IFAU
Institute for Evaluation of Labour
Market and Education Policy

Essays on pensions, retirement and tax evasion

Johannes Hagen

DISSERTATION SERIES 2016:1

Presented at the Department of Economics, Uppsala 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, Uppsala University, March 11, 2016. Essay 3 has recently been published by IFAU as Working paper 2016:11.

ISSN 1651-4149

Dissertation presented at Uppsala University to be publicly examined in Hörsal 2, Ekonomikum, Kyrkogårdsgatan 10, Uppsala, Friday, 11 March 2016 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Monika Bütler (University of St. Gallen, Department of Economics).

Abstract
Hagen, J. 2016. Essays on Pensions, Retirement and Tax Evasion. *Economic studies* 158. 195 pp. Uppsala: Department of Economics, Uppsala University. ISBN 978-91-85519-65-1.

Essay I: This essay provides an overview of the history of the Swedish pension system. Starting with the implementation of the public pension system in 1913, it outlines the key components of each major pension reform up until today along with a discussion of the main trade-offs and concerns that policy makers have faced. It also describes the historical background of the four largest occupational pension plans in Sweden and the mutual influence between these plans and the public pension system.

Essay II: Despite the fact that the increasing involvement of the private sector in pension provision has brought more flexibility to the pay-out phase of retirement, little is known about the characteristics of those who choose to annuitize their pension wealth and those who do not. I combine unique micro-data from a large Swedish occupational pension plan with rich national administrative data to study the choice between life annuities and fixed-term payouts with a minimum payout length of 5 years for 183,000 retiring white-collar workers. I find that low accumulation of assets is strongly associated with the choice of the 5-year payout. Consistent with individuals selecting payout length based on private information about their mortality prospects, individuals who choose the 5-year payout are in worse health, exhibit higher ex-post mortality rates and have shorter-lived parents than annuitants. Individuals also seem to respond to large, tax-induced changes in annuity prices.

Essay III: This essay estimates the causal effect of postponing retirement on a wide range of health outcomes using Swedish administrative data on cause-specific mortality, hospitalizations and drug prescriptions. Exogenous variation in retirement timing comes from a reform which raised the age at which broad categories of Swedish local government workers were entitled to retire with full pension benefits from 63 to 65. The reform caused a remarkable shift in the retirement distribution of the affected workers, increasing the actual retirement age by more than 4.5 months. Instrumental variable estimation results show no effect of postponing retirement on the overall consumption of health care, nor on the risk of dying early. There is evidence, however, of a reduction in diabetes-related hospitalizations and in the consumption of drugs that treat anxiety.

Essay IV (with Per Engström): The consumption based method to estimate underreporting among self-employed, introduced by Pissarides and Weber (1989), is one of the workhorses in the empirical literature on tax evasion/avoidance. We show that failure to account for transitory income fluctuations in current income may overestimate the degree of underreporting by around 40 percent. Previous studies typically use instrumental variable methods to address the issue. In contrast, our access to registry based longitudinal income measures allows a direct approach based on more permanent income measures. This also allows us to evaluate the performance of a list of instruments widely used in the previous literature. Our analysis shows that capital income is the most suitable instrument in our application, while education and housing related measures do not seem to satisfy the exclusion restrictions.

Keywords: Pensions, retirement, annuity, annuity puzzle, adverse selection, pension reform, instrumental variable, health, health care, mortality, tax evasion, engel curves, consumption, self-employment, permanent income

Johannes Hagen, , Department of Economics, Box 513, Uppsala University, SE-75120 Uppsala, Sweden.

ISSN 0283-7668
ISBN 978-91-85519-65-1
urn:nbn:se:uu:diva-274479 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-274479)

To Cajsa and Kerstin

Acknowledgments

Writing a thesis is like saving for retirement. You have a slight feeling that what you do is probably a good thing, but you don't have a clue whether the time and effort you put in is paying off. There is also a lot of uncertainty about the process of finishing and actually reaping those benefits. At this point, however, I truly feel that I'm approaching the retirement date of this thesis. I'm now going to make an attempt to sincerely thank the people around me who made this possible.

First of all, I would like to extend my deepest gratitude to my advisor Sören Blomquist. Sören has been central to my work from day one, meritoriously introducing me to the topic of pensions and retirement and to the research field of public economics, both here in Uppsala and in the wider world. He always struck the right balance between decisive guidance and giving me the leeway to try out new research ideas. Watching Sören, it has become increasingly clear to me that the Swedish practice of mandatory retirement indeed has its drawbacks.

Next, I would like to thank Håkan Selin, my co-advisor. Håkan was the natural starting point to look for potential research ideas and long-forgotten pension reforms. We really left no stone unturned in the search for nice, exogenous variation in retirement timing! Although no co-written paper has come out of this quest (yet), discussing with Håkan has had a huge impact on my work. I've particularly benefited from Håkan's impressive skills in econometrics and writing as well as his unerring sense of how to turn "dull" institutional details into interesting research questions.

I also wish to convey special thanks to my co-author and "shadow" advisor Per Engström. It is difficult to overestimate how much Per has meant to me during these years. Our work together stretches far beyond the paper included in this thesis, wandering deep into the bewildering woods of government agencies, randomized controlled trials and boat registers. Among all the things I learned from Per, always remaining patient and being generous with your time are the most important.

Several other persons have been fundamental to my time in academia. Thanks to Per Johansson who generously included me in his research program and made me think seriously about what saving for retirement really means. Thanks to James Poterba, who invited me to spend a very rewarding semester at MIT. Thanks to Henry Ohlsson who gave me the opportunity to work as a research assistant at the department and encouraged me to apply for the Ph.D. program. Also thanks to Mårten Palme, Karin Edmark and Paolo Sodini for insightful

comments at different stages of this thesis. I'm also indebted to Mikael Elinder who introduced me to the world of lecturing and spectacular conference runs.

I'm glad to have encountered many gifted people with an interest for pensions. Thanks to Pär Ola Grane, Lars Callert and Nikolaj Lunin at Alecta for data provision, instant feedback and many profitable discussions. Bo Könberg, Daniel Hallberg, Inger Johannisson, Yuwei de Gosson de Varennes and Hannes Malmberg also deserve recognition for bringing me new insights into this topic. The young generation of today, me included, are much indebted to Johannes Danielsson who brought "pensions and youth" onto the political arena.

My next thanks go to the fantastic Ph.D. cohort of 2011. In the company of these people, a helping hand, an intriguing discussion or a savory dinner have always been at close hand. Thanks to Jonas for always putting me in a better mood; Eskil for making our numerous trips to Stockholm worthwhile; Linuz for all the laughs and great conference company; Sebastian for showing that hard work pays off and for making our last year of writing a bit more tolerable; Jenny, Ylva and Mathias for the memories we've shared around the world. Last, thanks to Anna for being the best roommate I could imagine. I'm grateful that last year ended so well for both of us.

I also extend my sincere gratitude to all colleagues at the department for making my time here memorable. Oscar for all the lunch workouts and the tips about the do's and don'ts as a Ph.D. student and as a parent; Sebastian E, Jacob, Evelina, Kristin, Henrik, Fredrik, Lovisa, Oskar, Irina, Jon, Georg, Mattias, Arizo, Tobias, Jonas K, Eric, Susanne, Adrian and Spencer for great fellowship and cheerful lunches; PA, Eva, Katarina, Stina, Emma, Nina, Tomas, Javad, Åke and Ann-Sofie for being the backbone of the department.

I would also like to thank my friends and family. Thanks to Olof for being a trusted friend and the most efficient emailer I know; Ellen and Tobias for your everlasting kindness; Markus for numerous concerts and ping pong games. To all friends, thank you for keeping me grounded and putting up with all the nagging about pensions.

To my parents, Pernilla and Holger, thank you for your constant love and support. To my siblings, Joel, Lisa and Josef, there are so many things about you I treasure, thank you for everything! I also want to thank my in-laws for embracing me and Anders in particular for being a source of inspiration.

My biggest thanks go to my own little family. Cajsa, I feel immensely grateful for having met you. Spending the day with you is like living the best of dreams. To our daughter, Kerstin, thank you for blessing our lives. And thank you for being the single most important factor for making me finish this thesis, more or less, on time.

Uppsala, January 2016
Johannes Hagen

Contents

Introduction

This thesis consists of four self-contained essays. The first three essays relate to pensions and retirement in Sweden whereas the fourth essay deals with tax evasion among the self-employed. In this section, I introduce each of these two research fields and discuss how each essay contributes to the respective topic.

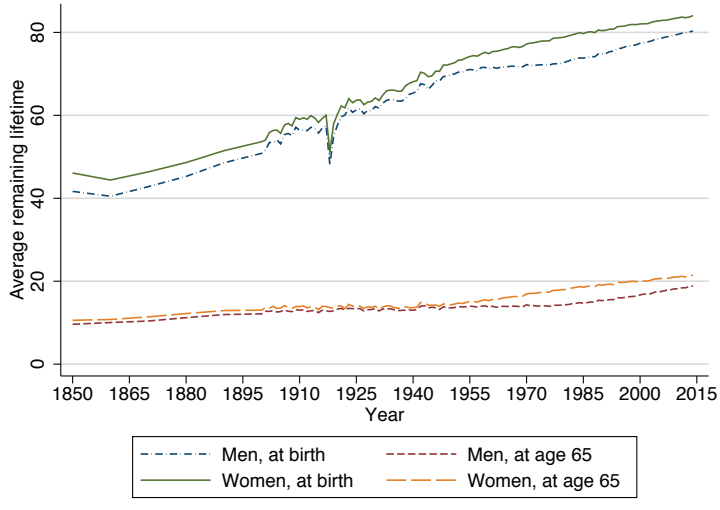
1 Pensions and retirement in Sweden

Figure 1 shows the evolution of the average remaining life expectancy at birth and at age 65 over the last 150 years for Swedish men and women. In 1913, the year in which the first public pension system was legislated in Sweden, the average life expectancy at birth was around 57 and 60 years for men and women, respectively. An average person who was fortunate to be alive at age 65 could expect to live another 13–14 years. Today, the average life expectancy is 80 years for men and 84 years for women, and those who live until the age of 65 can expect to live another 19–21 years, on average. In fact, a new-born today is more likely to reach the age of 80 than a 65-year-old was a century ago.

The expected length of life has important implications for the need and design of a pension system. Unless the actual retirement age increases at a similar rate, rising life expectancy necessarily translates into additional years spent in retirement that must be financed in some way. To get a picture of the historical development of the retirement age, Figure 2 plots gender-specific employment rates for three different age groups above the age of 60 for the time period 1961–2011. The left panel shows that the employment rate among men declined significantly in the first three decades. Although this trend has been reverted in the last 15–20 years or so, the current employment rates among elderly men are only at levels prevailing in the late 1970s and far from those of the early 1960s. The only group where the employment rate is higher today than 50 years ago are women aged 60–64. The share of individuals working in this group rose from around 30 percent in 1961 to almost 60 percent in 2011. Employment rates among women aged 65 and above, on the other hand, have been quite stable and rather low throughout this time period.

The increasing gap between average life expectancy and the actual retirement age provides an important backdrop to the retirement-related topics addressed in this thesis. Essay I provides a historical review of the development

Figure 1. Average remaining lifetime in years for men and women



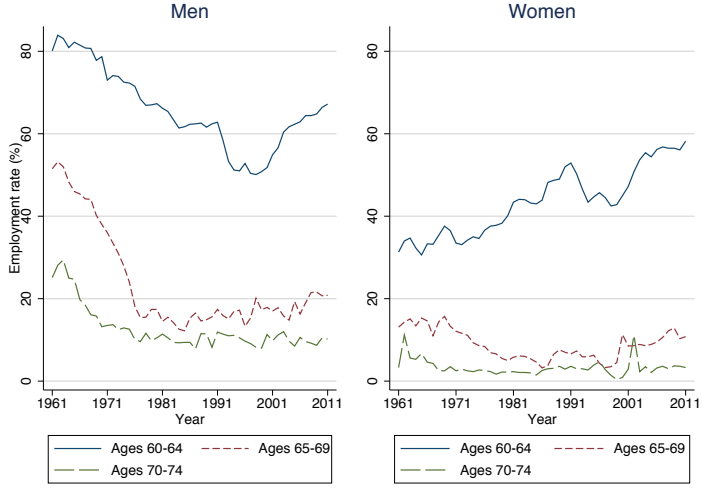
Source: SCB (2011, 2015)

of the Swedish pension system. It discusses the key components of each major pension reform up until today and how these have been shaped by the demographic and politico-economic context. Essay II deals with the decumulation of wealth during retirement or, more specifically, Swedish white-collar workers' preferences for the time period over which their occupational pension capital is paid out. From an individual viewpoint, how to use the accumulated pension assets becomes an increasingly important matter with the number of years spent in retirement. From the viewpoint of the government, the consequences of individuals failing to insure themselves against outliving their resources will be more severe in times of rising life expectancy. Essay III studies whether later retirement has an effect on health and, if so, which aspects of health. Rising life expectancy has been the main driver behind recent reform attempts to increase peoples' willingness to postpone retirement. Alongside the effects on labor supply, it is important to understand the health effects of such reforms to evaluate the potential effects on other parts of the welfare system.

1.1 A history of the Swedish pension system

The Swedish pension system, as we think of it today, has existed for about 100 years. During this period, the pension system has been subject to many changes, some more important than others. There exist detailed accounts of

Figure 2. Employment rates 1961–2011 for men and women aged 60–74



Source: Wadensjö (2011)

Swedish pension reforms in the second half of the 20th century[1], but less attention has been paid to the early development of the pension system. Essay I attempts to fill this gap. Starting with the implementation of the public pension system in 1913, it outlines the key components of each major pension reform up until today along with a discussion of the main trade-offs and concerns that policy makers have faced.

The implementation of the first public pension system in 1913 was foremost motivated by a need to provide poverty relief for the elderly. It was argued that a universal pension system was the best way to tackle the fiscal challenges associated with the growing ratio of old-to-young people and increasing life expectancy. Poverty relief, and soon also providing a minimum standard of living in retirement, has been at core of the rationale for the pension system ever since. However, the policy makers soon realized that the pace at which the new pension system enhanced the living conditions of the elderly was too slow. Many elderly chose not to participate in this pension plan as it would take many years for an individual to amass enough contributions to be able to claim a substantial pension.

To speed up the poverty reduction process and increase coverage, a pension component called the folkpension was introduced in 1935. The folkpension, a flat-rate benefit paid out to all retirees, raised the living conditions for many elderly, but was criticized for breaking the link between an individual's past earnings and the final benefit. The subsequent post-war expansion of the pen-

[1] See e.g. Kruse and Ståhlberg (1977); Palmer (2002); Sundén (2006) and Könberg et al. (2006).

sion system, starting with the introduction of the earnings-related ATP plan[2] in 1960, was driven by a desire to strengthen this link. It was argued that the public pension system should not only provide support for the elderly poor and redistribute resources from individuals with high lifetime earnings to those with low lifetime earnings, but also prevent large falls in income for individuals with different pre-retirement income levels. The often referred-to policy objectives of insurance against longevity and consumption smoothing (see Barr and Diamond (2008)) thus gained ground.

Clearly, an expansion of the pension system in combination with rising life expectancy and falling employment rates would increase the cost of pensions. However, contemporary projections about the future relationship between pension contributions and pension payments raised no concern about the long-run sustainability of the pension system. At the time, positive growth rates were more or less taken for granted and expectations of a long-run growth rate of 2–4 percent were reasonable. At this pace, the sum of contributions was projected to increase rapidly and the system could be maintained with low contribution rates. However, the projections done in the 1980s painted a much bleaker picture of the future of the pension system as a result of lower-than-expected growth rates and a rapidly growing old-age dependency ratio. The need of reform was taken seriously and an extensive overhaul of the public pension system was legislated in 1994 and subsequently implemented in 1999.

While the current pension system provides good conditions for long-run financial stability, the level of future pensions is being disputed. Since benefits are adjusted for changes in life expectancy, younger cohorts must work longer to have the same pension level as the older cohorts. As seen in Figure 2, reforming the public pension system is likely to have played a role in increasing people's willingness to work at older ages. Along with stricter eligibility rules in the sickness and unemployment insurance programs, several measures have been taken to increase old-age labor supply, including the use of a flexible retirement regime[3], raising the mandatory retirement age from 65 to 67 and the introduction of age-targeted tax credits for individuals aged 65 or above. However, there are still rules that reinforce the norm of retiring at the age of 65, such as eligibility for the minimum guarantee and the fact that occupational pension rights typically can be earned after age 65 only under special agreement between the employer and the employee. In fact, employment rates right below this age are among the highest in the world, but low above (Pensionsåldersutredningen, 2012).

The reform process has had a great impact on the reformation of the second-pillar occupational pension plans in the last two decades. Just like the public pension system, these plans have been or are underway of being transformed

[2]Den Allmänna Tilläggspensionen (ATP)

[3]Under the flexible retirement regime, benefits can be claimed and are actuarially increased from the age of 61. Also, there is no restriction on combination of work and pension income.

from defined benefit (DB) to defined contribution (DC).[4] The shift towards DC has primarily been motivated by a desire to reduce aggregate financial risk and make pensions more actuarial. Another important objective has been to increase individual choice. Individuals not only have the possibility to choose their own investment funds, but also flexibility over the time period over which to withdraw the accumulated savings at retirement. This is the topic of the second essay.

1.2 The determinants of annuitization

Economists have long been interested in how people accumulate wealth over the life-cycle. Recently, however, more interest has shifted to understanding what happens to those assets during retirement. Poterba et al. (2011) argue that the reason that interest has shifted is that the accumulation phase of the "baby boomers" is nearly over and that they have started to enter their retirement years. How the baby boomers spend down their retirement assets is important not only because of the sheer size of the babyboom generation, but also because they are expected to live longer than previous generations. The magnitude of the risk of outliving one's resources arguably rises with the expected number of years in retirement. The babyboom generation is also attracting attention because they are experiencing more flexibility with respect to how their assets can be withdrawn compared to earlier generations.

The increased flexibility during the payout phase of retirement is mainly a result of the ongoing shift in pension provision from DB to DC. The shift from DB to DC has been particularly evident in private sector pension plans. This is true both in countries, such as Sweden, Denmark, Norway, Switzerland and Australia, where private pensions are mandatory or quasi-mandatory, and in countries such as Canada, United Kingdom and the US, where voluntary private pensions are more predominant (OECD, 2013).

So far, the transition to DC has had greater implications for the degree of flexibility in voluntary pension plans than in mandatory pension plans. While voluntary pension plans typically offer a lump-sum option as an alternative to a traditional life annuity[5], mandatory pension plans sometimes put a cap on the amount of retirement assets that can be cashed out (e.g. Denmark) and sometimes provide no option to annuitization at all (e.g. the Netherlands). In the most recent decades, however, even mandatory pension plans have started to introduce more liquid payout options, of which the occupational pension plans in Switzerland, Australia and Sweden are notable examples. Payout decisions in these pension plans often involve substantial amounts of retirement savings and are important determinants of old-age economic security.

[4] See Appendix A of Essay I for a definition of these concepts.

[5] A life annuity is a series of payments at fixed intervals, paid to the annuitant for as long as he or she is alive.

In Essay II, "The determinants of annuitization: evidence from Sweden", I study payout choices in a large occupational pension plan for Swedish white-collar workers, the ITP plan. Similar to the other occupational pension plans in Sweden, the ITP plan has introduced fixed-term payouts as an alternative to annuitization. The fixed-term payout options allow individuals to withdraw their pension assets during a fixed number of years with a minimum of five years. A strength of this study is that I match data on actual payout decisions in the ITP plan with administrative data from Statistics Sweden, resulting in a unique data-set with rich individual background information on the retirees. Previous studies that have acquired data from private pension sponsors and life insurance companies are limited with respect to individual background information whereas studies that use survey-based data usually lack information on actual payout decisions.

Studies of payout decisions in private pension plans, including this one, often relate their findings to the so-called annuity market participation puzzle. The annuity puzzle means that fewer people choose to insure themselves against longevity through life annuities than theory would predict. The theory on how people should withdraw their wealth at retirement was pioneered by Yaari (1965) who concludes that rational individuals with no bequest motive are always better off by converting all of their wealth to an annuity than investing the money in a bond. Since then, a number of explanations, that also have been tested empirically, have been proposed to explain the low demand for life annuities.

First of all, I find that 76 percent of the retirees in my sample choose to annuitize their pension wealth. One explanation for why so many choose the life annuity is that participants are defaulted into the annuity if they have taken no action by age 65. Had any of the fixed-term payouts been the default, or if individuals were required to make an active choice, the annuitization rate would almost certainly be lower. However, the popularity of the fixed-term payouts has risen over time with 20 percent opting for any of these payout options in 2008 compared to 31 percent in 2013. The trend towards shorter payout horizons is likely to continue given that knowledge about the existence and implications of alternative payout options spreads and that interest rates have continued to fall. Moreover, I show that fixed-term payouts yield similar, or even higher expected returns than the life annuity.

I go on by studying one of the most common explanations for the annuity puzzle, namely adverse selection. Since an annuity's value is increasing in the length of time that an individual expects to be alive to receive annuity payments, longer-lived individuals have greater incentives to purchase annuities. If annuities are priced to reflect the longevity of annuitants, then annuities will not be actuarially fair from the standpoint of typical individuals (Finkelstein and Poterba, 2004). I find clear evidence of adverse selection of shorter-lived individuals and individuals in bad health into the most liquid payout option.

Taxes are another important source of variation in the price of annuities that

individuals might respond to. The progressivity of the tax schedule implies that the effective marginal tax rates decrease with the length of the payout. I show that the expected value of the 5-year payout could fall by as much as 20 percent relative to the life annuity when taxes are taken into account. In line with the hypothesis that individuals evaluate the benefit's net-of-tax value for different payout lengths rather than its gross value, I document low demand for the 5-year payout among individuals with high income and large capital stocks, in particular those whose total retirement income exceeds the central government income tax threshold only under the 5-year payout.

1.3 The health effects of postponing retirement

Most people agree that a key issue in dealing with the fiscal implications of increasing life expectancy is to prolong the careers of older workers. One of the most frequently used policy tools to accomplish this is to raise retirement age thresholds, such as the age at which individuals can first claim their benefit (minimum claiming age), claim a full benefit (normal retirement age) or are obliged to retire (mandatory retirement age). If the response to such reforms is to work longer, the fiscal viability of the pension system should strengthen. Indeed, a number of studies have shown that raising retirement age thresholds is likely to induce people to postpone retirement (e.g. Mastrobuoni (2009); Behaghel and Blau (2012); Atalay and Barrett (2015)). However, it is not enough to consider the effect on the pension system's finances in isolation from the potential effects on other parts of the welfare system. If these reforms have an impact not only on people's retirement behavior, but also on future health patterns and mortality rates, there are effects on health care costs that need to be taken into account.

Is there any evidence, a priori, about the sign of this effect? Does continued work lead to an improvement in health and a corresponding reduction in the cost burden for the health care sector, or is health more likely to be adversely affected by later retirement? There is no strong consensus regarding this effect, which may operate in different directions. On the one hand, continued work may buffer negative lifestyle shocks or a general decrease in physical and social activity, slowing the decline in health that naturally accompanies aging. On the other hand, continued work may adversely impact health through increased duration of work-related stress and strain (Insler, 2014). This question therefore becomes an empirical matter, the answer to which is likely to differ across job types, age groups and the type of reform. What is clear, though, is that the fiscal implications of retirement-driven health changes (if they exist) could be significant. In 2014, individuals aged 65 and over comprised 20 percent of the Swedish population, but they accounted for 40 percent of total drug prescriptions and 47 percent of all patient discharges from public hospitals (Socialstyrelsen, 2015a,b).

The aim of Essay III is to find out whether retirees' health is affected by continued work and, if so, which aspects of health. To do this, I study the effects of a 2-year increase in the normal retirement age on individuals' utilization of health care and mortality. The reform, which was implemented in year 2000, implied that local government workers who previously could claim a full pension benefit at the age of 63 now had to wait until age 65 to do this. To arrive at a causal estimate of the effect of this reform, in conjunction with longer working lives, on health, I use a difference-in-differences approach. Specifically, the health outcomes of the individuals in the affected worker categories are evaluated against the health outcomes of private sector workers of similar age who experienced no change in the retirement age during the period of study. This approach credibly deals with the simultaneous effects that may cloud the true impact of retirement on health; poor health is not only a potential outcome of retirement, but may also bring about retirement (McGarry, 2004).

The health outcomes are constructed using detailed administrative data on prescription drugs, hospital admissions and mortality. The data allows me to track individuals' consumption of health care and risk of dying many years after retirement and classify these events into medical causes that we know are related to retirement. The results indicate that postponing retirement has no impact on the overall consumption of health care, nor on the risk of dying early. There is evidence, however, of a reduction in health care utilization related to diabetes and anxiety.

This study makes an important contribution to the ongoing policy debate on retirement age. It suggests that raising retirement age thresholds would not have a serious impact on short to medium health outcomes on workers in the type of jobs considered. The focus on Swedish workers in low- to medium-paid public sector jobs does raise questions about the external validity of the results, but could nevertheless be considered a strength since various discussions of increasing retirement age thresholds deal primarily with the concern that such increases could adversely affect individuals in low-skilled jobs. The study also contributes to the literature on the relationship between retirement and health, which contains surprisingly little empirical evidence on the health effects of pension reforms that promote longer working lives. Most previous studies that use quasi-experimental variation in retirement timing to investigate the effect of retirement on health look at reforms that make early retirement more attractive. The general result from these studies is that (early) retirement is associated with an improvement in health. The contrasting results of this essay suggest that potential effects of a change in the actual retirement age due to an increase in the retirement age may be different from the corresponding effect that follows from introducing more generous early retirement rules.

2 Income underreporting among the self-employed

Measuring the extent of tax evasion in the economy is a difficult task. Self-reports of tax compliance are vulnerable to substantial underreporting because respondents are unwilling to admit the true extent of their participation in illegal activities. Tax administrators have also relied on the use of fiscal audits to create estimates of the aggregate "tax gap", that is the difference between the amount of tax that should be collected by the tax authorities against what is actually collected. Audits provide useful information about the patterns of noncompliance with respect to such variables as type of income, occupation, region of the country and age, but can be carried out by the tax authorities only at substantial resource cost.

Tax administrations and researchers have turned to more indirect measures of tax evasion due to the shortcomings of the direct measures described above.[6] A common approach to learning about tax evasion when no direct measure exists relies on traces of true income. The "traces-of-income" approach looks for a variable that is correlated with true income. If the researcher can predict an individual's true income, inferences about evasion can be done by comparing the prediction to what is actually reported.

The micro-based traces-of-income approach was pioneered by Pissarides and Weber (1989) (henceforth PW).[7] They focus on estimating the extent of tax evasion among the self-employed who clearly have much better opportunities to evade taxes than wage earners.[8] Using the ratio of food consumption to reported income as the trace of evasion, they argue that if self-employed spend a higher proportion of their reported income on food than wage earners with similar household characteristics and recorded incomes, then this reflects underreporting of income, not a higher propensity to consume food. Using UK data, they find that self-employed spend around 10 percent more on food relative to wage earners, which implies that they underreport their income by 55 percent.

The PW method has been applied in many other countries, including the US and Sweden, where Hurst et al. (2014) and Engström and Holmlund (2009) find that self-employed underreport their income by 25 and 30–35 percent, respectively. Feldman and Slemrod (2007) follow a similar approach using the relationship between charitable donations reported on income tax returns

[6] See Slemrod and Weber (2012) for a survey on this literature.

[7] This approach has also been applied at the macro level. The most prominent example is electricity use, which arguably is a function of true income. A high ratio of electricity use to formal income is an indication of a relatively large informal sector (for example, see Johnson et al. (1997) and Lackó (2000)).

[8] A common finding is that self-employed account for a large portion of the tax gap. In the UK, for example, just under half of the aggregate tax gap is accounted for by small and medium businesses (HM Revenue & Customs, 2015). In Sweden, as much as 85 percent of the estimated unreported income can be attributed to small businesses that together only account for 9 percent of reported income (Skatteverket, 2006).

and reported income as the trace of evasion. Under the assumption that self-employed are not inherently more charitable than wage earners, they report a self-employment noncompliance rate of 35 percent.

In Essay IV, which is joint with Per Engström, we address one of the key methodological problems of the PW method: researchers typically only have access to current income measures, while theory suggests that a more permanent measure of the household's consumption potential may be more relevant. The use of current income may lead to overestimation of underreporting among the self-employed as transitory income fluctuations attenuate the estimate of the income elasticity of food consumption. Previous studies acknowledge the importance of using more permanent measures when modeling food consumption, but given the typical cross-sectional design of survey data, it has proven difficult to come up with a good measure of permanent income.

The standard way of dealing with transitory income fluctuations has been instrumental variable (IV) techniques. We propose a more direct and intuitive solution. By merging survey data on consumption to rich panel data from official tax and income registers, we can move towards a measure of permanent income by averaging household income both forwards and backwards in time. We then investigate how the estimate of underreporting is affected as we extend the time window over which income is aggregated. The results are highly consistent with a substantial degree of attenuation bias. In fact, the estimated degree of underreporting falls by more than one-third as we move from current income to a 7-year average measure of household income. We conclude that it is empirically relevant to account for transitory income fluctuations when applying the PW method and that the preferred way of doing this is by constructing relevant measures of permanent income. However, if the researcher lacks panel data to do this, our analysis also shows that capital income performs well as an instrument for permanent income.

References

Atalay, K. and G. F. Barrett (2015). The impact of age pension eligibility age on retirement and program dependence: Evidence from an Australian experiment. *Review of Economics and Statistics 97*(1), 71–87.

Barr, N. and P. Diamond (2008). *Reforming pensions: Principles and policy choices*. Oxford University Press.

Behaghel, L. and D. M. Blau (2012). Framing social security reform: Behavioral responses to changes in the full retirement age. *American Economic Journal: Economic Policy 4*(4), 41–67.

Engström, P. and B. Holmlund (2009). Tax evasion and self-employment in a high-tax country: Evidence from Sweden. *Applied Economics 41*(19), 2419–2430.

Feldman, N. E. and J. Slemrod (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *The Economic Journal 117*(518), 327–352.

Finkelstein, A. and J. Poterba (2004). Adverse selection in insurance markets: Policyholder evidence from the UK annuity market. *Journal of Political Economy 112*(1), 183–208.

HM Revenue & Customs (2015). *Measuring tax gaps 2015 edition: Tax gap estimates for 2013-14*.

Hurst, E., G. Li, and B. Pugsley (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of economics and statistics 96*(1), 19–33.

Insler, M. (2014). The health consequences of retirement. *Journal of Human Resources 49*(1), 195–233.

Johnson, S., D. Kaufmann, A. Shleifer, M. I. Goldman, and M. L. Weitzman (1997). The unofficial economy in transition. *Brookings papers on economic activity 2*, 157–240.

Könberg, B., E. Palmer, and A. Sundén (2006). The NDC reform in Sweden: The 1994 legislation to the present. In Holzmann, R., Palmer, E. (Eds.), *Pension Reform: Issues and Prospects for Non-Financial Defined Contribution (NDC) Schemes*, pp. 449–466. Washington, DC: The World Bank.

Kruse, A. and A. C. Ståhlberg (1977). *Effekter av ATP: En samhällsekonomisk studie*. Lund: Lund Economic Studies.

Lackó, M. (2000). Hidden economy–an unknown quantity? Comparative analysis of hidden economies in transition countries, 1989–95. *Economics of Transition 8*(1), 117–149.

Mastrobuoni, G. (2009). Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities. *Journal of Public Economics 93*(11), 1224–1233.

McGarry, K. (2004). Health and retirement do changes in health affect retirement expectations? *Journal of Human Resources 39*(3), 624–648.

OECD (2013). Pensions at a glance 2013: OECD and G20 indicators. `http://dx.doi.org/10.1787/pension_glance-2013-en`.

Palmer, E. (2002). Swedish pension reform: How did it evolve, and what does it mean for the future? In Feldstein, M., Siebert, H. (Eds.), *Social Security pension reform in Europe*, pp. 171–210. University of Chicago Press.

Pensionsåldersutredningen (2012). *Slutbetänkande av Pensionsåldersutredningen: Åtgärder för ett längre arbetsliv*. Stockholm: Elanders Sverige AB.

Pissarides, C. A. and G. Weber (1989). An expenditure-based estimate of Britain's black economy. *Journal of public economics 39*(1), 17–32.

Poterba, J., S. Venti, and D. Wise (2011). The composition and drawdown of wealth in retirement. *The Journal of Economic Perspectives 25*(4), 95.

SCB (2011). Befolkning, demografisk analys, "Återstående medellivslängd för kvinnor och män 1900-2010" (2011-12-15). `http://www.scb.se/sv_/Hitta-statistik/Statistik-efter-amne/Befolkning/Befolkningsframskrivningar/Demografisk-analys/`.

SCB (2015). Befolkning, befolkningsstatistik, "Återstående medellivslängd för åren 1751–2014" (publ 2015-03-19). `http://www.scb.se/sv_/Hitta-statistik/Statistik-efter-amne/Befolkning/Befolkningens-sammansattning/Befolkningsstatistik/25788/25795/Helarsstatistik---Riket/25830/`.

Skatteverket (2006). *Svartköp och svartjobb i Sverige Del 1: Undersökningsresultat*. Stockholm: Skatteverket.

Slemrod, J. and C. Weber (2012). Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance 19*(1), 25–53.

Socialstyrelsen (2015a). Läkemedel – statistik för år 2014. `https://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/19768/2015-3-17.pdf`.

Socialstyrelsen (2015b). Sjukdomar i sluten vård 1989–2014. `https://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/19921/2015-10-1.pdf`.

Sundén, A. (2006). The Swedish experience with pension reform. *Oxford Review of Economic Policy 22*(1), 133–148.

Wadensjö, E. (2011). De äldres återkomst till arbetsmarknaden: Ett långsiktigt perspektiv. In *Arbetskraftsundersökningarna (AKU) 50 år*. Stockholm: Statistics Sweden Publishing.

Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies 32*(2), 137–150.

I. A history of the Swedish pension system

Note: This is a revised version of Hagen (2013). The overall structure of the paper is the same, but contains less institutional details than the original version.

Acknowledgments: I thank Sören Blomquist and Håkan Selin for valuable comments during the whole work progress of this report. Special thanks to Bo Könberg for interesting discussions and useful suggestions. I also thank Karl Birkholz, Inger Johannisson, Ann-Charlotte Ståhlberg, Per-Gunnar Edebalk, Margit Gennser, Anna Norén and Oskar Tysklind for reading tips and helpful comments. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

1 Introduction

The Swedish pension system, as we think of it today, has existed for about 100 years. The pension system that was legislated in 1913 and implemented in early 1914 was in fact the world's first pension system to cover a whole population, i.e. the first to introduce the principle of universality. Contemporary pension systems in other countries typically excluded the non-working population and only made certain subgroups or professions eligible for old-age income. Sweden is also one of the few countries in Europe to have implemented comprehensive pension reform to deal with the challenges of an aging population. The current pension system has served as a role model for many other countries because it maintains long-run financial stability and also provides increased work incentives.

There exist detailed accounts of the most recent pension reforms in Sweden, the implementation of the current pension system in particular[1], but less attention has been paid to the early development of the pension system. This paper attempts to fill this gap. Starting with the implementation of the public pension system in 1913, it discusses the key components of each major pension reform up until today and how these have been shaped by the demographic and politico-economic context. The aim of the paper is to enhance our understanding of why and how the Swedish pension system has become what it is today and the circumstances under which various aspects of pension design has been more or less successful.

The paper proceeds as follows. Section 2 discusses the origin of the public pension system. It discusses the political and demographic factors that led to its implementation and the influence of the existing foreign pension systems on Swedish policy makers. Section 3 explains why Sweden moved away from the Bismarckian retirement insurance design and instead chose to embark on a Beveridgean path towards the implementation of a pension system based on the principle of basic security.[2] Section 4 discusses the supplementary pension plan, ATP[3]. Particular attention is paid to the problems of the ATP plan that made the pension system unsustainable in the long run and resulted in the comprehensive pension reform of 1998. Section 5 discusses the implementation and the rules of the current public pension system. Section 6 discusses the development of the occupational pension system, the so-called second pillar. The occupational pension system is a separate system, but it is also supplementary in nature. Its history is therefore closely related to that of the public

[1] See e.g. Kruse and Ståhlberg (1977); Palmer (2002); Sundén (2006) and Könberg et al. (2006).

[2] The distinction between Bismarckian and Beveridgean pension regimes is a common classification of pension systems that I make frequent references to in my analysis. These regimes refer to the characteristics of the welfare programs associated with the German Chancellor, Otto von Bismarck, in the late 19th century and the British economist, William Beveridge, in the 1940s. See Appendix A for a definition of these two concepts as well as a number of other important pension concepts that are used throughout this paper.

[3] Den Allmänna Tilläggspensionen

pension system. Section 7 takes a forward-looking perspective and discusses what we can learn from the historical analysis about pension reform at a more general level and what it means for the future of the current pension system.[4]

2 The origin of the public pension system

2.1 Early pension systems

The first public pension system in Sweden was passed in 1913 by the Swedish Parliament. It is often referred to as the first universal pension system in the world because it, unlike its predecessors, extended beyond the working population. The first *formal* pension system, however, was introduced by the German chancellor Otto von Bismarck about 30 years earlier. What citizens of western democracies today take for granted thus seems to be a rather recent phenomenon, especially considering the great expansion of the public pension system during the latter half of the 20th century. However, the idea of transferring wealth or other kinds of benefits from the working generation to the old generation is in fact as old as modern civilization, although not formalized in the way we think about pensions and certainly not universal in character.[5]

In pre-industrial Sweden, the traditional retirement systems were founded on family and property. The church laws passed at the end of the 17th century put the responsibility for taking care of the poor, who were often old and unable-bodied, on the congregations. Some (but far from all) congregations abided these laws and built almshouses, in which the poor were lodged (Ottander and Holmqvist, 2003). Gradually, however, the responsibility of supporting the poor was shifted from the congregations to the local authorities and was formally codified in the Poor Law of 1847.

Private pension solutions based on occupation had been in place long before the implementation of the public pension system. Most significantly, military pensions have a long history in Western civilization and have often been used as an element to attract and motivate military personnel.[6] In Sweden, old-age benefits to ex-soldiers were introduced in the 17th century during a period of frequent warfare. Initially, crippled soldiers and their families were offered to stay in designated homes, but as the number of war victims increased, payments in the form of grains and eventually cash were paid out. The first

[4] Tables A.1 and A.2 summarize important dates in the history of the public pension system and the occupational pension system, respectively. Table A.3 provides an overview of major public pension reforms.

[5] Ancient Roman writings by Cicero and Horatius, among others, reveal to us that people possessing an exalted societal position or significant financial means, chose to "retire with dignity" rather than work throughout life.

[6] For example, the U.S. Congress used pensions to provide replacement income for soldiers injured in battle, to offer performance incentives and to arrange for orderly retirements (Clark et al., 2003).

pension fund was formed by the navy already in 1642, in which the employees agreed to abstain from a certain proportion of their wage and allocate this money to the fund.

Amplified urbanization and public sector growth resulted in the emergence of new civil professions that introduced occupational pension funds similar to those of the military and the navy. Teachers, civil servants, bankers, and later on postal service employees, health service employees, law enforcement employees, and railway workers were covered by profession-specific pension agreements financed through voluntary or mandatory contributions. In most cases, the funds were primarily designed to support widows and the replacement rates were generally very low. It is also important to note that the great majority of the Swedish population were not covered by any formal pension plan up until 1913 (Ottander and Holmqvist, 2003).

2.2 Political and demographic development

Retirement insurance and the economic situation of the elderly became an important political question at the end of the 19th century. One important reason for this was the rapidly changing demographic structure of the Swedish population. The number of elderly increased substantially in the wake of the industrialization process, which brought decreased infant mortality and a subsequent drop in fertility rates. The demographic change was reinforced by high emigration rates. Between 1870–1900, around 670 000 out of 4.2 million citizens emigrated, most of them in their twenties. By the end of this period, Sweden probably had the oldest population in the contemporary world (Edebalk and Olsson, 2010). The growing number of elderly poor put severe financial pressure on the financial situation of many municipalities, which were responsible for providing poor relief. Growing inequalities across districts gave rise to calls for transferring the financial burdens of poor relief to the central government. Eventually, the two issues of fixing the poor relief and spreading the financial burden for local authorities became interlinked, to which the introduction of a universal pension system appeared as a solution.

There was also growing awareness of the link between poverty and aging. In England, the distinction between "worthy" and "unworthy" poor took shape. The "worthy" poor consisted of people that were unable to work because of age and weakness (Edebalk, 1999). It was argued that a pension system would allow these elderly to age with "dignity".

In Sweden, there was no social movement or organization dedicated to the poverty question in the late 1800s. The first political platform for advocates of a revision of the poor relief legislation was arranged in 1906, the so-called Congress on Poverty[7]. As a response to the demands presented by the congress, the ruling right-wing government created a commission of inquiry,

[7]Fattigvårdskongressen

referred to as the Old-Age Insurance Commission[8], in 1907. The commission emphasized that "worthy" retirees and unable-bodied should be offered better and more dignified social support than what was provided by the existing poor relief (Elmér, 1960). It also suggested that virtually all people should be covered by a public pension system. Needy retirees should not have to depend on ordinary poor relief or other individuals.

2.3 Choosing a pension system

Two issues were at the core of the pension debate. The first issue was whether a new pension system should be based on mandatory or voluntary participation. The pros and cons of each type of system that were raised are highly generalizable. Mandatory participation in a pension system is paternalistic in the sense that it forces individuals to save according to the rules of the pension system and not according to their own preferences for inter-temporal consumption smoothing and risk-taking. Pension systems based on voluntary participation, however, run the risk of having low participation rates, especially among people that are in most need of a paternalistic setting to counter life-cycle myopia. In fact, participation rates had turned out to be very low in countries with pension systems based on voluntary participation, including Belgium, France and Italy. Another well-known justification for mandatory pension systems is to prevent free-riders from exploiting the altruism of others (Lindbeck and Persson, 2003).

The second issue was whether pensions should be paid out to all individuals (universal) or only to individuals who meet certain criteria (means-tested). The main advantage of means-tested benefits is that they can prevent poverty at a lower cost than a universal pension system by targeting those in need. However, means-tested benefits also create disincentives for individuals to save. Individuals might intentionally undersave during their working years so that, by gaming the system in this way, they will qualify for the means-tested benefit (Feldstein and Liebman, 2002).

The first proposal to introduce old-age pensions in Sweden came from two liberals, Erik Westin and Adolf Hedin. Hedin saw the creation of social insurance covering workers as a way to stop social discontent and emigration, which had reached unprecedented levels in the early 1880s. Hedin even claimed that a universal pension program should be considered. A commission was set up in 1884, which presented its findings five years later. The majority opinion supported a universal plan, but the commission's proposal never reached the parliament (Heclo, 1974).

Following the German adoption of old-age insurance in 1889, the momentum for a public pension system intensified. Sketched by an influential professor of mathematics, Anders Lindstedt, two proposals based on Bismarckian

[8] Ålderdomsförsäkringskommittén

principles were presented to the parliament in 1895 and 1898 respectively. These were not universal and included mandatory worker insurance plans against accidents as well as retirement insurance. However, the proposals were either significantly diminished to suit the opposition or not passed at all by the parliament. The critics argued that the German insurance-based pension system did not suit the predominantly agrarian Swedish society. Thus, when the Old-Age Insurance Commission was set up in 1907, new ideas on the design of the public pension were required in order to overcome the considerable political obstacles it faced.

In a report presented to the liberal government in 1912, the commission quickly ruled out a pension system based on voluntary participation (Heclo, 1974). The commission believed that such a solution would leave too many out of the system. Ever since, all major parties, both socialist and non-socialist, have approved of mandatory pension systems. Three retirement insurance alternatives based on mandatory participation stood out as realistic.

1. *A universal pension system with flat-rate benefits*
   Flat-rate benefits are related only to age and citizenship, not past earnings and contributions. This type of pension had not been fully implemented in any country at the time. There was broad consensus in Sweden that the state budget was too weak to provide decent replacement rates within a flat-rate benefit system, especially since the economy was expected to deteriorate in the near future.
2. *The Bismarckian model*
   The Bismarckain pension system in Germany was designed to extend the standard of living achieved during work life into retirement. Pension benefits were roughly proportional to labor income averaged over the entire life course and comprised very few redistributive properties (Börsch-Supan and Wilke, 2004). Pension systems characterized by a direct link between the level of contributions and received benefits are referred to as "Bismarckian". Pensions were called retirement insurance rather than social security and workers perceived their contributions as insurance premia rather than taxes. The insurance character was strengthened by treating the pension system as a separate entity from the government budget.
3. *A means-tested model*
   The Danish model, implemented already in 1891, provided elderly with means-tested pensions financed by tax revenue. Liberals and conservatives generally opposed a means-tested system for reasons related to market inefficiencies and demoralization. They argued that the state pension fund necessary to sustain a means-tested system would grow too large and inhibit capital formation. They also strongly opposed the idea that "unworthy" elderly – people showing no work effort and negligent parents – would receive pensions (Edebalk, 2003b).

The German model undoubtedly influenced Swedish policy makers, but the reform proposal that was presented by the Old-Age Insurance Commission to the parliament in 1913 contained a quite different pension system. The reasons for diverging from the Bismarckian system were political and demographic. First, excluding everyone but workers from the retirement insurance plans was politically impossible. The greater majority of the Swedish population lived in the countryside and would not be covered by a German-like pension system. The agrarian community was well represented in the parliament and made up an important voter base for all parties (Edebalk, 2003a). This, in combination with the presence of a relatively strong central government, made it possible to introduce a publicly financed pension system that covered the whole population. Second, excluding non-workers from a pension system would not tackle old-age poverty effectively, nor alleviate the financial burden of poor relief for the worst off local districts.

2.4 The 1913 pension system

At this time, the pension debate was not characterized by large party disagreements. The common viewpoint was that something needed to be done about growing fiscal inequalities among municipalities and deteriorating poor relief. In May 1913, the Swedish parliament voted unanimously in favor of the introduction of a universal public system in line with the proposal drawn up by the Old-Age Insurance Commission. The important work of the commission marked the beginning of a long tradition within the history of the Swedish pension system of consensus-seeking decision-making based on the work of cross-party investigation agencies.

The commission had chosen to present a combination of the Bismarckian model and the means-tested model, since both had its advantages and disadvantages. The system consisted of two components. The first and most important component was fully funded and based on individual contributions collected by the local governments.[9] The contribution level was a function of reported income and benefits were actuarially fair. This component resembled retirement insurance, aiming at extending the standard of living acquired during work life to retirement. The pension benefit was paid out from age 67.

The second component was supplementary and means-tested where benefits were paid out to all retirees "in need".[10] As opposed to the contributory pension benefits, benefits were tax-financed and were thus financed according to the pay-as-you go (PAYG) principle. A premium reserve system for managing the pension contributions was set up.[11]

The main objective of the pension system was to alleviate old-age poverty

[9] Avgiftspension

[10] Tilläggspension

[11] See Appendix A for a definition of premium reserve system.

and provide elderly with decent retirement conditions. The benefits of lifting retired workers out of poverty were weighed against the costs of creating saving disincentives through a mandatory government pension program and the risk of encouraging intentional undersaving and social demoralization. Poverty relief, and soon also providing a minimum standard of living in retirement, remained at the core of the rationale for the pension system up until the implementation of the earnings-related supplementary pension plan, ATP, in 1960.

So how generous was the pension system as a whole? Table 1 shows the average pension for two important worker categories, factory workers and farmers, with income from both the contributory component and the means-tested component. The means-tested benefits relate to individuals who claimed maximum benefits, which implies that the table reflects the relative size of the public pension system at its best. Replacement rates were nonetheless rather low, especially for factory workers whose pension only accounted for 8–16 percent of their previous wage (column 4). Moreover, farmers' pension increased as a share of the average wage level over time and thus seemed to fare better than factory workers (column 5), but this was partially explained by higher real wage growth rate in the industrial sector.[12]

Even if the fully funded component was designed to be the main source of pension income, most retirees received the bulk of their benefits from the means-tested supplementary pension. Since benefits in the fully funded component were directly linked to the contributions paid, it took many years for an individual to amass enough contributions to be able to claim a substantial pension. In fact, the fully funded component did not have any significant socioeconomic effects in the first 20–30 years or so (Edebalk, 2003b). This was one of the main reasons why the participation rate remained at very low levels long after the pension system was implemented; many people simply ignored claiming pension benefits because they were too low. As seen in column 6 of Table 1, the participation rate, defined as the share of population over 67 years of age with some kind of retirement income, reached 80% as late as 1936.

3 Leaning towards Beveridge – a universalistic pension system

In the beginning of the 20th century, two types of pension systems crystallized in western Europe, sometimes referred to as the "two worlds" of pension systems (Bonoli, 2003). First, there was the Bismarckian social insurance system adopted by countries like Germany, Italy, France and Switzerland. Second,

[12]Replacement rate comparisons between farmers and factory workers suffer from difficulties in measuring average wage rates and determining their real value. Moreover, the registered wage income most likely did not fully reflect the actual standard of living of farmers.

Table 1. *Public pension benefits in relation to the average yearly earnings for farmers and factory workers*

Year	Avge. factory worker's wage (SEK/year)	Avge. farm worker's wage (SEK/year)	% of factory worker's wage	% of farm worker's wage	Participation rate
1914	1,301	811	11.3	18.1	2
1916	1,479	987	13.9	20.8	40
1920	3,607	2,352	8.1	12.5	47
1921	3,363	1,649	8.8	17.9	-
1926	2,707	1,328	16.4	33.5	57
1931	2,767	1,247	16.4	36.3	73
1936	2,848	1,378	16.2	33.5	81
1937	2,974	1,471	21.9	44.2	85
1941	3,615	1,919	29.4	42.6	93
1946	4,790	3,246	30.8	37.1	96
1948	5,912	4,222	35.0	39.8	100
1951	7,600	5,026	31	39.1	100
1956	11,300	7,704	35	43.5	100

Source: Elmér (1960)

there was the redistributive Beveridgean pension system with flat-rate benefits introduced by Great Britain and Denmark among others. Sweden, as we have seen, endorsed characteristics of both systems in the public pension system of 1913. Over time, most countries reformed their pension systems only within the frameworks of the Bismarckian and the Beveridgean systems, respectively.[13] The difficulty of changing the fundamental characteristics of the pension system gave rise to the idea of path dependence with respect to the long-term development of pension systems.[14] If Sweden would stick to its universalistic, hybrid version or embark on any of the two major pension paths remained unclear even two decades after the public pension system was implemented. However, during the 1930s the Beveridgean ideals took hold and greatly characterized the pension reforms of 1935 and 1946.

3.1 Perspectives on pension reform

Apart from numerous minor changes, the fundamentals of the Swedish pension system were left unchanged between 1913 and 1935. However, as neither the left-wing nor the right-wing parties were completely content with the pension system, there was an ongoing debate about how it could be improved during these years.

[13]The exceptional case is the Netherlands that switched from a Bismarckian old-age insurance plan to a Beveridgean basic security system.

[14]Path dependence theory was originally developed by economists to explain technology adoption processes and industry evolution. See Pierson (2000) for a formalization of path dependence within political science.

The nature of the pension debate in the 1930s was quite different from that preceding the 1913 pension system legislation. The latter debate was ideologically heated because it concerned the design of the fundamental characteristics of the pension system. 20 years later, the debate focused on the practical weaknesses of the current system that had become apparent over time. Two opposite perspectives on pension reform dominated the debate. The first perspective was characterized by a fear that the pension system would grow too large and become financially unsustainable. The second perspective emphasized the insufficiency of current benefit levels.

The right-wing government that was formed in 1923 represented the first perspective. Their arguments were based on pessimistic projections of the performance of the Swedish economy and fears that a large social insurance system would severely harm free market mechanisms.[15] There were also fears that increased contribution rates would further crowd out private savings and place more funds under the supervision of the government. However, the right-wing government was replaced by a social democratic government before it could implement any of their reform proposals.

In the wake of the economic downturn in the 1920s, an increasing number of elderly poor were forced to rely on locally provided poverty relief for old-age support. This form of retirement was considered even more "unworthy" now than a few decades earlier because the welfare state had developed considerably in many other respects since then. In 1933, payments from the contributory component were still lower than what was considered a normal pension, amounting to only 9 percent of the average wage in the industrial sector (Schmidt, 1974). The Social Democrats praised the universality of the pension system, but called for a pervasive pension reform that would have a significant impact on benefit levels.

3.2 The 1935 reform

A new commission of inquiry called the Pension Insurance Commission was formed in 1928 to investigate the scope for improvement in the pension system. Most importantly, the commission had to agree on whether the insurance (Bismarckian) character of the pension should be increased or decreased.

The resulting reform proposal was a compromise between radical right-wing politicians' calls for a non-redistributive, fully contributory pension system and the social democrats' preference for a redistributive pension system financed by tax revenue. However, the commission emphasized the transition away from the Bismarckian insurance design by referring to the new pension system as the people's pension, or folkpension, rather than retirement insur-

[15] Gösta Bagge, an influential professor of economics and the leader of the Conservative Party between 1935–1944, said that the pension system would grow uncontrollably like Frankenstein's monster (Elmér, 1960).

ance as before.

The bill that was passed in 1935 under the social democratic government was very much in line with the reform proposals of the commission. The main changes were:

- The premium reserve system was partially abandoned, shifting the larger share of pension funding from the pension fund to general tax revenue. The increase in the share of pension costs financed by tax revenue and the partial abolishment of the premium reserve system implied that the pension system became more integrated with the normal state budget. Its insurance character was weakened and it was to a less extent seen as a separate, self-financing entity than before.
- The previous contributory pension benefits were changed into an annual basic pension of SEK 100 plus 10 percent of lifetime contributions for both men and women.[16] The basic pension of SEK 100 only amounted to 3–4 percent of the average wage of a factory worker as shown by Table 1. Loosening the relationship between contributions and benefits in this way illustrates the direction away from the Bismarckian insurance design towards the Beveridgean, flat-rate benefit system that was to be fully implemented in 1948.
- The supplementary, means-tested benefits were increased to achieve reasonable pension levels in shorter time.
- The most controversial element of the 1935 reform was regional heterogeneity in benefit generosity. Retirees in urban areas received higher pension benefits than retirees in rural areas for a given contribution level.[17]

3.3 The 1946 reform

Reform proposals

Only a few years after the 1935 reform, a new commission[18] was set up to investigate a wide spectrum of welfare issues. Pensions were initially not on the main agenda of the commission, but were brought up when the shortcomings of the 1935 reform became apparent. The most debated issue was the time lag in claiming full benefits from the contributory component. In the light of the recent development in the UK and the eventual legislation of flat-rate benefits in 1946, it is natural to believe that the commission was influenced by the Beveridge report that had been published four years earlier (Beveridge, 1942). However, two out of three reform alternatives presented by the com-

[16] Folkpension

[17] The country was divided into three parts based on expected cost of living (Dyrortsgruppering). People that lived in places where costs were high were entitled to more generous supplementary pension benefits. This categorization of regions based on relative cost level was also applied in direct income taxation where the size of possible deductions depended on place of living (Elmér, 1960).

[18] Socialvårdskommittén

mission in 1945 were not in line with the principals of the Beveridgean model. Thus, rather than having a direct influence on the legislative process, Beveridge's ideas gained ground in Sweden by influencing the thoughts of several key players like Gustav Möller.[19] Three reform proposals were presented by the commission:

1. The first alternative included a basic pension of SEK 200 and a means-tested supplementary pension of SEK 800. The basic pension would correspond to approximately 5 percent of the average factory wage, slightly more than under the existing system.
2. The second alternative also provided total pension benefits of SEK 1000, but reduced the size of the means-tested component to SEK 400.
3. The third alternative, that was also ultimately implemented, would provide everybody with a pension of SEK 1000 independent of past contributions and income level. Even though there were still means-tested components in the form of housing supplements and supplementary wife benefits, the third alternative resembled the Beveridgean model of universal flat-rate benefits to a larger extent than the previous alternatives.

The members of the commission unanimously rejected the first alternative. The allotted share of the universal basic pension under this alternative was too small to bring about a significant increase in the general living standard of the elderly. They could not, however, agree on whether alternative two or three should be preferred. Since a complete abolishment of the premium reserve system was embedded in all three alternatives, all pension costs were to be financed by tax revenue. This put the financial issue at the core of the debate. Advocates of including a means-tested component next to a basic pension (alternative two) emphasized the excessive costs of having a universal flat-rate benefit of SEK 1000 (alternative three). They also argued that a pension system funded by tax revenue legitimized the use of means-tested benefits to a larger extent than a fully funded pension system (Elmér, 1960). Those who opposed means-tested benefits, on the other hand, highlighted the administrative simplicity of alternative three.[20]

The politicians found it easier to agree on the third alternative than the experts in the commission. Liberal and conservative politicians believed that the distortionary and demoralizing consequences of means-tested benefits outweighed the financial costs of universal flat-rate benefits. Social democrats generally supported the third alternative, too, as raising the living standard of the elderly had been central to the party's agenda for a long time.

[19] Gustav Möller was Minister for Health and Social Affairs from 1924–1926, 1932–1936, 1936–1938 and 1939–1951. He was also a member of the Pension Insurance Commission that was set up in 1928.

[20] Having no means-tested benefit, except for the housing supplements, would make it possible to abolish the paper-based population register of contribution rates.

The reform

The main characteristics of the new pension system that was legislated in 1946 and implemented in 1948, were the following:

- The pension system was completely unfunded. The existing pension fund assets and the interest earned on these assets were used to finance outgoing pension payments during a transition period.
- The most important component was a universal flat-rate benefit of SEK 1000 per year, which substantially improved the financial situation for the elderly.[21] The average annual income for male and female workers in the agrarian sector amounted to SEK 4,259 and SEK 2,184, respectively (Socialstyrelsen, 1931).
- The contribution rate was one percent of total taxable income and pensions could be claimed at the age of 67.[22]
- The central government was responsible for paying out benefits that were exempt from means-testing, whereas local authorities administrated the means-tested, supplementary components.

As soon as the folkpension came into force in 1948, it became obvious that the real value of the flat-rate benefit of SEK 1000 had decreased as a result of an increase in the general price level. In 1950, the parliament passed a bill that marked a first step in the implementation of an automatic indexation mechanism that tied pension benefits to the current inflation rate. In fact, a common view among economists was that automatic inflation indexation of pension benefits was undesirable because this would signal governmental powerlessness against inflation. They soon acknowledged, however, that inflation indexation should not worsen the financial situation of the least well off.

There were discussions about whether public pension benefits should be tied to real wage increases rather than inflation. It was argued that when the working population fared better, so should the pensioners. In 1957, the parliament unanimously voted in favor of a gradual "standard" benefit increase over a period of ten years that was supposed to mirror the rising living standards in the economy. However, price inflation remained the main indexation measure within the public pension system for another 40 years.

The 1946 reform was an important milestone. There was no longer any relationship between contributions and benefits. All benefits were paid out independent of past earnings and contributions. Thus, the insurance character that was relatively prominent in the 1913 pension system, significantly weakened in 1935, was now completely erased.

The gradual shift from a pension system with a strong insurance character to a system characterized by Beveridgean ideals was clearly reflected in the composition of revenue sources of the public pension system. Soon after the

[21] Allmän folkpension (AFP)

[22] This retirement age was considered too high by many. There were also discussions about whether the retirement age should be fixed or flexible.

implementation of the first public pension system, individual contributions financed as much as 50 percent of total pension costs. This share was only 16 percent in 1953. Meanwhile, the share of total pension costs financed by central government taxes increased from 23 percent to 72 percent during the same period.

The increasing generosity of the pension system was also reflected in the pension fee structure. The contribution rate remained at one percent of taxable income until 1954 after which it was raised stepwise to four percent over a period of five years. The redistributive properties of the pension system were strengthened by raising the maximum fee payable and exempting individuals with low income from all pension payments. As pension levels rose, more people found claiming more worthwhile. As seen in Table 1, the participation rate rose from 81 to 96 percent between 1936 and 1946 and reached 100 percent by the time of the implementation of the folkpension in 1948.

The relatively generous basic pension substantially improved the living conditions for many elderly, which none of the previous reform attempts had achieved successfully. However, relating the increase in real pensions to the increase in real wages over the same period shows that recurrent calls for extending pension benefits often were legitimate. As seen in Table 1, factory workers experienced a modest replacement rate increase of 5 percentage points between 1941–1956, reaching 35 percent in 1956. Farmers claimed higher pensions as a share of their earned income, but given that the average income level in the agrarian sector was very low at the time, pension benefits were not that high after all. These replacement rates were still considered too low by many, especially by high-income earners whose acquired standard of living was far from sustained into retirement. One way to sustain the standard of living into retirement would be to introduce an earnings-related pension component.

4 The rise and fall of a defined benefit pension system

The introduction of an earnings-related pension component had been a recurrent political topic in the decades following the implementation of the public pension system in 1913. However, none of the reform proposals had been put into practice. In the meantime, large labor market groups tried to supplement the basic pension with negotiated occupational pension. Such pension plans had existed for quite some time already, although on a rather small scale.

The debate intensified as large differences in real retirement income between different income groups and professions emerged. The debate centered on the financial situation of retired blue-collar workers, who had been less successful in arranging supplementary private pension solutions than white-collar workers. While white-collar worker pensions were related to previous earnings, blue-collar workers normally received fixed pension benefits, inde-

pendent of the previous wage level. The widespread use of fixed pensions implied that among the few blue-collar workers who actually received occupational pension, very few received a pension that reflected the size of their previous earnings. Moreover, rising real wages in the economy implied that pensioners lagged behind the working generation. This laid the foundation for one of the greatest political battles in contemporary Swedish history; the ATP plan.

4.1 A non-conventional pension reform

In the mid-1950s, the coalition government of the Swedish Social Democratic Party and the Agrarian Party initiated a block overriding pension commission that they hoped would negotiate a proposal on a radical extension of the current folkpension and also some kind of supplementary, earnings-related pension component. The commission's proposal would then provide the basis for a traditional compromise solution (Hermansson, 1993).

The members of the commission fully agreed on a gradual increase of the real value of the folkpension over a period of ten years. There was, however, considerable disagreement on the design of the new supplementary pension, even among the government coalition partners. The political stalemate evoked demands for a national referendum to help resolve the situation. The people that were entitled to vote in the referendum that was subsequently held on October 13, 1957, could choose between three alternatives. The alternatives were referred to as *Linje 1*, *Linje 2* and *Linje 3* and corresponded to the policy preferences of the Social Democrats, the Agrarian Party and the bourgeois parties respectively:

- *Linje 1* – All employees would receive statutory supplementary pension based on previous earnings. Pension benefits would be financed by employer contributions and be tied to the nominal price level to secure their real value.
- *Linje 2* emphasized the voluntary character of the earnings-related pension component, which should be complementary to the existing basic pension. The basic pension would be raised and the government would guarantee the real value of the pension benefits.
- *Linje 3* – Accession to the supplementary pension would be achieved through individual, group or, most preferably, collective agreements. Labor market parties would agree to set up designated pension funds that would finance the pension entitlements of the employees. The government would not guarantee to uphold the real value of the pension benefits.

The outcome of the referendum was unfortunate as each side could regard themselves as winners in some sense.[23] The ambiguous outcome of the referendum caused political turmoil, which culminated in a re-election for the seats of the second chamber. *Linje 1* eventually won a majority with only one vote in the parliament.[24]

The ultimate implementation of ATP in 1960 was a great success for the Social Democrats. The ATP plan was one of the most important building blocks in the Swedish welfare state, for which the Social Democrats could claim full credit. It became a symbol for the Social Democrats' idealistic struggle against conservative forces, to which leading figures of the Social Democrats repetitively and effectively appealed to during the following decades (Lundberg, 2003). The 1960 reform was unconventional in the sense that one party alone was responsible for the design and the implementation of a major pension system. The reform marked an end of the long-lasting trend of traditional consensus-seeking decision-making based on the work of cross-party investigation agencies. The reform was also unconventional in the sense that the fundamental principle of the new ATP plan did not cohere with the traditional policy platform of the Social Democrats. From the very start of the Swedish pension debate in the 1890s, the Social Democrats had pushed for universal coverage and basic security for all citizens. These principles were set aside in the ATP plan in favor of the so-called loss-of-earnings principle. The loss-of-earnings principle would sustain the acquired standard of living into retirement by letting the size of the benefit depend on previous earnings. The introduction of an earnings-related component meant that the Bismarckian character of the pension system was strengthened.

4.2 Properties of ATP

The ATP reform was the last major change to the public pension system before the comprehensive pension reform in the late 1990s. ATP was a mandatory

[23] *Linje 1* received 45.8 percent of the votes, *Linje 2* 15.0 percent and *Linje 3* 35.3 percent. 3.9 percent of the votes were blank and the vote turnout was 72.4 percent (Elmér, 1960). Although *Linje 1* got more votes than the other alternatives, the share of votes for *Linje 1* was lower than the share of parliamentary seats currently held by the Communist Party and the Social Democratic Party together. In this sense, the outcome of the referendum was a failure for the Social Democrats rather than a victory. Meanwhile, the Agrarian Party did much better in the referendum than in any parliamentary election during the 1950s and therefore perceived the outcome of the vote as a great success. Furthermore, the opposition parties, representing *Linje 2* and *Linje 3*, together received greater support than the left-wing parties.

[24] The re-election made the two blocs equally large. However, prior to the decisive parliamentary vote on the ATP plan in May 1959, a member of the parliament of the Liberal Party, Ture Königson, completely changed the course of the game when he announced that he would abstain from voting. Despite massive critique from fellow party members, Ture Königson argued that it was more important to get some kind of supplementary pension system into place rather than a system according to his party line.

PAYG system covering all employees and operated as a separate pension plan next to the folkpension that had been in place since 1948. The size of the ATP benefit was determined by the so-called 15/30 rule: while only income during the top 15 income years mattered, 30 years of pensionable income was required to qualify for the maximum replacement rate of 60 percent. The benefit was reduced by 1/30 for each year the number of working years was less than 30. Together with the folkpension, the ATP plan provided a gross replacement rate of roughly 65 percent for an average worker. The average replacement rate in the OECD countries was 57 percent at the time (Queisser and Whitehouse, 2005)

Payments were financed by mandatory proportional payroll taxes levied on wages. No contributions were paid on wage portions above a certain threshold called the income ceiling.[25] Although there was no specific ATP fee for individuals, everyone still had to pay the pension fee for the folkpension. As seen in Table A.3, the fee was 4 percent just before the ATP reform and had risen to 5.86 percent in 1994 (Ståhlberg, 1993).

The fees were collected in designated pension funds, the so-called National Pension Funds[26]. The pension funds administered the pension capital and were responsible for, but did not guarantee, outgoing pension payments. During the initial years of the ATP plan, the contribution rate was set so that the system would build up a surplus to act as a buffer against cyclical shifts in contributions. The surplus would also help offset the expected decrease in private saving that would follow from making more capital available for lending (Sundén, 2006). The surplus could also be used to finance outgoing payments in case payroll tax revenue from the employers was insufficient.

From 1961 and 30 years onwards, the public pension system was subject to more than 50 changes. All changes, with a few notable exceptions[27], were referred to as "improvements", which in fact were "extensions", of the pension system. During this period, the Swedish public pension system expanded rapidly.

4.3 Problems with ATP

ATP rapidly received widespread support because it had an immediate and strong impact on the financial situation of the current elderly. Not even the bourgeois governments that were in power between 1976 and 1982 made any major changes to the ATP plan, partially because its design had turned out to be

[25] The income ceiling in the public pension system was 7.5 price base amounts, which in 1960 corresponded to annual earnings of SEK 31,500.

[26] AP-fonderna

[27] In 1980, under a bourgeois coalition government, an indexation change unexpectedly reduced the real value of pensions. Pensions were not fully adjusted to the inflation rate, which eroded the real value of the pension entitlements (Kruse, 2003). The 1988 decision to abolish the widow pension is another example.

particularly beneficial for people with shorter work histories and rising earnings profiles (typically liberal and conservative high-income workers). Hence, ideological forces were not the main contributory factor to the rise of a new debate on the need for pension reform in the beginning of the 1980s. Instead, deteriorating fiscal balances and sluggish growth exposed the financial instability inherent in the pension system.

The design of the ATP plan made the income distribution between the working population and the pensioners sensitive to changes in economic growth. Because benefits from ATP were indexed to follow prices, there was no link between the wage level of the working population and pensions. In times of rising real wages, contribution rates can be kept low and the standard of living of the working population rises relative that of the pensioners. However, when real wages fall and productivity lingers, contribution rates must be increased in order to finance the pension costs. In the 1980s and the early 1990s, Sweden experienced low or negative growth. As a result, earned pension rights and benefits rose faster than wages and contributions.[28]

Changes in economic growth also eroded the ATP plan as a source of income replacement. Because the income ceiling was indexed to follow consumer prices, real wage growth meant that successively larger proportions of the population earned wages above the ceiling (Sundén, 2006). In other words, the activation of the income ceiling for an increasingly larger proportion of the population meant that ATP gradually came to look more like an enhanced folkpension. The erosion of the ATP plan particularly disadvantaged workers who were not covered by occupational pensions that compensated for income above the ceiling. These were exclusively blue-collar workers.

Another problem was related to the redistributive properties of the ATP plan. At the center of this issue was the 15/30 rule, which had been hailed as the main foundation of the ATP plan that would ensure fairness and redistribute resources from high-income to low-income earners and from men to women. In practice, however, the ATP formula redistributed income from those with long working lives and a flat life-cycle income (typically low-income workers) to those with shorter work histories and rising earnings profiles (typically high-income workers) (Sundén, 2006). Table 2 shows that white-collar workers with high or middle positions in the 1944–1950 cohorts actually receive a higher pension as a share of paid contributions than low-income workers. The redistributional implications of ATP were regarded as "unjust" and undermined its political legitimacy. There were growing fears that future generations would refuse to fulfil the "implicit generational contract" implied by ATP by cutting outgoing pension benefits or raising costs for public service used by pensioners (Lindbeck, 1992).

[28]The excessive sensitivity of the income distribution to changes in the growth rate between these two groups was illustrated by Ståhlberg (1989), who compared the average net-of-tax pension to the average net-of-tax income in the economy under different assumptions about the growth rate. Her calculations showed that the share of average net pension income from ATP and the folkpension to average net income would amount to 119 percent in 2025 without economic growth, but only 61 percent under a growth rate of two percent.

Table 2. *The ratio of total contributions to outgoing pension benefits for different income groups, 1944–1950 cohorts*

Socio-economic group	Men	Women
Senior officials	0.84	1.06
Officials on middle level	0.88	0.88
Officials on lower level	0.84	0.73
Qualified workers	0.82	0.79
Unqualified workers	0.77	0.64
All	0.83	0.78

Source: SOU 1998:3

The ATP plan was also criticized for distorting individuals' labor supply decisions, which most likely contributed to the decline in labor force participation, particularly among men above the age of 60, witnessed during the second half of the 20th century. The introduction of the ATP plan had two opposite effects on the incentives to remain on the labor market. For some workers, increased pension benefit generosity created a wealth effect toward earlier exit from the labor market. This effect was amplified by the lowering of the normal retirement age from age 67 to 65 in 1976 and the extension of early withdrawal possibilities. However, ATP also created incentives to work longer as the benefit increased during the years of maturity of the pension plan. The actuarial addition, which was paid out to individuals who delayed retirement, also encouraged people to remain in the labor force.

The 15/30 rule had more explicit effects on the labor supply decision. Because extra pension rights were not earned for working years after 30 years on the labor market, unless this income exceeded the income of the 15 best years, increased labor supply did not necessarily translate into a larger benefit. Contributions paid during working years that did not affect the size of the pension benefit more or less worked as a pure tax on labor income.

5 The great compromise – a notional defined contribution system

5.1 The reform process

Although the abolishment of the ATP plan was not as politically heated as its implementation, the reform process leading up to the implementation of the new pension system in 1999 was long and extensive. A pension commission was set up in 1984 as a response to the growing number of reports that highlighted the need of reform due to the problems discussed in the previous section. The commission failed to produce explicit reform proposals due to political disagreement among its members and a widespread unwillingness to

change the fundamentals of the existing pension system.

The economic crisis that unfolded in the beginning of the 1990s pressed the new right-wing government to take fast and firm action. It was well-known that the ATP plan was vulnerable to economic recessions. A new cross-party investigatory body, referred to as the Working Group on Pensions, was formed.[29] The working group was instructed to formulate a reform proposal characterized by long-term considerations and stability. The new system should, among other things, make pensions more responsive to the general state of the economy, strengthen the link between contributions and benefits, contain more transparent redistribution mechanisms, allow for a flexible retirement age and encourage long-term saving (Pensionsarbetsgruppen, 1992).

The group's first report contained many important guiding principles of the design of the new pension system and it became clear that a rather extensive reform was underway. The main ingredients were the following (Pensionsarbetsgruppen, 1992):

- The adoption of the life-income principle implied that all life-time earnings would count towards the calculation of an individual's pension. It also implied a switch from defined benefit (DB) to defined contribution (DC).
- The reformed pension system, like the previous one, would be mandatory.
- A flexible retirement age that would make it possible to retire at any time between 60–70. This would provide increased work incentives and play down the age of 65 as the "normal" retirement age.
- Replace price indexation for pensions with wage indexation.
- The reformed pension system would either be PAYG or a combination of a PAYG and a premium reserve system (no consensus reached at the time).

The working group found it particularly hard to agree on two issues – the fully funded component and the income ceiling in the public pension system.

The introduction of a mandatory financial defined contribution (FDC) plan in the public pension system was central to the bourgeois parties. They argued that a pension plan based on private individual savings would strengthen the sense of ownership of individuals' earned pension rights and balance an increasingly powerful concentration of power within the economy by the state that had come about through the rapid growth of the National Pension Funds. The reform prospects engaged not only the political parties, but also powerful actors in the financial sector. The Social Democrats, on the other hand, defended the National Pension Funds and viewed them as an important collective saving instrument. They did agree to the introduction of an FDC component

[29] The new working group consisted only of members of the parliament and hence excluded labor market representatives and pensioner organizations. The composition of the working group was quite unique and contrasted sharply to the Swedish corporatist custom of formally incorporating labor market partners in social welfare reform processes.

provided that a separate and more important PAYG component was formed.

The second issue was the design of the income ceiling. While the Social Democrats preferred wage indexation to price indexation, the bourgeois parties were split. The gradual erosion of the income ceiling was the best example of how dysfunctional the ATP plan had become, but a pension system based on the principle of basic security was exactly what the bourgeois parties wanted. However, keeping a price index and thereby allowing for an eventual transformation of the system into a large basic pension was unacceptable for the Social Democrats. The working group eventually agreed to index the income ceiling to wages.

There was also disagreement about the redistributive properties of the income ceiling. The bourgeois parties wanted to levy contributions only on earnings below the income ceiling in order to strengthen the insurance character of the pension system. The Social Democrats, on the other hand, wished to maintain the progressive feature of the pension system and keep contributions above the income ceiling. A compromise solution was eventually reached.[30]

In June 1994, the parliament passed a bill of a reformed pension system, which was based on the final report of the Working Group on Pensions (SOU 1994:20). The bill was referred to as a "general proposal" because it contained some unresolved issues. Another four years of negotiations were needed to lay out all the main ingredients of the pension system, which was eventually legislated in June 1998.

There are several key explanations for why the reform process was successful. First, it was essential that the bourgeois parties and the Social Democrats could agree on what was portrayed as the most central property of the new pension system, the life-income principle. By introducing the life-income principle at the expense of the 15/30 rule, the bourgeois parties' preference for a stronger insurance character, and the Social Democrats' wish to end the redistribution from individuals with flat earnings profiles to individuals with steep earnings profiles could be satisfied at the same time (Könberg, 2008). In this sense, the reform had no clear winner. The reform can either be viewed as a necessary reduction of ATP, which main features are nonetheless kept intact – the new pension system is mandatory, earnings-related and provides benefits up to a certain income threshold. Or it can be viewed as the first introduction of a mandatory FDC plan.

The work of the small, but efficient and consensus-seeking investigative commission was the second key to the success of the reform process. Although the economic crisis may have helped the process by increasing the readiness of the policy makers to make tough decisions, the pension reform was the continuation of a long process that began in the mid-1980s (Marier,

[30]The pension contribution was split in half between employers and individuals. The payroll fee levied on employers is paid on all earnings, whereas the individual part, the so-called general pension contribution, is paid only on earnings below the income ceiling.

2002).

Third, opposite to what is often assumed, path dependence may have made pension reform more likely. Intuitively, path dependence seems likely to worsen the prospects for pension reform, as the policy makers are constrained by the institutional setup of the old system. However, given the state of the ATP plan and the general economy, all parties were aware that any credible reform proposal would have to be based on a financially and politically feasible transition plan. By making some actions "impossible", path dependence fostered broad political agreements.[31]

5.2 The three tiers of the new pension system

The public pension system consists of three tiers. The first tier, the minimum guarantee[32], is a means-tested pension supplement that ensures individuals with no or low pension income from the earnings-related component a minimum standard of living in retirement. The second tier is a notional defined contribution plan (NDC) referred to as the inkomstpension. The third tier, the premium pension, is an FDC plan with individual choice. The minimum guarantee is financed by general tax revenue, whereas the financing of the earnings-related component is shared between employers and employees. The contribution rate is 18.5 percent of earnings; 16 percent is credited to the notional account and 2.5 percent is contributed to the FDC plan. The first cohort to participate in the system is the group born in 1938; it will receive one-fifth of its benefit from the new system and four-fifths from the old system. Each cohort thereafter will then increase its participation in the new system by 1/20, so that those born in 1954 or later will participate only in the new system. In 2040, benefits will be completely paid from the new system (Sundén, 2006).

The minimum guarantee

The minimum guarantee replaced the folkpension and the so-called special supplement as the basic protection component in the new system. The minimum guarantee is indexed to prices, which implies that the role of the minimum guarantee decreases in times of high economic growth. The benefit is worth approximately 35 percent of the average wage of a blue-collar worker and ensures a minimum living standard of living in retirement.

The minimum guarantee is means-tested against the NDC component and is therefore only paid out to individuals that receive a small or no earnings-related pension. The reduction plan against the NDC component has been criticized for creating disincentives to work around retirement, especially for

[31] See Lundberg (2001), Lindbom (2001) and Lundberg (2003) for a thorough discussion on this issue.

[32] Garantipension

lower-paid workers with a small inkomstpension. Another criticized feature of the reduction plan is that it disregards income from the occupational pension system (Barr, 2013).

The NDC component

The main part of the current pension system is the NDC component, the inkomstpension. At the core of the NDC plan is the life-income principle. The main objective of the life-income principle is to reinforce incentives to work; additional years' of work should translate into higher benefits. Pension rights are accumulated on all types of earnings from age 16 with no upper age limit. Benefits from the inkomstpension can be withdrawn from age 61 and workers have statutory rights to work until age 67. This means that there is no formal retirement age in the new pension system. The NDC plan makes use of non-financial individual accounts to keep track of individual contributions. This means that annual contributions are used to finance current benefit obligations as in any PAYG system.

The rate of return in the NDC plan is determined by the per capita wage growth. The use of a wage index for benefits in payment places some of aggregate wage growth risk on retirees. It also imposes some aggregate financial risk on the NDC system. If the rate of growth of the wage bill is slower than that of average wages, for instance due to a fall in the size of the work force, total benefits grow faster than the contributions financing them. However, this indexation technique was considered the best way to keep a tight link between the living standard of the young and the old, which was priority.[33]

The inkomstpension was designed to bring financial stability into the pension system. An important component in this regard is the mechanism that adjusts pension benefits to changes in average life expectancy. An individual's annual benefit is calculated by dividing the balance in his or her notional account by an annuity divisor.[34] The divisor is determined by average life expectancy at retirement for a given cohort at the given retirement age. As long as life expectancy continues to increase, future cohorts will receive ever smaller monthly pension payments, as earned pension rights are distributed across more years.

Two issues with the way in which benefits are adjusted to longevity have

[33] A combined price-wage index would place less risk on retirees, but less indexing to wages, and so less correlation with revenues, would also increase the probability of a need for legislative intervention. An alternative indexing strategy is to use the change in the contribution wage sum as the measure of the rate of return. Auerbach and Lee (2009) shows that an NDC system in which rates of return are based on total rather than per capita economic growth is inherently more stable. Indexing to average wage growth might be more comprehensible from an individual's perspective, but the growth rate of the contribution wage sum provides a more relevant measure of the system's financial capacity because it takes the growth rate of the workforce into account.

[34] See Pensionsmyndigheten (2012a) for the mathematical representation of the annuity divisor.

received particular attention in the recent years. The first issue is that the annuity divisor converts NDC accounts into annuities using factors that depend on age and cohort, but not on gender or earnings level. Because women and high income earners, on average, live longer than men and low income earners respectively, the new pension system redistributes from men to women, and within each gender, from low income earners to high income earners. Moreover, those with shorter life expectancy will have an incentive to claim early, while those with longer life expectancy will have an incentive to delay claiming (Diamond, 1999). This means that the cohort-based longevity adjustment reduces the progressive feature of the pension system. However, the benefits of adjusting pensions along other dimensions than age and cohort should be weighed against its practical difficulties, such as classifying individuals into relevant groups and avoiding moral hazard issues.

Second, since benefits are adjusted for cohort life expectancy using the actual (period) mortality table rather than a projected (cohort) table, there is a risk that total benefit payments for the cohort will exceed their total contributions if they live longer than expected. In fact, De Gosson de Varennes (2016) shows that the cohort balances systematically fall short of paying fully for the cohort pool and that the deficits can amount to as much as 5–7 percent of cohort total pension costs. The implied intragenerational redistribution from future cohorts to current cohorts may be regarded as unfair and undermine the political legitimacy of the pension system.

To ensure financial stability, the policy makers also added an automatic mechanism that abandons indexation by average wage growth when the financial stability of the system is threatened. This happens when a measure that captures the ratio of the systems' assets to its liabilities falls below a certain threshold. The mechanism works automatically and does not require any political decisions. This goes in line with the objective of the pension reform to create a pension system autonomous from discretionary changes and minimize the risk of manipulation for political gain. However, the automatic balancing mechanism implies that the Swedish NDC plan does 100 percent of the adjustment on the side of benefits and zero on the side of taxes. Avoiding arbitrary tax changes and too much of new legislation makes the system more predictable and transparent, especially for employers, but it should be balanced against the risks falling on the elderly.

The individual account – the premium pension

The premium pension is the third component of the public pension system. The premium pension constitutes a relatively small portion of the new system. Of the 18.5 percent total contribution rate, 2.5 percentage points are allocated to individual financial accounts where the individuals choose how to invest their funds. The Swedish Pensions Agency keeps track of the accounts and executes the desired portfolio investments.

The premium pension was created for three purposes. First, funded indi-

vidual accounts were believed to increase overall savings in Sweden. Second, the policy makers wanted to allow participants to take account of the higher return in the capital markets as well as to tailor part of their pension to their risk preferences. Finally, an FDC plan is inherently immune against financial instability, as an individual's benefits are directly financed by her own accumulated contributions. Worker choice in the premium pension may be seen as another unnecessary source of risk for the elderly, but it can also be seen as a risk-spreading device. It implies that the rate of return on at least some part of an individual's total contributions to the public pension system is not dependent on the notional interest in the NDC plan.

The development of the share of active vs. passive savers in the premium pension underscores the importance of a well-designed default plan. In 2000, more than two-thirds made active choices. This share declined to 18 percent in 2001, 8 percent in 2005 and 1.5 percent in 2011 (Pensionsmyndigheten, 2012b). Rather than encouraging participants to make active portfolio choices, the authorities nowadays focus on the long-run performance of the default fund and improving the financial knowledge of the public so that participants can make good investment choices (Sundén, 2006).[35] The current default fund automatically decreases the risk level of the portfolio over the life-cycle, which helps overcome the potential problem of adverse selection of low-income workers into a low-risk default fund.

6 The second pillar

The history of Sweden's occupational pension system, referred to as the second pillar, begins long before a universal public pension system was even considered by the policy makers. Many of the questions that have repeatedly come back regarding the design and the objectives of the public pension system had already been addressed within an occupational pension framework. The section discusses the origin of the occupational pension plans and the mutual influence between these pension plans and the public pension system.

6.1 Early occupational pensions

The first worker category to be covered by a formal occupational pension plan was central government employees. The first central government pension plans were based on the principle that civil servants should be entitled to keep their jobs for a whole life-time. Already in 1778, it was decided that a civil servant who had turned 70 could choose to resign and keep his full salary throughout life. In 1877, retirement at age 70 was made mandatory and about two-thirds of the previous wage was paid out to a pensioner who retired at 65.

[35] For more information on investment behavior in the premium pension, see Engström and Westerberg (2003), Palme et al. (2007) and Säve-Söderbergh (2012).

Only minor amounts were paid out to individuals who retired earlier than this. However, because life expectancy at the time was around 50 years, not many people could expect to live long enough to claim a pension. The retirement age was subsequently lowered to 67 for men and 60 for women in 1907.

In 1907, the central government pension plan introduced a contributory component. One-third of the total pension cost of an individual was supposed to be covered by her own past contributions and the rest by the state. This component provided a model for the contributory component in the public pension system of 1913. The insurance character was subsequently weakened during the 1930s when the financial responsibility of paying out pensions was shifted back to the central government; how much the individual had contributed to the system was considered less important (Elmér, 1960). Despite these changes, the replacement rate of 65 percent was more or less upheld until the end of the 1980s.

In the private sector, most workers had to rely on their employer for old-age income. Farm workers and servants were considered to be part of their master's family, for which the master was legally required to provide old-age support. The first private sector pension plan that covered large groups of workers in different professions was introduced in 1917. Similar to the central government pension plan, two-thirds of the insurance premium was paid by the employer and the rest by the employee. The average replacement rate was 60 percent and benefits were generally sufficiently high as to disqualify for the means-tested benefit in the public system.

The early occupational pension plans were important for raising the standard of living in retirement, especially before the public pension system started to provide reasonable replacement rates in the mid-1930s. However, as the role of occupational pensions grew, so did differences in replacement rates between workers who were eligible for occupational pension and workers who were not. In the central government, only those who were "permanently" employed were eligible for central government pensions.[36] In a similar fashion, blue-collar workers were left out of the pension plan in the private sector, which covered white-collar workers only. The majority of the blue-collar working population was not part of any pension agreement and only had their basic pension to live off as pensioners.

To illustrate the importance of occupational pension of a covered individual's total pension, Table 3 compares the replacement rates for married railway and factory workers in Stockholm for selected years. Railway workers were employed by the central government and were thus entitled to occupational pension, whereas factory workers only received retirement income from the public pension. In 1922, railway workers had a replacement rate of 58 percent compared to 12 percent for factory workers. The difference became smaller over time as a result of the increased generosity of the public pension system, but even in 1956, the replacement rate of factory workers was only half of that of railway workers. As discussed in Section 4, the persistence of replacement rate differences between different groups on the labor market became a de-

[36]Permanent employment also implied that 35 years of service were needed to receive a central government pension (Schmidt, 1974). Non-permanent employees got covered in 1934.

Table 3. *Replacement rates for male lower-grade government employee and male factory worker without occupational pension in Stockholm for certain years*

Year	Railway worker	Factory worker
1922	58	12
1935	50	12
1946	51	30
1956	77	39

Source: Elmér (1960)

bated topic in the end of the 1950s and was one of the most important factors behind the introduction of ATP in 1960.

6.2 Implications of ATP

Two central issues in the debate about the design of the occupational pension system emerged over the next decades. The first issue was how to coordinate the occupational pension plans with the reformed public pension system, the ATP plan in particular. The second issue was the extension of occupational pension rights to blue-collar workers in the private sector.

There are two ways in which second-pillar pension plans are typically coordinated with the first pillar. The occupational pension can either be determined jointly with the public pension or paid out as a separate entity. These are referred to as the gross and net methods, respectively. The main advantages of the gross method are that it effectively achieves a certain replacement rate and that it is easy to understand for the plan participants. Net pensions, however, are advantageous from the plan administrator's perspective because they are easier to administrate and entail lower uncertainty surrounding future funding of pension obligations.

The central government plan, which was reformed in conjunction with the legislation of the ATP plan in 1959, applied the gross method.[37] Targeting the previous replacement rate of 65 percent, the occupational pension benefit was calculated as the difference between 65 percent of the final wage and the public pension benefit. The pension plan for private sector white-collar workers was also reformed in conjunction with the ATP reform. Benefits from this pension plan, referred to as ITP[38], were paid out according to the net method. This meant that individuals earned occupational pension rights for the salary above the income ceiling in the public pension system. ITP was based on a premium reserve system, which implied that the pension benefit was paid out from invested assets rather than current contributions as under a PAYG plan.

[37] The reformed pension plan was called SPR (Statens allmänna tjänstepensionsreglemente) and was in place for more than 30 years.

[38] Industrins tilläggspension för tjänstemän

In fact, ITP was the only occupational pension plan based on a premium reserve system until the 1990s.

Blue-collar workers still had less generous benefits than white-collar employees in the private sector and public sector employees after the ATP reform in 1960. They both had a lower total pension, since they lacked occupational pension coverage enjoyed by other groups, and they had no special arrangement that enabled them to exit the labor force at age 65 without having to claim an actuarially reduced public pension. In 1973, an agreement was reached between the central confederation of blue-collar workers, LO, and the Swedish Employers Association (SAF), to provide a supplement to the public system for blue-collar workers, the special supplementary pension (STP)[39]. The overall goal with the STP plan was to even out the replacement rate difference between white-collar and blue-collar workers by paying out benefits that corresponded to 10 percent of the final wage. In contrast to the other occupational pension plans, STP did not pay out pension benefits for earnings above the income ceiling.

6.3 Problems with the occupational pension plans

When plans for abolishing the ATP plan were drawn up in the early 1990s, a reconstruction of the occupational pension system became necessary. The occupational pension plans would have to be coordinated with the new rules of the public pension system. Calls for reform also intensified when the problems caused by some features of the occupational pension plans became more apparent. The occupational pension plans suffered from many of the problems that the ATP plan did because they had many institutional features in common.

First of all, the long-run financial stability of some of the pension plans was threatened by underfunding. Most financially troubled was the pension plan for local government employees.[40] A large share of the local government work force that had been recruited during the expansion of the public sector during the 1970s and 1980s now approached retirement. Increasing labor supply and rising labor force participation among women, who made up as much as 80 percent of the local government work force, also increased the pension costs of local governments. The municipalities were unable to alleviate the pressure by shifting pension means across municipalities, which piled up considerable pension debts that had to be financed by current budget means (Ståhlberg, 1993). Moreover, with a larger share of workers hitting the income ceiling, the share of the financial responsibility for providing pensions was shifted to the occupational pension plans. Underfunding was not a prob-

[39] Särskild tilläggspension

[40] Local government workers were covered by the central government pension plan, SPR, until the introduction of a local government-specific pension plan in 1985, called PA-KL.

lem in all plans, least of all in the fully funded plan for privately employed white-collar workers. Neither the central government pension plan suffered from particular underfunding issues, much thanks to the possibility of shifting pension means between different central government agencies.

Institutional complexity was another recognized problem. The complexity of the plans could be utilized by employers who were better informed than their employees. There are two examples of this. First, common for all major occupational pension plans was that persons who left the labor force before the normal retirement age were granted an annuity from age 65. This meant, however, that accumulated rights were not indexed during the interim. With high rates of inflation, early exit from the labor market normally resulted in a lower pension. This was generally not well understood by participants, but by employers, who could reduce the pension costs associated with a specific employee by making him/her choose this option (Palmer and Wadensjö, 2004). Second, the occupational pension plans contained a lower limit for the minimum number of working hours required to qualify for a pension. As a result, part-time workers, particularly younger employees and mothers with children, risked not acquiring pension rights, of which they were not always aware. Employers could utilize this feature of the system to save on overall pension costs.

The occupational pension plans were also criticized for distorting labor supply. Because the benefit level was determined by the wage during the most recent years before retirement, individuals were incentivized to work many hours just before retirement and less so during early years. Many people tried to find more favorable ways out of the labor force than simply reducing the number of hours worked. Of these, early retirement through the disability insurance program was the most attractive. The occupational pension that would be paid out from age 65, after having received the disability pension for several years, was calculated on basis of the income during the years preceding the granting of disability pension.

The progressiveness of pension premiums for employers with respect to age and wage was (and is still) a matter of debate. In a DB plan, the premium paid by the employer generally increases with the age of the employee. A wage increase in the years immediately preceding retirement typically has a much larger effect on the premium rate than a wage increase of the same magnitude earlier in the working career, since the increase in the defined benefit must be financed during a shorter time (Pensionsåldersutredningen, 2012). This can make it difficult for older persons to change jobs.

Finally, the distributional effects of the occupational pension plans were similar to those of the ATP plan, but of greater magnitude. The size of the occupational pension benefit was based on an average of the last several years' wages rather than on the best 15 years as in ATP. This implied that shifting from part-time work to full-time work at the end of the working career was even more beneficial within the occupational pension plans than under the ATP plan. Short careers and professions with a steep earnings-profile benefited ad-

ditionally from the fact that the income ceiling within all but the occupational pension plan for blue-collar workers was four times as high as in the public pension system.

6.4 Occupational pensions today

All major occupational pension plans have undergone significant changes during the last two decades. These were implemented to overcome the problems described in the previous section. Moreover, the structure of the new public pension system that was put in place in the late 1990s necessitated changes in the occupational pension plans to maintain previous retirement standards and to adjust coordination technicalities. Most importantly, all plans have changed from DB to DC or a mixture of the two. Each plan also includes a fully funded DC component, clearly inspired by the FDC component in the public pension system. Although the direction of reform has been the same across sectors, the timing has been different. Some of the plans were reformed early in the 1990s, whereas some underwent major changes 10–15 years later.

The central government pension plan has undergone two major changes during the last twenty years. The first reform in 1991 implied a shift from the gross to the net pension concept while keeping the DB structure. The structure of the new pension plan was a direct copy of the occupational pension plan for white-collar workers which had existed since 1960, providing higher replacement rates above the income ceiling than below. From 2003, the DB component only accrues to earnings above the income ceiling, while earnings below are DC. The current pension plan for local government employees was put in place in 2006 and closely resembles the central government pension plan.

In contrast to the current public sector pension plans, the new private sector pension plans do not contain any DB components. In 1996, the pension plan for blue-collar workers, STP, was replaced by a new DC pension plan called SAF-LO. Premium payments take place through the employer setting aside 4.5 percent of gross income up to the income ceiling. For wage portions above the ceiling, the premium is set at 30 percent. The new pension plan for white-collar workers is similar in construction to SAF-LO. Implemented in 2006, the ITP plan was last among the four major agreements to switch from DB to DC.

Around 90 percent of the total work force today are covered by the four major occupational pension plans (Pensionsåldersutredningen, 2012). Public sector employees are covered by law, which means that the remaining 10 percent work in the private sector. Because the occupational pension plans provide pension benefits above the income ceiling in the public pension system, the relative importance of the occupational pension typically rises with income. For individuals with earnings below the income ceiling, the occupational pen-

Table 4. *Occupational, public and private pension as a share of total pension income for individuals between 65 and 69 years of age*

Year	Cohorts aged 65-69	Public pension		Occupational pension		Private pension	
		Men	Women	Men	Women	Men	Women
1996	1927–1931	74.4	80.6	20.3	15.6	5.3	3.8
2002	1933–1937	67.9	76.2	24.2	15.6	5.3	3.8
2006	1937–1941	64.0	72.1	27.7	19.0	8.1	8.9
2007	1938–1942	62.2	70.3	29.4	20.3	8.4	9.4
2009	1940–1944	54.5	62.0	29.6	21.6	15.9	16.4
2010	1941–1945	53.3	60.7	31.2	23.0	15.5	16.3

Source: Pensionsåldersutredningen (2012)

sion amounts to 10–20 percent of total gross pension. For individuals with earnings above the ceiling, the occupational pension can account for as much as half of total pension income.

Table 4 clearly shows that the relative importance of the occupational pension has increased over time. For men born in the 1940s, one-third of total pension income is occupational pension income, which is 11 percentage points higher than for men born in the early 30s. The corresponding increase for women is 7 percent. One explanation for this trend is that more people have income that exceeds the income ceiling in the public pension system. Another explanation is the increase in female labor force participation. Moreover, all four major occupational pension plans have been converted, or are under way of being converted, from DB to DC. If the rate of return in fully funded systems exceeds the rate of return in the NDC component in the public system, which is indexed to the wage growth, the relative importance of occupational pensions will continue to increase in the future. Flood (2004) projects that the relative importance of the occupational income continues to increase in the future across sectors for both genders.[41]

Two issues with the design of the current occupational pension plans stand out in the debate. The first issue is how the occupational pensions affect the actual retirement age. Increasing the actual retirement age is a central policy objective around the world, including Sweden. The reformation of the public pension system in the 1990s and the years after that brought many changes to age-related regulations in order to increase labor supply among elderly.[42] There is a substantial risk that the rules and regulations of the occupational

[41] Table 4 also shows that private retirement savings have become more important as a source of pension income over the last 15 years. Private retirement savings are usually referred to as the third major source of pension income next to the public pension and the occupational pension. Savings in private retirement accounts are tax deductible. The tax deduction, however, disappeared in 2016.

[42] For example, the lowest age for early withdrawal was raised from 60 to 61. The lowest age for withdrawal of the minimum guarantee was raised from 60 to 65. Since 2001, employees have the right to remain employed until the age of 67 (prop. 2000/01:78).

pension plans might counteract the enhanced financial advantages of postponing retirement in the reformed public pension system (Pensionsåldersutredningen, 2012).

Even though most current occupational pension plans do not contain a formal retirement age, they presuppose a "normal" retirement age of 65. If no action is taken, the pension is paid out at the time the individual turns 65. Since income earned after the age of 65 does not matter for the size of the pension, the individual faces a pension accrual discontinuity at 65 and is highly disincentivized to work after this age. This "65-norm" is built in even in the DC components of the new occupational pension plans. Typically, pension rights can be earned after 65 only under special agreements between the employee and the employer (Pensionsåldersutredningen, 2012).

The second issue concerns the decumulation phase of retirement. In the last decade, all major occupational pension plans have introduced fixed-term payouts as an option to annuitization. The minimum time period over which the pension wealth can be withdrawn is five years.[43] Fixed-term payouts increase flexibility, as they allow individuals to invest, leave bequests or increase consumption during the early years of retirement, but raise concerns that they might trigger individuals to spend the money too rapidly for their own good. Another concern is that the consequences of individuals outliving their resources will fall on the rest of society. When the occupational pension payment becomes smaller or even zero, individuals may become eligible for means-tested benefits such as the housing supplement for elderly and the supplementary benefit that targets individuals who have no or very low pension income. Moreover, fixed-term payouts could potentially have a negative effect on the actual retirement age. In combination with the possibility of early withdrawal, time-limited payouts allow individuals to retire early and finance the first years out of the labor force by income from the occupational pension system.

7 Concluding remarks

This paper has provided an overview of the history of the Swedish pension system. It has discussed the political and economic background of each major public pension reform in the last 100 years as well as the parallel development of the second pillar.

This final section takes a more general, forward-looking perspective. It discusses three dimensions of pension design that have played an important

[43] In a recent paper, Hagen (2015) studies the payout choices of 183,000 white-collar workers who retired under the ITP plan between 2008–2013. He finds that the popularity of the fixed-term payouts has risen over time with 20 percent opting for any of these payout options in 2008 compared to 31 percent in 2013.

role in the development of the Swedish pension system, and the circumstances under which each of these has been more or less successful. Identifying circumstances under which pension reform has proven successful might yield important lessons for future policy.

The first dimension of pension design relates to the link between contributions and future benefits. The strength of this link has been central in all major pension reforms and the dimension along which ideological differences between the two party blocs have been most evident. Typically, liberal and right-wing parties have advocated a stronger link between benefits and contributions than socialist parties. They argue that a pension system characterized by a strong link, a so-called *Bismarckian* pension system, minimizes labor market distortions and sustains the standard of living acquired during work life into retirement for all workers. Socialists have favored a weaker link to increase the scope for redistribution and poverty prevention. Pension systems that aim at these things are said to have a strong *Beveridgean* character.

The Swedish pension history illustrates the importance of finding a good balance along this dimension. The predominantly Bismarckian system of 1913 failed to provide poverty relief and was soon replaced by a highly redistributive, flat-rate system based on the Beveridgean principle of basic security. In this system, however, there was no link between the standard of living acquired during work life and the standard of living in retirement, especially for workers who were not covered by a supplementary occupational pension plan. The implementation of the ATP plan in 1960 was the first serious attempt to provide both poverty relief and earnings-related benefits.

The ATP plan indeed raised the living standard of the elderly, but failed to provide a clear link between contributions and benefits. The fact that benefits were based only on a subset of years perverted the redistributive properties of the ATP plan. Instead of redistributing resources from high- to low-income individuals, or from men to women, the ATP plan redistributed resources from individuals with long working lives and a flat life-cycle income profile (mostly low-income workers) to those with shorter work histories and rising earnings profiles (mostly high-income workers). These features were regarded as unfair and eroded support for the ATP plan. Basing benefits on a subset of years also contributed to reduced labor market mobility close to retirement (lock-in effect) and weakened incentives to work additional hours at younger ages. On account of these properties, among others, the ATP plan was subsequently abolished in the late 1990s.

The fate of the ATP plan illustrates the importance of considering the whole range of effects of choosing a specific treatment of earnings in different years in determining the pension benefit. Basing benefits on a limited number of years strengthens the insurance character of the pension system as workers are allowed to exclude years with low or no earnings, but may also weaken its progressive properties and distort labor market decisions. To increase transparency around these issues, the current pension system bases benefits on con-

tributions in all years.

The second dimension relates to the choice of financing rules. A pension system is either funded or unfunded (or a mixture of the two). The Swedish case clearly illustrates the tradeoffs faced by policy makers in choosing the appropriate funding structure. A common argument in favor of funding is that funded systems always have sufficient reserves to pay all outstanding financial liabilities and hence ensure fiscal sustainability. In fact, the main motivation for having a large funded component in the public pension system of 1913 was to avoid too much pressure on the state budget. Many considered the state economy too weak to uphold a tax revenue-based pension system. However, funded systems may fail to produce adequate pensions for the oldest cohorts when a system starts or expands since they do not generally redistribute resources between generations. The subsequent implementation of the universal flat-rate basic pension in 1935 (extended in 1948) was motivated by the proven inability of the previous system to generate sufficient pension levels. Thus, support for unfunded alternatives may increase if funded pension systems fail to provide reasonable pension levels. The implementation of a funded component is therefore more likely to be successful if there exists an unfunded component that takes proper account of the financial situation of the current old.[44]

The issue of insufficient pensions relates more broadly to the overall objective of providing economic security in old age. A pension system should reduce poverty among the elderly and provide adequate retirement income. A pension system that fails to fulfil this basic task is unlikely to be sustained in the long run. In fact, inadequate attention to poverty relief in conjunction with too much focus on fiscal sustainability has been pointed out as a common policy error by the World Bank (Andrews, 2006). As for Sweden today, the concern is not so much about poverty relief among the elderly as about the pension system's ability to provide adequate replacement rates for retiring cohorts in the future. Future retirees must work longer and get a high rate of return on their funded pension assets in order to achieve similar replacement rates as today's retirees. Moreover, adjusting benefits to rising life expectancy at the age of withdrawal places heave reliance on rational behavior. If people continue to retire at broadly the same age as at present, benefits will over time become less adequate (Barr, 2013).[45] A gradual increase in the earliest eligi-

[44] A successful move towards more funding also requires that proper consideration is taken to the added burden of the younger workers who must not only pay their own contributions but also finance the outgoing pensions. The 1998 reform illustrates how well-designed transition rules can help overcome the political and financial challenges associated with moving from an unfunded system to a system based on individual contributions.

[45] Pensionsmyndigheten (2014) shows that the benefit reductions caused by increasing life expectancy are substantial. For example, the 1965 cohort must work until the age of 67 years and 9 months to receive the same pension level as those born in 1930 get at age 65. Thus, the fact that the average claiming age in the public pension system has been relatively stable dur-

bility age would raise the retirement age and assist sustainability.

The third and last dimension relates to how the size of the benefit is determined, that is, whether the pension system is defined benefit (DB) or defined contribution (DC). The development of the benefit structure of the public pension system can be divided into three parts: it started out as predominantly DC in 1913, gradually turned into DB with the introduction of the basic pension and the ATP plan, and became again (N)DC in the late 1990s. It is interesting to note that the arguments put forward in policy discussions in favor and against each benefit structure have been fairly similar across time. Arguments in favor of DB pensions center on the importance of providing a significant and immediate impact on the standard of living of the elderly. Advocates of DC pension plans, on the other hand, emphasize that they imply less risk for the central government (or the plan administrator) than DB pensions do. DC systems make all the adjustment to financial realizations on the side of benefits rather than contributions and hence put all the risk on plan participants. For example, a DC dominated regime was preferred in the early 20th century as the state budget was considered too weak to support universal flat-rate pension benefits or comprehensive means-tested benefits. A move towards DC was also warranted during the 1980s when the projected costs of the ATP plan soared due to a period of sluggish growth, amplified by an increase in the old-age dependency ratio.

These experiences suggest that the long-run political support for DC systems will primarily depend on two factors: their ability to generate sufficient pension levels and to preserve financial stability. Today, while the level of future pensions is being disputed, there are good conditions for long-run financial stability. The automatic balancing mechanism prevents the NDC system from running a systematic deficit by adjusting benefits so that the ratio of the system's assets to its liabilities never falls below a certain level. However, the legitimacy of the pension system may be undermined if pension payments become too volatile or unpredictable. The most important sources of financial imbalance that could contribute to the activation of the automatic balancing mechanism are the indexation of benefits to per capita wage growth and the use of actual rather than projected mortality tables.

ing the last decade (64.8 in 2014) raises concern about the level of future pensions. However, even if people do not seem to delay claiming, they do work longer. Conditional on being in the labor force at age 50, the average exit age rose from 63.1 to 63.8 between 2005 and 2014 (Pensionsmyndigheten, 2015).

References

Andrews, E. S. (2006). *Pension reform and the development of pension systems: An evaluation of World Bank assistance*. Washington, DC: World Bank Publications.

Auerbach, A. J. and R. Lee (2009). Notional defined contribution pension systems in a stochastic context: Design and stability. In Brown, J., Liebman, J., Wise, D. (Eds.), *Social Security Policy in a Changing Environment*, pp. 43–68. University of Chicago Press.

Barr, N. (2013). The pension system in Sweden. *Report to the Expert Group on Public Economics at the Swedish Ministry of Finance*.

Beveridge, W. (1942). The Beveridge Report. *His Majesty's Government: London*.

Bonoli, G. (2003). Two worlds of pension reform in Western Europe. *Comparative Politics 35*(4), 399–416.

Börsch-Supan, A. and C. B. Wilke (2004). The German public pension system: How it was, how it will be. *NBER Working Paper 10525*.

Clark, R., L. Craig, and J. Wilson (2003). *A history of public sector pensions in the United States*. University of Pennsylvania Press.

De Gosson de Varennes, Y. (2016). *Benefit design, retirement decisions and welfare within and across generations in defined contribution pension schemes*. Ph.D. thesis, Economic Studies 157, Department of Economics, Uppsala University.

Diamond, P. (1999). Social security reform with a focus on Sweden. *The Lindahl Lectures, Uppsala University*.

Edebalk, P. G. (1999). Beveridgeplanen – en klassiker. *Socialvetenskaplig tidskrift 1*, 159–171.

Edebalk, P. G. (2003a). Bismarck och de första socialförskringarna. *Socialvetenskaplig tidskrift 4*, 352–365.

Edebalk, P. G. (2003b). Folkpension och åldringsvård – om svensk socialpolitik 1903-1950. *Socialvetenskaplig tidskrift 2-3*, 131–150.

Edebalk, P. G. and M. Olsson (2010). Poor relief, taxes and the first universal pension reform: The origin of the Swedish welfare state reconsidered. *Scandinavian Journal of History 35*(4), 391–402.

Elmér, . (1960). *Folkpensioneringen i Sverige, med särskild hänsyn till ålderspensioneringen*. Lund: CWK Gleerup.

Engström, S. and A. Westerberg (2003). Which individuals make active investment decisions in the new Swedish pension system? *Journal of Pension Economics and Finance 2*(03), 225–245.

Feldstein, M. and J. B. Liebman (2002). Social security. *Handbook of Public Economics 4*, 2245–2324.

Flood, L. (2004). Vilka pensioner får framtidens pensionärer? *Ekonomisk debatt 3*, 16–30.

Hagen, J. (2013). A history of the Swedish pension system. *Working Paper 2013:7,* Department of Economics, Uppsala University.

Hagen, J. (2015). The determinants of annuitization: evidence from Sweden. *International Tax and Public Finance 22*(4), 549–578.

Heclo, H. (1974). *Modern social politics in Britain and Sweden: From relief to income maintenance*. Yale University Press New Haven.

Hermansson, J. (1993). *Politik som intressekamp: parlamentariskt beslutsfattande och organiserade intressen i Sverige*. Norstedts juridik.

Könberg, B. (2008). Först i världen med nytt pensionssystem – hur gick det till? *Frisinnad tidskrift 2*, 36–48.

Könberg, B., E. Palmer, and A. Sundén (2006). The NDC reform in Sweden: The 1994 legislation to the present. In Holzmann, R., Palmer, E. (Eds.), *Pension Reform: Issues and Prospects for Non-Financial Defined Contribution (NDC) Schemes*, pp. 449–466. Washington DC: The World Bank.

Kruse, A. (2003). Svenska pensionsreformer under 1900-talet. *25 år i täten – en vänbok till Gunnar Edebalk*, 83–98.

Kruse, A. and A. C. Ståhlberg (1977). *Effekter av ATP: En samhällsekonomisk studie*. Breviks Tryckeri.

Lindbeck, A. (1992). *Klarar vi pensionerna?* (1 ed.). Stockholm: SNS.

Lindbeck, A. and M. Persson (2003). The gains from pension reform. *Journal of Economic Literature 41*(1), 74–112.

Lindbom, A. (2001). De borgerliga partierna och pensionsreformen. In J. Palme (Ed.), *Hur blev den stora kompromissen möjlig*, pp. 50–87. Stockholm: Pensionsforum.

Lundberg, U. (2001). Socialdemokratin och 1990-talets pensionsreform. In J. Palme (Ed.), *Hur blev den stora kompromissen möjlig*, pp. 8–49. Stockholm: Pensionsforum.

Lundberg, U. (2003). *Juvelen i kronan: Socialdemokraterna och den allmänna pensionen*. Stockholm: Hjalmarson & Högberg.

Marier, P. (2002). *Institutional structure and policy change: Pension reforms in Belgium, France, Sweden and the United Kingdom*. Ph.D. thesis, University of Pittsburgh.

Ottander, C. and C. Holmqvist (2003). *En fyrk per daler: Pensioner i ett historiskt perspektiv*. Stockholm: Page One Publishing AB.

Palme, M., A. Sundén, and P. Söderlind (2007). How do individual accounts work in the Swedish pension system? *Journal of the European Economic Association 5*(2-3), 636–646.

Palmer, E. (2002). Swedish pension reform: How did it evolve, and what does it mean for the future? In Feldstein, M., Siebert, H. (Eds.)*, Social Security pension reform in Europe*, pp. 171–210. University of Chicago Press.

Palmer, E. and E. Wadensjö (2004). Public pension reform and contractual agreements in Sweden – From defined benefit to defined contribution. In Rein, M., Schmähl, W. (Eds.), *Rethinking the Welfare State*, pp. 226–248. Cheltenham: Edward Elgar.

Pensionsarbetsgruppen (1992). *Ett reformerat pensionssystem: Bakgrund, principer och skiss*. Stockholm: Allmänna förlaget.

Pensionsåldersutredningen (2012). *Underlagsrapport till Pensionsåldersutredningen: Tjänstepensioner och utträde från arbetslivet*. Stockholm: Elanders Sverige AB.

Pensionsmyndigheten (2012a). Orange rapport: Pensionssystemets årsredovisning 2011.
Pensionsmyndigheten (2012b). Premiepensionen: Pensionsspararna och pensionärerna 2011.
Pensionsmyndigheten (2014). Orange rapport: Pensionssystemets årsredovisning 2014.
Pensionsmyndigheten (2015). Medelpensioneringsålder och utträdesålder m.m. *Rapport VER 2015-2*.
Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review 94*(02), 251–267.
Proposition (1959:100). Med förslag till lag om försäkring för allmän tilläggspension m.m. Stockholm: Riksdagen.
Proposition (1987/88:171). Om reformering av den allmänna försäkringens efterlevandeförmåner m.m. Stockholm: Riksdagen.
Proposition (1997/98:151). Inkomstgrundad ålderspension, m.m. Stockholm: Riksdagen.
Proposition (1997/98:152). Garantipension m.m. Stockholm: Riksdagen.
Proposition (2000/01:70). Automatisk balansering av ålderspensionssystemet. Stockholm: Riksdagen.
Proposition (2000/01:78). Rätt arbeta till 67 års ålder. Stockholm: Riksdagen.
Queisser, M. and E. Whitehouse (2005). Pensions at a glance: Public policies across OECD countries. *MPRA Paper No. 10907*.
Schmidt, F. (1974). *Allmänna och privata pensioner: Mål och medel*. Norstedts.
Socialstyrelsen (1931). Lönestatistisk årsbok för Sverige.
SOU (1994:20). *Reformerat pensionssystem: Betänkande*. Stockholm: Fritzes offentliga publikationer.
SOU (1998:3). *Sammanställning av remissyttranden över betänkanden som legat till grund för propositionerna om reformerat ålderspensionssystemet.* Stockholm: Fritzes offentliga publikationer.
SOU (2013:25). *Slutbetänkande av Pensionsåldersutredningen: åtgärder för ett längre arbetsliv*. Stockholm.
Ståhlberg, A. (1989). Problemen med ATP. *Ekonomisk debatt 5*, 351–361.
Ståhlberg, A.-C. (1993). *Våra pensionssystem* (1 ed.). Stockholm: SNS.
Sundén, A. (2006). The Swedish experience with pension reform. *Oxford Review of Economic Policy 22*(1), 133–148.
Säve-Söderbergh, J. (2012). Self-directed pensions: Gender, risk, and portfolio choices. *The Scandinavian Journal of Economics*.

Appendix

A Concepts and definitions

- **Basic pension** – A flat-rate state pension paid to all who meet the minimum contribution requirement. A universal basic pension was introduced in Sweden in 1935 and has remained an important feature, although in different versions, of the Swedish pension system ever since.
- **Beveridgean pension system** – Public pension arrangement based on means-tested or universal flat-rate benefits. The Beveridgean pension model stems from the Beveridge plan (Beveridge, 1942) presented in the UK in 1942. It is often referred to as the most influencing social policy reform proposal of all time. The plan states that social insurance systems should be universal and mandatory and guarantee existential minimum. Benefits are financed by flat-rate contributions and consist of simple cash transfers. Since there is no, and has never been, a pension system designed completely along the lines of the Beveridge plan, the pension systems we refer to as *Beveridgean* exhibit much variation. Benefits can be financed by tax revenue and Beveridgean components may co-exist with Bismarckian components within the same pension system and so forth.
- **Bismarckian pension system** – Public pension arrangement based on earnings-related social insurance, typically financed by wage-based contributions. There is a close relationship between benefits and contributions, which is why pensions are referred to as *retirement insurance*. There is little redistribution and benefits are seldom universal. The term *Bismarckian* refers to the German Chancellor, Otto von Bismarck, who implemented the first formal pension system in the world in the late 19th century.
- **Collective agreement** – An agreement between employers and employees which regulates the terms and conditions of employees in their workplace, their duties and the duties of the employer. There are presently four large agreement-based occupational pension systems in Sweden, covering privately employed blue-collar workers, privately employed white-collar workers, central government employees and local government employees respectively.
- **Contribution rate** – The amount of money that is contributed (monthly) to a specific pension plan by law. Sometimes referred to as the premium fee.

- **Defined contribution (DC)** – In a DC pension plan, individual accounts are set up for participants and benefits are based on the amounts credited to these accounts. In the pension literature, DC plans are therefore referred to as "individual account plans". A DC plan can either be financial defined contribution (FDC) or notional defined contribution (NDC). Individual account balances grow with annual contributions and the rate of return on the account. The rate of return depends on whether the plan is NDC or FDC.
- **Defined benefit (DB)** – In a DB pension plan, the state or the employer promises a specified monthly benefit on retirement that is predetermined by a formula based on the employee's earnings history, tenure of service and age. It is the converse of a defined contribution plan, where the pension benefit is determined by investment returns or the accumulated amount of contributions.
- **Flat-rate benefits** – These are benefits that are related only to age and citizenship, not past earnings and contributions. They usually have an anti-poverty objective and are used to ensure everybody with a certain minimum standard of living. They are either financed by tax revenue or by contributions. The main advantage of universal flat-rate benefits is that they effectively can prevent poverty in old age with relatively little direct effect on saving incentives. However, they entail large costs for the state.
- **Financial defined contribution (FDC)** – An FDC plan works as a DC plan, where contributions to individual accounts are invested in market assets. The final benefit thus depends on the contribution plus the investment's return.
- **Full funding** – In a fully funded pension plan, current contributions are set aside and invested in order to finance the future pensions of current contributors. Many company plans are fully funded. Public pay-as-you-go pensions may be partially pre-funded when the government raises the contribution rate above what is necessary to finance current benefits, in order to accumulate a fund to help pay future benefits. The designated pension fund(s) is (are) sometimes referred to as a *premium reserve system*.
- **Gross occupational pension** – Gross pension plans are coordinated with the public pension system to guarantee the individual a certain *total* pension level (see *net* occupational pension for its converse).
- **Life-income principle** – The life-income principle implies that an individual should earn pension rights on all earnings, and not only on specific types of income or on income earned during a limited number of years.
- **Loss-of-earnings principle** – The insurance compensation should be based on the income of the insured. In other words, accumulated pension rights should be directly linked to previous earnings. This principle was

at the core of the supplementary pension plan, ATP, in which benefits were regarded as "deferred earnings" rather than a handout.

- **Indexation** – A system whereby pensions are automatically increased at regular intervals by reference to a specific index of prices or earnings.
- **Income ceiling** – The public pension system contains a ceiling on the income qualifying for pension rights. The ceiling is currently at 7.5 income base amounts. For 2016, this means that no pension rights are earned for the monthly wage portion that exceeds SEK 37,062. Supplementary occupational pensions typically provide pension benefits for income over the ceiling.
- **Means-tested benefits** – Benefits that are paid only if the recipient's income falls below a certain level. Means-tested benefits effectively target the poor and can potentially alleviate old-age poverty at a smaller cost than universal flat-rate benefits. However, a large bureaucratic apparatus is required to manage benefit applications that are subject to means-testing. Means-tested benefits may also create incentives for some individuals to intentionally undersave or underreport earned income during the working years in order to claim benefits they are in fact not eligible for.
- **Net occupational pension** – Net pension plans provide benefits that "float on top" of the public pension. They contain no direct coordination with the public pension system.
- **Notional defined contribution (NDC)** – An NDC plan works as a DC plan, where contributions to individual accounts are recorded but not invested in market assets. NDC plans are PAYG, where annual contributions finance current pension benefit obligations. The rate of return in NDC plans differ according to the indexation choice of the policy maker. The rate of return in the Swedish NDC plan, the inkomstpension, is determined by the per capita wage growth.
- **Occupational pension** – Access to occupational pension plans is linked to an employment or professional relationship between the plan member and the entity that establishes the plan (the plan sponsor). Occupational pension plans may be established by employers or groups thereof and labor or professional associations, jointly or separately. The plan may be administered directly by the plan sponsor or by an independent entity. In the latter case, the plan sponsor may still have oversight responsibilities over the operation of the plan. These are often regarded and designed as supplementary to the public pension system. In the Swedish case, participation is mandatory for employers who are part of some kind of collective agreement. Employers must set up (and make contributions to) occupational pension plans which employees will be required to join.
- **Pay-as-you-go (PAYG)** – An arrangement under which benefits are paid out of current revenues and no funding is made for future liabilities. PAYG-systems are therefore unfunded.

- **Pension plan**[46] – A legally binding contract with an explicit retirement objective. This contract may be part of a broader employment contract, it may be set forth in the plan rules or documents, or it may be required by law. In addition to having an explicit retirement objective, pension plans may offer additional benefits, such as disability, sickness, and survivors' benefits.
- **Pensionable income** – Income measure on which contributions to a certain pension plan is paid. Pensionable income in the Swedish public pension system includes wages as well as payments from social security and unemployment insurance systems.
- **Premium reserve system** – System for creating a premium reserve used in different kinds of insurance contexts. Most occupational pension plans make use of a premium reserve system, in which individuals contribute repeatedly, as pension rights are earned, to an actuarial liability that should guarantee the pension obligations. Premium reserve systems are also referred to as *fully funded* pension systems.
- **Public pension system** – Refers to the pension system that is administered by the government.
- **Replacement rate** – The ratio of an individual's (or a given population's) (average) pension in a given time period and the (average) income in a given time period. The replacement rate reflects the relative generosity of a pension plan.
- **Social security** – Also referred to as social insurance, where people receive benefits or services in recognition of contributions to an insurance program. These services typically include provision for retirement pensions, disability insurance, survivor benefits and unemployment insurance. Should not be mixed with the term's meaning in the United States, where social security refers to a specific social insurance program for the retired and disabled.

[46] In this report, the term *pension system* refers to *a set* of pension plans administered by a specific public or private actor, and therefore has a broader meaning than the term *pension plan*.

B Time lines

Table A.1. *Important events in the history of the public pension system*

Year	Event
1913	Public pension system legislated
1935	Universal basic pension, folkpension
1937	Regional heterogeneity in pension benefits – predecessor to housing supplements
1946	Big increase in pension generosity – increased basic pension
1959	ATP legislated
1969	Special supplement introduced
1976	Retirement age lowered from 67 to 65
1976	Partial pension introduced
1991	More restrictive disability pension
1994	General proposal of new pension system passed by parliament
2001	First pension payments from new system
2001	Right to work until the age of 67
2010	Automatic balance mechanism activated for the first time

Table A.2. *Important events in the history of the occupational pension plans*

Year	Event
1907	Pension act for permanent central government employees
1917	First major private sector pension plan (SPP)
1934	Non-permanenent civil servants covered by state pension plan
1959	New gross pension plan (SPR) for central government employees introduced
1960	ITP plan for white-collar workers introduced, premium reserve system
1973	First occupational pension plan (STP) for blue-collar workers
1991	New net pension plan (PA-91) for central government employees
1996	First occupational pension plan with no DB component – SAF-LO for for blue-collar workers
1998	DC plan (PFA-98) for local government employees; DB above income ceiling)
2003	DC (PA-03) for central government employees; DB above income ceiling
2006	New plan for white-collar workers (ITP1); no DB component

Table A.3. *Major public pension system reforms*

Year	Commissions	Political color	Main components	Type	Contribution rate	Replacement rate	Financing rules	Indexation
1913	Old-Age Insurance Commission, 1907	Conservative government Unanimity	1. Avgiftspension 2. Pensionstillägg	DC Means-test, DB	At most 1.1%, ind. contribution	30%/24% of total contributions, men/women SEK 150/140, men/women	Premium res. Tax/PAYG	
1935	Pension Insurance Commission, 1928	Social Dem. government Unanimity	1. Allmän folkpension 2. Pensionstillägg	Universal flat-rate benefit	1%, inc. tax	SEK 100+ 10% of total contr.	Tax/PAYG/ premium res.	
1946 (1948)	Social Welfare Commission, 1937	Social Dem. government Unanimity	1. Allmän folkpension 2. Means-tested housing supplement	Universal flat-rate benefit	1% (1946) - 4% (1959), inc. tax	SEK 1000	Tax/PAYG	Prices (1950)
1959 (1960)	Åkesson Commission, 1947 Public pension commission, 1956	Social Dem. government Referendum One vote majority	1. Folkpension 2. ATP	Universal flat-rate benefit Suppl. earnings-related, DB	4% (1959) - 5.86% (1994), inc. tax 3% – 11 % Payroll fee	90%/70% of 1 BA unmarried/married 60% of wage during 15 best years	Tax PAYG	Prices (1968) Prices
1994 (1998)	The Pension Commission, 1984 Working Group on Pensions, 1991 Implementation Group, 1994	Right-wing government Compromise Unanimity Implemented under Social Dem. government	1. Garantipension 2. Inkomstpension 3. Premiepension	Universal minimum guarantee NDC FDC	7%, general contr., deductible State old-age contribution 10.21%, payroll fee	2.13/1.9 BA, singles/couples	Tax PAYG Fully funded	Prices Real wage growth Return on investments

II. The determinants of annuitization: evidence from Sweden

Acknowledgments: This paper has benefited from comments by Sören Blomquist, Håkan Selin, James Poterba, Jeffrey Brown, Jonathan Reuter, Paolo Sodini, Mårten Palme, Svetlana Paschenko, Tarmo Valkonen, Raymond Gradus, Jacob Lundberg, Ponpoje Porapakkarm, Yuwei De Gosson de Varennes, Maria Polyakova, Jean-Marie Lozachmeur, Matthew Zaragoza-Watkins, Mikael Elinder, Per Johansson, Per Engström, Hannes Malmberg and seminar participants at the Swedish Social Insurance Inspectorate (ISF), the 2014 IIPF Congress, Uppsala Brown Bag, and the 2014 Topics in Public Economics in Uppsala. I also thank two anonymous reviewers and guest editor Monika Bütler. Special thanks to Alecta for providing data and to Pär Ola Grane, Nikolaj Lunin and Lars Callert for very helpful discussions. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

Int Tax Public Finance (2015) 22:549–578
DOI 10.1007/s10797-015-9360-5

The determinants of annuitization: evidence from Sweden

Johannes Hagen[1]

Published online: 19 June 2015

Abstract Despite the fact that the increasing involvement of the private sector in pension provision has brought more flexibility to the pay-out phase of retirement, little is known about the characteristics of those who choose to annuitize their pension wealth and those who do not. I combine unique micro-data from a large Swedish occupational pension plan with rich national administrative data to study the choice between life annuities and fixed-term payouts with a minimum payout length of 5 years for 183,000 retiring white-collar workers. I find that low accumulation of assets is strongly associated with the choice of the 5-year payout. Consistent with individuals selecting payout length based on private information about their mortality prospects, individuals who choose the 5-year payout are in worse health, exhibit higher ex-post mortality rates and have shorter-lived parents than annuitants. Individuals also seem to respond to large, tax-induced changes in annuity prices.

Keywords Annuity puzzle · Longevity insurance · Occupational pension · Adverse selection · Administrative data

JEL Classification D91 · H55 · J26 · J32

Electronic supplementary material The online version of this article (doi:10.1007/s10797-015-9360-5) contains supplementary material, which is available to authorized users.

✉ Johannes Hagen
johannes.hagen@nek.uu.se

[1] Department of Economics, Uppsala University, Box 513, 751 20 Uppsala, Sweden

1 Introduction

The ongoing shift in pension provision from defined benefit (DB) to defined contribution (DC) has brought more flexibility not only to the accumulation phase of retirement, but also to the decumulation phase. Flexibility during the decumulation phase manifests itself primarily through the introduction of more liquid payout options, such as lump sums and phased withdrawals, alongside the traditional life annuity.[1] Payout phase design involves a trade-off between flexibility and protection from longevity risk. Liquid payout options allow individuals to invest, buy an annuity, leave bequests or increase consumption during the early years of retirement, but raise concerns that they may trigger individuals to spend the money too rapidly for their own good (Barr and Diamond 2008). Despite the fact that many countries and private pension plan sponsors have referred to this trade-off to motivate their specific payout phase design, little is known about the characteristics of those who choose to annuitize and those who do not.

The main reason for the limited amount of empirical research on the demand for different payout options is the lack of reliable and comprehensive data. Private pension sponsors and life insurance companies are often reluctant to disclose individual choices, and most public pension systems impose mandatory annuitization. Survey-based data contain rich background information on the retirees, but usually lack actual payout decisions. Studies that use data provided by private annuity companies do contain rich information on annuity choices and prices, but are limited with respect to individual background information.

This paper uses unique micro-data from a large Swedish occupational pension plan to study the payout decision at retirement. The data are supplied by the second largest occupational pension sponsor in Sweden and include real payout decisions of about 183,000 white-collar workers. The payout decision involves substantial amounts of retirement savings, as workers are required to contribute a fraction of the wage to whatever occupational pension plan the employer is affiliated to. The company data are merged with national administrative data from Statistics Sweden to get rich individual background information, such as labor market history, education level, health status and parent longevity. To my knowledge, this is the first paper to combine data from a private life insurance company with administrative data to study the determinants of annuitization.

Previous empirical studies on the determinants of annuitization have tried to explain the so-called annuity market participation puzzle. The precise nature of the annuity puzzle is not well defined, but traditionally refers to the contradictory nature of low annuitization rates in the private market for annuities in the USA and the theoretical predictions of standard standard neoclassical life-cycle model (Benartzi et al. 2011). A seminal paper by Yaari (1965) shows that risk averse individuals without bequest motives will always prefer to hold their assets in actuarial notes (buy an annuity) rather than ordinary notes. A number of explanations have been proposed to explain the low demand for life annuities; the presence of load factors arising from administrative

[1] With lump sums, individuals receive the entire value of the accumulated retirement as a single payment, whereas phased withdrawals allow individuals to agree on a schedule of period fixed or variable payments (Antolin 2008).

costs, incomplete markets and adverse selection (Mitchell et al. 1999; Finkelstein and Poterba 2002, 2004); bequest motives (Friedman and Warshawsky 1990; Brown 2001; Inkmann et al. 2011; Ameriks et al. 2011; Lockwood 2012); annuity prices (Warner and Pleeter 2001; Fitzpatrick 2012; Chalmers and Reuter 2012); means-tested government benefits (Bütler et al. 2011; Pashchenko 2013); and pre-annuitized first-pillar pension income (Bernheim 1992; Dushi et al. 2004; Beshears et al. 2011). More behaviorally oriented phenomena, such as loss aversion, default provision and framing, have also been put forward as potential explanations for the annuity puzzle (Brown 2007; Brown et al. 2008; Agnew et al. 2008; Benartzi et al. 2011; Beshears et al. 2014).

The payout decision I study in this paper concerns whether individuals withdraw their occupational pension as a life annuity or during a fixed number of years. Under the life annuity, the individual's pension capital is converted into a monthly payment stream that is paid out as long as the individual is alive. In the fixed-term payout options, the individual specifies the time period during which the pension capital should be withdrawn. Payments cease after the preferred time period has expired. Payments also cease if the individual dies before the specified payout period has elapsed.[2] The fastest rate at which the pension capital can be withdrawn is over 5 years.

The average annuitization rate over the whole period is 76 %. This number is comparable to the annuitization rates in papers that study similar pension settings (Chalmers and Reuter 2012; Bütler and Teppa 2007). Thus, studies on payout decisions in mandatory second pillar pension plans, including this one, raise less concerns about the existence of an annuity puzzle than what the size of private annuity markets does.[3] Under the period of study, 2008–2013, the fraction of retirees choosing fixed-term payout options rose from 20 to 31 %. The 5-year payout is by far the most popular fixed-term payout option, chosen by more than two-thirds of the individuals who did not choose the life annuity.

I carefully analyze some of the determinants of annuitization that have been discussed in the literature. I pay particular attention to the role of health and life expectancy (adverse selection), retirement wealth and the tax consequences of choosing different payout options. As for adverse selection, I analyze whether there are systematic relationships between the length of the payout on the one hand and ex-post mortality on the other. I also investigate whether individuals in bad health self-select into fixed-term payouts. Health is proxied by the number of days an individual has been absent from work due to illness during a pre-specified age interval prior to retirement. Finally, I use information on parent mortality to create proxies for an individual's life expectancy.

The results show that individuals in bad health are more likely to choose the 5-year payout. Ex-post mortality rates also signal the presence of adverse selection. Individuals who choose the 5-year payout are 59 % more likely to die within two years after claiming than annuitants. I also find that the parents of annuitants live longer than

[2] In DC schemes, an individual can buy survivor insurance, which means that the remaining pension capital will be paid out to his or her partner or children. The combination between a fixed-term payout and survivor insurance is sometimes referred to as a *fixed-term annuity*.

[3] Annuitization rates in US DB pension plans that offer a lump sum are typically lower than this, ranging from 25 to 50 % (Mottola and Utkus 2007; Benartzi et al. 2011; Previtero 2014; Banerjee 2013). Similar results are found in Brown et al. (2015) who study an irrevocable choice between a more immediate and a more deferred payment stream in Croatia, where 30 % choose the deferred payment.

the parents of those who choose the 5-year payout. This suggests that individuals form expectations about how long they are likely to live based on the life-span patterns of their parents and take this into account when they decide whether to annuitize or not.

Small stocks of pension capital are more likely to be withdrawn during a fixed number of years. Although small capital stocks often signal low pre-retirement income, this effect seems to be mainly driven by channels associated with the size of the capital stock rather than by the income level of the individual. Payout preferences are fairly constant across the income distribution, except among individuals at the very top. High-income individuals, particularly those with large capital stocks, are much less likely to choose the 5-year payout. One potential explanation for this is that they want to avoid a higher effective marginal tax rate. I develop a tax-adjusted measure of the money's worth ratio (MWR) of the life annuity to study the effect of tax-induced changes in annuity prices on the decision to annuitize.

The rest of the paper is organized as follows. Section 2 provides a brief description of the Swedish pension system, with emphasis on the structure of the occupational pension plan for white-collar workers. The potential determinants of the demand for different payout options are discussed in Sect. 3. Section 4 describes the data, and Sect. 5 reports the results from several empirical specifications. Section 6 concludes.

2 Background information

In this section, I describe the main components of the Swedish pension system. I pay particular attention to the structure of the occupational pension system, in which individuals face different payout options.[4]

2.1 The structure of the Swedish pension system

Sweden's pension system has two main pillars, a universal public pension system and an occupational pension system for workers whose employer is tied to some occupational pension plan. Participation in the second pillar is mandatory for these employees. The public pension is the most important source of pension income, amounting to 50–80 % of an individual's total pension income. Mandatory annuitization applies to all pension wealth in the public pension system. The public pension system has in itself three tiers, of which two are earnings-related and DC. They insure income up to a certain threshold level called the "income ceiling".[5] The third tier is a means-tested pension supplement that ensures individuals with no or low pension income from the earnings-related component a minimum standard of living in retirement.

The second pillar consists of a number of different occupational, employer-provided pension plans. Occupational pension plans are constructed and thought of as supplements to the public pension system, as they provide pension benefits above the income ceiling in the public pension system. The occupational pension therefore plays a more important role for total retirement income for individuals with earnings above the ceil-

[4] A detailed description of the Swedish pension system can be found in Hagen (2013).

[5] The ceiling is currently at 7.5 income base amounts (IBA). This means that no pension rights are earned for the monthly wage portion that exceeds SEK 35 375 (1 USD = 6.5 SEK in 2013).

ing than for those below. Contributions to the second pillar are essentially proportional to insured income up to the income ceiling. The employer is mandated to contribute between 4 and 6 % of the wage portion of the insured that does not exceed the income ceiling. For wage portions above this threshold, contribution rates are much higher, typically around 30 %. There is also a third pillar for voluntary savings available to anyone who cares to supplement the retirement income provided by the first two pillars.

2.2 Occupational pension for white-collar workers

Most occupational pension plans are designed and implemented at the union level. There are four large agreement-based occupational pension plans that cover around 90 % of the total work force. Two of these plans cover workers employed in the public sector. The other two pension plans cover white-collar workers and blue-collar workers in the private sector, respectively. This study focuses on payout patterns in the pension plan for white-collar workers.

In the last two decades, all major occupational pension plans have undergone significant changes. Most importantly, all plans have been changed from DB to DC or a mixture of the two. In the old DB-dominated regime, individuals had relatively little control over their occupational pension assets. Transferability of accumulated assets was limited, and benefits after retirement were typically received in the form of life annuities. The transition to DC has had important implications for the control individuals have over their assets. In particular, all pension plans have introduced fixed-term payouts as an alternative to lifelong annuities. Since the transition to DC is still ongoing, the majority of today's retirees still have some part of their occupational pension wealth in old DB plans where annuitization is mandatory. White-collar workers, however, face no restriction on the fraction of wealth that can be withdrawn during a fixed number of years.[6] More on this below.

The pension plan for white-collar workers is called ITP. The transition of the ITP plan to DC began as late as in 2006, which implies that all cohorts in my sample belong to the old pension plan.[7] The most important source of occupational pension income for white-collar workers in this pension plan is DB. This component is referred to as ITP2.[8] Mandatory annuitization applied to all ITP2 benefits up until 2008 when fixed-term payouts were introduced. In fact, ITP2 is the only DB plan in Sweden that allows

[6] For example, the pension plan for blue-collar workers, SAF-LO, only allows for fixed-term withdrawals of pension wealth accumulated after 1996.

[7] The first cohort to be affected by the new DC pension plan, called ITP1, is those born in 1979.

[8] Benefits from ITP2 are calculated based on the pensionable wage w, which, in effect, is the final wage. Pensionable wage also incorporates benefits in kind, compensation for regular shift work, time on call and standby time at the time of retirement. The replacement rate is higher for pre-retirement income that exceeds the income ceiling in the public pension system. The ITP2 benefit is calculated according to the following equation, where w_i denotes the wage portion related to IBA i:

$$\text{ITP2} = 0.1w_{<7.5\ \text{IBA}} + 0,65w_{7.5-20\ \text{IBA}} + 0.325w_{20-30\ \text{IBA}}$$

For full ITP2, 30 whole entitlement years are required. An entitlement year can be earned from age 28 and is earned if the individual worked at least 20 % of full time. Earned ITP2 entitlement years are transferred to the new employer if the individual changes job.

for fixed-term payouts. Because the payout decision accrues to all ITP2 pension wealth, substantial amounts are at stake in the payout decision among this group of workers.

Employer-sponsored contributions to the second pillar are managed by some occupational pension company. Each pension plan has its own "default" managing company, meaning that if the individual takes no action, her pension assets will be managed by that pension plan's default company. Importantly, the ITP plan does not allow individuals to transfer DB assets between pension companies. Self-selection out of the default managing company, Alecta, which provides the data used in this paper, is therefore not an issue. Neither is adverse selection in the sense that only individuals who expect to live long insure themselves an issue because the ITP plan covers around 90 % of the population of Swedish white-collar workers.[9]

The normal retirement age for white-collar workers is 65. A few months before plan participants reach this age, they receive information about the available payout options from the managing pension company. The information letter clearly states the size of the monthly benefit under each payout option. If the participant takes no action, the ITP2 pension is paid out as a life annuity from age 65. However, opting out is easy and requires no time-consuming paperwork. The individual simply ticks the box that corresponds to the preferred payout option in the information letter.

The life annuity guarantees the retiree a stream of money right up until the point of death, whereas payments cease after a certain date under the fixed-term payout options. Fixed-term payments also cease if the individual dies before the end of the term. There are different conversion factors for each payout option, i.e., the factor at which the accumulated pension capital is converted into a monthly payment. The conversion factors depend on assumptions about average life expectancy at each claiming age and the rate of return on the pension capital, but is independent of gender and marital status. The resulting monthly benefit is increased with the inflation rate. The occupational pension plan I study in this paper offers five payout options: a life annuity, or fixed-term payouts over 5, 10, 15 or 20 years.[10]

2.3 Tax treatment of occupational pension income

Retirement income from public and private pension plans is subject to the same tax rules as income from labor.[11] A proportional local tax rate applies to all earned income and taxable transfers which includes pension income. The mean local income tax in 2013 was 31.73 % with a minimum rate of 28.89 and a maximum rate of 34.52. For pension incomes above a certain threshold (SEK 450,300 in 2013; 1 USD = 6.5

[9] There is also a DC component within the ITP plan called ITPK. Similar to ITP2, ITPK assets can be paid out during a fixed number of years or as a life annuity. This study focus on payout choices in the DB component for two reasons. First, ITPK is less important and normally amounts to 10 % of the final wage if t is paid out over five years. Second, the data only contains information on ITPK assets that were accumulated prior to 2008 (*Ursprunglig* ITPK).

[10] In practice, retirees are allowed to withdraw their pension during any number of years with a minimum of 5 years. However, only 1 % choose some other payout length than the pre-specified fixed-term payout lengths (5, 10, 15 and 20 years) and the life annuity.

[11] See Edmark et al. (2012) for a detailed description of the Swedish income tax system.

SEK), the taxpayer also has to pay a central government (state) income tax. The state income tax schedule consists of two brackets; the marginal tax rates in each bracket are 20 % (for incomes between 450,200 and 620,600 in 2013) and 25 % (for incomes above 620,600), respectively. Before computing the individual's tax liability, a basic deduction is made against the individual's total income. The basic deduction is phased in at lower income levels and phased out at higher income levels with consequences for the marginal tax rate in these income intervals.[12]

The progressivity of the tax schedule will have implications for individuals' valuation of different payout alternatives in the occupationalpension, especially for high-income earners with incomes above the state tax threshold.

3 Empirical predictions

This section summarizes the potential determinants of annuitization that have been discussed in the literature and that can be tested empirically with the data at hand.

The choice between the life annuity and any of the fixed-term payout options should depend on the expected present discounted value (EPDV) of each payout option. The EPDV of a particular payout depends in turn on the price of that payout option, the discount rate and on the characteristics of the individual. Asymmetric information about retiree life expectancy is the most natural source of variation in the EPDV of a given payout option across individuals.

In a standard life-cycle model, life annuities provide higher rates of return than other risk-free investments, because they transfer assets from those who die to those who survive. The additional rate of return on life annuities is referred to as the "mortality premium".[13] However, because fixed-term payouts also transfer assets from those who die before the preferred time period has elapsed to those who survive it is not clear that the life annuity should provide a higher rate of return. As shown below, the EPDV of the ITP life annuity is close to the EPDV of the fixed-term payouts. Moreover, if there is an option value associated with holding liquid assets, fixed-term payouts should become even more attractive.

3.1 Health, mortality and life expectancy

An individual's health condition and life expectancy should influence the annuitization decision because annuities hedge longevity risk. If individuals recognize and respond to variation in the relative value of different payout options that is due to differences in longevity, individuals who expect to live long should prefer life annuities to fixed-term payouts with short payment horizons and vice versa.

[12] From 2009, the basic deduction for individuals aged 65 and above is higher than for those below this age. This implies that the thresholds in the state income tax schedule are somewhat lower for individuals below age 65. In 2013, the basic deduction for individuals aged 65 and above was phased in between SEK 45,000 and SEK 166,900 and phased out between SEK 212,300 and 538,700.

[13] Davidoff et al. (2005) extend Yaari (1965) and show that individuals benefit from converting a significant fraction of their assets into life annuities even in the presence of incomplete markets for life annuities. Feigenbaum et al. (2013), on the other hand, show that the welfare effect of annuitization is ambiguous in general equilibrium on account of pecuniary externalities.

The most compelling evidence of the presence of adverse selection in life annuity markets is when ex-post mortality rates are lower among those who buy life annuities than among those who do not. Finkelstein and Poterba (2004) report evidence of adverse selection of long-lived individuals into private annuity markets in the UK. Because fixed-term payouts were introduced in 2008, I can track mortality only within a few years after the claim was made. If there is adverse selection, ex-post mortality rates should be higher among individuals who choose fixed-term payouts.

I get information about an individual's health status from social insurance register data. The most straightforward health measure is based on the number of days an individual has been absent from work due to sickness. The medical literature has shown that sickness absence can be used as an integrated measure of physical, psychological and social functioning in studies of working populations (Marmot et al. 1995; Kivimäki et al. 2003).

However, health is a non-perfect measure of subjective life expectancy. Two individuals with similar health status can have very different beliefs about how long they are likely to live relative to the cohort average. Such differences in beliefs might reflect family-specific rather than individual-specific characteristics. A very intuitive and simple way for an individual to get information about how long she is likely to live is to look at life-span patterns of family members. It has been shown that individuals take (same-sex) parent longevity into account when assessing their own life horizons (Van Solinge and Henkens 2009). I use information about the age at death of individual *i*'s parents to proxy individual *i*'s life expectancy.

3.2 Retirement wealth

Wealth has been shown to be an important determinant of the payout decision.[14] Ideally, wealth should be measured before annuitization takes place and include both pension and non-pension financial wealth. However, because the data do not contain information on non-pension financial wealth, the analysis is restricted to (occupational) pension wealth. Non-pension wealth is proxied by a measure of permanent income.

Pension income from other occupational pension plans and the public pension system should be thought of as pre-annuitized wealth.[15] These may act as substitutes for a life annuity in the ITP plan. Because the marginal value of insurance declines with the level of insurance, the value that a retiree attaches to the incremental life annuity should fall with the level of pre-annuitized income.

I predict that retirees who are more reliant on ITP benefits because they spent most of their career working for ITP employers should be more likely to choose the life annuity. The marginal effect of income from other occupational pension plans on the probability of choosing the life annuity should therefore be negative. However, the

[14] Some studies that find that (retirement) wealth and pre-annuitized income are important determinants of the payout decision are Inkmann et al. (2011), Chalmers and Reuter (2012), Bütler and Teppa (2007) and Pashchenko (2013).

[15] Pension wealth in occupational pension plans other than the ITP plan can be viewed as pre-annuitized wealth since these plans only allow limited amounts of DC capital to be withdrawn as fixed-term payouts for the cohorts studied in this paper.

relationship between retirement wealth and payout preferences is potentially more complex than this. First, it has been shown that small outcomes are discounted at higher rates than greater ones are (Frederick et al. 2002). Individuals with low levels of occupational pension wealth might find life annuities unattractive since their wealth would be translated into a very small payment stream. Low levels of retirement wealth would then be associated with a higher propensity to cash out the money at the fastest possible rate, i.e., over 5 years. At the same time, low levels of capital might be the result of high discount rates due to low investment in education. Second, if individuals prefer to cash out small amounts of retirement wealth and if the researcher has an incomplete picture of individuals' payout decisions in different pension plans, low annuitization rates might reflect the distribution of account balances rather than the distribution of preferences (Benartzi et al. 2011).

Another reason why we would expect individuals to withdraw their pension during a fixed number of years is to become eligible for different kinds of means-tested benefits after the pension payments have ceased. Indeed, Bütler et al. (2011) show that the availability of means-tested benefits can reduce the desired annuitization levels substantially. This "moral hazard" issue is, however, difficult to investigate empirically in the context of this paper. Not enough time has passed since the introduction of fixed-term payouts to see whether individuals who choose the 5-year payout are more likely to receive means-tested benefits 5 years after claiming.

3.3 Annuity pricing and tax treatment

Variation in the value of life annuities arise not only from differences in retiree characteristics, but also from differences in annuity pricing. An individual should prefer the annuity if the price of the annuity is lower than the expected benefit from smoother lifetime consumption. Chalmers and Reuter (2012) show that retirees respond very little or nothing at all to changes in the price of the life annuity option. They argue that low price elasticities of demand are due to the complexity of evaluating the EPDV of different payout options. However, based on the relatively high demand for the better than actuarially fair-priced life annuity that they observe, they suggest that retirees may still respond strongly to large, salient changes in annuity prices. This notion is supported by Bütler et al. (2013) who find that a large 20 % reduction in the rate at which capital is translated into an annuity reduced the propensity to annuitize by about 8 percentage points.

Large changes in the value of life annuities can arise from the tax treatment of retirement income. The progressivity of the tax schedule implies that the effective marginal tax rates under the fixed-term payouts are higher than under the life annuity. In this section, I extend the traditional measure of an annuity's value, the "money's worth ratio" (MWR), to account for the tax treatment of pension income.[16]

The MWR of a life annuity is the ratio of the EPDV of the flow of payments made by an annuity to the money paid for an annuity, where the money paid for an annuity

[16] MWRs of annuities have been used in a number of earlier studies, including Friedman and Warshawsky (1988), Mitchell et al. (1999), Brown and Poterba (2000), Finkelstein and Poterba (2004) and Chalmers and Reuter (2012).

typically refers to the foregone lump sum payment. In my setting, the value of the money paid cannot be determined since there is no lump sum option, nor information in the data on the exact amount of contributions paid to the ITP plan. This implies that the EPDV of the life annuity must be related to the EPDV of each fixed-term payout option in order to find the MWR of the life annuity. The EPDV of payout option p purchased by an individual of gender g and age a in year t is given by:

$$\text{EPDV}^{p}_{g,a,t}(B) = B^{p} \sum_{i=1}^{T} \pi_{g,a,a+i}(1 + r_t)^{-i} \tag{1}$$

where $\pi_{a,a+i}$ is the probability of someone living i more years, believed at age a, B^p is the annual gross benefit received by an individual under payout option p, r_t is the appropriate discount rate for payments received in year t, expressed at an annual rate, and T is the last period of payment.[17] For life annuities, T is chosen so that $\pi_{t,T} \approx 0$, which happens when $T = 45$ (i.e., no one lives beyond age 110 years, assuming $a = 65$). For the 5-year payout option, $T = 5$, for the 10-year payout option $T = 10$ and so on.

The MWR of the life annuity is then the ratio of the EPDV of the life annuity to the EPDV of any of the fixed-term payout options. For example, the MWR of the life annuity with respect to the 5-year payout option is expressed as:

$$\text{MWR}^{\text{Annuity/5-year}} = \frac{\text{EPDV}^{\text{Annuity}}}{\text{EPDV}^{\text{5-year}}} \tag{2}$$

If this expression is equal to one, the life annuity and the 5-year payout have equal EPDVs.

The *net-of-tax* MWR is defined as the ratio between the EPDV of the net benefit under the life annuity and the EPDV of the net benefit under any of the fixed-term payout options. I replace the gross benefit B^p in Eq. 1 with the net benefit $B^{p,net}$. The net benefit is given by

$$B^{p,net} = B^{p} - \left[T(B^{p}, I, t) - T(0, I, t)\right] \tag{3}$$

where $T(B^p, I, t)$ is total taxes paid under payout option p, $T(0, I, t)$ is total taxes paid when the individual has no pension income from the ITP plan, I is other pension

[17] When estimating the expected present discounted value of each payout option, I use gender-specific mortality tables published by Statistics Sweden for years 2009–2013. To account for the fact that payments of different payout length might be discounted at different rates, I use the mean yield on 5-year Treasury notes in year t-1 for the 5-year payout option and the mean yield on 10-year notes for the remaining payout options. The mean yield on 10-year Treasury notes in 2012 was 1.59 % compared to 1.12 % for 5-year notes.

Table 1 Tax-adjusted money's worth ratios (MWR) of life annuities

	Low income	Middle income	High income
Gross annuity	2,000	6,000	12,000
Net-of-tax annuity	1,437	3,970	7,850
Average tax, 5-year	.303	.352	.477
Average tax, 10-year	.288	.342	.413
Average tax, 15-year	.282	.341	.364
Average tax, annuity	.282	.338	.346
Gross $MWR^{Annuity/5-year}$	.958	.958	.958
Gross $MWR^{Annuity/10-year}$	.958	.958	.958
Gross $MWR^{Annuity/15-year}$	.961	.961	.961
Net $MWR^{Annuity/5-year}$	.988	.978	1.20
Net $MWR^{Annuity/10-year}$	.967	.964	1.07
Net $MWR^{Annuity/15-year}$	.962	.965	.989

This table reports tax-adjusted calculations of the money's worth ratio (MWR) for three representative individuals. The MWR is defined as the ratio of the expected present discounted value of the net-of-tax benefit under the life annuity to the expected present discounted value of the net-of-tax benefit under a given fixed-term payout. Each individual retires at age 65 and has no labor income. The public pension for the low-, middle- and high-income individual amounts to SEK 10,000, 15,000 and 20,000, respectively. The relative importance of the ITP benefit increases with the income level. The first row reports the ITP benefit paid out as an annuity. The second row reports the net-of-tax ITP benefit. The benefit formulas and the tax schedule of year 2013 are used

income and t is the claim year.[18] The average tax rate is then given by:

$$\tau = \frac{T(B^p, I, t) - T(0, I, t)}{B^p} \tag{4}$$

Table 1 reports average tax rates and MWRs for three representative individuals. The first row reports the gross benefit, B^p, under the life annuity. The second row reports the corresponding net benefit. Due to the progressivity of the tax schedule, the average tax rate decreases with the length of the payout for a given level of income. Short payout horizons result in higher monthly income and higher marginal taxes. Note that the average tax rate for the low- and middle-income individuals is only two percentage points higher under the 5-year option than under the life annuity compared to more than 13 percentage points for the high-income individual. For the high-income individual, this corresponds to an increase in the average tax rate of 38 %. The reason for this is that the high-income individual reaches the state income tax threshold under the 5- and 10-year payout options.

The table also reveals that the gross MWR of the life annuity with respect to each fixed-term payout is somewhat below one. In fact, the EPDV of the life annuity is larger

[18] As mentioned in Sect. 2.3, from 2009 and onwards, the basic deduction is higher for individuals aged 65 or above than for those below. The tax paid, T, should therefore be a function of both t and a. However, because almost 90 % claim at age 65, I simplify by making the tax function independent of age.

than the EPDV of the 5-year payout only for discount rates below 1.1 %. Because fixed term payouts yield similar, or even higher, expected returns than the life annuity, it is hard to motivate why low annuity demand in this setting would reflect an "annuity puzzle".

The picture changes when taxes are taken into account. The net-of-tax MWR calculations show that the fixed-term payouts become less attractive relative to the life annuity. This is particularly true for the 5-year payout. For example, the MWR of the life annuity with respect to the 5-year payout for the middle-income individual increases from 0.96 to 0.98 when taxes are accounted for. The corresponding net MWR for the high-income individual amounts to as much as 1.20. These changes reflect average tax rate differences across income levels and payout options.

The variation in net MWR across individuals will be used in the empirical analysis. First, I calculate the net MWR of the life annuity with respect to each fixed-term payout option for each individual.[19] Following the literature, the net MWR is then added as a regressor to predict payout choice.[20]

Another way to investigate the importance of the tax effect is to look at individuals who are exposed to different marginal tax rates depending on their payout choice. For this reason, I identify individuals who are subject to state income tax only under the 5-year option. Because the net EPDV of the 5-year payout is substantially lower than the corresponding gross EPDV due to a higher marginal tax rate, these individuals should be less likely to choose the fixed-term payouts.

3.4 Bequest motives, socioeconomic background and demographic characteristics

Differences in retiree characteristics that relate to socioeconomic background and demographics can also generate cross-sectional differences in the expected utility associated with life annuity payments.

Bequest motives have been put forward as one of the most important explanations for the annuity puzzle. In the absence of bequest motives, any wealth that an individual is holding at death does not contribute to utility, so there is no reason to forego the higher rate of return on annuities that arises from the mortality premium (Brown 2001). However, the preference for leaving bequests may counteract the insurance benefits of annuities and make more liquid payout options relatively more attractive.

Empirical findings on the effects of bequest motives on payout decisions are mixed. To test for the effect of bequest motives, researchers typically proxy intentional bequest motives with the presence of children. Brown (2001) and Inkmann et al. (2011) find insignificant effects of the number of children on the payout decision. Bütler and Teppa (2007) lack data on the presence of children, but find that divorced/widowed men cash

[19] When calculating individual MWRs, I take into account the tax system of the relevant year, pension income from other occupational pension plans and private pension accounts, public pension benefits, gender and the age at retirement. To simplify, I assume zero labor earnings after claiming.

[20] Brown (2001), Bütler et al. (2011, 2013) regress payout choice on the annuity equivalent wealth (AEW), an alternative measure of the value of an annuity. Chalmers and Reuter (2012) use the money's worth concept.

out more than single men, which is indicative for the presence of a bequest motive. I follow the literature and proxy bequest motives with the presence of children.[21]

As mentioned, the conversion factors do not depend on gender. Because women live longer than men on average, there is an actuarial bias that makes annuitization more attractive for women. The gross MWR of the life annuity with respect to the 5-year payout for women is 0.99 compared to 0.92 for men, which means that the EPDV of the life annuity is almost 9 % higher for women. The results in the literature with respect to gender are mixed. Inkmann et al. (2011) and Bütler and Teppa (2007) report higher cash-out rates for women, probably reflecting availability of alternative sources of income, whereas Chalmers and Reuter (2012) find the opposite.

Labor market participation at the time of withdrawal might also impact the decision to annuitize. The tax consequences of positive labor income after retirement, particularly in combination with large pension capital stocks, might drive down the demand for fixed-term payouts.

4 The data

4.1 The data sets

I use data from a pension company called Alecta. Alecta manages occupational pensions for approximately two million private customers, making it the second largest occupational pension company in Sweden and one of the largest owners on the Stockholm Stock Exchange. Alecta is the default managing company of the ITP plan which means that they administrate pension contributions and pension payouts for private-sector white-collar workers whose employer is part of the ITP plan.

The data consist of information on all Alecta's customers that retired over a six-year period ending in 2013. The sample includes 182,808 individuals born between 1943 and 1951. Most importantly, it contains information on the year and month each retiree claimed the occupational pension and under which payout option it is withdrawn. It also contains the relevant conversion factors for claims at age 65. The payout decision concerns pension wealth in the DB component, ITP2, which is tied to Alecta.

The company data are merged with register data from Statistics Sweden to obtain rich individual background information. The data are derived from the Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA). LISA covers the period 1990–2011, which implies that I can create panels of past income streams for each individual in the sample. From the panel structure of the register data, I construct one observation per retiree with variables that are likely to be determinants of the payout decision. The data also contain birth and mortality information for the biological parents of these individuals.

Since the register data end in 2011, there is no contemporaneous income information for individuals who retire in 2012 and 2013. In order to get a full picture of the post-retirement financial situation of these individuals, I restrict the sample to pay-

[21] Kopczuk and Lupton (2007) show that children are an imperfect proxy of bequest motives. An alternative proxy of bequest motives is the purchase of life or survivor insurance (Bernheim 1992). Survivor insurance in the ITP plan is optional. Unfortunately, the data do not include information on whether the individual has purchased this insurance or not.

out choices made in or before year 2010 when the analysis includes current income streams, such as labor market earnings, public pension benefits and pension income from other occupational pension plans. Losing three years of data reduces the sample size from 182,808 to 82,066.

Information on parents' date of birth and mortality is available for 89 % of the sample. The remaining 11 % have parents who never lived in Sweden and do not show up in the registers for this reason. These individuals are excluded from the sample in the analysis of parent mortality and payout choice, which reduces the sample size to 162, 537.

4.2 Descriptive statistics

Table 2 reports descriptive statistics for all key variables. All numerical variables are expressed in constant 2013 Swedish Crowns (SEK).

The average monthly benefit from the DB component, ITP2, amounts to SEK 4074. The large variation in the size of the benefit across individuals (standard deviation of SEK 6272) is a direct result of large differences in pre-retirement income and the construction of the ITP plan, which replaces more income above the income ceiling.

Pre-retirement income is defined as the average income from labor during the 5 years preceding retirement. The average monthly pre-retirement income is SEK 22,437. Permanent income is defined as the average monthly income between age 51 up to retirement. Permanent income should reflect lifetime earnings better than pre-retirement income, and is also more strongly correlated with retirement wealth. Active participation in the labor market after retirement ("Working" in Table 2) is defined as having labor earnings greater than two income base amounts in the year succeeding the claiming year.[22] In fact, more than 20 % are classified as working based on this definition.

We also see from the table that some white-collar workers receive pension income from other occupational pension plans. Other occupational pension income is defined as non-ITP occupational pension income that the individual receives in the last year that he or she is observed in the data. Because most people claim at age 65, this implies that other pension income is typically measured between ages 66–68 for those who claimed before 2011. On average, retirement benefits from the ITP plan make up 55.4 % of the total occupational pension income. 93 % receive income from some other occupational pension plan.[23] For 20 % of the sample, this income is three times larger than what they get from the ITP plan. Public pension benefits are on average 33 % larger than total occupational pension income.

For those who claim at age 65, the mean gross MWR of the life annuity with respect to the 5-year payout is 0.94. The corresponding tax-adjusted MWR is equal to 0.96. As expected, the tax treatment of occupational pension income increases the EPDV

[22] This corresponds to a monthly income of SEK 7 400 or 30 % of the average permanent income. This measure also includes self-employment income.

[23] 12, 18 and 38 % receive income from the occupational pension plan for central government workers, local government workers and blue-collar workers in the private sector, respectively. As many as 82 % receive pension income that is not classified by LISA as coming from any of the four major occupational pension plans (variable "OvrTjp").

Table 2 Summary statistics

	Mean	SD	Min	Max	Observations
Demographics					
- Female	.392	.488	0	1	182,808
- Married	.677	.468	0	1	182,808
- Single	.323	.468	0	1	182,808
- Widowed	.044	.205	0	1	182,808
- Elementary school	.204	.403	0	1	182,808
- High school	.56	.496	0	1	182,808
- College	.236	.425	0	1	182,808
- Has children	.745	.436	0	1	182,808
Pension income					
- Monthly ITP2 benefit	4,074	6,272	114	294,245	182,808
- ITP's share of occupational pension	.552	.289	.0011	1	82,066
- ITP's share of total pension	.212	.196	.00105	1	82,066
- Has other pension plan	.931	.254	0	1	82,066
- Larger benefit from other plans	.202	.401	0	1	82,066
Money's worth ratios (MWR)					
- Gross MWR annuity/5-year	.936	.0492	.832	1.05	161,130
- Net MWR annuity/5-year	.961	.0731	.816	1.37	72,734
- Average tax rate	.332	.0869	0	.567	82,066
- Pays state income tax	.101	.301	0	1	82,066
- State tax only under 5-year payout	.122	.327	0	1	82,066
Retirement age					
- Claim age - ITP2	64.8	.746	57	68	182,808
- Claim age - public pension	64.2	1.46	61	68	78,119
Pre-retirement income					
- Permanent income (monthly)	24,265	19,312	0	930,985	182,808
- Pre-retirement income (monthly)	22,437	21,694	0	1514837	182,808
- Working	.205	.404	0	1	82,066

Table 2 continued

	Mean	SD	Min	Max	Observations
Health and mortality					
- Mean sick leave days per year	15.7	30.2	0	304	182,430
- At least one day on sick leave	.569	.495	0	1	182,808
- 15 sick leave days per year or more	.251	.434	0	1	182,808
- Dead within 2 years after claim	.022	.147	0	1	82,066
- Dead within 4 years after claim	.043	.203	0	1	18,627
- Same-sex parent dead at age 65	.119	.324	0	1	162,537
- Same-sex parent alive at age 90	.224	.417	0	1	162,537

This table reports summary statistics for white-collar workers in the private sector who retired from the ITP plan between 2008 and 2013. Income variables are in the 2013 price level, 1 USD = 6.5 SEK

of the life annuity relative to the fixed-term payouts. The average tax rate, as given by Eq. 4, is 33 % and around 10 % pay the additional state marginal tax rate of 20 %. State tax status is contingent on choosing the 5-year payout for 12.2 % of the sample.

The average claiming age is 64.8. ITP2 benefits can be withdrawn from age 55 with an early retirement penalty of 7 % per year. Almost 90 % claim the occupational pension when they turn 65. A number of factors contribute to this pattern: Pension rights are not earned after the age of 65, benefits are automatically paid out at this age, and 65 is usually referred to as the "normal retirement age". Thus, there is a strong norm that retirement should happen at this age. As seen in Table 2, the average age at which public pension benefits are claimed (64.2) is somewhat lower than the average age at which ITP benefits are claimed (64.8).[24]

39.3 % of the sample are women. An individual's marital status is based on what is observed at the time of retirement. 67.7 and 32.3 % of the sample are married[25] and singles, respectively, and 4.4 % are widowed. A high school degree (12 years of schooling) is the highest education level for the majority (56 %). 23.6 % has a college degree (at least two years in college or university), and 20.4 % finished only elementary school (9 years or less).

Information on health and mortality is also collected from LISA. I use individual-specific information on sickness absence as a proxy for health.[26] I look at sickness absence incidence between age 51 up to retirement. The mean number of days per year of sickness absence in the sample is 15.7. About 43 % of the sample never received sickness benefits during this period.

[24] Remember that the sample is restricted to retirees who claimed ITP benefits before 2011 when we look at non-ITP income sources (N = 82,066). The public pension claiming age is not observed for about 5 % which means that the true average claiming age might be somewhat higher than 64.2.

[25] Includes co-habiting partners.

[26] An employee that calls in sick receives sick pay from the employer the first 14 days of absence. After 14 days, the employer reports the case to the Swedish Social Insurance Agency (Försäkringskassan), who pays out a sickness benefit that amounts to about 80 % of the individual's salary. The data in this study contain information on the number of days individuals receive sickness benefits from Försäkringskassan.

Mortality data are available up to year 2012. For this reason, the sample is restricted to claims made before 2011 and 2009 when I look at mortality within two and four years after the claim date, respectively. 2.2 % of the sample die within two years after claiming and 4.3 % within 4 years. These mortality rates are sufficiently high to allow meaningful study of retiree mortality and adverse selection.

The ideal measure of life expectancy based on parent mortality would be to relate the age at death of the individual's same-sex parent to the average age at death of that parent's cohort. However, not enough time has passed to determine cohort mortality for all parent cohorts in the sample.[27] Instead, I create two dummy variables that take on the value of one if the individual's same-sex parent is deceased/alive at age 65/90. An individual whose same-sex parent died before age 65 is likely to expect to live shorter than an individual whose same-sex parent was alive at age 90.[28]

4.3 Payout choices (the dependent variable)

The payout choice variable is the outcome variable and takes on five different values as explained in Sect. 2: Individuals can withdraw their pension wealth over 5, 10, 15 or 20 years (fixed-term payouts) or as a life annuity.

Relative frequencies of the choice variable by year are reported in Table 3. Between 2008, the year in which the ITP plan introduced fixed-term payouts, and 2013, the fraction of people that chooses the life annuity decreased from 80.9 to 70.5 %. This averages to 75.9 % over the whole period. The 5-year payout option is the most popular alternative among the fixed-term payouts. As seen in column (7), 16.1 % of all retirees choose this option. However, the 10- and 15-year payout options have grown more in relative terms. The 10-year payout option, for example, was preferred by only 3.9 % in 2008, but by 7.8 % in 2013. Very few people (0.2 %) choose the 20-year payout.

There are two potentially important explanations for the decreasing demand of the life annuity. First, life annuities have become less attractive as a result of declining interest rates. When interest rates fall, annuity payouts are reduced for the same amount of pension capital. However, while the mean yield on ten-year Treasure notes fell from 3.90 to 2.12 % between 2008 and 2013, Alecta's underlying assumption about the interest rate was changed only once during this period. In December 2012, the interest rate assumption was reduced from 3 % to 2 %, which generated a substantial drop in the gross MWR of about 7 percentage points. This relative price change possibly contributed to the falling demand for the life annuity in 2013, but given that the trend towards shorter payout horizons seem to have started more than two years before this, there might be other important forces at play.

The second explanation is that the trend toward shorter payout horizons is driven by an information or learning effect. Even though there is no direct empirical evidence of such an effect, knowledge about the existence of alternative payout options in second pillar pension plans has probably become more widespread in the most recent years

[27] Complete cohort-specific mortality tables are available for cohorts born in 1910 or earlier (SCB 2010).

[28] Various definitions of parent longevity are possible. I also create dummies for whether both (and any) of the parents are deceased/alive at age 65/90. The results turn out to be robust to these definitions.

Table 3 Relative frequencies of the choice variable, reported by year

	(1) 2008	(2) 2009	(3) 2010	(4) 2011	(5) 2012	(6) 2013	(7) All years
5-Year payout	0.139	0.152	0.147	0.163	0.175	0.176	0.161
10-Year payout	0.039	0.040	0.036	0.048	0.072	0.078	0.054
15-Year payout	0.012	0.018	0.017	0.023	0.038	0.037	0.025
20-Year payout	0.000	0.001	0.001	0.002	0.002	0.004	0.002
Annuity	0.809	0.789	0.799	0.764	0.712	0.705	0.759
Observations	18,627	30,397	33,042	34,023	33,680	33,039	182,808

This table shows relative frequencies of the choice variable reported by year (columns 1–6). Column (7) pools all years

as a result of more media attention. This implies that more people make active decisions and are less likely to be defaulted into the life annuity. Both from the viewpoint of the individual and the policy maker, the payout decision becomes an increasingly more important matter as the share of occupational pension wealth that is not subject to mandatory annuitization grows over time in all major occupational pension plans.

Table 4 reports relative frequencies of the choice variable by a number of individual characteristics. We see that the distribution of preferences varies substantially across some characteristics and less across others.

There are large differences in payout preferences between income groups. 76 % of the individuals in the bottom permanent income quartile choose the life annuity compared to 79.5 % in the top quartile. Only 9.7 % in the top quartile choose the 5-year payout, whereas 19 % do so in the bottom quartile. Differences are even more pronounced across the account balance distribution.[29] Only 68.3 % in the bottom account balance quartile annuitize (28.6 % choose the 5-year payout) compared to 82.9 % in the top quartile (6.2 % choose the 5-year payout). Individuals that continue to work after collecting the ITP benefit are more likely to annuitize than those who retire completely from the work force. Given the actuarial bias in favor of the life annuity for women, it is somewhat surprising that women choose the 5-year payout option more than men (18.7 vs. 14.4 %). There is also preliminary evidence that individuals compare the net-of-tax value of different payout options. Only 5.3 % of those who are subject to state income tax only under the 5-year payout choose this payout length. Individuals who die within 2 years of claiming are much more likely to choose the 5-year option, providing preliminary evidence of the presence of adverse selection.

[29] Specifically, the account balance distribution refers to the monetary value of the pension payment under the life annuity. For each individual, I observe the monthly benefit under the preferred payout option. The monetary value of the life annuity is calculated as the product of the given benefit and the relevant conversion factor. For example, if the monthly benefit under the 5-year payout is SEK 2000 and the (hypothetical) conversion factor is 3.5, the value of the life annuity is equal to SEK 714.3 ($2500 \div 3.5 = 714.3$).

Table 4 Relative frequencies of the choice variable, reported by demographic and socioeconomic characteristics

	5-Year	10-Year	15-Year	Annuity	Obs
Female	.187	.0477	.0178	.747	71,616
Male	.144	.0575	.03	.766	111,192
Has children	.16	.0554	.0267	.756	136,243
No children	.163	.0485	.0208	.766	46,565
Large benefit from other plans	.171	.0323	.0114	.785	16,547
Bottom ITP capital quartile	.269	.0392	.00868	.683	46,440
Upper ITP capital quartile	.0623	.0571	.047	.829	45,702
Pays state income tax	.144	.0563	.0317	.765	8299
State tax only under 5-year payout	.0528	.0493	.0349	.86	9987
Early retirement	.286	.107	.055	.542	21,678
Normal retirement age	.144	.0465	.0212	.788	161,130
Upper income quartile	.0965	.0588	.0444	.795	45,607
Lower income quartile	.19	.0368	.0131	.76	45,608
Working	.122	.029	.0134	.835	16,831
Not working	.154	.0408	.0167	.788	65,235
15 sick leave days per year	.195	.0493	.0186	.736	45,906
No sick leave	.133	.0519	.029	.783	78,804
Mortality within 2 years	.236	.0261	.00943	.729	1803
Same-sex parent dead at age 65	.173	.0599	.0269	.739	19,419
Same-sex parent alive at age 90	.156	.0497	.0222	.771	36,404
All individuals	.161	.0536	.0252	.759	182,808

This table reports relative frequencies of the choice variable by several demographic and socioeconomic characteristics. The 20-year payout option is excluded because very few people choose this option

5 Empirical results

This section investigates the individual-level determinants of annuitization. Table 5 reports marginal effects from a multinomial logit analysis applied to the pooled sample of individuals who claimed their ITP benefit between years 2008–2013. The reference category is the life annuity. The marginal effect associated with each coefficient can be interpreted as the variable' incremental effect on the probability of choosing a fixed-term payout option.[30]

Two sets of results are reported for each fixed-term payout. The model underlying the estimates in columns (1), (3) and (5) is estimated for the complete sample, whereas columns (2), (4) and (6) restrict the sample to individuals who claimed their ITP pen-

[30] The demand for fixed-term payout options can also be estimated in an OLS framework. I estimate a logit model where the dependent variable equals one if individual i chooses fixed-term payout option k and zero otherwise. The OLS estimates turn out to be very similar in magnitude to the marginal effects from the multinomial logit regressions and the inferences are unchanged. I do not report the estimation results for the 20-year payout. The estimates are not informative because only 0.2 % choose this payout option.

Table 5 Results from Multinomial Logit regressions (marginal effects)

	5-Year		10-Year		15-Year	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.001	−0.005^{+}	−0.006**	−0.010**	−0.005**	−0.006**
	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Married	−0.007**	−0.007*	0.004**	0.003^{+}	0.001	0.000
	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Widowed	−0.005	−0.004	−0.005^{+}	−0.003	−0.004	−0.001
	(0.004)	(0.006)	(0.003)	(0.004)	(0.002)	(0.003)
Has children	0.003	0.000	0.001	−0.000	0.002^{+}	0.002
	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Elementary school	0.014**	0.010**	0.002	−0.001	−0.004**	−0.002
	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
College	−0.013**	−0.001	−0.008**	−0.003	−0.002*	0.001
	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Early claimer	0.168**	0.145**	0.043**	0.032**	0.019**	0.011**
	(0.002)	(0.004)	(0.001)	(0.002)	(0.001)	(0.001)
Log permanent income	0.002**	0.007**	0.004**	0.003**	0.002**	0.001*
	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Bottom ITP capital quartile	0.110**	0.134**	−0.013**	−0.002	−0.021**	−0.009**
	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)
Top ITP capital quartile	−0.141**	−0.113**	−0.007**	−0.006**	0.014**	0.007**
	(0.003)	(0.005)	(0.001)	(0.002)	(0.001)	(0.001)
15 sick leave days per year or more	0.029**	0.021**	0.002	−0.003	−0.001	−0.002^{+}
	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
Same-sex parent dead at age 65	0.010**	0.008*	0.003*	0.003	−0.000	−0.003*
	(0.003)	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)
Same-sex parent alive at age 90	−0.009**	−0.010**	−0.003*	−0.003	−0.002*	−0.005**
	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)
State tax only under 5-year payout		−0.051**		0.006*		0.005**
		(0.007)		(0.002)		(0.001)
Working		−0.017**		−0.011**		−0.003**
		(0.003)		(0.002)		(0.001)

sion no later than in year 2010.[31] All specifications include a separate fixed effect for each year of claiming to control for variation in payout choices that is related to specific year effects. I also include a dummy for claiming before age 65 to control for variation in payout choices that is related to early retirement. All specifications also control

[31] All regression models restrict the sample to individuals for whom we know the date of birth of both biological parents. The results are similar when all individuals are included (excluding the variables of parent longevity from the model).

Table 5 continued

	5-year		10-year		15-year	
	(1)	(2)	(3)	(4)	(5)	(6)
Large benefit from other plans		−0.060** (0.004)		−0.005* (0.002)		0.001 (0.002)
Observations	162,537	73,555	162,537	73,555	162,537	73,555

This table reports marginal effects estimated via multinomial logit. The marginal effect associated with each coefficient can be interpreted as the variable's incremental ffect on the probability of choosing the fixed-term payout. All specifications include year fixed-effects, a control for log permanent income, two dummies for parent longevity, and dummies to indicate whether the individual is female; is married; is widowed; has children; finished elementary school or college; retires before 65; belongs to the bottom or top quartile of the ITP capital distribution; has more than 15 sick leave days per year on average. Columns (2), (4) and (6) add dummies to indicate whether the individual has large non-ITP pension income; works after claiming; pays state income tax only under the 5-year payout. The model in columns (1), (3) and (5) is estimated for the full sample of retirees who claimed their ITP benefit between 2008 and 2013, whereas the model in columns (2), (4) and (6) restricts the sample to individuals who claimed their ITP pension between 2008 and 2010. Standard errors in parentheses

$^{+}$ Statistically significant at the 10 % level

* Statistically significant at the 5 % level

** Statistically significant at the 1 % level

Table 6 Retiree mortality and payout choice

	Probability of dying within:	
	(1) Four years	(2) Two years
5-Year payout	0.017**	0.013**
	(0.004)	(0.001)
10-Year payout	-0.019^{+}	-0.006^{+}
	(0.010)	(0.003)
15-Year payout	−0.012	-0.010^{+}
	(0.016)	(0.005)
Observations	18619	82005
Mean dep. var.	0.043	0.022

Cols. 1 and 2 report marginal effects from a logit model of the probability of dying within two and four years of claiming, respectively. The excluded category is the life annuity. All regressions include, in addition to the covariates shown above, indicator variables for year of claiming, gender and a dummy for claiming before age 65. Since mortality data are available only up to year 2012, the analyses on mortality within two and four years are restricted to individuals who claimed prior to year 2011 and 2009, respectively. Standard errors in parentheses

$^{+}$ Statistically significant at the 10 % level

* Statistically significant at the 5 % level

** Statistically significant at the 1 % level

for gender, marital status, the presence of children, education level, early retirement, log permanent income, sick leave absence and two dummies for parent longevity. In addition, columns (2), (4) and (6) add dummy variables to indicate whether the individual is working or not after retirement, whether the individual's total pension

income exceeds the state income tax threshold under the 5-year payout option only, and whether more than 75 % of the individual's total occupational pension income comes from some other pension plan than the ITP plan. I also test for adverse selection by estimating logit models of the probability of dying within two and four years of claiming.[32] Table 6 reports the results from these estimations.

To evaluate the role of annuity pricing, I exploit tax-induced variation in the net MWR of the life annuity with respect to the fixed-term payouts. Adding the net MWR as a right-hand side variable would be straightforward in case the decision involved only two payout options. For example, studies that look at the choice between a lump sum and a life annuity use some appropriate measure of the relative value of the life annuity with respect to the lump sum to predict the probability to annuitize (Brown 2001; Bütler et al. 2011; Chalmers and Reuter 2012). However, with four payout alternatives to the life annuity, there are four potential individual-level regressors of the net MWR. Because of the popularity of the 5-year payout, I use the net MWR of the life annuity with respect to the 5-year payout.[33] If individuals recognize and respond to tax-induced variation in life annuity pricing, then the demand for the fixed-term payout should be lower when the value of the life annuity is higher. The marginal effects from these estimations are reported in Table 7. Because the net MWR is a function of the level of ITP capital, other pension income, I, and permanent income, I estimate the model both with and without controlling for these variables. I also estimate the model with and without controlling for gender because gender is the second source of variation in the net MWR alongside taxes. I report results for the 5- and 10-year payouts.[34]

5.1 Health, mortality and life expectancy

The results with respect to health and life expectancy are broadly consistent with our predictions. Column (1) of Table 5 shows that individuals with more than 15 sick leave days per year on average are 2.9 percentage points more likely to choose the 5-year payout option. The estimate is somewhat lower but still significant when the sample is restricted to payout decisions made in 2010 or earlier. The corresponding effect on the demand for the 10-year payout option is very close to zero, and even negative for the last specification, as seen in columns (3) and (4).

The estimated coefficients on same-sex parent mortality are significant and have the predicted signs. The marginal effect of same-sex parent mortality at age 65 on the probability of choosing the 5-year payout option is 1 percentage point. Retirees whose

[32] Finkelstein and Poterba (2004) apply a similar LPM model framework as well as a hazard model framework to study annuitant mortality in the UK annuity market.

[33] Using the relative value of the 5-year payout to predict the demand for the 10- and 15-year payout is not ideal since there might be within-individual variation in the net MWR across payout options. The within-individual variation is too small to motivate the use of conditional logit or mixed logit models. Such models allow for the explanatory variables to include attributes of the choice alternatives, which in our case would be the payout-specific net MWRs.

[34] As seen in Table 1, there is very limited variation in the net MWR across the income distribution. The relative value of the 5-year payout is therefore not a good proxy for the relative value of the 15-year payout.

same-sex parent was alive at age 90 decrease the demand for the 5-year payout by 0.9 percentage points. The corresponding effects for the 10-year payout are again smaller and amount to 0.3 and −0.3 percentage points, respectively. Thus, the parents of annuitants live longer than the parents of those who choose a fixed-term payout, which can be interpreted as evidence of adverse selection. A caveat of this interpretation could be that parent longevity picks up some of the variation in non-pension wealth that is not captured by any of the other variables.[35] We know that wealthier individuals tend to live longer (see, e.g., Smith 1999) and that wealth is an important determinant of annuitization. Thus, the possibility that some of the effect of parent longevity operates through non-pension wealth cannot be ruled out.

Table 6 reports marginal effects from a logit model on ex-post mortality and payout choice. Again, the estimated coefficients provide evidence pointing in the direction of adverse selection. Specifically, individuals who choose the 5-year payout are 1.7 and 1.3 percentage points more likely to die within four and two years after claiming, respectively. This corresponds to percentage effects of 40 and 59 %. In contrast, there is no self-selection of individuals who die shortly after retirement into any of the other fixed-term payout options. If anything, mortality rates among retirees who choose 10- or 15-year payouts are lower than among annuitants. These results support the general finding that individuals who enter retirement with low life expectancy or bad health tend to minimize the time period over which they withdraw their pension wealth.[36]

5.2 Retirement wealth and annuity pricing

The coefficients on log permanent income in columns (1) and (2) of Table 5 show that higher permanent income is associated with a marginally positive effect on the probability of choosing the 5-year payout, *given* a certain level of retirement wealth. Because income is highly correlated with retirement wealth, it is difficult to evaluate the role of income when the model controls simultaneously for ITP retirement wealth. In fact, the sign of the log permanent income coefficient changes when the retirement wealth variables are excluded from the regression model (not reported here).

Based on the descriptive analysis in Sect. 4.3 and the regression estimates in Table 5, it is nevertheless clear that the size of the ITP benefit is an important predictor of payout choice. Columns (1) and (2) reveal that retirees in the bottom quartile of the account balance distribution are 11 percentage points more likely to choose the 5-year payout than those in the second and third quartile. Retirees in the top quartile, on the other hand, are 14.1 percentage points less likely to choose the 5-year payout on the margin. The demand for 10-year payouts is less related to the size of the ITP benefit. The coefficients on the dummies for being in the bottom and top quartile reflect marginal effects of −1.3 and −0.1 percentage points, respectively.

[35] Non-pension wealth is imperfectly proxied by permanent income if wealth is inherited, and ITP wealth might underestimate the overall wealth level of the individual if she is a member of multiple occupational pension plans.

[36] OLS estimates from a linear probability model of ex-post mortality yield similar results.

An alternative way to measure the impact of the size of the ITP benefit on the demand for fixed-term payouts is to replace the existing account balance dummies with the monetary value of the ITP benefit under the life annuity. Again, the larger the annuity payment, the lower the demand for the 5-year payout. The estimated coefficients (not reported here) are statistically significant and also economically meaningful with a one-standard deviation increase in the monthly life annuity payments (SEK 4860) decreasing the demand for the 5-payout by 17 percentage points. Bütler and Teppa (2007) also find that the marginal effect of pension wealth on the demand for life annuities is positive, but smaller than in this study. Chalmers and Reuter (2012), however, report that a one-standard-deviation increase in the life annuity payments *increase* the demand for the lump sum between 3.0 and 5.2 percentage points.

Columns (2), (4) and (6) of Table 5 show that the demand for fixed-term payouts decreases when retirees receive income from some other occupational pension plan next to the ITP plan. More specifically, if at least 75 % of the total occupational pension income is paid out from a non-ITP pension plan, the demand for the 5- and 10-year payout decreases by 6.0 and 0.5 percentage points, respectively. This result contradicts the prediction that retirees that are less reliant on ITP benefits should be less likely to annuitize. A possible explanation to this result is that individuals with multiple sources of occupational pension devote more attention and effort to the payout decisions of other, more important pension plans. The default option becomes particularly important for retirees who have contributed to different plans during their career. Note also that the magnitude of the coefficient on the bottom pension quartile becomes larger when we control for other occupational pension income. This suggests that retirees that are located in the bottom quartile of the account balance distribution because of low lifetime earnings are more likely to choose short payout periods than retirees who have small account balances because they spent large part of their careers working for employers outside the ITP plan.

As seen in column (1), retirees whose total retirement income exceeds the state income tax threshold only under the 5-year payout are more than 5 percentage points less likely to choose the 5-year payout. This suggests that individuals evaluate the benefit's net-of-tax value for different payout lengths, rather than its gross value, at least when the tax effect is large and salient.

Turning to Table 7, which presents the marginal effects from a multinomial logit model of the net MWR on the demand for each fixed-term payout, we find further evidence that individuals respond to variation in the relative price of the life annuity. The marginal effect of the logit in specification (1) is -0.40 and is highly significant. This means that a one percentage point increase in the net MWR of the life annuity with respect to the 5-year payout leads to a 0.40 percentage point decrease in the probability of choosing the 5-year payout.

Controlling for gender (specification (2)) alters the impact of the net MWR substantially. The estimated marginal effect of -0.96 is both statistically and economically significant with a one-standard deviation increase in the net MWR of the life annuity being associated with a 7.0 percentage point reduction in the demand for the 5-year payout option. Because gender and taxes are the sole sources of variation in the net MWR, conditioning on gender implies that the estimate of the net MWR reflects tax-

induced changes in the relative value of the life annuity only.[37] Adding income and retirement wealth controls (specification 3) reduces the estimated marginal effect of the net MWR to −0.39. The difference is due to the fact that there is not much variation left in the net MWR when we condition linearly on retirement wealth and permanent income. Ultimately, identification hinges on functional form. There is no variation left in the net MWR if one controls for retirement wealth and income in a very flexible way.

The demand for the 10-year payout also falls when the net-of-tax value of the life annuity increases relative to the 5-year payout. Specifically, column (5) reveals that a one percentage point increase in the net MWR of the life annuity leads to 0.03 percentage point decrease in the probability of choosing the 10-year payout.

The results can be compared to the findings of previous papers that study the relationship between annuity demand and the relative value of the annuity. Contrary to the findings of this paper and what economic theory would predict, Chalmers and Reuter (2012) find that an increase in the money's worth of the life annuity leads to an increase in the demand for the lump sum. One reason for these opposite findings might be that tax-induced variation in the relative value of the life annuity is more salient and more easily understood than variation that stems from the use of multiple benefit formulas as in Chalmers and Reuter (2012).

The results of this paper are more in line with the findings of the group of papers that uses a related (utility-based) measure of the annuity's value, namely the annuity equivalent wealth (AEW). Brown (2001) and Bütler et al. (2013) report that a one percentage point increase in the AEW leads to a one and 0.88 percentage point increase in the annuitization rate, respectively. Bütler and Teppa (2007) reports a somewhat lower marginal effect of 0.44 %. The column (2) estimate in Table 7 of −0.96 is thus close to previous studies, although the interpretation of these estimates is different to the extent that the MWR and the AEW capture different aspects of the value of a life annuity.

5.3 Bequest motives, socioeconomic background and demographic characteristics

I find no evidence of bequest motives having an effect on the payout decision. All coefficients on the presence of children in Table 5 for the 5- and 10-year payouts are close to zero and insignificant. Marital status also turns out to be an unimportant predictor of the payout decision. Columns (1) and (2) in Table 5 show that married retirees are 0.7 percentage points less likely to choose the 5-year option. Previous studies on the demand for life annuities have found significant and important effects of marital status on the demand for life annuities (Bütler and Teppa 2007; Inkmann et al. 2011; Chalmers and Reuter 2012). One potential reason why I find zero or very small effects is that the rich set of control variables pick up variation in the demand for life annuities that are correlated with marital status but not due to it.

[37] Another way to isolate tax-induced variation in the net MWR is to use gender-independent mortality tables when calculating EPDVs according to Eq. 1. This yields similar marginal effects of −0.96.

Table 7 Impact of the net MWR on the demand for fixed-term payouts

	5-year			10-year		
	(1)	(2)	(3)	(4)	(5)	(6)
Net MWR annuity/5-year	−0.400**	−0.959**	−0.389**	−0.068**	−0.027*	−0.053**
	(0.023)	(0.039)	(0.042)	(0.012)	(0.013)	(0.015)
Gender	No	Yes	Yes	No	Yes	Yes
Wealth controls	No	No	Yes	No	No	Yes
Observations	65035	65035	65035	65035	65035	65035

This table reports marginal effects of the net money's worth ratio (MWR) of the life annuity with respect to the 5-year payout estimated via multinomial logit. Wealth and income controls refer to log permanent income, two dummies to indicate whether the individual belongs to the bottom or top quartile of the ITP capital distribution and a dummy for having large non-ITP pension income. All specifications include year fixed effects and the set of control variables used in specifications (2) and (4) of Table 5. The sample is restricted to claims made at age 65 between 2008 and 2010. Standard errors in parentheses

$^{+}$ Statistically significant at the 10 % level

* Statistically significant at the 5 % level

** Statistically significant at the 1 % level

From the results in Table 5, it is difficult to draw conclusions about the effect of being female *per se*. The estimates reveal almost zero marginal effects of being female on the demand for the 5-year option and a small negative effect on the demand for the 10-year payout. Thus, the raw gender difference of 4 percentage points in the demand for the 5-year payout that we saw in Sect. 4.3 can be explained by other variables, such as income and retirement wealth.[38] The high demand for fixed-term payouts among women is surprising given that the money's worth of the life annuity is considerably higher for women than for men.

Labor market participation at the time of withdrawal has a negative impact on the probability of choosing any of the fixed-term payouts. The marginal effect on the demand for 5- and 10-year payouts of being classified as working after claiming amounts to −1.7 and −1.1 percentage points, respectively. Individuals who simultaneously receive labor and pension income may find fixed-term payouts unattractive for tax reasons.

Education also affects the payout decision. Retirees with low education (elementary school) are 1.0–1.4 percentage points more likely to choose the 5-year payout. Retirees with a college degree are less likely to choose the 5-year payout with an impact at the margin of about 1.3 percentage points. The positive correlation between education level and annuitization rates might reflect higher life expectancy among college graduates. Education is less important in explaining the demand for 10- and 15-year payouts.

Finally, individuals who retire before the normal retirement age are much less likely to annuitize. The estimated coefficient reveals that "early claimers" are 17 percentage points more likely to choose the 5-year payout and about 4.3 percentage points more

[38] Excluding the two retirement wealth variables from specification (1) yields a significant estimate of female of 0.035.

likely to choose the 10-year payout than the life annuity. Chalmers and Reuter (2012), who find similar results for public servants in Oregon, suggest that such behavior might be explained by younger retirees using the lump sum to acquire new skills and re-renter the labor force. Among white-collar workers in Sweden, however, the combination of short payout horizons and early retirement is more likely driven by individuals who want to maximize their consumption possibilities during the years of early retirement. Compared to those who claim at age 65, early retirees receive larger benefits from the ITP plan, have significantly higher permanent income and are in better health.[39]

6 Conclusion

This paper provides empirical evidence on the determinants of annuitization. The study is based on a unique data set which includes real payout decisions of about 183,000 white-collar workers in Sweden who started to withdraw their second pillar retirement wealth between 2008 and 2013. The data are supplied by the default managing company for the occupational pension plan for white-collar workers, ITP. The data are merged with national administrative data to obtain rich individual background information. The main analysis focuses on salient and easily understood variation in the value of life annuities and fixed-term payouts due to differences in retiree characteristics.

Participation in the second pillar is mandatory. The payout decision involves large amounts of money, focusing on retirement wealth in the defined benefit component, ITP2. Retirees face five payout options, including a life annuity and four fixed-term payouts withdrawn over 5, 10, 15 or 20 years, respectively. Fixed-term payouts should be an attractive alternative to the life annuity for individuals who prefer to invest the money on their own or enhance consumption possibilities during the early years of retirement. Approximately 76 % of the individuals in the sample choose to annuitize their pension wealth. This is a surprisingly high annuitization rate given that the EPDV of the fixed-term payouts is at least as high as that of the life annuity. Two-thirds of all fixed-term payouts are accounted for by the 5-year payout option.

I find clear evidence of adverse selection of shorter-lived individuals and individuals in bad health into the most liquid payout option. All else equal, expos mortality is higher among retirees who choose the 5-year payout than among annuitants. In fact, individuals who choose the 5-year payout are 59 % more likely to die within two years after claiming than annuitants. These individuals also have higher sickness absence rates prior to retirement. I find a negative relationship between same-sex parent longevity and the probability of choosing the 5-year payout. This suggests that individuals form expectations about how long they are likely to live based on the lifespan patterns of their parents and take this into account when they decide whether to annuitize or not.

[39] Summary statistics for these two groups are provided in the Online Appendix (Table A1). The online appendix also provides results from several predictions that are made to illustrate the findings in Table 5. Specifically, I use the coefficient estimates that underlie the marginal probability effects to predict probabilities of choosing each payout option for individuals with different attributes.

Compared to other similar settings that have been studied in the literature, annuitization rates vary more by account balance, but less by individual characteristics such as marital status, gender and income. Individuals in the top account balance quartile are 14–15 percentage points more likely to choose the life annuity than the 5-year payout, whereas low accumulation of retirement assets is strongly associated with choosing the 5- or 10-year payouts. One potential explanation for this pattern, also found in Bütler and Teppa (2007) and Hurd and Panis (2006), is differences in the rate of time preferences across individuals. Another one is that small outcomes are discounted at a higher rate than greater ones are.

Taxes are also likely to play an important role here. Due to the progressivity of the tax schedule, the effective marginal tax rates decreases with the length of the payout. I use tax-induced variation in the money's worth ratio (MWR) of the life annuity with respect to the 5-year payout to study the impact of the annuity price on the decision to annuitize. In the preferred specification, a one percentage point increase in the net MWR of the life annuity is associated with a 0.96 percentage point decrease in the demand for the 5-year payout. The negative tax effect on the expected present discounted value of the fixed-term payouts is particularly large for high-income individuals with large capital stocks. This could explain why demand for the 5-year payout is so low among this group. Demand for the 5-year payout is also low among retirees whose total retirement income exceeds the state income tax threshold only under the 5-year payout. This suggests that individuals evaluate the benefit's net-of-tax value for different payout lengths, rather than it's gross value, at least when the tax effect is large and salient.

High demand for the life annuity may also be driven by the fact that individuals are defaulted into annuitization if they take no action before age 65. Someone who sticks with the default may do so because he or she interprets the default provision as implicit advice from the plan sponsor, or simply because he or she is too uninformed about or even unaware of the imminent payout decision. Future research should try to identify the causal effect of defaults on payout decisions. Existing evidence of the role of defaults is only suggestive and inconclusive.[40]

Acknowledgments This paper has benefited from comments by Sören Blomquist, Håkan Selin, James Poterba, Jeffrey Brown, Jonathan Reuter, Paolo Sodini, Mårten Palme, Svetlana Paschenko, Tarmo Valkonen, Raymond Gradus, Jacob Lundberg, Ponpoje Porapakkarm, Yuwei De Gosson de Varennes, Maria Polyakova, Jean-Marie Lozachmeur, Matthew Zaragoza-Watkins, Mikael Elinder, Per Johansson, Per Engström, Hannes Malmberg and seminar participants at the Swedish Social Insurance Inspectorate (ISF), the 2014 IIPF Congress, Uppsala Brown Bag, and the 2014 Topics in Public Economics in Uppsala. I also thank two anonymous reviewers and guest editor Monika Bütler. Special thanks to Alecta for providing data and to Pär Ola Grane, Nikolaj Lunin and Lars Callert for very helpful discussions. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

[40] Bütler and Teppa (2007) and Benartzi et al. (2011) suggest that defaulting workers into a life annuity would increase annuitization rates. In contrast, in an experimental study, Agnew et al. (2008) found that the presence of a default option did not affect payout decisions.

References

Agnew, J. R., Anderson, L. R., Gerlach, J. R., & Szykman L. R. (2008). Who chooses annuities? An experimental investigation of the role of gender, framing, and defaults. *The American Economic Review*, *98*(2), 418–422.

Ameriks, J., Caplin, A., Laufer, S., & Van Nieuwerburgh, S. (2011). The joy of giving or assisted living? Using strategic surveys to separate public care aversion from bequest motives. *The Journal of Finance*, *66*(2), 519–561.

Antolin, P. (2008). Policy options for the payout phase. In *OECD Working Papers on Insurance and Private Pensions No.25*.

Banerjee, S. (2013). Annuity and lump-sum decisions in defined benefit plans: The role of plan rules. *EBRI Issue Brief*, p. 381.

Barr, N., & Diamond, P. (2008). *Reforming pensions: Principles and policy choices*. Oxford: Oxford University Press.

Benartzi, S., Previtero, A., & Thaler, R. H. (2011). Annuitization puzzles. *The Journal of Economic Perspectives*, *25*(4), 143–164.

Bernheim, B. D. (1992). How strong are bequest motives? Evidence based on estimates of the demand for life insurance and annuities. *Journal of Political Economy*, *99*, 899–927.

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Behavioral economics perspectives on public sector pension plans. *Journal of Pension Economics and Finance*, *10*(02), 315–336.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., & Zeldes, S. P. (2014). What makes annuitization more appealing? *Journal of Public Economics*, *116*, 2–16.

Brown, J. R. (2001). Private pensions, mortality risk, and the decision to annuitize. *Journal of Public Economics*, *82*(1), 29–62.

Brown, J. R. (2007). Rational and behavioral perspectives on the role of annuities in retirement planning. *NBER Working Paper* 13537.

Brown, J. R., Ivković, Z., & Weisbenner, S. (2015). Empirical determinants of intertemporal choice. *Journal of Financial Economics*, *116*(3), 473–486.

Brown, J. R., Kling, J. R., Mullainathan, S., & Wrobel, M. V. (2008). Why don't people insure late-life consumption? A framing explanation of the under-annuitization puzzle. *American Economic Review*, *98*(2), 304–309.

Brown, J. R., & Poterba, J. M. (2000). Joint life annuities and annuity demand by married couples. *Journal of Risk and Insurance*, *67*(4), 527–553.

Bütler, M., K. Peijnenburg, & Staubli, S. (2011). How much do means-tested benefits reduce the demand for annuities? In *CESifo working paper: Empirical and Theoretical Methods, No. 3493*.

Bütler, M., Staubli, S., & Zito, M. G. (2013). How much does annuity demand react to a large price change? *The Scandinavian Journal of Economics*, *115*(3), 808–824.

Bütler, M., & Teppa, F. (2007). The choice between an annuity and a lump sum: Results from swiss pension funds. *Journal of Public Economics*, *91*(10), 1944–1966.

Chalmers, J., & Reuter, J. (2012). How do retirees value life annuities? Evidence from public employees. *Review of Financial Studies*, *25*(8), 2601–2634.

Davidoff, T., Brown, J. R., & Diamond, A. P. (2005). Annuities and individual welfare. *American Economic Review*, *95*(5), 1573–1590.

Dushi, I., et al. (2004). Household annuitization decisions: Simulations and empirical analyses. *Journal of Pension Economics and Finance*, *3*(2), 109–143.

Edmark, K., Liang, C.-Y., Mörk, E., & Selin, H. (2012). Evaluation of the swedish earned income tax credit. *IFAU Working Paper 2012*, p. 1.

Feigenbaum, J., Gahramanov, E., & Tang, X. (2013). Is it really good to annuitize? *Journal of Economic Behavior & Organization*, *93*, 116–140.

Finkelstein, A., & Poterba, J. (2002). Selection effects in the United Kingdom individual annuities market. *The Economic Journal*, *112*(476), 28–50.

Finkelstein, A., & Poterba, J. (2004). Adverse selection in insurance markets: Policyholder evidence from the UK annuity market. *Journal of Political Economy*, *112*(1), 183–208.

Fitzpatrick, M. D. (2012). How much do public school teachers value their retirement benefits?. *American Economic Journal: Economic Policy*, *91*(1), 35–53.

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351–401.

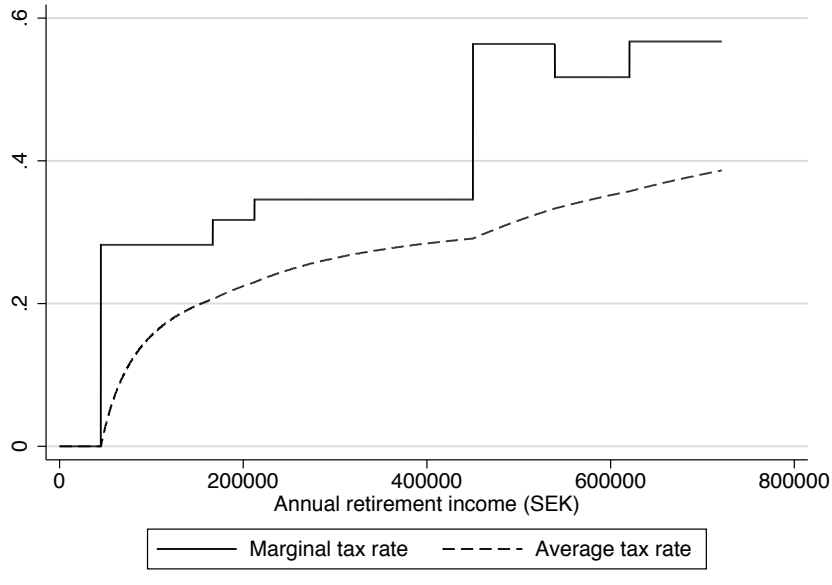
Friedman, B. M., & Warshawsky, M. (1988). Annuity prices and saving behavior in the united states. In *Pensions in the US Economy* (pp. 53–84). Chicago: University of Chicago Press.
Friedman, B. M., & Warshawsky, M. (1990). The cost of annuities: Implications for saving behavior and bequests. *Quarterly Journal of Economics*, *105*, 135–154.
Hagen, J. (2013). A history of the swedish pension system. *UCFS Working Paper No 2013:7*, Uppsala University.
Hurd, M., & Panis, C. (2006). The choice to cash out pension rights at job change or retirement. *Journal of Public Economics*, *90*(12), 2213–2227.
Inkmann, J., Lopes, P., & Michaelides, A. (2011). How deep is the annuity market participation puzzle? *Review of Financial Studies*, *24*(1), 279–319.
Kivimäki, M., Head, J., Ferrie, J. E., Shipley, M. J., Vahtera, J., & Marmot, M. G. (2003). Sickness absence as a global measure of health: Evidence from mortality in the Whitehall II prospective cohort study. *BMJ*, *327*(7411), 364.
Kopczuk, W., & Lupton, J. P. (2007). To leave or not to leave: The distribution of bequest motives. *The Review of Economic Studies*, *74*(1), 207–235.
Lockwood, L. M. (2012). Bequest motives and the annuity puzzle. *Review of Economic Dynamics*, *15*(2), 226–243.
Marmot, M., Feeney, A., Shipley, M., North, F., & Syme, S. (1995). Sickness absence as a measure of health status and functioning: From the UK Whitehall II study. *Journal of Epidemiology and Community Health*, *49*(2), 124–130.
Mitchell, O. S., Poterba, J. M., & Warshawsky, M. J. (1999). New evidence on the money's worth of individual annuities. *American Economic Review*, *89*(5), 1299–13018.
Mottola, G. R., & Utkus, S. P. (2007). Lump sum or annuity? An analysis of choice in DB pension payouts. *Vanguard Center for Retirement Research*, *30*(89), 2.
Pashchenko, S. (2013). Accounting for non-annuitization. *Journal of Public Economics*, *98*, 53–67.
Previtero, A. (2014). Stock market returns and annuitization. *Journal of Financial Economics*, *113*(2), 202–214.
SCB. (2010). *Cohort mortality in Sweden-mortality statistics since 1861*. Statistics Sweden.
Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, *13*(2), 144.
Van Solinge, H., & Henkens, K. (2009). Living longer, working longer? The impact of subjective life expectancy on retirement intentions and behaviour. *The European Journal of Public Health*, *29*, 47–51
Warner, J. T., & Pleeter, S. (2001). The personal discount rate: Evidence from military downsizing programs. *American Economic Review*, *91*(1), 33–53.
Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, *32*(2), 137–150.

Appendix

The income tax system in Sweden

Figure 1 provides a graphical representation of the Swedish income tax system discussed in Sect. 2.3 in the main text. Specifically, it shows the marginal and average tax rate as functions of retirement income for individuals who have turned 65 and have no labor income. The discrete jump in the marginal tax rate at SEK 450,200 marks the state income tax, which I pay particular attention to in the analysis of the impact of tax-induced variation in the relative price of the life annuity on annuity demand.

Figure 1. Marginal and average tax rate in 2013



Characteristics of early claimers

Table 1 reports summary statistics for individuals who claim their occupational pension at or after the normal retirement age of 65 and those who claim at earlier ages. We see that the individuals who claim early receive almost twice as large ITP2 benefits (SEK 6,994 vs. SEK 3,681), have significantly higher

permanent income (SEK 32,014 vs. SEK 23,223) and are in better health than those who retire at age 65. The mean sick leave days per year among early retirees is only half of that of those who retire at age 65.

Predictions

To illustrate the findings in Table 5, this section reports the probability of choosing each payout option for individuals with different attributes. These probabilities are computed using the coefficient estimates that underlie the marginal probability effects in Table 5.

Table 2 reports predicted probabilities for two hypothetical individuals. I consider "low-quartile" individuals who belong to the bottom permanent income quartile, received sickness benefits prior to retirement and attained no higher education level than elementary school, and "high-quartile" individuals who belong to the top permanent income quartile, never received sickness benefits and have a college degree. For low-quartile individuals, the predicted probability of choosing the life annuity is 72.3 %. For high-quartile individuals, the predicted probability is 80.7 %. Preference differences across the two groups are more visible among those who choose fixed-term payouts. The predicted probability of choosing the 5-year option for low-quartile individuals is 22.2 % compared to 9.4 % for high-quartile individuals. High-quartile individuals are also more likely to choose 10- and 15-year payouts than low individuals. These summary measures support the previous finding that individuals in good health and with high levels of lifetime income are more likely to choose the life annuity or any of the fixed-term payouts with long payout horizons.

To further evaluate the role of permanent income, Table 3 reports predicted probabilities of choosing each payout option for different income quartiles. The predicted probability of choosing the life annuity in the highest income quartile is 79.1 % compared to 75.0 % in the lowest income quartile. Individuals in the middle part of the income distribution are, however, less likely to annuitize. The demand for fixed-term payouts is more clearly related to income. The demand for the 5-year payout falls with the level of permanent income, whereas the demand for 10- and 15-year payouts increases with income. The most compelling result is that individuals in the top income quartile have significantly lower demand for the 5-year payout.

We know that the size of the ITP benefit plays an important role in the payout decision. Table 4 reports predicted probabilities of choosing each payout option for different account balance quartiles. Col. (1) reports the size of the ITP benefit under the life annuity option. The predicted probability of choosing the life annuity for individuals in the bottom quartile is 67.8 % compared to 82.9 % in the top quartile. The next column reveals considerable heterogeneity in the demand for 5-year payouts. 27.4 % in the bottom quartile are predicted to choose the 5-year payout compared to only 6.2 % in the top quartile.

Table 1. *Who are the early claimers?*

	(1) Normal retirement age	(2) Early claimers
Demographics		
Female	0.401	0.323
	(0.490)	(0.468)
Married	0.668	0.743
	(0.471)	(0.437)
Single	0.332	0.257
	(0.471)	(0.437)
Widowed	0.0464	0.0260
	(0.210)	(0.159)
Elementary school	0.208	0.178
	(0.406)	(0.383)
High school	0.555	0.593
	(0.497)	(0.491)
College	0.237	0.229
	(0.425)	(0.420)
Has children	0.740	0.786
	(0.439)	(0.410)
Pension income		
Monthly ITP2 benefit (DB)	3680.8	6993.7
	(5793.4)	(8523.3)
Bottom ITP capital quartile	0.275	0.0995
	(0.446)	(0.299)
Top ITP capital quartile	0.223	0.392
	(0.416)	(0.488)
Retirement age		
Claim age ITP2	65	62.97
	(0)	(1.024)
Claim age public pension	64.36	62.80
	(1.406)	(1.140)
Preretirement income		
Permanent income (monthly)	23222.7	32014.2
	(18602.9)	(22468.8)
Working	0.223	0.0648
	(0.416)	(0.246)
Health status		
Mean sick leave days per year	16.65	8.748
	(31.06)	(21.84)
15 sick leave days per year or more	0.265	0.152
	(0.441)	(0.359)
Observations	161117	21678

Note: Col. (1) reports summary statistics for individuals who claim their ITP pension at or after the normal retirement age of 65. Col. (2) reports summary statistics for those who claim at earlier ages.

Table 2. *Estimated probability of choosing each payout option, for selected characteristics*

	5-Year	10-Year	15-Year	Annuity
Low-quartile individual	.222	.0436	.0114	.723
High-quartile individual	.094	.0543	.0392	.807

Note: This table shows the combined relationship between different sets of individual characteristics and the probability to choose each payout option using the coefficient estimates that underlie the marginal probability effects of specifications (1), (3) and (5) in Table 5. Individuals are split into two groups: "low-quartile" individuals belong to the bottom permanent income quartile, received sickness benefits prior to retirement and attained no higher education level than elementary school, and "high-quartile" belong to the top permanent income quartile, never received sickness benefits and have a college degree. The remaining variables are evaluated at their sample means.

Table 3. *Estimated probability of choosing each payout option, by permanent income*

	Permanent income (monthly, SEK)	5-Year	10-Year	15-Year	Annuity
Quartile 1	6,710	.194	.0419	.0138	.75
Quartile 2	18,292	.188	.0534	.0191	.739
Quartile 3	26,329	.162	.0586	.0255	.752
Quartile 4	45,929	.102	.0608	.0417	.791

Note: This table shows the predicted probability of choosing each payout option by permanent income quartile. Col. (2) reports the average monthly income between age 51 and retirement.

Table 4. *Estimated probability of choosing each payout option, by pension amount*

	Monthly pension (SEK)	5-Year	10-Year	15-Year	Annuity
Quartile 1	283	.274	.0395	.00817	.678
Quartile 2	1,192	.15	.0563	.0214	.771
Quartile 3	2,593	.161	.0623	.0237	.752
Quartile 4	8,994	.062	.0572	.047	.829

Note: This table shows the predicted probability of choosing each payout option by pension amount quartile. Col. (2) shows the size of the ITP benefit under the life annuity.

III. What are the health effects of postponing retirement? An instrumental variable approach

Acknowledgments: I thank James Poterba, Sören Blomquist, Håkan Selin, Mårten Palme, Karin Edmark, Jonathan Reuter, Sergio S. Urzua, Josef Zweimüller, Hannes Malmberg, Matthew Zaragoza-Watkins, Ludovica Gazze, Daan Streuven, Kathleen Easterbrook, Bart Zhou Yueshen, Ponpoje Porapakkarm, Johan Wikström, Per Engström, Arizo Karimi, Daniel Waldenström, Per Johansson, Adrian Adermon, Daniel Hallberg and seminar participants at IIPF 2015, IFAU, Sudswec, Uppsala Brown Bag and the Interdisciplinary Workshop on Ageing and Health in Uppsala for their comments. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

1 Introduction

Many countries have responded to increasing life expectancy by raising retirement age thresholds while others have announced future increases.[1] The key rationale for such reforms is to improve the fiscal stability of pension systems through increased labor force participation rates among older workers. However, critics argue that the positive consequences must be weighed against the potential adverse effects of working longer on health. If workers are unable to work until the raised retirement age or if their health deteriorates at faster rate due to continued work, the fiscal burden might simply be shifted from the pension system to other parts of the welfare system. Understanding the health effects of retirement age increases, in conjunction with longer working lives, is therefore a crucial issue in pension policy design.

Empirically investigating the causal effects of retirement on health is a difficult task because the retirement decision is endogenous to health. Workers in good health are more likely to retire late, meaning that the simple correlation between health and retirement is likely to be negative. To properly assess the effect of retirement itself on health we need independent variation in retirement timing. The most credible approach is to use quasi-experimental variation in retirement induced by some policy change. While a number of studies have used variation from reforms that make early retirement more attractive, the evidence from reforms that promote longer working lives is surprisingly scarce.[2] The general result from studies that look at reforms promoting early retirement suggests that increasing the retirement age would contribute to a deterioration in population health.[3] However, such generalizations may be misleading if the potential effects of a change in the actual retirement age due to an increase in the retirement age are different from the corresponding effect that follows from introducing more generous early retirement rules. Early retirement reforms often contain elements of involuntary retirement, which makes it difficult to separate the potential effects of the reform itself from those of a change in the actual retirement age. They also target select groups of workers in industries or occupations in need of re-structuring.

This paper uses exogenous variation in retirement timing that comes from a large reform that increased the normal retirement age (NRA) from 63 to 65 for Swedish local government workers to assess the physical and mental health

[1] See e.g. Feldstein and Siebert (2009) and Holzmann (2005) for a discussion of recent pension reforms around the world and Andersen et al. (2014) for a focus on the Nordic countries.

[2] Atalay and Barrett (2014) exploit variation across birth cohorts in the eligibility age for women from the Australian 1993 Age Pension reform and find that retirement has a positive impact on health. Lalive and Staubli (2014) find no strong evidence of an effect on mortality from a Swiss reform that raised women's full retirement age from 62 to 64.

[3] Coe and Lindeboom (2008); Neuman (2008); Bloemen et al. (2013) and Hallberg et al. (2014) find that (early) retirement is associated with an improvement in well-being. A notable exception is Kuhn et al. (2010), who find that access to more generous early retirement rules increased mortality among male blue-collar workers in Austria.

effects of postponing retirement. Prior to year 2000, these workers could retire at age 63 with full pension benefits and an average replacement rate of 73 percent. As of 2000, those born before 1938 could continue to retire under the old rules, but those born in or later than 1938 had to work until the age of 65 to claim a full benefit. The new rules incentivized these workers to retire later as each month of retirement before age 65 implied a benefit reduction of 0.4 percent. The reform caused a remarkable shift in the retirement distribution, increasing the actual retirement age by more than 4.5 months. Given that the reform was announced in 1999, only one year prior to its implementation, workers had little opportunity to increase their retirement income in ways other than retiring later. I estimate the health effects of postponing retirement in an Instrumental Variable (IV) framework, where retirement is instrumented by several interaction terms between being born in 1938 or later and working in the local government sector. Because there were very few men in the affected worker categories, the analysis focuses exclusively on women. The control group is made up of female private sector workers of similar age. These workers experienced no major change in retirement incentives during the period of study and are similar to the local government workers along several background covariates.

I study health outcomes up to 11 years after the implementation of the reform using detailed Swedish administrative data on health care utilization and mortality. I study two major types of health care utilization: consumption of prescription drugs and hospitalizations. The drug register contains the universe of prescription drug purchases with information on the exact substance, the defined daily dose (DDD) and the date the drug was prescribed. The hospitalization data contains information about the arrival and discharge date and diagnoses codes for each hospitalization event at Swedish hospitals. The health data are then merged to administrative data on individual demographics and labor market status. Retirement is defined on a monthly basis using information from employers' employment records.[4]

Access to cause-specific data on mortality and health care utilization is valuable because retirement has been shown to be empirically related to various aspects of both physical and mental health. I examine several medical causes based on their known relationship with retirement in the previous medical and health-economic literature. Regarding physical health, I first look at the relationship between retirement and diseases of the circulatory system. These conditions can be related to stress and are often caused by correctable health-related behavior, such as an unhealthy diet, lack of exercise, being overweight and smoking. This analysis is complemented by an examination of health events that are directly related to alcohol and tobacco consumption as well as type 2 diabetes. I then study musculoskeletal system diseases and the con-

[4]A similar definition of retirement is used in the Austrian context by Kuhn et al. (2010) and Manoli and Weber (2014).

sumption of analgesic drugs (painkillers) to investigate whether postponing retirement has an effect on physical body functions.[5] I also study the effects of working longer on the consumption of mental health drugs with a particular focus on drugs that treat depression, anxiety and sleeping disorders.[6]

Results from the IV analysis are reported separately for drug prescriptions, hospital admissions and mortality. I document no effect on the probability of being prescribed a non-zero quantity of drugs, nor on total drug consumption. There is also no effect on the probability of being hospitalized due to any cause, nor on the number of days spent in hospital. Moreover, tracking mortality up to year 2011 (age 76 for the oldest cohort), I fail to reject the null hypothesis of no causal effect of working longer on mortality. The estimates are precisely estimated, which allows me to bound the effect sizes to a narrow range around zero.

The overall zero effect breaks down when I look at health care related to specific medical causes. First, I show that postponing retirement causes a reduction in the consumption of drugs that treat anxiety. The effect is quantitatively important with one year of extra work reducing the number of prescribed packages of anxiolytics by 24 percent. Since there is no effect for the extensive margin, i.e. the probability of being prescribed a non-zero quantity of anxiolytics, this result suggests that postponing retirement might alleviate short-term anxiety and depressive symptoms among elderly with pre-existing mental health issues. This interpretation is supported by the heterogeneity analysis, which shows that the reduction in the consumption of anxiolytics is driven by workers with previous sickness absence. Second, I show that postponing retirement significantly reduces the probability of being hospitalized due to diabetes. There is also evidence, although less robust, that postponing retirement reduces the risk of being treated for diseases of the circulatory system. Taken together, the two latter results suggest that continued work at older ages might provide individuals with better opportunities to preserve a healthy

[5]The empirical evidence on the effect of retirement on these aspects of physical health is mixed. Retirement has been shown to increase the risk of both heart disease (Behncke, 2012), stroke (Moon et al., 2012), obesity (Godard, 2016) and diabetes (Dave et al., 2008). In contrast, Bloemen et al. (2013) and Hallberg et al. (2014) report that retirement reduces the risk of heart-related mortality, and Insler (2014) shows that the observed beneficial influence of retirement on health could be explained by a reduction in smoking. These results reflect the fact that retirement might impact health in many, perhaps conflicting, ways. On the one hand, new retirees may lose some incentive to invest in their health, as their income is no longer dependent on health. Retirement might also lead to a general decline in physical activity if work constitutes the primary form of exercise. On the other hand, retirees have more leisure time with which to engage in physical activity or healthier diets.

[6]Just as for physical health, the direction of the effect of retirement on mental health is not clear a priori. On the one hand, retirement may have a positive impact on mental health through increased sleep duration (Eibich, 2014; Vahtera et al., 2009) and diminished work stress (Midanik et al., 1995). On the other hand, retirement might increase the risk of social isolation and depression (Dave et al., 2008; Börsch-Supan and Schuth, 2010; Szinovacz and Davey, 2004).

lifestyle than retirement and hence reduce the risk of lifestyle diseases. The conclusion from the cause-specific analysis is that continued work is more likely to affect certain medical conditions than overall health.

Although the empirical framework is based on Swedish public sector workers, the results should be of more general interest. First, the worker categories that were affected by the reform are characterized by demanding work environment and relatively high rates of sickness absence, including personal care workers, nursing professionals, cleaners and restaurant service workers. The focus on low- to medium-paid public sector jobs is relevant from a policy perspective since various discussions of increasing the retirement age thresholds deal primarily with the concern that such increases could adversely affect individuals in low-skilled jobs. Second, since retirees have equal access to publicly provided health service and medical care as employed individuals, the estimates are likely to capture the direct effect of working longer on health care utilization and mortality, rather than indirect effects, such as access to health insurance. The effects are also unlikely to operate through a loss of income as the long-run effect of the reform on disposable income is small. Finally, health care utilization is arguably the most important health dimension in estimating the fiscal impact of reforms that promote longer working lives. In 2014, individuals aged 65 and over comprised 20 percent of the Swedish population, but they accounted for 40 percent of total drug prescriptions and 47 percent of all patient discharges from public hospitals (Socialstyrelsen, 2015a,b).

This paper relates more broadly to a literature that tries to estimate the causal effect of retirement on health. To get a picture of the results in this literature, Table A.2 gives a brief summary of the empirical methods and key findings of 23 selected articles in the health-economic literature. Two of these studies report zero effects of retirement. Eleven studies report that retirement has a positive effect on health, whereas the remaining ten conclude that retirement is in fact associated with a decline in health. Even though these papers differ along several important dimensions, such as the population being studied, health outcomes and empirical methodology, these contrasting results are also likely to stem from the lack of convincing empirical strategies to deal with endogenous selection into retirement.

The most frequently used instrument is age-specific retirement incentives, such as early retirement windows or eligibility age-thresholds. This strategy has been used both in cross-country studies (e.g. Rohwedder et al. (2010); Sahlgren (2012); Mazzonna and Peracchi (2012); Coe and Zamarro (2011); Godard (2016)) and in within-country studies (e.g. Charles (2004); Bound and Waidmann (2007); Coe and Lindeboom (2008); Neuman (2008); Bonsang et al. (2012); Gorry et al. (2015)). The identifying assumption is that the instruments affect health only indirectly through their effects on the age of retirement. This is a strong assumption because workers who are subject to different retirement rules may differ with respect to unobserved variables (Kuhn et al., 2010).

The remaining part of the paper is organized as follows. Section 2 provides a theoretical framework for health and retirement. Section 3 discusses the details of the reform and section 4 describes the data. The econometric framework is outlined in section 5 and sections 6 and 7 present first and second-stage results, respectively. Section 8 presents additional results and section 9 concludes.

2 Theoretical framework

In this section I formulate a lifetime labor supply model of health and retirement. The purpose of this simple model is to illustrate how individuals evaluate both the financial effects and the health effects of working longer when they choose the optimal date of retirement. The role of health is highlighted by extending the basic lifetime labor supply model without health by making the length of life endogenous to the retirement age. The model does not yield any predictions about the sign of the effect of retirement on health, but provides framework for thinking about how we can isolate the health effects of retirement through exogenous variation in the budget constraint.

Individual preferences are defined over two goods, consumption c and work l. Utility is assumed to be additively separable so that $u_t(c_t, l) = v(c_t) - \phi_t \times l$ where ϕ_t is the disutility from working in period t and l takes the value one if the individual works in that period and zero otherwise. Letting the interest rate and the discount rate equal zero, the lifetime utility is given by

$$\int_{t=0}^{R} [v(c_t) - \phi_t]\, dt + \int_{t=R}^{T(R)} v(c_t) dt \tag{1}$$

where $T(R)$ is the last period of life. I make retirement and the length of life endogenous by assuming that the length of life depends on the age of retirement, that is, T is written as a function of R, $T(R)$. The lifetime compensation is the sum of wage earnings net of pension contributions, and pension wealth, where pension wealth is the sum of pension income collected in retirement. The budget constraint for an individual who participates in a defined benefit pension system can be written as

$$C \equiv \int_{t=0}^{T(R)} c_t dt = \int_{t=0}^{R} w_t (1-\tau)\, dt + \int_{t=R}^{T(R)} B(R)\, dt \tag{2}$$

where C is lifetime consumption and τ is the contribution rate. Using Leibniz rule, the slope of the budget constraint is then given by

$$\frac{dC}{dR} = w_R \times (1-\tau) + \frac{d}{dR} \int_{t=R}^{T(R)} B(R)\, dt \tag{3}$$

$$= w_R \times (1-\tau) - B(R) + \int_{t=R}^{T(R)} \frac{\partial B}{\partial R}\, dt + \frac{\partial T(R)}{\partial R} B(T(R),R)$$

The first term in Equation 3 captures the extra wage earnings that arises from postponing retirement R by one year. The total change to pension wealth for this additional year of work, referred to as the pension wealth accrual, is measured by the second and third terms of Equation 3. The second term is the forfeited pension benefit that could have been collected had the individual retired in the current year. The third term is the change in the size of the pension benefit accumulated over the retirement period. These two terms are determined by the rules of the pension system.

The last term in Equation 3 reflects another source of dynamics in this model, namely how work choices may affect health and the length of life. More specifically, this term illustrates the net effect of retirement on health (length of life), scaled by the pension benefit $B(T(R),R)$. We do not want to impose the sign of this term a priori. If health is positively affected by later retirement, this term will be positive and lifetime consumption increases for a given retirement age. Conversely, if the effect is negative, lifetime consumption decreases for a given retirement age, which promotes earlier retirement. Thus, even though the individual is incentivized to postpone retirement through high monetary gains, she might still choose to retire if continued work has a large negative effect on health. This means that changes to the parameters of the pension system will not only affect work decisions, but potentially also the health condition of the retirees.

A simple example can illustrate the relationship between health and retirement more clearly. Suppose that the wage is the same in all periods of work so that $w(t) = w$, $\forall t$. Furthermore, assume that the pension benefit is given by $B(R) = b$, $\forall t$ and that it is not financed contributions ($\tau = 0$). Further assume that $T(R) = \alpha + \beta R$ and that $R^* < T(R)$, where R^* is the optimal retirement age. If $\beta > 0$ later retirement had a positive impact on health. The lifetime budget constraint becomes $\int_{t=0}^{\alpha+\beta R} c dt = (w-(1-\beta)b)R + M$ where $M = \alpha b$ and its slope $\frac{dC}{dR} = w-(1-\beta)b$. Assuming $v(.)$ is concave with respect to c_t, the individual will maximize utility for any retirement date by perfectly smoothing consumption over the lifecycle so $c_t = \frac{C}{T(R)}$ for all t. Then the lifetime utility can be written as $T(R) \times v\left(\frac{C}{T(R)}\right) - \int_0^R \phi_t dt$. Substituting for $T(R)$ and the lifetime consumption C, an individual who chooses the optimal retirement age R solves the following maximization problem:

$$\max_R U(R) = (\alpha+\beta R) \times \psi\left[\frac{[w-(1-\beta)b]R + M}{\alpha+\beta R}\right] - \int_0^R \phi_t dt \qquad (4)$$

The solution to this problem will give us $R = R(w,b,\alpha,\beta,M)$. The empirical part of the paper aims at estimating β, that is, how retirement affects health,

by exploiting an exogenous change in the individual's budget constraint that is not related to health.

Before turning to the empirical analysis, I will make some remarks about how this model relates to the Grossman (1972) model of health production. In the Grossman (1972) model, health is both a consumption good that yields direct utility and an investment good that yields utility indirectly through increased productivity and higher wages. In my model, health has a more subtle, yet equally important, role for the retirement decision: health increases utility by raising the length of life T and thereby lifetime consumption. It is clear that allowing for a separate health process, which the individual can influence in many different ways, is more realistic than just allowing individuals to affect their health condition through their retirement decision. However, adding health as a separate argument in the utility function, modeled by some health production function, would not alter the model's basic prediction: a retiring individual will consider both the monetary gains and the health consequences of continued work when she decides on the optimal retirement age.

3 The occupational pension system

3.1 Retirement benefits in Sweden

Sweden's pension system has two main pillars, a universal public pension system and an occupational pension system. Swedish retirees generally receive most of their pension income from the public pension system, but the occupational pension system is an important complement. The occupational pension system consists of a number of different pension plans that are negotiated at the union-level and cover large group of workers. In fact, the four largest agreement-based occupational pension plans cover around 90 percent of the total work force. These include the pension plan for blue-collar private sector workers, white-collar private sector workers, local government workers and state-level government workers, respectively. The focus of this study is the pension plan for local government workers. The control group is made up of private sector workers. For the cohorts being studied, i.e. 1935 to 1942, there were no major changes in the private sector pension plans.[7]

[7]There are two large occupational pension plans in the private sector: one for blue-collar workers (SAF-LO) and one for white-collar workers (ITP). The ITP plan was mainly defined benefit and the same rules applied to all birth cohorts studied in this paper. The SAF-LO plan, on the other hand, is a pure defined contribution scheme. The implementation of the SAF-LO plan in 1996 implied that blue-collar workers born between 1932–1967 were subject to special transitional rules. However, cross-cohort differences in retirement incentives are minor because of the long transition period. See Hagen (2013) for a more detailed description of these pension plans.

3.2 The occupational pension reform for local government workers

The pre-reform occupational pension plan for local government workers, called PA-KL, covered local government workers born before January 1, 1938. PA-KL was defined benefit and directly coordinated with the public pension system. PA-KL stipulated that the sum of the annual occupational pension benefit and the public pension benefit should amount to a certain fraction of the individual's pre-retirement income. The occupational pension would always pay out the residual amount net of the public pension benefit to reach a certain replacement rate. In year 2000, the gross replacement rate amounted to 73 percent for a female local government employee with an average wage rate who retired at the age of 63. If the public pension accounted for 60 percent, the occupational pension benefit would amount to 13 percent of her qualifying income. Thus, local government workers only needed to know about the gross replacement rates to get a full picture of their retirement income.[8]

In the pre-reform pension plan, retirement was mandatory for everyone at the age of 65 unless the employer offered a prolongation. However, the age at which full or unreduced retirement benefits could be withdrawn, i.e. the normal retirement age (NRA), was different for different occupations. The NRA was either 63 or 65. Early withdrawals could be made from the age of 60, but the penalty rate at a given claiming age, i.e. the reduction in the gross replacement rate, was different depending on what NRA the worker faced. Here, I focus on workers who had a NRA of 63.

Workers who faced a NRA of 63 could retire at this age with a full benefit. The benefit was not actuarially increased for claims made after 63, which means they had little incentive to work past this age. Selin (2012) shows that these workers lost SEK 169,000 (USD 1 = SEK 7.2 in 2010) in pension wealth from continuing working an additional year after turning 63. Broad categories of workers had a NRA of 63, including personal care workers, nurses, preschool teachers, restaurant service workers and cleaners.

In 1998 a new agreement, PFA98, was signed for Swedish local government workers. The most important change was that the NRA was set to 65 for all local government workers. This was achieved by introducing equal early retirement penalty rates for all occupations. The new penalty rates implied that the pension was reduced by 0.4 percent per month of retirement before age 65. Rather than receiving a full benefit, retiring at age 63 as compared to 65 now implied a substantial benefit reduction of 9.6 percent (0.4*12*2=9.6).

The reform implied two additional changes to the pension plan. First, there was a partial shift from defined benefit to defined contribution. For earnings below the ceiling of 7.5 increased price base amounts, the pension was en-

[8]The appendix provides more information on this pension plan. Selin (2012) has used this reform to study spousal spillover effects on retirement behavior.

tirely defined contribution.[9] The contribution differed slightly over time and also between employers and type of tenure, but centered on 3.4–3.5 percent for wage portions below the income ceiling and 1–1.1 percent for earnings above. Individuals with earnings above the ceiling got an additional defined benefit pension.[10] The individual could always increase her pension wealth by postponing retirement until the age when she was obliged to retire or until no more pension rights could be earned. Second the occupational pension was not residual to the public pension anymore, but paid out as a separate entity. Workers in the new plan were thus directly exposed to the early retirement penalty rates in the public pension system.[11]

The new PFA98 agreement came into effect on January 1, 2000 for those born in 1938 or later. Those born in 1937 and earlier were completely unaffected by the occupational reform and would still be covered by the old plan. The reform was implemented rather quickly and without much media coverage.[12] The purpose of the reform was to harmonize rules across all worker categories in the local government sector and to adjust rules to the new defined contribution structure.

While the reform substantially increased the incentives to postpone retirement beyond the age of 63, it did not change the stock of already accumulated occupational pension wealth. The reason for this was a transition rule that would compensate workers in post-reform cohorts for potential benefit reductions due to the new rules. The pension wealth earned up to December 31, 1997 was converted into a life annuity that corresponded to the annual pension benefit that the individual would have received if she had retired by that date. Pension rights earned after this date were accredited the new pension plan. If the resulting pension from these two components was lower than the corresponding pension in the absence of a reform, workers received the difference from the employer. As a result, the pension wealth at age 65 was more or less unchanged for the transition cohorts. Importantly, workers who retired before 65 were not eligible for this compensation, which implies that the most important effect of the reform was to raise the NRA from 63 to 65 for workers who had a NRA of 63 in the pre-reform pension plan.

It should be noted that the first post-reform cohort in the empirical analysis, i.e. those born in 1938, are also the first cohort to participate in the new public pension system. The 1938 cohort receives one-fifth of its benefit from the

[9] 1 increased price base amount equaled SEK 43,300 in 2010.

[10] This defined benefit component amounted to 62.5 percent of earnings between 7.5 and 20 base amounts and 31.25 percent between 20 and 30 base amounts.

[11] The monthly penalty rate in the public pension system was 0.5 percent.

[12] Selin (2012) reports that a search in the online press archive Presstext, which covers the biggest daily newspapers in Sweden, reveals that the first article mentioning PFA98 is written in the fall of year 2000. Low media coverage, however, does not rule out the possibility that the reform may have become known among the affected individuals through unions informing or word-of-mouth information.

new system and four-fifths from the old system. Each cohort then increases its participation in the new system by 1/20, so that those born in 1954 will participate only in the new system (Hagen, 2013). Benefits from the new system were paid our for the first time in 2001, three years after it was legislated. The transition rules apply to all individuals born after 1937 and are controlled for in the empirical analysis by including cohort fixed-effects in the regression model.

4 Data

4.1 Data on retirements

Individual demographics and labor market information is collected and maintained by Statistics Sweden. The *Longitudinal Database on Education, Income and Employment* (LOUISE) provides demographic and socioeconomic information. The data covers the entire Swedish population between 16 and 65 during the period 1987–2000, and individuals aged 16 to 74 between 2001–2010. The population of interest is local government workers whose NRA was increased from 63 to 65 in 2000. The main sample analyzed is composed of individuals born between 1935 and 1942. Those born in 1938 were the first ones to be affected by the new rules. The control group is made up of private sector workers in the same birth cohorts.

Importantly, there is information in the data which allows me to distinguish these workers from other workers in the local government sector who had a NRA of 65 both before and after the reform. I use the Swedish Standard Classification of Occupations (SSYK-96) to identify workers in occupations who had a NRA of 63.[13] Individuals who are observed working in any of these occupations between ages 61–63 are included in the treatment group. I identify workers with a NRA of 65 in the same way. If an individual is observed working in both occupation categories, I use the most recent observation to determine the NRA. SSYK codes are available from 1996, which means that those born in 1935 is the oldest cohort for whom occupation status is known at age 61.[14] I define someone as working in the private sector if she has not been employed in the public sector between ages 61–63.[15]

[13] *Standard för svensk yrkesklassificering* (SSYK-96). SSYK-96 is based on the International Standard Classification of Occupations (ISCO-88).

[14] It is not possible to determine the NRA for all local government workers. The NRA cannot be determined for SSYK codes that map simultaneously to occupations with different NRAs. For example, pre-school teachers and after-school teachers have the same SSYK code (3310), but different NRAs. I therefore restrict the treatment group to workers in occupations where the SSYK code maps exclusively to a NRA of 63.

[15] It is more difficult to determine private sector affiliation from the data. The data which contain the SSYK codes only contain a small representative sample of private sectors workers (around 23 percent). In contrast, the universe of public sector workers is included in this data.

I make four restrictions to the sample of local government and private sector workers born between 1935 and 1942. First, because the affected worker categories in the local government sector were dominated by women, male workers are excluded from the analysis. In fact, only 3 percent of these workers are men. Second, I restrict the analysis to individuals registered as employed for 12 full months in the year of their 61st birthday. This restriction is done for two reasons. First, it excludes individuals who exited the labor force early and whose retirement decision is unlikely to have been affected much by the reform. Second, it ensures that I observe at least one SSYK code for each local government worker. In order for the SSYK code to be reported, the individual must be employed during the "reference month", which typically occurs at the end of the year. In effect, this restriction implies that the first month in which individuals are allowed to retire is the month in which they turn 62. Third, I restrict the sample to individuals who have five years of consecutive employment prior to age 61 (at any work place). Finally, I also exclude individuals who are registered as self-employed at some point between ages 61–63. The final sample consists of 133,026 individuals of whom 57,415 are local government workers.

Table 1 reports the distribution of workers in the most numerous worker categories, and the corresponding SSYK codes, in the treatment and control group, respectively. The majority of the treatment group work within personal care. These include child-care workers, assistant nurses, home-based personal care assistants and dental nurses. Other important worker categories in the treatment group are restaurant service workers, nursing professionals and cleaners. The number of worker categories in the control group is larger since it includes both blue-collar and white-collar workers in the private sector. The most numerous worker categories in the control group are blue-collar jobs, including salespersons, plant and machine operators, manufacturing laborers and craft workers. White-collar workers are foremost represented in the categories "other associate professionals", "professionals" while "clerks" include both. Three of the treatment group occupations are found in the control group, too (personal care-related workers, restaurant service workers, and helpers and cleaners).

The retirement definition reflects the month in which an individual retires completely from the work force. This definition uses records of employment spells, which are obtained from the the *Register-Based Labor Market Statistics* (RAMS). The information in RAMS is based on reports that all employers submit to the Swedish Tax Agency. For each employee, the employer must report how much wages and benefits have been paid out, how much taxes have been drawn and, most importantly, during which months the employee has been employed at the firm. This information allows me to infer in what month and year an individual exits the labor market. The decision to retire is equated with the month in which the individual's last employer reports the employment contract to be officially ended. The outcome variable in the first-stage analysis

Table 1. Occupations in the treatment and control group

Treatment group (local government)		Control group (private sector)	
Occupation	SSYK-96	Occupation	SSYK-96
Personal care & related workers (64%)	513	Salespersons (31%)	52
Restaurant service, housekeeping (15%)	512, 913	Plant & machine operators (15%)	8
Nursing & midwifery professionals (13%)	223, 323	Clerks (16%)	4
Helpers & cleaners (8%)	912	Manufacturing laborers (6%)	932
Physiotherapists (< 1%)	5141	Helpers & cleaners (9%)	912
Hairdressers (< 1%)	3226	Craft & related trade workers (6%)	7
		Restaurant service, housekeeping (7%)	512, 913
		Other associate professionals (3%)	34
		Personal care & related workers (3%)	513
		Professionals (2%)	2

Note: The first column reports the occupations in the local government sector that had a NRA of 63 before the reform (the treatment group). The third column reports the most common private sector occupations (the control group). The corresponding SSYK codes are listed in the second and fourth columns, respectively. The share of workers in each occupation are reported in parentheses. A worker is classified into an occupation if she is observed working in that occupation at any time between ages 61–63. The occupations are therefore not mutually exclusive. SSYK codes are only available for a representative sample of the private sector workers. The shares in this group are adjusted for sampling probabilities.

on the retirement effects of the reform is defined as the number of months an individual is registered as employed between ages 62 and 68. The upper limit of age 68 is chosen because it is the oldest age to which the youngest cohort can be tracked.[16]

4.2 Data on health care utilization and mortality

I study mortality outcomes and two major types of health care utilization: hospitalizations and consumption of prescription drugs. Three register based data sources are used for this purpose.

The analysis of drug prescriptions are based on data from the *Prescription Drug Register*, which contains information about all over-the-counter sales of prescribed medical drugs between 2005–2009. For each occasion when a prescription drug was bought, the data contains detailed information about the the Anatomical Therapeutic Chemical (ATC) code of the drug, and the number of defined daily doses (DDDs) purchased over the entire period. The analysis of mortality is based on information from the *Cause of Death Register*. Causes of death are classified using the International Classification of Diseases (ICD). Hospitalizations are studied using information about inpatient care available in the *National Patient Register*.[17] For each hospitalization event, the register

[16]The employer records have been used in the Swedish context by Laun (2012). She studies the retirement effects of two age-targeted tax credits in 2007 using the number of remunerated months at age 65. Kreiner et al. (2014) use monthly payroll data on wages and salaries to study year-end tax planning in Denmark.

[17]Information on hospital admissions is provided by the National Board of Health and Welfare and covers all inpatient medical contacts at public hospitals from 1987 through 1996. From

has information about the arrival and discharge date, and diagnoses codes in ICD format. Inpatient records exist from 1964 to 2010, while the mortality data ends in 2011.

The analysis focuses on the extensive and the intensive margins of health care utilization. For the extensive margin, I define a set of binary outcome variables that equal 1 one if the individual consumes a non-zero quantity of drugs or is hospitalized for at least one night during a pre-specified time period. The intensive margin outcome variables for drug prescriptions are given by the product of the DDD per package and the number of prescribed packages, summed over the years 2005–2009 for each individual. Since DDDs are not directly comparable across drug types, I complement this analysis by looking only at the prescribed number of packages. Information on the number of days spent in hospital is used to construct intensive margin outcomes for inpatient care. The intensive margin adds important variation to the quantity of consumed health care, especially for individuals with previous records of health care utilization.

Because the different health registers cover different years, the pre-specified time period over which health outcomes are defined will vary across the type of health event. The mortality data ends in 2011, which means that the maximum age up to which all cohorts can be tracked is 69. The outcome variable is thus set equal to 1 if the individual died before reaching this age. To make use of all the data at hand, I also look at mortality by year 2011. In a similar fashion, the hospitalization outcomes are either based on an individual's hospital admissions between ages 65–68 or between age 65 and year 2010. The latter time period implies that the length of the follow-up period decreases with the year of birth of the individual. Age 65 is chosen as the lower age limit because our primary interest lies in estimating the effects on health care utilization after the individual is retired. Finally, all drug outcomes are based on prescriptions made between 2005–2009.

I adopt a framework used by Cesarini et al. (2015) to classify mortality and health care utilization events into a number of relevant medical causes. Specifically, I examine deaths and health care utilization events into two cause categories: common causes and hypotheses-based causes. The common causes are cancer, respiratory disease, diseases of the circulatory system and other. The hypotheses-based causes are chosen based on their appearance in previous economic and medical literature on the association between retirement and health. These include ischemic heart disease, hypertension, cerebrovascular disease, musculoskeletal disease, alcohol and tobacco consumption, injury and type 2 diabetes. While the common causes are the same, Cesarini et al. (2015) choose the hypotheses-based causes based on their known relationship with wealth rather than retirement.

1997 onward, the register also includes privately operated health care. Before 1997, virtually all medical care in Sweden was performed by public agents (Hallberg et al., 2014).

Mortalities and hospital admissions are readily classified into each of these causes using the ICD codes. Only primary diagnoses are used to classify hospital admissions and deaths into one of the common causes. These are therefore mutually exclusive. In the hypotheses-based causes, the cause-of-death or hospitalization variable is set equal to 1 if the condition matches at least one of the (first five) listed diagnosis codes on the discharge record or the death certificate. All codes are included because some of the hypotheses-based causes are rarely listed as the primary cause of death or primary diagnosis.

I use the ATC codes to classify prescription drugs into categories that closely resemble the common causes and hypotheses-based causes.[18] The prescription data is also used to define a number of (hypotheses-based) mental illnesses. These include depression, psychosis, anxiety, sleeping disorders and Parkinson. Table A.1 describes the aggregation of ATC and ICD codes for each cause.

4.3 Descriptive statistics

Table 2 shows descriptive statistics for pre- and post-reform cohorts for the treatment and control group, respectively.

The first row shows that post-reform cohorts in the treatment group retire about 5.5 months later than the pre-reform cohorts. The corresponding difference in the control group is very small, which yields a raw difference-in-differences (DD) estimate of 5.3 months. This is strong preliminary evidence that the reform had a positive impact on the retirement age. The second and third rows show that similar results are obtained for two alternative measures of retirement. According to the first alternative definition, individuals who receive a positive amount of pension income are classified as retired. The second definition is income-based. According to this definition, the individual retires the year before her annual earnings fall below 1 price base amount ($\approx$ USD 5,900 in 2010). Since these definitions are measured at the yearly level, the raw DD estimate in the second row of 0.56 reflects an increase in the claiming age of more than 6.5 months. The income-based definition of retirement reflects a smaller, yet sizable, effect of about 3 months.[19]

Table 2 shows that the two groups are similar in terms of several background characteristics, including marital status, the probability of having children (of

[18]The hypotheses-based classification is amended in several ways. First, following Cesarini et al. (2015), I merge ischemic heart disease and hypertension into a single category ("Heart") because many drugs are used to treat both these illnesses. Second, I limit the set of drugs used to treat musculoskeletal diseases (ATC code "M") to analgesics (painkillers). Third, I do not try to classify drugs into "Alcohol and Tobacco" due to the complexity of the prescription data.

[19]Individuals are allowed to be retired from the age of 56 according to these definitions, which helps explain why the average retirement ages implied by these definitions are lower than the average retirement age implied by the main definition. The sample restrictions explained in section 3 apply nonetheless.

any age) at home and immigrant status. The two groups also have similar pre-retirement health status. Sickness absence, measured as the number of years an individual has been absent from work for more than 14 consecutive days between ages 56–60, is only marginally higher in the treatment group, just like the probability of having been hospitalized during the same period.[20] Differences apply mainly to education level and pre-retirement earnings. Local government workers have, on average, 0.5–0.6 years more of schooling and somewhat higher pre-retirement earnings than the private sector workers. The income distribution of the local government workers is, however, more compressed.

Table 2 gives some idea about how these characteristics change over time in the treatment and control group, respectively. To estimate these changes more formally, I employ a DD framework and regress the characteristic of interest on several interaction terms between cohort $j = 1935, ..., 1942$ and the local government dummy as well as the local government dummy itself and cohort-fixed effects (the 1937 cohort as reference). The estimated interaction terms are reported in Table A.3. While the results give little reason to be concerned about differential trends with respect to most background characteristics and pre-retirement health, it is clear that the level of education in the treatment group increases at a slower rate than in the control group. The differential trend with respect to education generates important differences also in terms of pre-retirement earnings, at least for the three youngest cohorts. I will account for these trends in the regression analysis by including interaction terms between education/income and birth cohort and sector.

Going back to Table 2, the average occupational strain is somewhat higher in the treatment group, but the control group exhibits larger variation. This is expected since the control group includes workers in occupations with very high strain (e.g. plant and machine operators) and occupations with very low strain (e.g. business and legal professionals) while the majority in the treatment group work within personal care.

Turning to our health outcomes, we see that around 30 percent of the local government workers are hospitalized for at least one night between ages 65–68. This fraction increases to almost 60 percent when the follow-up period is extended to year 2010. The private sector workers exhibit very similar hospitalization rates. The differences amount to less than one percentage point. The two groups are also similar with respect to drug consumption. More than 90 percent of the individuals in the pre-reform cohorts are prescribed a non-zero quantity of drugs between 2005–2009. Around 40 percent are prescribed mental drugs. Note, however, that the cross-cohort decline in drug consump-

[20] Specifically, our measure of sickness absence is the number of years the variable "sjukpp" in the LOUISE database takes on a non-zero positive value between ages 56–60. "Sjukpp" includes sickness benefits that are paid out by the Swedish Social Insurance Agency (Försäkringskassan). The Social Insurance Agency is responsible for paying out sickness benefits to individuals who have been sick for more than 14 consecutive days.

Table 2. *Descriptive statistics*

	Local government workers				Private sector workers			
	Pre		Post		Pre		Post	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Retirement								
Employment, nr of months btw 62–68	36.19	(24.09)	41.56	(21.85)	41.30	(23.61)	41.42	(23.07)
Retirement age (claim age)	63.75	(1.976)	64.19	(1.808)	63.83	(2.478)	63.71	(2.501)
Retirement age (income-based)	63.23	(2.071)	63.75	(2.508)	62.95	(2.911)	63.23	(3.089)
Demographics								
Married	0.623	(0.485)	0.620	(0.485)	0.613	(0.487)	0.609	(0.488)
Single	0.0535	(0.225)	0.0592	(0.236)	0.0612	(0.240)	0.0702	(0.255)
Divorced	0.169	(0.375)	0.194	(0.395)	0.183	(0.387)	0.207	(0.405)
Widow	0.155	(0.362)	0.127	(0.333)	0.143	(0.350)	0.114	(0.318)
Immigrant	0.103	(0.304)	0.0886	(0.284)	0.112	(0.315)	0.108	(0.311)
Children at home	0.127	(0.333)	0.113	(0.317)	0.112	(0.315)	0.108	(0.311)
Years of schooling	10.30	(2.717)	10.69	(2.691)	9.724	(2.859)	10.38	(2.973)
Average physical strain	6.497	(0.915)	6.421	(0.925)	5.623	(2.559)	6.873	(2.035)
Average social strain	7.597	(1.233)	7.348	(1.326)	6.149	(3.086)	7.536	(2.501)
Income								
Log(average earnings 56–60)	11.92	(0.336)	12.04	(0.368)	11.88	(0.681)	12.04	(0.682)
Log(std. dev. earnings 56–60)	9.261	(0.838)	9.645	(0.802)	9.624	(0.987)	9.888	(0.965)
Pre-retirement health								
Hospitalized ages 56–60	0.254	(0.435)	0.240	(0.427)	0.250	(0.433)	0.241	(0.428)
Sickness benefits (years)	1.209	(1.269)	1.000	(1.235)	1.023	(1.236)	0.838	(1.196)
Health outcomes								
Prescribed any drug	0.927	(0.261)	0.944	(0.230)	0.925	(0.263)	0.949	(0.219)
Prescribed mental drug	0.397	(0.489)	0.358	(0.479)	0.403	(0.491)	0.364	(0.481)
Hospitalized ages 65–68	0.294	(0.456)	0.276	(0.447)	0.299	(0.458)	0.277	(0.448)
Hospitalized age 65–year 2010	0.578	(0.494)	0.355	(0.478)	0.579	(0.494)	0.352	(0.477)
Mortality by age 69	0.0453	(0.208)	0.0447	(0.207)	0.0450	(0.207)	0.0413	(0.199)
Mortality by year 2011	0.103	(0.304)	0.0561	(0.230)	0.103	(0.304)	0.0541	(0.226)
Observations	18561		38854		23895		51716	

Note: The sample includes female local government (treatment group) and private sector (control group) workers born between 1935–1942 who have five years of consecutive employment prior to age 61 (at any work place) and are registered as employed for 12 full months in the year of their 61st birthday. The sample of local government workers is restricted to workers in occupations whose NRA was increased from 63 to 65 in 2000. Earnings are in the 2010 price level. Retirement variables right-censored at age 68. Columns (1)–(4) display statistics for the treatment group, while columns (5)–(8) consider the control group. Pre-reform cohorts refer to those born before 1938. Average strain is derived from the occupational demand scales provided by Kroll (2011) by taking the average of the strain values associated with the jobs held prior to retirement.

tion and hospital admissions is larger in the treatment group than in the control group. The opposite pattern is seen for our two measures of mortality, i.e. the probability of being dead by the age of 69 or by year 2011. In sum, it is difficult to draw any conclusions about the existence of an effect of the reform on mortality and health care utilization based on these raw DD estimates.

5 Econometric framework

The primary interest of this paper is to estimate the health effects of postponing retirement. The regression model of interest can be written as:

$$y_i = \alpha + \beta R_i + \varepsilon_i \tag{5}$$

where y_i is a measure of health for individual i and ε_i is the error term. R_i is a continuous measure of labor force participation, which I define as the number of months in employment between ages 62–68 conditional on being

employed at the age of 61. The coefficient of interest is β, the causal effect of an additional month of employment on health. This coefficient is the empirical counterpart of the β term in the theoretical model in section 2. R_i is endogenous because individuals who retire later are more likely to be in good health.

To estimate the causal effect of continued work we need variation in retirement timing that is exogenous to health. For this purpose I utilize the above described reform, which raised the NRA from 63 to 65 for local government workers born in 1938 or later. I use an instrumental variable (IV) framework to assess the causal relationship between postponing retirement and health, where R_i is instrumented by an interaction term between being born in 1938 or later (post-reform cohorts, $CH = 1$) and being employed in the local government sector ($LG = 1$). This means that I estimate the causal effect for those individuals who postpone retirement due to the reform, i.e. the compliers. Assuming heterogeneous effects of postponing retirement on health, the 2SLS estimator estimates the local average treatment effect (LATE) instead of the average treatment effect (ATE).

For individual i in cohort j in sector s, the first-stage DD equation is then written as

$$R_{i,j,s} = \alpha + \delta \left(LG_s \times CH_{j\in[1938,1942]} \right) + \phi LG_s + \lambda_j + \mathbf{X}_{i,j,s}\theta + u_{i,j,s} \qquad (6)$$

where λ_j denotes cohort-fixed effects and the vector $\mathbf{X}_{i,j,s}$ is a set of control variables which includes the number of years of schooling, region of residence and month of birth fixed-effects, dummies for being single, divorced or widowed (married reference group), immigrant status and having children at home. The set of control variables also include the log of the average of yearly earnings, the log of the standard deviation of yearly earnings and the number of years with more than 14 consecutive sick leave days between ages 56–60. To account for differential trends in educational attainment/income, I also add interactions between years of schooling/income and cohort and years of schooling/income and local government. The coefficient of interest is δ, which reflects the reform effect on the number of months employed before exiting the labor market, comparing local government workers born in 1938 or later to private sector workers in the same birth cohorts. Differences in employment across the treatment and control group are captured by the term ϕ. The maximum value of $R_{i,j,s}$ is 72 because of right-censoring at age 68.

To capture the heterogeneity in the effect of the instrument on the first-stage outcome, I allow for cohort-specific effects by including interaction terms between the local government dummy and cohort j in the first-stage equation

$$R_{i,j,s} = \alpha + \sum_j \delta_j \left(LG_s \times CH_{j\in[1938,1942]} \right) + \phi LG_s + \lambda_j + \mathbf{X}_{i,j,s}\theta + e_{i,j,s} \qquad (7)$$

which implies there are five instruments for the main specification (post-reform cohorts: 1938–1942). Each cohort-specific reform effect δ_j is evaluated against a pooled sample of the pre-reform cohorts.[21] The IV model is just-identified if the first-stage equation is given by (6). The model is over-identified if the first-stage equation is given by (7), i.e. the number of instruments is larger than than number of endogenous regressors. The over-identified model better captures the causal effect of postponing retirement if there are differences across birth cohorts in their retirement response to the reform. The causal effect from the over-identified model averages IV estimates using the instruments one at a time, where the weights depend on the relative strength of each instrument in the first stage.

The reduced-form is given by replacing the dependent variable in any of the first-stage equations by a measure of health, $y_{i,j,s}$. The reduced-form equation for the just-identified model is then given by:

$$y_{i,j,s} = \alpha + \psi \left(LG_s \times CH_{j\in[1938,1942]} \right) + \phi LG_s + \lambda_j + \mathbf{X}_{i,j,s}\theta + u_{i,j,s} \qquad (8)$$

The reduced-form estimate ψ is referred to as the intention-to-treat effect. This specification is similar to a DD model, where we compare health outcomes of local government workers and private sector workers across birth cohorts.

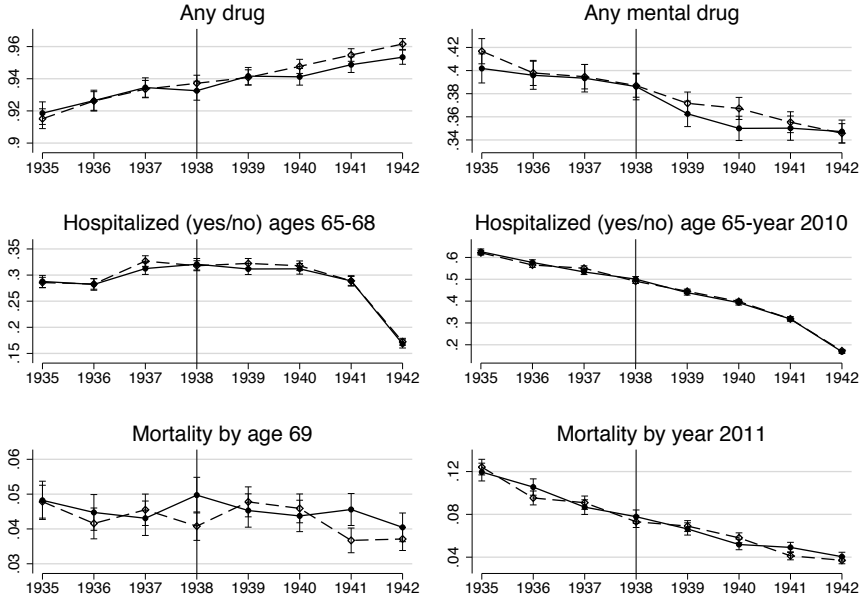
5.1 Identifying assumptions

Technically, identification requires two assumptions. First, the increase in the NRA must have an impact on the retirement age, that is $\delta \neq 0$ in Equation 6 (instrument relevance). This assumption is carefully analyzed in section 6. Second, exposure to the increase in the NRA affects health only through the number of months in employment (the exclusion restriction). This means that the instrument must be uncorrelated with the error term in the second-stage equation.

Concerns about the existence of other channels through which the reform might affect health can be alleviated if we can show that the parallel trends assumption hold. This assumption implies that the outcome variable evolved in the same way in the treated group as in the control group in absence of the reform. Figure 1 plots series of average outcomes for the treatment and control group before and after the reform for some of the key health measures. The two top panels show that the probability of being prescribed a non-zero quantity of drugs evolved similarly for pre-reform cohorts in the two groups. The lower panels show that post-retirement hospitalization and mortality also seem to satisfy the parallel trends assumption. It is also reassuring that the levels are similar across the two groups.

[21] This specification is analogous to the specification by Atalay and Barrett (2015) who study the impact of age pension eligibility age on retirement in Australia with the difference that they exploit variation in birth cohort and gender rather than birth cohort and sector as this paper does.

Figure 1. Comparing the treatment and control group



Note: This figure plots the means and the corresponding 95 percent confidence intervals of several health outcomes by cohort and treatment status. Solid and dashed lines refer to the treatment and control group, respectively. The confidence interval is obtained by regressing each health outcome on a constant, separately by cohort and treatment status.

We can test the parallel trends assumption more formally by estimating pre-reform trends in the reduced-form framework. Specifically, I extend Equation 7 by adding two interaction terms between the local government dummy and cohorts $j = 1936, 1937$. For the parallel trends assumption to hold, the estimated δ_j coefficients for these two cohorts should be close to zero and insignificant. Table 3 reports the estimation results for the extensive margin health outcomes shown in Figure 1 along with the corresponding intensive margin measures of health care utilization. The upper panel of Table 3 shows that the pre-reform trends with respect to drug consumption are similar. Only two of the coefficients are statistically different from zero (at the 10 percent level) and they also relate to different cohorts. The lower panel raises some concern of a positive trend in the health status of the treatment group individuals as two of the estimates for the 1936 cohort are positive and significant; the probability of having been hospitalized between age 65 and year 2010 and mortality by year 2011. However, given that all other hospitalization outcomes are insignificant and the relatively large year-to-year fluctuations in mortality, I conclude that the assumption of parallel trends in mortality and health care utilization seems to hold.

The exclusion restriction could be violated if the reform coincides with differential trends in occupation-specific work environment. One concern for the

Table 3. *Estimating pre-reform trends in health care utilization and mortality*

	All drugs			Mental drugs		
	(1) Any	(2) Dose	(3) Packages	(4) Any	(5) Dose	(6) Packages
Cohort 1936 * LG	-0.135	1858.7	8.157*	-0.481	204.7	0.104
	(0.606)	(1531.9)	(4.377)	(1.149)	(225.9)	(0.270)
Cohort 1935 * LG	-0.0717	1615.0	5.207	-2.109*	-166.9	-0.290
	(0.631)	(1551.2)	(7.045)	(1.165)	(226.1)	(0.270)
N	133026	133026	133026	133026	133026	133026
Mean dep. var.	94.033	46848.811	48.086	37.384	1623.839	4.146

	Hospitalized (yes/no)		Hospital days		Mortality	
	(1) Ages 65–68	(2) Age 65–year 2010	(3) Ages 65–68	(4) Age 65–year 2010	(5) By age 69	(6) By year 2011
Cohort 1936 * LG	1.423	2.783**	0.0135	0.771	0.643	1.591**
	(1.086)	(1.176)	(0.365)	(0.645)	(0.488)	(0.697)
Cohort 1935 * LG	1.220	1.615	-0.194	-0.155	0.473	0.170
	(1.101)	(1.175)	(0.360)	(0.682)	(0.505)	(0.738)
N	133026	133026	133026	133026	133026	133026
Mean dep. var.	28.325	42.484	3.662	7.146	4.352	7.038

Note: This table shows estimates from estimating Equation 7 after adding two pre-reform interaction terms between cohort $j = 1935, 1936$ and the local government dummy. See Tables 2 and 4 for more information on the sample of analysis and controls. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

existence of differential trends in work environment between public sector and private sector occupations is the large scale retrenchment of the public sector during the late 1990s (Angelov et al., 2011). If the work environment deteriorated across cohorts in the treatment group as a result of this, and there was no corresponding decline in the control group, we might capture effects on post-retirement health that are not only due to continued work, but to changes in work environment, too.

I do two things to investigate this issue. First, I look at occupation-specific sick leave patterns for women around the reform. Using data from the Social Insurance Agency, the upper panel of Figure A.1 plots the fraction of female workers that were absent from work for more than 60 days in a given year for five important worker categories. Personal care workers, the most numerous worker category in the treatment group, exhibit much higher absence rates than the other worker categories, but the trends look similar. The trends are also similar when we look at the average number of sick leave days, as can be seen from the lower graph in Figure A.1.

Second, I compare sick leave patterns of younger workers in the treatment and control group occupations in the years surrounding the reform. The advantage of this approach is that younger local government workers' sick leave patterns should be unrelated to the pension reform itself, yet indicative of the work environment situation. For each year between 1996 and 2005, I sample

all women aged 45–50 who are either observed working in any of the treatment group occupations listed in Table 1 or in the private sector. Here, an individual is defined as working in the private sector if she is neither self-employed nor working in the public sector. Figure A.2 plots the fraction of individuals with more than 14 consecutive days of sick leave in each of these two groups. Re-assuringly, we see that the sickness absence rates evolve similarly both in the years prior to and after the reform.

One additional assumption is needed for the 2SLS estimator to capture the LATE, namely that the instrument has a monotone impact on the endogenous variable, i.e. there are no defiers in the population. The existence of defiers could also harm the reduced-form approach if the composition of the treatment and control group changed in a way that is related to health because of the reform. Local government workers could avoid the new rules by retiring prior to the implementation of the new pension plan on January 1, 2000. Given that the reform was agreed on in mid-1998, those born in 1938 and 1939 were given some room to retire under the old rules.

The preferred way to test for the presence of such anticipatory behavior would be to apply a similar DD framework as in the main analysis and look specifically at retirement behavior at ages 60–61 for the affected cohorts. However, a simultaneous reform in the public pension system makes such an analysis difficult. In 1998, the minimum claiming age in the public pension system was raised from 60 to 61 (Palme and Svensson, 2004). As a result, individuals born in 1938 had to wait an additional year before they could claim public pension benefits. In contrast to private sector workers who were directly exposed to the new minimum claiming age, local government workers were unaffected by this reform as long as they retired under the pre-reform rules. Thus, we would not know to what extent a DD estimator would reflect anticipatory behavior among local government workers on the one hand, and later retirement among private sector workers on the other. Instead, I do two things to deal with this issue. First, by conditioning on being employed for 12 full months in the year of their 61st birthday, I exclude most individuals who potentially retire in anticipation of the reform. Second, I test whether the results are robust to excluding the 1938 and 1939 cohorts.

The composition of the treatment and control group might also change if some local government workers change occupation because of the reform. Occupations in which workers were able retire at 63 with a full pension might become less attractive relative to occupations which NRA was unchanged. However, the potential compositional effects of "disillusioned" local government workers should be minor given that job turnover rates are generally low for age groups close to retirement (Andersson and Tegsjö, 2010). Additionally, defined benefit pension plans of this kind may also reduce labor market flexibility through back-loading and limited portability of pension rights (Munnell and Sunden, 2004). One way to investigate whether we should worry about job changes of this sort is to look at the fraction of local government workers

who are seen working in occupations with different NRAs. It turns out that only 2.0 percent of all local government workers in my sample are observed working in both categories after the age of 60. Moreover, the probability of transitioning to an occupation with a NRA of 65 is roughly similar to the probability of transitioning to an occupation with a NRA of 63.

6 The impact of the reform on retirement

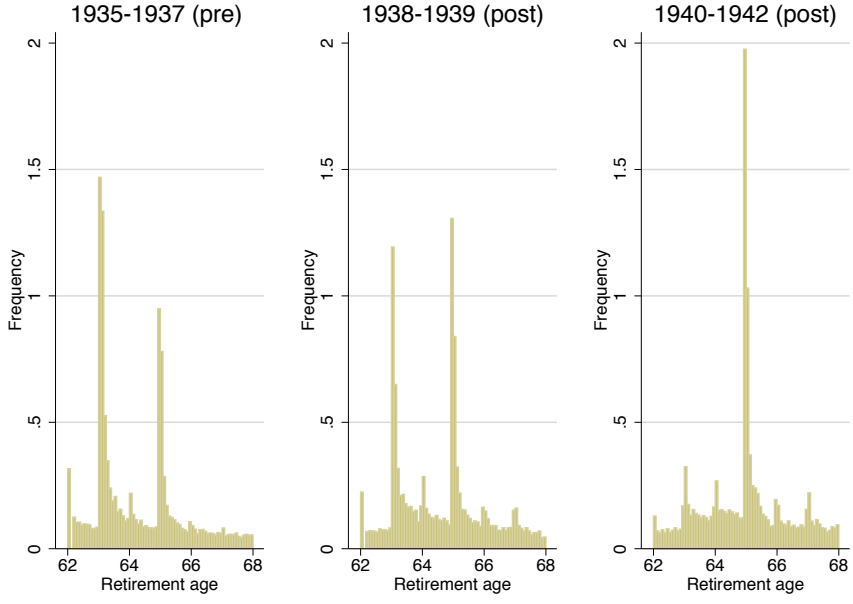
We know from the descriptive statistics in section 4.3 that post-reform cohorts in the treatment group retire more than 5.3 months later than the corresponding birth cohorts in the control group. This section aims at quantifying the impact of the reform on retirement in more detail.

The retirement effects of the reform are perhaps best illustrated in a histogram. Figure 2 shows the retirement distribution for pre- and post-reform cohorts in the treatment group. Most evident in the left-most panel is the spike of retirements around age 63. The spike around 65 is also pronounced, which means that many workers continue to work past the age at which they become entitled to full pension benefits. The two oldest post-reform cohorts, i.e. those born in 1938 and 1939, seem to retire later than the pre-reform cohorts, but the spike around 63 is only marginally smaller. Remarkably, it almost vanishes for the 1940–1942 cohorts. These graphs provide clear evidence that the reform increased the actual retirement age.

I proceed by presenting first-stage OLS estimates of Equations 6 and 7. The results are presented in Table 4. Column (1) presents the common treatment effect from the just-identified case, while column (2) allows for heterogeneous effects across birth cohorts, i.e. the over-identified case. The common treatment effect amounts to 4.5 months, providing clear evidence that the first assumption of the IV model holds. Column (2) shows that the effect on retirement is largely driven by the youngest cohorts. For example, those born in 1942 retire more than 6.2 months later than the pre-reform cohorts as compared to 1.4 months for those born in 1939.

How can we be sure that this movement in the retirement mass is not only the result of a general trend towards longer working lives? Figure 3 shows retirement distributions for the control group. Except for a slight decrease in the mass of retirements at ages 62 and 63, little seems to happen across these birth cohorts. I also estimate pre-reform trends for the retirement age in a similar fashion as for health in the previous section. Column (3) of Table 7 reports the estimation results after adding two interaction terms between pre-reform cohort $j = 1935, 1936$ and the local government dummy to the specification in column (2). The estimated coefficients imply that local government workers born in 1935 and 1936 retire 0.5 and 1.1 months earlier than those born in 1937, respectively, accounting for the the corresponding change in the control group. The coefficient for the 1935 cohort is significant at the 10 percent level while the coefficient for the 1936 cohort is insignificant. These results support the interpretation that the first-stage effects are the result of the reform itself rather than a differential underlying trend in retirement age between the treat-

Figure 2. Retirement distribution for local government workers (by cohort)



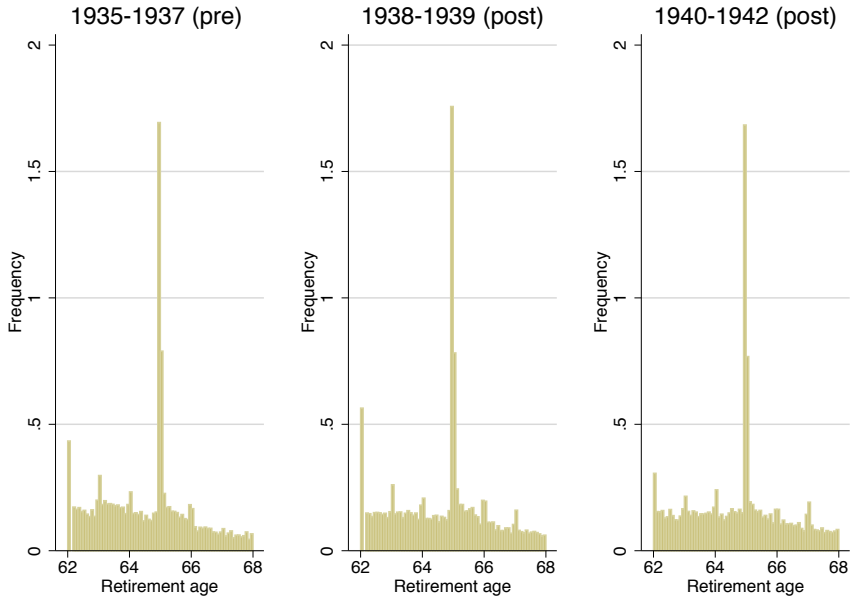
Note: Histogram of retirements in the treatment group. The decision to retire is equated with the month in which the individual's last employer reports the employment contract to be officially ended. See Table 2 for more information on the sample of analysis.

ment and control group.

There are several plausible explanations for the between-cohort differences in labor supply response observed in columns (2) and (3). First, if norms adjust slowly in response to a change in the NRA we should expect the labor supply adjustments to increase over time. In this specific case, though, the importance of norms should not be exaggerated. Sixty-five was already the NRA in all other major occupational pension plans as well as in the public pension system. Second, an immediate adjustment in response to changes in incentives could be prevented by adjustment costs or frictions (Gelber et al., 2013). Although the financial incentives to retire before 65 changed very quickly with the reform, there might be large non-financial costs of changing the retirement plans on short notice. Such costs should be higher for older cohorts that received news about the new rules just before they reached their intended retirement age.

Table A.4 shows the first-stage effects for the alternative retirement definitions. Columns (1) and (2) show that the reform also had a significant impact on claiming behavior. The common treatment effect of 0.49 translates into an increase in the actual claiming age of 5.9 months. The income-based definition of retirement yields an estimate of 0.37 years or 4.4 months. Again we see that the effect is driven by the youngest cohorts. Thus, these results verify

Figure 3. Retirement distribution for private sector workers (by cohort)



Note: Histogram of retirements in the control group. The decision to retire is equated with the month in which the individual's last employer reports the employment contract to be officially ended. See Table 2 for more information on the sample of analysis.

that the first-stage effect is robust to various definitions of retirement and that individuals indeed work longer as a result of the reform.[22]

[22] As an additional robustness check, I estimate the first-stage equations using Tobit regression. OLS estimates of the slope coefficients might be inconsistent and downward-biased when the dependent variable is right-censored. The Tobit estimates turn out similar in magnitude to the OLS estimates in Table 4.

Table 4. *First-stage results*

	(1)	(2)	(3)
Post-reform CH * LG	4.474***		
	(0.274)		
Cohort 1942 * LG		6.217***	5.684***
		(0.379)	(0.488)
Cohort 1941 * LG		5.371***	4.836***
		(0.395)	(0.501)
Cohort 1940 * LG		5.691***	5.156***
		(0.397)	(0.502)
Cohort 1939 * LG		1.893***	1.358***
		(0.418)	(0.520)
Cohort 1938 * LG		2.695***	2.159***
		(0.424)	(0.525)
Pre-reform cohorts			
Cohort 1936 * LG			-0.603
			(0.551)
Cohort 1935 * LG			-1.057*
			(0.557)
Observations	133,026	133,026	133,026
Mean dep. var.	40.708	40.708	40.708
F-statistic	265.869	81.565	45.792

Note: Column (1) shows first-stage estimates from Equation 6 and columns (2) and (3) from Equation 7. Column (3) adds two pre-reform interaction terms between cohort $j = 1935, 1936$ and the local government dummy to the specification in column (2). Robust standard errors in parentheses. Dependent variable: number of months employed from age 62 to 68. Estimated using OLS. Dependent variable right-censored at 72 (age 68). All regressions include cohort-fixed effects, regional dummies and dummies for month of birth. Additional control variables are the log of the average of yearly earnings between ages 56 and 60, the log of the standard deviation of yearly earnings between ages 56-60, number of years of schooling, dummies for immigrant status and having children at home, the number of years with more than 14 consecutive days of sick leave between ages 56–60, and interactions between schooling years/income and cohort and schooling years/income and local government. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively. See Table 2 for more information on the sample of analysis.

7 Results

7.1 Drug prescriptions

I begin with the consumption of prescription drugs because it is the most common form of health care utilization. The analyses of drug prescriptions are based on data about all over-the-counter sales of prescribed medical drugs between 2005–2009. The estimation sample is limited to individuals who were alive at the end of 2009. I first look at the consumption of drugs that treat physical illnesses and then turn to the mental health drugs.

Table 5 reports OLS and 2SLS estimates of the effect of an additional month of employment on the probability of being prescribed a non-zero quantity of drugs in a given category between 2005–2009. The outcome variables have been multiplied by 100 so that the coefficients can be interpreted in percentage points. The mean of each variable is reported under its estimated regression coefficient to facilitate interpretation. Column (1) reports results for total drug prescriptions, a category that includes all types of drugs while columns (2)–(10) report results for drugs categorized into any of the common or hypotheses-based causes. I report 2SLS estimates from the over-identified model to utilize cross-cohort variation in the strength of the first stage. The intensive margin estimates are reported in Table 6, where the upper panel uses the sum of DDDs as the dependent variable and the lower panel the number of prescribed packages.

The first column of Table 5 shows that more than 96 percent of the sample are prescribed a non-zero quantity of drugs during this period. It is therefore not surprising that the OLS estimate reflects a zero correlation between continued work and overall drug consumption. The correlation between specific drug categories and continued work is stronger. For example, the OLS estimate of -0.02 in column (9) implies that one year of extra work is associated with a -3.3 percent reduction in the probability of being prescribed drugs related to diabetes.[23] For cerebrovascular disease, the corresponding reduction amounts to 1.9 percent.[24] These OLS estimates reflect the negative correlation that is typically observed between retirement age and health (those who retire early tend to be in worse health), but cannot be used to make any causal claims about the effect on retirement on health due to non-random selection into retirement. It should be noted that the OLS estimates related to drugs that treat respiratory disease and analgesics have the opposite sign, i.e. the probability of being prescribed these drugs increases with retirement age. However, the amount consumed is lower for those who work longer. In fact, all intensive margin OLS estimates in Table 6 are negative and, with one exception only, highly significant.

[23] I.e. $-0.021 \times 12/7.449 = -0.034$

[24] Cerebrovascular disease refers to a group of conditions that affect the circulation of blood to the brain, causing limited or no blood flow to affected areas of the brain. This is commonly a stroke.

Table 5. *The effect of postponing retirement on drug utilization: the extensive margin*

		Common causes				Hypotheses-based causes				
	(1) Any drug	(2) Cancer	(3) Resp	(4) Circ	(5) Other	(6) Heart	(7) Cere	(8) Analgesics	(9) Diabetes	(10) Mental
OLS: Employment, *R*	0.000944	-0.00199	0.0596***	-0.0545***	0.00157	-0.0591***	-0.0442***	0.0134**	-0.0207***	-0.0347***
	(0.00235)	(0.00150)	(0.00621)	(0.00605)	(0.00275)	(0.00610)	(0.00562)	(0.00602)	(0.00332)	(0.00603)
2SLS: Employment, *R*	-0.0426	-0.0388	0.0668	-0.103	-0.0564	-0.109	-0.177*	-0.0205	-0.0767	0.111
	(0.0418)	(0.0266)	(0.109)	(0.106)	(0.0494)	(0.107)	(0.0985)	(0.106)	(0.0586)	(0.106)
N	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822
Mean dep. var.	96.180	1.400	51.886	61.891	94.615	59.916	27.792	61.340	7.449	37.733

Note: This table reports OLS and 2SLS estimates of postponing retirement by one month on the utilization of prescription drugs. Drug utilization is scaled so that a coefficient of 1.00 denotes a one percentage point increase in the likelihood of prescription. Circ: circulatory disease. Resp: respiratory disease. Table A.1 describes the aggregation of ATC codes for each cause. See Tables 2 and 4 for more information on the sample of analysis and controls. The sample is restricted to individuals who were alive at the end of 2009. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

Table 6. *The effect of postponing retirement on drug utilization: the intensive margin*

		Common causes			Hypotheses-based causes				
	(1) Any drug	(2) Resp	(3) Circ	(4) Other	(5) Heart	(6) Cere	(7) Analgesics	(8) Diabetes	(9) Mental
	Panel A: Sum of daily doses prescribed								
OLS: Employment, *R*	-112.6***	-2.770**	-81.50***	-28.32***	-81.50***	-2.801***	-9.280***	-0.970**	-8.410***
	(7.894)	(1.299)	(6.259)	(3.888)	(6.259)	(0.670)	(2.233)	(0.484)	(1.051)
2SLS: Employment, *R*	-138.3	13.57	-84.69	-67.16	-84.69	-4.657	-30.63	-3.653	-1.927
	(138.9)	(22.26)	(110.8)	(68.07)	(110.8)	(11.92)	(38.89)	(8.314)	(19.00)
N	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822
Mean dep. var.	48062.954	2282.055	28615.946	17164.953	28615.946	933.174	6319.871	430.062	1644.933
	Panel B: Number of prescribed packages								
OLS: Employment, *R*	-0.0900***	-0.00387***	-0.0245***	-0.0614***	-0.0244***	-0.00300***	-0.00240	-0.00523***	-0.0137***
	(0.0217)	(0.00123)	(0.00175)	(0.0212)	(0.00175)	(0.000382)	(0.00155)	(0.00104)	(0.00143)
2SLS: Employment, *R*	-0.139	-0.00498	-0.0288	-0.0994	-0.0308	-0.0117*	0.00186	-0.0197	-0.0226
	(0.378)	(0.0208)	(0.0310)	(0.370)	(0.0309)	(0.00665)	(0.0287)	(0.0181)	(0.0248)
N	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822
Mean dep. var.	48.004	3.784	10.918	33.198	10.777	1.376	5.222	1.731	4.215

Note: This table reports OLS and 2SLS estimates of postponing retirement by one month on the quantity of prescribed drugs. Drugs related to cancer have been left out because the data does not include information on the number of DDDs nor on the number of prescribed packages for this category. Circ: circulatory disease. Resp: respiratory disease. Table A.1 describes the aggregation of ICD codes for each cause. See Tables 2 and 4 for more information on the sample of analysis and controls. The sample is restricted to individuals who were alive by the end of year 2009. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

There seems to be a negative relationship between continued work and the consumption of prescription drugs that treat physical illnesses even when health selection into retirement is controlled for. Most 2SLS estimates are negative, yet statistically insignificant and quite small. For example, the 2SLS estimate of -0.043 for all drugs in column (1) translates into a small relative effect of -0.53 percent of one year of extra work. The estimate is precisely estimated with the 95 percent confidence interval of one year of extra work ranging from -1.6 percent to 0.01 percent (-0.06 to 0.02 SD units). The only statistically significant coefficient is the coefficient related to drugs that treat cerebrovascular disease (significant at the 10 percent level). Specifically, the point estimate in column (7) implies that one year of extra work reduces the probability of being prescribed a non-zero amount of drugs that treat cerebrovascular disease by –7.7 percent.[25] The 95 percent confidence interval associated with this yearly effect ranges from -16.0 percent 0.7 percent (-0.09 to 0.004 units).

The intensive margin 2SLS estimates in Table 6 are in line with the extensive margin results. Drug consumption is only marginally affected by continued work and the effect is negative in most cases. The only drug category that is significantly affected by continued work are, again, drugs that treat cerebrovascular disease.

Table 7 reports the extensive margin results for the selection of drugs that treat mental illnesses. Again, the importance of controlling for negative health selection becomes clear when we compare the OLS estimates to the 2SLS estimates. While the OLS estimates are negative and highly significant, none of the 2SLS estimates are significant and some of them have the opposite sign. For example, the OLS estimate for antidepressants in column (6) reflects a relative reduction of 3.4 percent from one year of extra work. The corresponding 2SLS estimate translates into a corresponding yearly reduction of around 1.1 percent only. The 95 percent confidence interval associated with this yearly effect ranges from –12.9 to 10.6 percent (-0.06 to 0.05 SD units).

Similar to drugs that treat physical illnesses, the prescribed quantity of mental health drugs is more strongly correlated with retirement age than the probability of being prescribed non-zero quantities. Specifically, the OLS estimates in Table 8 for the sum of DDDs and the number of prescribed packages for the category "any mental drug" reflect relative effects of -4 to -6 percent per additional year of work. The 2SLS point estimates imply even larger effects, but are not precise enough to reject the null hypothesis of no effect.

Two sub-categories of mental health drugs are, however, significantly affected by continued work. First, column (4) reveals a significant reduction in the number of prescribed packages of anxiolytics. The point estimate of -0.015 translates into a yearly relative effect of -23.9 percent. The corresponding estimate for the sum of DDDs translates into a relative effect of the

[25] $-0.177 \times 12/27.79 = -0.077$

Table 7. *The effect of postponing retirement on mental health drugs: the extensive margin*

	(1) Any mental drug	(2) Parkinson	(3) Antipsychotics	(4) Anxiety	(5) Hypnotics & sedatives	(6) Antidep. (N06A)	(7) N06AA	(8) N06AB	(9) N06AX
OLS: Employment, *R*	-0.0347***	-0.00892***	-0.0140***	-0.0312***	-0.0136**	-0.0456***	-0.0115***	-0.0365***	-0.0151***
	(0.00603)	(0.00201)	(0.00169)	(0.00456)	(0.00543)	(0.00462)	(0.00229)	(0.00397)	(0.00284)
2SLS: Employment, *R*	0.111	-0.0342	-0.00577	-0.0487	0.0484	-0.0154	0.0207	-0.0369	-0.0551
	(0.106)	(0.0355)	(0.0295)	(0.0790)	(0.0942)	(0.0805)	(0.0398)	(0.0692)	(0.0495)
N	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822
Mean dep. var.	37.733	2.585	1.823	15.345	24.568	16.086	3.323	11.128	5.277

Note: This table reports OLS and 2SLS estimates of postponing retirement on the utilization of drugs that treat mental illnesses. Drug utilization is scaled so that a coefficient of 1.00 denotes a one percentage point increase in the likelihood of prescription. Table A.1 describes the aggregation of ATC codes for each cause. See Tables 2 and 4 for more information on the sample of analysis and controls. The sample is restricted to individuals who were alive at the end of 2009. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

Table 8. *The effect of postponing retirement on mental health drugs: the intensive margin*

	(1) Any mental drug	(2) Parkinson	(3) Antipsychotics	(4) Anxiety	(5) Hypnotics & sedatives	(6) Antidep. (N06A)	(7) N06AA	(8) N06AB	(9) N06AX
			Panel A: Sum of daily doses prescribed						
OLS: Employment, *R*	-8.410***	-2.368***	-0.791***	-1.283***	-1.827***	-2.141***	-1.812***	-0.260***	-0.0730
	(1.051)	(0.587)	(0.180)	(0.343)	(0.564)	(0.461)	(0.421)	(0.0384)	(0.178)
2SLS: Employment, *R*	-1.927	5.555	-0.962	-6.800	6.800	-6.520	-5.577	0.343	-1.316
	(19.00)	(10.76)	(3.315)	(6.912)	(9.675)	(8.151)	(7.358)	(0.671)	(3.384)
N	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822
Mean dep. var.	1644.933	263.786	57.545	323.602	562.765	437.235	274.022	75.086	87.900
			Panel A: Number of packages						
OLS: Employment, *R*	-0.0137***	-0.00256***	-0.00119***	-0.00331***	-0.00176***	-0.00492***	-0.000817***	-0.00260***	-0.00158***
	(0.00143)	(0.000565)	(0.000253)	(0.000438)	(0.000629)	(0.000570)	(0.000182)	(0.000381)	(0.000327)
2SLS: Employment, *R*	-0.0226	0.000745	0.00219	-0.0154**	-0.00224	-0.00789	-0.00237	0.00396	-0.0101*
	(0.0248)	(0.00997)	(0.00419)	(0.00769)	(0.0109)	(0.00994)	(0.00313)	(0.00665)	(0.00583)
N	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822	125,822
Mean dep. var.	4.215	0.297	0.152	0.775	1.676	1.314	0.162	0.802	0.346

Note: This table reports OLS and 2SLS estimates of postponing retirement by one month on the quantity of prescribed drugs that treat mental illnesses. See Tables 2 and 4 for more information on the sample of analysis and controls. The sample is restricted to individuals who were alive by the end of year 2009. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

same magnitude, but is estimated with less precision and turns out insignificant. The second category that has a significant coefficient is "other antidepressants" (N06AX), which includes antidepressants not fitting into the established classes of serotonin reuptake inhibitors (SSRIs; N06B) and tricyclic antidepressants (N06AA).[26]

These results suggest that continued work might alleviate short-term anxiety and depressive symptoms among elderly with pre-existing mental health issues. This interpretation is based on three observations. First, both anxiolytics and "other antidepressants" are primarily used when SSRIs yield insufficient treatment effects or in cases of severe depression or anxiety. Second, the vast majority of prescribed anxiolytics belong to a class of medications called benzodiazepines, which are typically recommended for short-term ($<$1 month) relief of anxiety (Allgulander and Nutt, 2010).[27] Third, the number of prescribed packages is a particularly relevant measure of the length and intensity of the treatment with drugs that treat anxiety because they are often taken "as needed" rather than on a daily basis.

One potential mechanism for this result is that the transition from work to retirement might be associated with a decline in the degree of social interactions, which may affect mental health. Social networks formed at work may protect individuals from shocks that may otherwise impact health.[28] Another potential mechanism is that continued work is associated with better sleeping habits than retirement since anxiolytics, benzodiazepines in particular, are often used to treat insomnia in the elderly (Holbrook et al., 2000).

7.2 Hospitalizations

We now turn to the analysis of hospital admissions. Table 9 reports OLS and 2SLS estimates of the effect of an additional month of work on the probability of being hospitalized between ages 65–68 (Panel A) and between age 65 and year 2010 (Panel B). I follow Hallberg et al. (2014) and estimate reduced-form effects on the number of days spent in hospital using a Poisson regression model. The results from this analysis are presented in Table 10. I restrict the estimation sample to individuals who were alive for the entire period over

[26] SSRIs are the most common and also the first-choice antidepressants since they are easy to handle and have fewer and less severe side effects than the other two classes of antidepressants (Turnheim, 2004).

[27] The use of benzodiazepines increases steadily with age despite more worrisome side effects among the elderly, including higher risk of falls (Woolcott et al., 2009) and delirium (Clegg and Young, 2011). It is also twice more likely among women than among men (Allgulander and Msghina, 2010).

[28] It is not clear that retirement always leads to a decline in social interactions as one has more leisure time with which to establish new social contacts outside work. Börsch-Supan and Schuth (2010) report that social networks shrink after retirement while Eibich (2014) find that the number of good friends increases.

which a hospitalization variable is defined.

As seen from the OLS estimates in Table 9, hospitalizations are more strongly correlated with retirement age than drug prescriptions. All estimates are negative and highly significant. The point estimates for the main outcomes in column (1), i.e. the probability of being hospitalized due to any cause, imply that one additional month of employment decreases the probability of being hospitalized by around 0.03–0.04 percentage points, which corresponds to a relative effect of 1–2 percent per additional year of work. The hypotheses-based estimates imply even stronger correlations. For example, one year of extra work increases the probability of being hospitalized due to diabetes by 7.2 percent.

The 2SLS estimates provide further evidence that the causal effect of postponing retirement on overall health care utilization is likely to be small. The column (1) estimates for the probability of being hospitalized due to any cause are insignificant, yet precisely estimated. The point estimate of -0.096 (0.010) in Panel B translates into a relative effect of -2.9 percent per additional year of work. The 95 percent confidence interval associated with this yearly effect ranges from -8.9 percent to 3.2 percent (-0.07 to 0.03 SD units).

Next, I explored in detail the diagnoses codes to see whether the small effect on hospitalizations masks any heterogeneous effects with respect to the hospitalization cause. Columns (2)–(12) of Table 9 report 2SLS estimates of the probability of having been hospitalized for each of the common and hypotheses-based causes. As seen in Panel A, only the coefficient for cerebrovascular disease is significantly different from zero. The implied relative effect is large; the probability of being hospitalized due to cerebrovascular disease is reduced by almost 40 percent per additional year of work. This result is in line with the previous result on drugs that treat cerebrovascular disease.

For the second time period (Panel B), we see that continued work reduces the probability of being hospitalized due to diabetes. The point estimate in column (12) implies that one month of additional month of employment is associated with a 0.11 percentage point decrease in the likelihood of being hospitalized with diabetes as one of the listed causes for admission. In relative terms, one year of extra work decreases this likelihood by 37.5 percent.

Many of the risk factors for cerebrovascular disease and diabetes are related to lifestyle, including weight, diet, stress and exercise. These results thus suggest that continued work at older ages might provide individuals with better opportunities to preserve a healthy lifestyle than retirement and hence reduce the risk of lifestyle diseases. An alternative interpretation is that the reduced risk of being treated for cerebrovascular disease is a direct result of fewer or less severe diabetes related complications. In fact, the risk of stroke is increased by 150–400 percent for patients with diabetes (Franco et al., 2007). Hospitalization for reasons related to diabetes usually requires that the patient suffers from life-threatening metabolic complications or severe chronic complications that require intensive treatment or close monitoring.

A third interpretation to the reduction in diabetes related hospitalizations is that it is the result of the previously documented effect on the consumption of anxiolytics and antidepressants. The medical literature has shown that depression is a risk factor for the onset of diabetes (Knol et al., 2006). Several factors associated with depressive symptoms, including obesity-promoting health behaviors and inflammatory responses (resulting in increased cortisol), can induce insulin resistance and the development of type 2 diabetes (Golden et al., 2008). There is also evidence that depression could be the outcome of diabetes rather than the other way around (Anderson et al., 2001). Depressive symptoms may occur as a result of the hardships of dealing with the complications of diabetes. Although the potential physical and psychological mechanisms underlying the bidirectional association between diabetes and depression are not entirely clear, proper lifestyle interventions including adequate weight management and regular physical activity are recommended to lower the risk of both conditions (Pan et al., 2010).

Turning to the intensive margin estimates in Table 10, continued work has no significant effect on the number of days spent in hospital due to cerebrovascular disease, nor due to diabetes. The estimates are negative and indicate quite sizable effects, but they are not statistically distinguishable from zero. There is, however, a significant effect on the number of days spent in hospital due to ischemic heart disease.[29] Since cardiac ischemia and cerebrovascular disease have many risk factors in common, this result supports the interpretation that continued work might reduce the risk of being treated for lifestyle diseases.

The result that postponing retirement has no effect on the overall consumption of health care, and even reduces health care related to certain medical causes, contrasts two previous papers that study the effect of retirement on health care utilization. First, in a recent paper that uses similar data on hospital admissions as this paper does, Hallberg et al. (2014) report that the introduction of a generous early retirement program for Swedish army employees reduced the number of days in hospital between ages 61–70 by more than 40 percent. These findings suggest that the health implications of targeted early retirement offers may be very different from those of strengthened financial incentives to work longer. It should be noted, however, that the population of female local government workers studied in this paper differ in many respects to the sample of male army employees in Hallberg et al. (2014). Second, using data from the German Socio-Economic Panel Study (SOEP), Eibich (2014) finds that retirement is associated with a significant reduction in the probability of seeing a doctor or staying in a hospital. My results are more in line with a study by Gorry et al. (2015) who use US data from the Health and Retirement Study (HRS) and variation in eligibility for Social Security and private

[29]Cardiac ischemia occurs when plaque and fatty matter narrow the inside of an artery to a point where it cannot supply enough oxygen-rich blood to meet the heart's needs.

Table 9. *The effect of postponing retirement on cause-specific hospitalization*

		Common causes				Hypotheses-based causes						
	(1) Any visit	(2) Cancer	(3) Resp	(4) Circ	(5) Other	(6) Ischemic	(7) Hypertension	(8) Cerebro	(9) Muscu	(10) Alcohol	(11) External	(12) Diabetes
					Panel A: Ages 65–68							
OLS: Employment, *R*	-0.0389***	-0.00831***	-0.00846***	-0.0213***	-0.0297***	-0.00802***	-0.0227***	-0.00902***	-0.0160***	-0.00677***	-0.00688***	-0.0114***
	(0.00548)	(0.00250)	(0.00161)	(0.00290)	(0.00511)	(0.00193)	(0.00299)	(0.00145)	(0.00287)	(0.00132)	(0.00217)	(0.00171)
2SLS: Employment, *R*	-0.00184	0.00156	-0.0106	-0.0234	-0.0551	0.00311	0.0825	-0.0478*	0.0804	0.0129	-0.0236	-0.00889
	(0.0967)	(0.0446)	(0.0278)	(0.0514)	(0.0896)	(0.0342)	(0.0530)	(0.0263)	(0.0507)	(0.0230)	(0.0380)	(0.0302)
N	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166
Mean dep. var.	26.924	4.324	1.667	5.888	21.606	2.531	6.329	1.447	5.725	1.176	3.174	1.904
					Panel B: Age 65–year 2010							
OLS: Employment, *R*	-0.0320***	-0.00402	-0.00802***	-0.0215***	-0.0210***	-0.00973***	-0.0264***	-0.0132***	-0.0131***	-0.00563***	-0.00800***	-0.0137***
	(0.00595)	(0.00322)	(0.00214)	(0.00387)	(0.00578)	(0.00262)	(0.00428)	(0.00208)	(0.00392)	(0.00184)	(0.00300)	(0.00233)
2SLS: Employment, *R*	-0.0962	0.0405	0.0148	-0.0848	-0.221**	-0.0620	-0.0729	-0.0387	0.0123	0.0187	-0.0732	-0.107***
	(0.103)	(0.0566)	(0.0372)	(0.0680)	(0.101)	(0.0462)	(0.0749)	(0.0369)	(0.0686)	(0.0314)	(0.0532)	(0.0411)
N	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550
Mean dep. var.	40.108	6.905	2.866	10.479	33.344	4.420	13.263	2.766	10.771	2.092	6.003	3.423

Note: This table reports OLS and 2SLS estimates of postponing retirement on hospitalization from specific causes. Hospitalization is scaled so that a coefficient of 1.00 denotes a percentage point increase in the likelihood of being hospitalized from this cause. Circ: circulatory disease. Resp: respiratory disease. Ischemic: Ischemic heart disease. Cerebro: cerebrovascular disease. Muscu: musculoskeletal disease. Table A.1 describes the aggregation of ICD codes for each cause. See Tables 2 and 4 for more information on the sample of analysis and controls. The sample is restricted to individuals who were alive at age 68 (Panel A) and by the end of year 2010 (Panel B). Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

Table 10. *Number of hospital days*

		Common causes				Hypotheses-based causes						
	(1) All	(2) Cancer	(3) Resp	(4) Circ	(5) Other	(6) Ischemic	(7) Hypertension	(8) Cerebro	(9) Muscu	(10) Alcohol	(11) External	(12) Diabetes
					Panel A: Ages 65–68							
Post-reform CH * LG	0.0157	-0.144*	0.00494	-0.0400	0.0378	-0.198*	0.154	-0.200	0.0761	0.0947	0.0461	-0.109
	(0.0500)	(0.0873)	(0.154)	(0.0818)	(0.0345)	(0.110)	(0.0989)	(0.158)	(0.0763)	(0.203)	(0.100)	(0.160)
N	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166	128,166
Mean dep. var.	2.714	0.670	0.234	0.986	6.289	0.349	0.642	0.453	0.900	0.231	0.422	0.304
					Panel B: Age 65–year 2010							
Post-reform CH * LG	0.0190	-0.0159	0.0684	-0.0714	0.0268	-0.126	0.0170	-0.139	0.0216	0.0694	-0.0101	-0.161
	(0.0372)	(0.0768)	(0.127)	(0.0626)	(0.0313)	(0.0865)	(0.0584)	(0.120)	(0.0571)	(0.131)	(0.0747)	(0.116)
N	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550	124,550
Mean dep. var.	5.012	0.855	0.333	1.435	7.718	0.565	1.297	0.695	1.423	0.367	0.666	0.532

Note: This table reports reduced-form Poisson estimates on the number of hospital days from specific causes. Circ: circulatory disease. Resp: respiratory disease. Ischemic: Ischemic heart disease. Cerebro: cerebrovascular disease. Muscu: musculoskeletal disease. See Tables 2 and 4 for more information on the sample of analysis and controls. The sample is restricted to individuals who were alive at age 68 (Panel A) and by the end of year 2010 (Panel B). Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

defined benefit plans. They find that even though retirement improves both self-reported health and life satisfaction, there is little evidence that retirement influences health care utilization.

7.3 Mortality

Mortality is the most objective measure of health. One would expect small or even zero effects of continued work on mortality given the small overall effects on hospitalizations and drug prescriptions. Panel A of Table 11 reports estimates for mortality by age 69. Results for mortality by year 2011 are reported in Panel B. All dependent variables have been multiplied by 100 so that the coefficients can be interpreted in percentage points.

The OLS estimates are again negative and significant. Specifically, the point estimate in column (1) of Panel A reveals that one month of extra work is associated with a decrease in the risk of dying before age 69 by 0.10 percentage points. One year of extra work is then associated with a 1.22 percentage point decrease in mortality, which in relative terms correspond to a 28.0 percent reduction in mortality by age 69.[30] For mortality by year 2011, the corresponding reduction in mortality amounts to 20.0 percent. The 2SLS estimates, on the other hand, are insignificant and much smaller. Thus, while the OLS estimate suggests that people who work longer also live longer, the 2SLS estimates show that this relationship is not a causal one.

I go on by analyzing mortality with respect to the cause of death. Using the familiar categorization into common and hypotheses-based causes, Table 11 reports 2SLS estimates for cause-specific mortality by age 69 and by year 2011. The magnitude of the effects varies across causes, but none of them are statistically distinguishable from zero. Thus, I find no evidence that retirement affects the probability of death due to any of the listed causes.

Even if all causal estimates on mortality are insignificant, we cannot rule out that later retirement has an impact on mortality. A key issue in ruling out effect sizes of important magnitude is the precision of the estimates. Given that only 4.4 percent of the individuals in the sample are deceased by age 69, it comes as no surprise that the standard errors are quite large. For example, the 0.34 percentage point reduction in mortality by age 69 from one additional year of work has a 95 percent confidence interval of -0.73 to 1.41. This translates into a range of relative effects of -16.7 to 32.3 percent. The corresponding confidence interval for mortality by year 2011 ranges from -21.6 percent to 16.5 percent.

But how does the precision of the estimates compare to other studies on the relationship between retirement and mortality? The most comparable study is Hernaes et al. (2013) who investigate the mortality effects of lowering the

[30] $-0.102 \times 12/4.352 = -0.281$

Table 11. *The effect of postponing retirement on cause-specific mortality*

		Common causes				Hypotheses-based causes						
	(1) All	(2) Cancer	(3) Resp	(4) Circ	(5) Other	(6) Ischemic	(7) Hyper	(8) Cerebro	(9) Muscu	(10) Alcohol	(11) External	(12) Diabetes
					Panel A: Mortality by age 69							
OLS: Employment, *R*	-0.102***	-0.0611***	-0.00466***	-0.0211***	-0.0143***	-0.0120***	-0.00328***	-0.00845***	-0.00106***	-0.0282***	-0.00181***	-0.00409***
	(0.00241)	(0.00193)	(0.000521)	(0.00109)	(0.000874)	(0.000826)	(0.000479)	(0.000676)	(0.000273)	(0.00130)	(0.000327)	(0.000454)
2SLS: Employment, *R*	0.0283	0.0101	0.00288	0.0177	-0.00176	0.0113	0.00357	0.0151	-0.00621	0.0102	0.00400	0.00820
	(0.0454)	(0.0360)	(0.00902)	(0.0203)	(0.0159)	(0.0155)	(0.00904)	(0.0129)	(0.00478)	(0.0241)	(0.00601)	(0.00882)
N	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026
Mean dep. var.	4.352	2.700	0.168	0.844	0.558	0.489	0.166	0.326	0.054	1.220	0.073	0.147
					Panel B: Mortality by year 2011							
OLS: Employment, *R*	-0.118***	-0.0648***	-0.00626***	-0.0246***	-0.0184***	-0.0140***	-0.00474***	-0.0104***	-0.00127***	-0.0305***	-0.00169***	-0.00563***
	(0.00305)	(0.00231)	(0.000672)	(0.00136)	(0.00114)	(0.00103)	(0.000628)	(0.000826)	(0.000339)	(0.00156)	(0.000433)	(0.000603)
2SLS: Employment, *R*	-0.0152	-0.000254	0.00784	0.00558	0.00951	0.0136	0.00182	-0.00105	0.00387	0.0115	0.000197	0.00498
	(0.0570)	(0.0432)	(0.0121)	(0.0259)	(0.0213)	(0.0195)	(0.0122)	(0.0166)	(0.00620)	(0.0290)	(0.00797)	(0.0116)
N	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026
Mean dep. var.	7.038	3.850	0.282	1.325	0.915	0.761	0.289	0.504	0.086	1.723	0.121	0.251

Note: This table reports 2SLS estimates of postponing retirement on mortality from specific causes. Mortality is scaled so that a coefficient of 1.00 denotes a percentage point increase in the likelihood of dying from this cause. Circ: circulatory disease. Resp: respiratory disease. Ischemic: Ischemic heart disease. Hyper: hypertension. Cerebro: cerebrovascular disease. Muscu: musculoskeletal disease. Table A.1 describes the aggregation of ICD codes for each cause. See Tables 2 and 4 for more information on the sample of analysis and controls. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

early retirement age for a group of Norwegian workers using a similar DD strategy as in this paper. In addition, the sample sizes are quite similar. They find that a one-year increase in the actual retirement age results in a 0.2 percentage point increase in mortality by age 70. The estimated (insignificant) effect is thus somewhat smaller than what I find for mortality by age 69, but is estimated with similar precision with the 95 percent confidence interval ranging from -0.78–1.18 percentage points. In another related study, Kuhn et al. (2010) find that the introduction of more generous early retirement rules for Austrian blue-collar workers had a significant effect on mortality among male workers, but no effect among female workers. For women, Kuhn et al. (2010) report that one additional year spent in early retirement results in a 0.02 percentage point increase in mortality at age 67. The effect is thus very close to zero, but the confidence interval ranges from -1.84 to 1.88 percentage points.

8 Additional results

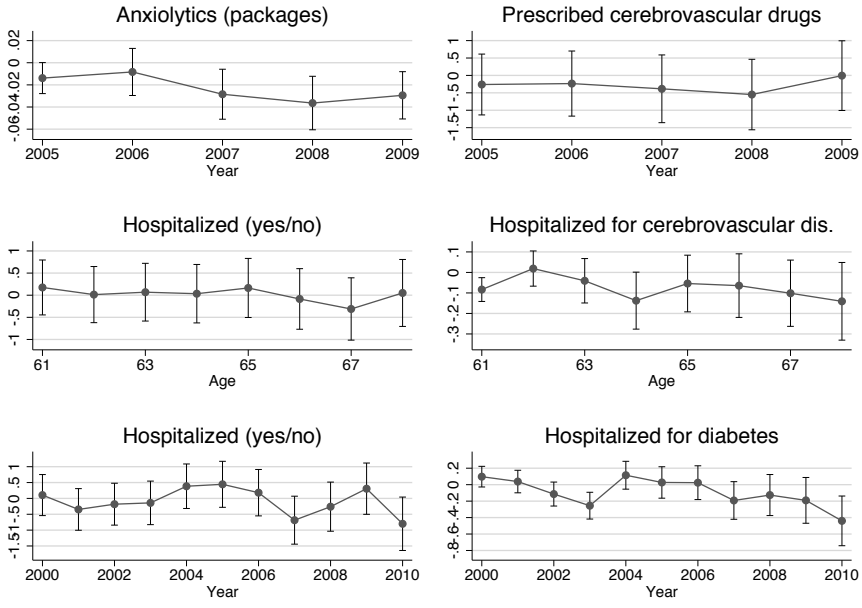
8.1 Dynamics

To investigate whether the observed effects on cause-specific health care utilization are more likely to occur at younger or older ages, I estimate the reduced-form equation (8) with a dummy variable that equals 1 if the individuals experienced a specific health event at a particular age, a, or in a particular year, y. Due to data availability, the effect on drug outcomes is estimated separately for years $y = 2005, ..., 2009$ while hospital admissions are studied at ages $a = 61, ..., 68$ and in years $y = 2001, ..., 2010$. The estimated interaction terms and the corresponding 95 percent confidence intervals are plotted in Figure 4. The hospitalization analysis is extended to ages before 65 to investigate the potential effects of the reform on health care utilization prior to retirement.

The first two panels show results for the number of prescribed packages of anxiolytics and the probability of being prescribed a non-zero quantity of drugs related to cerebrovascular disease, respectively. While the effect on cerebrovascular disease exhibits little variation over time, the reduction in anxiolytics is clearly driven by the most recent years.

The remaining four panels show results for hospitalizations due to any cause and hospitalizations due to cerebrovascular disease and diabetes. There is no particular time pattern with respect to the probability of being hospitalized due to any cause, but the effects on cerebrovascular disease and diabetes grow slightly with age/over time. These results suggest that the effects on lifestyle illnesses, such as type 2 diabetes, take time to manifest and that a longer follow-up period may reveal even larger effects. However, the results should be interpreted with some caution given that the magnitude of the differences over time and age are rather small and that these estimates capture both time and cohort effects.

Figure 4. Dynamics



Note: This figure plots reduced-form estimates and the corresponding 95 percent confidence intervals for a series of health outcomes. Separate regressions by year or age. The binary outcome variables have been multiplied by 100 so that the y-axis denotes percentage points.

8.2 Heterogeneous treatment effects

This section analyzes whether postponing retirement differentially affects different types of workers. I present reduced-form estimates of the effect of the reform on a selection of health outcomes for different subsamples. The health outcomes are the same as in the previous section (8.1) except that I replace the extensive margin hospitalization variables for any cause and for cerebrovascular disease with mortality by age 69 and the probability of being prescribed a non-zero quantity of diabetes-related drugs. Figures 5–7 plot the point estimates and the corresponding 95 percent confidence intervals. Consequently, the effect is not significant at the 5 percent level if the confidence interval includes zero. For scaling reasons, the number of prescribed packages of anxiolytics has been standardized to have mean zero and a standard deviation of one and then divided by 5. I report p-values from an interaction model that test whether the estimated difference between two subsamples is statistically significant. Specifically, I re-estimate the reduced-form equation (8) and interact each right-hand side variable with a dummy for belonging to a certain subsample. Moreover, since the 2SLS second-stage estimates can be understood as a re-scaling of the reduced-form, I discuss the strength of the first stage whenever the two groups differ substantially in their retirement response

to the reform. The p-value of the interaction term between the group dummy variable and the DD estimator then indicates whether the reform had a differential impact on the two groups.

I first estimate the model separately for individuals with income below and above the median of pre-retirement income. Individuals with low income are more likely to have had more physically demanding jobs than individuals with high income. Retiring from such jobs may imply a larger reduction in work-related strain and better health outcomes.[31] However, for many individuals, work-related activities may constitute the primary form of exercise and physical activity. If the prevalence of engagement in physical activity is similar or lower for individuals with physically demanding jobs, retirement could lead to more adverse health effects for these workers.[32] Continued work would then help preserve the health of individuals with demanding occupations.

Figure 5 provides little evidence of effect heterogeneity by income level. The differences with respect to the utilization of health care are quite small and also statistically insignificant. Moreover, even if the two mortality estimates have the opposite sign, none of them, as well as the the difference between them, are statistically different from zero. The reduced-form estimates are fairly representative of the 2SLS second-stage estimates. High-income individuals retire somewhat later due to the reform with a common treatment effect of about 4.5 months compared to 4.0 months for the low-income group.

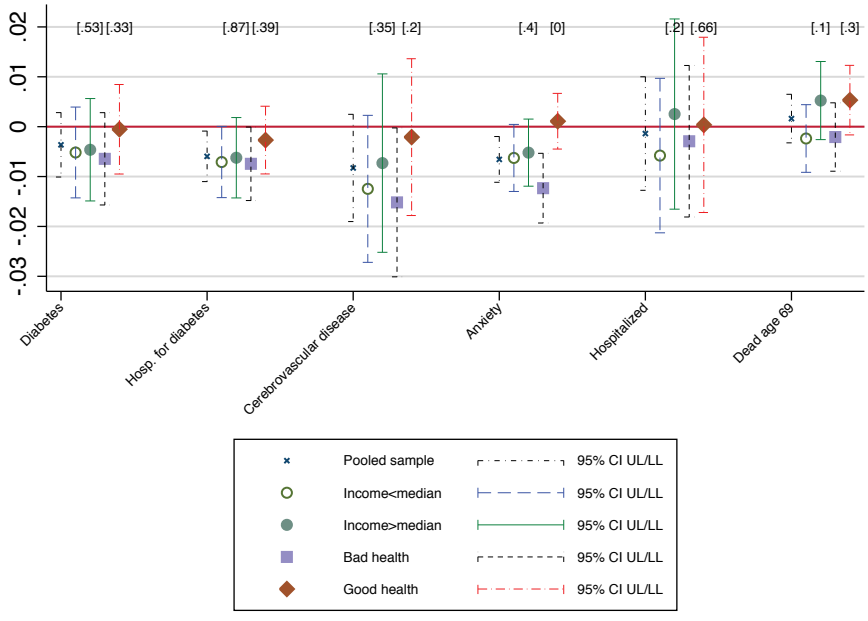
Next, I investigate effect heterogeneity with respect to ex-ante health status. Individuals in bad health might suffer more if the existing health problems accelerate the health decline that follows from additional work. However, if individuals in bad health have less opportunity to invest in their health after retirement, they might benefit more from continued work relative to individuals with better ex-ante health status. My measure of ex-ante health status is based on whether the individual was on sick leave for at least 14 consecutive days in any year between ages 56–60. Based on this definition, close to 50 percent are classified into each subsample. The results, also shown in Figure 5, show that there is more important effect heterogeneity along this dimension than the income dimension. In fact, all estimates imply better health outcomes for individuals in bad ex-ante health. As indicated by the p-values in the right-hand side brackets, the differences are particularly large for cerebrovascular drugs and anxiolytics. It should be noted, however, that the second-stage differences in health outcomes are smaller as individuals in bad health respond more strongly to the reform than individuals in good health (common treatment effect of 5.6 and 3.6 months, respectively).

I go on by checking specifically for effect heterogeneity with respect to oc-

[31] For example, Hallberg et al. (2014) find stronger positive health effects of early retirement for workers with low education and low income. Eibich (2014) reports that the positive effect on physical health is stronger among individuals who retire from physically demanding jobs.

[32] For example, Dave et al. (2008) find that the overall negative effect of retirement on health is larger for individuals with physically demanding work.

Figure 5. Effect heterogeneity by income and health



Note: The figure shows the reduced-form estimates for the effect of the reform on various outcomes by income level and health status. The dots mark the point estimates and the lines provide 95 percent confidence intervals. The brackets above each health outcome report *p*-values from an interaction model that tests the equality of the coefficients between two subsamples. The hospitalization and prescription drug variables are defined over the time period age 65 to year 2010 and years 2005-2009, respectively. For scaling reasons, the number of prescribed packages of anxiolytics has been standardized to have mean zero and a standard deviation of one and then divided by 5.

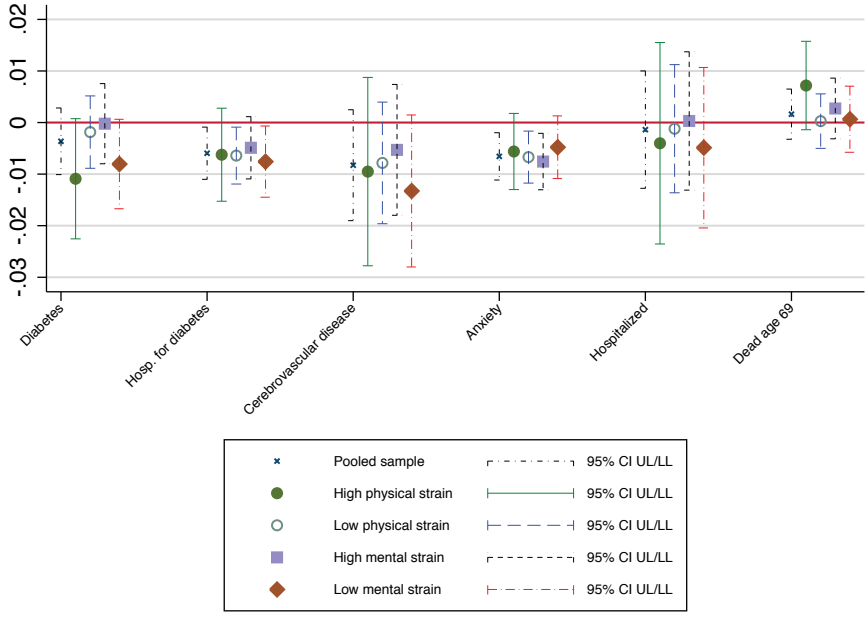
cupational strain. Using data on occupational strain provided by Kroll (2011), I estimate the model separately for workers in occupations with high/low physical strain and occupations with high/low psycho-social strain.[33] Some occupational groups in the treatment group are classified as having both physically and psycho-socially straining jobs. These include helpers and cleaners (SSYK code 912), medical care nurses (3231), housekeepers and restau-

[33] Kroll (2011) estimates so called Job Exposure Matrices (JEMs) for different occupational groups using survey data on working conditions of about 20,000 employees in Germany. These matrices are then used to create one "physical job index" and one "psycho-social job index" which range between 1 and 10 with higher values meaning higher burden. The index values are matched to my data via the 4 digit SSYK codes. I classify an occupation as physically or psycho-socially demanding if the corresponding index is larger than 8. As a result, 23 percent in the treatment group work in occupations that are classified as having high physical strain while 55 percent work in psycho-socially demanding occupations. In the control group, 37 percent work in physically straining occupations while 67 percent work in psycho-socially straining occupations.

rant workers (913 & 512). Some occupations, including home-based personal care (5133) and nursing associate professionals (323), have high psycho-social strain but low physical strain. The remaining occupations are neither physically or psycho-socially straining.[34] Because the data on occupational strain is limited only to a representative sample of private sector workers for which there is information on occupation, all control group workers are included in the regressions. For this reason I cannot test the equality of the coefficients between the two subsamples. Figure 6 shows the results.

The results show some evidence that individuals in occupations with high

Figure 6. Effect heterogeneity by occupational strain



Note: The figure shows the reduced-form estimates for the effect of the reform on various outcomes by occupational strain. The data on occupational strain is provided by Kroll (2011).

physical strain consume less health care due to the reform than individuals in occupations with low physical strain, diabetes-related drugs in particular. This result is in line with the hypothesis that individuals in demanding occupations experience a larger drop in physical activity at retirement and hence more adverse health effects. The second-stage differences are even larger because individuals in physically straining occupations respond less to the reform, exhibiting a first-stage effect of 2.8 months compared to 4.9 months for individuals in non-straining occupations. High mental strain, on the other hand, is associated with worse health outcomes in all cases but one, the consumption

[34]These include all other personal care-related jobs (513), nursing and midwifery professions (223), hairdressers (5141) and physiotherapists (3226).

of anxiolytics. Overall, though, the differences in health outcomes by occupational strain are quite small and most likely statistically insignificant.

Finally, the health effects of continued work could also depend on marital status. Social support from a spouse may help to buffer various shocks associated with retirement, diminished social interactions in particular.[35] Furthermore, married individuals may obtain greater pleasure from retirement than non-married individuals if there is complementarity of leisure between spouses. Continued work should then be more beneficial for non-married than married individuals. However, the opposite case is also conceivable. Increasing the normal retirement age essentially implies that individuals have to work longer to achieve the same pension level. Receiving news about such a reform at short notice could cause mental stress, which may be better dealt with in the presence of a spouse. I also look at heterogeneity with respect to the retirement status of the partner.[36]

The results from estimating the sample separately for married and non-married individuals, and separately for married individuals with and without a retired partner are shown in Figure 7. As for marital status, we see that two coefficients for married individuals reflect significant reductions in health care utilization (hospitalization due to diabetes and the consumption of anxiolytics) compared to none for non-married individuals. However, the estimates for the non-married individuals are less precise and in several cases quite close in magnitude to the corresponding estimates for the married individuals. In fact, the p-values from the interaction model imply that the two groups do not differ significantly with respect to any outcome.

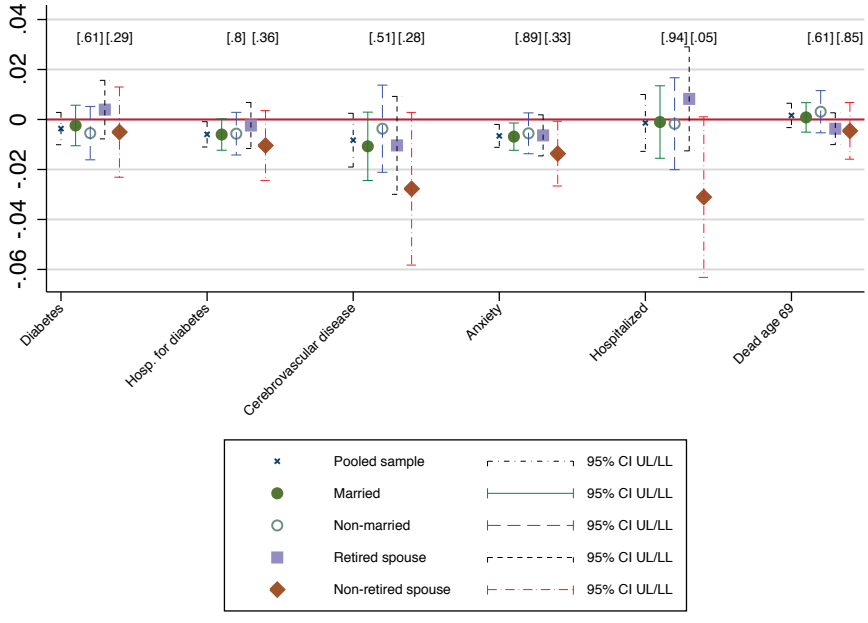
There is larger effect heterogeneity with respect to the retirement status of the partner. Individuals with a non-retired spouse are more likely to experience a reduction in health care utilization as compared to individuals with a retired spouse. The difference is most evident in the case of the probability of being hospitalized due to any cause. This difference is also statistically significant at the 5 percent level. This result suggests that the benefits of postponing retirement are larger when the partner is not yet retired. An analysis of the first stage reveals that individuals with a non-retired spouse respond more strongly to the reform with respect to the retirement age. For these individuals, the common first-stage effect amounts to 5.6 months compared to only 2.5 months for individuals with a retired spouse. Apart from providing evidence in favor of the complementarity-in-leisure hypothesis, this result implies that

[35] For example, Dave et al. (2008) find that the adverse health effects of retirement are mitigated if the individuals is married.

[36] The retirement status of the partner is based on the year in which the wife retires. Because the LOUISE data only covers individuals aged 16-65 during the period 1987-2000 and individuals aged 16-74 from 2001 and onwards, the information on retirement status of the partner is incomplete for married women who retire prior to year 2001 and whose husbands are born in 1934 or earlier. As a result, I only observe the retirement status of the partner for 82 percent of the married workers in the sample.

the second-stage differences in health outcomes are smaller than those of the reduced-form above.

Figure 7. Effect heterogeneity by marital status



Note: The figure shows the reduced-form estimates for the effect of the reform on various outcomes by marital status. The brackets above each health outcome report p-values from an interaction model that tests the equality of the coefficients between two subsamples.

8.3 Adjusting for multiple hypothesis testing

One concern with the analysis is that we may find spurious effects due to the number of outcomes considered. Following Persson and Rossin-Slater (2014) and Kling et al. (2007), I address this issue by creating two outcome indexes, one for physical health and one for mental health. The physical health index consists of all extensive and intensive margin measures of drugs that treat physical diseases (Tables 5 and 6, except for column (9)) and hospitalizations between age 65 and year 2010 (lower panels of Tables 9 and 10) as well as mortality by year 2011 (52 outcomes). The mental health index consists of all extensive and intensive margin measures of mental health drug consumption (Tables 7 and 8, 27 outcomes).

First, I invert each outcome so that a higher value represents a better outcome. Then, I standardize each modified outcome by subtracting the control group mean and dividing by the control group standard deviation. Finally, I take an equally weighted average of the standardized outcomes.

Table 12 presents OLS and 2SLS estimates using the two indexes as out-

Table 12. *The effect of postponing retirement on physical and mental health indexes*

	Physical Health Index		Mental Health Index	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
Employment, R	0.000523***	0.00130*	0.000468***	0.000838
	(0.0000388)	(0.000704)	(0.0000486)	(0.000884)
N	133,026	133,026	133,026	133,026
Mean dep. var.	0.057	0.057	0.053	0.053

Note: See text in section 8.3 for more information on how the indexes are constructed and Tables 2 and 4 for more information on the sample of analysis and controls. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

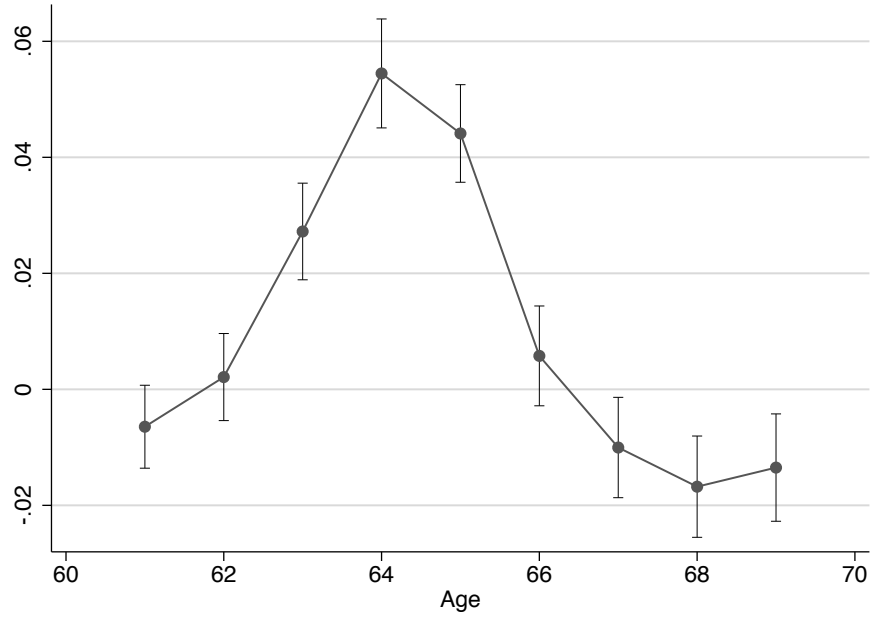
comes. Both 2SLS estimates are positive and also larger than the corresponding OLS estimate, but only statistically significant in the case of the physical health index. The results suggest that if postponing retirement has an impact on overall health, then this effect is likely to be small. Retirement is more likely to have an impact on certain medical conditions, physical health conditions in particular.

8.4 Income effects

One important aspect of estimating the health effects of reforms that promote later retirement is that these effects may operate through changes in lifetime income.[37] To illustrate the effect of the reform on lifetime income, I replace the dependent variable in Equation 8 with log disposable income at age a and estimate it for ages $a = 61, ..., 69$. The DD estimates from these regressions are shown in Figure 8. There is a positive and significant effect on disposable income of about 2–5 percent at ages 63–66, which corresponds to an annual increase in disposable income of SEK 3,500 to SEK 8,500. From age 67, the effect is negative and barely statistically different from zero. The conclusion from these results is that the income effects should be rather small and that potential health effects are more likely to operate through other channels.

[37] There is a large literature on the health effects of income loss due to unemployment (e.g. Black et al. (2012) and Eliason and Storrie (2009)). The direction of the income effect on health is not clear, however. Jensen and Richter (2004) show that an unexpected pension benefit reduction among Russian pensioners increased mortality, whereas Ruhm (2000) show that fatalities decline during recessions.

Figure 8. The effect on disposable income of the reform



Note: This figure plots the coefficient estimates from the interaction term in a DD specification of the effect of the reform on log disposable income at age a. See Tables 2 and 4 for more information on the sample of analysis and controls.

8.5 Robustness

I conducted a number of sensitivity analyses of the main findings.

To verify that the insignificant results for mortality do not merely result from the linear probability framework, I supplement the analysis with estimates from duration models. I estimate an exponential proportional hazard model in which the hazard of death individual i faces at age t is assumed to be given by,

$$h_i(t) = h_0(t)\exp(\psi\left(LG_s \times CH_{j\in[1938,1942]}\right) + \phi LG_s + \lambda_j + \mathbf{X}_{i,j,s}\theta) \qquad (9)$$

where the exponentiated parameters are defined similarly as in Equation 8. The key assumption of Equation 9 is that these covariates proportionally affect the baseline hazard $h_0(t)$.

Panel A of Table A.5 shows the estimated effects on mortality from different causes when survival time is right-censored at age 69. The estimates are shown as hazard ratios, so the (insignificant) estimate in column (1) of 1.034 (95% CI 0.91–1.15) means that the mortality risk among the affected workers increased by 3.4 percent as a result of the reform. The first estimate in Panel B shows that the hazard rate from the full sample (i.e. using mortality data up to 2011) is 1.008 (95% CI 0.92-1.09). Columns (2)–(12) report hazard ratios

for cause-specific mortality. Again, there is no evidence that postponing retirement affects the probability of death due to any of the listed causes. Thus, the duration model estimates confirm the conclusion from the main analysis of no effect on mortality from postponing retirement.

I make four additional robustness tests. First, I check whether the hospitalization and prescription results are robust to including individuals who died during the follow-up period. Second, I exclude individuals born in 1938 and 1939 who had the possibility of avoiding the new pension rules by retiring early (see section 5.1). If individuals in bad health truly retired early in order to avoid the new rules, there is an overrepresentation of healthy individuals in the 1938 and the 1939 cohorts in the treatment group which could explain why we find non-negative health effects of postponing retirement. Third, I test whether the results are sensitive to the choice of instrument, i.e. whether the just-identified model produces different results from the over-identified model. Fourth, I check whether the results are robust to the exclusion of control variables. Reassuringly, each of these tests produce results very similar to those presented in the paper.[38] While postponing retirement does not have an impact on mortality and overall health care utilization, it does seem to reduce the probability of being treated for anxiety, depression and diabetes. The results for cerebrovascular disease are robust to the two former tests, but less so for the choice of instrument and the exclusion of control variables.

9 Conclusion

This paper studies the causal effect of postponing retirement on health care utilization and mortality. Retirement is endogenous to health because individuals who retire later are more likely to be in good health. To estimate the causal effect of postponing retirement on health we need variation in retirement timing that is exogenous to health. For this purpose I utilize variation in retirement timing from a pension reform that raised the age at which broad categories of Swedish local government workers were entitled to retire with full pension benefits from 63 to 65. Because the new rules applied to individuals born in 1938 or later, I use both birth cohort and sector variation for identification.

The analysis is restricted to women because there were very few men in the largest worker categories that were affected by the reform, which include personal care-related workers, nursing professionals, cleaners and restaurant service workers. The control group is made up of female private sector workers of similar age. I use an IV-approach to estimate the effects of postponing retirement on post-retirement health, where retirement is instrumented by several interaction terms between being born in 1938 or later and working in the local government sector.

[38] The results are available upon request.

Health outcomes are measured using detailed individual-level administrative data on drug prescriptions, hospital admissions and mortality. I focus on the extensive and intensive margin of health care utilization, i.e. the prevalence and total consumption of inpatient care and drug prescriptions during a pre-specified period of time. ICD and ATC codes are used to classify health events into a number of relevant medical causes based on their appearance in previous literature on the association between retirement and health.

I first show that retirement age is negatively correlated with the consumption of health care and mortality – those who retire later consume less health care and are less likely to die early. The results from the IV analysis emphasize the importance of controlling for this selection issue. They reveal zero effects on the risk of dying early and on the overall consumption of health care.

The cause-specific analyses suggest that continued work is more likely to affect certain medical conditions than overall health. First, I document a significant reduction in the quantity of prescribed drugs that treat anxiety. Since there is no effect for the extensive margin, this result suggests that postponing retirement might alleviate short-term anxiety and depressive symptoms among elderly with pre-existing mental health issues. Second, postponing retirement also causes a significant decrease in the probability of being hospitalized due to diabetes. Third, there is evidence, although less robust, of a reduction in the risk of being treated for cerebrovascular disease. The mechanisms behind these results are difficult to investigate given the data at hand. However, since many of the risk factors for diseases of the circulatory system and diabetes are related to lifestyle, one interpretation is that old people are more likely to maintain a more healthy lifestyle (e.g. better diet and more physical exercise) at work than at home. Continued work may also uphold social interactions and good sleeping habits, which could have a positive impact on mental well-being.

The findings of this paper have timely and direct policy implications. First, the paper adds to the literature on the relationship between pensions and retirement, suggesting that the magnitude of the behavioral response to financial incentives could be quite large. The adjustment process might be quick if the new retirement age is already considered the norm in many other occupations. Second, and most importantly, the results suggest that raising retirement age thresholds would not have a serious impact on short to medium term health outcomes. In fact, retiring at a later age might even have positive effects on mental well-being and physical lifestyle. The external validity of the results could very well be questioned given that we focus on Swedish women in low- to medium-paid public sector jobs. Although a select group, the features of their pension are similar to those of many other sector-specific pension plans. Moreover, since various discussions of increasing the retirement age thresholds deal primarily with the concern that such increases could adversely affect individuals in low-skilled jobs, this focus could also be considered a strength.

References

Allgulander, C. and M. Msghina (2010). Nytta och risker med bensodiazepiner. *Läkartidningen 41*, 2025–2029.

Allgulander, C. and D. Nutt (2010). Benzodiazepines. In *Encyclopedia of Psychopharmacology*, pp. 218–223. Springer.

Andersen, T. M., N. Määttänen, and T. Valkonen (2014). The Nordic model – challenged but capable of reform. In Valkonen, T., Vihriälä, V. (Eds.), *The Nordic model – challenged but capable of reform*, pp. 113–144. Copenhagen: Nordic Council of Ministers.

Anderson, R. J., K. E. Freedland, R. E. Clouse, and P. J. Lustman (2001). The prevalence of comorbid depression in adults with diabetes a meta-analysis. *Diabetes care 24*(6), 1069–1078.

Andersson, J. and B. Tegsjö (2010). En rörlig arbetsmarknad – dynamiken bland jobb, individer och företag. In *Fokus på näringsliv och arbetsmarknad våren*, pp. 27–40. Örebro: Statistics Sweden.

Angelov, N., P. Johansson, E. Lindahl, and E.-A. Lindström (2011). Kvinnors och mäns sjukfrånvaro. *IFAU Rapport 2011:2*.

Atalay, K. and G. F. Barrett (2014). The causal effect of retirement on health: New evidence from Australian pension reform. *Economics Letters 125*(3), 392–395.

Atalay, K. and G. F. Barrett (2015). The impact of age pension eligibility age on retirement and program dependence: Evidence from an Australian experiment. *Review of Economics and Statistics 97*(1), 71–87.

Behncke, S. (2012). Does retirement trigger ill health? *Health economics 21*(3), 282–300.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2012). Losing heart? The effect of job displacement on health. *NBER Working Paper 18660*.

Bloemen, H., S. Hochguertel, and J. Zweerink (2013). The causal effect of retirement on mortality: Evidence from targeted incentives to retire early. *IZA Discussion Paper 7570*.

Bonsang, E., S. Adam, and S. Perelman (2012). Does retirement affect cognitive functioning? *Journal of health economics 31*(3), 490–501.

Börsch-Supan, A. and M. Schuth (2010). Early retirement, mental health and social networks. In Börsch-Supan, A., Brandt, M., Litwin, H., Weber, G. (Eds.), *Active ageing and solidarity between generations in Europe: First results from SHARE after the economic crisis*, pp. 218–223. Berlin: De Gruyter.

Bound, J. and T. Waidmann (2007). Estimating the health effects of retirement. *Working Paper 2007–168.* University of Michigan, Michigan Retirement Research Center.

Cesarini, D., E. Lindqvist, R. Östling, and B. Wallace (2015). Wealth, health, and child development: Evidence from administrative data on Swedish lottery players. *Quarterly Journal of Economics, Forthcoming*.

Charles, K. K. (2004). Is retirement depressing? Labor force inactivity and psychological well-being in later life. *Research in Labor Economics 23*, 269–299.
Clegg, A. and J. B. Young (2011). Which medications to avoid in people at risk of delirium: a systematic review. *Age and ageing 40*(1), 23–29.
Coe, N. and M. Lindeboom (2008). Does retirement kill you? Evidence from early retirement windows. *CentER Discussion Paper 2008-93*.
Coe, N. B. and G. Zamarro (2011). Retirement effects on health in Europe. *Journal of health economics 30*(1), 77–86.
Dave, D., I. Rashad, and J. Spasojevic (2008). The effects of retirement on physical and mental health outcomes. *Southern Economic Journal 75*(2), 497–523.
Eibich, P. (2014). Understanding the effect of retirement on health using regression discontinuity design. *SOEP papers on Multidisciplinary Panel Data Research No. 669*.
Eliason, M. and D. Storrie (2009). Does job loss shorten life? *Journal of Human Resources 44*(2), 277–302.
Feldstein, M. and H. Siebert (2009). *Social security pension reform in Europe*. Chicago: University of Chicago Press.
Franco, M., P. Orduñez, B. Caballero, J. A. T. Granados, M. Lazo, J. L. Bernal, E. Guallar, and R. S. Cooper (2007). Impact of energy intake, physical activity, and population-wide weight loss on cardiovascular disease and diabetes mortality in Cuba, 1980–2005. *American journal of epidemiology 166*(12), 1374–1380.
Försäkringskassan (2012). Sjukskrivningar i olika yrken under 2000-talet. Antal ersatta sjukskrivningsdagar per anställd 2002–2010. Sjukförsäkringsrapport 2012:14. http://www.forsakringskassan.se/wps/wcm/connect/53a6e173-bad8-4ee2-84df-d909665bfc13/socialforsakringsrapport_2012_14.pdf?MOD=AJPERES.
Försäkringskassan (2015). Sjukskrivningar 60 dagar eller längre. En beskrivning av sjukskrivna åren 1999–2014 efter kön, ålder, arbetsmarknadsstatus, yrke, sjukskrivningslängd och diagnospanorama. Socialförsäkringsrapport 2015:1. https://www.forsakringskassan.se/wps/wcm/connect/d7d4b78e-39fa-4c2f-bed9-ade979b5ff23/socialforsakringsrapport_2015_1.pdf?MOD=AJPERES.
Gelber, A. M., D. Jones, and D. W. Sacks (2013). Earnings adjustment frictions: Evidence from the social security earnings test. *NBER Working Paper 19491*.
Godard, M. (2016). Gaining weight through retirement? Results from the SHARE survey. *Journal of health economics 45*, 27–46.
Golden, S. H., M. Lazo, M. Carnethon, A. G. Bertoni, P. J. Schreiner, A. V. D. Roux, H. B. Lee, and C. Lyketsos (2008). Examining a bidirectional association between depressive symptoms and diabetes. *Jama 299*(23), 2751–2759.
Gorry, A., D. Gorry, and S. Slavov (2015). Does retirement improve health and life satisifaction? *NBER Working Paper 21236*.
Grossman, M. (1972). On the concept of health capital and the demand for health. *The Journal of Political Economy 80*(2), 223–255.
Hagen, J. (2013). A history of the Swedish pension system. *Working Paper 2013:7,* Department of Economics, Uppsala University.
Hallberg, D., P. Johansson, and M. Josephson (2014). Early retirement and post-retirement health. *IFAU Working Paper 2014:12*.

Hernaes, E., S. Markussen, J. Piggott, and O. L. Vestad (2013). Does retirement age impact mortality? *Journal of health economics 32*(3), 586–598.

Holbrook, A. M., R. Crowther, A. Lotter, C. Cheng, and D. King (2000). Meta-analysis of benzodiazepine use in the treatment of insomnia. *Canadian Medical Association Journal 162*(2), 225–233.

Holzmann, R. (2005). *Old-age income support in the 21st century: An international perspective on pension systems and reform*. Washington, DC: World Bank Publications.

Insler, M. (2014). The health consequences of retirement. *Journal of Human Resources 49*(1), 195–233.

Jensen, R. T. and K. Richter (2004). The health implications of social security failure: Evidence from the Russian pension crisis. *Journal of Public Economics 88*(1), 209–236.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Knol, M., J. Twisk, A. Beekman, R. Heine, F. Snoek, and F. Pouwer (2006). Depression as a risk factor for the onset of type 2 diabetes mellitus. a meta-analysis. *Diabetologia 49*(5), 837–845.

Kreiner, C. T., S. Leth-Petersen, and P. E. Skov (2014). Year-end tax planning of top management: Evidence from high-frequency payroll data. *The American Economic Review 104*(5), 154–158.

Kroll, L. E. (2011). Construction and validation of a general index for job demands in occupations based on ISCO-88 and KldB-92. *Methoden Daten Analysen 5*(63-90).

Kuhn, A., J.-P. Wuellrich, and J. Zweimuller (2010). Fatal attraction? Access to early retirement and mortality. *IZA Discussion Paper 5160*.

Lalive, R. and S. Staubli (2014). How does raising women's full retirement age affect labor supply, income, and mortality? Evidence from Switzerland. *mimeo*.

Laun, L. (2012). The effect of age-targeted tax credits on retirement behavior. *IFAU Working Paper 2012:18*.

Manoli, D. and A. Weber (2014). Nonparametric evidence on the effects of financial incentives on retirement decisions. *CESifo Working Paper 4619*.

Mazzonna, F. and F. Peracchi (2012). Ageing, cognitive abilities and retirement. *European Economic Review 56*(4), 691–710.

Midanik, L. T., K. Soghikian, L. J. Ransom, and I. S. Tekawa (1995). The effect of retirement on mental health and health behaviors: The kaiser permanente retirement study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 50*(1), S59–S61.

Moon, J. R., M. M. Glymour, S. Subramanian, M. Avendaño, and I. Kawachi (2012). Transition to retirement and risk of cardiovascular disease: Prospective analysis of the US health and retirement study. *Social science & medicine 75*(3), 526–530.

Munnell, A. H. and A. E. Sunden (2004). *Coming up short: The challenge of 401 (k) plans*. Washington, DC: Brookings Institution Press.

Neuman, K. (2008). Quit your job and get healthier? The effect of retirement on health. *Journal of Labor Research 29*(2), 177–201.

Palme, M. and I. Svensson (2004). Income security programs and retirement in Sweden. In Gruber, J., Wise, D. (Eds.), *Social Security Programs and Retirement around the World: Micro-Estimation*, pp. 579–642. Chicago: University of

Chicago Press.
Pan, A., M. Lucas, Q. Sun, R. M. van Dam, O. H. Franco, J. E. Manson, W. C. Willett, A. Ascherio, and F. B. Hu (2010). Bidirectional association between depression and type 2 diabetes mellitus in women. *Archives of Internal Medicine 170*(21), 1884–1891.
Persson, P. and M. Rossin-Slater (2014). Family ruptures and intergenerational transmission of stress. *IFN Working Paper 1022*.
Rohwedder, S., R. J. Willis, et al. (2010). Mental retirement. *Journal of Economic Perspectives 24*(1), 119–38.
Ruhm, C. J. (2000). Are recessions good for your health? *The Quarterly Journal of Economics 115*(2), 617–650.
Sahlgren, G. (2012). Work 'til you drop: Short-and longer-term health effects of retirement in Europe. *IFN Working Paper 928*.
Selin, H. (2012). What happens to the husband's retirement decision when the wife's retirement incentives change? *CESifo Working Paper No 3772*.
Socialstyrelsen (2015a). Läkemedel – statistik för år 2014. `https://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/19768/2015-3-17.pdf`.
Socialstyrelsen (2015b). Sjukdomar i sluten vård 1989–2014. `https://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/19921/2015-10-1.pdf`.
Szinovacz, M. E. and A. Davey (2004). Retirement transitions and spouse disability: Effects on depressive symptoms. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 59*(6), S333–S342.
Turnheim, K. (2004). Drug therapy in the elderly. *Experimental gerontology 39*(11), 1731–1738.
Vahtera, J., H. Westerlund, M. Hall, N. Sjösten, M. Kivimäki, P. Salo, J. E. Ferrie, M. Jokela, J. Pentti, A. Singh-Manoux, et al. (2009). Effect of retirement on sleep disturbances: The GAZEL prospective cohort study. *Sleep 32*(11), 1459.
Woolcott, J. C., K. J. Richardson, M. O. Wiens, B. Patel, J. Marin, K. M. Khan, and C. A. Marra (2009). Meta-analysis of the impact of 9 medication classes on falls in elderly persons. *Archives of internal medicine 169*(21), 1952–1960.

Appendix

Table A.1. *Aggregation of ICD and ATC classifications*

	Hospitalization and mortality (ICD-10)	Drug prescription (ATC)
Common causes		
Cancer	C00–D48	L01
Respiratory	J00–J99	R
Circulatory	I00–I99	C
Other	All other	All other
Hypothesis-based causes		
Physical health		
Ischemic heart disease	I20–I25	
Hypertension	I10–I15	
Heart disease		C01,C02,C03,C04,C07,C08,C09,C10
Cerebrovascular	I60–I69	B01
Musculoskeletal / Analgesics	M	A03D,A03EA,M01,M02A,M03
External causes	S00–T35	
Diabetes type 2	E10–E14	A10
Alcohol & tobacco	C0–C16,C22,C25,C30–C34,C53, C64–C65,C67,C92.0,C92.4,C92.5, E24.4, F10, G31.2, G62.1, G72.1, I42.6, K29.2, K70, K85.2, K86.0, T51, X45, X65, Y15	
Mental health		
Parkinson		N04
Antipsychotics		N05A
Anxiety		N05B
Hypnotics & sedatives		N05C
Antidepressants		N06A

Note: This table describes the aggregation of the World Health Organization's International Classification of Diseases (ICD) diagnoses and ATC Anatomic Classsification Codes (ATC) codes into common and hypotheses-based causes.

Table A.2. *Literature overview*

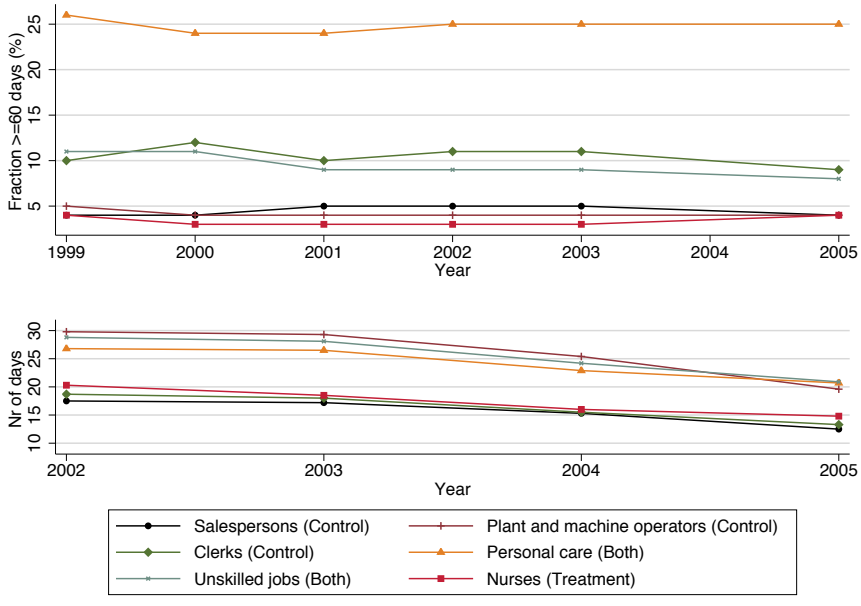
Article	Data and population	Empirical strategy	Health effect of retirement
Atalay & Barrett, 2014	Australian Health Surveys	IV; eligibility age reform	(+) Positive effect
Behncke, 2012	ELSA, 3 waves, UK	IV, age retirement incentives	(−) Increases risk of chronic condition
Bloemen et. al, 2013	Dutch register data	IV; early retirement reform	(+) Early retirement decreases mortality
Bonsang et. al, 2012	HRS 1998–2008, US	FE; IV, eligibility age for SS	(−) Negative effect on cognitive functioning
Börsch-Supan & Schuth, 2013	SHARE	IV; age retirement incentives	(−) Cognitive decline
Bound & Waidmann, 2007	ELSA, 2nd wave, UK	RD; health before and after retirement	(+) Positive effect on physical health for men
Charles, 2004	HRS, 2nd and 3rd wave	FE; IV, age retirement incentives	(+) Positive effect on subjective well-being
Coe & Lindeboom, 2008	HRS, 7 waves	IV; early retirement windows	(+/−) No long-term effects
Coe & Zamarro, 2011	SHARE	IV; age-specific retirement incentives	(+) Positive effect on general subjective health
Dave et. al, 2006	HRS, 7 waves	Panel data methodologies	(−) Negative effect on physical and mental health
Eibich, 2014	SOEP, Germany	RD based on German pension incentives	(+) Positive effect on self-reported and mental health
Godard, 2015	SHARE	IV; age retirement incentives	(−) Retirement increases obesity among men
Gorry et. al, 2015	HRS, US	IV; age retirement incentives	(+) Positive/no effect on life satisfaction/health care cons.
Grip et. al, 2012	Survey, Dutch public sector empl.	Unexpected change in replacement rate	(+) Positive effects on pre-retirement mental health
Hallberg et. al, 2014	Swedish register data, military	Introduction of early retirement scheme; DD model	(+) Reduces mortality and inpatient care
Hernaes et. al, 2013	Norwegian register data	Introduction of early retirement scheme; DD and IV	(+/−) No effect on mortality
Insler, 2014	HRS, 1992 wave	IV; self-reported retirement probabilities	(+) Positive effects
Kuhn et. al, 2010	Austrian registers, male blue-collar	Extension of early retirement scheme; DD and IV	(−) Increases mortality
Lalive & Staubli, 2014	Swiss administrative data, women	Full retirement age increase	(−) Mortality increases due to the reforms
Mazzonna & Peracchi, 2012	SHARE	IV; age retirement incentives	(−) Rate of cognitive decline increases
Neuman, 2008	HRS, US	IV; age retirement incentives, spouse's labor supply	(+) Positive effect on subjective health
Rohwedder & Willis, 2010	ELSA, SHARE and HRS	IV; age retirement incentives	(−) Negative effect on cognitive ability
Sahlgren, 2012	SHARE	FE and IV; age retirement incentives	(−) Overall negative effects

Table A.3. *Changes in pre-determined individual characteristics between treatment and control group*

	(1) Married	(2) Single	(3) Divorced	(4) Widow	(5) Immigrant	(6) Children	(7) Schooling	(8) Earnings	(9) Sick leave	(10) Hospitalized ages 56–60
Pre-reform cohorts										
Cohort 1935 * LG	0.00604	0.00422	0.00140	-0.0117	0.00399	0.00362	0.261***	0.0192	0.111***	0.000281
	(0.0116)	(0.00550)	(0.00905)	(0.00857)	(0.00740)	(0.00776)	(0.0664)	(0.0123)	(0.0294)	(0.0104)
Cohort 1936 * LG	-0.00788	0.00737	0.00830	-0.00779	0.00846	-0.000905	0.140**	0.0113	0.0499*	-0.00746
	(0.0115)	(0.00552)	(0.00909)	(0.00840)	(0.00735)	(0.00770)	(0.0661)	(0.0124)	(0.0284)	(0.0103)
Post-reform cohorts										
Cohort 1938 * LG	0.00127	0.00730	0.00472	-0.0133*	0.00284	0.00000820	0.0246	-0.00323	-0.0188	-0.00858
	(0.0112)	(0.00535)	(0.00886)	(0.00805)	(0.00704)	(0.00734)	(0.0643)	(0.0122)	(0.0269)	(0.00995)
Cohort 1939 * LG	0.00221	0.00259	0.00299	-0.00779	-0.00731	-0.0108	-0.0478	-0.0148	-0.0236	-0.0171*
	(0.0111)	(0.00531)	(0.00886)	(0.00782)	(0.00702)	(0.00720)	(0.0632)	(0.0121)	(0.0264)	(0.00980)
Cohort 1940 * LG	0.00760	-0.00385	-0.00232	-0.00143	-0.0111	-0.0108	-0.141**	-0.0308***	0.0126	-0.00660
	(0.0109)	(0.00529)	(0.00875)	(0.00762)	(0.00680)	(0.00710)	(0.0622)	(0.0118)	(0.0266)	(0.00974)
Cohort 1941 * LG	-0.0139	0.00754	0.00848	-0.00210	-0.00647	-0.0122*	-0.142**	-0.0373***	0.0286	-0.00366
	(0.0108)	(0.00531)	(0.00871)	(0.00756)	(0.00685)	(0.00705)	(0.0623)	(0.0117)	(0.0268)	(0.00956)
Cohort 1942 * LG	0.00942	-0.00836	0.00592	-0.00698	-0.00848	-0.0124*	-0.222***	-0.0647***	0.0928***	-0.00310
	(0.0105)	(0.00515)	(0.00848)	(0.00727)	(0.00645)	(0.00685)	(0.0604)	(0.0113)	(0.0267)	(0.00931)
N	133,026	133,026	133,026	133,026	133,026	133,026	133,026	133,026	133,026	133,026
Mean dep. var.	0.615	0.063	0.193	0.129	0.102	0.113	10.340	11.995	0.971	0.244

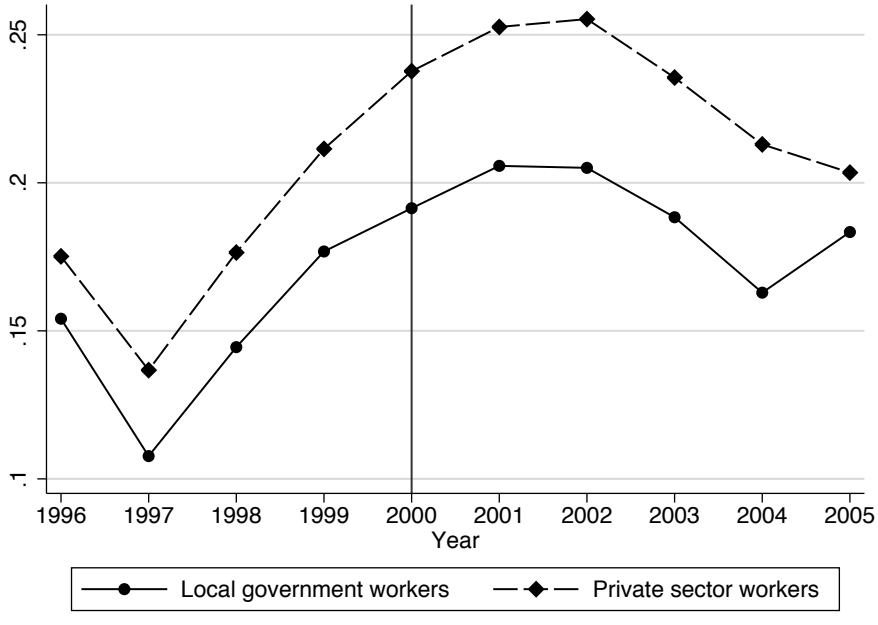
Note: This table reports estimation results from regressing the characteristic of interest on several interaction terms between cohort $j = 1935, ..., 1942$ and the local government dummy as well as the local government dummy itself and cohort-fixed effects (the 1937 cohort as reference). See Tables 2 and 4 for more information on the sample of analysis and controls. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

Figure A.1. Sick leave patterns by occupation



Note: The upper figure plots the fraction of female workers with more than 60 sick leave days, by year and occupation. The lower figure plots the average number of sick leave days. The data includes all employed female workers in the age group 18–64. "Unskilled jobs" refer to SSYK codes 91–93. The SSYK codes for the remaining occupations are provided in Table 1. Source: Försäkringskassan (2012, 2015).

Figure A.2. Sick leave patterns of younger workers



Note: This figure plots the fraction of female workers aged 45–50 with more than 14 consecutive days of sick leave for several years surrounding the reform. The solid and dashed lines denote individuals observed working in any of the treatment group occupations (see Table 1) and in the private sector, respectively. The dip in 1997 is the result of a temporary increase in the number of consecutive days an employee had to be sick in order to receive sickness benefits from the Social Insurance Agency rather than the employer. Between January 1 1997 and March 31 1998, the 14-day period was extended to 28 days. Since employer-provided sickness benefits are not observed in the data, sickness rates are lower during this period by definition.

Table A.4. *First-stage results for alternative retirement definitions*

	Claim age		Income-based	
	(1)	(2)	(3)	(4)
Post-reform CH * LG	0.493***		0.365***	
	(0.0252)		(0.0256)	
Cohort 1938 * LG		0.223***		0.198***
		(0.0397)		(0.0404)
Cohort 1939 * LG		0.397***		0.268***
		(0.0388)		(0.0403)
Cohort 1940 * LG		0.586***		0.464***
		(0.0370)		(0.0393)
Cohort 1941 * LG		0.604***		0.419***
		(0.0364)		(0.0395)
Cohort 1942 * LG		0.615***		0.446***
		(0.0349)		(0.0385)
Controls	Yes	Yes	Yes	Yes
Observations	133,026	133,026	133,026	133,026
Mean dep. var.	63.880	63.880	63.880	63.880
F-statistic	384.359	101.839	202.792	49.333

Note: Dependent variable: income-based retirement age, columns (1)–(2); claim age, columns (3)–(4). Estimated using OLS. Dependent variables right-censored at age 68. Columns (1) and (3) show estimates from Equation 6 and columns (2) and (4) from Equation 7. See Tables 2 and 4 for more information on the sample of analysis and controls. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

Table A.5. *Proportional hazard model estimates of cause-specific mortality*

		Common causes				Hypotheses-based causes						
	(1) All	(2) Cancer	(3) Resp	(4) Circ	(5) Other	(6) Ischemic	(7) Hyper	(8) Cerebro	(9) Muscu	(10) Alcohol	(11) External	(12) Diabetes
				Panel A: Mortality by age 69								
Post-reform CH * LG	1.034	1.014	1.018	1.145	1.088	1.083	1.006	1.194	0.651	0.953	1.140	1.426
	(0.0593)	(0.0736)	(0.303)	(0.143)	(0.188)	(0.180)	(0.290)	(0.243)	(0.358)	(0.105)	(0.529)	(0.435)
N	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026
				Panel B: Mortality by year 2011								
Post-reform CH * LG	1.008	1.042	1.177	1.083	0.986	1.085	1.180	1.012	0.946	0.994	1.008	1.364
	(0.0425)	(0.0597)	(0.247)	(0.106)	(0.117)	(0.141)	(0.247)	(0.161)	(0.399)	(0.0863)	(0.327)	(0.310)
N	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026	133026

Note: This table shows the estimated hazard ratios from estimating Equation 9. Circ: circulatory disease. Resp: respiratory disease. Ischemic: Ischemic heart disease. Hyper: hypertension. Cerebro: cerebrovascular disease. Muscu: musculoskeletal disease. See Tables 2 and 4 for more information on the sample of analysis and controls. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% level respectively.

More details on the reform

This section provides a more detailed description of the pre-reform occupational pension plan for local government sectors than section 3.2.

In the pre-reform pension plan, each worker received a lifetime annuity with an annual value calculated according to the following formula:

$$B(R,S) = \bar{q}(R) \times \gamma_j \times P_k(R) \times \min\left(\frac{S}{30}, 1\right) \tag{10}$$

This pension benefit is a function of years in service S, retirement age R, qualifying income $\bar{q}$, a benefit factor γ_j and a penalty factor $P_k(R)$. Qualifying income is the average salary, denoted in price base amounts (BA), earned during the best 5 of the last 7 years of employment. The value of the benefit factor represents the replacement rate for each income bracket j presented in Table A.6. The penalty factor $P_k(R)$ depends on the retirement age R and the group k to which the worker belongs. It reduces the pension $(P_k < 1)$ if the individual retires before the NRA. The benefit is also reduced if the individual has less than 30 years of qualifying income $(S < 30)$. The pre-reform pension plan guaranteed that the individual would get the pension benefit B by paying out the difference between B and the public pension.[39]

Table A.6. Gross replacement rates in the pre-reform pension plan (PA-KL)

Qualifying income $\bar{q}(R)$	Replacement rate γ_j
0 - 1 BA	96%
1 BA - 2.5 BA	78.5%
2.5 BA - 3.5 BA	60%
3.5 BA - 7.5 BA	64%
7.5 BA - 20 BA	65%
20 BA - 30 BA	32.5%
30 BA -	0%

Relating to the model in section 2, we can derive the slope of the budget constraint for an individual that chooses the optimal retirement age under this pension plan. To simplify, I assume that the worker has 30 years of qualifying income so that $B(R) = \bar{q}(R) \times \gamma_j \times P_k(R)$. For illustrative purposes, I also assume that health is exogenous to retirement. Substituting Equation 10 for

[39] The level of the public pension benefit PB was determined in the following way: $PB = 0.6 \times \bar{w} \times \min\left(\frac{S}{30}, 1\right) + \omega BA$ where $\bar{w}$ is qualifying income and ω is a parameter that determines the level of basic pension.

$B(R)$ in the budget constraint (2), gives the following slope:

$$\begin{aligned}\frac{dC}{dR} &= w_R \times (1-\tau) + \frac{d}{dR}\int_{t=R}^{T} (\bar{q}(R) \times \gamma_j \times P_k(R))\ dt \\ &= w_R \times (1-\tau) - \bar{q}(R) \times \gamma_j \times P_k(R) \\ &+ \left[\frac{\partial \bar{q}(R)}{\partial R} \times P_k(R) \times \frac{\partial P(R)}{\partial R} \times \bar{q} \times \gamma_j\right](T-R) \end{aligned} \tag{11}$$

The first term in Equation (11) captures the extra wage earnings that arises from postponing retirement by one year. The total change to pension wealth for this additional year of work, referred to as the pension wealth accrual, is measured by the second and third terms in Equation (11). The second term is the forfeited pension benefit that could have been collected had the individual retired in the current year. The third term is the change in the size of the pension benefit accumulated over the retirement period. This last term is determined by how the qualifying income $\bar{q}(R)$ changes due to continued work, and the rate at which early retirement is punished, that is $\frac{\partial P_k(R)}{\partial R}$. The penalty factor is the key feature of the analysis because it determines what the NRA is.

Table A.7 presents the penalty rates for two groups of local government workers: those who had a NRA of 63 and those who had a NRA of 65. The sum of the product between the entry in each cell and 12 (months) gives the penalty rate for a worker who retires at a given age.[40]

Table A.7. Adjustments for early/late withdrawals (in %) in the pre-reform pension plan for local government workers

Age	60	61	62	63	64	65	66
Normal retirement age = 63	−0.5	−0.45	−0.35	0	0	0.1	0.1
Normal retirement age = 65	−0.5	−0.5	−0.4	−0.3	−0.3	0.1	0.1

Note: The table reports penalty rates in the pre-reform pension plan, PA-KL, for workers who had a NRA of 63 (row 1) and for workers who had a NRA of 65 (row 2).

[40]For example, if a worker with a NRA of 63 retired at age 62 the gross pension benefit was reduced by $12 \times 0.35 = 3.85\%$. The corresponding reduction for a worker with a NRA of 65 would be $11 \times 0.4 + 12 \times 0.3 + 12 \times 0.3 = 11.6\%$.

IV. Income underreporting among the self-employed: a permanent income approach

Co-authored with Per Engström

Acknowledgments: We thank Bertil Holmlund, Sören Blomquist, Adrian Adermon, Torsten Persson, Spencer Bastani, Niels Johannesen, Christian Wittrock, Mikael Elinder, Che-Yuan Liang, Katinka Hort, Thor Olav Thoresen, and Håkan Selin for valuable comments. We also thank seminar participants at the Swedish Tax Agency, UCFS Workshop in Public Economics and Uppsala Brown Bag for their comments. Jan Wallander and Tom Hedelius Foundation (Johannes Hagen), Swedish Research Council for Health, Working Life and Welfare (FORTE grant: 217421002) and the Uppsala Center for Fiscal Studies (UCFS) are acknowledged for their financial support (Per Engström).

1 Introduction

Recent years have seen an increase in the application of indirect measures of tax evasion. Slemrod and Weber (2012) even describe it as an "explosion in empirical research on tax evasion", and entreat researchers on tax evasion to enlist in the "credibility revolution" (Angrist and Pischke, 2010) in empirical econometrics.

One of the modern workhorses in empirical research on tax evasion is the Pissarides and Weber (1989) method (henceforth PW). It is a clever indirect method of estimating the degree of income underreporting by self-employed individuals, who arguably have much better opportunities to evade taxes than wage earners do. In a nutshell, the method is based on using excess food consumption among self-employed as smoking gun evidence of income underreporting. Based on survey data on consumption (and income, if the researchers do not have access to registry based income measures), the consumption and income relations (Engel curves) may shed light on the true incomes of self-employed.

In this paper, we address one of the key methodological problems of the PW method: researchers typically only have access to current income measures, while theory suggests that a more permanent measure of the household's consumption potential may be more relevant. We remain agnostic to which income measure that is the most empirically relevant. It seems unlikely that current consumption would only be related to total lifetime income. Theory would only suggest this when we abstract from credit restrictions, uncertainty and other realistic features. However, it is also unlikely that consumption would only be related to current income. Our presumption is that the truth lies somewhere in between these two extremes – i.e. we presume that the most relevant income measure is not yearly (current) income but a more permanent measure, which we, for expositional convenience, simply denote permanent income.

It is hard to account for the fact that current income is a noisy measure of permanent income, without access to permanent income. Transitory income fluctuations attenuate the estimate of the income elasticity of food consumption which in turn may lead to overestimation of underreporting among the self-employed. Previous studies acknowledge the importance of using more permanent income measures when modeling food consumption, but given the typical cross-sectional design of survey data, it has proven difficult to come up with a good measure of permanent income. Pissarides and Weber (1989) try to account for this through instrumental variable (IV) techniques, which has subsequently become the standard way of approaching the problem. However, the IV solution relies on somewhat arbitrary exclusion restrictions; it is very hard to find variables that are closely correlated with permanent income but have no independent association with consumption.

We try to solve this problem in a more direct way by using a unique feature

of our consumption data. By merging the survey data on consumption to rich panel data from official tax and income registers, we can move towards a measure of permanent income by averaging household income both forwards and backwards in time. We then investigate how the estimate of underreporting is affected as we extend the time window over which we aggregate income. Specifically, does the PW method overestimate underreporting due to transitory income fluctuations, and if so, by how much?

Our approach is closely related to the analysis in Hurst et al. (2014) who exploit the panel dimension of the Panel Study of Income Dynamics to mitigate the effects of transitory income fluctuations. In line with a potential story of reduced attenuation bias, they document an increase in the food income elasticity as they move from current income to a three-year average income measure. The increased income elasticity, however, does not carry over to a lower estimate of underreporting. Having access to longer panel data, we take a more systematic approach to investigate this issue. We isolate the effects of transitory income fluctuations on the estimate of underreporting by keeping the sample of households intact across all income definitions.

Our results are highly consistent with a substantial degree of attenuation bias. The estimated food income elasticity increases by more than 40 percent as we move from current income to a 7-year average measure of household income. As a result, the estimated degree of underreporting falls by more than one-third.

The second part of our paper addresses the usual way to deal with transitory income fluctuations, namely IV estimation.[1] Most previous studies in the literature have used IV (Apel, 1994; Engström and Holmlund, 2009; Schuetze, 2002; Kim et al., 2009; Pissarides and Weber, 1989; Kukk and Staehr, 2014; Johansson et al., 2005; Martinez-Lopez, 2013).[2] Given the difficulty to find instruments that satisfy the exclusion restriction, we believe that finding a relevant measure of permanent income is a better way to deal with transitory income fluctuations than IV. However, in cases where data limitations rule out the former strategy, the researcher must decide on the most suitable set of instruments. So far, we know very little about the relative performance of different instruments.

Since we have access to a good measure of permanent income, we are in a unique position to evaluate the performance of different sets of instruments that have been used in the literature. To our knowledge, this is the first paper that systematically does so. Our results show that capital income appears to

[1] Due to the high quality of our registry based income data, we assume that the transitory fluctuations can be fully attributed to true transitory variation in household income, and not to classical measurement error in reported income.

[2] Kukk and Staehr (2014)) is, however, in some ways similar in spirit to our paper since they have access to "regular income" which they interpret as a more permanent income measure. Lyssiotou et al. (2004) develop a related demand system approach, extending the basic PW model that also relies on an instrumental variable approach.

be the best available instrument in terms of satisfying the exclusion restriction and producing estimates of the food income elasticity and income underreporting that are close to the "true" estimates using OLS and multiple-year averages of household income. Variables related to education and housing yield unreasonably large income elasticities and seem to belong both in the income and food equations.

The remaining of the paper is organized as follows. Section 2 explains how we account for transitory income fluctuations in the PW method. Section 3 describes the data and the key variables that relate to food consumption, income and self-employment. Section 4 discusses the OLS and IV results. Section 5 provides some robustness checks and section 6 concludes.

2 Estimating underreporting of income of the self-employed

This section briefly describes the expenditure-based estimation approach, originally developed by Pissarides and Weber (1989). We then show how we account for the effects of transitory income fluctuations by using panel data on household income. We borrow notation from Engström and Holmlund (2009).

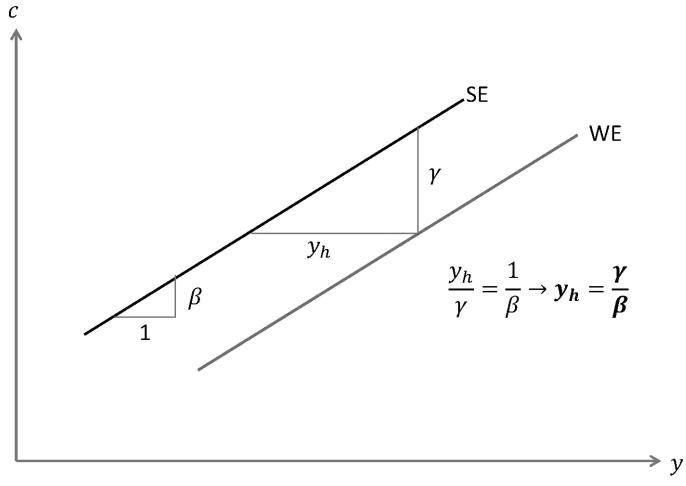
2.1 The basic model

The PW approach is illustrated by figure 1. Let c denote log food consumption, $c = \ln C^F$, and y log disposable income, $y = \ln Y^D$. The figure shows two log-linear Engel curves, one for self-employed households and one for wage earners. The intercept difference, γ, between these two curves reflect the degree of income underreporting among the self-employed.

Four central assumptions underlie this interpretation.[3] First, the elasticity of consumption with respect to income, β, is equal for the two groups. This is illustrated by the curves having the same slope. Second, there is no systematic misreporting of expenditures on food consumption between the two groups. The item of expenditure that most likely fulfills this assumption is food. There is little reason to lie about food consumption and it is also easy to report. Third, self-employed households systematically underreport their income by a constant factor, whereas wage earners are assumed to report their true incomes. If wage earners also underreport their income, the PW method will only provide a lower bound estimate. Finally, the researcher needs to assume that individuals who misreport their income in surveys misreport their income in the same way to the tax authorities. This is not a concern in our

[3] Since the primary focus of this paper is the distinction between current and permanent income, we will not provide a thorough examination of the validity of all these assumptions.

Figure 1. Engel curves for wage earners (WE) and self-employed (SE)



study since we collect information on household income from registry based data.

Pooling the data for self-employed households and wage earners, we can estimate the degree of underreporting among the self-employed by estimating an equation of the following form:

$$c_{it} = \mathbf{X}_i\alpha + \beta y_{it} + \gamma SE_{it} + \varepsilon_{it} \tag{1}$$

where subscript i denotes household i, subscript t denotes year t, $\mathbf{X_i}$ is a vector of variables affecting consumption, SE_{it} is a dummy variable for self-employed households and ε_{it} is a random error term. The parameter, γ, captures the intercept difference between the two Engel curves. The fraction of true income reported by the self-employed, κ, is identified as $\exp(-\gamma/\beta)$ and we form the estimate $\hat{\kappa}$ as $\exp(-\hat{\gamma}/\hat{\beta})$ using the estimated coefficients from Eq. 1. Following Hurst et al. (2014), we express our results in terms of the amount that the self-employed underreport their income, which is given by $1-\kappa$.

The derivation follows directly from figure 1. The degree of underreporting, or hidden income y_h, is given by $y_h = y_T - y_r$, where y_T is the log of true disposable income and y_r is the log of reported disposable income. The figure shows that $y_h = \gamma/\beta$, which implies that $y_T = \gamma/\beta + y_r$. In "un-logged" form the relation can be written $Y_T = e^{\gamma/\beta}Y_r$ which gives $Y_r/Y_T = e^{-\gamma/\beta} \equiv \kappa$, where Y_r and Y_T denote reported and true incomes, respectively. It follows directly that $Y_h/Y_T = 1-\kappa$, since $Y_T \equiv Y_h + Y_r$, where Y_h denotes hidden disposable income.

2.2 Accounting for transitory income

According to the permanent income hypothesis, the relevant income measure to include in the estimated equation is permanent income as opposed to current income. This causes a problem since measures of permanent income are less often available. Most studies have instead relied on current income measures. This may result in overestimation of the income underreporting by self-employed. The reason for this can be derived as follows.

Suppose that permanent income y_{it}^p is what matters for consumption in the Engel curve above (Eq. 1), and that current income is just a "noisy" version of permanent income, so that $y_{it} = y_{it}^p + \omega_{it}$. Further, assume that $\mathrm{E}(\omega_{it} = 0)$, $\mathrm{Var}(\omega_{it}) = \sigma_\omega^2$ and that ω_{it} is uncorrelated with y_{it}^p and ε_{it}. Then we can re-write Eq. 1 as:

$$c_{it} = \mathbf{X}_i\alpha + \beta y_{it} + \gamma SE_{it} + v_{it} \tag{2}$$

where $v_{it} = -\beta\omega_{it} + \varepsilon_{it}$. Transitory income fluctuations introduce attenuation bias in our estimates of β because $\mathrm{Cov}(y_{it}v_{it}) = -\beta\sigma_\omega^2 \neq 0$. Since the estimated degree of underreporting decreases in β, the attenuation bias will lead to overestimation of the true underreporting by self-employed.

Since measures of permanent income are rare, this causes a serious practical problem when applying the PW method. Several studies have used instrumental variable techniques to mitigate the effects of transitory income and measurement error. However, finding instruments that affect consumption only through permanent income is difficult, and it is impossible to directly test the exclusion restriction without access to permanent income.

We use a unique feature of our consumption data to solve this problem. By merging the consumption data to rich panel data from official tax and income registers, we can observe past and future income streams for each household that participates in the household survey. Specifically, we move towards a measure of permanent income by averaging household income both forwards and backwards in time. If current income is a noisy proxy for permanent income, we expect to see an increase in the size of $\hat{\beta}$ and a corresponding decline in our estimate of the degree of underreporting, $1 - \hat{\kappa}$, as we extend the time window over which we average income. Our access to a measure of permanent income also makes it possible to directly evaluate the performance of a set of instruments widely used in the previous literature.

The discussion above abstracts from a highly related problem addressed in Pissarides and Weber (1989). Pissarides and Weber (1989) recognize that current income may be a better approximation for permanent income for wage earners compared to self-employed. Self-employed typically have higher income fluctuations from year to year. They account for this by scaling (down) the estimate of underreporting by a factor that incorporates the relative income volatility between these two groups. However, this method does not address the fact that current income might be a bad proxy for permanent income for *both* groups, which is the focus of our study. By applying the correction

procedure suggested by Pissarides and Weber (1989) we find that the higher volatility of self-employed income is a problem of second order compared to the problem of using current instead of permanent income. Table A.2 in the appendix show that moving from current towards permanent income has a much larger effect on the underreporting estimate than what controlling for between-group differences in income volatility has.

3 Data

3.1 Consumption survey (HUT)

The consumption data comes from the Swedish Household Budget Survey (Hushållensutgifter, HUT). The household data is presented annually by Statistics Sweden. Around 4,000 randomly selected households are approached each year, of which slightly more than half participate in the survey. We use data from 2003–2009 with a total number of households of 15,044. The HUT data contain no panel elements.[4]

The participating households are asked to report their consumption expenditures during randomly selected two-week periods using a detailed expense manager. The expenditures are then multiplied by 26 to represent annual consumption. The households should also note whether the expenditures are associated with a certain household member. Various other questions are asked as to get information on household characteristics, including employment status, age, occupation, type of housing and number of children.

3.2 Income data (LINDA)

To calculate household income, we use the register based longitudinal database LINDA, constructed to be cross-sectionally representative of the Swedish population each year.[5] The data set is large; it contains 3.35 percent of the Swedish population each year corresponding to over 300,000 individuals. Information about individuals' incomes comes from official tax reports, so that the income variables are free from measurement errors that are common in survey data. Swedish register data on income are of very high quality because they are automatically third party reported (for wage earners) and are reported separately for different types of income.

We use LINDA from 2000–2012, which means that we can observe both past and future income streams of the households in the consumption survey. An additional advantage of using register data for incomes is that we may

[4]The design and main results of the HUT studies are presented in reports from Statistics Sweden (2003, 2004,... 2009).

[5]Edin and Fredriksson (2000) provide a detailed account of the data collection process for LINDA.

directly interpret the results in terms of tax evasion/avoidance. Most studies using the PW method relies on survey data instead of register data. It is not obvious, however plausible, that a tax evading self-employed individual also underreport incomes in surveys. See Hurst et al. (2014) for an in depth discussion on this matter.

Because LINDA is at the individual level, we aggregate income for the members of a given household to get household income. By construction of the HUT survey, one household member, referred to as the "sampled individual", is always part of LINDA. However, since LINDA and HUT use different household definitions, the remaining household members are not always part of LINDA. HUT households are self-reported and consist of individuals who share residence and have a common household budget. In LINDA, individuals must share residence and be registered as partners, or have children, in order to form a household. We will restrict the sample to households whose members are all part of LINDA.

3.3 Key variables and sample restrictions

The two key variables are annual food consumption and annual disposable income. Our main measure of food consumption is reported in the data as “food and alcohol purchases plus meals out". There are several reasons why food consumption is a suitable measure when applying the expenditure based approach. First, it is assumed to be mundane enough for individuals not to be afraid of reporting truthfully. An attractive feature is also that food needs to be bought often – yearly food expenditure is most likely better approximated by two weeks purchases than what e.g. yearly clothes expenditure is. Finally, it is less likely to be registered as a business expense, in which case it would be unclear how the expense was reported to the HUT survey, than most other expenditure categories.

Our main measure of household income is disposable income. Current income is defined as the household’s disposable income for the relevant HUT year. Disposable income is based on all types of (register based) income, including transfers, income from labor and self-employment and capital income. Taxes are deducted from gross income to get household disposable income.[6] We use past and future income records to create multiple-year average measures of income. This approach has been used by many others in the literature to construct measures of permanent income (Solon, 1992; Gottschalk et al., 1994; Hurst et al., 2014). For each household in year t, we compute income measures that average income between $t-i$ and $t+i$ where i ranges from 0 to 3. The effect of moving from current income to permanent income is il-

[6]Transfers consist of both taxable transfers, such as sick pay and unemployment benefits, and tax-free transfers, such as child allowances and social assistance payments. Capital income refers to interest income, dividends and net capital gains.

lustrated by extending the time window from $i = 0$ (no lag) to $i = 3$ (7-year average). For this to work, we restrict the sample to "stable" households that exist in LINDA in all 7 years between $t-3$ and $t+3$.[7] 3,052 households are dropped as a result.

We impose that household income must be positive not only for the current year, but also for all measures of permanent income. We also exclude households with income from farming. Households with negative income or zero reported food expenditures are also dropped from the sample. We restrict the sample to households where the household head is between 18 and 67 years old. As a result of these sample restrictions, 1,705 households are dropped. As mentioned in section 3.2, we restrict the sample to households that are the same in HUT and LINDA. 1,119 households are dropped as a result.[8] We are then left with 9,164 households.

Self-employment status of the household is based on information in LINDA. We define self-employed households as being a household where at least one of the adult members either report positive active business income or are considered as being connected to a closely held corporation (CHC).[9] As a robustness check, we employ two alternative definitions of self-employment that have been used in the literature. First, we define self-employment based on self-reported employment status in the HUT survey. Second, we define self-employed households as households with reported income from self-employment of at least 25 percent of total reported income.

As a first step, a household is classified as self-employed if the household is classified as self-employed in year t. We refer to this as the "unrestricted" definition of self-employment as households might transition from self-employment to employment (or vice versa) during the other years. To account for such transitions, our second definition is more restrictive, requiring that households are consistently classified as self-employed in all seven years. Dropping households that transition between self-employment and em-

[7]There are three potential reasons why households do not exist in LINDA in all 7 years. First, a household is removed from LINDA if the sampled household member dies or emigrates. Second, the household of a sampled person is removed if the adult members of the household no longer share residence or are no longer registered as partners. Third, new individuals are added to LINDA each year to ensure that each cross-section of LINDA is representative for the whole population (Edin and Fredriksson, 2000).

[8]Cohabiting couples without common children should be overrepresented in this group. We therefore expect these households to be younger and smaller in size. We also expect their incomes to be lower as we aggregate income over an incomplete set of household members. Table A.1 in the appendix confirm these observations. Koijen et al. (2015) also note that HUT and LINDA use different household definitions. Using HUT data from 2007, they find that 85 percent of the households have the same number of adults in LINDA, which comes very close to what we find.

[9]The relevant variables are nakte (”inkomst av aktiv enskild näringsverksamhet”), nakthb (”inkomst av aktiv näringsverksamhet för delägare i handelsbolag”) and bfoab (”kod för samgranskning med fåmansföretag”). This definition is used by Bastani and Selin (2014).

ployment, we are ultimately left with 7,728 households, of which 811 are self-employed. We refer to the resulting sample as the "restricted" sample. Since the empirical relevance of this distinction is not known a priori, we report estimation results for both samples.

3.4 Differences between wage earners and the self-employed

Table 1 reports descriptive statistics for the self-employed and the wage earners in the unrestricted sample (columns (1) and (2)) and in the restricted sample (columns (3) and (4)). On average, self-employed in the restricted sample are somewhat older, have marginally higher incomes and have larger houses than the self-employed in the unrestricted sample. Wage earners, on the other hand, do not differ much across the two samples.

Table 1. *Descriptive statistics*

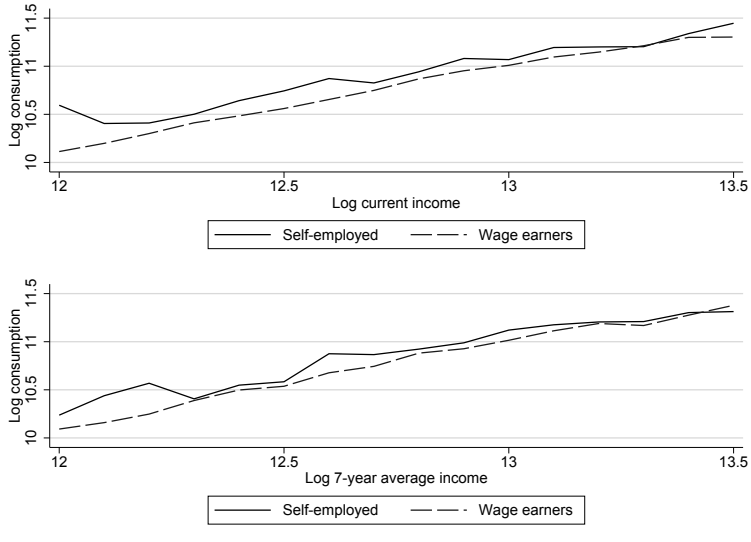
	Unrestricted sample		Restricted sample	
	(1)	(2)	(3)	(4)
	Self-employed	Wage earners	Self-employed	Wage earners
Food expenditure	11.01	10.83	11.03	10.82
	(0.600)	(0.622)	(0.590)	(0.625)
Current disposable income	12.97	12.86	12.97	12.85
	(0.584)	(0.522)	(0.589)	(0.514)
Disposable income 3-year average	12.99	12.86	12.99	12.85
	(0.549)	(0.496)	(0.535)	(0.493)
Disposable income 5-year average	13.00	12.86	13.01	12.85
	(0.534)	(0.490)	(0.530)	(0.489)
Disposable income 7-year average	13.00	12.86	13.01	12.85
	(0.520)	(0.485)	(0.513)	(0.484)
Age	47.94	46.76	49.32	46.75
	(10.08)	(11.02)	(9.575)	(11.05)
Household size	3.330	3.001	3.309	2.969
	(1.367)	(1.417)	(1.360)	(1.419)
High school	0.459	0.454	0.480	0.456
	(0.499)	(0.498)	(0.500)	(0.498)
College	0.383	0.399	0.334	0.396
	(0.486)	(0.490)	(0.472)	(0.489)
Single family house	0.662	0.582	0.671	0.573
	(0.473)	(0.493)	(0.470)	(0.495)
Size of house (m2)	138.1	116.3	140.4	114.6
	(52.47)	(47.54)	(51.86)	(46.63)
Property tax	7.162	5.675	7.462	5.544
	(3.559)	(4.147)	(3.323)	(4.179)
Capital income	6.444	4.783	6.733	4.663
	(4.422)	(4.179)	(4.454)	(4.147)
Observations	1454	7710	811	6917

Note: This table reports summary statistics for self-employed and wage earners. All numerical variables are converted to constant 2013 Swedish Crowns (SEK; 1 USD = 6.5 SEK in 2013) and are reported in logged form. For capital income and property taxes, we add one before taking logs. In the unrestricted sample, a households is classified as self-employed if the household is classified as self-employed in year t. The restricted sample consists of households that are consistently classified as either self-employed or wage earners in all 7 years between $t-3$ and $t+3$.

Focusing on between-group differences, spending on food is higher among the self-employed. The self-employed also have higher incomes, but they consume more relative to their income than wage earners. The difference amounts to around 1 percentage point, which is similar to what Hurst et al. (2014) find. As expected, the self-employed have more volatile income than wage earners. They are also slightly older, more likely house owners and have larger households. They have the same education level as wage earners, on average.

Figure 2 plots the relationship between food consumption and disposable income for wage earners and self-employed, respectively. Specifically, we plot the average of log food consumption for equally spaced income bins for households with incomes between the 5th and 95th percentiles. The upper panel uses current income and the lower panel uses our 7-year average income measure. First we see that the Engel curve for the self-employed lies above the Engel curve for the wage earners over essentially the entire income range. In other words, for a given amount of income, self-employed spend more on food than wage earners. Second, the Engel curves converge at lower income levels when we move from current income to 7-year average income. This is consistent with a potential story that the reduction in attenuation bias that follows from taking out transitory income fluctuations is most important at the lower end of the income distribution. Third, from visual inspection the central assumption of equal slopes seem to hold. We formally test this by adding an interaction term between the self-employment dummy and the relevant income measure to Eq. 2. As seen in Table A.3 in the appendix, the estimates of the interaction term are insignificant across all income measures.

Figure 2. Relationship between food consumption and income (bins)



4 Estimation results

In this section, we present the results from estimating Eq. 2. That is, we regress log food consumption on log disposable income and a set of control variables. The latter include age, age squared, number of children, number of adults, six dummies for H-region and year dummies. The OLS regressions also control for the set of variables that are used as instruments in the IV analysis, including two dummies for high school and college (the reference category is elementary school), single family house, size of house (square meters), log property taxes and log capital income. We report four sets of estimates of both β and γ and the corresponding estimated amount of underreporting $1-\kappa$, one for each income measure.

We begin by discussing the results from the unrestricted specification, that is, we define self-employed households as households in which at least one member was self-employed in year t. The results are reported in Table 2. The estimates of γ are positive and significant in all four specifications. The interpretation is that self-employed spend around 7 percent more on food relative to wage earners with the same reported income. The γ estimates are roughly stable when moving towards permanent income.

Table 2. *Estimation results, unrestricted sample*

	(1) Current income	(2) 3-year average	(3) 5-year average	(4) 7-year average
β	0.259***	0.345***	0.368***	0.398***
	(0.0358)	(0.0318)	(0.0337)	(0.0306)
γ	0.0728***	0.0726***	0.0672***	0.0678***
	(0.0226)	(0.0227)	(0.0227)	(0.0226)
1-κ	.245	.19	.167	.157
se $1-\kappa$	.069	.054	.052	.049
R2	0.408	0.417	0.420	0.423
No of SE	1454	1454	1454	1454
Obs	9164	9164	9164	9164

Note: This table shows the OLS estimates of the log-linear Engel curve from Eq. 2. The income elasticity is denoted as β and the coefficient on the self-employment dummy is denoted γ. We estimate the regression for different measures of household income (indicated across the columns). $1-\kappa$ is the estimated amount of underreporting of the self-employed, using the estimates of β and γ. The standard list of controls include age, age squared, number of children, number of adults, six dummies for H-region and year dummies. In addition, all OLS regressions control for the set of variables that are used as instruments in the IV analysis, including two dummies for high school and college, a dummy for single family house, size of house, log property taxes and log capital income. Robust standard errors in parentheses. We use the delta method to calculate standard errors (se) for the degree of underreporting $1-\kappa$. Household sample weights are used in all estimations.

We also see that the estimates of β increase substantially as the time window used to construct the income measure is extended. Column (1) shows

that the food income elasticity amounts to 0.26 when we use current income, while the 7-year permanent income measure gives a β estimate of 0.40. Furthermore, the β estimate increases substantially when moving from current income to the 3-year measure of permanent income. When subsequently moving to the 5 and 7-year measures the additional increases in the β estimate become smaller and smaller, indicating that the estimate has more or less stabilized when using the 7-year measure.

The estimates of underreported income, $1-\kappa$, mirror the pattern seen for the β estimates; we see a large initial drop from 0.25 to 0.19 when moving from current income to the 3-year average, followed by increasingly smaller drops as we extend the time window over which income is aggregated.

The overall results are highly consistent with a substantial degree of attenuation bias in the β estimate based on current income. The attenuation bias is remedied when moving towards a measure of permanent income – the β estimate increases and the estimate of underreported income decreases. Due to the covariance between estimates from the different specifications, it is not straightforward to determine whether the drop in estimated underreporting is statistically significant. When disregarding the potential covariance, we get a t-value of 2.95 when comparing the estimated income elasticity between current income and the 7-year average income measure. However, since the covariance is likely positive, this is a conservative estimate that probably underestimates the statistical difference in the β estimates.

The panel dimension of the data, and the ambition to think in terms of permanent measures as opposed to current measures, highlight a general problem with the PW method. The method generally relies only on current self-employment status. If this status is a bad proxy for permanent status the possibilities to underreport income may change dramatically over years. As an extreme example we can imagine a household that has wage earner status all years except the HUT year, in which it has self-employment status and reports much lower income than the true income. When adding past and future incomes for such a household, we are adding the true incomes. The estimate of underreporting may therefore be attenuated when not controlling for past and future employment status. In the next subsection we will therefore restrict the sample to households that remain self-employed or wage earners for the whole period, i.e. from 3 year before the HUT year until 3 years after the HUT year.

4.1 Persistence in self-employment status

In this subsection we will address the problem that self-employment status is not permanent. As seen from Table 3 there is rather high persistence in the employment status. The results from the table can be summarized as: if a household is self-employed in period t, the chance that it is self-employed in $t+1$ or $t-1$ is roughly 85 percent. The persistence in self-employment status is thus high but still far from permanent.

Table 3. *Persistence of self-employment status*

	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$
Fraction self-employed	0.693	0 .762	0.872	1	0.848	0.734	0.657

Note: Fraction of households self-employed in HUT year that were also self-employed in year $t-i$

We therefore restrict the sample to households that are categorized as self-employed or wage earners throughout the whole 7-year period, and reestimate the specifications in Table 2. The results are reported in Table 4. Using the restricted sample yields higher estimates of γ. For example, column (1) shows that self-employed now spend 11 percent more on food relative to wage earners with the same reported current income. This is what we would expect when we switch from a noisy indicator of self-employment to a more stable measure. As a result, the estimates of $1-\kappa$ estimates are somewhat larger than the corresponding estimates in Table 2. However, the pattern seen in Table 2 when moving towards a measure of permanent income is perfectly reproduced in Table 4 – the β estimate increases and the estimate of underreported income decreases.

Table 4. *Estimation results, restricted sample*

	(1) Current income	(2) 3-year average	(3) 5-year average	(4) 7-year average
β	0.291***	0.365***	0.384***	0.418***
	(0.0354)	(0.0390)	(0.0407)	(0.0364)
γ	0.110***	0.108***	0.101***	0.102***
	(0.0262)	(0.0264)	(0.0260)	(0.0260)
1-κ	.316	.257	.232	.217
se $1-\kappa$	.064	.056	.054	.05
R2	0.418	0.425	0.428	0.431
No of SE	811	811	811	811
Obs	7728	7728	7728	7728

Note: This table shows the OLS estimates of the log-linear Engel curve from Eq. 2. The sample is restricted to households that are categorized as self-employed or wage earners throughout the whole 7-year period. The income elasticity is denoted as β and the coefficient on the self-employment dummy is denoted γ. $1-\kappa$ is the estimated amount of underreporting of the self-employed, using the estimates of β and γ. Same control variables as in Table 2. Robust standard errors in parentheses. We use the delta method to calculate standard errors (se) for the degree of underreporting $1-\kappa$.

Thus, using a more persistent measure of self-employment instead of current self-employment status is empirically irrelevant for the magnitude of the decline in underreporting that is due to reduced attenuation bias. However, because the estimated *level* of underreporting depends on this definition, we will use the more restrictive definition in the remaining analyses of this paper.

4.2 IV results

The usual way of dealing with the attenuation bias in the estimate of β is through instrumental variable methods. The list of instruments used in the literature is very long: housing related variables, education, employment of the spouse, nationality and gender of the household head, and measures of capital income. As always when using IV methods, the suitability of the chosen instruments hinges on the exclusion restriction. The instruments are often somewhat arbitrarily excluded from the list of covariates and elevated to instrument variable status. To our knowledge, no study has systematically evaluated the performance of different sets of instruments.[10] We are in a unique position to do so. Since we have access to a good proxy for permanent income, we are able to evaluate the performance of different sets of instruments and explicitly test the exclusion restriction. We focus on three commonly used sets of potentially valid instruments: education, housing related variables and capital income.

In Table 5 we treat the estimates from Table 4 above as "the truth". Column (1) of Table 5 replicates the OLS estimates of column (4) in Table 4. We then compare the performance of the three separate families of instruments that are most frequently used in the previous literature. The instruments are separated into three categories: housing, education and capital income. The housing category includes log property taxes[11], house size and a dummy for single family house; the education category is captured by two dummies for high school and college, respectively; and the capital income measure is defined as log taxable capital income.[12] The estimation results from these IV specifications are reported in columns (3), (5) and (7). As before, we use the estimates of β and γ to construct estimates of $1-\kappa$. We also report first-stage F-statistics and p-values from the Sargan test of overidentifying restrictions.

Because the instruments we use are part of the usual set of controls, the set of controls will be different across the IV specifications. The IV controls will also be different from those used in the previous OLS analysis. To facilitate comparison, columns (2), (4) and (6) report the OLS estimates that are based on the same set of controls used in the IV specifications of columns (3), (5) and (7), respectively. As seen from the table, the exact choice of controls only has a marginal effect on the OLS estimates.

The results show that the estimates of the food income elasticity vary substantially across the IV specifications. The education instrument yields an

[10] Hurst et al. (2014) evaluate the performance of educational attainment as an instrument for household income. They do this by comparing the IV estimate of the food income elasticity for current income to the OLS estimate using a three-year average income measure. They find that the IV and OLS specifications produce similar estimates of β, which they interpret as evidence of instrument validity.

[11] The property tax amounts to approximately 1 percent of the assessed value of the house. The property tax was replaced by a (fixed) property fee in 2008. The relevant variables in LINDA are sfast ("fastighetsskatt") and sfavg ("fastighetsavgift"). Around 66 percent of the households pay property tax.

[12] The relevant variable is kkaps ("statlig taxerad inkomst av kapital"). This variable is restricted to be positive. Around 62 percent of the households have positive taxable income from capital.

Table 5. *IV estimation results, restricted sample*

	Ref.	Education		Housing		Capital income	
	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) IV	(6) OLS	(7) IV
β	0.418	0.445	1.281	0.441	0.657	0.421	0.399
γ	0.102	0.096	0.218	0.110	0.146	0.105	0.122
$1-\kappa$	0.217	0.194	0.156	0.220	0.199	0.221	0.263
se $1-\kappa$	0.050	0.048	0.028	0.048	0.038	0.050	0.065
First stage F-stat	.	.	47.724	.	98.956	.	83.227
Sargan p-value	.	.	0.655	.	0.001	.	.

Note: This table shows the IV estimates of the log-linear Engel curve from Eq. 2 for different instruments. The instruments used in columns (3),(5) and (7), respectively, are education (two dummies for high school and college), housing (log property taxes, single family house and square meters) and log capital income with the corresponding OLS estimates given in columns (2), (4) and (6). Each specification includes the complete set of controls (see Table 2) minus the relevant instruments. Column (1) reproduces column (4) of Table 2, our main OLS specification using the 7-year average income measure. The last row reports the p-value from the Sargan test for over-identification (only for multiple instruments).

estimate of β greater than one compared to 0.66 for housing. Using capital income as instrument yields a β estimate of 0.40, which comes very close to the "true" estimate of 0.42 given in column (6). It is also consistent with many empirical studies that find that food is a relative necessity.

The estimates of underreported income, $1-\kappa$, again mirror the pattern seen for the β estimates. The estimated underreporting is 15.6 percent and 19.9 percent for the education and housing estimates, respectively, which is lower than the corresponding OLS estimates. The negative effect of higher β on the estimated underreporting is partially offset by higher estimates of γ. The estimated underreporting for the capital income specification is 26.3 percent.

The fact that the β estimates are very different across instruments raise concern about the validity of the instruments. It is clear from the F-statistics that the instruments are relevant, i.e. have an impact on log disposable income. Capital income and housing related variables are most strongly correlated with income with F-statistics of 83.2 and 99.0, respectively. However, it is less clear that the exclusion restrictions hold for all the instruments. In general, it is impossible to test the exclusion restriction. However, since we have a good measure of permanent income we can provide an explicit test of the exclusion restrictions, under the assumption that our 7-year average captures the relevant relation between income and consumption. We do this by estimating Engel curves based on the 7-year average income measure and the three different sets of instruments. If the exclusion restriction holds, the instruments should be statistically insignificant. We perform F-tests of joint significance of the separate sets of instruments. The results are reported in Table 6.

We first note that education has a significant effect on food consumption. The F-test rejects the null hypothesis that the two education dummies are jointly equal to zero. The corresponding F-test for the housing variables is

also statistically significant with a p-value close to 0. Capital income, on the other hand, does not have a direct effect on food consumption when controlling for permanent income. The p-value of the F-test is equal to 0.33.

The housing and education variables appear to be directly related to food expenditure, in addition to their relation to income, and they also produce unreasonably large income elasticities. If they belong in the estimated model, they should be used as controls rather than instruments. Capital income appears to be the best available instrument in terms of satisfying the exclusion restriction and producing income elasticities that are close to the "true" elasticity.

Table 6. *Testing the exclusion restriction*

	(1)
Disposable income 7-year average	0.418***
	(0.0364)
Education	
High school	0.0548**
	(0.0236)
College	0.142***
	(0.0253)
Housing	
Single family house	0.0183
	(0.0260)
Size of house (m2)	-0.000180
	(0.000281)
Property tax	0.00823**
	(0.00340)
Capital income	
Capital income	0.00192
	(0.00197)
F-test education	18.871
p-value	0.000
F-test housing	6.502
p-value	0.000
F-test capital	0.953
p-value	0.329
R2	0.431
Obs	7728

Note: This table reports estimates from regressing log food consumption on our 7-year average income measure and the sets of instruments used in Table 5. We also include the standard list of controls (see table 2).

5 Robustness

Some studies, including this one, define self-employment based on reported business income (e.g. Pissarides and Weber (1989); Schuetze (2002); Kukk and Staehr (2014)). An alternative approach is to use self-reported employment status to determine whether a household is self-employed or not (e.g. Engström and Holmlund (2009); Hurst et al. (2014); Johansson et al. (2005); Kim et al. (2009)). Furthermore, Kukk and Staehr (2015), compare the two approaches based on Estonian data and find substantial differences in the resulting level of underreporting. In this section we therefore complement the above analyzes with a treatment based on the self-reported measure of employment status.

As an alternative to the income-based definition, we define self-employed households as households where at least one of the adult members consider themselves as being self-employed.[13] Even though self-reported employment status is only available for the current HUT year, this definition is more likely to capture persistent self-employment than our unrestricted, income-based definition of self-employment. This is confirmed by Table 7, which reports the fraction of self-employed households for each self-employment definition and across all sample restrictions. The fraction of self-employed households using the income-based definition is consistently higher across all sample restrictions but the last.

Table 7. *Self-employment shares across sample restrictions*

	Obs	Share SE (income-based def.)	Share SE (self-reported def.)
Original sample	15044	0.142	0.110
Stable households	11992	0.150	0.115
General restrictions	10287	0.154	0.130
HUT $\subset$ LINDA (unrestr.)	9164	0.159	0.130
Persistent SE status (restr.)	7728	0.105	0.102

Note: Column (1) describes the restrictions we impose to the original sample, column (2) reports the share of self-employed households according to our main definition of self-employment and column (3) reports the share of self-employed households according to self-reported self-employment status.

We reestimate the specifications in Table 4 on the restricted sample and use self-reports to define self-employment. The results are presented in Table 8. All β and γ estimates are remarkably close to the corresponding estimates in Table 4. Again we see that the estimated underreporting decreases by more than 30 percent when we move from current to permanent income.

We also reestimate the IV specifications in Table 5 to check whether the IV results are robust to using self-reported self-employment status. Table 9 reports the results. It turns out that the estimates are very similar to the previous IV estimates, which strengthens the conclusion that capital income is prefer-

[13] The survey asks individuals to report their main occupation. The alternatives include: employed, self-employed in unincorporated company ("enskild näringsidkare"), self-employed in an incorporated company ("aktiebolag"), farmer, student, unemployed, retired and "other".

Table 8. *Self-reported self-employment status*

	(1) Current income	(2) 3-year average	(3) 5-year average	(4) 7-year average
β	0.291***	0.365***	0.385***	0.418***
	(0.0355)	(0.0390)	(0.0407)	(0.0364)
γ	0.115***	0.116***	0.109***	0.108***
	(0.0280)	(0.0282)	(0.0281)	(0.0280)
1-κ	.326	.272	.248	.228
se $1-\kappa$	.068	.059	.058	.054
R2	0.418	0.425	0.428	0.432
No of SE	788	788	788	788
Obs	7728	7728	7728	7728

Note: Same as Table 4 except that we use self-reported employment status to define self-employed households. Restricted sample.

Table 9. *IV estimation results, self-reported SE status*

	Ref.	Education		Housing		Capital income	
	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) IV	(6) OLS	(7) IV
β	0.418	0.445	1.322	0.441	0.654	0.421	0.393
γ	0.108	0.093	0.218	0.116	0.147	0.111	0.125
$1-\kappa$	0.228	0.189	0.152	0.232	0.201	0.232	0.272
se $1-\kappa$	0.054	0.052	0.027	0.051	0.039	0.053	0.070
First stage F-stat	.	.	47.724	.	98.956	.	83.227
Sargan p-value	.	.	0.595	.	0.001	.	.

Note: The same as Table 9 except that we use self-reported employment status to define self-employed households.

able to the other two sets of instruments.

Finally, we test whether the results are robust to defining households as self-employed if they report having business income above a certain share of total income. The advantage of this definition is that we capture households that have a sizable part of their income from self-employment. Specifically, we follow Pissarides and Weber (1989) and define self-employment as consisting of all households with reported income from self-employment of at least 25 percent of total income from labor. The results from reestimating Table 4 using the share of business income to define self-employment are reported in Table 10. The sample of self-employed households is restricted to households that are consistently classified as self-employed in all 7 years. The sample of wage earners includes households with zero reported income from self-employment in all 7 years. As expected, using a less generous definition of self-employment, the share of self-employed households drops. Out of 7,481

households, 565 are now defined as self-employed. The results are consistent with our previous results.

Table 10. *Defining self-employment based on the share of business income to total income*

	(1) Current income	(2) 3-year average	(3) 5-year average	(4) 7-year average
β	0.292***	0.367***	0.382***	0.415***
	(0.0374)	(0.0410)	(0.0422)	(0.0377)
γ	0.0932***	0.0952***	0.0866***	0.0884***
	(0.0312)	(0.0315)	(0.0315)	(0.0315)
1-κ	.273	.228	.203	.192
se $1-\kappa$	.08	.068	.068	.062
R2	0.415	0.423	0.425	0.428
No of SE	565	565	565	565
Obs	7481	7481	7481	7481

Note: Same as Table 4 except that we define self-employed households as households with at least 25 percent of total reported income from labor being self-employment income in all 7 years. Wage earners include households with zero reported income from self-employment in all years. For CHC-owners, self-employment income is defined as the sum of income from self-employment, dividend payments and wage income.

6 Conclusions

In this paper, we analyze to what extent the Pissarides and Weber (1989) expenditure-based approach to tax evasion overestimates income underreporting among the self-employed due to transitory income fluctuations. A unique feature of our data allows us to merge the survey data on consumption to rich panel data from official tax and income registers. Transitory income fluctuations may cause an attenuation bias in the estimate of the food income elasticity. We mitigate this bias by moving towards a measure of permanent income by averaging household income both forwards and backwards in time.

Our results are highly consistent with a substantial degree of attenuation bias. The estimated food income elasticity increases by more than 40 percent as we move from current income to a 7-year average measure of household income. As a result, the estimated degree of underreporting falls by more than one-third. The results are robust to various definitions of self-employment.

Previous studies try to account for transitory income fluctuations through instrumental variable techniques. Since we have access to a good measure of permanent income, we evaluate the performance of different sets of instruments that have been used in the literature. Our results show that capital income appears to be the best available instrument in terms of satisfying the exclusion restriction and producing estimates of the food income elasticity and income underreporting that are close to the "true" estimates using OLS and

multiple-year averages of household income. Variables related to education and housing yield unreasonably large income elasticities and seem to belong both in the income and food equations.

We conclude that it is empirically relevant to account for transitory income fluctuations when applying the PW method. The preferred way of doing this is by constructing relevant measures of permanent income. However, when lacking panel data, our analysis also show that capital income performs well as an instrument for permanent income, while education and housing related variables do not satisfy the exclusion restriction in our application. The latter conclusion contrasts Hurst et al. (2014), who find, in a similar analysis using US data, that education measures performs well as instruments of permanent income.

References

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Apel, M. (1994). An expenditure-based estimate of tax evasion in Sweden. In *Essays on Taxation and Economic Behavior*, Economic Studies 18. Department of Economics: Uppsala University.

Bastani, S. and H. Selin (2014). Bunching and non-bunching at kink points of the Swedish tax schedule. *Journal of Public Economics 109*, 36–49.

Edin, P.-A. and P. Fredriksson (2000). LINDA – longitudinal individual data for Sweden. *Working Paper 2000:19.* Department of Economics, Uppsala University.

Engström, P. and B. Holmlund (2009). Tax evasion and self-employment in a high-tax country: Evidence from Sweden. *Applied Economics 41*(19), 2419–2430.

Gottschalk, P., R. Moffitt, L. F. Katz, and W. T. Dickens (1994). The growth of earnings instability in the US labor market. *Brookings Papers on Economic Activity 2*, 217–272.

Hurst, E., G. Li, and B. Pugsley (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of economics and statistics 96*(1), 19–33.

Johansson, E. et al. (2005). An estimate of self-employment income underreporting in Finland. *Nordic Journal of Political Economy 31*(1), 99–109.

Kim, B., J. Gibson, and C. Chung (2009). Using panel data to exactly estimate income under-reporting by the self employed. *Korea Institute for International Economic Policy Working Paper 09–02*.

Koijen, R., S. Van Nieuwerburgh, and R. Vestman (2015). Judging the quality of survey data by comparison with "truth" as measured by administrative records: Evidence from Sweden. In Carroll, C., Crossley, T., Sabelhaus, J. (Eds.), *Improving the Measurement of Consumer Expenditures*. University of Chicago Press.

Kukk, M. and K. Staehr (2014). Income underreporting by households with business income: Evidence from Estonia. *Post-Communist Economies 26*(2), 257–276.

Kukk, M. and K. Staehr (2015). Identification of households prone to income underreporting employment status or reported business income? *Public Finance Review*, 1091142115616182.

Lyssiotou, P., P. Pashardes, and T. Stengos (2004). Estimates of the black economy based on consumer demand approaches. *The Economic Journal 114*(497), 622–640.

Martinez-Lopez, D. (2013). The underreporting of income by self-employed workers in Spain. *SERIEs 4*(4), 353–371.

Pissarides, C. A. and G. Weber (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics 39*(1), 17–32.

Schuetze, H. J. (2002). Profiles of tax non-compliance among the self-employed in Canada: 1969 to 1992. *Canadian Public Policy 28*(2), 219–238.
Slemrod, J. and C. Weber (2012). Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance 19*(1), 25–53.
Solon, G. (1992). Intergenerational income mobility in the United States. *The American Economic Review 82*(3), 393–408.

Appendix

Table A.1. *Descriptive statistics matched and non-matched households*

	(1) Not matched	(2) Matched
Self-employed	0.123	0.159
	(0.328)	(0.365)
Food expenditure	10.90	10.86
	(0.544)	(0.622)
Current disposable income	12.40	12.88
	(0.673)	(0.533)
Disposable income 3-year average	12.44	12.88
	(0.633)	(0.507)
Disposable income 5-year average	12.46	12.88
	(0.612)	(0.500)
Disposable income 7-year average	12.47	12.88
	(0.609)	(0.493)
Age	44.28	46.95
	(12.59)	(10.88)
Household size	2.820	3.053
	(1.239)	(1.415)
High school	0.465	0.455
	(0.499)	(0.498)
College	0.397	0.396
	(0.490)	(0.489)
Single family house	0.455	0.595
	(0.498)	(0.491)
Size of house (m2)	110.6	119.7
	(56.03)	(49.00)
Property tax	4.017	5.911
	(4.155)	(4.095)
Capital income	3.539	5.047
	(4.391)	(4.262)
Observations	1123	9164

Note: This table compares households that did not have the same composition in HUT and LINDA (not matched) to those that had (matched).

Table A.2. *Alternative estimates of* $1-\kappa$ *with PW income volatility adjustment*

	(1) Current income	(2) 3-year average	(3) 5-year average	(4) 7-year average
β	0.291***	0.365***	0.384***	0.418***
	(0.0354)	(0.0390)	(0.0407)	(0.0364)
γ	0.110***	0.108***	0.101***	0.102***
	(0.0262)	(0.0264)	(0.0260)	(0.0260)
1-κ	.316	.257	.232	.217
se $1-\kappa$	.064	.056	.054	.05
$\hat{\sigma}^2_{yS}-\hat{\sigma}^2_{yW}$	0.128	0.092	0.092	0.083
PW $1-\kappa$	0.270	0.222	0.195	0.184
R2	0.418	0.425	0.428	0.431
No of SE	811	811	811	811
Obs	7728	7728	7728	7728

Note: This table shows the OLS estimates of the log-linear Engel curve from Eq. 2. This table is a replica of Table 2 and adds the estimates of $1-\kappa$ using the PW income volatility adjustment as outlined by Hurst et al. (2014) (online appendix).

Table A.3. *Equality of slopes*

	(1) Current income	(2) 3-year average	(3) 5-year average	(4) 7-year average
β	0.299***	0.377***	0.396***	0.431***
	(0.0397)	(0.0429)	(0.0449)	(0.0396)
$\beta * \gamma$	-0.0451	-0.0765	-0.0718	-0.0806
	(0.0536)	(0.0590)	(0.0604)	(0.0590)
t-test slope (p-value)	0.400	0.195	0.234	0.172
R2	0.418	0.426	0.428	0.432
No of SE	811	811	811	811
Obs	7728	7728	7728	7728

Note: This table reports the OLS estimate of the log-linear Engel curve from Eq. 2 after adding an interaction term between the self-employment dummy and the relevant income variable. The estimates of the interaction term reflect whether the assumption of equal slopes between self-employed and wage earners holds or not under each income measure.