previous earnings above the 2007 tax liability threshold. The second column presents the results for males only. Since the effects appeared to be driven primarily by men, these results give an indication of the public finance implications if women could be influenced to respond as much as men. The analysis accounts for the benefits and costs only during the tax year immediately following the 65th birthday for the treatment group, combined over the years 2007 to 2009.

The statistics used for the calculations are presented in panel A. To perform the analysis, I need to make a number of simplifying assumptions. I only consider the estimated effect on the employment outcome variable. To measure the behavioral effect, I use the employment estimate to calculate the number of individuals who were added to the labor force as a result of the reform. I assume that the individuals who were encouraged to work by the reform have the same distribution of earnings as the individuals who would have worked regardless of the reform. Thereby, the realized average earned income tax credit, average payroll tax credit, average labor income taxes and average net labor income in the treatment group during 2007–2009 can be taken as representative for all individuals. The analysis is simplified by the fact that the public pension system is actuarially fair. A delayed labor force exit therefore has no public finance implications in terms of public pension payments. Workers above age 65 are also not eligible for social programs such as unemployment and disability benefits.

Panel B calculates the mechanical cost of the reform for individuals in the treatment group during 2007–2009, disregarding the behavioral effect. The costs are the payments of the earned income tax credit and the payroll tax credit, given by the average size of the tax credit payments times the number of individuals attributed to the mechanical component. I assume that all tax credit payments are used for consumption, which implies that there are also benefits of the reform in terms of value added tax payments. Following Pirttilä and Selin (2011), I use an effective value added tax rate of 21 percent. Adding the costs from the tax credit payments and subtracting the benefits from the value added tax payments gives the total mechanical cost of the reform, amounting to 405 million SEK for the full sample and 258 million SEK for the male sample.

Panel C calculates the benefits of the reform that can be attributed to the behavioral component. In terms of income taxes, the benefits are the additional personal income tax and payroll tax payments of individuals induced to work by the reform. Assuming that all income is used for consumption, there are also benefits from value added taxes on net labor income. Adding these bene-

fits together gives the total benefits of the reform that can be attributed to the behavioral component, amounting to about 54 million SEK for both the full and the male sample. The benefits are similar for the two samples since the behavioral effects primarily could be attributed to men.

	Men and women (1)	Men (2)
A. Statistics used for calculations		
Employment estimate	0.015	0.024
No. individuals	35,142	18,207
No. employed individuals	11,951	7,140
of which attributed to mechanical component	11,422	6,698
of which attributed to behavioral component	529	442
Mean EITC, SEK	8,486	8,475
Mean payroll tax credit, SEK	36,423	40,349
Mean labor income taxes, SEK	69,872	87,708
Mean net labor income, SEK	155,517	161,974
B. Mechanical component		
Costs, EITC, million SEK	96.9	56.8
Costs, payroll tax credit, million SEK	416.0	270.3
Benefits, VAT on EITC, million SEK	20.4	11.9
Benefits, VAT on payroll tax credit, million SEK	87.4	56.8
Total costs, mechanical component, million SEK	405.0	258.3
C. Behavioral component		
Benefits, labor income taxes, million SEK	37.0	38.8
Benefits, VAT on net labor income, million SEK	17.3	15.0
Total benefits, behavioral component, million SEK	54.2	53.8
D. Behavioral benefits as a share of mechanical costs		
Share, effect from EITC and PT	0.134	0.208
Share, effect only from EITC	0.708	1.200
E. Estimates required for a fully financed reform		
Employment estimate, effect from EITC and PT	0.112	0.117
Employment elasticity, effect from EITC and PT	1.762	1.634
Employment estimate, effect only from EITC	0.021	0.020
Employment elasticity, effect only from EITC	0.948	0.832

Individuals turning 65 Nov–Dec 2007–2009 with previous earnings above 17,100 (2007 SEK). Degree of self financing is the sum of extra income taxes, extra VAT income and extra VAT tax credits total divided by the total cost. Values in 2007 SEK.

Table 5.11: Public finance implications of the age-targeted tax credits

Panel D calculates the degree to which the benefits from the behavioral changes counteract the mechanical costs of the reform. This can be thought of as the reform's degree of "self financing". The first row presents the results of this exercise under the assumption that both tax credits contributed equally to the observed changes in retirement behavior, whereas the second row presents the results assuming that all of the effect was due to the earned income tax credit. The costs of the earned income tax credit and the payroll tax credit are large, and the behavioral changes can only account for 13 percent of the

mechanical costs of the reform for the full sample, and 21 percent for the male sample. If attributing all of the behavioral changes to the earned income tax credit, and disregarding the costs of the payroll tax credit, the degree of self financing increases to 70 percent for the full sample and 120 percent for the male subsample. Under those assumptions, the behavioral changes for males are in fact large enough to fully finance the reform.

Finally, panel E presents calculations of the employment estimates and elasticities that would be required for the benefits from the behavioral component to fully finance the costs of the mechanical component of the reform. Considering the total costs of the earned income tax credit and the payroll tax credit, in the first two rows, would require employment estimates of 0.11–0.12 and employment elasticities of 1.6–1.8, which are arguably very large. Only considering the earned income tax credits, in the two last rows, reduces the required employment estimates to about 0.02 and the employment elasticities to 0.8–0.9, which are of more reasonable sizes compared to estimates in the previous literature.

The assumptions made in the analysis can of course be questioned. First, the estimated effect on taxable earnings might be a more appropriate benchmark than the effect on employment. The effect on earnings could not be significantly separated from zero in the analysis, which implies that the behavioral component would be zero. Since the other measures indicated a positive labor supply response, however, there is reason to believe that there is a behavioral response that the earnings variable is unable to recover. Second, even if the effect on employment is used as a benchmark, the assumption that the realized earnings distribution is representative for individuals induced to work by the reform can be questioned. It might be more likely that the individuals who entered employment as a result of the reform work less than those who would have continued to work regardless of the reform. Both of these concerns would imply that the behavioral component is overrated, and that the degree of self financing is even lower.

There are also other caveats to the analysis. First, the control group can be argued to also be treated by the reform. The tax credits will accrue to the control group with a delay, one year later than to the treatment group, which might induce them to remain in employment until becoming eligible for the tax credits. Such behavior would attenuate the behavioral effects. In addition, the results from the main analysis suggest that the responsiveness to the reform has increased over time, and may hence be even larger in the future. Furthermore, the analysis only considers the costs and benefits of the reform during the year immediately following the 65th birthday, for individuals in the narrowly defined treatment group. This is the only effect that can be established in the paper. A full analysis would take into account the costs and benefits for all

individuals affected by the reform. Finally, the analysis does not take welfare, but only public finance considerations, into account. A full welfare analysis may give very different implications, but is beyond the scope of this paper.

7. Conclusion

This paper studies the impact of an earned income tax credit and a payroll tax credit targeted at workers above age 65 that were simultaneously introduced in Sweden in 2007. The identification strategy exploits that the eligibility for the two tax credits was governed by the year of birth, while the social security system is governed by age at retirement, which creates a discontinuity in labor tax rates that is due to the timing of the 65th birthday around the new year. Other differences between the groups are controlled for using a difference-in-differences approach, to the extent that they do not change discontinuously at the time of the reform. The sample consists of individuals who turned 65 from November to February between 2001 and 2009 and had annual earnings above the 2007 tax liability threshold three to five years earlier.

The results suggest a positive effect of the age-targeted tax credits on labor market outcomes in the estimation sample. In the year immediately following the 65th birthday, employment in the treatment group increased by 1.5 percentage points, annual earnings as a share of previous earnings increased by 1.8 percentage points and the number of remunerated months increased by 0.13, although the effect on taxable labor earnings is insignificant. The results are shown to be robust to a variety of robustness checks. An analysis of heterogeneous responses indicates that the effects can primarily be attributed to men, increase with education and are particularly large for self-employed. Other institutions may discourage a delayed retirement for women and low educated workers, and the tax credits may need to be combined with other reforms to impact the retirement behavior of these groups. To study the relative contribution of the two types of labor tax credits I use the variation in the size of the credits by income and year, but the analysis gives no conclusive result.

I also evaluate the size of the effects found in the analysis. To facilitate comparison with previous studies, the estimated effects are translated into elasticities. Attributing the behavioral changes to both tax credits, which is appropriate since there was no evidence of a differential impact, gives implied elasticities of 0.25 for employment, 0.38 for annual earnings as a share of previous earnings and 0.15 for the number of remunerated months. In addition, a rough analysis of the public finance implications of the reform suggest that the benefits from the behavioral changes amount to only a small share of reform costs for the estimation sample. Much of the previous research on earned income tax credits has focused on low income earners. This paper provides

new evidence of the effectiveness of tax credits for workers at the margin of retirement. The implied elasticities found in this study are small compared to the elasticities in the U.S. earned income tax credit literature. Interestingly, since the costs of the two tax credits were substantial, a fully financed reform would require unusually large employment elasticities, compared to what has typically been found in the labor supply literature.

Two previous reports to the Swedish government have analyzed the impact of the age-targeted tax credits on labor market outcomes using a difference-in-differences approach. Pirttilä and Selin (2011) compare the employment rates from the Labor Force Survey of the age groups 65–74 and 55–64 before and after the reform in 2007, and find an increase in employment of 2 percentage points or 19.4 percent in the older age group. The authors emphasize that the analysis is suggestive and merely points in interesting directions for future work. The Ministry of Finance (2012) compare labor market outcomes between individuals aged 66 and 64 and between individuals who turned 65 in the first and the last quarter of the year, before and after 2007. The former analysis suggests an employment increase of 7 percentage points, whereas the estimates in the latter analysis are insignificant. The advantage of the present analysis compared to these studies is that the control and treatment groups are more similar due to the narrow age span of only a few months. With less similar groups it may be difficult to account for differential trends over time. The results in this study suggest that the previous estimates are likely to overstate the change in employment.

Although the tax credits, in terms of foregone tax revenues, showed to be a costly way to increase labor supply at older ages, changes in retirement behavior may have long-run benefits that are not accounted for in this study and that may interact with other possible future policy changes. There are also other limitations to the analysis in this paper. Since the control group eventually becomes eligible, the analysis captures the effect of receiving the tax credits one year earlier. If the control group delays their labor force exit until becoming eligible, the estimated effects will be attenuated. In addition, only the short-term effects during the year immediately following the 65th birthday are being analyzed. Instead of a static model, one could consider a dynamic model in which the tax credits affect the option value of retirement. This may reduce the change in financial incentives and lead to larger implied elasticities. Considering the dynamic effects of the age-targeted tax credits on retirement incentives is an avenue for future research.

References

Aaberge, R., and L. Flood (2008), "Evaluation of an In-Work Tax Credit Reform in Sweden: Effects on Labor Supply and Welfare Participation of Single Mothers," IZA Discussion Paper No. 3736, Institute for the Study of Labor.

Baker, M. (2002), "The Retirement Behavior of Married Couples. Evidence from the Spouse's Allowance," *Journal of Human Resources*, vol. 37(1), pages 1–34.

Bennmarker, H., E. Mellander, and B. Öckert (2009), "Do Regional Payroll Tax Reductions Boost Employment?," *Labour Economics*, vol. 16, pages 480–489.

Bohm, P., and H. Lind (1993), "Policy Evaluation Quality – A Quasi-Experimental Study of Regional Employment Subsidies in Sweden," *Regional Science and Urban Economics*, vol. 23, pages 51–65.

Blundell, R., and H. Hoynes (2004), "Has 'In-Work' Benefit Reform Helped the Labor Market?," in D. Card, R. Blundell, and R. B. Freeman (eds.), *Seeking a Premier Economy: The Economic Effects of British Economic Reforms 1980–2000*, Chicago: The University of Chicago Press, pages 411–460.

Blundell, R., and T. MaCurdy (1999), "Labor Supply: A Review of Alternative Approaches," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics Volume 3*, Elsevier Science B. V., pages 1559–1695.

Chetty, R., and E. Saez (2009) "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients," NBER Working Paper No. 14836, National Bureau of Economic Research.

Donald, S. G., and K. Lang (2007), "Inference with Difference-in-Differences and other Panel Data," *The Review of Economics and Statistics*, vol. 89(2), pages 221–233.

Edmark, K., C-H. Liang, E. Mörk, and H. Selin (2012), "Evaluation of the Swedish Earned Income Tax Credit," IFAU Working Paper No. 2012:1, Institute for Evaluation of Labour Market and Education Policy.

Eissa, N., and H. Hoynes (2006), "Behavioral Responses to Taxes: Lessons from the EITC and Labor Supply," in J. M. Poterba (ed.), *Tax Policy and the Economy 20*, Cambridge: MIT Press.

Ericson, P., and L. Flood (2009), "A Microsimulation Approach to an Optimal Swedish Income Tax," IZA Discussion Paper No. 4379, Institute for the Study of Labor.

Fredriksson, P., and B. Öckert (2009), "The Effect of School Starting Age on School and Labor Market Performance," mimeo, Stockholm University.

French, E. (2005), "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies*, vol. 72, pages 395–427.

Gruber, J. (1997), "The Incidence of Payroll Taxation: Evidence from Chile," *Journal of Labor Economics*, vol. 15(3), pages S72–S101.

Gruber, J., and D. A. Wise (1999), *Social Security and Retirement Around the World*, Chicago: The University of Chicago Press.

Holmlund, H. (2007), "A Researcher's Guide to the Swedish Compulsory School Reform," Working Paper No. 9/2007, Swedish Institute for Social Research.

Hotz, V. J. and J. K. Scholz (2003), "The Earned Income Tax Credit," in R. A. Moffit (ed.), *Means-tested transfer programs in the United States*, Chicago: The University of Chicago Press.

Hurd, M. D. (1990), "The Joint Retirement Decision of Husbands and Wives," in D. A. Wise (ed.), *Issues in the Economics of Aging*, Chicago: The University of Chicago Press.

Huttunen, K., J. Pirttilä, and R. Uusitalo (2010), "The Employment Effects of Low-Wage Subsidies," IZA Discussion Paper No. 4931, Institute for the Study of Labor.

Immervoll, H., H. J. Kleven, C. T. Kreiner, and E. Saez (2007), "Welfare Reform in European Countries: A Microsimulation Analysis," *Economic Journal*, vol. 117, pages 1–44.

Lehmann, E., F. Marical, and L. Rioux (2011), "Labor Earnings Respond Differently to Income-Tax and to Payroll-Tax Reforms," mimeo, CREST.

Liebman, J. B. (1998), "The Impact of the Earned Income Tax Credit on Incentives and Income Distribution," in J. M. Poterba (ed.), *Tax Policy and the Economy 12*, Cambridge: MIT Press.

Lumsdaine, R. L., and O. S. Mitchell (1999), "New Development in the Economic Analysis of Retirement," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics Volume 3*, Elsevier Science B. V., pages 3261–3307.

Meghir, C., and M. Palme (2005), "Educational Reform, Ability and Family Background," *American Economic Review*, vol. 95(1), pages 414–424.

Palme, M., and I. Svensson (2007), "Financial Implications of Income Security Reforms in Sweden," in J. Gruber and D. A. Wise (eds.), *Social Security Programs and Retirement Around the World: Fiscal Implications of Reform*, Chicago: The University of Chicago Press.

Pirttilä, J., and H. Selin (2011), "Tax Policy and Employment: How Does the Swedish System Fare?" Working Paper 2011:2, Department of Economics, Uppsala University.

Riksrevisionen (2009), "Jobbskatteavdraget," RiR 2009:20.

Sacklén, H. (2009), "Arbetsutbudseffekter av reformer på inkomstskatteområdet 2007–2009," Report, Swedish Ministry of Finance.

Schirle, T. (2008), "Why Have the Labor Force Participation Rates of Older Men Increased since the Mid-1990s?," *Journal of Labor Economics*, vol. 26(4), pages 549–594.

SOU 2012:28, *Längre liv, längre arbetsliv*, Delbetänkande av Pensionsåldersutredningen.

Summers, L. H. (1989), "Some Simple Economics of Mandated Benefits," *American Economic Association Papers and Proceedings*, vol. 79, pages 177–183.

Zweimüller, J., R. Winter-Ebmer, and J. Falkinger (1996) "Retirement of Spouses and Social Security Reform," *European Economic Review*, vol. 40, pages 449–472.

Appendix

	Treatment Group (1)	Control Group (2)
Female	0.485	0.488
Less than High School	0.349	0.331
High School	0.399	0.401
College	0.252	0.268
Immigrant	0.108	0.113
Previously sick	0.395	0.399
No spouse	0.309	0.307
Older spouse	0.314	0.303
Younger spouse	0.378	0.390
Self-employed	0.068	0.066
Previous earnings	240,837	242,403
Observations	86,161	95,023

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Treatment and control groups consist of individuals turning 65 in Nov–Dec and Jan–Feb, respectively, 2001–2009, with previous earnings above 17,100 (2007 SEK).

Table A-1: Summary statistics

Variables	Reform indicator (1)	Included in the estimation sample (2)
Included	0.0016	
	(0.0010)	
Female		-0.008***
		(0.002)
High School		0.075***
		(0.002)
College		0.170***
		(0.002)
Immigrant		-0.183***
		(0.002)
Self-employed		0.211***
		(0.004)
Previously sick		-0.197***
		(0.002)
Older spouse		0.043***
		(0.002)
Younger spouse		0.082***
		(0.002)
Constant	0.2026***	0.711***
	(0.0017)	(0.002)
Year dummies	Yes	No
Age dummies	Yes	No
R-squared	0.6088	0.122
Observations	257,104	257,104

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Nov–Feb 2001–2009.

Table A-2: The effect of being included in the estimation sample on the reform indicator and the effect of covariates on the probability of being included in the sample

Variables	Employ-ment (1)	Taxable labor earnings (2)	Share of previous earnings (3)	Remune-rated months (4)
Reform	0.002	399	0.097	0.009
	(0.002)	(267)	(0.125)	(0.036)
Constant	0.015***	3,114***	0.320***	0.664***
	(0.002)	(312)	(0.101)	(0.043)
R-squared	0.006	0.009	0.001	0.017
Observations	75,920	75,920	75,920	75,920

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Nov–Feb 2001–2009 with previous earnings below 17,100 (2007 SEK). Employed if earnings>1 income base amount. Includes the controls in Table 5.3.

Table A-3: The effect of the age-targeted tax credits on labor market outcomes for the excluded sample

Variables	Employ-ment (1)	Taxable labor earnings (2)	Share of previous earnings (3)	Remune-rated months (4)
Reform	0.015*	1,529	0.018***	0.128**
	(0.007)	(1,532)	(0.004)	(0.041)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Nov–Feb 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount. Includes the controls in Table 5.3.

Table A-4: The effect of the age-targeted labor tax credits on labor market outcomes using the Donald and Lang (2007) two-step approach

Variables	Employ-ment (1)	Taxable labor earnings (2)	Share of previous earnings (3)	Remune-rated months (4)
Reform	0.015***	1,493	0.018***	0.133**
	(0.004)	(1,389)	(0.005)	(0.053)
Constant	0.180***	31,736***	0.143***	3.336***
	(0.004)	(963)	(0.004)	(0.049)
Controls	No	No	No	No
Year dummies	Yes	Yes	Yes	Yes
Age dummies	Yes	Yes	Yes	Yes
R-squared	0.024	0.021	0.015	0.015
Observations	181,184	181,184	181,184	181,184

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Nov–Feb 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount.

Table A-5: The effect of the age-targeted labor tax credits on labor market outcomes, without controls

Sample	Employ-ment (1)	Taxable labor earnings (2)	Share of previous earnings (3)	Remune-rated months (4)
December-January	0.011*	2,720	0.017***	0.141*
	(0.006)	(1,933)	(0.006)	(0.072)
November-February	0.015***	1,518	0.018***	0.133***
	(0.004)	(1,335)	(0.004)	(0.051)
October-March	0.014***	-232	0.012***	0.122***
	(0.003)	(1,076)	(0.004)	(0.041)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Oct–Mar 2001–2009 with previous earnings above 17,100 (2007 SEK). Employed if earnings>1 income base amount. Includes the controls in Table 5.3.

Table A-6: The effect of the age-targeted labor tax credits on labor market outcomes for different sample selections

Variables	> 0 (1)	> 0.5 BA (2)	> 1 BA (3)	> 1.5 BA (4)	> 2 BA (5)	> 3 BA (6)
Reform	0.0170***	0.0222***	0.0150***	0.0107***	0.00332	0.00279
	(0.00458)	(0.00449)	(0.00428)	(0.00407)	(0.00387)	(0.00351)
Constant	0.451***	0.338***	0.282***	0.239***	0.208***	0.163***
	(0.00575)	(0.00543)	(0.00506)	(0.00474)	(0.00447)	(0.00397)
R-squared	0.070	0.083	0.086	0.085	0.084	0.081
Observations	181,184	181,184	181,184	181,184	181,184	181,184

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Oct–Mar 2001–2009 with previous earnings above 17,100 (2007 SEK). Includes the controls in Table 5.3. BA=income base amount, 45,900 SEK in 2007 ($ 6,600).

Table A-7: The effect of the age-targeted labor tax credits on the probability of earning more than a certain cutoff in income base amounts

Variables	> 0 months (1)	> 3 months (2)	> 6 months (3)	> 9 months (4)
Reform	0.0213***	0.00977**	0.0111**	0.0109**
	(0.00460)	(0.00453)	(0.00448)	(0.00442)
Constant	0.388***	0.338***	0.316***	0.306***
	(0.00568)	(0.00550)	(0.00541)	(0.00534)
R-squared	0.075	0.073	0.071	0.069
Observations	181,184	181,184	181,184	181,184

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Individuals turning 65 Oct–Mar 2001–2009 with previous earnings above 17,100 (2007 SEK). Includes the controls in Table 5.3.

Table A-8: The effect of the age-targeted labor tax credits on the probability of working more than a certain number of months

Chapter 6

Wage Dynamics and Firm-Level Shocks*

Benjamin Friedrich Lisa Laun Costas Meghir Luigi Pistaferri

Abstract

This paper proposes a framework for introducing the firm into empirical models of the dynamic income process. The firm side has typically been neglected in the literature studying the nature of the income process, but the recent availability of matched employer-employee data gives rise to major new opportunities in this direction. A key contribution is that the model allows for studying the extent to which firm-level productivity shocks are transmitted to wages. To disentangle shocks from endogenous responses to shocks, selection into employment and between jobs is explicitly modeled. The proposed model allows for a rich stochastic structure of the income process, both at the individual and the firm-worker match level, and is consistent with most findings in the literature. Based on the method of indirect inference, we present a strategy for estimation and identification of the model.

Keywords: Income process, Wage dynamics, Firm dynamics
JEL-codes: H51, H55, I18, J26

*We are grateful for comments from Mårten Palme and seminar participants at the Institute for Evaluation of Labour Market and Education Policy, the Conference on Matched Employer-Employee Data: Developments since AKM, the 3rd Workshop on Industry and Labor Market Dynamics and the IFAU/UCLS Workshop on Linked Employer-Employee Data. Lisa Laun gratefully acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation.

1. Introduction

There has been an increased interest in understanding pay policies of firms, and in particular the extent to which firm-level productivity shocks are transmitted to workers' wages. Such departures from perfect competition and the law of one price have been motivated by the developments in search theory starting by the seminal models of Burdett and Mortensen (1998) and Mortensen and Pissarides (1994). In a competitive labor market, workers only bare the risk of shocks to their productivity, but in the presence of search frictions, there are multiple sources of risk distinct from workers' productivity shocks.[1] While the theoretical justifications for departures from the law of one price are compelling, the empirical evidence is not quite there.

First, most equilibrium search models that have been estimated on empirical data assume no productivity shocks. These include recently Postel-Vinay and Robin (2002) and Cahuc, Postel-Vinay and Robin (2006). An exception is the model of Lise, Meghir and Robin (2012) which allows for the effect of firm-level productivity shocks in a context of a model with productive complementarities. However, their model is estimated on individual-level data and hence cannot measure directly the productivity shocks, but infers them from the structure of the model. The recent availability of matched employer-employee data gives rise to major new opportunities in this direction. Second, models that have been estimated on matched employer-employee data, without a specific economic structure, such as Abowd, Kramarz and Margolis (1999), have focused on sorting and firm-worker heterogeneity rather than the dynamics of shocks. A recent paper by Guiso, Pistaferri and Schivardi (2005) has indeed measured the impact of firm-level shocks on wages using matched employer-employee data. However, their approach is limited by the fact that they ignore job to job mobility and transitions between employment and unemployment. Such transitions may well hide the impact of firm-level shocks on wages because a worker may quit instead of suffering too large a pay cut.

In this paper, we propose a framework for introducing the firm in empirical models of the dynamic income process. In particular, the proposed model allows for studying the extent to which firm-level productivity shocks are transmitted to wages. The key innovation is that we account both for job to job transitions and for transitions between employment and unemployment to capture the role of job mobility and labor force participation in hiding the impact of shocks. The model allows for a rich stochastic structure of the income process, both at the individual and the firm-worker match level, and is consistent with most findings in the literature. Based on the method of indirect inference,

[1] Low, Meghir and Pistaferri (2010) illustrate the importance of such distinctions for the welfare effects of risk.

we present a strategy for estimation of the model and identification of the key parameters.

The proposed approach is not structural in the sense that we do not specify a model that defines the way pay setting is carried out. While an equilibrium model defines clearly the way that shocks are transmitted, a tractable model comes with a number of strong assumptions both on the form of contracting and on the structure of production.[2] Thus our paper is the opening investigation into an agenda that will lead to richer structural models taking the firm side more seriously and allowing for a rich dynamic stochastic structure of the income process.

The model of the income process is inspired by the increased availability of matched employer-employee data. Although the paper does not contain an empirical contribution, the data we have in mind for estimation is a matched employer-employee panel data set at a quarterly level. On the worker side, the data should contain hourly wages, employment status, firm identifiers and a rich set of individual characteristics such as gender, age, education, marital status, the presence of children and financial income. On the firm side, the data should contain a measure of firm performance such as revenue or value-added, and firm characteristics such as sector and firm size.

The paper proceeds as follows. Section 2 provides a discussion of the related literature. Section 3 presents the model of the income process. Section 4 presents a strategy for estimation and identification of the model. Section 5 concludes the paper.

2. Previous Literature

Many countries have experienced increased earnings inequality over the last decades.[3] Whether this is due to increased permanent inequality or increased volatility has important implications for aggregate savings and consumption as well as for policy design. Dating back to Friedman (1957), theory predicts that consumption responds fully to unanticipated permanent changes in income but not at all to anticipated changes.[4] Furthermore, the consumption response to unanticipated transitory income changes will equal the annuity value of the shock, which is typically very small. The distinctions between anticipated changes and shocks and between permanent and transitory changes to income are therefore key to understanding consumption and savings behavior. Also the

[2]See Lamadon, Lise, Meghir and Robin (2012) for an effort in this direction.

[3]See Meghir and Pistaferri (2011) for a comprehensive review of the literature on the nature of the income process.

[4]See Meghir (2004) for an extensive discussion of Friedman's permanent income hypothesis.

policy responses to changes in inequality differ depending on the nature of risk. Permanent inequality can be alleviated by, e.g., education reforms, whereas earnings volatility can be mitigated by, e.g., income maintenance programs. As exemplified by Meghir and Pistaferri (2004), transitory shocks to income include over-time labor supply, piece-rate compensation and bonuses, whereas permanent shocks can occur due to job mobility, long-term unemployment, health shocks or promotions.

In the literature on earnings dynamics, the income process is typically described as

$$\ln Y^e_{i,a,t} = d^e_t + \beta'^e X_{i,a,t} + u_{i,a,t},$$

where $Y_{i,a,t}$ is a measure of income, such as annual earnings, for individual i of age a in period t and $X_{i,a,t}$ include individual characteristics, most importantly a polynomial in age. The index e denotes a specific group defined by, e.g., gender and education, and d^e_t capture group-specific aggregate time effects, reflecting the price of human capital. The interest is to study the time series properties of the error term $u_{i,a,t}$, which is typically represented by the residual logarithm of income from an OLS regression, $\hat{y}_{i,a,t} = \ln Y_{i,a,t} - \hat{d}_t - \hat{\beta}'_{i,a,t} X_{i,a,t}$.

In this literature, earnings are considered exogenous and income risk is the volatility in earnings that cannot be explained by observable characteristics, represented by the variance of the error term. The error term is usually modeled as the sum of a number of random components. A specification that covers most of the aspects in the literature includes a permanent component, a transitory component, an individual-specific growth term and a measurement error. The stochastic structure is evaluated by studying the covariance matrix of the changes in residual income. Different time series structures have different implications for the variance and auto covariances of the error term.

Much effort has been devoted to studying the persistence of shocks to income. Although transitory shocks are mean-reverting, they may be somewhat persistent. The order of serial correlation in the transitory component can be evaluated based on the order to which changes in residual income are correlated. Several studies, including Lillard and Willis (1978), MaCurdy (1982), Abowd and Card (1989) and Meghir and Pistaferri (2004), have found the transitory component to be best modeled as serially correlated with low persistence, implying that transitory shocks may last for longer than one period.

The first study of the income process, Lillard and Willis (1978), modeled the permanent component of the income process as a fixed effect. Later studies, including MaCurdy (1982), Abowd and Card (1989) and Meghir and Pistaferri (2004), have argued that the permanent component may be subject to random shocks. This implies that the evolution of earnings is a random walk with drift, which can explain the empirical observation that the variance of earnings increases over the life cycle. The implication of a unit root in income is that

innovations in the past do not predict earnings growth, i.e., that the current wage, adjusted for predictable growth due to observable characteristics, is the best predictor of future wages. As is shown by Meghir and Pistaferri (2004), knowing the order of serial correlation of the transitory component is enough to recover the variance of the permanent shock. The key moment condition for identifying the variance of the permanent shock is:

$$E\left(\zeta_{i,a,t}^2\right) = E\left[g_{i,a,t}\left(\sum_{j=-(1+q)}^{(1+q)} g_{i,a+j,t+j}\right)\right],$$

where $\zeta_{i,a,t}$ is the permanent shock, $g_{i,a,t}$ is the change in residual log earnings and q is the order of serial correlation in the transitory component.

Identifying the variance of the transitory shock to income is more difficult. Many studies of the income process use survey data in which measurement errors are not negligible and classical measurement error will be subsumed by the transitory component. Meghir and Pistaferri (2004) point out that earlier studies have ignored the inability to disentangle the variance of the transitory shock, the variance of the measurement error and the parameters of the transitory process. The distinction has economic implications, however, since measurement error is pure noise while transitory shocks reflect uncertainty that may give rise to economic responses. The authors suggest two ways of handling this issue, by obtaining bounds for the unidentified variances or by using an external estimate of the measurement error to recover the variance of the transitory shock.

A controversy in the literature regards whether there are permanent individual-specific components in the growth rates of earnings, as proposed by Lillard and Weiss (1979) and Hause (1980). In contrast to the random walk specification of the permanent component, this would mean that earnings growth of an individual is correlated over time. Mistakenly leaving out the random growth rate term would overestimate the variance of the permanent shock. The random growth model implies that the auto covariances of earnings changes contain fixed positive components of the cross-sectional variance in individual-specific growth rates. Higher-order auto covariances should therefore be positive and constant, capturing the variance of the random growth term.

Based on such a test, MaCurdy (1982), Abowd and Card (1989) and Meghir and Pistaferri (2004) reject the existence of an individual-specific random growth term. Guvenen (2009) points out, however, that the power of the test might be low since most panel data sets used in the literature follow individuals for a limited period of time or is subject to substantial attrition, which implies that higher-order auto covariances are estimated on a small sample of individuals. He further shows that the random growth model is consistent with the observed

pattern of increased earnings variance over the life cycle if one allows for learning about individual-specific growth rates. In principle, an individual-specific growth term in addition to a permanent component and a transitory component is identifiable.

The literature reviewed so far has focused on investigating the nature of the income process, and takes the estimated unexplained variability in earnings to represent earnings risk. In these models, labor supply is taken as exogenous. A new branch of the literature tries to distinguish the underlying sources of risk and recognizes that part of the variability in earnings might be induced by endogenous responses to shocks, such as employment changes or job mobility. Disentangling exogenous shocks to income from endogenous responses to shocks is important for understanding the sources of income uncertainty, in particular since some types of shocks may be insurable. Abowd and Card (1989) take a step in this direction by simultaneously studying changes in earnings and changes in hours. They find that income changes cannot primarily be attributed to changes in productivity, but rather to fluctuations in hours worked.

More recently, Low, Meghir and Pistaferri (2010) use a structural approach in which they allow for different sources of uncertainty and for endogenous responses to shocks. They broadly divide risk into productivity risk and employment risk. Productivity risk is captured by permanent shocks to wages while employment risk comes from exogenous job destruction, the arrival rate of job offers when employed and unemployed and variation in the quality of the firm-worker match. Individuals may respond to these shocks by changing employer or by moving into unemployment, which is captured by selection equations into work and between jobs that depend on the wage offers being made. In this framework, the authors study the welfare effects of risk and find that individuals value a decrease in productivity risk higher than a decrease in employment risk. They also find that making job mobility and employment choices endogenous reduces the variance of the permanent shock compared to earlier studies. A related paper is Altonji, Smith and Vidangos (2009), who specify a model of employment, hours, wages and earnings in order to distinguish different sources of risk. Selection into employment and between jobs is modeled in a similar way as in Low, Meghir and Pistaferri (2010).

The recent availability of matched employer-employee data allows for an extended analysis of the role of the firm for the dynamics of earnings. An additional source of risk is shocks to firm-level productivity. Using a matched employer-employee data set from Italy, Guiso, Pistaferri and Schivardi (2005) estimate the amount of income insurance provided by firms. They find that firms insure their workers fully against temporary shocks to firm-level productivity but only partially against permanent shocks. The authors focus on a panel of workers continuously employed at a given firm and do not account

for employment changes or job mobility. However, workers may respond to a wage shock by leaving the employment for another firm or for unemployment. Such behavior would truncate the distribution of the shock and could lead to an overestimation of the amount of insurance provided by firms. This paper proposes a strategy for assessing the impact of firm-level productivity shocks on wages, taking endogenous selection into employment and between jobs into account and allowing for a rich stochastic structure of the income process.

3. The Model

This section presents the model of the income process that we propose for estimating the impact of firm-level productivity shocks on wages. We begin by presenting the structure of wages. Thereafter, we discuss how to take endogenous responses to shocks into account by modeling transitions between jobs and to employment.

3.1. Wages

We assume that wages for a particular individual i of age a who is working in a firm j at time t are determined by:

$$\ln w_{i,a,t} = d_t^g + x_{i,a,t}'\gamma^g + P_{i,a,t}^g + \varepsilon_{i,a,t}^g + \upsilon_{i,j(a_0),a,t}^g, \tag{6.1}$$

where $w_{i,a,t}$ is the real hourly wage, d_t is an aggregate time effect and $x_{i,a,t}$ are observable characteristics explaining wages, most importantly a quartic function in age. The stochastic structure of wages includes a permanent component $P_{i,a,t}$, a transitory component $\varepsilon_{i,a,t}$ and a time-varying match-specific component $\upsilon_{i,j(a_0),a,t}$, where $j(a_0)$ indexes the firm that the worker joined at age $a_0 \leq a$. The superscript g indicates that the individual belongs to a specific group defined by education and gender, and implies that the stochastic terms may be drawn from distributions that depend on the group. From now on, the superscript g is omitted to avoid cumbersome notation.

The permanent component of wages is assumed to follow a random walk process:

$$P_{i,a,t} = P_{i,a-1,t-1} + \zeta_{i,a,t} = P_i^{init} + \sum_{s=1}^{a} \zeta_{i,s,t-a+s}, \tag{6.2}$$

where the individual-specific initial permanent component, P_i^{init}, is normally distributed with mean zero and variance σ_P^2 and the shock to the permanent process, $\zeta_{i,a,t}$, is normally distributed and independent over time with mean zero and variance σ_ζ^2.

The transitory individual-specific component $\varepsilon_{i,a,t}$ is assumed to be i.i.d. for now with mean zero and variance σ_ε^2. It subsumes any measurement error.

As noted by Low, Meghir and Pistaferri (2010), the transitory shock when working with annual earnings data is important because of unemployment spells, but this source of shocks is modeled explicitly through the job mobility and employment equations. We can allow for a richer stochastic structure of the transitory component if the analysis supports such a specification.[5]

The firm enters the model through the firm-worker match-specific component, which is specified as follows:

$$\upsilon_{i,j(a_0),a,t} = \begin{cases} \rho\,\upsilon_{i,j(a_0),a-1,t-1} + \xi_{i,j(a_0),a,t} & \text{if } J_{i,a,t} = 0 \\ \upsilon^{init}_{i,j(a),a,t} & \text{if } J_{i,a,t} = 1 \end{cases}, \tag{6.3}$$

where $J_{i,a,t}$ is an indicator variable that takes the value one if the individual changes employer in period t, else zero. For individuals who remain at their firm, i.e., for $J_{i,a,t} = 0$, the match-specific effect follows an AR(1) process. The quality of the match is thus allowed to develop over the firm-worker relationship and the parameter ρ describes the persistence of the firm-worker match. This is an extension compared to Low, Meghir and Pistaferri (2010) and Altonji, Smith and Vidangos (2009), who modeled the match effect as fixed within jobs.

Each time a worker joins a new firm, i.e., for $J_{i,a,t} = 1$, they get a new draw from the distribution of initial matches υ^{init}. This distribution will be specific to the characteristics of the firm $j(a_0)$ from which the person's offer at age a comes. More specifically, we define mutually exclusive firm bins $s = \{1,\ldots,S\}$ based on sector and firm size and assume an identical stationary distribution $\upsilon^{init}_s \sim N\left(\mu_s, \sigma^2_{\upsilon,s}\right)$ for all firms within the same bin. Even though we assume these underlying distributions to be uncorrelated across firm bins, the match-specific effects of successive accepted jobs will be correlated because of the job acceptance strategy of the individual. This will be controlled for using the job change probability.

The model at this point includes firm information just to identify job mobility. We now consider the way that firm-level productivity shocks can affect individual wages. The aim is to provide a link between firm performance and wages based on the match-specific effect. Ultimately, this should be achieved based on a structural model, but such a model, where there are potential complementarities between the firm and the worker, is not available at this point and is likely to be highly complicated, in particular when we wish to account for shocks. The key difficulty is that we need to model how wage negotiations

[5] We can further generalize the model to include a "random growth" term g_i, i.e.,

$$\ln w_{i,a,t} = d^g_t + x'_{i,a,t}\gamma^g + P^g_{i,a,t} + \varepsilon^g_{i,a,t} + \upsilon^g_{i,j(a_0),a,t} + g_i a.$$

For the time being this is omitted, although it is in principle identifiable.

take place as workers arrive and depart from the firm and as the firm and the worker receive productivity shocks.

The match-specific component in our model is subject to shocks, $\xi_{i,j(a_0),a,t}$, which allow wages to fluctuate within the firm. In a competitive framework these shocks will be purely idiosyncratic and will not be related to firm-level shocks. One of the key issues we are interested in is whether the shocks $\xi_{i,j(a_0),a,t}$ are related to productivity shocks at the firm level. In other words, we can consider as an alternative to purely idiosyncratic shocks the following process:

$$\xi_{i,j(a_0),a,t} = \kappa q_{j(a_0),t} + \psi_{i,j(a_0),a,t}, \tag{6.4}$$

where $q_{j(a_0),t}$ are firm-level productivity shocks and $\psi_{i,j(a_0),a,t}$ is i.i.d. with mean zero and variance σ^2_ψ. Note that if the firm-level productivity shock admits a permanent-transitory decomposition, such that $q_{j(a_0),t} = q^P_{j(a_0),t} + q^T_{j(a_0,t}$, we can rewrite (6.4) as:

$$\xi_{i,j(a_0),a,t} = \kappa^P q^P_{j(a_0),t} + \kappa^T q^T_{j(a_0),t} + \psi_{i,j(a_0),a,t},$$

and follow the identification strategy for extracting the permanent and transitory shocks outlined in Guiso, Pistaferri and Schivardi (2005).

The model as specified allows for a large variety of patterns for the shocks and is consistent with most findings in the literature. The existence of a match-specific effect has been motivated theoretically within the search and matching framework and empirically by papers such as Topel and Ward (1992) and Abowd, Kramarz and Margolis (1999). Most studies on earnings dynamics, however, have not explicitly modeled the firm side. Low, Meghir and Pistaferri (2010) include a match-specific component in the wage process, but in contrast to this paper, the match is not allowed to develop within the firm-worker relationship and is not subject to shocks that could be related to firm-level productivity. These additions to the match-specific component are one of the contributions of our work compared to earlier studies.

3.2. Selection into Employment and Job Mobility

A key issue is controlling for selection into employment and for job mobility. Wages are only observed for those working and employment and mobility decisions may well be endogenous to wages. Ignoring this selection issue may truncate the distribution of shocks and bias the variance estimates of the income process. To handle this problem, we specify two selection equations similar to Low, Meghir and Pistaferri (2010) and Altonji, Smith and Vidangos (2009). They are "reduced form" in nature since we do not fully specify the structure of the problem in terms of utility and ignore dynamics.

Define $E_{i,a,t} = 1$ if the individual works in period t and zero otherwise. Then we assume that employment is governed by:

$$E_{i,a,t} = \mathbf{1}\{z'_{i,a,t}\delta + \phi_1(P_{i,a,t} + \varepsilon_{i,a,t,} + \upsilon_{i,j(a_0),a,t}) + u^E_{i,a,t} > 0\}, \tag{6.5}$$

where $z_{i,a,t}$ are exogenous characteristics which include the variables in $x_{i,a,t}$ from the wage equation as well as exclusion restrictions that will be excluded from the wage equation. Employment also depends on all the stochastic elements of the wage equation and a random error term $u^E_{i,a,t}$.

Among those employed in two successive periods, further define $J_{i,a,t} = 1$ if the individual changes employer in period t and zero otherwise. Then we assume that job mobility is governed by:

$$\begin{aligned} J_{i,a,t} =& \mathbf{1}\{z'_{i,a,t}\theta + b_1(\upsilon_{i,j(a_0),a,t} - \upsilon^{init}_{i,j(a),a,t}) \\ &+ b_2\mathbf{1}\{s(a) \neq s(a_0)\} + u^J_{i,a,t} > 0\}, \end{aligned} \tag{6.6}$$

where $z_{i,a,t}$ again include the variables in $x_{i,a,t}$ and exclusion restrictions. Potential exclusion restrictions are marital status, number of children and financial income. The idea is that these variables affect the probability of changing employer or exiting employment, without affecting the wage rate.

Importantly, while we allowed the employment equation to depend on all the stochastic elements of the wage equation, job mobility only depends on the difference between the match-specific effect in the current firm and a potential new employer. Since the non-firm-specific stochastic components can be interpreted as elements that are portable across firms they should not influence the decision to move. We further allow switching costs that depend on the firm bin s to which the alternative firm $j(a)$ belongs. In particular, we assume a switching cost b_2 for moving to a firm in a different bin. Since firm bins are defined by sector and firm size, the switching cost reflects, for example, the loss of industry-specific human capital or the change in the degree of vocational specialization across firm sizes.

We assume that the error terms u^E and u^J are normally distributed with unit variances (a normalization) and that they are uncorrelated to all other shocks present in the model. Hence, we assume that the variance matrix of $(u^E, u^J, \zeta, \varepsilon, \psi, \upsilon^{init}, P^{init})$ is diagonal.

In a first step we need to determine the probability that an unemployed worker obtains any offer at all. We can fix this as λ^U, i.e., the arrival rate of offers when unemployed. For those currently employed, we can fix the probability of receiving an outside offer as λ^E, i.e., the arrival rate of alternative offers when employed. Both rates can be made dependent on the education-gender group to which the worker belongs.

Given a wage offer we need to determine what the offer consists of and from where it is coming. According to our model, the offer is a match-specific effect $\upsilon^{init}_{i,j(a),a,t}$. All firms within a firm bin are drawing shocks to the match-specific effect from the same distribution. It still matters to which firm the individual will be allocated because of the realization of the firm-level shock. We therefore need to allocate the worker to an actual firm. We do this by specifying the probability of any one firm as

$$p_j\left(m_{s(j)}\right) = \frac{1}{K_s} \cdot \frac{\omega_{s(j)}\left(m_{s(j)}\right)}{\sum\limits_{\text{all bins } s'} \omega_{s'}\left(m_{s'}\right)} = \frac{1}{K_s} \cdot \frac{\omega' m_{s(j)}}{\sum\limits_{\text{all bins } s'} \omega' m_{s'}} \tag{6.7}$$

where K_s are the number of firms in firm bin s. The first term implies that firms within a given bin are drawn randomly conditional on receiving an offer from bin s. The second term is the probability of an offer from firm bin s as a function of individual and firm characteristics m. Now we need to specify the weights $\omega_{s(j)}\left(m_{s(j)}\right)$, which will be parameters to estimate. They will be a function of firm size and sector, defining the firm bins. Furthermore, the weights may depend on individual characteristics to reflect that different types of workers are observing different firms.

Finally, in addition to endogenous quits, we assume that jobs are separated at an exogenous rate φ. Workers whose jobs are separated become unemployed and can only receive a new offer in the next period. With this, the model is complete.

4. Estimation

Although the model of the income process outlined in the previous section is not exactly structural, for estimation purposes we refer to it as the "structural model". The estimation of the model is quite complex because of the dynamics and because of transitions between work and different jobs. The proposed estimation strategy first estimates the dynamics of wages, ignoring the firm as well as the dynamics of the productivity shocks ignoring wages. In the next step the two are brought together.

The proposed estimation strategy is based on the method of indirect inference, which is a simulation-based method for estimating the parameters of economic models. The model was introduced by Smith (1990, 1993) and extended by Gourieroux, Monfort and Renault (1993) and Gallant and Tauchen (2009). Indirect inference relies on matching moments from an approximate model, the "auxiliary model", whose parameters are estimated by minimizing the distance between the parameters of the auxiliary model estimated from the

observed data and the parameters of the auxiliary model estimated from the simulated data.[6]

For now, we consider a simple structure for the firm-level productivity shock q, obtained from changes in firm performance over time. For example, it can be constructed as the change in the logarithm of revenue per worker, normalized around zero in each firm bin. For empirical purposes, we have in mind six different groups of individuals g, defined by gender and the education groups less than high school, high school and some college. Likewise, we consider 16 firm bins defined by firm size and industry. Firm size is determined by the initial number of employees and consists of the categories 5–20, 20–50, 50–100 and above 100 employees, whereas industry consists of the categories construction, manufacturing, retail trade and services.

In this section, we first describe how to construct the simulated data set from the structural model. Thereafter, we discuss the auxiliary model and identification.

4.1. Simulation of the Model

We first need to create a simulated data set of the structural model that was described by equations (6.1) through (6.7). We begin by making a guess about the vector of parameters in the model:

$$\beta^0 = \left\{\vec{\gamma}, \vec{\delta}, \vec{\theta}, \rho, \kappa, \phi_1, b_1, b_2, \sigma_P^2, \sigma_\zeta^2, \sigma_\varepsilon^2, \sigma_\psi^2, \vec{\mu}_s, \vec{\sigma}_{\upsilon,s}^2, \lambda^U, \lambda^E, \vec{\omega}, \varphi\right\}.$$

We would already have allocated individuals into groups g defined by education and gender and firms into firm bins s defined by firm size and industry. Consider a cohort of individuals entering the labor market at some time t and age 0. For each individual, we start by drawing $P_{i,0,t}^{init}$ from $N\left(0, \sigma_P^2\right)$ and keep this fixed across the entire life cycle. We then draw whether there is an employment offer or not with probability λ^U. If an offer is not available, we set $E_{i,0,t} = 0$ and the actual wage to missing. If an employment offer is drawn, we need to define from which firm the offer is coming. We hence draw a firm $j(0)$ with probability $p(m_j)$. The drawing of the firm also implicitly assigns the firm productivity shock $q_{j(0),t}$. Next, we get draws of the wage shocks $\left(u_{i,0,t}^E, u_{i,0,t}^J, \zeta_{i,0,t}, \varepsilon_{i,0,t}, \psi_{i,j(0),0,t}, \upsilon_{i,j(0),0,t}^{init}\right)$ from the underlying distributions.

We then construct the offered wage and the employment indicator:

$$\begin{aligned}
\ln w_{i,0,t} &= d_t + x_{i,0,t}'\gamma + P_{i,0,t} + \varepsilon_{i,0,t} + \upsilon_{i,j(0),0,t}^{init} \\
E_{i,0,t} &= \mathbf{1}\{z_{i,0,t}'\delta + \phi_1(P_{i,0,t} + \varepsilon_{i,0,t} + \upsilon_{i,j(0),0,t}^{init}) + u_{i,0,t}^E > 0\}.
\end{aligned}$$

[6]See, e.g., Guvenen and Smith (2010) and Altonji, Smith and Vidangos (2009) for recent applications.

If $E_{i,0,t} = 0$, we set the actual wage received to missing, otherwise we set it equal to the offered wage.

From age 1 onward, the simulation algorithm works as follows. If the individual was unemployed in the previous period, the algorithm is repeated exactly as described for age 0. If the individual was employed in the previous period, we first draw an exogenous separation shock with probability φ. If there is no separation, we get draws of the shocks for that period, $\left(u^E_{i,1,t+1}, u^J_{i,1,t+1}, \zeta_{i,1,t+1}, \varepsilon_{i,1,t+1}, \psi_{i,j(0),1,t+1}, \upsilon^{init}_{i,j(1),1,t+1}\right)$. We then construct the firm-worker match-specific effect and the permanent component as:

$$\begin{aligned}\upsilon_{i,j(0),1,t+1} &= \rho\,\upsilon^{init}_{i,j(0),0,t} + \kappa q_{j(0),t+1} + \psi_{i,j(0),1,t+1}\\ P_{i,1,t+1} &= P_{i,0,t} + \zeta_{i,1,t+1}.\end{aligned}$$

We determine whether an alternative offer is made with probability λ^E. If an alternative offer is not available, we define the offered wage and the employment indicator as:

$$\begin{aligned}\ln w_{i,1,t+1} =& d_{t+1} + x'_{i,1,t+1}\gamma + P_{i,1,t+1} + \varepsilon_{i,1,t+1} + \upsilon_{i,j(0),1,t+1}\\ E_{i,1,t+1} =& \mathbf{1}\{z'_{i,1,t+1}\delta + \phi_1(P_{i,1,t+1} + \varepsilon_{i,1,t+1} + \upsilon_{i,j(0),1,t+1}) + u^E_{i,1,t+1} > 0\}.\end{aligned}$$

If $E_{i,1,t+1} = 0$, we set the actual wage received to missing, otherwise we set it equal to the offered wage.

If instead an alternative offer is available, we draw an alternative firm, $j(1)$, with probability $p(m_j)$. We then construct the job mobility indicator:

$$\begin{aligned}J_{i,1,t+1} =& \mathbf{1}\{z'_{i,1,t+1}\theta + b_1(\upsilon_{i,j(0),1,t+1} - \upsilon^{init}_{i,j(1),1,t+1})\\ &+ b_2\mathbf{1}\{s(1) \neq s(0)\} + u^J_{i,1,t+1} > 0\}.\end{aligned}$$

The offered wage and the employment indicator are then given by:

$$\begin{aligned}\ln w_{i,1,t+1} =& d_{t+1} + x'_{i,1,t+1}\gamma + P_{i,1,t+1} + \varepsilon_{i,1,t+1}\\ &+ \upsilon^{init}_{i,j(1),1,t+1}J_{i,1,t+1} + \upsilon_{i,j(0),1,t+1}(1 - J_{i,1,t+1})\\ E_{i,1,t+1} =& \mathbf{1}\{z'_{i,1,t+1}\delta + \phi_1(P_{i,1,t+1} + \varepsilon_{i,1,t+1} + \upsilon^{init}_{i,j(1),1,t+1}J_{i,1,t+1}\\ &+ \upsilon_{i,j(0),1,t+1}(1 - J_{i,1,t+1})) + u^E_{i,1,t+1} > 0\}.\end{aligned}$$

If $E_{i,1,t+1} = 0$, we set the actual wage received to missing, otherwise we set it equal to the offered wage. Hence, for a given draw of the structural parameters b_0, we have a simulated data set where for each individual and age we have a {0,1} employment indicator, an accepted wage which is set to missing if $E_{i,a,t} = 0$, and a {0,1} job mobility indicator, which is set to missing if the worker is not employed in the current and the previous period.

4.2. Auxiliary Model and Identification

To estimate the dynamics of wages we construct moments from the observed data, which we match the equivalent moments derived from simulated data. To capture the effects of mobility, we will construct moments depending on whether the individual changed jobs, moved to work from unemployment, or remained in the same job. Such moments will also be conditioned on the exogenous variables driving mobility and participation. To capture the impact of the firm-level productivity shocks on wages we will subsequently add cross moments between the firm and the worker. The auxiliary model consists of a set of dynamic equations and parameters that are closely related to the structural model we have in mind.

The first auxiliary equation is an employment equation:

$$E_{i,a,t} = \mathbf{1}\{z'_{i,a,t}m + e^E_{i,a,t}\}, \tag{6.8}$$

where z is understood to include individual characteristics $x_{i,a,t}$, as well as exclusion restrictions such as number and ages of children, marital status and financial income. Clearly, there is a close correspondence between the auxiliary parameter m in equation (6.8) and the structural parameter δ in equation (6.5). We define the residual from this regression as $\tilde{e}^E = E_{i,a,t} - z'_{i,a,t}\hat{d}$, which is understood to include the stochastic terms of the wage equation and a random error term.

The next auxiliary equation is a job mobility equation:

$$J_{i,a,t} = \mathbf{1}\{z'_{i,a,t}n + e^J_{i,a,t}\}, \tag{6.9}$$

where the auxiliary parameter n in equation (6.9) is linked to the structural parameter θ in equation (6.6). We define the residual from this regression as $\tilde{e}^J = J_{i,a,t} - z'_{i,a,t}\hat{n}$, which will include the difference between the match-specific effect in the current firm and at a potential new employer and a switching cost if the individual moves to a firm in another firm bin.

We also estimate a wage equation:

$$\ln w_{i,a,t} = d_t + x'_{i,a,t}g + e^W_{i,a,t}, \tag{6.10}$$

where there is a close correspondence between the auxiliary parameter g in equation (6.10) and the structural parameter γ in equation (6.1). We next define the residual of this regression as $\tilde{e}^W_{i,a,t} = \ln w_{ia} - x'_{i,a,t}\hat{g} - d_t$, which includes all stochastic components of the wage equation.

Next, we consider as an auxiliary equation the regression of the wage residual against the firm-level productivity shock:

$$\tilde{e}^W_{i,a,t} = hq_{j,t} + \eta_{i,a,t}, \tag{6.11}$$

which allows us to link the auxiliary parameter h to the structural parameter κ given by equation (6.4). Furthermore, $var(\tilde{e}^W_{i,a,t} - \hat{h}q_{j,t})$ is related to the structural parameter σ^2_ψ.

We also consider the variance of the unobserved component of wages at the time of entry in the labor market, $var(\tilde{e}_{i,0,t})$, as an auxiliary moment that can be used to identify the structural parameter σ^2_P, i.e., the variance of the initial permanent component in equation (6.2).

The variances and covariances of wage growth between and within jobs:

$$\begin{aligned}
&var\left(\Delta\tilde{e}^W_{i,a,t}|J_{i,a,t}=0\right)\\
&var\left(\Delta\tilde{e}^W_{i,a,t}|J_{i,a,t}=1\right)\\
&cov\left(\Delta\tilde{e}^W_{i,a-l,t-l},\Delta\tilde{e}^W_{i,a,t}|J_{i,a,t}=0\right)\\
&cov\left(\Delta\tilde{e}^W_{i,a-l,t-l},\Delta\tilde{e}^W_{i,a,t}|J_{i,a,t}=1\right)
\end{aligned}$$

for lags $l = 1,...,4$ as well as the variance of the job mobility residual and the covariances of the employment and job mobility residuals and the wage residual:

$$\begin{aligned}
&var\left(\tilde{e}^J_{i,a,t}\right)\\
&cov\left(\tilde{e}^E_{i,a,t},\tilde{e}^W_{i,a,t}|J_{i,a,t}=0\right)\\
&cov\left(\tilde{e}^E_{i,a,t},\tilde{e}^W_{i,a,t}|J_{i,a,t}=1\right)\\
&cov\left(\tilde{e}^J_{i,a,t},\tilde{e}^W_{i,a,t}|J_{i,a,t}=0\right)\\
&cov\left(\tilde{e}^J_{i,a,t},\tilde{e}^W_{i,a,t}|J_{i,a,t}=1\right)
\end{aligned}$$

should help identifying respectively $\sigma^2_\zeta, \sigma^2_\varepsilon, \sigma^2_{\upsilon,s}$, μ_s, ρ, ϕ_1 and b_1. Note that all these variance-covariance moments are computed only for the workers. The same is done in the simulated data.

In addition, following Moulton (1986), we use the spatial correlation of residual wage growth within a firm as an additional auxiliary moment to help identify κ and σ^2_ψ.

Average transitions from employment into work and across firms are used to identify λ^U, λ^E and φ, and average transitions between and within firm bins are used to identify the switching costs. We also estimate matching probabilities to firm type by worker characteristics to identify the structural parameters ω. With this, the identification of the model is complete.

5. Conclusion

A large literature has studied the nature of the income process, but the firm side has typically been neglected in these studies. The firm may play a central role,

however, for the amount and characteristics of income risk individuals face. Most studies have also not been explicit about disentangling income shocks from responses to shocks, such as job mobility or changes in employment status. The recent availability of matched employer-employee data allows for extensions of the literature in these directions. These data make it possible to introduce the firm in the dynamic income process and to disentangle different sources of income uncertainty.

This paper is the opening investigation into an agenda that will lead to richer structural models taking the firm side more seriously and allowing for a rich dynamic stochastic structure of the income process. The paper proposes a framework for introducing the firm in empirical models of the dynamic income process. In particular, the proposed model allows for analyzing the extent to which firm-level productivity shocks are transmitted to wages. To disentangle the shocks from endogenous responses to shocks, selection into employment and between jobs is explicitly modeled.

The firm is introduced through a match-specific effect which is allowed to develop throughout the worker-firm relationship. To study the importance of firm-level shocks, we analyze whether shocks to the match-specific effect can be related to productivity shocks at the firm level. Based on the method of indirect inference, we propose a strategy for estimation and identification of the model. We specify an auxiliary model that consists of a set of dynamic equations and parameters that are closely related to the structural model we have in mind. Based on the empirical strategy outlined in this paper, next on the research agenda is to bring the model to data.

References

Abowd, J., and D. Card (1989), "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, vol. 57, pages 411–445.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999), "High Wage Workers and High Wage Firms," *Econometrica*, vol. 67, pages 251–333.

Altonji, J. G., A. A. Smith, and I. Vidangos (2009), "Modeling Earnings Dynamics," NBER Working Paper No. 14743, National Bureau of Economic Research.

Burdett, K., and D. T. Mortensen (1998), "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, vol. 39(2), pages 257–73.

Cahuc, P., F. Postel-Vinay, and J-M. Robin (2006), "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, vol. 74(2), pages 323–364.

Friedman, M. (1957), *A Theory of the Consumption Function*, Princeton: Princeton University Press.

Gallant, A. R., and G. Tauchen (2009), "Simulated Score Methods and Indirect Inference for Continuous Time Models," in L. P. Hansen and Y. Ait-Sahalia (eds.), *Handbook of Financial Econometrics Volume 1*, Amsterdam: North Holland, pages 427–477.

Gourieroux, C., A. Monfort, and E. Renault (1993), "Indirect Inference," *Journal of Applied Econometrics*, vol. 8, pages S85–S118.

Guiso, L., L. Pistaferri, and F. Schivardi (2005), "Insurance within the Firm," *Journal of Political Economy*, vol. 113, pages 1054–1087.

Guvenen, F. (2009), "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, vol. 12, pages 58–79.

Guvenen, F., and A. Smith (2010), "Inferring Labor Income Risk from Economic Choices: An Indirect Inference Approach," NBER Working Paper No. 16327, National Bureau of Economic Research.

Hause, J. C. (1980), "The Fine Structure of Earnings and the On-the-Job Training Hypothesis," *Econometrica*, vol. 48(4), pages 1013–1029.

Lamadon, T., J. Lise, C. Meghir, and J. M. Robin (2012), "Worker and Firm Effect in an Economy with Search Friction," mimeo.

Lillard, L., and Y. Weiss (1979), "Components of Variation in Panel Earnings Data: American Scientists 1960–1970," *Econometrica*, vol. 47(2), pages 437–454.

Lillard, E. L., and R. Willis (1978), "Dynamic Aspects of Earnings Mobility," *Econometrica*, vol. 46, pages 985–1012.

Lise, J., C. Meghir, and J-M. Robin (2012), "Matching, Sorting and Wages," mimeo.

Low, H., C. Meghir, and L. Pistaferri (2010), "Wage Risk and Employment Risk over the Life Cycle," *American Economic Review*, vol. 100(4), pages 1432–67.

MaCurdy, T. E. (1982), "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, vol. 18, pages 82–114.

Meghir, C. (2004), "A Retrospective on Friedman's Theory of Permanent Income," *The Economic Journal*, vol. 114(June), pages F293–F306.

Meghir, C., and L. Pistaferri (2004), "Income Variance Dynamics and Heterogeneity," *Econometrica*, vol. 72(1), pages 1–32.

Meghir, C., and L. Pistaferri (2011), "Earnings, Consumption and Life Cycle Choices," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics Volume 4*, Elsevier Science B. V., pages 773–854.

Mortensen, D. T., and C. A. Pissarides (1994), "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, vol. 61(3), pages 397–415.

Moulton, B. R. (1986), "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, vol. 32, pages 385–397.

Postel-Vinay, F., and J-M. Robin (2002), "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, vol. 70(6), pages 2295–2350.

Smith, A. A., Jr. (1990), *Three Essays on the Solution and Estimation of Dynamic Macroeconomic Models*, Ph.D. thesis, Duke University.

Smith, A. A., Jr. (1993), "Estimating Nonlinear Time-Series Models using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, vol. 8, pages S63–S84.

Topel, R., and M. Ward (1992), "Job Mobility and the Careers of Young Men," *Quarterly Journal of Economics*, vol. 107(2), pages 439–479.

Publication series published by IFAU – latest issues

Rapporter/Reports

2012:20 Ulander-Wänman Carin ”Flexibilitet – en dominerande diskurs i de anställningsvillkor som rör sysselsättningstrygghet för arbetstagare i kommun och landsting”

2012:21 Wikström Christina and Magnus Wikström ”Urval till högre utbildning – Påverkas betygens prediktionsvärde av ålder?

2012:22 Karimi Arizo, Erica Lindahl and Peter Skogman Thoursie ”Effekter av föräldrapenning på arbetsutbud”

2012:23 Sibbmark Kristina ”Arbetsmarknadspolitisk översikt 2011”

2012:24 Liljeberg Linus, Sara Martinson and Jonas Thelander ”Vad innebär det att bli coachad? En utvärdering av jobbcoachningen vid Arbetsförmedlingen”

2012:25 van den Berg Gerard J., Petter Lundborg and Johan Vikström ”De långsiktiga ekonomiska och sociala konsekvenserna av att förlora ett barn”

2012:26 Hägglund Pathric, Per Johansson and Lisa Laun ”Rehabiliteringsgarantin”

2012:27 Engdahl Mattias and Olof Åslund ”Ekonomiska drivkrafter och studieresultat – effekter av Sfi-bonus”

2012:28 Lindgren Karl-Oskar ”Arbetsplatsstorlek och sjukfrånvaro”

Working papers

2012:20 Angelov Nikolay and Arizo Karimi ”Mothers’ income recovery after childbearing”

2012:21 Wikström Christina and Magnus Wikström ”University entrance selection and age at admission”

2012:22 Karimi Arizo, Erica Lindahl and Peter Skogman Thoursie ”Labour supply responses to paid parental leave”

2012:23 van den Berg Gerard J., Petter Lundborg and Johan Vikström ”The economics of grief”

2012:24 Åslund Olof and Mattias Engdahl ”The value of earning for learning: performance bonuses in immigrant language training”

2012:25 Collet Francois and Peter Hedström ”Endogenous tie formation mechanisms in a directed network generated by employee mobility”

2012:26 Lindgren Karl-Oskar ”Workplace size and sickness absence transitions”

Dissertation series

2012:1 Laun Lisa “Studies on social insurance, income taxation and labor supply”