VATT Publications 72

Essays on public economics

Tuuli Paukkeri

VATT INSTITUTE FOR ECONOMIC RESEARCH

VATT PUBLICATIONS

72

Essays on public economics

Tuuli Paukkeri

Valtion taloudellinen tutkimuskeskus
VATT Institute for Economic Research
Helsinki 2018

Tuuli Paukkeri, VATT Institute for Economic Research and Aalto University School of Business

ISBN 978-952-274-219-3 (nid.)
ISBN 978-952-274-220-9 (PDF)

ISSN 0788-4990 (nid.)
ISSN 1795-3332 (PDF)

http://urn.fi/URN:ISBN: 978-952-274-220-9

Valtion taloudellinen tutkimuskeskus
VATT Institute for Economic Research
Arkadiankatu 7, 00100 Helsinki, Finland

Juvenes Print Oy
Helsinki, May 2018

Aalto University

Essays on public economics

DOCTORAL DISSERTATION

Tuuli Paukkeri

Aalto University
School of Business
Department of Economics

A doctoral dissertation completed for the degree of Doctor of Science (Economics and Business Administration) to be defended, with the permission of the Aalto University School of Business, at a public examination held at the lecture hall B-200 in the main building of the school (Runeberginkatu 14-16, Helsinki, Finland) on 8 June 2018 at 12 noon.

Main dissertation advisor
Pertti Haaparanta, professor, Aalto University, Finland

Co-dissertation advisors
Matti Sarvimäki, associate research professor, Aalto University, Finland
Natalia Zinovyeva, assistant professor, Aalto University, Finland

Previous co-dissertation advisors
Yao Pan, assistant professor, Aalto University, Finland
Jukka Pirttilä, professor, University of Tampere, Finland

Preliminary examiners
Pierre Koning, professor, VU University Amsterdam, Netherlands
Tuomas Kosonen, research director, Labour Institute for Economic Research, Finland

Opponent
Pierre Koning, professor, VU University Amsterdam, Netherlands

Aalto University publication series
DOCTORAL DISSERTATIONS 81/2018

ISBN 978-952-60-7965-3 (printed)
ISBN 978-952-60-7966-0 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-7966-0

Abstract

This doctoral thesis is a collection of four essays in public economics that look at various public policies and their impacts on low-income and otherwise vulnerable individuals. The essays share the general aim of studying the effectiveness of public policies in achieving their stated goals. The first essay is single-authored by the candidate, and the latter three are collaborations with one or more co-authors.

In the first essay, I use a unique dataset compiled from Finnish registers and surveys to provide a comprehensive characterisation of the take-up behaviour of Finnish welfare benefits (housing allowance and social assistance) using descriptive methods. I provide various stylised facts on take-up and discuss how income dynamics matter for understanding take-up and benefit targeting.

The second essay focuses on the impact of information on benefit take-up. We study the information campaign in the context of the introduction of the guarantee pension program in Finland in 2011 and find that receiving a mailed information letter and application form significantly increased take-up compared to non-recipients.

In the the third essay, we analyse the impact of employers' disability insurance (DI) contributions on the incidence of disability pensions among their workers. Experience rating is used in DI in Finland in order to increase employers' incentives to prevent disabilities among their workers. We use detailed data and an empirical strategy that allows us to identify the causal effect of experience rating on disability inflow. Our analysis finds that the policy is not effective in reducing disabilities.

The fourth essay uses a theoretical framework to provide optimal tax and transfer rules for poverty reduction in developing countries. We modify the standard optimal tax framework by restricting tax instruments to be linear, which are more feasibly implemented in countries with a lower administrative capacity. We show that when we change from the standard objective of welfare maximisation to that of poverty minimisation, which better depicts the concrete objectives of such countries, the optimal tax and transfer rules are changed.

Tiivistelmä

Tämä väitöskirja koostuu neljästä julkistaloustieteen alaan kuuluvasta esseestä, jotka käsittelevät erinäisiä politiikkainstrumentteja ja niiden vaikutuksia pienituloisiin ja muilla tavoin haavoittuviin yksilöihin. Esseiden tavoite on vastata kysymykseen, kuinka hyvin nämä politiikkainstrumentit saavuttavat niiden eksplisiittiset tavoitteet. Ensimmäinen essee on väittelijän yksin kirjoittama, muut kolme on kirjoitettu yhteistyössä eri kirjoittajien kanssa.

Ensimmäisessä esseessä luon laajan katsauksen suomalaisten viimesijaisten sosiaalietuuksien, asumistuen ja toimeentulotuen, alikäyttöön. Esseessä käytetään kuvailevia menetelmiä yhdessä ainutlaatuisen, eri rekisteriaineistoista kootun aineiston kanssa. Muodostan niiden avulla useita tyyliteltyjä faktoja alikäytöstä ja tutkin, miten kohderyhmän tulodynamiikka vaikuttaa alikäyttöön ja tukien kohdentumiseen.

Toinen essee käsittelee informaation vaikutusta tukien alikäyttöön. Tutkimme suomalaisen takuueläkkeen voimaantulon yhteydessä vuonna 2011 toteutettua informaatiokampanjaa. Tutkimus osoittaa, että informaatiokirjeen ja hakulomakkeen saaminen postissa vaikutti merkittävästi tuen hakemisalttiuteen.

Kolmannessa esseessä tutkimme työnantajien työkyvyttömyyseläkevakuutusmaksujen vaikutusta työntekijöiden työkyvyttömyyseläkkeiden yleisyyteen. Suomessa on käytössä maksuluokkamalli, jonka tulisi kasvattaa työnantajien kannustimia ehkäistä työntekijöidensä terveysongelmia ennalta. Käyttämämme empiirinen menetelmä sekä aineistomme tarkkuustaso mahdollistavat vakuutusmaksujen kausaalivaikutuksen tunnistamisen. Analyysimme perusteella maksuluokkamalli ei toimi tavoitellulla tavalla työkyvyttömyyseläkkeiden vähentämisessä.

Neljännessä esseessä käytämme teoriamallia tarkastellaksemme optimaalista verojen ja tulonsiirtojen rakennetta köyhyyden vähentämiseen kehittyvissä maissa. Muokkaamme tavallista optimiveromallia ottamaan huomioon kehittyvien maiden heikomman hallinnollisen kapasiteetin rajoittamalla veroinstrumentit lineaarisiksi. Tutkimus osoittaa, että veromallin tavoitefunktion vaihtaminen tyypillisestä hyvinvoinnin maksimoinnista kehitysmaiden tavoitteita paremmin kuvastavaan köyhyyden vähentämiseen vaikuttaa optimaalisiin vero- ja tulonsiirtosääntöihin.

List of essays

This thesis consists of an introduction and the following four essays:

Essay 1: Take-up of welfare benefits: combining a static and dynamic perspective. *Unpublished manuscript.*

Essay 2: Does information increase the take-up of social benefits? Evidence from a new benefit program. *Joint with Tuomas Matikka. Unpublished manuscript. Available as VATT Working Paper 83/2016.*

Essay 3: Using a kinked policy rule to estimate the effect of experience rating on disability inflow. *Joint with Tomi Kyyrä. Unpublished manuscript. Parts of this research have been published in Finnish in report 07/2015 of the Finnish Centre for Pensions.*

Essay 4: Optimal taxation and public provision for poverty reduction. *Joint with Ravi Kanbur, Jukka Pirttilä and Matti Tuomala. Published in International Tax and Public Finance, Vol. 25 (1), pp 64–98. Open access publication, reprinted here under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).*

Acknowledgements

Producing this doctoral dissertation has taken a lot of encouragement and support from various individuals and institutions. I wish to extend my gratitude to all those who have contributed towards the finalisation of this extensive project.

Firstly, I wish to thank my supervisor Pertti Haaparanta for his encouragement, not only during the entire PhD process, but already during my M.Sc. degree. Jukka Pirttilä initially served as my co-supervisor, for which I am extremely grateful, as well as for the eventual co-authorship. Later, Pertti, Matti Sarvimäki, Natalia Zinovyeva, and originally also Yao Pan, served as members in my thesis committee, providing helpful comments and some healthy pushing. I am very lucky to have had supervision from so many talented researchers.

Three of my dissertation essays are co-authored. Tomi Kyyrä, Tuomas Matikka, Ravi Kanbur, Jukka Pirttilä and Matti Tuomala – I am extremely grateful for your co-authorship and for the encouragement, valuable lessons, and fun moments working with you has provided me. These joint research projects were the most instructive and enjoyable part of my degree.

The Department of Economics at Aalto University School of Business was my home base during my studies in the Finnish Doctoral Program in Economics. In addition to the academic support from the department, I wish to thank Maria Nyberg and Jutta Heino for their help on the administrative side. I also wish to thank Mikko Mäkinen and Otto Toivanen for the enjoyable teaching experiences as your teaching assistant.

A big part of my PhD has involved the VATT Institute for Economic Research. In addition to the co-authorships with Tomi Kyyrä and Tuomas Matikka, I am especially grateful to Tomi for offering me such a great opportunity to work on my thesis and pursue my research interests in the future. I also thank my colleagues for the support for the finalisation of my dissertation, and Tiina Heinilä and Nina Intonen for their help with the publishing process. And a big thanks goes to D3 corridor for the great company.

I also spent considerable time alongside my studies as a research assistant at the World Institute for Development Economics Research (UNU-WIDER). I would like to thank Finn Tarp and Tony Addison for the opportunity to get to know this international research institute, and the researchers and support staff for the warm atmosphere and research support.

I would like to extend my gratitude towards my pre-examiners Pierre Koning and Tuomas Kosonen for the time, effort, and expertise you have put into inspecting my work, and the valuable insight you provided in your examination reports. I thank Pierre Koning for additionally agreeing to act as my opponent.

During my studies, I had the opportunity to visit London School of Economics as a visiting graduate student. I am grateful for Henrik Kleven for hosting me, and for Jukka Pirttilä for encouraging and arranging the visit. I wish to thank the professors whose classes I took and whose help I received, especially Henrik Kleven and Johannes Spinnewijn, as well as fellow students, Anne Brockmayer, Anders Jensen, Francisco Costa, Sam Marden, Laura Derksen, Jon de Quidt and others, whose company and help I enjoyed.

Another international visit that mattered significantly for my research career was my undergraduate exchange at Pontificia Universidad Católica de Chile in 2006. The course 'Desarrollo Económico de America Latina' by Felipe Larraín sparked my interest in development issues and academic research, which eventually led me to pursue a PhD degree. My interests have since moved from development towards public economics, which occurred largely due to the inspiring course 'Public Economics I' taught by Jukka Pirttilä and Matti Tuomala at our doctoral program.

I received financial support from various sources, enabling me to focus on the work itself. I wish to thank the Finnish Doctoral Programme in Economics, Yrjö Jahnsson Foundation, HSE Foundation and OP-Pohjola Foundation for supporting financially my studies, the visit to LSE, and the acquisition of my unique dataset. Essay 2 benefited additionally from funding from the Strategic Research Council at the Academy of Finland, Essay 3 from the Finnish Centre for Pensions, and Essay 4 from the Academy of Finland, which I and my co-authors gratefully acknowledge. I would also like to thank Statistics Finland, the Social Insurance Institution, Finnish Centre for Pensions, Finnish Armed Forces and UNU-WIDER for providing data and research cooperation.

Thank you fellow students and friends Francesca Valentini, Liisa Laine, Mikaela Carlström, Robin Stitzing, Sinikka Parviainen, Rolle Magnusson, James

Corbishley, Anna Sahari, Hanna Virtanen, Milla Nyyssölä, Kaisa Alavuotunki, Nelli Valmari, Lassi Ahlvik, Maija Sirkjärvi, Michaela Schmöller, Paula Mäkelä, Terhi Ravaska, and many others, for the fun moments in and outside Economicum. It has also been inspiring to witness the ambition you each have for your research, and the hard work you put into it.

Thank you Mum Erja and Dad Pekka, for your never-ending support for me, and for the interest you always showed towards my work. You also made my graduation possible in a very concrete way, by lovingly caring for our son when I prepared for the defence.

Finally – thank you Iivo, for an immeasurable amount of support over these years. Being able to share the process with you has been an important resource for me. Your own ambition and methodicalness have also been a huge inspiration. Thank you for the ideas, troubleshooting, and encouragement, as well as the much needed counterweight to my work.

Table of contents

Introduction

This doctoral thesis is a collection of four essays that look at various public policies and their impacts on low-income and otherwise vulnerable individuals. The first two essays study the take-up of income transfer programs targeted at low-income individuals. The third essay studies the impact of employers' insurance contributions on the prevalence of disability pensions among their workers. The fourth essay uses a theoretical framework to provide optimal tax and transfer rules for poverty reduction in developing countries. The essays share the general aim of studying the effectiveness of public policies in achieving their stated goals (e.g. income support, prevention of disabilities).

The essays contribute to the field of public economics. The theme of the third essay additionally overlaps with labour economics, and the fourth one with development economics. The first essay is single-authored by the candidate, and the latter three are collaborations with one or more co-authors.

1 The take-up of social benefits

The public sectors of various countries offer an array of income transfer programs with various goals, such as providing income security in the face of adverse life events or reducing poverty and inequality in the society. When implementing such programs, policymakers must decide on several details: the target population, the application process, the level of the benefit, and so on. All these decisions matter for the program's impact, but in addition to the direct effect (for example, a narrow target group or low benefit level could result in a small aggregate impact) they may also have unintended effects through program participation. Formulating detailed and complex rules for social benefit programs can create high costs for the target group to learn about eligibility or to go through the application process. The literature studying benefit take-up typically considers information, transaction, and stigma costs (Currie, 2006), and recent research has additionally identified various psychological costs in program participation (e.g. Bertrand, Mullainathan and Shafir, 2006; Bhargava and Manoli, 2015). The balance of application costs and the expected size of the monetary benefit can become negative for some individuals, resulting in non-take-up.

It is indeed a wide-spread feature of means-tested social benefit programs that some proportion of the targeted individuals do not apply for the benefits they are entitled to. Currie's (2006) and OECD's (Hernanz, Malherbert and Pellizzari, 2004) overviews show that take-up rates across countries and programs can range from close to 100% to far below 50%. As suggested by Bhargava and Manoli (2015), the non-take-up of benefits should be viewed as a policy problem, which can be influenced by parameters chosen by the policy-makers. It needs to be recognised that the details of program implementation affect take-up costs and benefits, and consequently people's willingness and awareness to apply for benefits, which in turn affects the effectiveness of the program in reaching its goal.

The first two essays provide two different perspectives to the issue of non-take-up of social benefits. In the first essay, I characterise the take-up behaviour related to two benefits targeted at the poorest households in Finland: housing allowance and social assistance. The detailed and varied analysis allows me to characterise the overall importance of take-up costs for different kinds of households, relative to the benefit size, and the resulting impact on poverty alleviation. In the second essay, the focus is specifically on information and transaction costs. We use the introduction phase of the Finnish guarantee pension program and the related information campaign to illustrate how a simple information treatment impacted on the take-up rate of vulnerable individuals outside the labour force.

The first essay, studying housing allowance and social assistance, contributes to the literature on benefit take-up firstly by providing a comprehensive picture of welfare take-up in a static set-up, and secondly by taking income dynamics into account. Take-up literature typically studies the static context, but some researchers have suggested that the dynamics of the eligible households' circumstances could also matter for take-up behaviour (e.g. Blundell, Fry and Walker, 1988; Blank and Ruggles, 1996). Combining these two perspectives allows me to characterise take-up behaviour from various angles. I provide several stylised facts on take-up, and show that the benefits reach the main target population – households with long-term low incomes – whereas those experiencing short-term low income are more often left out due to non-take-up. Previous research on these particular benefits has studied take-up in a more narrow static set-up, but not the income patterns over time (Bargain, Immervoll and Viitamäki, 2012; Lyytikäinen, 2008). The results help to form a comprehensive picture of take-up behaviour and the nature of take-up costs, complementing the findings of Blank and Ruggles (1996), one of few studies considering the dynamic aspect of take-up.

Households with more variable income might anticipate becoming ineligible in the near future, which reduces the expected size of the benefit. For a short-term need, then, the costs of take-up are more likely to remain higher than the expected benefit. Households with short expected eligibility could for example try to find other means to cope without claiming the benefit. This is consistent with the standard economic hypothesis that eligible households calculate expected benefits and weigh them against take-up costs.

For this study, I construct a unique dataset from various registers and surveys of Statistics Finland as well as the Finnish Defence Forces. An important feature of the data is that I link income information and other characteristics over a longer time period to each annual data set in order to follow the eligible population's behaviour and experiences over a longer time window than allowed by the static, annual, data. With these data, I use graphical and statistical descriptive methods to portray take-up behaviour. In order to estimate take-up rates, I use microsimulation methods to determine which households are eligible for the benefits. I also study how the various simulation choices affect the take-up estimates, thus contributing to the literature using microsimulation methods as well as illustrating the robustness of the results.

Although the findings are informative of take-up behaviour among the population studied, part of the target population cannot be analysed with existing data. Some groups excluded from the current analysis could potentially have a lower propensity to take up or behave differently in the dynamic set-up. As better data becomes available, it would be important to study the take-up of these groups in a dynamic setting as well.

Whereas the first essay provides a wide view to the take-up behaviour of Finnish welfare benefits, the second essay focuses specifically on information and transaction costs in take-up. The context of this essay is the introduction of an entirely new social benefit, the Finnish guarantee pension program in 2011, and the Social Insurance Institution's (SII) campaign to raise awareness of the program among the eligible population. While the campaign consisted of various kinds of tools, its main component that we study was the directed mailing to part of the eligible population. The mailing consisted of a short information letter and a pre-populated application form together with a postage-paid return envelope, and was sent to those individuals the SII could most easily and reliably recognise as eligibles: the recipients of a full national pension. With our detailed data from the SII and Statistics Finland on all pensioners, we can identify the rest of the eligible pool and compare how the take-up behaviour of the recipients of the January 2011 mailing (treatment group) compares to those who did not receive this particular mailing (the control group). We provide both illustrative graphical and descriptive evidence, as well as causal regression estimates, of the impact of this mailing on take-up.

The overall take-up rate of the guarantee pension was very high: 93% of all eligible pensioners had applied for it by the end of 2011. Nevertheless, we find that the take-up rate as well as the speed of take-up of the January mailing recipients and non-recipients differed significantly. Using our main regression specification, we estimate the impact of the letter to be 33 percentage points, causing the treatment group take-up rate to be more than 50% larger than that of the control group. Our data set also contains several variables that inform on the health status of the target individuals, which allows us to study the impact of this kind of a simple information treatment on individuals with varying health status. We find that pensioners with medical expenses for severe or long-term illnesses do not react differently from those without such medication, suggesting that deteriorated health itself does not reduce the take-up effect of the mailing. Furthermore, severely ill pensioners who are less likely to manage their financial issues by themselves, respond even more strongly to the letter.

The essay's contribution to the literature lies especially in that we study individuals outside the labor force, namely the low-income elderly and disability pensioners. Earlier literature on take-up, and the literature on information provision, have mostly analysed groups with a tighter connection to the labour market or education (e.g. Bettinger, Long, Oreopoulos and Sanbonmatsu 2012; Bhargava and Manoli 2015; Liebman and Luttmer 2015). However, knowledge of how this kind of inexpensive and simple information provision affects take-up among non-working individuals – many with poorer health – is very relevant for practical policy making, as many social benefit programs target benefits to such vulnerable individuals.

2 Incentivising employers to reduce disability pension incidence

Whereas the previous two essays looked at income transfer programs that are targeted to individuals with low incomes, the third one looks at a program that provides income security in the face of a disability that prevents the individual from working full-time or at all. Instead of studying the take-up behaviour regarding disability pensions, the essay takes a different perspective on the efficiency of disability policies: how well does the current policy encourage the proactive prevention of such pensions. In many countries, disability benefit costs have been rising over recent years, which creates a burden on public budgets and pension systems. Work-disabling conditions are also a personal tragedy to those who encounter them. This has prompted governments to search for ways to curb the growth of expenses (OECD, 2010). Regarding this health-based benefit, one channel is to try to improve the health of workers so that disability benefits are needed less often. In this endeavour, in addition to the workers themselves, employers can potentially play a role as well. For example, they can take care of the working conditions of their workers, provide part-time work to suit workers' health needs, and so on.

However, since it is costly for the employers, they may invest too little in such disability-preventing measures from the society's point of view. In Finland and the Netherlands, experience rating of disability insurance (DI) premiums is used to increase employers' incentives for such investment. In experience rating systems, the employer's insurance premium reacts to the prevalence of disability benefit claims among its workforce: employers with a high disability risk face higher insurance costs whereas employers with a low disability risk pay lower insurance premiums. This system should ideally cause employers' incentives to be aligned closer to those of the society's, encouraging them to invest more in measures that reduce disabilities. In this essay, we estimate what effect the experience rating system in DI has on the incidence of sickness and disability in Finland.

To study the effect of experience rating, we take advantage of an institutional feature that allows us to identify its impact from other confounding effects. In Finland, a firm's degree of experience rating depends on its size: the smallest firms are not subject to experience rating at all, whereas the largest firms are fully experience-rated. For firms in between, the degree of experience rating increases from 0 to 1 with the firm's size. These discontinuities ("kinks") in the experience rating rule at the threshold values for small and large firms allow us to use a regression kink design (Nielsen, Sørensen and Taber, 2010; Card, Lee, Pei and Weber, 2015) to identify a causal effect from experience rating. We also benefit from the availability of detailed register data from the Finnish Centre for Pensions and Statistics Finland that cover all private-sector firms and their employees over the period 2007–2013, as well as worker sick leave periods and disability pensions, together with their medical diagnoses.

Due to the rareness of experience rating in disability insurance across countries, there is not much previous research on its efficiency in this context, and the existing evidence is inconclusive (Koning, 2009; van Sonsbeek and Gradus, 2013; Kyyrä and Tuomala, 2013; de Groot and Koning, 2016). The Finnish institutional set-up, and our detailed data, allow us to contribute to understanding better the efficiency of incentives provided by experience rating in the context of DI.

Our analysis suggests that experience rating does not help to reduce sick leaves or disability benefit claims. Different sized firms have differential incentives to prevent sickness and disability incidence, but we find no differences in incidence rates between them. Using the regression kink analysis, we find no evidence of the degree of experience rating having an impact on the incidence of sick leave or disabilities. Our data also allow us to disaggregate the inflows to different types of disability benefits, and also by medical condition, but the results do not vary along these dimensions, either.

Thus, our analysis casts doubt on the efficiency of experience rating in DI as a way to reduce disabilities. A possible explanation is that the current design of the Finnish experience rating scheme is too complex for employers to properly grasp the impact of disability incidence on their insurance premiums, thus hindering their incentives to act proactively.

3 Optimal tax and transfer policies for poverty reduction

The first three essays study individual public programs and their effectiveness in reaching their goals. It is also important to consider the public sector as a whole, incorporating taxes, transfers and other tools together in the analysis. This is naturally very difficult to do empirically, but can be done in a comprehensive manner using a theoretical framework. This exercise is relevant for developed economies as well, but in the fourth essay we take the viewpoint of a developing country, where policymakers are often focused on poverty alleviation but the administrative capacity of the public sector can be more limited, and modify the framework accordingly.

The optimal taxation literature (in the tradition of Mirrlees (1971)) assumes that the public sector is capable of implementing complex instruments such as non-linear income taxes. Many of the results of this literature are therefore not directly useful for policy recommendations in developing countries. We therefore study linear instead of nonlinear income taxes in our model and also consider the administrative requirements that different instruments impose on the public sector. For example, linear income taxation can be implemented by combining a proportional income tax and a lump-sum transfer. Such taxes can be withheld at source, which is administratively easier than to realise a fully non-linear income tax schedule requiring the accounting of all incomes over the course of the year from all sources. We study all of the most relevant redistributive instruments from this perspective: income taxation, income transfers, taxes and subsidies on commodities, and public provision of public and private goods.

Another motivation for the focus on the developing country context is that in such countries there is often an explicitly expressed goal to reduce poverty and to distribute the fruits of economic growth more evenly, by the countries' governments themselves as well as the wider development community. In fact, one of the Sustainable Development Goals (SDGs) formulated by United Nations in 2015 is simply "End poverty in all of its forms everywhere". However, standard optimal tax analyses typically assume that the government is instead interested in well-being, modeled as the maximisation of a social welfare function which aggregates individual utilities in the society. We study how the optimal tax rules are affected when the goal of poverty reduction is explicit in the model.

We find that, compared to the welfare-maximising optimal linear income tax, the poverty-minimising optimal tax formula includes additional pressure towards lowering the marginal tax rate in order to boost earnings to reduce income poverty. Our numerical simulations however show that this mechanism is offset by the redistributive concerns, as the optimal tax rates are higher under the poverty minimisation objective than under welfare maximisation. We observe a more drastic result when studying commodity taxation: setting poverty minimisation as the government's objective changes completely the conditions under which uniform commodity taxation is optimal, as set out by Deaton (1979). Under poverty minimisation, uniform commodity taxation is unlikely to be ever optimal. In practice, however, the administrative difficulties of differentiated commodity taxes should also be taken into account.

The findings complement earlier studies on optimal linear taxation (e.g. Tuomala, 1985; Piketty and Saez, 2013) and on non-welfaristic objectives (e.g. Kanbur, Keen and Tuomala, 1994; Pirttilä and Tuomala, 2004). The analysis also illustrates how the theoretical framework can be modified to bring out conclusions that are relevant to the policy context at hand.

Bibliography

Bargain, Olivier, Herwig Immervoll, and Heikki Viitamäki, "No claim, no pain. Measuring the non-take-up of social assistance using register data," *Journal of Economic Inequality*, 2012, *10* (3), 375–395.

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir, "Behavioral economics and marketing in aid of decision making among the poor," *Journal of Public Policy & Marketing*, 2006, *25(1)*, 8–23.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu, "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment," *Quarterly Journal of Economics*, 2012, *127* (3), 1205–1242.

Bhargava, Saurabh and Dayanand Manoli, "Psychological frictions and the incomplete take-up of social benefits: evidence from an IRS experiment," *American Economic Review*, 2015, *105* (11), 3489–3529.

Blank, Rebecca M. and Patricia Ruggles, "When do women use Aid to Families with Dependent Children and Food Stamps? The dynamics of eligibility versus participation," *The Journal of Human Resources*, 1996, *31* (1), 57–89.

Blundell, Richard, Vaness Fry, and Ian Walker, "Modelling the take-up of means-tested benefits: The case of housing benefits in the United Kingdom," *The Economic Journal*, 1988, *98(390)*, 58–74.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber, "Inference on causal effects in a generalized regression kink design," *Econometrica*, 2015, *83* (6), 2453–2483.

Currie, Janet, "The take-up of social benefits," in A. Auerbach, D. Card, and J.M. Quigley, eds., *Public policy and the income distribution*, Russell Sage Foundation, 2006, pp. 80–148.

de Groot, Nynke and Pierre Koning, "Assessing the effects of disability insurance experience rating. The case of The Netherlands," *Labour Economics*, 2016, *41*, 304–317.

Deaton, Angus, "Optimally uniform commodity taxes," *Economics Letters*, 1979, *2*, 357–361.

Hernanz, Virginia, Franck Malherbert, and Michele Pellizzari, "Take-up of welfare benefits in OECD countries: A review of the evidence," OECD Social, Employment and Migration Working Papers 17, OECD 2004.

Kanbur, Ravi, Michael Keen, and Matti Tuomala, "Optimal non-linear income taxation for the alleviation of income-poverty," *European Economic Review*, 1994, *98*, 701–719.

Koning, Pierre, "Experience rating and the inflow into disability insurance," *De Economist*, 2009, *157* (3), 315–335.

Kyyrä, Tomi and Juha Tuomala, "Does experience rating reduce disability inflow?," VATT Working Paper 46, VATT Institute for Economic Research 2013.

Liebman, Jeffrey B. and Erzo F.P. Luttmer, "Would people behave differently if they better understood social security? Evidence from a field experiment," *American Economic Journal: Economic Policy*, 2015, *7* (1), 275–299.

Lyytikäinen, Teemu, "Studies on the effects of property taxation, rent control and housing allowances," VATT Research reports 140, VATT Institute for Economic Research 2008.

Mirrlees, James A, "An exploration in the theory of optimum income taxation," *Review of Economic Studies*, 1971, *38* (114), 175–208.

Nielsen, Helena Skyt, Torben Sørensen, and Christopher R. Taber, "Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform," *American Economic Journal: Economic Policy*, 2010, *2* (2), 185–215.

OECD, *Sickness, disability and work: Breaking the barriers*, OECD Publishing, 2010.

Piketty, Thomas and Emmanuel Saez, "Optimal labor income taxation," in Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, eds., *Handbook of Public Economics, vol. 5*, Vol. 5 of *Handbook of Public Economics*, Elsevier, 2013, pp. 391 – 474.

Pirttilä, Jukka and Matti Tuomala, "Poverty alleviation and tax policy," *European Economic Review*, 2004, *48*, 1075–1090.

Tuomala, Matti, "Simplified formulae for optimal linear income taxation," *Scandinavian Journal of Economics*, 1985, *87*, 668–672.

van Sonsbeek, Jan-Maarten and Raymond H.J.M. Gradus, "Estimating the effects of recent disability reforms in the Netherlands," *Oxford Economic Papers*, 2013, *65* (4), 832–855.

The take-up of welfare benefits: combining a static and dynamic perspective

Abstract

Incomplete take-up is a common phenomenon across various social benefit programs. Understanding why eligible individuals do not claim their benefits can help in the design of efficient public programs. In this paper I use unique and detailed data from Finland to provide new stylised facts on the take-up of welfare benefits. I extend the standard static framework by taking eligible households' income dynamics into account. I find that eligibility to the benefits is concentrated among the worst-off households, but that eligible households who do not take up the benefits are experiencing only a short-term fall in income, from which they recover shortly afterwards. On the contrary, households who claim benefits typically have permanently low income. The findings are consistent with households reacting to the benefit schemes in a rational manner, weighting expected benefits to take-up costs. Take-up costs thus do matter, but do not seem to screen out the most needy families.

Keywords: Take-up of social benefits, social assistance, housing allowance, targeting of benefits

JEL classification codes: I38, D31, H53

1 Introduction

For many social benefits, the proportion of eligibles actually claiming and receiving the benefit is well below 100%. Incomplete take-up can significantly reduce the potential of a social program to achieve its goals. On the other hand, hurdles that reduce take-up can also be seen as useful for targeting the benefits to the most needy and keeping public budgets in check. To be able to design an optimal policy that balances these two opposing aspects, it is important to understand the drivers of take-up behaviour in detail.

In this paper, I study the take-up of Finnish welfare using a unique dataset combined from Finnish registers and surveys. I focus on two welfare benefits targeted at the low-income population: social assistance and housing allowance. The study contributes to the literature on benefit take-up firstly by providing a comprehensive picture of welfare take-up in a static set-up, and secondly by taking income dynamics into account. I characterise take-up from various angles, and show that the benefits reach the main target population, the chronically low-income households, whereas those experiencing short-term low income are more often left out due to non-take-up. Previous research on these benefits has studied take-up in a more narrow static set-up, but not the income patterns over time (Bargain, Immervoll and Viitamäki, 2012; Lyytikäinen, 2008).

The literature on benefit take-up has traditionally studied the impact of benefit size and claiming costs on take-up (Currie, 2006). The costs are often divided into information costs, transaction costs and stigma costs, and the importance of each cost type varies across countries and social program contexts. Previous studies have utilised reforms or experiments to show that these costs can be changed to influence take-up, and the claiming costs should be seen as a policy variable that can be decided upon just as benefit levels and other characteristics of the programs can (Bettinger et al., 2012; Bhargava and Manoli, 2015; Zantomio, 2015). Claiming costs can also be used to screen out applicants for whom the benefits are not intended. Kleven and Kopczuk (2011) suggest that there is an optimal level of take-up costs that balances between improving take-up of the eligibles and preventing the leaking of benefits to non-eligibles. While I am not able to distinguish between different types of take-up costs, I characterise the overall importance of these costs for different kinds of households by illustrating who is screened out by take-up costs.

The first set of results relate to the static take-up patterns. First, take-up rates are estimated to be varying around 70–90% for both benefits. These relatively high rates likely reflect the focus on a sample for whom eligibility can be more accurately estimated, and from which some groups with potentially lower propensity to take-up are excluded. Second, take-up probabilities are increasing in the size of the benefit, which is consistent with optimising behaviour over a fixed take-up cost and the benefit size. Third, both benefit eligibility and take-up are concentrated among the worst-off households – those with the lowest incomes, lower education, more unemployment, less earnings, and so on. The received benefits reduce the overall share of low-income (in relative terms in the Finnish context) households in the population by 1.8 percentage points, and the share could be slightly reduced further if non-take-up could be eliminated. The eligible samples forego on average 4% of their disposable income in housing allowance benefits and 7% in social assistance benefits.

Whereas the results listed above are illustrative of the cost side of the take-up decision, I characterise the benefit side in the decision by relating the take-up behaviour to household income and other characteristics over time. For example, households accustomed to more variable incomes are likely to expect a shorter benefit duration and, therefore, should be less prone to take up. Several researchers have suggested incorporating such dynamic elements into the assessment of take-up. Blundell, Fry and Walker (1988, p.66) suggest that individuals who expect "their circumstances to improve in the near future may not consider it worth claiming", and that "we might expect those with fluctuating incomes or experiencing frequent changes of circumstances to take up less than those whose circumstances are more constant". Anderson and Meyer (1997) also note the importance of expected unemployment duration and benefit duration in unemployment benefit take-up. Blank and Ruggles (1996) establish that better-off women often have short eligibility spells for welfare programs in the US, that they become quickly ineligible because of increasing income, and are less likely to take up the benefit. Worse-off women experience longer eligibility spells and also claim the benefits more often. Beyond these articles, there is not much research on the issue of dynamics in take-up, potentially due to a lack of suitable data.

To further our knowledge on income dynamics and take-up, I take advantage of the availability of data on individuals' income and labour market participation over several years and combine this dynamic information with the static take-up analysis described above. This analysis gives the fourth key finding: I illustrate that households who do not claim the housing allowance benefits in a particular year despite being eligible are on average experiencing a short-term income fall, which they recuperate during the next year. The claiming households, on the other hand, seem to have permanently low income, and they are also typically long-term benefit recipients. The non-claiming households also have more variability in their income over time. 30% of the eligible non-claimers experience a major income drop from the previous year, whereas only 7% of eligible claimers experience a similar drop. The non-claimers also experience a sharp rise in their income in the following year more often than the claimers (46% vs. 20%, respectively). A similar pattern also emerges for social assistance take-up behaviour. However, when controlling for other characteristics at the same time, the variance of household income in the years preceding the take-up decision is correlated with non-take-up of social assistance, but not housing allowance. I also study the impact of income changes just before or after the take-up decision as well as labour market status changes and household composition changes over a longer time.

This observed pattern regarding income and take-up is consistent with the standard economic hypothesis that eligible households calculate expected benefits and weigh them against take-up costs – since claiming is costly, households with more variable income might anticipate being ineligible in the following year and therefore do not take up this year despite their current, short-term, need. For a short-term need, the cost of take-up is more likely to remain higher than the expected benefit. Households might also try to find other means to cope without claiming the benefit. There could also be more confusion about income support availability and eligibility rules when incomes fluctuate or change suddenly.

To say something further about the potential underlying reasons behind non-take-up, I link additional information to the data to study two potential mechanisms. I use information on parental income to measure the availability of outside resources such as loans from close relatives, and on school and ability test scores to roughly measure the households' general ability to cope with the welfare system. Using a sub-sample for whom these measures are available, this analysis provides the fifth main result: I find evidence that a lower test score of the household is correlated with a higher take-up probability. This suggests that non-take-up does not (at least to a large extent) occur because of an inability to cope with the system. To the contrary, the benefits reach on average more disadvantaged households. Furthermore, parental resources do not correlate with take-up, indicating that at least this channel of outside resources is not typically used as an alternative to applying for welfare benefits.

Even though these last results can only be suggestive due to the smaller sample size, together with the dynamic income patterns they help to form a wider picture of take-up behaviour and the nature of take-up costs, complementing the findings of Blank and Ruggles (1996).

Using the dynamic data I can also deepen our understanding of the impact the benefits have on relative poverty.[1] In addition to looking at static yearly shares of low-income households in the population, I find that households who are chronically low-income over the observation period exit poverty due to their benefit income much less often than those whose incomes are more variable around the low-income threshold. The share of low-income households could be further reduced by eliminating the non-take-up of the benefits, but many individuals would still remain close to the relative poverty line, at just above or below, at risk of transient poverty.

By characterising the take-up of Finnish welfare benefits in this varied way, I uncover features in take-up behaviour that have clear policy relevance. The observation that non-claiming families mostly suffer from low income only in the short term is useful for evaluating the targeting accuracy of the benefit to the most needy. The current analysis suggests that benefits are concentrated among households with the lowest incomes and those with long-term need, whereas those with temporary low income are reached less often. The suggestive findings regarding ability and take-up reinforce the finding that the benefits are targeted quite successfully towards the most needy families. The take-up costs thus do not seem to screen out the most needy families. It needs to be kept in mind, however, that the sample studied here is a selected one, and leaves out some groups that could potentially have a lower propensity to take-up, or behave differently in the dynamic set-up. The findings are thus not conclusive about the take-up behaviour of these Finnish welfare benefits.

The paper proceeds as follows. I first describe the nature of the Finnish welfare benefits – housing allowance and social assistance – and the broader context of income support. I then describe the data and methods used in this study, and then turn to the analysis of the take-up of housing allowance and

[1] I study relative income-poverty in the Finnish context: a household is defined as low-income if their income is less than 60% of the national median.

social assistance. The final section concludes.

2 Institutions

Finnish social benefits form two tiers: the first tier consists of benefits granted on the basis of the individual's current health, family, or labour market condition, and second tier benefits are based primarily on low income. First tier benefits include e.g. student allowances, sickness benefits, maternity benefits, child benefits, and unemployment benefits[2]. These benefits often entail earnings-related components, and they are typically granted by the Finnish Social Insurance Institution (SII). The second tier consists of housing allowance and social assistance, which can be considered to form the minimum living standard for Finnish residents. Both are means-tested and tax-free benefits that depend on household composition and income level.[3]

Housing allowance. Housing allowance (HA) is intended to help low-income families cover their housing costs consisting of the rent or maintenance fee for owner-occupiers, heating and water use costs.[4] The legislation is very detailed, determining maximum acceptable housing costs per square meter based on a host of factors: the number of household members, the surface area of the dwelling, the municipality where the dwelling is located, the year when the building was constructed or fundamentally remodelled, and the type of heating system in the building. Housing costs in excess of the maximum limits are not covered by the benefit. In addition, maximum dwelling sizes are also stipulated in the rules (37 m^2 for a single, 57 m^2 for two persons, and so on), and if the household lives more spaciously, the costs allocated to the square meters surpassing these limits are not covered.

[2]There are three kinds of unemployment benefits. The earnings-related unemployment insurance benefits (UI) are available only to those who participate in a private unemployment fund. Those who are not eligible for earnings-related benefits but fulfil a work history condition, are eligible for basic unemployment allowance (UA) from the Social Insurance Institution (SII). These benefits last for a maximum of 500 days. If the unemployed person uses up these benefits and is still not re-employed, he is eligible for labour market subsidy (LMS), which has a lower benefit level but no maximum duration. Individuals who don't fulfil the work history condition of UI or UA, such as the long-term unemployed, and the young who only entered the labour market recently, are also eligible for LMS.

[3]The description in this section corresponds to the legislation that was in place during the study period 2003–2011. Monetary amounts are illustrated in 2011 levels.

[4]Two separate housing allowance schemes exist for pensioners and students, but in this paper I focus on the general housing allowance scheme that is targeted at working-age households.

The above rules determine 'reasonable housing costs' that can be covered by the benefit. In addition, household income affects the size of the benefit. The benefit formula can be described as:

$$HA = (reasonable\, housing\, costs - deductible) \times 80\,\%. \tag{1}$$

The deductible in the formula refers to a means test, which is based on household composition and average long-term gross income and assets. Long-term income refers to income that can be expected to last for at least the following 5 months, or if the household experiences considerable fluctuations in income, expected average income for the following 12 months. Social assistance is excluded from the income measure. If the household's assets, such as savings and investments in financial instruments, are above limits set in the regulation, 15% of the assets above the limit are calculated as annual income. For the poorest applicants, the deductible is zero. In any case, the benefit always covers a maximum of 80% of acceptable costs, reflecting the principle that households should cover at least a part of the costs themselves. The benefit is granted by the SII, typically for one year at a time.

The general housing allowance is an important benefit for low-income families, and typically especially so for the unemployed. Hannikainen-Ingman et al. (2012) estimated that in November 2009, 53% of labour market subsidy (LMS) recipients and 37% of basic unemployment allowance (UA) recipients received housing allowance benefits. They also find single mothers to be overrepresented among housing allowance recipients. The take-up of housing allowance hasn't been studied much, but Lyytikäinen (2008) estimates take-up to be in the range between 64–78% in 2005.

Social assistance. Social assistance is an income transfer for poor households with the purpose of guaranteeing a minimum living standard when all other means fall short of that minimum level. It is called a 'last resort' benefit, which means that all other sources of income should be considered before applying for social assistance. It is intended for a short-term need, and therefore the benefit is typically granted for one month at a time, and has to be reapplied each month. The benefit is granted by the municipality's social work office.

The benefit formula for basic social assistance[5] (SA) can be described as:

$$\begin{aligned} SA =& family\, needs - family\, income \\ =& \sum (family\, weights \times unit\, need) + housing\, costs + other\, costs \\ & - (income - disregard) \end{aligned} \quad (2)$$

'Unit need' refers to a measure of living standards defined in the social assistance legislation. It is defined in terms of the monetary amount that a single adult living alone is considered to need to cover his or her basic living costs such as food, clothing, hygiene, minor medical needs, utility bills, and so on. The total needs of the family are then based on the family composition as a weighted sum according to specific weights set out in the regulation. For example, a family with a husband and wife and three children aged 15, 12 and 8 years, are considered to need 3.83 times the unit need to cover their normal living costs.[6]

In 2011, the unit need was 419.11 eur/month, so the family of the example above would be considered to need $3.83 \times 419.11 = 1,605.19$ euros to cover their monthly basic needs. Housing costs and other costs can additionally be taken into account in benefit calculation. The housing costs of the family are typically accepted in full, but there can be some discretionary consideration regarding what is a 'reasonable' level of housing in the region, although no strict guidelines exist such as those for housing allowance. Other costs such as day care costs or larger medical bills can be covered on a discretionary basis.

If the family's disposable income falls short of their needs, they should be granted the difference as social assistance. All income is taken into account in the eligibility consideration, except for a small disregard. Each family is allowed to keep 20% of their monthly net labour earnings up to 150 euros (100 euros before 2005) without it reducing their social assistance benefits. Because the benefit is intended for short-term need, the calculation is based on income in the current month. For example, suppose that the net income of the example family is 2,000 euros per month (suppose part of the income comes from labour earnings), and their monthly rent is 1000 euros. They would then be eligible for $(1,605.19 + 1,000) - (2,000 - 150) = 755.19$ euros

[5] I focus on basic social assistance in this study. I ignore the supplementary benefit called preventive social assistance, which is granted in special circumstances and for which eligibility cannot easily be determined from the data.

[6] The first adult is given weight 1, second and further adults a weight of 0.85. A first child aged 10–17 has weight 0.7 and a first child aged 0–9 has weight 0.63, and second children in the categories have 0.05 smaller weights. Adult children living with their parents have a weight of 0.73. The example family would thus have a total weight of $1 + 0.85 + 0.7 + 0.65 + 0.63 = 3.83$.

per month as social assistance.

In addition to the principles underlying formula (2), eligibility is restricted by further regulations. Since all other income sources are overriding relative to social assistance, the adults of the family are expected to work or actively seek for work, and also to apply for all other social benefits they might be eligible for (including housing allowance) before applying for social assistance. Besides earnings and other social benefits, the families should also first use up any liquid assets they might have before they can be eligible for the benefit. This refers especially to savings in bank accounts and similar assets.[7]

Whereas housing allowance is granted for the entire household, social assistance is targeted at the nuclear family: if there are many adults living in the same dwelling, only one married or cohabiting couple is considered to belong to the same family (together with their under-aged their children, if any), and any other adults form families of their own (possibly together with their spouses and children) and should apply for the benefit separately. This applies even to adult children who live with their parents.

Furthermore, students, conscripts and those in non-military service are typically not eligible for social assistance because there are other benefit schemes that provide a basic living standard for these groups. Pensioners are often not eligible for social assistance because the pensioners' housing allowance is more generous than the regular housing allowance, so their incomes are typically lifted above social assistance standards. They can still be granted social assistance, for example to cover large medical costs. These restrictions have some implications for the eligibility simulation as will be discussed further in the next section.

The receipt of social assistance is typically connected to receipt of other benefits. Hannikainen-Ingman et al. (2012) estimate that in November 2009, 35% of LMS recipients and 21% of UA recipients received social assistance. Around half of LMS and UA recipients who also received housing allowance, received additionally social assistance. Official statistics of the same period report that out of all social assistance recipient households, 69% were housing allowance recipients and 40% were LMS recipients (THL, 2010).[8]

[7]Being the owner of one's home is not considered as having liquid assets in the guidelines of the benefit, so even owner-occupier families can be eligible for the benefit. Mortgage downpayments are not accepted as housing costs in equation (2), but mortgage interest payments are.

[8]Descriptive studies on social assistance often focus on disentangling where the need to claim the benefit arises. The role of housing costs has been discussed a lot (e.g. Heinonen, 2010; Kauppinen et al., 2015; Honkanen, 2010), as well as the link to unemployment benefits (Hannikainen-Ingman et al., 2012; Hiilamo et al., 2005). Many have concluded that the low level of these other social

The take-up of Finnish social assistance has not been studied much even though it is a recurring theme in Finnish media and policy context. The first analyses were made with small survey samples, estimating take-up rates based on self-assessed eligibility and reported use of the benefit (Kuivalainen, 2007). Bargain et al. (2012) use Finnish Income Distribution Survey data to produce more reliable estimates on benefit eligibility and take-up. Their estimates of take-up vary between 56% in 1996 and 49% in 2003, suggesting much lower take-up rates than Lyytikäinen's (2008) estimates for housing allowance. As for characteristics correlated with take-up, they find that a longer ongoing unemployment spell and the receipt of LMS benefits predict higher take-up probability. They also find that household characteristics such as being a single parent, having low education, and not owning one's dwelling also increase take-up, whereas the age of the household head or the number of children did not have a significant impact on take-up.

Kuivalainen and Sallila (2013) study how much social assistance benefits decrease poverty among the benefit recipients. In the section on poverty impact of the benefits, I add to their results by combining the two benefits, taking into account the non-take-up estimates as well as the dynamic perspective.

3 Data and empirical methodologies

3.1 Income Distribution Survey

My main data source is the Income Distribution Survey (IDS) from Statistics Finland, which covers 25,000–30,000 individuals in 10,000–11,000 households yearly (approx. 0.5% of the population) and is weighted-representative of the Finnish population.[9] A key benefit of the IDS data is the availability of actual household composition and housing cost information from the interviews, which are not available at the population level from any registers. Beyond the information originating from the survey interviews, most of the data content of the IDS comes from registers. The data is therefore very accurate and does not suffer from reporting effects regarding benefit receipt, for example.

I use the IDS waves from 2003–2011. The data are at the individual level, but because I focus on benefits that are determined for the household as a

benefits leads to the need to apply also to social assistance on top of them.

[9]This is the same dataset as used in Bargain et al. (2012) for the study of social assistance non-take-up.

whole, I collapse the data to household level.

3.2 Additional variables

Using the individual identifiers contained in the IDS datasets, I can link information to each individual in each IDS wave. This allows me to provide new information on factors that have not been studied in the take-up literature before. At least in Finland, the common discourse often suggests that households eligible for welfare benefits often prefer to resort to help from relatives. Very high stigma or transaction costs could lead individuals to rather survive on their own, even amounting to borrowing from relatives or friends. To study this phenomenon, I link the taxable income of the parents of each individual to measure the household's access to outside resources.[10]

Another theme often referred to in the take-up discourse is the role of ability. Welfare rules and procedures are often accused of being too complex, so that transaction costs rise too high especially for the less able households. A lower ability could be linked to low income and thus benefit eligibility, but also lower propensity to take-up. As proxies for ability, I link the elementary school grade point average as well as the test scores from Finnish Defence Forces' Basic Skills Test. These variables are only available for a subset of the households, so the link between ability and take-up can only be estimated for a subsample.[11]

To gain more insight into the behaviour of the households, I also combine

[10] A parent's income is available if the parent is alive and living in Finland, and has any taxable income that year. I set the value to zero if no income is found. For each adult, I measure the average of their two parents' income. If there are two adults living in the family, I take the average parental income over both individuals.

[11] Elementary school GPA is available for 1991–2011. Finnish elementary school ends after 9th grade, when the pupils typically turn 16. The GPA ranges from 4–10. The Military's Basic Skills Test results are available for 1982–2013. Military service is compulsory for all Finnish males. Typically the service is done at the age of 19–20, and it has to be completed between the ages 17–29. All conscripts take a Basic Skills Test at the beginning of the service. The test consists of three categories: verbal, arithmetic, and logical reasoning. Each subtest and the total are given between 0–40 points, and then standardised to a 9-point scale. This scale follows the normal distribution so that each year the mean of the test results is 5 points, and standard deviation is two points. See Pekkala Kerr, Pekkarinen and Uusitalo (2013) for a more detailed description. Due to the nature of these data, the variables are available to particular subsets of the IDS households. School GPA is available to households with young adults, or older adults with children in the suitable age range. The Basic Skills Test score is available to households with males (very few women partake in military service) in specific age ranges. Note that this excludes for example single women or single mothers with very young children. I assume that the GPA or test score of one individual is descriptive of the average ability of the family. In cases where there are several households members for whom these variables are available, I use the highest GPA or test score in the household.

data from other registers of Statistics Finland to all the IDS individuals three years before I observe them in the IDS data, the year of the IDS, and three years after. This allows me to illustrate the dynamics of household income and activity over a longer time horizon, even though I can estimate eligibility and hence take-up for only one year for each household from the IDS. These variables include annual disposable income (eur/year), gross amount of all social benefits received (eur/year), receipt of certain benefits (including social assistance and housing allowance, eur/year and days (or months)/year), and each person's main labour market activity either during the year or at the end of the year (as defined by Statistics Finland). All these information are at the individual level, and can be aggregated up to the household level.

As discussed in more detail in section 3.3, I pool all the IDS datasets together to increase sample size. I refer to this pooled IDS as year t. Year t is the point in time when I can estimate the eligibility for the benefits based on the information provided in the IDS data and observe whether they were benefit claimers or not. I then find the income and labour market information for each individual/household in the pooled data set for three preceding years ($t-1$ to $t-3$) and three subsequent years ($t+1$ to $t+3$).[12]

One important drawback to this approach is that household composition can change over the course of these seven years. Household members may move out, or new children may be born into the household, and so on. Measures such as "household income" in a year $s \neq t$ using the household composition in year t IDS is therefore not always an accurate description of the household's actual situation. Unfortunately the accurate composition can only be found from the IDS survey information for year t. However, measuring the income of the household members this way does capture the income dynamics for those household members that actually form the household in year t, when I analyse their take-up decision: it captures both the previous experiences of these members, as well as their future development (which they might be expecting in year t), and may therefore affect their take-up decision in year t, whether or not they actually form a household together in those years. (See section 3.5 and Appendix B for sensitivity analyses.)

[12]Because the last year observed in these data is 2012, I can follow IDS 2010 individuals only two years, and IDS 2011 only one year after their IDS participation. This means there are slightly fewer observations for pooled years $t+2$ and $t+3$.

3.3 Simulation of benefit eligibility and sample formation

From each yearly IDS dataset, I can estimate the households' eligibility for social assistance and housing allowance by simulating a potential benefit to the household based on their characteristics and the eligibility rules.[13]

As described in section 2, social assistance benefits are mainly not intended for students, pensioners or conscripts, and students and pensioners also have their own housing subsidy schemes. I therefore do not simulate benefits to any households containing such individuals. In addition, since social assistance benefits are determined for the nuclear family, I exclude households where several adults (other than spouses) live together, including adult children living with their parents. After this trimming, I am left with households consisting of nuclear families with working-age adults: one- and two-parent families with only under-aged children, as well as couples and singles without children. There are between 4,000–5,000 such households in each IDS dataset.[14]

For housing allowance, the simulation consists of determining the composition of the household, the municipality of residence, housing costs, and various housing details provided in the IDS survey (square meters, type of residence, year of construction, etc.) and household income, as defined in the eligibility rules. The simulation of social assistance is based on a similar set of variables.

The IDS dataset provides information at the annual level, whereas social assistance eligibility is determined monthly, and housing allowance eligibility spells can start at any time during the calendar year. Thus the simulated eligibility is based on average monthly income calculated from total yearly income, which is an imperfect estimate of actual monthly income if incomes fluctuate during the year. Note however that the estimation error this causes means that some truly eligible households will not be captured by the simulation, but all those simulated eligible should be truly eligible: if income is constant throughout the year and below the eligibility threshold, the household is eligible each month. If income is higher in some months and lower in some months but on average below the thresholds, then the household is eligible in the months when income is lower. On the other hand, households eligible for a couple of months during the year, or for very small amounts, can

[13] For both benefits, I take advantage of the SISU code by Statistics Finland which includes the accurate coding of Finnish laws and regulations in SAS language.

[14] I discuss the impact of this restriction in Appendix B.1.

be left out of the eligibility simulation. My take-up estimates will therefore be an upper bound as the smallest benefits are not considered, for which take-up propensity is likely to be lower. Having annual data also means that the number of eligible months cannot be estimated, but only the total amount of benefits the household is eligible for during the year. This describes the annual average shortfall of the family's income from the regulative thresholds, and indicates the relative need of the household.

The comparison of household income and costs gives the estimated benefit size for each household. I then restrict the eligible samples further to achieve a sample where eligibility is more reliably estimated. I also study the sensitivity of the results to these simulation choices; see section 3.5 and Appendix B.

Firstly, I require the household to be eligible for at least 500 euros per year to be considered as eligible, in order to rule out the smallest estimated benefits, which can be more prone to estimation error due to yearly data instead of monthly.

Second, I further restrict eligibility based on the type of income. I take into account the notion that assets affect the level of housing allowance, and are likely to restrict social assistance eligibility. Unfortunately, the data are not detailed enough to accurately calculate the impact of assets on housing allowance benefits according to the legislation, and do not contain information on savings. Instead, I use information on taxable capital income to form a categorical measure for whether the household had any assets. Even though capital income is an imperfect measure for assets, it is an income source that is likely to affect benefit eligibility, consisting of e.g. dividend and interest income from financial instruments, income from selling such instruments, as well as rental income. With this crude measure, I categorically restrict eligibility from households with non-negligible capital income (over 50 euros per year).

Third, I account for the last-resort nature of social assistance. In principle, the household should have applied for all other benefits before applying for social assistance. This includes housing allowance, and since both benefits are targeted to the poor, households eligible for social assistance typically would be eligible for housing benefits as well. I therefore restrict the social assistance eligibility to households already receiving housing allowance. Even though there are situations when a household can be eligible for social assistance when they are not eligible for housing allowance[15], social assistance eligibility

[15]For example, a very short-term income shock could make the household eligible for social

is more reliably estimated for households who already are eligible for and receive housing allowance. The downside of this choice is that it restricts the social assistance eligibles to a specific group: those who have already decided to take up housing allowance.

In sum, I define a household as eligible for housing allowance or social assistance in a given year if their simulated benefit is at least 500 euros per year, and they had at most very little taxable capital income that year. For social assistance, I also require the household to be a recipient of housing allowance. These definitions form my main eligible samples for housing allowance and social assistance. There are around around 100–300 households in each yearly IDS dataset that fulfill these requirements.[16] If the household is additionally observed to have received the benefit that year, they are defined as having claimed the benefit ('take-up'), otherwise they are categorised as non-claimers ('non-take-up').

To increase sample size for meaningful estimation, I pool all the IDS datasets together. This results in 2,261 households eligible for housing allowance and 1,221 households eligible for social assistance in the main sample. Table 1 provides summary statistics of the main eligible samples, contrasting them to the entire IDS sample (which is representative of Finland averaged over 2003–2011), and in the Results section I illustrate how these characteristics differ by take-up status (Tables 2–3).

3.4 Methodological considerations for analysing survey data

The IDS dataset is a yearly survey sample involving sampling weights, stratification, and a rotating panel structure.[17] Using these survey elements, the datasets can be analysed as representative of the Finnish population at the yearly level. Nationally representative estimates of population characteristics

assistance for a short period but not for housing allowance, which is based on longer-term need.

[16] In Appendix B.2, I illustrate the simulation accuracy for both benefits.

[17] The IDS survey is a rotating panel, where each household is asked to participate in two consecutive years (four since 2009). All members of each household are included in the sample, and the data is at the individual level. The survey design is stratified. There are 13 strata based on a combination of the type and level of household income for each rotational group, so there are 26 strata in a given year (52 since 2009). The yearly IDS survey sample covers around 0.5% of the Finnish population, and even though some groups are over-sampled (e.g. the low-income and the very high-income), the sampling fraction is well below 5% in each stratum. With such small sampling fraction, the finite population correction can be ignored (Cochran, 1977). The individual level data is clustered at the household level since the whole household is always included. The dataset provides sampling weights for both the individual and household level that represent the sampling design as well as non-response and attrition.

Table 1: Descriptive statistics: eligible sample characteristics compared to overall IDS mean.

	overall mean in IDS (N=95,177)	main sample eligible for HA (N=2,216)	main sample eligible for SA (N=1,221)
Couple, no children	0.30	0.03	0.04
Couple, with children	0.23	0.11	0.14
Single adult, no children	0.39	0.70	0.70
Single adult, with children	0.05	0.16	0.12
# of HH members	2.1	1.7	1.7
# of underaged children living at home	0.4	0.5	0.5
Age of HH head	50.7	40.5	40.9
Dwelling size, m^2	87.6	54.5	55.0
Dwelling size, m^2 per equivalence unit	59.5	44.5	44.7
HH highest education level: basic	0.22	0.35	0.37
HH highest education level: medium	0.41	0.51	0.49
HH highest education level: high	0.37	0.14	0.14
HH disposable income, equivalence scale	€ 23,462	€ 10,607	€ 10,767
Average income of own parents	€ 11,731	€ 12,044	€ 11,200
Average income of own and spouse's parents	€ 9,100	€ 6,759	€ 6,482
HH has labour earnings	0.69	0.42	0.35
HH has entrepreneurial income	0.18	0.05	0.03
HH member unemployed	0.22	0.77	0.80
HH member unemployed: receives LMS	0.08	0.61	0.68
HH member unemployed: receives UA	0.02	0.10	0.08
HH member unemployed: receives ER	0.13	0.13	0.11
Average days HH members unemployed	19.1	158.4	173.5
HH member receives sick leave benefits	0.06	0.13	0.14
Average days HH members on sick leave	3.9	9.8	9.5
HH member receives parental leave benefits	0.09	0.13	0.14
Average days HH members on parental leave	8.2	20.1	20.4
Metropolitan area	0.20	0.22	0.20
Major university town	0.15	0.20	0.19
Other large town	0.09	0.13	0.14
Other, town-like region	0.23	0.22	0.23
Other, densely populated region	0.15	0.11	0.12
Other, rural regions	0.17	0.12	0.12
HH receives HA	0.08	0.87	1.00
HH receives SA	0.08	0.68	0.81
Actual HA receipt, eur/year (where > 0)	€ 2,288	€ 2,889	€ 2,819
Actual SA receipt, eur/year (where > 0)	€ 2,368	€ 3,199	€ 3,820
Simulated HA, eur/year	–	€ 2,313	–
Simulated SA, eur/year	–	–	€ 3,023

Note: Pooled household level observations from IDS datasets 2003–2011. Weighted means using IDS survey weights. Household (HH) head is defined as the person with highest earnings. Equivalence units refer to the modified OECD equivalence scale, where the first adult has weight 1, further members above 14 years old have weight 0.5 each, and further members below the age of 14 each get a weight of 0.3. Own parental income is averaged over two potential parents. Overall parental income is averaged over four potential parents: own and spouse's parents. LMS: labour market support. UA: unemployment assistance. ER: earnings-related unemployment benefit. Number of days of given benefits are averaged over two adults for couples.

such as mean income or the poverty rate can be produced using the sampling weights. Weight and stratum information are also needed in standard error calculation, which is typically carried out using Taylor series linearization (Heeringa, West and Berglund, 2010).

In economic analysis, the case for weighting survey data is not always clear-cut. As Solon, Haider and Wooldridge (2015) discuss, it depends on the type of inference to be made from the analysis whether weighted estimates should be produced or not. They recommend that descriptive statistics as well as descriptive regressions (as opposed to e.g. causal estimates of treatments) should be weighted to make them representative of the population of interest and avoid biases caused by the survey design such as under- or over-sampling of certain groups. Since my analysis is in this sense descriptive in nature, I present weighted estimates. Weighting corrects for the over- or underrepresentation of some groups in the sampling design, as well as differential non-response and attrition rates. A motivation for not weighting would be that I am focusing on a particular subsample, the poorest households, and in principle unweighted estimates could be considered as representative of them. The unweighted sensitivity analysis shows that the main conclusions do not differ (see Appendix B.4).

The main analysis is thus weighted using IDS survey weights. Note that while some parts of the analysis are weighted to make the estimates representative of the entire population (such as poverty rates), in many cases the estimates are weighted to be representative of the relevant sub-population (the eligibles). In some parts I present yearly estimates, but to save space and to benefit from the larger sample size of the pooled data, I mostly present averages over the period 2003–2011. In places where I present yearly weighted estimates, I employ the stratum information in standard error calculation. However, when presenting pooled averages over the study period, I do not account for stratification. The strata are designed for each yearly survey sample to make the observations within strata more similar with each other, and more different across strata, and this idea does not naturally extend to the pooled setting. Since employing stratification typically gives more accurate estimates, leaving it out gives more conservative standard error estimates.

Note that while the IDS sampling weights are designed to make yearly estimates representative of the population of interest in that specific year (year t), in the graphical dynamic analysis I follow the IDS individuals from $t-3$ to $t+3$. The weights in a given IDS do not translate directly to other years, so they cannot be used to make estimates representative of the calendar years represented by $t - 3$ to $t + 3$. I therefore present unweighted graphs in the graphical analysis. Weighting with year t IDS weights produces essentially identical results, but it has to be kept in mind that this makes the relative year $s \neq t$ estimates representative of the population in year t, not s.

A final issue arising from the survey nature of the data is that pooling the yearly surveys together causes the household-level observations to be clustered by household. Because of the rotating nature of the IDS survey, households can participate in more than one survey wave, which leads to some households appearing several times in the pooled data. Note that some households that participate in several waves might be estimated eligible in some waves and ineligible in some, reducing the number of times they appear in the pooled data of the eligible sample. The pooled setting therefore involves within-household correlation, which I account for by clustering standard errors at the household level.

3.5 Sensitivity analyses

To sum up the various methodological considerations brought up in the previous sections, I list here the main features of the analysis set out in the previous sections, and briefly explain how I study the sensitivity of the results to the various choices. All the sensitivity checks are presented in Appendix B.

I only consider households that are relevant from the perspective of the two benefits: households with no adult children or other adults besides cohabiting partners, and households with no students, pensioners or conscripts. For this "relevant" sub-sample, I simulate households' eligibility for housing allowance and social assistance using the information in the yearly IDS datasets. The comparison of income and costs gives the simulated benefit for all households, and I further restrict the set of households to arrive at my main eligible samples:

(i) the household has to be eligible for over 500 euros of benefits per year (both HA and SA samples)

(ii) the household cannot have more than 50 euros/year of taxable capital income (both HA and SA samples)

(iii) to be included in the social assistance eligible sample, the household has to be a recipient of housing allowance.

I run sensitivity checks on all these choices: (i) I use 1000e and 1500e as the eligibility thresholds; (ii) I do not restrict the sample based on capital income at all; and (iii) I do not require social assistance eligibles to be recipients of housing allowance. In part (iii), I still include simulated housing allowances (for those not receiving any) as income when simulating social assistance. This is a good description of actual eligibility determination practice.

I also perform the following sensitivity checks on the entire analysis:

(iv) To control for household clustering, I cluster the standard errors by household in the main analysis. As a sensitivity check, I restrict the sample to the first appearance of each household in the sample, so that no clustering takes place.

(v) Since the family composition is not necessarily constant over time, but I form a measure of "household income" over the follow-up period $t-3$ to $t+3$, I also study the results with a sub-sample where no household changes are observed in any of the years. As mentioned, I only observe at a very crude level whether there were any changes in the composition of the *residents* of a given address. This unfortunately captures also changes that are not necessarily relevant for the family composition (e.g. a subletting person moving out and a new person in) and does not allow identifying new-born children. In addition to using this measure as a control in the regression analysis, I employ it to form a sub-group for whom there are no changes in the resident and thus family composition.

(vi) I run both weighted and unweighted analyses to study the impact survey weighting has on the results.

All the sensitivity checks are provided in Appendix B. The main finding is that the conclusions are not sensitive to the sample definitions and methodologies used.

4 Results

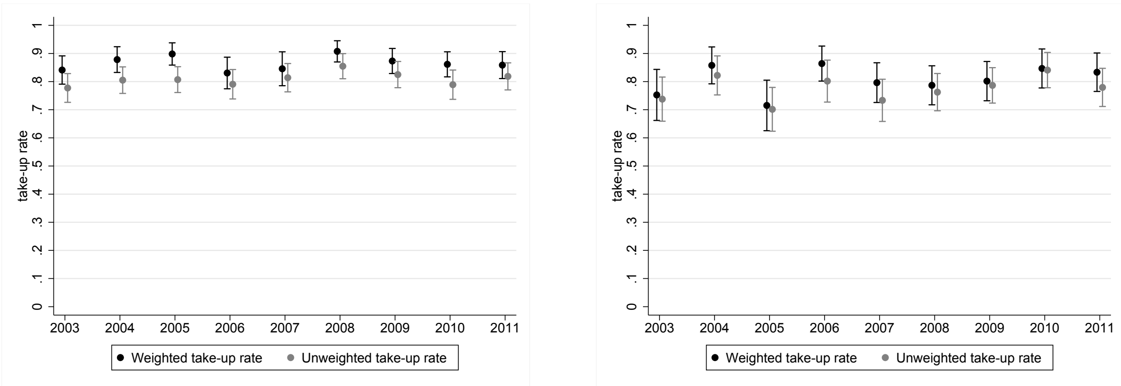
In this section, I first present graphical descriptive results, then turn to descriptive multivariate regressions, and finally turn to the analysis of the benefits' impact on relative poverty in the society.

4.1 Graphical analysis and descriptive statistics on benefit take-up

4.1.1 Average take-up rates

Figure 1: Average yearly take-up rates in 2003–2011.

(a) Housing allowance (b) Social assistance



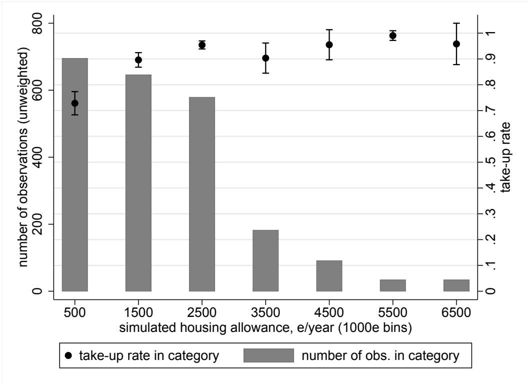
Note: Main eligible samples, weighted with IDS survey weights and unweighted estimates. Standard errors corrected for household clustering.

I find take-up of both benefits to be quite high, varying between 70–90% over 2003–2011 with no clear time trend (Figure 1). These take-up estimates are considerably higher than previously estimated in Finland, especially for social assistance. The main reason for this is likely in the selection of the eligible sample. As I restrict the sample to exclude households with more complex life situations such as adult children living with their parents, there should be less error in the eligibility estimation, which increases the estimated take-up rate. Restricting the eligible sample to households already receiving housing allowance has the same effect: including households with no housing allowance in the eligible sample reduces the yearly average take-up rates of social assistance by 10–15 percentage points. The capital income restriction has a much smaller effect on the take-up estimates. (see Figure B.3) The estimates are thus likely more accurate for this sub-sample, but the downside is that I cannot estimate take-up rates for other groups of interest, such as young adults living with their parents. It is likely that households with more complex circumstances are more often non-claimers of social assistance, even though the discretionary rules could allow them to receive it.

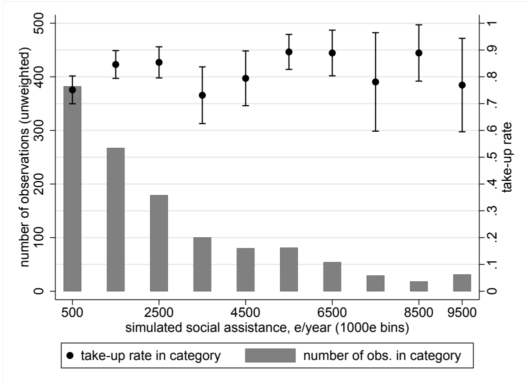
Most benefit recipients receive quite small yearly amounts of social assistance. Standard economic theory would suggest individuals weigh the claiming

Figure 2: Take-up rate as a function of simulated benefit size (in 1000e bins).

(a) Housing allowance



(b) Social assistance



Note: Pooled observations from IDS datasets 2003–2011, main eligible samples. Standard errors corrected for household clustering.

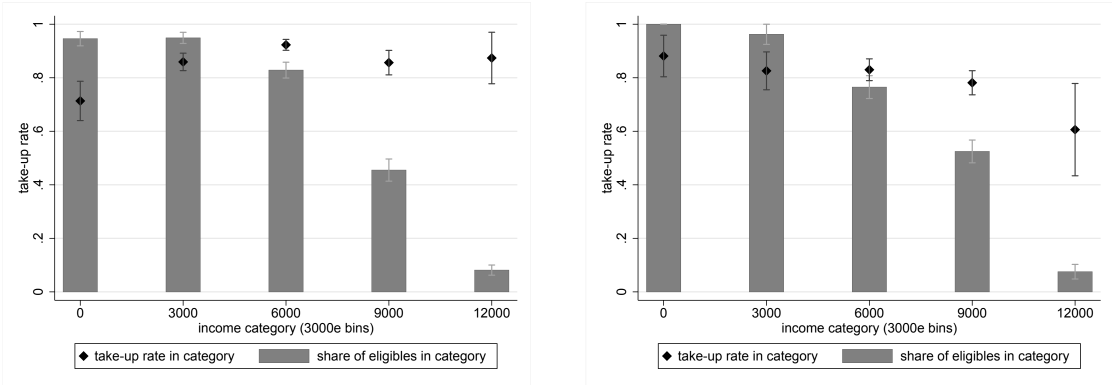
costs against the utility from claiming the benefit, and so one would expect take-up to be increasing in benefit size, and decreasing in income. Figure 2 illustrates that for housing allowance, an increasing relationship with benefit size clearly emerges, but for social assistance take-up the relationship is less clear. On the contrary, there is no clear pattern of housing allowance take-up rate across income ranges, whereas for social assistance a decreasing relationship with income does emerge; see Figure 3.[18] The eligible group falls almost entirely into the poorest 5% of the population, and eligibility is clearly concentrated in the lowest income bins. It is surprising that even among the poorest households, there is no full take-up. It could signify that these households have some other coping strategies available that are not observed in the data but that they prefer to employ rather than tackle the costly take-up process.

Tables 2 and 3 provide descriptive statistics of the eligible samples divided into benefit claimers and non-claimers. For many variables, the comparison between the take-up and non-take-up groups suggests that the non-claimers are slightly better off than the claimers: they have more often earnings, are less often and shorter spells unemployed or on sick leave, have more often a member with higher education, and so on. However, in the housing allow-

[18]The income term in the Figure is *household equivalent income* excluding housing allowance and/or social assistance. In all income comparisons, I use household equivalent income to make the income levels of families of different sizes comparable. It is calculated as the household total income divided by the sum of consumption units in the household. I use the modified OECD equivalence scale, where the first adult has weight 1, further members above 14 years old have weight 0.5 each, and further members below the age of 14 each get a weight of 0.3.

Figure 3: Eligibility and take-up rates as a function of household equivalent pre-benefit income (in 3000e bins).

(a) Housing allowance (b) Social assistance



Note: Pooled observations from IDS datasets 2003–2011, main eligible samples. Standard errors corrected for household clustering. Income categories are in terms of pre-benefit income: (disposable income − HA − SA) for HA take-up, (disposable income − SA) for SA take-up. Equivalent income is obtained by dividing this income measure by household consumption units. Eligibility shares are calculated from the relevant population (relevant household types, no capital income, and for SA, receiving HA).

ance sample, the non-claiming households have lower pre-benefit income than the claiming households. For social assistance it is the opposite. The next subsection studies the income patterns more closely.

4.1.2 Income patterns

In Figure 4a, I divide the group eligible for housing allowance in year t into households that take up the benefit that year (t) and those who do not, and plot their income development over the seven-year horizon from $t-3$ to $t+3$. An important difference is observable. Whereas those households who take up the benefit in t have a fairly constant pre-benefit income level over the time period, those not claiming the benefit are on average experiencing a short-term fall in income in year t. Non-claiming households' pre-benefit income was higher than that of the claiming households in the previous years, falls below the income level of the claiming households in year t (as Table 2 illustrated), and rises back up afterwards.

Table 4 illustrates the prevalence of large income changes just around the year of eligibility and take-up decision. 51% of housing allowance non-claimers are experiencing a considerable drop in household income (net of both benefits) from the previous year, whereas only 23% of the benefit claimers

Table 2: Descriptive statistics of the main sample eligible for housing allowance.

	overall mean (N=2,216)	HA non-take-up (N=432)	HA take-up (N=1,829)	
Couple, no children	0.03	0.06	0.03	**
Couple, with children	0.11	0.10	0.11	
Single adult, no children	0.70	0.78	0.68	***
Single adult, with children	0.16	0.05	0.18	***
# of HH members	1.7	1.5	1.7	***
# of underaged children	0.5	0.3	0.6	***
Age of HH head	40.5	40.9	40.4	
Dwelling size, m^2	54.5	65.3	52.9	***
Dwelling size, m^2 per equivalence unit	44.5	56.2	42.7	***
HH highest education level: basic	0.35	0.36	0.35	
HH highest education level: medium	0.51	0.44	0.52	*
HH highest education level: high	0.14	0.19	0.13	**
HH disposable income, equivalence scale	€ 10,607	€ 6,252	€ 11,281	***
HH disposable income excl. SA & HA, equiv.	€ 6,752	€ 5,811	€ 6,897	***
Simulated amount of HA, e/year	€ 2,313	€ 1,612	€ 2,421	***
HH receives SA	0.68	0.18	0.75	***
Average income of own parents	12,044	14,233	11,705	**
Average income of own and spouse's parents	6,759	8,102	6,552	**
HH has labour earnings	0.42	0.50	0.41	**
HH has entrepreneurial income	0.05	0.19	0.03	***
HH member unemployed	0.77	0.41	0.82	***
HH member unemployed: receives LMS	0.61	0.21	0.67	***
HH member unemployed: receives UA	0.10	0.09	0.10	
HH member unemployed: receives ER	0.13	0.12	0.13	
Average days HH members unemployed	158.4	60.8	173.3	***
HH member receives sick leave benefits	0.13	0.08	0.13	**
Average days HH members on sick leave	9.8	6.1	10.3	
HH member receives parental leave benefits	0.13	0.07	0.14	***
Average days HH members on parental leave	20.1	6.8	22.1	***
Metropolitan area	0.22	0.23	0.22	
Major university town	0.20	0.12	0.21	***
Other large town	0.13	0.10	0.13	
Other, town-like region	0.22	0.25	0.22	
Other, densely populated region	0.11	0.10	0.11	
Other, rural regions	0.12	0.20	0.10	***

Note: Pooled observations from IDS datasets 2003–2011. Weighted means using IDS survey weights. Household (HH) head is defined as the person with highest earnings. Equivalence units refer to the modified OECD equivalence scale, where the first adult has weight 1, further members above 14 years old have weight 0.5 each, and further members below the age of 14 each get a weight of 0.3. Own parental income is averaged over two potential parents. Overall parental income is averaged over four potential parents: own and spouse's parents. LMS: labour market support. UA: unemployment assistance. ER: earnings-related unemployment benefit. Number of days of given benefits are averaged over two adults for couples. The final column displays the significance of a t-test for difference between take-up and non-take-up group means (*** p<0.01, ** p<0.05, * p<0.1). Standard errors corrected for household clustering.

Table 3: Descriptive statistics of the main sample eligible for social assistance.

	overall mean (N=1,221)	SA non-take-up (N=277)	SA take-up (N=944)	
Couple, no children	0.04	0.03	0.04	
Couple, with children	0.14	0.18	0.13	*
Single adult, no children	0.70	0.63	0.72	**
Single adult, with children	0.12	0.16	0.11	*
# of HH members	1.7	1.9	1.7	**
# of underaged children	0.5	0.7	0.5	**
Age of HH head	40.9	39.0	41.3	**
Dwelling size, m^2	55.0	57.6	54.4	
Dwelling size, m^2 per equivalence unit	44.7	43.9	44.9	
HH highest education level: basic	0.37	0.28	0.40	***
HH highest education level: medium	0.49	0.49	0.48	
HH highest education level: high	0.14	0.23	0.12	***
HH disposable income, equivalence scale	€ 10,767	€ 8,783	€ 11,246	***
HH disposable income excl. SA & HA, equiv.	€ 5,877	€ 6,780	€ 5,659	***
Simulated amount of SA, e/year	€ 3,023	€ 2,731	€ 3,093	*
Average income of own parents	€ 11,200	€ 12,498	€ 10,886	
Average income of own and spouse's parents	€ 6,482	€ 7,385	€ 6,263	
HH has labour earnings	0.35	0.51	0.32	***
HH has entrepreneurial income	0.03	0.09	0.02	***
HH member unemployed	0.80	0.74	0.81	*
HH member unemployed, receives LMS	0.68	0.53	0.71	***
HH member unemployed: receives UA	0.08	0.12	0.08	
HH member unemployed: receives ER	0.11	0.14	0.10	
Average days HH members unemployed	173.5	129.4	184.1	***
HH member receives sick leave benefits	0.14	0.09	0.15	**
Average days HH members on sick leave	9.5	4.1	10.9	***
HH member receives parental leave benefits	0.14	0.17	0.14	
Average days HH members on parental leave	20.4	25.9	19.1	
Metropolitan area	0.20	0.25	0.19	*
Major university town	0.19	0.24	0.18	*
Other large town	0.14	0.11	0.15	
Other, town-like region	0.23	0.22	0.24	
Other, densely populated region	0.12	0.08	0.12	
Other, rural regions	0.12	0.09	0.13	

Note: Pooled observations from IDS datasets 2003–2011. Weighted means using IDS survey weights. Household (HH) head is defined as the person with highest earnings. Equivalence units refer to the modified OECD equivalence scale, where the first adult has weight 1, further members above 14 years old have weight 0.5 each, and further members below the age of 14 each get a weight of 0.3. Own parental income is averaged over two potential parents. Overall parental income is averaged over four potential parents: own and spouse's parents. LMS: labour market support. UA: unemployment assistance. ER: earnings-related unemployment benefit. Number of days of given benefits are averaged over two adults for couples. The final column displays the significance of a t-test for difference between take-up and non-takeup group means for the variable in question (*** p<0.01, ** p<0.05, * p<0.1). Standard errors corrected for household clustering.

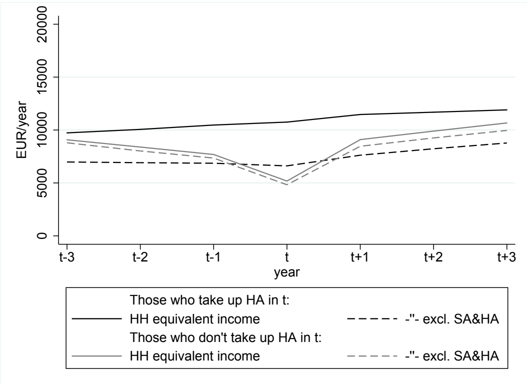
experience such a large drop that year. For 57% of the non-claimers and only 35% of the claimers, pre-benefit income rises considerably from year t to the following year $t + 1$. These patterns imply that housing allowances are not applied for a short-term need. The Table also displays big differences in the variability of income between benefit claimers and non-claimers, measured by the coefficient of variation of pre-benefit income.

Figure 4b illustrates the same patterns for social assistance eligibles. Here, however, the definition of the eligible sample matters for the conclusions to be made. The main eligible sample restricts the eligibles to those who already receive housing allowance benefits. But as Figure 4a illustrated, this group is characterised by steady and low pre-benefit income over time. Dividing such a sample to social assistance claimers and non-claimers does not give us much information on potential differences in the patterns. Table 4 confirms that for the main sample, there are no significant income changes around year t for this sample, nor do the shares differ between the take-up and non-take-up groups. But there is some variabiliy hidden behind the averages that Figure 4b does not convey: the non-claiming households experience significantly more variation in their pre-benefit income than claiming households.[19] Furthermore, loosening the restriction of being a housing allowance recipient (Figure B.5) suggests that similar income dynamics might be in place among social assistance eligibles as there are for housing allowance eligibles: the households are experiencing a short-term fall in income when they are observed to be eligible and non-claiming, and they recover their earlier income level shortly after. This could further signify that households with less clear circumstances, in terms of benefit eligibility rules (such as those not eligible for housing allowance), could be deterred from applying for social assistance for a short-term need. It is also possible that households eligible for both benefits simply view the take-up costs too high to apply for either of them when they expect their need to be of shorter duration.

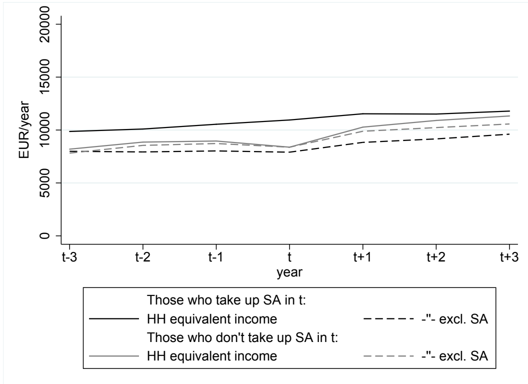
[19]Note that with annual data, I am only able to study income fluctuations at the annual level, not monthly variation within the year. Some families could have permanently low average annual income, but experience fluctuations within the year due to unemployment spells or job seasonality. My data does not allow identifying these kinds of fluctuations.

Figure 4: Income development over years $t-3$ to $t+3$ for year t eligible claimers and non-claimers.

(a) Housing allowance eligibles



(b) Social assistance eligibles



Note: Main eligible samples, pooled observations. Unweighted means. Eligibility and take-up categories, and household (HH) composition, observed in year t and kept constant throughout the graph.

Table 4: Income variability for housing allowance and social assistance eligibles.

	eligible for HA			eligible for SA		
	non-take-up	take-up		non-take-up	take-up	
Income decreases by >20% from $t-1$ to t	0.51	0.23	***	0.21	0.16	
Income decreases by >50% from $t-1$ to t	0.30	0.07	***	0.04	0.04	
Income increases by >20% from t to $t+1$	0.57	0.35	***	0.31	0.30	
Income increases by >50% from t to $t+1$	0.46	0.20	***	0.14	0.15	
CV of income over $t-3$ to t	56 %	36 %	***	40 %	28 %	***
CV of income over $t-3$ to $t+3$	74 %	48 %	***	40 %	34 %	**

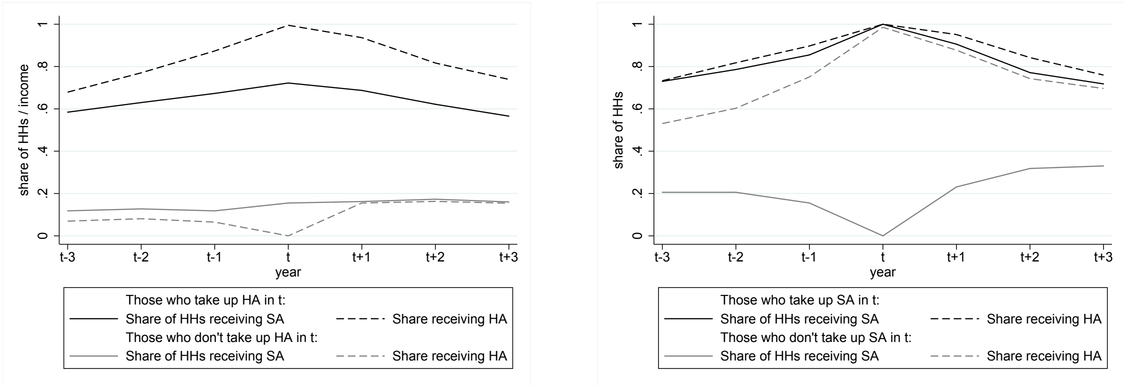
Note: Pooled observations from main eligible samples, 2003–2011. Weighted means using IDS survey weights. Eligibility and take-up categories, and household (HH) composition, observed in year t and kept constant throughout $t-3$–$t+3$. Pre-benefit income is (disposable income − HA − SA) for HA take-up, (disposable income − SA) for SA take-up. Coefficient of variation (CV) of income is calculated as the ratio of the standard deviation to the mean and indicated as a percentage. The final column displays the significance of a t-test for difference between take-up and non-takeup group means (*** p<0.01, ** p<0.05, * p<0.1). Standard errors corrected for household clustering.

4.1.3 Persistence of benefit use

As Figure 4 already suggested, those who claim their housing allowance or social assistance benefits in year t also receive some benefit income in the preceding and following years (illustrated by the difference between the solid and dashed lines). Figure 5 confirms the persistence of benefit use. For both housing allowance and social assistance eligibles, year t claimers typically also received the benefit in previous years, and continue receiving in the following

Figure 5: Receipt of social assistance and housing allowance among eligible claimers and non-claimers.

(a) Housing allowance eligibles (b) Social assistance eligibles



Note: Main eligible samples, pooled observations. Unweighted means. Eligibility and take-up categories, and household (HH) composition, observed in year t and kept constant throughout the graph.

years. They also typically receive these benefits for the most part of the year, more than 6 months on average (not displayed in the graph). Around 60% of housing allowance users top up the benefit with social assistance, and vice versa, most social assistance claimers receive housing allowance over many years. (Social assistance eligibles are by construction restricted to receive housing allowance only in year t.) It is also noteworthy that around 20% of both eligible non-claimers eventually claim the benefits in subsequent years.

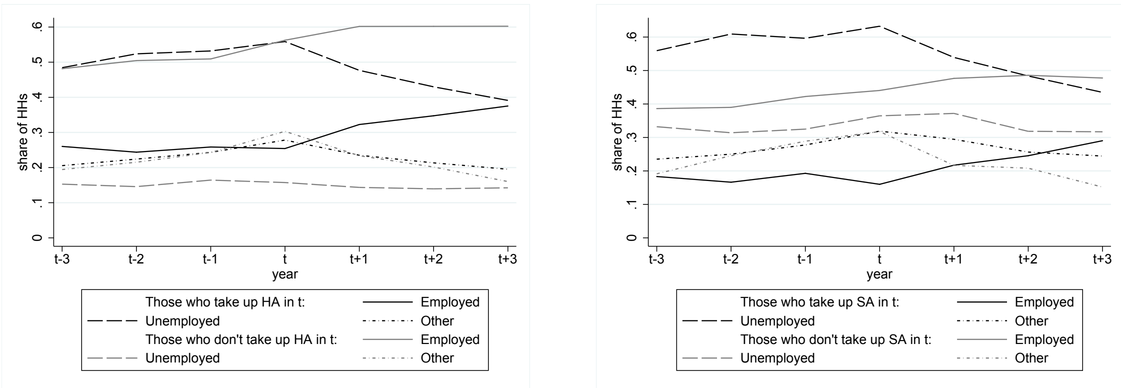
4.1.4 Labour market status

Variability in the labour market status of the family could be related to the variability in income observed above. Figure 6 again divides the households eligible for housing allowance into year-t benefit claimers and non-claimers, and displays the share of households with adult members in a particular labour market status in each group.[20] Here the group "other" refers to other reasons for being outside the labour market than unemployment, such as child care. What clearly emerges is the considerable share of employed adults among the housing allowance non-claimers (gray solid line). Unemployment is instead dominating in the claiming group (black dashed line). The situation is similar among social assistance eligibles, but there is less employment and

[20] These graphs use the indicator for the longest labour market status during the year, but results are identical using the main labour market status at the end of the year (indicators defined by Statistcs Finland).

Figure 6: Household labour market status development over time, eligible claimers vs. eligible non-claimers.

(a) Housing allowance eligibles (b) Social assistance eligibles



Note: Main eligible samples, pooled observations. Unweighted means. Eligibility and take-up categories, and household (HH) composition, observed in year t and kept constant throughout the graph. Figures indicate the share of households with an adult member in the given labour market status.

more unemployment among both claimers and non-claimers than for in the housing allowance sample. The figure does not reveal any clear causes for the rapid income changes oberved in Figure 4.

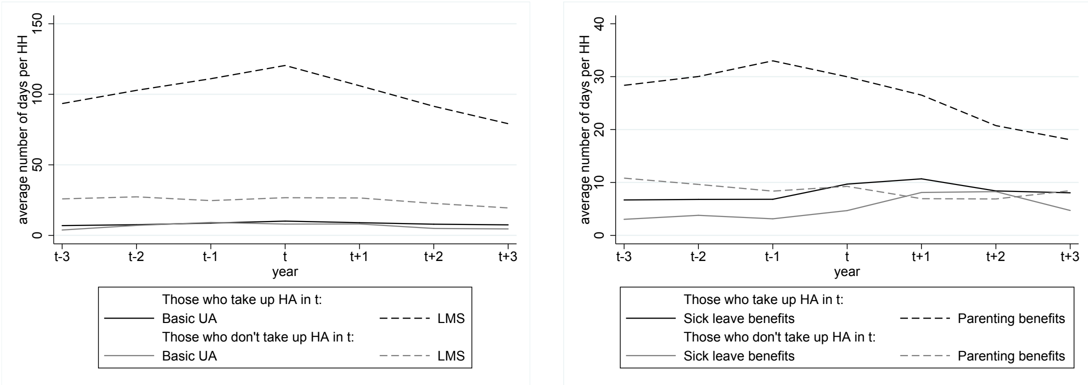
As my dataset includes the number of days per year the households receive certain benefits, I can deepen the picture regarding the labour market status. Figure 7 illustrates that long spells of labour market support (LMS) benefits are very typical among housing allowance and social assistance claimers. The graph also suggests that the group “other” in Figure 6 is composed more of family members on parental leave rather than sick leave. Long sick leaves are not very common, although this is an imperfect proxy for the general health and wellbeing of the target population.

4.1.5 Access to outside resources and ability measures

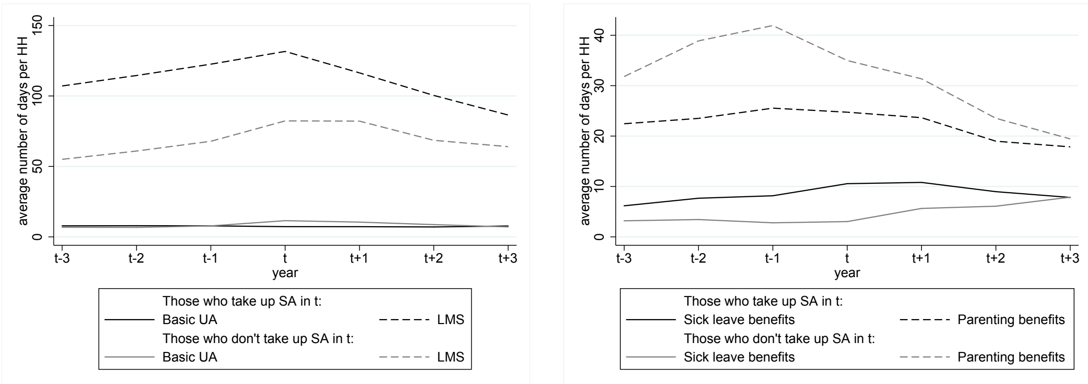
Table 5 displays the sample means for the ability measures for those households for whom each measure is available. There is a clear difference in the means by take-up: the non-claiming households have on average higher elementary school grades and ability test scores. Especially for the military’s Basic Ability Test scores, the differences are relatively large, as differences in the total score and mathematical score between non-claiming and claiming families are close to one point. Since the results are standardised, this can be interpreted as a half a standard deviation (see footnote 11).

Figure 7: Household labour market status development over time, based on the number of days household members received specific benefits. Eligible claimers vs. eligible non-claimers.

(a) Housing allowance eligibles



(b) Social assistance eligibles



Note: Main eligible samples, pooled observations. Unweighted means. Eligibility and take-up categories, and household (HH) composition, observed in year t and kept constant throughout the graph. Figures indicate the average number of days adult household members received benefits.

4.2 Regression analysis

4.2.1 Baseline regressions

To study the correlation of various household characteristics with benefit take-up while controlling for their confounding effects, I perform a standard linear probability OLS regression for benefit take-up. I use the IDS weights and and correct the standard errors for household clustering. Columns (1)–(3) in Table 6 display the results for housing allowance take-up, and Columns (4)–(6) for social assistance take-up. Appendix Tables A.1–A.2 display the bivariate correlations of each variable with benefit take-up.

Some factors have a clear correlation with both housing allowance and social assistance take-up, but differences also do emerge. Most of the variables

Table 5: Average values of ability measures in the eligible samples.

	mean in HA sample (N=982)	HA non-take-up (N=170)	HA take-up (N=812)	
Elementary school GPA	7.3	7.4	7.3	
	mean in HA sample (N=498)	HA non-take-up (N=107)	HA take-up (N=391)	
Basic Ability Test: total	4.4	5.0	4.3	**
Basic Ability Test: logical	4.6	5.1	4.4	**
Basic Ability Test: verbal	4.4	4.8	4.2	*
Basic Ability Test: mathematical	4.3	5.1	4.1	***
	mean in SA sample (N=516)	SA non-take-up (N=130)	SA take-up (N=386)	
Elementary school GPA	7.3	7.7	7.2	***
	mean in SA sample (N=275)	SA non-take-up (N=58)	SA take-up (N=217)	
Basic Ability Test: total	4.3	5.0	4.1	**
Basic Ability Test: logical	4.4	5.6	4.2	***
Basic Ability Test: verbal	4.3	4.7	4.2	
Basic Ability Test: mathematical	4.1	4.9	4.0	**

Note: Pooled observations from main eligible samples, 2003–2011. Weighted means using IDS survey weights. The final column displays the significance of a t-test for difference between take-up and non-takeup group means (*** p<0.01, ** p<0.05, * p<0.1). Standard errors corrected for household clustering. School GPA is graded 4-10. Basic Ability Test total and subsection scores are a standardised score between 0–9 that follows the normal distribution with mean 5 and standard deviation 2.

indicating household types (age, single/couple, children/no children) are not significant for either of the benefits. Education level (measured from highest degree in the family) has a stronger negative correlation with take-up for social assistance than for housing allowance. Having any labour or entrepreneurial earnings during the year suggests lower take-up probabilities, whereas receiving unemployment benefits during the year increases take-up. The different unemployment benefit types have different correlations with social assistance and housing allowance, as only the receipt of labour market support correlates with social assistance take-up but all three types are relevant in the housing allowance equation. The receipt of other benefits – indicating absence from the labour market – has varying significance on the take-up of the benefits.

Households that live more spaciously[21] are less likely to be housing allowance claimants. This is curious since the square meter limits of housing allowance rules do not preclude households living in more spacious dwellings from being eligible, but means that the benefit will cover a smaller portion

[21]The measure divides the surface area of the dwelling by the equivalence units of the family, using the same equivalence scales as for equivalent income, see footnote 18. The housing allowance rules stipulate maximum acceptable square meter limits based on the number of individuals living in the dwelling, without account for whether they are adults or children ($37m^2$ for singles, $57m^2$ for two individuals, etc.). The limits correspond roughly to 30–39 m^2 per equivalence unit.

of their total housing costs, since costs due to the 'extra' square meters are discarded in benefit calculation. Living spaciously does not correlate with social assistance take-up, which could indicate the more discretionary rules of the benefit compared to housing allowance.

As expected, and indicated by Figure 2, take-up probability is increasing in the predicted size of the benefit, though only very moderately for social assistance. The unintuitive pattern in Figure 3 is visible in Table 6, too, as current income has a positive relationship with the take-up of housing allowance. The coefficient is negative for social assistance in most of the specifications.

I also incorporate indicators describing the income and labour market patterns into the regression analysis. To measure the experienced variability in income illustrated in Figure 4 and Table 4 above, I use the coefficient of variation (CV) of household pre-benefit income to capture the variability in income. I also consider immediate (pre-benefit) income changes: the change in income in year t from the previous year, as well as the change in income from year t to the following year. These capture more directly potential expectations about the household's income development in the immediate future. To characterise the variability of labour market status over time – a categorical variable – I first construct a dummy for each year indicating whether the individual had the same main labour market status this year as he or she had in the previous year. I then calculate the average share of years that the adults of the household remain in the same main labour market status in two consecutive years, as the mean of these dummies over the adults of the household. The higher the share, the more stable the labour market status of the family is. I also include the crude proxy for variability in the family composition as described in section 3.5.

In Table 6, I control for variability both before the eligibility year (years $t-3$ to t; columns (2) and (5)) and the entire seven-year period (years $t-3$ to $t+3$; columns (3) and (6)).[22] The first measures the experienced variability in the past, which is likely to affect future expectations. The latter set of variables measures both the experiences as well as the potential expectations about the future, assuming the families make accurate predictions about future circumstances.

The regression shows that the variability in household circumstances over

[22] I exclude the 2011 IDS wave from this latter analysis as these individuals can only be followed until 2012 ($t+1$).

Table 6: Regression results for housing allowance (HA) and social assistance (SA) take-up using the main eligible samples.

	(1) HA	(2) HA	(3) HA	(4) SA	(5) SA	(6) SA
Age of HH head (x 10 years)	0.096	0.023	0.011	-0.004	-0.062	-0.072
	(0.059)	(0.058)	(0.061)	(0.087)	(0.089)	(0.095)
Age squared	-0.011	-0.003	-0.002	0.001	0.008	0.009
	(0.007)	(0.007)	(0.007)	(0.011)	(0.011)	(0.012)
Couple	-0.141	-0.118	-0.079	0.044	0.039	0.038
	(0.055)***	(0.056)**	(0.053)	(0.054)	(0.053)	(0.062)
Children	0.010	0.030	0.057	-0.043	-0.044	-0.029
	(0.025)	(0.026)	(0.027)**	(0.049)	(0.048)	(0.051)
Couple x children	-0.008	-0.024	-0.065	-0.090	-0.079	-0.106
	(0.061)	(0.062)	(0.059)	(0.075)	(0.073)	(0.083)
Education level, medium	0.002	-0.020	-0.031	-0.063	-0.071	-0.072
	(0.018)	(0.019)	(0.019)	(0.029)**	(0.029)**	(0.032)**
Education level, high	-0.038	-0.042	-0.039	-0.153	-0.144	-0.161
	(0.028)	(0.026)	(0.026)	(0.044)***	(0.044)***	(0.047)***
Labour market support recip.	0.189	0.184	0.181	0.172	0.153	0.159
	(0.024)***	(0.026)***	(0.026)***	(0.037)***	(0.036)***	(0.038)***
Basic UA recipient	0.085	0.062	0.052	0.024	0.010	0.035
	(0.036)**	(0.036)*	(0.038)	(0.055)	(0.053)	(0.056)
Earnings-related UI recipient	0.054	0.034	0.031	0.029	0.016	0.029
	(0.025)**	(0.025)	(0.026)	(0.050)	(0.051)	(0.054)
Has labour income	-0.037	-0.029	-0.032	-0.070	-0.072	-0.091
	(0.018)**	(0.017)*	(0.018)*	(0.031)**	(0.031)**	(0.034)***
Has entrepreneurial income	-0.259	-0.240	-0.271	-0.313	-0.309	-0.308
	(0.051)***	(0.051)***	(0.054)***	(0.088)***	(0.090)***	(0.098)***
Sickness benefit recipient	0.019	0.021	0.007	0.058	0.062	0.066
	(0.022)	(0.022)	(0.024)	(0.034)*	(0.033)*	(0.037)*
Parental leave benefit recipient	0.076	0.075	0.056	0.078	0.072	0.119
	(0.024)***	(0.025)***	(0.025)**	(0.047)*	(0.046)	(0.048)**
m^2 per equivalence unit	-0.003	-0.003	-0.003	0.000	0.000	0.000
	(0.001)***	(0.001)***	(0.001)***	(0.001)	(0.001)	(0.001)
HH pre-benefit equiv. income	0.017	0.019	0.018	-0.007	-0.017	-0.020
(x 1000 eur)	(0.005)***	(0.006)***	(0.006)***	(0.009)	(0.009)*	(0.010)*
Simulated benefit (x 1000 eur)	0.066	0.061	0.059	0.020	0.017	0.016
	(0.010)***	(0.011)***	(0.011)***	(0.008)**	(0.008)**	(0.008)*
Pre-benefit income change		0.000	0.000		-0.000	-0.000
between $t-1$, t (x 10%)		(0.000)	(0.000)		(0.000)	(0.000)
Pre-benefit income change		-0.000	-0.000		-0.003	-0.007
between t, $t+1$ (x 10%)		(0.000)	(0.000)***		(0.002)**	(0.003)**
Family composition changes		-0.054			0.040	
over $t-3$ to t		(0.019)***			(0.031)	
CV of pre-benefit income		-0.003			-0.014	
over $t-3$ to t (x 10%)		(0.003)			(0.005)***	
Constancy of labour mkt status		-0.007			-0.006	
over $t-3$ to t (x 10%)		(0.003)**			(0.005)	
Family composition changes			-0.080			0.023
over $t-3$ to $t+3$			(0.021)***			(0.035)
CV of pre-benefit income			0.000			-0.010
over $t-3$ to $t+3$ (x 10%)			(0.004)			(0.007)
Constancy of labour mkt status			-0.011			-0.007
over $t-3$ to $t+3$ (x 10%)			(0.004)***			(0.007)
Constant	0.395	0.632	0.708	0.676	0.968	1.002
	(0.123)***	(0.135)***	(0.144)***	(0.176)***	(0.195)***	(0.231)***
N (unweighted)	2,261	2,059	1,844	1,221	1,189	1,049
R^2	0.28	0.28	0.31	0.13	0.15	0.16

Note: Pooled observations from IDS datasets 2003–2011, main eligible samples. Weighted means using IDS survey weights. Standard errors correcting for household clustering in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression includes year and regional dummies. Pre-benefit income is (disposable income − HA − SA) for HA take-up, (disposable income − SA) for SA take-up. Coefficient of variation (CV) of income is calculated as the ratio of the standard deviation to the mean. In columns (3) and (6) year 2011 IDS sample is excluded.

Table 7: Regression results for housing allowance (HA) and social assistance (SA) take-up for households eligible for both benefits.

	(1) HA	(4) HA	(2) SA	(5) SA	(3) HA&SA	(6) HA&SA
Age of HH head (x 10 years)	0.084	0.020	0.030	0.013	0.079	0.016
	(0.072)	(0.072)	(0.077)	(0.085)	(0.090)	(0.097)
Age squared	-0.009	-0.002	-0.004	-0.001	-0.008	-0.000
	(0.009)	(0.009)	(0.009)	(0.010)	(0.011)	(0.012)
Couple	-0.110	-0.088	0.018	0.039	-0.069	-0.039
	(0.066)*	(0.069)	(0.057)	(0.058)	(0.072)	(0.073)
Children	-0.026	0.008	-0.055	-0.044	-0.038	-0.009
	(0.037)	(0.039)	(0.044)	(0.045)	(0.049)	(0.050)
Couple x children	-0.001	-0.022	-0.042	-0.059	-0.037	-0.062
	(0.072)	(0.076)	(0.072)	(0.073)	(0.084)	(0.085)
Education level, medium	-0.006	-0.028	-0.045	-0.061	-0.052	-0.083
	(0.023)	(0.023)	(0.028)	(0.030)**	(0.031)*	(0.032)**
Education level, high	-0.049	-0.053	-0.155	-0.145	-0.170	-0.166
	(0.034)	(0.032)*	(0.038)***	(0.041)***	(0.042)***	(0.044)***
Labour market support recip.	0.202	0.193	0.131	0.121	0.277	0.258
	(0.030)***	(0.033)***	(0.034)***	(0.038)***	(0.036)***	(0.040)***
Basic UA recipient	0.101	0.077	0.016	0.004	0.079	0.050
	(0.056)*	(0.055)	(0.050)	(0.050)	(0.060)	(0.060)
Earnings-related UI recipient	0.099	0.075	0.032	0.030	0.092	0.073
	(0.035)***	(0.035)**	(0.051)	(0.053)	(0.053)*	(0.054)
Has labour income	-0.067	-0.055	-0.048	-0.053	-0.092	-0.086
	(0.023)***	(0.023)**	(0.029)	(0.030)*	(0.032)***	(0.032)***
Has entrepreneurial income	-0.194	-0.183	-0.231	-0.251	-0.293	-0.289
	(0.055)***	(0.055)***	(0.054)***	(0.056)***	(0.054)***	(0.057)***
Sickness benefit recipient	0.021	0.021	0.067	0.082	0.078	0.092
	(0.026)	(0.025)	(0.033)**	(0.032)**	(0.037)**	(0.037)**
Parental leave benefit recipient	0.087	0.079	0.070	0.072	0.124	0.124
	(0.031)***	(0.032)**	(0.042)*	(0.042)*	(0.047)***	(0.047)***
m^2 per equivalence unit	-0.003	-0.002	-0.000	0.000	-0.002	-0.002
	(0.001)***	(0.001)***	(0.001)	(0.001)	(0.001)***	(0.001)**
HH pre-benefit equiv. income	-0.004	-0.002	-0.009	-0.015	-0.011	-0.014
(x 1000 eur)	(0.008)	(0.009)	(0.007)	(0.008)*	(0.009)	(0.010)
Simulated HA (x 1000 eur)	0.099	0.087	0.007	0.005	0.069	0.057
	(0.013)***	(0.013)***	(0.014)	(0.015)	(0.016)***	(0.017)***
Simulated SA (x 1000 eur)	-0.022	-0.022	0.005	0.001	-0.006	-0.008
	(0.007)***	(0.007)***	(0.007)	(0.007)	(0.008)	(0.008)
HA recipient			0.519	0.487		
			(0.042)***	(0.048)***		
Pre-benefit income change		0.000		0.000		0.000
between $t-1$, t (x 10%)		(0.000)		(0.000)*		(0.000)**
Pre-benefit income change		-0.000		-0.000		-0.000
between t,$t+1$ (x 10%)		(0.000)		(0.000)		(0.000)
CV of pre-benefit income		-0.003		-0.008		-0.010
over $t-3$ to t (x 10%)		(0.003)		(0.004)**		(0.004)**
Constancy of labour mkt status		-0.006		-0.002		-0.006
over $t-3$ to t (x 10%)		(0.003)*		(0.005)		(0.005)
Family composition changes		-0.041		0.037		-0.005
over $t-3$ to t		(0.024)*		(0.031)		(0.032)
Constant	0.455	0.690	0.169	0.310	0.314	0.598
	(0.145)***	(0.169)***	(0.157)	(0.194)	(0.183)*	(0.219)***
N (unweighted)	1,513	1,340	1,513	1,340	1,513	1,340
R^2	0.34	0.35	0.34	0.33	0.25	0.26

Note: Pooled observations from IDS datasets 2003–2011. Households eligible for both benefits, SA eligibles not restricted to receiving HA. Weighted regression using IDS survey weights. Standard errors correcting for household clustering in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression includes year and regional dummies. Pre-benefit income is (disposable income − HA − SA) for both benefits. Coefficient of variation (CV) of income is calculated as the ratio of the standard deviation to the mean. In columns (2), (4) and (6) year 2011 IDS sample is excluded.

a longer time period does matter for take-up, but there are differences among housing allowance and social assistance take-up patterns. In the housing allowance regression, changes in the household composition and labour market changes have a significant relationship with take-up, but the income variability measures do not. For social assistance, income variance is more relevant as both the coefficient of variation of past income and the immediate income increase after year t are statistically significant, with negative signs. It also seems that for housing allowance, also the future realisations play a role, whereas for social assistance it is rather the past experiences that have significant correlations with take-up. This is confirmed by an F-test for the joint significance of the dynamic variables.[23]

Overall, these coefficients are quite small. For instance a 10-percentage point increase in the coeffienct of variation of past income (the average difference between claiming and non-claiming households in Table 4) is associated with just a 1.4 point reduction in social assistance take-up probability. For the labour market status variable, the coefficient is of the opposite sign than would be expected: households with more stable labour market situations are more prone to take up. It is possible that the variable captures differences in more stable employment situation of the non-claimers (as suggested by Figure 6).

The evidence thus lends some support to claims such as those made by Blundell et al. (1988) and findings by e.g. Blank and Ruggles (1996). Households with more variability in their family composition and income histories are less likely to take up these welfare benefits.

Table 7 displays similar results for households eligible for both benefits, looking at the take-up of housing allowance and social assistance separately, and the take-up of both benefits jointly. (In these regressions the social assistance eligibles are not restricted to receiving housing allowance.)

4.2.2 Access to outside resources and ability measures

To gain some insight on the importance of outside resources and ability on take-up behaviour, I perform separate regressions using the subset of individuals for whom these measures are available. Due to the smaller sample size, some of the baseline results of Table 6 become statistically insignificant,

[23] As was evident from graphs 4b and B.5, sample definitions for the social assistance eligibles affect the importance of income dynamics. This is evident in the regression results with different sample definitions as well; see Tables B.7–B.8 in the Appendix.

but adding the variable of interest (parental income, test scores) to the regression does not further change the coefficients in an important way. Table 8 displays the coefficients of interest of these separate regressions.

Table 8: Additional regression results for housing allowance and social assistance take-up for a sub-sample.

	(1)	(2)	(3)	(4)
	HA	HA	SA	SA
Average parental income (x 10,000 eur)	-0.002	-0.002	0.001	0.000
	(0.001)	(0.001)	(0.002)	(0.002)
N	2,261	2,059	1,221	1,189
Basic Skills Test: total	-0.018	-0.015	-0.032	-0.034
	(0.010)*	(0.010)	(0.016)**	(0.016)**
N	498	456	275	272
Basic Skills Test: logical	-0.016	-0.012	-0.045	-0.045
	(0.009)*	(0.008)	(0.017)***	(0.017)***
N	496	454	274	271
Elementary school GPA	-0.01	-0.01	-0.084	-0.084
	(0.016)	(0.015)	(0.025)***	(0.025)***
N	982	910	516	501
Regression includes dynamic variables	no	yes	no	yes
Regression specification corresponds to column ... in Table 6	(1)	(2)	(4)	(5)

Note: Pooled observations from IDS datasets 2003–2011, main eligible samples. Weighted regression using IDS survey weights. Standard errors correcting for household clustering in parentheses (*** p<0.01, ** p<0.05, * p<0.1). School GPA is graded 4-10. Basic Ability Test total and subsection scores are a standardised score between 0–9 that follows the normal distribution with mean 5 and standard deviation 2.

Parental income does not have a significant correlation with either of the benefits, suggesting that outside resources are not important for the decision to take up. The ability measures have a strongly significant coefficient for social assistance take-up, but the significance and also magnitudes are weaker in the housing allowance regressions. As the Basic Skills Test scores are standardised, a one-point change in the score can be interpreted as a 0.5 standard deviation change in the underlying test result (see footnote 11). Thus, a one standard deviation increase in the overall test score (the logical reasoning score) is associated with approximately 6 (9) points lower probability to take up social assistance. Elementary school GPA also has a strong correlation, with two points higher GPA on a 4-10 scale being associated with up to 16.8 points lower take-up probability.

Although the analysis can only be performed to a very small and specific subsample of families with males or children in the correct age group, the results are encouraging in the sense that lower ability does not seem to preclude the eligible families from applying for welfare benefits. Rather, the benefits seem to successfully reach also the families with lower capacities. Parental background also does not seem to correlate with benefit receipt.

4.3 Impact of welfare benefits on relative income-poverty

Since one of the main reasons for having social benefits like social assistance and housing allowance is to alleviate income-poverty, it is useful to study how well they achieve this goal. This entails both the question how well the benefits are targeted towards the poorest households, as well as how much poverty alleviation they achieve. Figure 3 already illustrated that eligibility and benefit receipt are quite srongly focused among the poorest households. In this section, I study how much poverty alleviation the benefits achieve and the relationship with take-up patterns. I first discuss static poverty measures and then add the perspective of income dynamics.

4.3.1 Static poverty measures

To measure the prevalence and depth of low incomes, I use the P_α class of poverty measures developed by Foster, Greer and Thorbecke (1984):

$$P_\alpha = \frac{1}{N}\sum_{i=1}^{H}\left(\frac{z-y_i}{z}\right)^\alpha \tag{3}$$

$$= \frac{1}{N}\sum_{i=1}^{H}\left(1-y_i^*\right)^\alpha , \tag{4}$$

where z is the poverty line and y_i is equivalent income of individual i. The measure can be expressed using y_i^*, income normalised by the poverty line. The summation is over H individuals who are below the poverty line (for them $z - y_i > 0$), and the sum is averaged over the entire population, N. Setting $\alpha = 0$, the measure is reduced to the poverty headcount ratio, $P_0 = \frac{H}{N}$. With $\alpha = 1$, the measure denotes the average poverty gap, which represents the amount of equivalent income, as a share of the poverty line, that poor individuals would need to be lifted above the poverty line. I focus mostly on the poverty headcount ratio for its ease of interpretation.[24] To define 'low income', I use the relative poverty line, which is typically set at 60% of median equivalent income.[25] The discussion is thus illustrative of relative

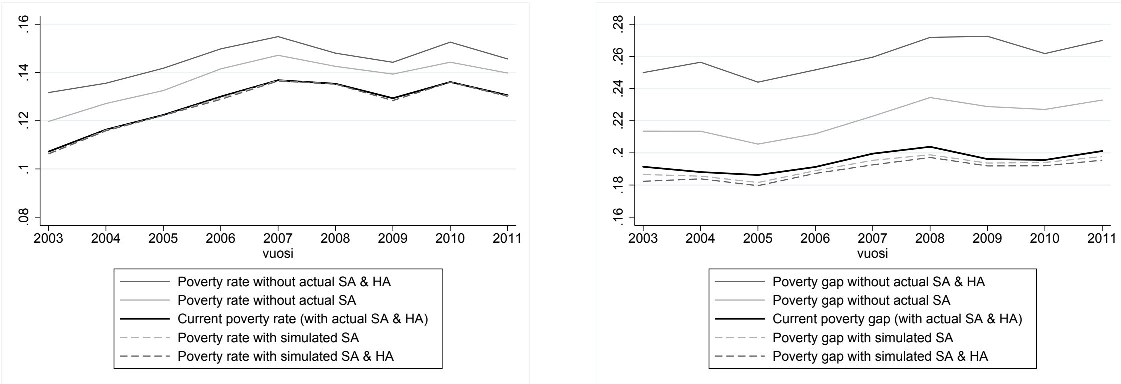
[24] I do not study relative inequality among the poor in this analysis. This aspect can be accounted for by setting $\alpha \geq 2$.

[25] I also analyse the impact of the two benefits on poverty by calculating poverty measures with observed income and excluding any received housing allowance or social assistance income. In reality, adding or removing benefit income from the population would also change median income

underpriviledge in the Finnish context, rather than absolute poverty.

Figure 8: Poverty measures.

(a) Poverty headcount rate in the population (b) Average poverty gap among the poor



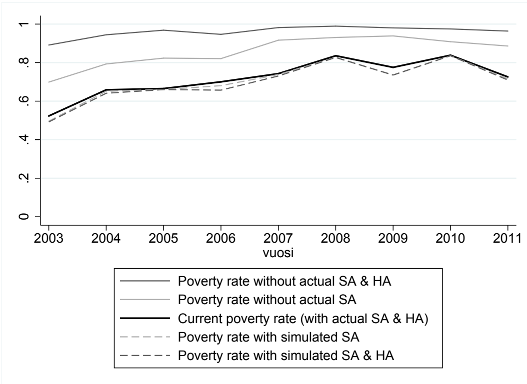
Note: Poverty line defined as 60% of median equivalent income. Poverty headcount ratio is the share of individuals below the poverty line. Poverty gap is average amount of equivalent income each household member would need in order to be lifted above poverty line, expressed as a percentage of the poverty line. The displayed poverty gaps are average gaps among individuals below the poverty line.

The benefits achieve an important amount of poverty reduction in the population. The poverty headcount ratio, based on observed equivalent disposable income (including the amount of housing allowance and social assistance actually received, if any), is on average 13% over the years. In Figure 8a, the solid lines indicate that social assistance benefits reduce poverty by around 1 percentage point, and housing allowance a further 0.8 percentage points, on average. Figure 8b illustrates the average depth of poverty among the poor. Without social assistance and housing allowance, the poverty gap would be around 26% of the poverty line, but is reduced to around 20% with the receipt of the two benefits. The dashed lines in Figure 8 illustrate the poverty measures that would prevail if the estimated non-take-up was eliminated so that all eligible households received the benefit amount they were calculated in the simulation. Since the analysis focuses on such a small sample of households, excluding e.g. pensioners, students and young adults living with their parents, the effect of eliminating non-take-up does not affect the population-wide measures by much.

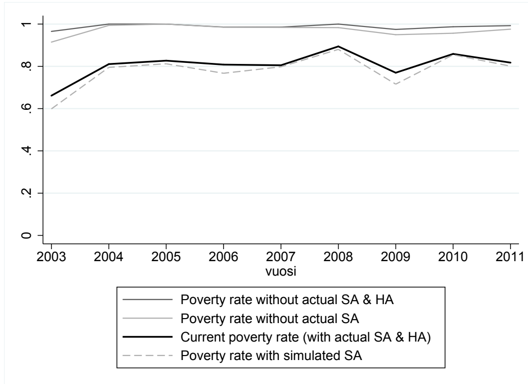
and thus the poverty line. The change in the poverty line is however very small and would not affect the conclusions, so I hold it fixed throughout the analysis. It also makes the analysis more accessible when the point of comparison is held fixed. Incomes of a given year are however always compared to the poverty line of the year in question.

Figure 9: Poverty headcount ratio and poverty gap among the main eligible samples.

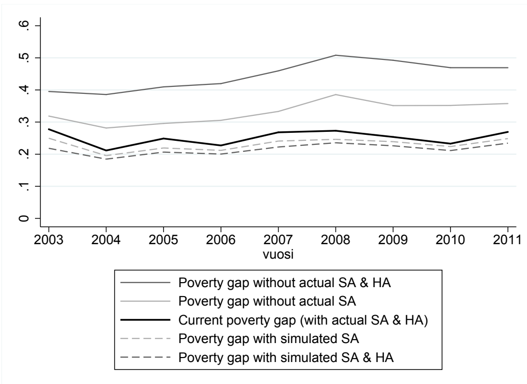
(a) Poverty headcount ratio among housing allowance eligibles



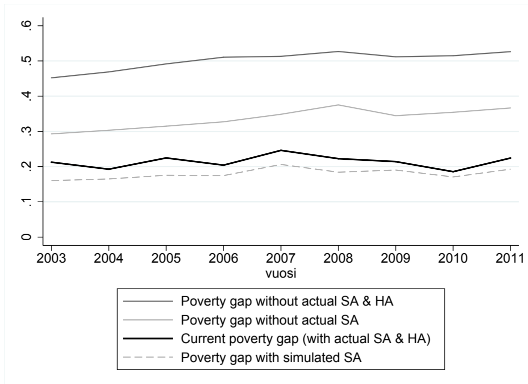
(b) Poverty headcount ratio among social assistance eligibles



(c) Poverty gap among housing allowance eligibles below the poverty line



(d) Poverty gap among social assistance eligibles below the poverty line



Note: Poverty line defined as 60% of median equivalent income. Poverty headcount ratio is the share of individuals below the poverty line. Poverty gap is average amount of equivalent income each household member would need in order to be lifted above poverty line, expressed as a percentage of the poverty line. The displayed poverty gaps are average gaps among individuals below the poverty line.

Figure 9 zooms into poverty within the eligible samples. It is apparent that almost all would be poor without their received benefits, and the received benefits reduce poverty considerably. In graph (a), on average 92% of housing allowance eligibles would be categorised as poor without their housing allowance benefits, and only 72% remain poor after accounting for both housing allowance and social assistance benefits received. Eliminating non-take-up of housing allowance would reduce the poverty rate by a further 0.5 percentage points. On average, the non-claiming households forego 4% of their disposable income in terms of housing allowance benefits. Compared to the benefit claimers, for whom the received housing allowances correspond to 18% of their disposable income, the average foregone amount is not very high.

Graph (b) shows that without housing allowance and social assistance income, practically all social assistance eligibles in the sample would be below the poverty line. Their pre-benefit incomes are also so low that most remain poor even after housing allowance income. On average 19% of social assistance eligibles are lifted above the poverty line by the two benefits. If there was no social assistance non-take-up among this sample, the poverty rate would fall a further 2.5 percentage points on average, to 78%. The non-claiming families forego 7% of their disposable income in unclaimed social assistance benefits (benefit claimers receive 20% of their disposable income in social assistance benefits).

To further study the impact on the depth of poverty, Figure 10 shows where the eligibles are situated relative to the poverty line. The received benefits move the income distribution towards the right. It is also noteworthy that adding simulated benefits to the sample eligibles would further move the distribution up, but there is still a concentration close to the poverty line. The benefits thus do raise the lowest incomes but keep the population still below the poverty line, or at least vulnerable to falling back below it.

4.3.2 Income dynamics and poverty measures

In the section on income dynamics, Figure 4 illustrated an important point about poverty among the samples: there are households whose incomes are low permanently during the seven-year follow-up period, but also those whose incomes fall only in the short term. The poverty analysis literature typically terms these as chronic poverty and transient poverty. In the following, I take advantage of my dynamic data set-up and employ the approach suggested by Jalan and Ravallion (1998) and Ravallion (1998) to analyse overall, chronic and transient poverty. Note that I still consider relative deprivation in Finnish standards, rather than absolute poverty, which is the focus of most of this literature. The distinction is useful in the Finnish context as well, since it helps to disentangle how much of observed cross-sectional poverty is permanent and how much transitory, and policy measures can be designed accordingly. It should be emphasised that this analysis also relies heavily on the assumption of constant household composition (discussed in section 3.5), as it relies on household equivalent income over time. The analysis is therefore only indicative of true poverty dynamics.

As the poverty measure, I use the P_α poverty measures as above, focusing on the headcount ratio with $\alpha = 0$. Equations (5)–(7) display the formulas. I form an overall, intertemporal poverty measure $\bar{P}_\alpha$ by calculating the average share of years over the seven-year follow-up period when the household's equivalent income is below the poverty line. A household is defined as chronically poor if their average income over the seven-year period is below the poverty line. The difference between overall intertemporal poverty $\bar{P}_\alpha$ and chronic poverty P_α^C is defined to be the amount of transient poverty, P_α^T.

$$\bar{P}_\alpha = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{H_t} (1 - y_{it}^*)^\alpha \tag{5}$$

$$P_\alpha^C = \frac{1}{N} \sum_{i=1}^{H^C} (1 - \bar{y}_i^*)^\alpha \tag{6}$$

$$P_\alpha^T = \bar{P}_\alpha - P_\alpha^C, \tag{7}$$

where $\bar{y}_i^* = \frac{1}{T} \sum_{t=1}^{T} y_i^*$ in equation 6 is individual i's average normalised income over the time period, and H^C is the number of households for whom income stays on average below the poverty line during the period.

Table 9: Chronic and transient poverty in the main eligible samples.

	eligible for HA	eligible for SA
Pre-benefit poverty:		
Intertemporal poverty	100 %	100 %
Chronic poverty	94 %	98 %
Transient poverty	6 %	2 %
Post-benefit poverty:		
Intertemporal poverty	96 %	98 %
Chronic poverty	78 %	84 %
Transient poverty	19 %	14 %
Impact of HA & SA on post-benefit poverty:		
From poor to nonpoor	4 %	2 %
From chronically poor to nonpoor	2 %	2 %
From chronically poor to transient poor	15 %	13 %
From transient poor to nonpoor	23 %	18 %

Note: Pooled observations from IDS datasets 2003–2011, main eligible samples. Weighted means using IDS survey weights.

Table 9 illustrates the shares of each type of poverty among main samples eligible for housing allowance or social assistance before and after accounting for benefit income. Most of the sample is defined as chronically poor without their benefit income, but the shares are quite different when including benefit income. 4% of the housing allowance sample exit poverty entirely after both housing allowance and social assistance income. This happens mostly through moving transient poor households out of poverty. Only 2% of the chronically poor exit poverty entirely but 15% swift to transient poverty due to the benefit income. 23% of pre-benefit transient poor exit poverty after the benefits. The shares are very similar for the social assistance eligible sample.

As the analysis in sections 4.1.2 and 4.2 illustrated, income dynamics matter for benefit take-up. Take-up patterns can also be analysed based on these definitions of chronic and transient poor; see Table 10. Among the housing allowance eligibles, 90% of the chronically poor take up the benefit, whereas only 57% of the transient poor claim their benefit. For social assistance eligibles, 79% of chronically poor and 42% of transient poor take up. Eligible non-claimants of a particular year (year t in the context of section 4.1.2) escape poverty the following year (year $t + 1$) more often than the claimants. Practically all housing allowance eligibles are below the poverty line when estimated eligible, but whereas 90% of the claiming households remain below in the following year (82% in three years), only 74% of non-claimants are below the line one year after non-take-up (65% three years after).

Looking at chronic and transient poverty prevalence among the benefit claimers – 25% of all housing allowance recipients and 21% of social assistance recipients are transient poor –, it can be said that both benefits are strongly targeted towards the long-term poor.

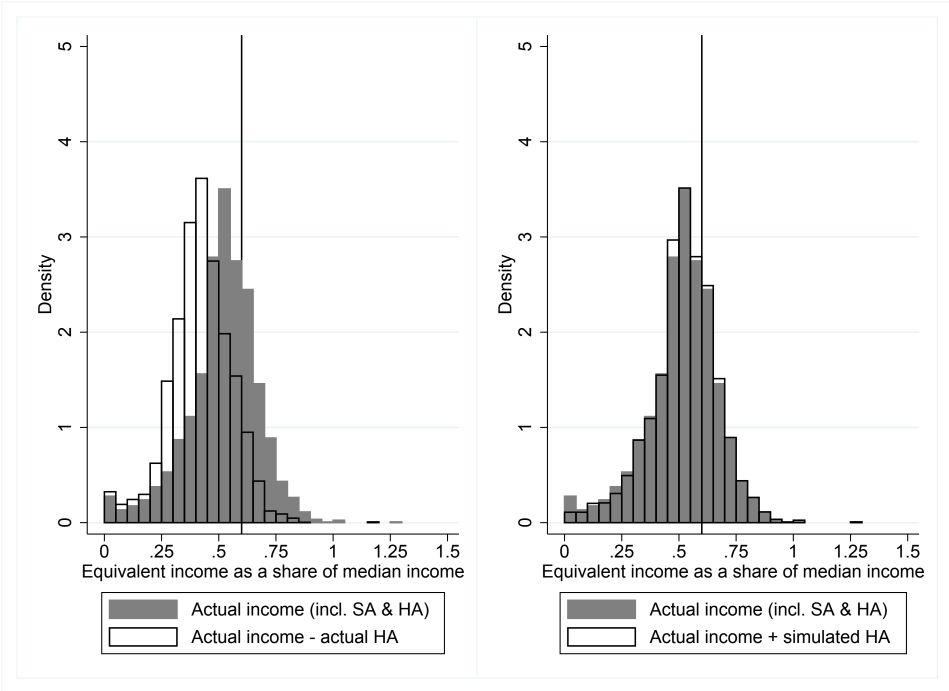
Table 10: Benefit take-up rates among the chronic and transient poor in the main eligible samples.

	eligible for HA	eligible for SA
Overall take-up rate	87 %	81 %
Pre-benefit chronic poor	90 %	79 %
Pre-benefit transient poor	57 %	42 %

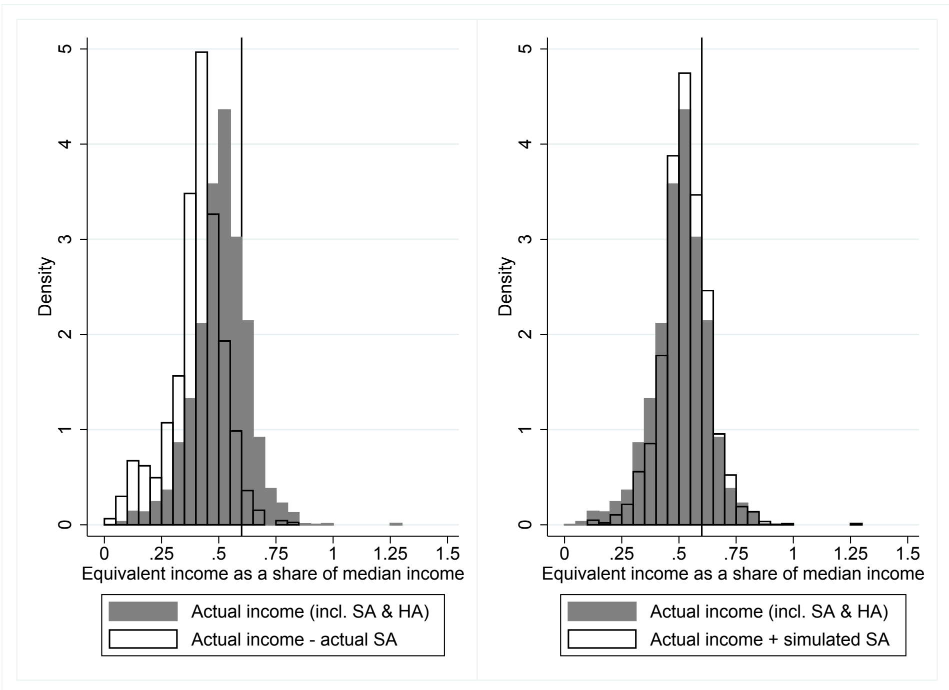
Note: Pooled observations from IDS datasets 2003–2011, main eligible samples. Weighted means using IDS survey weights.

Figure 10: Impact of housing allowance and social assistance on the income distribution among the eligible samples.

(a) Housing allowance



(b) Social assistance



Note: Pooled observations from IDS datasets 2003–2011, main eligible samples. Weighted frequencies using IDS survey weights. The vertical line indicates the location of the poverty line (60% of median equivalent income). Some larger observations removed for data sensitivity reasons.

5 Concluding remarks

The analysis illustrates that when estimating how well a benefit program is concentrated among the target population, both static and dynamic information can be needed. Studying current poverty rates and other characteristics of the benefit claimers and non-claimers is important but provides only part of the story. The dynamic context provides additional information on take-up behaviour and thus on take-up costs and program targeting effectiveness. In the context of Finnish welfare benefits studied here, the static analysis finds that benefit eligibility and take-up are concentrated among the worst-off households. Incorporating the dynamic setting to the analysis reveals that due to non-take-up, benefits do not reach the short-term poor households as often but are concentrated among the long-term poor, who are arguably in larger need of state support. There is less relief of short-term poverty in the society than of long-term poverty.

The benefit and drawback of the current analysis is the strict selection of the eligible sample. This both helps to get more reliable estimates for benefit take-up among the chosen sample, but also leaves out some groups of interest, especially families where young adults live with their parents. It is possible that households with more complex life situations (compared to the benefit rules) apply for the benefits less often, but this analysis cannot shed light on their behaviour. The Appendix discusses this and other impacts of the sample selection on the results.

Bibliography

Anderson, Patricia M. and Bruce D. Meyer, "Unemployment insurance takeup rates and the after-tax value of benefits," *Quarterly Journal of Economics*, 1997, *112* (3), 913–937.

Bargain, Olivier, Herwig Immervoll, and Heikki Viitamäki, "No claim, no pain. Measuring the non-take-up of social assistance using register data," *Journal of Economic Inequality*, 2012, *10* (3), 375–395.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu, "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment," *Quarterly Journal of Economics*, 2012, *127* (3), 1205–1242.

Bhargava, Saurabh and Dayanand Manoli, "Psychological frictions and the incomplete take-up of social benefits: evidence from an IRS experiment," *American Economic Review*, 2015, *105* (11), 3489–3529.

Blank, Rebecca M. and Patricia Ruggles, "When do women use Aid to Families with Dependent Children and Food Stamps? The dynamics of eligibility versus participation," *The Journal of Human Resources*, 1996, *31* (1), 57–89.

Blundell, Richard, Vaness Fry, and Ian Walker, "Modelling the take-up of means-tested benefits: The case of housing benefits in the United Kingdom," *The Economic Journal*, 1988, *98(390)*, 58–74.

Cochran, William G., *Sampling techniques*, John Wiley & Sons, 1977.

Currie, Janet, "The take-up of social benefits," in A. Auerbach, D. Card, and J.M. Quigley, eds., *Public policy and the income distribution*, Russell Sage Foundation, 2006, pp. 80–148.

Foster, James, Joel Greer, and Erik Thorbecke, "A class of decomposable poverty measures," *Econometrica*, 1984, *52* (3), 761–766.

Hannikainen-Ingman, Katri, Heikki Hiilamo, Pertti Honkanen, Susan Kuivalainen, and Pasi Moisio, "Perus- ja vähimmäisturvan yleisyys ja päällekkäisyys 2000-2009," Working paper 33/2012, Social Insurance Institution of Finland (Kela) 2012.

Heeringa, Steven G., Brady T. West, and Patricia A. Berglund, *Applied survey data analysis*, Chapman & Hall/CRC, 2010.

Heinonen, Hanna-Mari, "Asumismenojen huomioon ottaminen toimeentulotuessa ja yleinen toimeentulotukitilanne kunnissa ja kuntayhtymissä syksyllä 2009," Sosiaali- ja terveysturvan selosteita 73/2010, Social Insurance Institution of Finland (Kela) 2010.

Hiilamo, Heikki, Helka Hytti, and Pentti Takala, "Työikäiset toimeentuloturvan vähimmäisetuuksien saajina," Sosiaali- ja terveysturvan selosteita 42/2005, The Social Insurance Institution of Finland (Kela) 2005.

Honkanen, Pertti, "Asumiskustannukset toimeentulotuessa," Sosiaali- ja terveysturvan selosteita 72/2010, Social Insurance Institution of Finland (Kela) 2010.

Jalan, Jyotsna and Martin Ravallion, "Transient poverty in postreform rural China," *Journal of Comparative Economics*, 1998, *26* (2), 338–357.

Kauppinen, Timo M., Katri Hannikainen-Ingman, Seppo Sallila, and Veera Viitanen, "Pienituloisten asuinolot," Working paper 22/2015, National Institute for Health and Welfare (THL) 2015.

Kerr, Sari Pekkala, Tuomas Pekkarinen, and Roope Uusitalo, "School tracking and development of cognitive skills," *Journal of Labor Economics*, 2013, *31* (3), 577–602.

Kleven, Henrik and Wojciech Kopczuk, "Transfer program complexity and the take-up of social benefits," *American Economic Journal: Economic Policy*, 2011, *3* (1), 54–90.

Kuivalainen, Susan, "Toimeentulotuen alikäytön laajuus ja merkitys," *Yhteiskuntapolitiikka*, 2007, *72* (1), 49–56.

_ **and Seppo Sallila**, "Toimeentulotuen saajien köyhyys ja toimeentulotuen köyhyyttä vähentävä vaikutus 1990-2010," in Susan Kuivalainen, ed., *Toimeentulotuki 2010-luvulla. Tutkimus toimeentulotuen asiakkuudesta ja myöntämiskäytännöistä*, THL Reports 9/2013, National Institute for Health and Welfare (THL), 2013, p. Chapter 2.

Lyytikäinen, Teemu, "Studies on the effects of property taxation, rent control and housing allowances," VATT Research reports 140, VATT Institute for Economic Research 2008.

Ravallion, Martin, "Expected poverty under risk-induced welfare variability," *The Economic Journal*, 1998, *98* (393), 1171–1182.

Solon, Gary, Steven J. Haider, and Jeffrey Wooldridge, "What are we weighting for?," *The Journal of Human Resources*, 2015, *50* (2), 301–316.

THL, "Social assistance 2009," Statistical report 31/2010, National Institute for Health and Welfare (THL) 2010.

Zantomio, Francesca, "The route to take-up: Evidence from the UK pension credit reform," *Oxford Bulletin of Economics and Statistics*, 2015, *77* (5), 719–739.

A Appendix: Additional tables

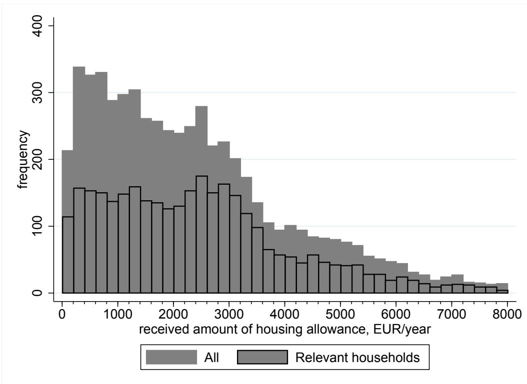
Tables A.1–A.2 display the univariate regression coefficients for the variables included in Table 6.

B Appendix: Sensitivity analysis

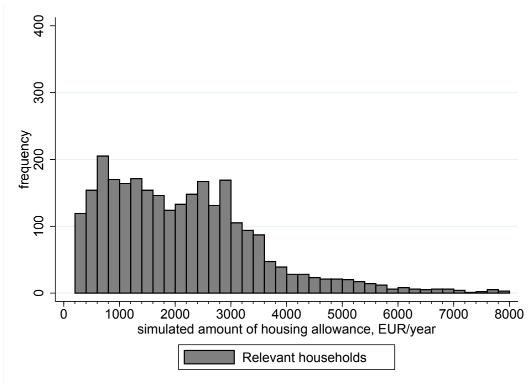
B.1 Distribution of benefits and excluded households

Figure B.1: Benefit size distribution for housing allowance (top) and social assistance (bottom), actual observed in data (left) and simulated (right).

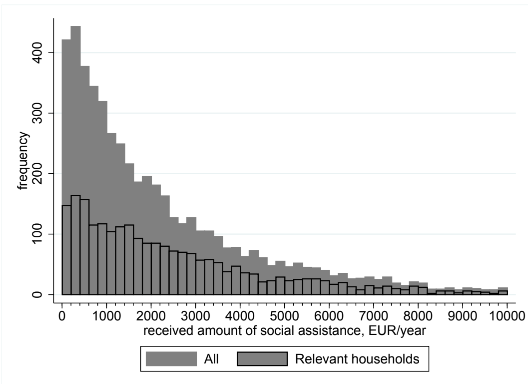
(a) Housing allowance, true



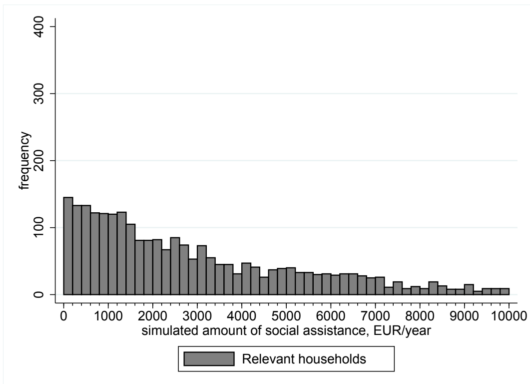
(b) Housing allowance, all simulated



(c) Social assistance, true



(d) Social assistance, all simulated



Note: Pooled observations from IDS datasets 2003–2011, main eligible samples. Unweighted frequencies. Graphs are capped at 8,000 and 10,000 euros for data sensitivity reasons. Graphs (b) and (d) display the distribution of all simulated benefits before the restrictions to the main eligible sample discussed in section 3.5.

Figure B.1 illustrates the distribution of social assistance and housing allowance benefit sizes at the household level. The grey bars in the background in graphs (a) and (c) illustrate the actual distribution of benefits in the IDS

Table A.1: Regression coefficients for each explanatory variable separately for housing allowance take-up.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Age of HH head (x 10 years)	0.200 (0.062)***																		
Age squared	-0.025 (0.008)***																		
Couple		-0.101 (0.059)*																	
Children		0.093 (0.017)***																	
Couple x children		0.029 (0.064)																	
Education level, medium			0.020 (0.021)																
Education level, high			-0.052 (0.032)																
Labour market support recip.				0.244 (0.022)***															
Basic UA recipient				0.092 (0.037)**															
Earnings-related UI recipient				0.061 (0.027)**															
Has labour income					-0.046 (0.019)**														
Has entrepreneurial income						-0.391 (0.053)***													
Sickness benefit recipient							0.058 (0.025)**												
Parental leave benefit recipient								0.075 (0.019)***											
m^2 per equivalence unit									-0.005 (0.001)***										
HH pre-benefit equiv. income (x 1000 eur)										0.013 (0.004)***									
Simulated HA (x 1000 eur)											0.060 (0.008)***								
Pre-benefit income change between $t-1$,t (x 10%)												0.000 (0.000)***							
Pre-benefit income change between t,$t+1$ (x 10%)													-0.000 (0.000)						
Family composition changes over t-3 to t														-0.060 (0.019)***					
CV of pre-benefit income over t-3 to t (x 10%)															-0.012 (0.003)***				
Constancy of labour mkt status over t-3 to t (x 10%)																-0.009 (0.003)***			
Family composition changes over t-3 to t+3																	-0.068 (0.019)***		
CV of pre-benefit income over t-3 to t+3 (x 10%)																		-0.014 (0.003)***	
Constancy of labour mkt status over t-3 to t+3 (x 10%)																			-0.020 (0.004)***
Constant	0.509 (0.119)***	0.852 (0.012)***	0.863 (0.018)***	0.700 (0.022)***	0.885 (0.011)***	0.886 (0.009)***	0.859 (0.010)***	0.856 (0.011)***	1.077 (0.029)***	0.776 (0.028)***	0.728 (0.024)***	0.891 (0.011)***	0.919 (0.011)***	0.929 (0.024)***	0.875 (0.009)***	0.874 (0.009)***	0.900 (0.012)***	0.936 (0.014)***	0.999 (0.025)***
N (unweighted)	2.261	2.261	2.261	2.261	2.261	2.261	2.261	2.261	2.261	2.261	2.261	2.261	2.192	2.215	2.126	2.146	2.261	2.252	2.231
R^2	0.01	0.01	0.00	0.11	0.00	0.06	0.00	0.01	0.06	0.01	0.05	0.01	0.03	0.01	0.00	0.00	0.01	0.04	0.02

Table A.2: Regression coefficients for each explanatory variable separately for social assistance take-up.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Age of HH head (x 10 years)	-0.058																		
	(0.083)																		
Age squared	0.010																		
	(0.010)																		
Couple		0.016																	
		(0.055)																	
Children		-0.078																	
		(0.042)*																	
Couple x children		-0.020																	
		(0.076)																	
Education level, medium			-0.053																
			(0.030)*																
Education level, high			-0.178																
			(0.046)***																
Labour market support recip.				0.117															
				(0.032)***															
Basic UA recipient				-0.045															
				(0.055)															
Earnings-related UI recipient				-0.046															
				(0.047)															
Has labour income					-0.130														
					(0.030)***														
Has entrepreneurial income						-0.394													
						(0.084)***													
Sickness benefit recipient							0.068												
							(0.033)**												
Parental leave benefit recipient								-0.044											
								(0.037)											
m^2 per equivalence unit									0.001										
									(0.001)										
HH pre-benefit equiv. income (x 1000 eur)										-0.017									
										(0.005)***									
Simulated SA (x 1000 eur)											0.010								
											(0.005)*								
Pre-benefit income change between $t-1$,t (x 10%)												-0.000							
												(0.000)							
Pre-benefit income change between t,$t+1$ (x 10%)													-0.003						
													(0.002)						
Family composition changes over t-3 to t														-0.025					
														(0.028)					
CV of pre-benefit income over t-3 to t (x 10%)															-0.013				
															(0.004)***				
Constancy of labour mkt status over t-3 to t (x 10%)																0.001			
																(0.005)			
Family composition changes over t-3 to t+3																	-0.041		
																	(0.028)		
CV of pre-benefit income over t-3 to t+3 (x 10%)																		-0.011	
																		(0.005)**	
Constancy of labour mkt status over t-3 to t+3 (x 10%)																			-0.001
																			(0.006)
Constant	0.861	0.825	0.856	0.735	0.851	0.817	0.796	0.812	0.774	0.946	0.776	0.815	0.848	0.805	0.808	0.814	0.825	0.844	0.811
	(0.156)***	(0.018)***	(0.022)***	(0.029)***	(0.016)***	(0.014)***	(0.016)***	(0.016)***	(0.041)***	(0.041)***	(0.024)***	(0.019)***	(0.018)***	(0.033)***	(0.015)***	(0.015)***	(0.021)***	(0.022)***	(0.039)***
N (unweighted)	1.221	1.221	1.221	1.221	1.221	1.221	1.221	1.221	1.221	1.221	1.221	1.221	1.207	1.212	1.194	1.217	1.221	1.221	1.217
R^2	0.01	0.01	0.02	0.02	0.02	0.03	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.00

datasets. Both benefits are focused towards smaller yearly amounts. The bars with black borderlines in the same graph display the distribution of benefits in the sub-population that is relevant for eligibility estimation: when households with adult children living with their parents are excluded, as well as households with students, pensioners or conscripts. A considerable part of benefit recipients are excluded with this choice. For housing allowance, most of these excluded households were excluded because there was at least one student (note that they may be students for only a part of the year). Families with students are also an important share of the excluded households in the social assistance sample, followed by pensioner households. The rest are explained with households containing conscripts or other adults besides the nuclear family (such as adult children).

Table B.1: Difference between observed benefit recipients and relevant sample for estimation.

	HA	SA
# receiving benefits	5,790	5,166
..of which:		
household with students	2,061	1,660
household with pensioners	454	1,172
household with conscripts	180	168
household with other adults	706	960
# receiving benefits in relevant sample	3,220	2,323

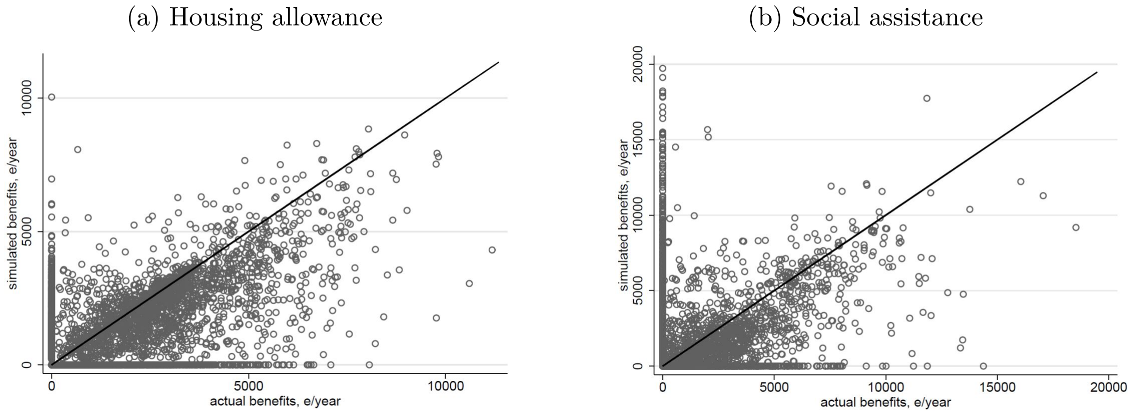
Note: Pooled observations from IDS datasets 2003–2011. Unweighted frequencies. Displayed household categories are overlapping.

B.2 Simulation accuracy

Figure B.2 describes how well simulated benefits compare to actual benefits. There is variation along the 45-degree line, especially for social assistance, but considering the amount of discretion and complexity in the benefit rules, the approximation seems quite good. There are more points below the 45-degree line than above it, which means that benefits are more often underestimated than overestimated. This is natural since households can have some additional costs which are covered by social assistance but are not visible in the data (such as health care costs, or a sudden need to replace a broken household appliance).

For 59% of the observations where the household is recorded to receive some social assistance, the simulated yearly benefit is within ±1000 euros of the observed benefit size. The share is 77% for housing allowance, where there is less discretion in benefit rules. The points along the vertical axis, where actual benefits are zero, show that non-take-up occurs at all benefit levels. Yearly scatter plots (not displayed here) show that simulation accuracy is similar across the years.

Figure B.2: Simulation accuracy for housing allowance and social assistance.



Note: Simulated and true benefit sizes among all relevant households types, 2003–2011 pooled. Unweighted observations. Some larger observations removed for data sensitivity reasons.

B.3 Impact of sample selection

Table B.2 displays differences in eligible sample household characteristics when the eligibility threshold is raised from 500 euros of estimated benefit per year to 1000 and 1500 euros per year. The rest of the tables and graphs focus on the main eligibility threshold of 500 euros per year. Tables B.4 and B.3 illustrate the impact of the other sample restrictions on sample characteristics.

Figure B.3 illustrates the impact of the eligible sample restrictions on the take-up rate estimates. This figure illustrate that sample definition does have an impact on the average yearly take-up rate estimates. The biggest impact is made among social assistance eligibles by the restriction to be receiving housing allowance. But Figures B.4 and B.5 on the other hand illustrate that the different samples still behave in a similar manner: eligible housing allowance claimers experience stable low incomes and non-claimers a temporary fall in income in year t. Here, too, restricting social assistance eligibles to households already receiving housing allowance does affect the picture. The restriction removes in large part households with higher and fluctuating annual incomes (who do not claim housing allowance, as Figure B.4a illustrates). Thus in the main social assistance eligible sample, the year t non-claimers are experiencing much more stable and low income patterns. Removing this restriction (Figures B.5c and e), the non-claimers' income patterns are again visibly fluctuating with a large drop in year t.

One key element that this dynamic analysis relies on is that changes in household composition during the follow-up period $t - 3$ to $t + 3$ does not matter. The right-hand panel of Figures B.4 and B.5 verifies that restricting the eligible sample to households with no residence changes during the seven-year period does not change the overall patterns.

Since results using a sample with no household clustering (restricting households to appear only once in the data) are identical to the main eligible sample, they are not displayed here.

Tables B.5–B.8 show regression results with different sample definitions: table B.5 uses housing allowance eligibles when the sample has not been restricted with the requirement of no capital income; table B.6 uses both the main eligible sample (cf. Table 6 in the main text) and the non-capital-income-restricted sample when both are restricted to households where there are no changes in the composition of the residents. Tables B.7–B.8 does the same modifications for the social assistance eligible sample, additionally including households with no previous housing allowance receipt. The main conclusions do not change from these tables.

B.4 Unweighted results

Table B.9 displays unweighted regression results for benefit take-up using the same specifications as in Table 6. The main conclusions do not change from these tables.

Table B.2: Descriptive statistics of housing allowance and social assistance eligible samples with different eligibility thresholds.

sample eligible for: eligibility threshold (EUR/year):	HA 500	HA 1000	HA 1500	SA 500	SA 1000	SA 1500
Couple, no children	0.03	0.03	0.03	0.04	0.04	0.04
Couple, with children	0.11	0.11	0.12	0.14	0.15	0.17
Single adult, no children	0.70	0.69	0.66	0.70	0.69	0.67
Single adult, with children	0.16	0.17	0.19	0.12	0.12	0.13
# of HH members	1.7	1.7	1.8	1.7	1.8	1.8
# of underaged children	0.5	0.6	0.6	0.5	0.6	0.6
Age of HH head	40.5	40.1	39.8	40.9	40.8	40.6
Dwelling size, m^2	54.5	53.5	53.7	55.0	56.5	58.3
Dwelling size, m^2 per equivalence unit	44.5	43.1	42.6	44.7	45.4	46.1
HH highest education level: basic	0.35	0.36	0.37	0.37	0.39	0.40
HH highest education level: medium	0.51	0.50	0.50	0.49	0.47	0.48
HH highest education level: high	0.14	0.14	0.13	0.14	0.13	0.13
HH disposable income, equivalence scale	€ 10,607	€ 10,625	€ 10,722	€ 10,767	€ 10,746	€ 10,679
HH disposable income excl. SA & HA, equiv.	€ 6,752	€ 6,434	€ 6,125	€ 5,877	€ 5,551	€ 5,183
Average income of own parents	€ 12,044	€ 12,030	€ 12,081	€ 11,200	€ 10,876	€ 10,841
Average income of own and spouse's parents	€ 6,759	€ 6,752	€ 6,830	€ 6,482	€ 6,301	€ 6,391
HH has labour earnings	0.42	0.38	0.35	0.35	0.33	0.31
HH has entrepreneurial income	0.05	0.05	0.04	0.03	0.03	0.04
HH member unemployed	0.77	0.77	0.77	0.80	0.77	0.74
HH member unemployed: receives LMS	0.61	0.63	0.65	0.68	0.66	0.64
HH member unemployed: receives UA	0.10	0.09	0.08	0.08	0.08	0.08
HH member unemployed: receives ER	0.13	0.12	0.10	0.11	0.08	0.07
Average days HH members unemployed	158.4	166.6	170.3	173.5	166.9	147.4
HH member receives sick leave benefits	0.13	0.13	0.13	0.14	0.14	0.14
Average days HH members on sick leave	9.8	10.0	10.2	9.5	9.6	8.5
HH member receives parental leave benefits	0.13	0.14	0.16	0.14	0.15	0.16
Average days HH members on parental leave	20.1	22.0	25.4	20.4	21.4	22.2
Metropolitan area	0.22	0.23	0.24	0.20	0.21	0.24
Major university town	0.20	0.21	0.22	0.19	0.20	0.19
Other large town	0.13	0.12	0.12	0.14	0.14	0.13
Other, town-like region	0.22	0.23	0.23	0.23	0.22	0.22
Other, densely populated region	0.11	0.10	0.10	0.12	0.11	0.11
Other, rural regions	0.12	0.11	0.09	0.12	0.12	0.11
HH receives HA	0.87	0.90	0.93	1.00	1.00	1.00
HH receives SA	0.68	0.71	0.74	0.81	0.82	0.83
Actual HA receipt, eur/year (where > 0)	€ 2,889	€ 3,029	€ 3,221	€ 2,819	€ 2,916	€ 3,003
Actual SA receipt, eur/year (where > 0)	€ 3,199	€ 3,330	€ 3,544	€ 3,820	€ 4,124	€ 4,515
Simulated HA, eur/year	€ 2,313	€ 2,585	€ 2,876	-	-	-
Simulated SA, eur/year	-	-	-	€ 3,023	€ 3,477	€ 4,042

Note: Pooled observations from IDS datasets 2003–2011. Weighted means using IDS survey weights. Eligible households (relevant household types, no capital income, and for SA, receive HA) with different minimum thresholds for simulated benefit size.

Table B.3: Descriptive statistics of housing allowance eligibles with different sample definitions.

	(1)	(2)	(3)	(4)
Couple, no children	0.03	0.02	0.04	0.02
Couple, with children	0.11	0.05	0.12	0.05
Single adult, no children	0.70	0.86	0.69	0.84
Single adult, with children	0.16	0.08	0.16	0.08
# of HH members	1.7	1.3	1.7	1.3
# of underaged children	0.5	0.2	0.5	0.2
Age of HH head	40.5	45.7	40.9	46.1
Dwelling size, m^2	54.5	49.7	56.3	51.8
Dwelling size, m^2 per equivalence unit	44.5	45.2	45.6	46.7
HH highest education level: basic	0.35	0.32	0.35	0.33
HH highest education level: medium	0.51	0.54	0.50	0.53
HH highest education level: high	0.14	0.14	0.15	0.14
HH disposable income, equivalence scale	€ 10,607	€ 10,456	€ 10,372	€ 10,174
HH disposable income excl. SA & HA, equiv.	€ 6,752	€ 6,374	€ 6,714	€ 6,307
Average income of own parents	€ 12,044	€ 9,616	€ 11,961	€ 9,545
Average income of own and spouse's parents	€ 6,759	€ 5,067	€ 6,831	€ 5,079
HH has labour earnings	0.42	0.35	0.41	0.34
HH has entrepreneurial income	0.05	0.04	0.08	0.07
HH member unemployed	0.77	0.83	0.74	0.79
HH member unemployed: receives LMS	0.61	0.70	0.58	0.67
HH member unemployed: receives UA	0.10	0.10	0.09	0.09
HH member unemployed: receives ER	0.13	0.13	0.13	0.13
Average days HH members unemployed	158.4	201.9	151.7	192.7
HH member receives sick leave benefits	0.13	0.13	0.12	0.13
Average days HH members on sick leave	9.8	10.8	9.5	10.3
HH member receives parental leave benefits	0.13	0.03	0.13	0.03
Average days HH members on parental leave	20.1	1.3	20.0	1.2
Metropolitan area	0.22	0.21	0.22	0.21
Major university town	0.20	0.19	0.20	0.18
Other large town	0.13	0.14	0.12	0.13
Other, town-like region	0.22	0.22	0.22	0.22
Other, densely populated region	0.11	0.10	0.11	0.10
Other, rural regions	0.12	0.15	0.12	0.15
HH receives HA	0.87	0.90	0.83	0.86
HH receives SA	0.68	0.72	0.64	0.68
Actual HA receipt, eur/year (where > 0)	€ 2,889	€ 2,606	€ 2,891	€ 2,605
Actual SA receipt, eur/year (where > 0)	€ 3,199	€ 2,887	€ 3,195	€ 2,891
Simulated HA, eur/year	€ 2,313	€ 2,157	€ 2,275	€ 2,114

Note: Pooled observations from IDS datasets 2003–2011. Weighted means using IDS survey weights. Samples: (1) main eligible sampl;, (2) sample (1) + no HH composition changes during $t-3$–$t+3$; (3) sample (1) + allow eligibles with capital income > 50eur/year; (4) sample (3) + no HH composition changes during $t-3$–$t+3$.

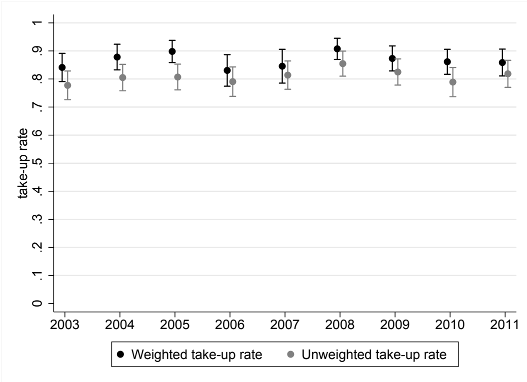
Table B.4: Descriptive statistics of social assistance eligibles with different sample definitions.

	(1)	(2)	(3)	(4)	(5)	(6)
Couple, no children	0.04	0.02	0.05	0.03	0.06	0.04
Couple, with children	0.14	0.05	0.15	0.07	0.16	0.08
Single adult, no children	0.70	0.90	0.69	0.87	0.67	0.84
Single adult, with children	0.12	0.03	0.11	0.03	0.11	0.03
# of HH members	1.7	1.3	1.7	1.3	1.8	1.4
# of underaged children	0.5	0.2	0.5	0.2	0.6	0.3
Age of HH head	40.9	46.4	40.8	46.5	41.3	47.0
Dwelling size, m 2	55.0	49.5	59.6	54.4	63.6	59.1
Dwelling size, m 2 per equivalence unit	44.7	45.8	48.0	48.9	50.3	51.8
HH highest education level: basic	0.37	0.32	0.36	0.32	0.35	0.32
HH highest education level: medium	0.49	0.53	0.48	0.52	0.48	0.51
HH highest education level: high	0.14	0.15	0.16	0.16	0.17	0.17
HH disposable income, equivalence scale	€ 10,767	€ 10,611	€ 9,917	€ 9,907	€ 9,603	€ 9,485
HH disposable income excl. SA & HA, equiv.	€ 5,877	€ 5,504	€ 5,932	€ 5,524	€ 5,964	€ 5,491
Average income of own parents	€ 11,200	€ 9,105	€ 11,889	€ 9,141	€ 11,850	€ 9,208
Average income of own and spouse's parents	€ 6,482	€ 4,785	€ 6,961	€ 4,953	€ 7,189	€ 5,135
HH has labour earnings	0.35	0.28	0.40	0.31	0.40	0.32
HH has entrepreneurial income	0.03	0.03	0.06	0.05	0.11	0.10
HH member unemployed	0.80	0.84	0.71	0.76	0.67	0.71
HH member unemployed: receives LMS	0.68	0.75	0.57	0.66	0.53	0.61
HH member unemployed: receives UA	0.08	0.06	0.08	0.06	0.08	0.06
HH member unemployed: receives ER	0.11	0.10	0.11	0.11	0.11	0.10
Average days HH members unemployed	173.5	215.7	146.9	191.8	136.9	177.7
HH member receives sick leave benefits	0.14	0.13	0.12	0.13	0.12	0.13
Average days HH members on sick leave	9.5	10.1	8.5	9.3	8.1	8.8
HH member receives parental leave benefits	0.14	0.03	0.14	0.03	0.14	0.03
Average days HH members on parental leave	20.4	1.2	18.8	1.2	18.7	1.2
Metropolitan area	0.20	0.19	0.19	0.18	0.19	0.19
Major university town	0.19	0.16	0.18	0.16	0.17	0.15
Other large town	0.14	0.15	0.13	0.14	0.12	0.13
Other, town-like region	0.23	0.23	0.24	0.24	0.24	0.23
Other, densely populated region	0.12	0.10	0.12	0.11	0.13	0.12
Other, rural regions	0.12	0.17	0.14	0.18	0.15	0.18
HH receives HA	1.00	1.00	0.80	0.85	0.74	0.78
HH receives SA	0.81	0.82	0.68	0.73	0.62	0.66
Actual HA receipt, eur/year (where >0)	€ 2,819	€ 2,589	€ 2,819	€ 2,589	€ 2,804	€ 2,586
Actual SA receipt, eur/year (where >0)	€ 3,820	€ 3,490	€ 3,761	€ 3,452	€ 3,757	€ 3,460
Simulated SA, eur/year	€ 3,023	€ 2,683	€ 3,218	€ 2,889	€ 3,299	€ 2,996

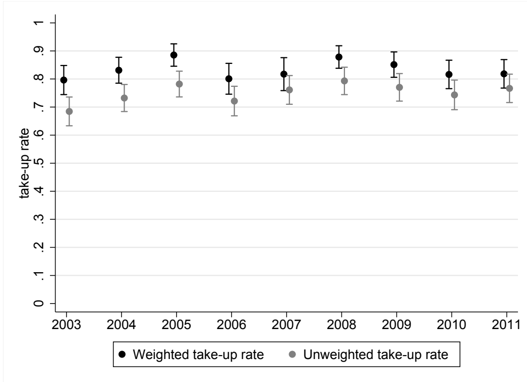
Note: Pooled observations from IDS datasets 2003–2011. Weighted means using IDS survey weights. Samples: (1) main eligible sample; (2) sample (1) + no HH composition changes during $t-3$–$t+3$; (3) sample (1) + allow eligibles with no HA receipt; (4) sample (3) + no HH composition changes during $t-3$–$t+3$; (5) sample (3) + allow eligibles with capital income > 50eur/year; (6) sample (5) + no HH composition changes during $t-3$–$t+3$.

Figure B.3: Average yearly take-up rates in 2003–2011 with different eligible sample selection.

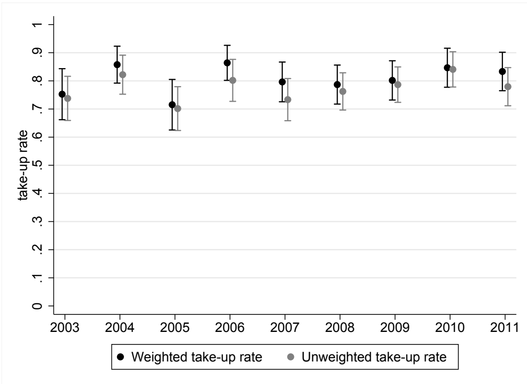
(a) Housing allowance eligibles, main sample



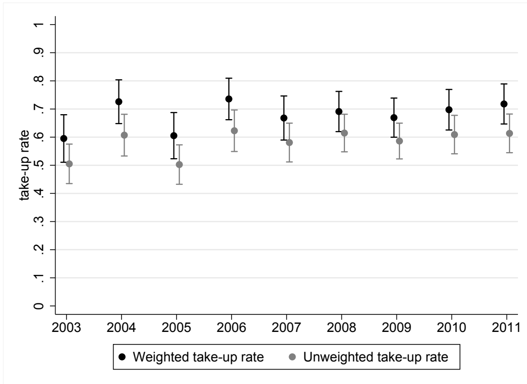
(b) Housing allowance eligibles when allowing capital income



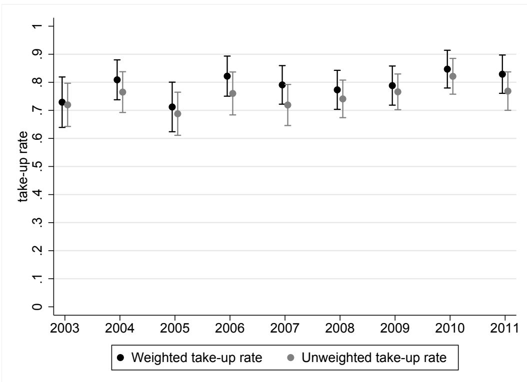
(c) Social assistance eligibles, main sample



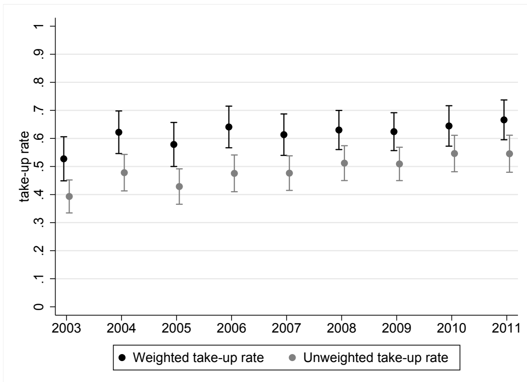
(d) Social assistance eligibles when allowing no housing benefit receipt



(e) Social assistance eligibles when allowing capital income



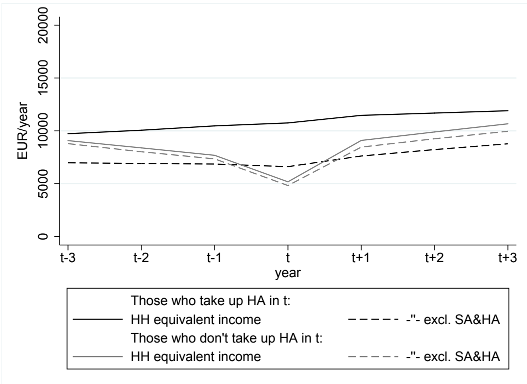
(f) Social assistance eligibles when allowing capital income and no housing benefit receipt



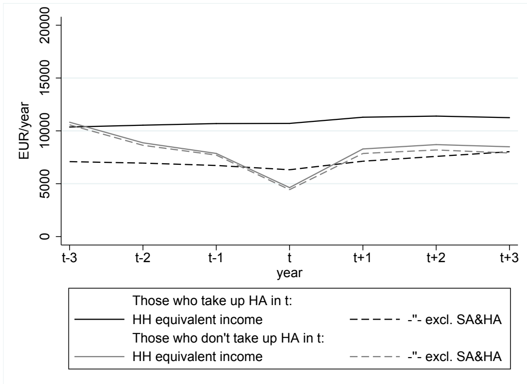
Note: Eligible samples, weighted with IDS survey weights and unweighted estimates. Standard errors corrected for household clustering.

Figure B.4: Household labour market status development over time, based on the number of days household members received specific benefits. Eligible claimers vs. eligible non-claimers using different samples.

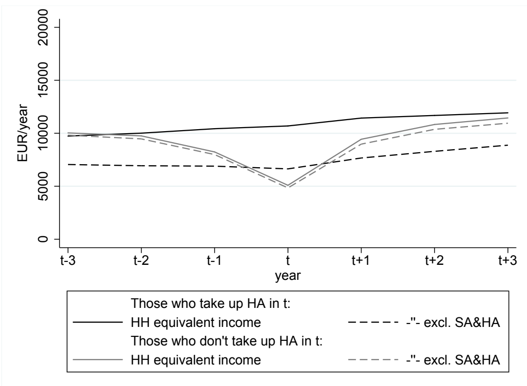
(a) Housing allowance eligibles, main sample



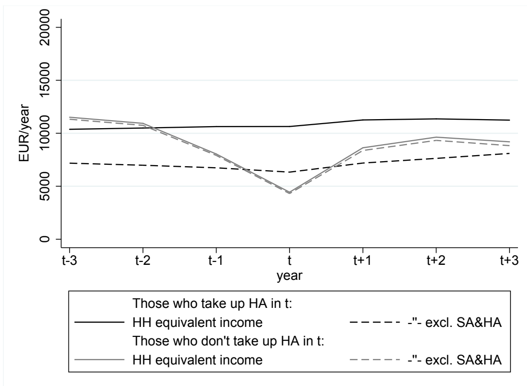
(b) Housing allowance eligibles, restricting to households with no composition changes during $t-3$–$t+3$



(c) Housing allowance eligibles when allowing capital income



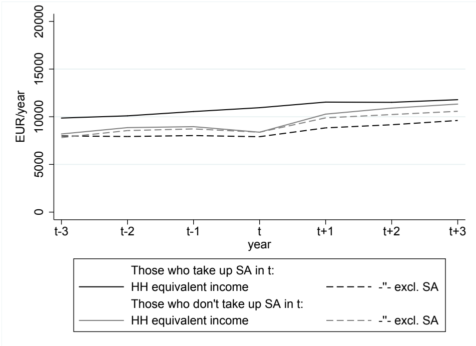
(d) Housing allowance eligibles, allowing capital income and restricting to households with no composition changes during $t-3$–$t+3$



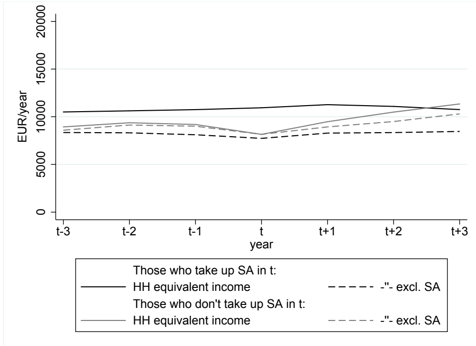
Note: Main eligible samples, pooled observations. Unweighted means. Eligibility and take-up categories, and household (HH) composition, observed in year t and kept constant throughout the graph.

Figure B.5: Household labour market status development over time, based on the number of days household members received specific benefits. Eligible claimers vs. eligible non-claimers using different samples.

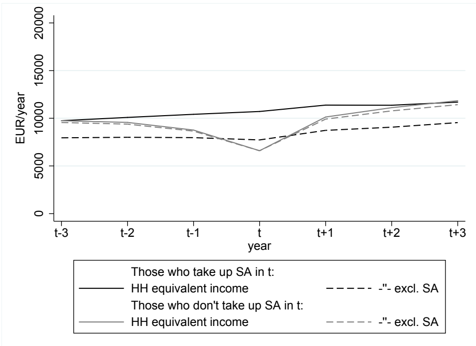
(a) Social assistance eligibles, main sample



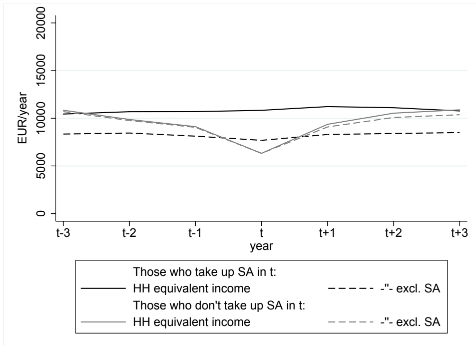
(b) Social assistance eligibles, restricting to households with no composition changes during $t-3$–$t+3$



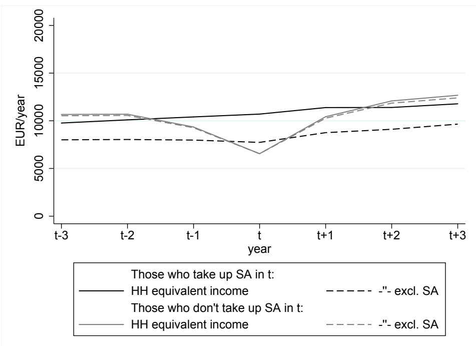
(c) Social assistance eligibles, allowing households with no HA receipt



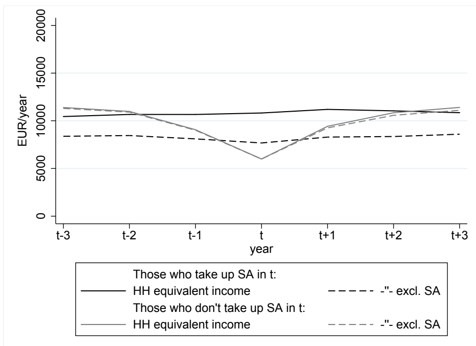
(d) Social assistance eligibles, allowing households with no HA receipt and restricting to households with no composition changes during $t-3$–$t+3$



(e) Social assistance eligibles, allowing capital income and no HA receipt



(f) Social assistance eligibles, allowing capital income and no HA receipt, and restricting to households with no composition changes during $t-3$–$t+3$



Note: Main eligible samples, pooled observations. Unweighted means. Eligibility and take-up categories, and household (HH) composition, observed in year t and kept constant throughout the graph.

Table B.5: Regression results for housing allowance take-up with different sample definitions.

	(1) HA	(2) HA	(3) HA
Age of HH head (x 10 years)	0.100	0.022	0.008
	(0.057)*	(0.057)	(0.059)
Age squared	-0.012	-0.004	-0.002
	(0.007)*	(0.007)	(0.007)
Couple	-0.149	-0.128	-0.088
	(0.048)***	(0.049)***	(0.047)*
Children	0.017	0.033	0.059
	(0.025)	(0.026)	(0.027)**
Couple x children	0.003	-0.011	-0.053
	(0.054)	(0.055)	(0.053)
Education level, medium	0.007	-0.013	-0.022
	(0.018)	(0.018)	(0.018)
Education level, high	-0.050	-0.050	-0.044
	(0.028)*	(0.027)*	(0.027)
Labour market support recip.	0.195	0.189	0.188
	(0.024)***	(0.025)***	(0.026)***
Basic UA recipient	0.085	0.063	0.056
	(0.036)**	(0.035)*	(0.038)
Earnings-related UI recipient	0.055	0.033	0.035
	(0.025)**	(0.025)	(0.026)
Has labour income	-0.031	-0.024	-0.024
	(0.017)*	(0.017)	(0.018)
Has entrepreneurial income	-0.204	-0.187	-0.206
	(0.042)***	(0.041)***	(0.044)***
Sickness benefit recipient	0.018	0.020	0.007
	(0.022)	(0.022)	(0.024)
Parental leave benefit recipient	0.073	0.075	0.051
	(0.024)***	(0.025)***	(0.026)**
m^2 per equivalence unit	-0.003	-0.003	-0.003
	(0.001)***	(0.001)***	(0.001)***
HH pre-benefit equiv. income	0.017	0.019	0.017
(x 1000 eur)	(0.005)***	(0.005)***	(0.006)***
Simulated benefit (x 1000 eur)	0.071	0.067	0.066
	(0.010)***	(0.010)***	(0.010)***
Has capital income	-0.172	-0.175	-0.145
	(0.043)***	(0.043)***	(0.045)***
Pre-benefit income change		0.000	0.000
between $t-1$,t (x 10%)		(0.000)	(0.000)
Pre-benefit income change		-0.000	-0.000
between t,$t+1$ (x 10%)		(0.000)	(0.000)***
Family composition changes		-0.054	
over $t-3$ to t		(0.019)***	
CV of pre-benefit income		-0.003	
over $t-3$ to t (x 10%)		(0.003)	
Constancy of labour mkt status		-0.007	
over $t-3$ to t (x 10%)		(0.003)**	
Family composition changes			-0.079
over $t-3$ to $t+3$			(0.020)***
CV of pre-benefit income			-0.003
over $t-3$ to $t+3$ (x 10%)			(0.004)
Constancy of labour mkt status			-0.012
over $t-3$ to $t+3$ (x 10%)			(0.004)***
Constant	0.370	0.634	0.710
	(0.119)***	(0.131)***	(0.140)***
N	2,586	2,374	2,137
R^2	0.36	0.38	0.39

Note: Pooled observations from IDS datasets 2003–2011. Weighted regression using IDS survey weights. Samples: columns (1)–(3): HA eligibles, allowing for capital income > 50eur/year. Standard errors correcting for household clustering in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression includes year and regional dummies. Pre-benefit income is (disposable income – HA – SA). Coefficient of variation (CV) of income is calculated as the ratio of the standard deviation to the mean. In column (3) year 2011 IDS sample is excluded.

Table B.6: Regression results for housing allowance take-up with different sample definitions.

	(1) HA	(2) HA	(3) HA	(4) HA	(5) HA	(6) HA
Age of HH head (x 10 years)	-0.077	-0.144	-0.145	-0.048	-0.112	-0.117
	(0.082)	(0.072)**	(0.071)**	(0.080)	(0.072)	(0.076)
Age squared	0.008	0.016	0.016	0.005	0.011	0.012
	(0.010)	(0.009)*	(0.008)*	(0.009)	(0.009)	(0.009)
Couple	-0.320	-0.308	-0.268	-0.257	-0.236	-0.207
	(0.121)***	(0.118)***	(0.104)**	(0.100)**	(0.098)**	(0.086)**
Children	-0.034	-0.046	-0.034	-0.023	-0.041	-0.022
	(0.038)	(0.038)	(0.039)	(0.040)	(0.039)	(0.041)
Couple x children	0.107	0.113	0.029	0.033	0.028	-0.044
	(0.133)	(0.129)	(0.117)	(0.113)	(0.110)	(0.102)
Education level, medium	0.016	-0.016	-0.029	0.024	-0.005	-0.014
	(0.022)	(0.021)	(0.020)	(0.022)	(0.021)	(0.021)
Education level, high	-0.022	-0.035	-0.041	-0.024	-0.028	-0.031
	(0.036)	(0.033)	(0.031)	(0.035)	(0.033)	(0.033)
Labour market support recip.	0.188	0.195	0.211	0.189	0.197	0.218
	(0.041)***	(0.046)***	(0.047)***	(0.039)***	(0.044)***	(0.045)***
Basic UA recipient	0.083	0.063	0.083	0.094	0.075	0.098
	(0.051)	(0.053)	(0.055)	(0.050)*	(0.052)	(0.054)*
Earnings-related UI recipient	0.066	0.045	0.046	0.073	0.052	0.055
	(0.042)	(0.042)	(0.044)	(0.040)*	(0.040)	(0.043)
Has labour income	-0.008	-0.004	-0.017	-0.015	-0.012	-0.023
	(0.023)	(0.023)	(0.025)	(0.023)	(0.023)	(0.024)
Has entrepreneurial income	-0.226	-0.214	-0.258	-0.154	-0.144	-0.164
	(0.069)***	(0.068)***	(0.071)***	(0.055)***	(0.055)***	(0.058)***
Sickness benefit recipient	-0.008	-0.005	-0.039	-0.010	-0.007	-0.040
	(0.028)	(0.028)	(0.030)	(0.028)	(0.029)	(0.031)
Parental leave benefit recipient	0.090	0.087	0.095	0.044	0.046	0.043
	(0.043)**	(0.040)**	(0.045)**	(0.048)	(0.044)	(0.048)
m^2 per equivalence unit	-0.002	-0.002	-0.002	-0.003	-0.002	-0.002
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
HH pre-benefit equiv. income	0.015	0.019	0.024	0.017	0.021	0.025
(x 1000 eur)	(0.007)**	(0.008)**	(0.008)***	(0.007)**	(0.007)***	(0.008)***
Simulated benefit (x 1000 eur)	0.081	0.077	0.082	0.087	0.085	0.089
	(0.014)***	(0.014)***	(0.015)***	(0.014)***	(0.013)***	(0.014)***
Has capital income				-0.231	-0.219	-0.195
				(0.060)***	(0.061)***	(0.065)***
Pre-benefit income change		0.000	-0.000		0.000	-0.000
between t-1,t (x 10%)		(0.000)	(0.000)**		(0.000)	(0.000)**
Pre-benefit income change		0.000	-0.000		0.000	-0.000
between t,$t+1$ (x 10%)		(0.000)	(0.000)		(0.000)	(0.000)
CV of pre-benefit income		-0.002			-0.005	
over $t-3$ to t (x 10%)		(0.004)			(0.004)	
Constancy of labour mkt status		-0.008			-0.009	
over $t-3$ to t (x 10%)		(0.003)***			(0.003)***	
CV of pre-benefit income			0.006			0.004
over $t-3$ to $t+3$ (x 10%)			(0.004)			(0.004)
Constancy of labour mkt status			-0.012			-0.012
over $t-3$ to $t+3$ (x 10%)			(0.004)***			(0.004)***
Constant	0.766	0.956	0.885	0.695	0.875	0.812
	(0.176)***	(0.166)***	(0.175)***	(0.173)***	(0.165)***	(0.175)***
N	973	884	775	1,114	1,022	901
R^2	0.30	0.31	0.36	0.42	0.45	0.47

Note: Pooled observations from IDS datasets 2003–2011. Weighted regression using IDS survey weights. Samples: columns (1)–(3): HA eligibles, restricting to households with no household composition changes during $t-3$–$t+3$; columns (4)–(6): HA eligibles, allowing for capital income > 50eur/year, restricting to households with no household composition changes during $t-3$–$t+3$. Standard errors correcting for household clustering in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression includes year and regional dummies. Pre-benefit income is (disposable income − HA − SA). Coefficient of variation (CV) of income is calculated as the ratio of the standard deviation to the mean. In columns (3) and (6) year 2011 IDS sample is excluded.

Table B.7: Regression results for social assistance take-up with different sample definitions.

	(1) SA	(2) SA	(3) SA	(4) SA	(5) SA	(6) SA
Age of HH head (x 10 years)	0.030	-0.047	-0.038	0.046	-0.020	-0.013
	(0.074)	(0.079)	(0.083)	(0.068)	(0.072)	(0.077)
Age squared	-0.004	0.006	0.005	-0.006	0.002	0.001
	(0.009)	(0.010)	(0.010)	(0.008)	(0.009)	(0.009)
Couple	0.030	0.036	0.040	0.016	0.018	0.016
	(0.047)	(0.048)	(0.051)	(0.040)	(0.041)	(0.043)
Children	-0.082	-0.069	-0.063	-0.085	-0.075	-0.067
	(0.039)**	(0.041)*	(0.043)	(0.036)**	(0.037)**	(0.039)*
Couple x children	-0.064	-0.073	-0.097	-0.037	-0.042	-0.053
	(0.064)	(0.065)	(0.068)	(0.055)	(0.056)	(0.058)
Education level, medium	-0.049	-0.057	-0.065	-0.057	-0.061	-0.069
	(0.027)*	(0.027)**	(0.030)**	(0.026)**	(0.026)**	(0.028)**
Education level, high	-0.149	-0.145	-0.157	-0.150	-0.143	-0.158
	(0.036)***	(0.038)***	(0.040)***	(0.032)***	(0.034)***	(0.036)***
Labour market support recip.	0.137	0.118	0.112	0.140	0.122	0.117
	(0.032)***	(0.032)***	(0.034)***	(0.029)***	(0.030)***	(0.032)***
Basic UA recipient	0.007	-0.012	0.004	0.013	-0.002	0.007
	(0.049)	(0.048)	(0.052)	(0.046)	(0.046)	(0.049)
Earnings-related UI recipient	0.038	0.025	0.029	0.029	0.016	0.020
	(0.046)	(0.047)	(0.050)	(0.042)	(0.043)	(0.047)
Has labour income	-0.070	-0.077	-0.088	-0.072	-0.082	-0.091
	(0.027)**	(0.028)***	(0.031)***	(0.025)***	(0.026)***	(0.028)***
Has entrepreneurial income	-0.216	-0.240	-0.239	-0.146	-0.169	-0.165
	(0.046)***	(0.048)***	(0.049)***	(0.034)***	(0.036)***	(0.037)***
Sickness benefit recipient	0.052	0.055	0.048	0.060	0.064	0.056
	(0.033)	(0.032)*	(0.036)	(0.031)*	(0.030)**	(0.034)*
Parental leave benefit recipient	0.076	0.067	0.110	0.063	0.050	0.089
	(0.039)*	(0.040)*	(0.042)***	(0.034)*	(0.035)	(0.037)**
m^2 per equivalence unit	-0.001	-0.000	-0.000	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
HH pre-benefit equiv. income	-0.002	-0.010	-0.009	0.001	-0.006	-0.004
(x 1000 eur)	(0.006)	(0.006)	(0.007)	(0.005)	(0.005)	(0.006)
Simulated benefit (x 1000 eur)	0.012	0.009	0.009	0.011	0.010	0.009
	(0.006)**	(0.006)	(0.006)	(0.005)**	(0.005)**	(0.005)*
HA recipient	0.509	0.479	0.470	0.489	0.459	0.457
	(0.036)***	(0.040)***	(0.043)***	(0.033)***	(0.036)***	(0.037)***
Has capital income				-0.157	-0.173	-0.177
				(0.032)***	(0.033)***	(0.035)***
Pre-benefit income change		0.000	-0.000		0.000	0.000
between $t-1$,t (x 10%)		(0.000)	(0.000)		(0.000)***	(0.000)***
Pre-benefit income change		-0.000	-0.000		-0.000	-0.000
between t,$t+1$ (x 10%)		(0.000)*	(0.000)**		(0.000)**	(0.000)***
Family composition changes		0.031			0.035	
over $t-3$ to t		(0.028)			(0.025)	
CV of pre-benefit income		-0.013			-0.012	
over $t-3$ to t (x 10%)		(0.004)***			(0.004)***	
Constancy of labour mkt status		-0.004			-0.003	
over $t-3$ to t (x 10%)		(0.004)			(0.004)	
Family composition changes			0.012			0.006
over $t-3$ to $t+3$			(0.031)			(0.028)
CV of pre-benefit income			-0.009			-0.007
over $t-3$ to $t+3$ (x 10%)			(0.005)			(0.005)
Constancy of labour mkt status			-0.004			-0.004
over $t-3$ to $t+3$ (x 10%)			(0.006)			(0.006)
Constant	0.133	0.451	0.421	0.112	0.383	0.346
	(0.144)	(0.178)**	(0.196)**	(0.134)	(0.163)**	(0.176)**
N	1,754	1,614	1,439	2,182	2,030	1,829
R^2	0.37	0.36	0.36	0.44	0.45	0.44

Note: Pooled observations from IDS datasets 2003–2011. Weighted regression using IDS survey weights. Samples: columns (1)–(3): SA eligibles, allowing for households not receiving HA; columns (4)–(6): SA eligibles, allowing for households not receiving HA and with capital income > 50eur/year. Standard errors correcting for household clustering in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression includes year and regional dummies. Pre-benefit income is (disposable income − SA). Coefficient of variation (CV) of income is calculated as the ratio of the standard deviation to the mean. In columns (3) and (6) year 2011 IDS sample is excluded.

Table B.8: Regression results for social assistance take-up with different sample definitions.

	(1) SA	(2) SA	(3) SA	(4) SA	(5) SA	(6) SA	(7) SA	(8) SA	(9) SA
Age of HH head (x 10 years)	0.294	0.259	0.144	0.309	0.260	0.193	0.307	0.239	0.178
	(0.174)*	(0.177)	(0.188)	(0.161)*	(0.157)*	(0.175)	(0.145)**	(0.140)*	(0.156)
Age squared	-0.031	-0.027	-0.014	-0.033	-0.028	-0.020	-0.033	-0.025	-0.018
	(0.019)	(0.020)	(0.021)	(0.018)*	(0.018)	(0.020)	(0.016)**	(0.015)*	(0.017)
Couple	-0.147	-0.143	-0.141	0.060	0.063	0.056	0.024	0.033	0.016
	(0.144)	(0.149)	(0.170)	(0.107)	(0.108)	(0.110)	(0.082)	(0.081)	(0.083)
Children	-0.014	-0.012	-0.046	-0.067	-0.051	-0.107	-0.124	-0.115	-0.151
	(0.100)	(0.100)	(0.112)	(0.085)	(0.085)	(0.100)	(0.074)*	(0.075)	(0.082)*
Couple x children	0.092	0.076	0.121	-0.049	-0.078	-0.018	0.056	0.034	0.090
	(0.185)	(0.188)	(0.211)	(0.142)	(0.142)	(0.154)	(0.111)	(0.111)	(0.118)
Education level, medium	-0.037	-0.043	-0.038	-0.030	-0.033	-0.043	-0.032	-0.032	-0.042
	(0.043)	(0.042)	(0.047)	(0.040)	(0.040)	(0.045)	(0.036)	(0.036)	(0.041)
Education level, high	-0.113	-0.125	-0.115	-0.117	-0.117	-0.102	-0.128	-0.119	-0.117
	(0.054)**	(0.057)**	(0.060)*	(0.047)**	(0.051)**	(0.056)*	(0.042)***	(0.045)***	(0.049)**
Labour market support recip.	0.225	0.225	0.265	0.153	0.148	0.169	0.141	0.129	0.161
	(0.061)***	(0.062)***	(0.068)***	(0.051)***	(0.052)***	(0.060)***	(0.046)***	(0.046)***	(0.053)***
Basic UA recipient	0.053	0.069	0.066	-0.032	-0.018	-0.028	-0.017	-0.008	-0.006
	(0.071)	(0.075)	(0.081)	(0.064)	(0.067)	(0.076)	(0.060)	(0.062)	(0.071)
Earnings-related UI recipient	0.081	0.082	0.094	0.030	0.031	0.042	0.007	0.002	0.012
	(0.077)	(0.078)	(0.088)	(0.071)	(0.072)	(0.079)	(0.066)	(0.067)	(0.074)
Has labour income	-0.086	-0.069	-0.117	-0.074	-0.059	-0.081	-0.079	-0.075	-0.086
	(0.048)*	(0.050)	(0.059)**	(0.043)*	(0.045)	(0.053)	(0.039)**	(0.040)*	(0.048)*
Has entrepreneurial income	-0.311	-0.306	-0.243	-0.240	-0.251	-0.200	-0.159	-0.172	-0.146
	(0.136)**	(0.143)**	(0.169)	(0.076)***	(0.083)***	(0.089)**	(0.051)***	(0.053)***	(0.056)***
Sickness benefit recipient	0.002	0.010	-0.000	0.031	0.044	0.028	0.051	0.063	0.051
	(0.048)	(0.047)	(0.053)	(0.047)	(0.046)	(0.054)	(0.044)	(0.043)	(0.050)
Parental leave benefit recipient	-0.088	-0.085	0.071	-0.027	-0.024	0.104	-0.025	-0.029	0.076
	(0.120)	(0.118)	(0.113)	(0.109)	(0.107)	(0.107)	(0.085)	(0.084)	(0.085)
m^2 per equivalence unit	0.000	0.000	-0.000	-0.001	-0.000	-0.001	-0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
HH pre-benefit equiv. income (x 1000 eur)	-0.001	-0.006	-0.004	0.001	-0.006	-0.000	0.004	-0.001	0.004
	(0.014)	(0.014)	(0.017)	(0.010)	(0.010)	(0.011)	(0.008)	(0.008)	(0.009)
Simulated benefit (x 1000 eur)	0.035	0.035	0.039	0.015	0.015	0.019	0.013	0.014	0.018
	(0.015)**	(0.015)**	(0.017)**	(0.011)	(0.011)	(0.012)	(0.008)	(0.009)	(0.009)*
HA recipient				0.511	0.481	0.493	0.491	0.463	0.475
				(0.060)***	(0.068)***	(0.074)***	(0.051)***	(0.056)***	(0.059)***
Has capital income							-0.160	-0.181	-0.175
							(0.045)***	(0.046)***	(0.051)***
Pre-benefit income change between $t-1,t$ (x 10%)		0.005	0.002		0.004	-0.000		0.000	0.000
		(0.005)	(0.004)		(0.004)	(0.004)		(0.000)**	(0.001)
Pre-benefit income change between $t,t+1$ (x 10%)		0.001	-0.006		-0.000	-0.000		-0.000	-0.000
		(0.002)	(0.005)		(0.000)**	(0.000)**		(0.000)**	(0.000)**
CV of pre-benefit income over $t-3$ to t (x 10%)		-0.012			-0.015			-0.013	
		(0.010)			(0.008)**			(0.006)**	
Constancy of labour mkt status over $t-3$ to t (x 10%)		0.004			0.006			0.005	
		(0.007)			(0.007)			(0.006)	
CV of pre-benefit income over $t-3$ to $t+3$ (x 10%)			-0.004			-0.007			-0.006
			(0.011)			(0.007)			(0.006)
Constancy of labour mkt status over $t-3$ to $t+3$ (x 10%)			0.007			0.009			0.009
			(0.010)			(0.009)			(0.009)
Constant	-0.093	0.043	0.191	-0.512	-0.332	-0.318	-0.502	-0.306	-0.305
	(0.388)	(0.397)	(0.466)	(0.362)	(0.362)	(0.418)	(0.329)	(0.323)	(0.369)
N	550	541	462	723	680	587	899	853	746
R^2	0.18	0.19	0.21	0.37	0.37	0.37	0.47	0.48	0.48

Note: Pooled observations from IDS datasets 2003–2011. Weighted regression using IDS survey weights. Samples: columns (1)–(3): main SA eligible sample + restricting to households with no composition changes during $t-3$–$t+3$; (4)–(6): SA eligibles, allowing for households not receiving HA + restricting to households with no composition changes during $t-3$–$t+3$; (7)–(9): SA eligibles, allowing for households not receiving HA and with capital income > 50eur/year + restricting to households with no composition changes during $t-3$–$t+3$. Standard errors correcting for household clustering in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression includes year and regional dummies. Pre-benefit income is (disposable income – SA). Coefficient of variation (CV) of income is calculated as the ratio of the standard deviation to the mean. In columns (3), (6) and (9) year 2011 IDS sample is excluded.

Table B.9: Unweighted regression results for housing allowance and social assistance take-up using the main eligible samples.

	(1) HA	(2) HA	(3) HA	(4) SA	(5) SA	(6) SA
Age of HH head (x 10 years)	0.234	0.104	0.086	0.024	-0.028	-0.001
	(0.052)***	(0.054)*	(0.055)	(0.087)	(0.090)	(0.097)
Age squared	-0.027	-0.012	-0.010	-0.001	0.004	0.001
	(0.006)***	(0.007)*	(0.007)	(0.011)	(0.011)	(0.012)
Couple	-0.157	-0.131	-0.104	0.042	0.044	0.054
	(0.043)***	(0.045)***	(0.045)**	(0.046)	(0.048)	(0.054)
Children	-0.025	0.035	0.052	-0.062	-0.056	-0.032
	(0.024)	(0.025)	(0.026)**	(0.047)	(0.046)	(0.049)
Couple x children	0.006	-0.028	-0.060	-0.047	-0.048	-0.079
	(0.050)	(0.051)	(0.052)	(0.066)	(0.067)	(0.073)
Education level, medium	-0.003	-0.026	-0.034	-0.093	-0.100	-0.117
	(0.017)	(0.016)	(0.017)**	(0.028)***	(0.028)***	(0.029)***
Education level, high	-0.051	-0.057	-0.049	-0.161	-0.154	-0.181
	(0.026)**	(0.024)**	(0.025)*	(0.041)***	(0.041)***	(0.043)***
Labour market support recip.	0.229	0.222	0.222	0.191	0.168	0.174
	(0.021)***	(0.022)***	(0.023)***	(0.033)***	(0.033)***	(0.034)***
Basic UA recipient	0.089	0.073	0.065	0.037	0.025	0.039
	(0.030)***	(0.030)**	(0.032)**	(0.052)	(0.052)	(0.055)
Earnings-related UI recipient	0.093	0.067	0.064	-0.006	-0.011	-0.004
	(0.022)***	(0.021)***	(0.022)***	(0.046)	(0.047)	(0.050)
Has labour income	-0.047	-0.050	-0.056	-0.077	-0.078	-0.088
	(0.016)***	(0.016)***	(0.017)***	(0.029)***	(0.029)***	(0.031)***
Has entrepreneurial income	-0.217	-0.199	-0.220	-0.334	-0.341	-0.338
	(0.037)***	(0.038)***	(0.040)***	(0.063)***	(0.064)***	(0.068)***
Sickness benefit recipient	0.035	0.038	0.030	0.078	0.083	0.086
	(0.020)*	(0.020)*	(0.021)	(0.031)**	(0.030)***	(0.034)**
Parental leave benefit recipient	0.116	0.092	0.081	0.059	0.060	0.101
	(0.024)***	(0.024)***	(0.026)***	(0.044)	(0.044)	(0.046)**
m^2 per equivalence unit	-0.003	-0.003	-0.003	-0.001	-0.000	0.000
	(0.000)***	(0.001)***	(0.001)***	(0.001)	(0.001)	(0.001)
HH pre-benefit equiv. income	0.027	0.029	0.029	-0.005	-0.016	-0.019
(x 1000 eur)	(0.004)***	(0.005)***	(0.005)***	(0.008)	(0.008)*	(0.010)**
Simulated benefit (x 1000 eur)	0.080	0.070	0.070	0.017	0.015	0.015
	(0.008)***	(0.008)***	(0.009)***	(0.008)**	(0.008)**	(0.007)**
Pre-benefit income change		0.000	0.000		0.000	0.000
between$t-1$, t (x 10%)		(0.000)*	(0.000)		(0.000)	(0.000)
Pre-benefit income change		-0.000	-0.000		-0.004	-0.006
between t,$t+1$ (x 10%)		(0.000)*	(0.000)***		(0.001)***	(0.003)**
Family composition changes		-0.047			0.007	
over $t-3$ to t		(0.017)***			(0.029)	
CV of pre-benefit income		-0.001			-0.012	
over $t-3$ to t (x 10%)		(0.003)			(0.005)***	
Constancy of labour mkt status		-0.007			-0.007	
over $t-3$ to t (x 10%)		(0.002)***			(0.005)	
Family composition changes			-0.072			-0.018
over $t-3$ to $t+3$			(0.018)***			(0.032)
CV of pre-benefit income			-0.001			-0.008
over $t-3$ to $t+3$ (x 10%)			(0.004)			(0.007)
Constancy of labour mkt status			-0.013			-0.011
over $t-3$ to $t+3$ (x 10%)			(0.004)***			(0.006)*
Constant	-0.068	0.295	0.394	0.650	0.962	0.913
	(0.104)	(0.124)**	(0.132)***	(0.171)***	(0.190)***	(0.221)***
N	2,261	2,059	1,844	1,221	1,189	1,049
R^2	0.36	0.36	0.38	0.16	0.19	0.19

Note: Pooled observations from IDS datasets 2003–2011, main eligible samples. Unweighted regression. Standard errors correcting for household clustering in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression includes year and regional dummies. Pre-benefit income is (disposable income − HA − SA) for HA take-up, (disposable income − SA) for SA take-up. Coefficient of variation (CV) of income is calculated as the ratio of the standard deviation to the mean. In columns (3) and (6) year 2011 IDS sample is excluded.

Does information increase the take-up of social benefits? Evidence from a new benefit program*

Abstract

The effectiveness of transfer programs can be significantly reduced if eligible individuals fail to apply for them. In this paper we study the impact of information provision on the take-up of social benefits. We exploit the implementation of the guarantee pension program in Finland in 2011, which offered a minimum monthly pension (688 euros) to low-income pensioners. The Finnish Social Insurance Institution sent information letters and application forms to a portion of the eligible population a month before implementation. We find clear evidence that this mailing significantly increased take-up and prompted pensioners to apply sooner, showing that simple and inexpensive mailings can have a large effect on benefit take-up among individuals outside the labor force.

Keywords: Take-up of social benefits, information

JEL classification codes: H24, H55, D03

*This paper is joint with Tuomas Matikka (VATT Institute for Economic Research and CESifo) and has been published as VATT Working Paper 83/2016. We thank Konsta Lavaste for excellent research assistance.

1 Introduction

A prominent feature of various social benefit programs is that some proportion of the targeted individuals fail to apply for the benefits they are entitled to (Currie 2006). This incomplete take-up impairs the effectiveness of benefit programs and limits the ability to reduce poverty and increase well-being. Therefore, knowledge of how to affect the take-up rate is a key issue in implementing such policies.

Various factors could explain why individuals do not apply for the benefits they are entitled to. These have traditionally been categorized into information costs, transaction costs and stigma related to applying for the benefit, assuming that people rationally judge these costs against the size of the benefit (Currie 2006). Recently, different types of cognitive costs that could interfere in the take-up decision have gained attention, such as the complexity and non-transparency of many programs (Chetty, Friedman and Saez 2013; Chetty and Saez 2013; Liebman and Zeckhauser 2004). Therefore, evidence of the effectiveness of various means to increase the availability of information and the clarity of benefit rules and application procedures is relevant in terms of policy-making, as these factors are likely to play a significant role in affecting take-up.[1]

We study how a simple information letter sent to eligible individuals affects the take-up rate. We study the introduction of the guarantee pension program in Finland in 2011. The program offers a minimum monthly pension of 688 euros for all pensioners below this pension income level, targeted at pensioners with no or very little work history. For a typical eligible pensioner, the guarantee pension increased monthly income by 100–170 euros.

As part of an information campaign, the Finnish Social Insurance Institution (SII) sent information letters together with an application form to a group of eligible pensioners prior to implementation. This provides us quasi-experimental variation in receiving this letter that enables a compelling analysis of the impact of the letter on the take-up rate of this benefit.

We contribute to the scarce literature on the effectiveness of information by studying individuals outside the labor force, namely the low-income elderly and disability pensioners. Earlier literature on take-up has mostly concen-

[1] Complex social programs that lead to incomplete take-up can be effective in targeting the benefit at those with the greatest need (see e.g. Kleven and Kopczuk 2011). However, this is not necessarily the case when considering benefits that are specifically targeted at non-working individuals at the very low end of the income distribution with clear eligibility criteria, such as the guarantee pension program in Finland.

trated on analyzing low-income working individuals who are entitled to social benefits or tax credits, such as the Earned Income Tax Credit (EITC) in the US (Bhargava and Manoli 2015; Chetty et al. 2013; Chetty and Saez 2013). However, knowledge of how inexpensive and simple information provision affects take-up among non-working individuals is very policy-relevant, as incomplete take-up is bound to impair the effectiveness of welfare policies targeted at these vulnerable individuals in great economic need. We also compare the effectiveness of the information letter to that of press coverage in affecting the take-up rate, thus providing evidence of the effectiveness of different measures of information provision. In addition, by utilizing detailed register data on medicine reimbursements and medical diagnosis, we provide new evidence of the effects of directed information for individuals with different states of health.

The SII sent the letters to a selected group of eligible pensioners in late January 2011, a month before actual implementation of the program. The mailing included information about the guarantee pension program and the eligibility criteria, details on how to apply for the benefit, and an application form and a postage-paid return envelope.

The SII targeted the letters at pensioners who were most likely eligible for the guarantee pension: recipients of a full national pension from the SII. This group constitutes 65% of the (estimated) eligible population, leaving 35% of eligible individuals without the letter. The treatment group (letter in January) and the control group (no letter in January) are very similar, consisting of eligible low-income pensioners with no or small employment pensions (earnings-related pensions), with both groups also receiving other social benefits from the SII and eligible for varying amounts of guarantee pension benefits. In addition, during August and September 2011 the SII contacted all remaining pensioners who were potentially eligible for the guarantee pension but had not applied for it by then.

The overall take-up rate of the guarantee pension was very high, as 93% of all eligible pensioners applied for it by the end of 2011. Nevertheless, our results show that the letter had a significant impact on take-up, and prompted eligible pensioners to apply sooner. We find a clear difference in the take-up rate of eligible pensioners whose probability of receiving the letter is higher, compared to individuals with similar pension income who are eligible for similar benefit amounts but were less likely to receive the letter. This shows that selection to the treatment group based on national pension income level is not driving the results. Furthermore, we observe a clear increase in take-up in the control group after the SII contacted all remaining eligible pensioners during August–September of 2011 who had not applied by then. This evidence also illustrates the impact of personal information.

Our regression results show that receiving the letter increased the average probability of applying for the guarantee pension before August by 33 percentage points – 55% relative to the control group take-up rate. In addition, the January letter increased the take-up rate by 12 percentage points (15% relative to control group) by the end of 2011, implying a significant effect of early information on the overall take-up that persists despite the second round of information campaigning in August–September.

We find larger responses among severely ill or disabled pensioners receiving a specific benefit intended for assistance in normal daily activities (cooking, eating, dressing etc.). This finding suggests that the effect of the letter was greater for those who were more likely not to manage their financial issues solely by themselves but rather receive assistance from, for example, close relatives. In contrast, we find no significant differences in responses between individuals with or without medical expenses for severe or long-term illnesses, nor between individuals with or without prescribed medicine for mental illnesses. This suggests that deteriorated health does not confound the effect of the letter, at least within the population with prescribed medicine for their illnesses. Overall, the letter had a large and significant effect on take-up in various subgroups, implying that the response is not limited to any specific types of pensioners. Finally, we find no evidence of the impact of press coverage on take-up, implying that informing the eligible population through the mass media is significantly less effective than personal mailings.

Our work is related to recent empirical literature that has found information provision to play a key role in take-up. Bhargava and Manoli (2015) show that clarifying the details and simplification of the application form for the EITC in the US increased claiming for this tax credit, while attempts to reduce perceived stigma and claiming costs did not. Mastrobuoni (2011) finds that the official social security statement sent to people close to their retirement age in the US increased the accuracy of their estimates of their future social security benefits. In a similar vein, Liebman and Luttmer (2015) show that simple information interventions increased the understanding of means-tested social security benefits in the US, and increased labor supply among individuals close to retirement age. Bettinger, Long, Oreopoulos and

Sanbonmatsu (2012) find that mere information treatments did not increase the take-up of college financial aid in the US, whereas the combined treatment of information and personal assistance did. However, Zantomio (2015) finds that a bundle of measures intended to encourage take-up, including simplified claiming procedures, application assistance and targeted mailings, had no impact on take-up of the pension credit, which replaced the existing minimum income guarantee program in the UK in 2003.[2]

We add to the literature by finding that information has a distinctive role in the take-up of social benefits, particularly among vulnerable individuals outside the labor force. Providing simple information on eligibility and an application form had a very significant effect on take-up, and prompted eligible individuals to apply for the new benefit sooner. From a policy perspective, providing this type of a simple treatment to eligible individuals provides a cost-effective way to increase take-up and the effectiveness of social programs targeted at the very low end of the income distribution.

The paper proceeds as follows: Section 2 describes the guarantee pension program and the information campaign. Section 3 introduces the data and presents the descriptive analysis. Section 4 presents the regression results, and Section 5 concludes the study.

2 Guarantee pension and the information campaign

2.1 Guarantee pension program

Prior to the implementation of the guarantee pension program, the main sources of pension income for low-income pensioners were typically the state-paid national pension and employment pensions (earnings-related pensions) based on contributions accumulated during working life. The national pension provides the basic subsistence for pensioners with little or no employment pension income due to short work histories. The maximum amount is 586.46 euros per month for single persons and 520.19 euros for those living with a

[2]Information provision can also reduce take-up. For example, Hertel-Fernandez and Wenger (2013) find that providing accurate information about unemployment insurance (UI) benefits reduced self-reported willingness to apply for UI benefits. They hypothesize that the experiment participants had an overly optimistic view of UI generosity and the ease of application, and by correcting those beliefs, the new information reduced participation intentions.

partner.[3] If the person has employment pension income above 51.79 euros per month, this gradually reduces the national pension such that each euro of employment pension income reduces the national pension by 50 cents.[4]

The guarantee pension was implemented on March 1, 2011. The goal of the guarantee pension program is to reduce pensioner poverty by raising all pensioners' incomes to a minimum level – it is a top-up benefit for low pension incomes. The program is targeted at approximately 100,000 low-income pensioners out of a total of 1.3 million pensioners in Finland. The target group is pensioners with little or no work history, who therefore have accumulated very little or no employment pension. The target population consists mainly of low-income old-age pensioners, and the disabled who collect a permanent disability pension. The guarantee pension is administered by the Finnish Social Security Institution (SII), which administers most welfare benefits in Finland, including national pensions.

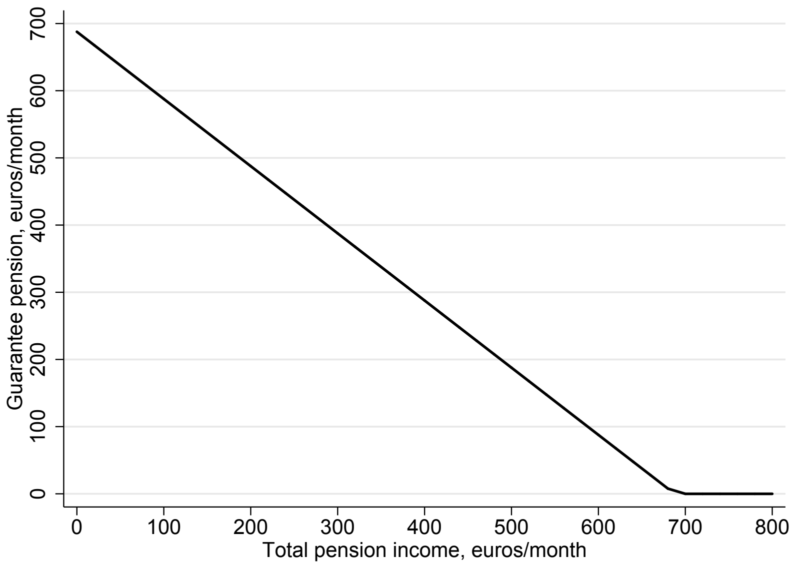
The main eligibility requirement for the guarantee pension is having total pension income below 687.74 euros per month. Figure 1 illustrates the simple linear relationship between total pension income and guarantee pension entitlement. If an eligible pensioner has zero total pension income, she is eligible for the maximum benefit equal to the threshold. Each euro of pension income reduces the benefit by one euro. The minimum payable amount is 6.23 euros.

Only pension income affects the guarantee pension – other types of social benefits do not. Low-income pensioners may be eligible for pensioners' housing allowance from the SII or social assistance from their municipality. The SII also provides health-related benefits such as reimbursements for medical expenses and a care allowance for those who live at home but need constant assistance in daily activities due to severe illness or disability, but these benefits do not affect guarantee pension eligibility either. Also, the applicant's earnings (earned income and capital income) and wealth do not affect eligibility, nor do their spouse's pension income, earnings, or wealth. Additional eligibility requirements are that the pensioner is currently living in Finland, and has lived in the country for at least 3 years between the ages of 16–65. Also, part-time employment pensions do not give entitlement to guarantee pension benefits.

[3]Since our analysis focuses on 2011, the year of implementation, we express all monetary amounts in 2011 terms. The amounts of national pension and guarantee pension are increased annually according to a cost-of-living index.

[4]In addition, the national pension can be lowered due to, for example, having taken up an early old-age pension or having lived abroad for long periods during one's working life.

Figure 1: Guarantee pension eligibility based on total pension income.



An application to the SII is required in order to receive the guarantee pension. In the Finnish welfare system, an application to the SII is always required in order to claim a benefit.[5] Applying for the guarantee pension became possible one month before implementation, on February 1 2011. There is no re-application requirement, meaning that once a person has been granted the benefit, she receives it monthly until her retirement ends or she passes away. However, the benefit amount may of course change if the person's pension income changes.

If an individual does not apply for the guarantee pension immediately, she can receive the benefits retrospectively for up to six months. Beyond that time older benefit months are lost permanently. Thus if a pensioner was eligible when the benefit was implemented in March 2011 and applied by September 30 of that year, she could receive the total sum of benefits she was eligible for, starting from March.

The application process was made extremely simple. The application form requires only the applicant's name and bank account information, and a declaration of any received pension income not paid out by the SII or Finnish pension funds. The application form is presented in Figure A.2 in the Appendix. In addition to the paper form, the benefit can be applied for using the SII online platform, calling the SII service number or visiting an SII office.

[5]This requirement even includes child benefits that every household with children under the age of 17 is eligible for, irrespective of household income or any other characteristics.

In addition to low transaction costs, stigma costs are also likely to be low. Implementation of the benefit was based very much on a notion of fairness, which was clearly visible in the news coverage of the political process during 2008–2010. Therefore, negative feelings about take-up are likely to be low, especially relative to other more discretionary benefits such as social assistance. Therefore, the program constitutes a suitable testing ground for the impact of the letter and personal information, as other types of take-up costs are low.

2.2 Information campaign

January letters: information prior to implementation. The SII wanted to ensure that the eligible population became aware of the benefit and would know how to apply for it. For our purposes, the main feature of the information campaign was the targeted information letter sent to a portion of the eligible population prior to implementation. The SII aimed to avoid personally informing ineligible pensioners about the new benefit. Therefore, they targeted the mailing at a subpopulation that was most likely to be eligible: those receiving a full national pension. For this group of pensioners, the total amount of pension income can easily and reliably be checked in the SII registers, and thus the room for errors in eligibility status was small. Therefore, this procedure was not originally selected by the SII in order to study how the letter affects take-up, but rather to avoid informing potentially ineligible pensioners about the program in the early stages of implementation.

The SII sent the letters on January 24–28, 2011. No pensioners other than the recipients of full national pensions were approached in this way, so their knowledge of the benefit relied on, for example, reading the news or perhaps visiting an SII office for other reasons. The SII originally estimated that there would be around 115,000 eligible pensioners in 2011, 75,400 of whom were sent an information letter.

The mailing consisted of a short letter explaining the existence of the new benefit and the eligibility criteria, and that the recipient was likely to be eligible. In addition, the letter included an application form and a postage-paid return envelope. Thus, in addition to information provision, the letter presumably reduced the costs of applying for the benefit by providing the application form and a return envelope. The information letter and the application form are presented in the Appendix (Figures A.1 and A.2). Since the SII sends out letters to its clients frequently (e.g. regarding benefit application decisions), there are standardized procedures for mass mailings in place, and therefore this particular information letter mailing was relatively inexpensive to carry out.

We utilize the quasi-experimental variation in receiving the letter to identify the effect of the letter treatment on take-up. We evaluate the impact of receiving the letter on overall take-up, but focus especially on the impact on *early* take-up – applying before August 2011 – as we will clarify in detail below. Since the letters were targeted at those on a full national pension, letter recipient status among the eligible population is not random. Nevertheless, the treatment group – pensioners who received the letter in January – and the control group – eligible pensioners who did not receive the letter in January – both consist of low-income pensioners receiving some pension benefits from the SII, mainly varying amounts of national pension. This implies that both groups were very accustomed to dealing with the SII, and knew how social benefits are applied for in Finland. Furthermore, since selection to the treatment is based on observable characteristics related to the level and composition of total pension income, we can control for any of these differences between the treatment and control groups. We discuss the differences between the treatment and control groups in more detail in Section 3.

August information campaign: reminders after implementation. In addition to the January letter, the SII conducted a second targeted information campaign in August–September 2011. As the guarantee pension benefit can be granted retrospectively for up to six months, at this point in time the SII wanted to ensure they reached the entire pool of eligible persons by the end of September. SII district offices were instructed to contact those deemed eligible but who had not yet applied for the benefit between August 9–September 14. Persons deemed to be eligible were either sent a letter (similar to that in January) or phoned by local SII officials. The SII estimated that this active campaigning meant that all eligible individuals were likely to have heard of the benefit by mid-September 2011.

Our main analysis focuses on estimating the impact of the January letter on take-up by the end of July 2011. Later take-up is affected by the second round of information in August–September, and would at least partly confound our analysis. Moreover, as guarantee pension recipients consist of very low-income individuals with limited access to any outside income, there are no obvious incentives to postpone application for several months. Therefore, it is reasonable to assume that the potential effect of the January letters materialized soon after they were received. However, we also analyze the effect of the January letter beyond September 2011, and study how the August–September information campaign itself affected take-up.

Media coverage. In addition, the SII's information campaign included strong visibility in various media. The SII sent out several press releases over the course of the year, published news items in the SII's customer magazine and website, and distributed information brochures at local SII offices. Overall, the SII published 14 press releases between the falls of 2010 and 2011. Their own media monitoring found 130 hits related to guarantee pensions in various media in January-October 2011, and several others towards the end of 2010.

For those who did not receive the letter in January, their awareness of the benefit and their potential eligibility would thus rely on being exposed to these other information channels. Since almost all eligible pensioners received at least some national pension income and very often other types of benefits from the SII, it is likely that most of the eligible population would have been exposed to some form of information on the benefit during 2011. Receiving a personal letter with an application form was, however, likely to be a much stronger prompt about eligibility than the more general information and publicity.

Furthermore, given the wide media coverage and other potential sources of information, our results on receiving the January letter can be interpreted as the effect of personal information within an environment of widespread general knowledge of the benefit. In addition, we conduct a separate analysis on the effectiveness of press coverage in increasing take-up in Section 4.4.

3 Data and descriptive analysis

3.1 Data

We use detailed register data from the SII covering both the national pensions that the SII administers as well as all other pension data that the Finnish Centre for Pensions collects from pension funds. The base population consists of all pensioners in 2011 with a valid pension at the end of the year. For pensions paid out by the SII, we observe monthly payments, but for employment pensions we can only observe the information at the end of the year. Therefore, we use the amount paid in December as our measure of monthly pension income.[6]

The data also include background characteristics such as gender, place of residence and age. Also, we have data on individual medicine expenses reimbursed by the SII, and, for severe illnesses, the medical diagnosis related to these prescription medicines.[7] Additionally, we have merged information from the registers of Statistics Finland regarding education level and other income types.

We limit our estimation sample to pensioners who we estimate to be eligible for the guarantee pension benefit based on the eligibility criteria described above. We discuss our eligibility estimation in more detail below. In order to reliably study the effect of the January letter on take-up, we further restrict our sample to pensioners who were retired on January 2011 and continued their retirement until at least December 2011. With these restrictions, we have a sample of 105,574 eligible pensioners. The treatment group consists of 68,655 pensioners who received the January information letter.

Table 1 shows the average characteristics of the sample. Old-age pension and full disability pension are the most common pension types. Almost all pensioners (97%) in the sample receive national pension income from the SII. There is a small minority who do not receive any national pension, which can be due to, for example, receiving other pension income or having lived abroad for many years. Less than half of the eligible pensioners have employment pension income. Other sources of pension income are rare.

Average total pension income is 574 euros per month in the sample, and average guarantee pension benefits received are around 102 euros per month. Approximately half of the pensioners receive pensioners' housing allowance from the SII, and a similar number receive pensioners' care allowance. 98% of eligible pensioners receive some of these benefits or the national pension

[6]Changes in monthly pension payments within a year are very rare.

[7]The national health insurance system covers part of individual medicine expenses. Basic medicine expenses cover prescribed medicine for a wide variety of illnesses, such as exanthema and allergies, and common prescribed medicines such as antibiotics. Special medicine expenses cover prescribed medicine for severe and long-term illnesses, such as cancer and cardiovascular diseases. Basic reimbursement is directly subtracted from the selling price at the pharmacy. The eligibility for special reimbursements is based on a medical certificate, and once the reimbursement status is granted, the reimbursed amount is directly subtracted from the selling price at the pharmacy. For more information on the special medicine reimbursement scheme, see http://www.kela.fi/web/en/reimbursements-for-medicine-expences_special-reimbursement (accessed 30.8.2016).

Table 1: Descriptive statistics for the total sample of pensioners eligible for the guarantee pension, and for the treatment group (January letter) and the control group (no January letter).

	Total sample (N=105,574)		Treatment group (N=68,655)		Control group (N=36,919)	
	share	mean	share	mean	share	mean
National pension, eur/month	0.97	528	1.00	566	0.91	451
Employment pension, eur/month	0.43	139	0.17	33	0.91	177
Other pension income, eur/month	0.02	138	0.01	109	0.03	156
Total pension income, eur/month	–	574	–	572	–	576
Guarantee pension (observed), eur/month	0.92	102	0.97	108	0.82	88
Old-age pensioners	0.39	–	0.26	–	0.63	–
Disability pensioners	0.51	–	0.64	–	0.26	–
Other pension types	0.12	–	0.10	–	0.16	–
Pensioners' housing allowance, eur/month	0.48	238	0.58	238	0.30	239
Pensioners' care allowance, eur/month	0.44	143	0.56	147	0.21	125
Basic medicine expense reimbursement	0.81	348	0.78	355	0.85	335
Special medicine expense reimbursement	0.57	1,159	0.58	1,286	0.56	911
Any mental illness medicine reimbursement	0.24	–	0.30	–	0.12	–
Male	0.38	–	0.44	–	0.26	–
Has a spouse	0.34	–	0.20	–	0.61	–
Age, years	–	59	–	53	–	69
Level of education (1–5)	–	1.4	–	1.4	–	1.5
Labor earnings in 2011	0.06	4,899	0.05	4,817	0.08	4,997
Capital income in 2011	0.30	2,883	0.24	2,342	0.41	3,467
Entrepreneurial income in 2011	0.04	6,064	0.02	6,000	0.06	6,105
Foreign income in 2011	0.03	1,236	0.02	772	0.04	1,728
Net income in 2011	–	11,820	–	12,096	–	11,307

Notes: Means for monetary variables are calculated for positive observations and presented in euros (in 2011). 'Share' denotes the relative share of recipients within each group. Other pension income includes veterans' pensions, pensions from traffic injuries and accidents, and foreign pension income. Other pension types include early old-age pensions, temporary rehabilitation benefits, unemployment pensions, and farm closure compensation pensions. Guarantee pension refers to observed amount of guarantee pension for recipients in 2011. Medicine reimbursements for mental illnesses include special reimbursements based on a medical diagnosis under the category of mental disorders, such as dementia, psychosis, paranoia and schizophrenia, and medicine prescribed for Alzheimer's disease. Level of education refers to a categorical measure of education level (1=elementary school, 5=graduate or post graduate degree). Net income information is available for 105,555 pensioners and income information by income type for 105,549 pensioners.

from the SII, indicating that the entire sample is accustomed to dealing with the SII. Even though guarantee pension eligibility does not depend on income from other sources (labor, capital or entrepreneurial income), non-pension income is not very common among the sample, except for small amounts of capital income. However, variation within the sample is notable, as there are some individuals with relatively high non-pension income.

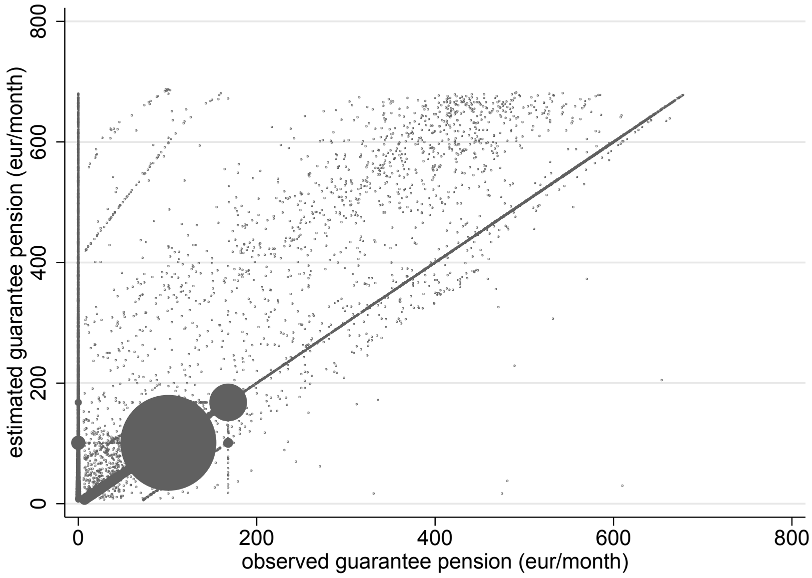
As explained above, the treatment group to whom the January letter was sent was not chosen randomly among the pool of eligible pensioners, but was based on receiving a full national pension. The treatment and control groups are therefore likely to differ in some characteristics. Table 1 shows that the treatment group has more disability pensioners, and the control group more old-age pensioners. This also affects the difference in the average age between the groups. Disability pensioners are also typically male and unmarried, which is visible in the group averages. Unsurprisingly, only 17% of pensioners in the treatment group have positive employment pension income, and the average level is also very low. However, average pension income is similar in both groups, which indicates they are eligible to similar amounts of guarantee pension benefits. There are also no notable differences regarding other sources of income between the treatment and control groups, except perhaps the prominence of small amounts of annual capital income in the control group. Still, it is important that both groups consist of pensioners with low total pension income who receive benefits from the SII.

3.2 Descriptive analysis

In order to analyze the effect of the January letter on take-up, we first need to estimate eligibility for the guarantee pension program among all pensioners. As described above, the eligibility criteria are rather straightforward, as eligibility mainly depends on total pension income, which we can accurately observe from the register data. Figure 2 presents a weighted scatter plot of estimated guarantee pensions (vertical axis) and observed guarantee pensions (horizontal axis). The majority of observations lie on the 45-degree line. The estimated guarantee pension is equal (+/- 10 euros) to the observed amount for 97% of pensioners receiving the guarantee pension. The points at the zero level of actual guarantee pension describe individuals not receiving a guarantee pension but who we estimate to be eligible for it. According to our estimation, incomplete take-up occurs at all levels of the guarantee pension, implying that the level of the benefit does not drive the estimated incomplete take-up.

Figure 3 shows the distribution of observed guarantee pensions and our estimate of the distribution based on the eligibility criteria. The left-hand side of the figure shows that there are two clear spikes in the actual distribution. These mark the guarantee pension received by pensioners with a typical full national pension from the SII (101 euros for single pensioners and 168 euros for cohabiting pensioners). The right-hand side shows that the estimated

Figure 2: Eligibility estimation.



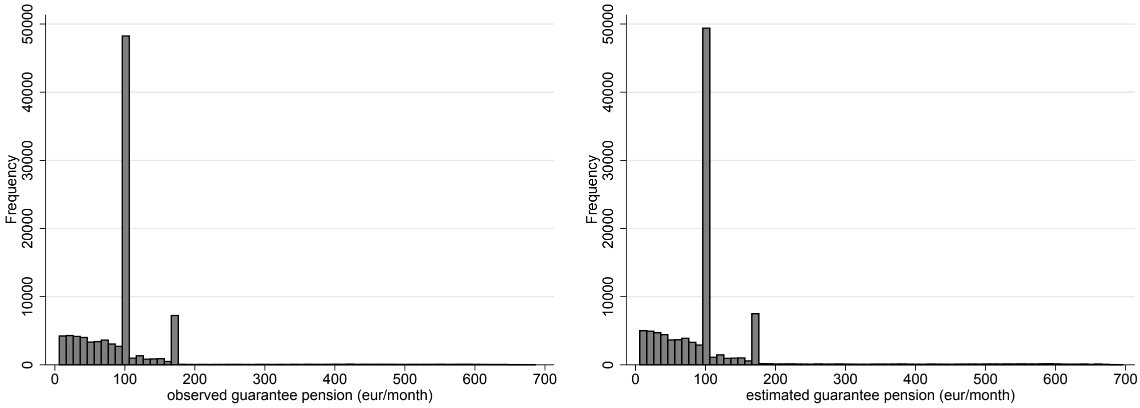
Notes: The figure presents a weighted scatter plot of estimated guarantee pensions (vertical axis) and observed guarantee pensions (horizontal axis). The majority of observations lie on the 45-degree line (line not drawn). The largest concentrations on the 45-degree line correspond to the guarantee pension entitlement of typical recipients of a full national pension (101 euros for single pensioners and 168 euros for cohabiting pensioners).

distribution is very similar to the observed distribution, implying that we can very accurately capture the distribution of guarantee pension benefits.

Our measure of take-up is a dummy indicating whether we observe that a person applied for the guarantee pension by some specific point in time. We measure take-up as having applied, irrespective of the benefit decision of the SII. The number of rejected applications is very small, and thus including rejections does not affect our results in a significant way. Among our sample of 105,574 estimated eligible pensioners, 1,186 applied but were not granted the benefit (438 in the treatment group and 748 in the control group).

Table 2 shows the take-up rates of the guarantee pension program among the eligible population. The overall take-up rate by the end of 2011 is relatively high, as 93% of all eligible individuals had applied for the benefit. The table shows that the take-up rate increased over time during 2011, but was already at 77% in March of 2011 when the benefit was introduced. The average recipient applied for the benefit right after its implementation in week 11 (March 14–18, 2011). The table also reveals clear differences in the behavior of letter recipients and the control group, which we discuss in more detail below.

Figure 3: Actual guarantee pension distribution (left-hand side) and estimated guarantee pension distribution (right-hand side).



Notes: The figure shows the distribution of observed guarantee pensions (left-hand side) and our estimate of the distribution based on the eligibility criteria (right-hand side). The distinctive spikes in both distributions mark the guarantee pension received by pensioners with a full national pension from the SII (101 euros for single pensioners and 168 euros for cohabiting pensioners).

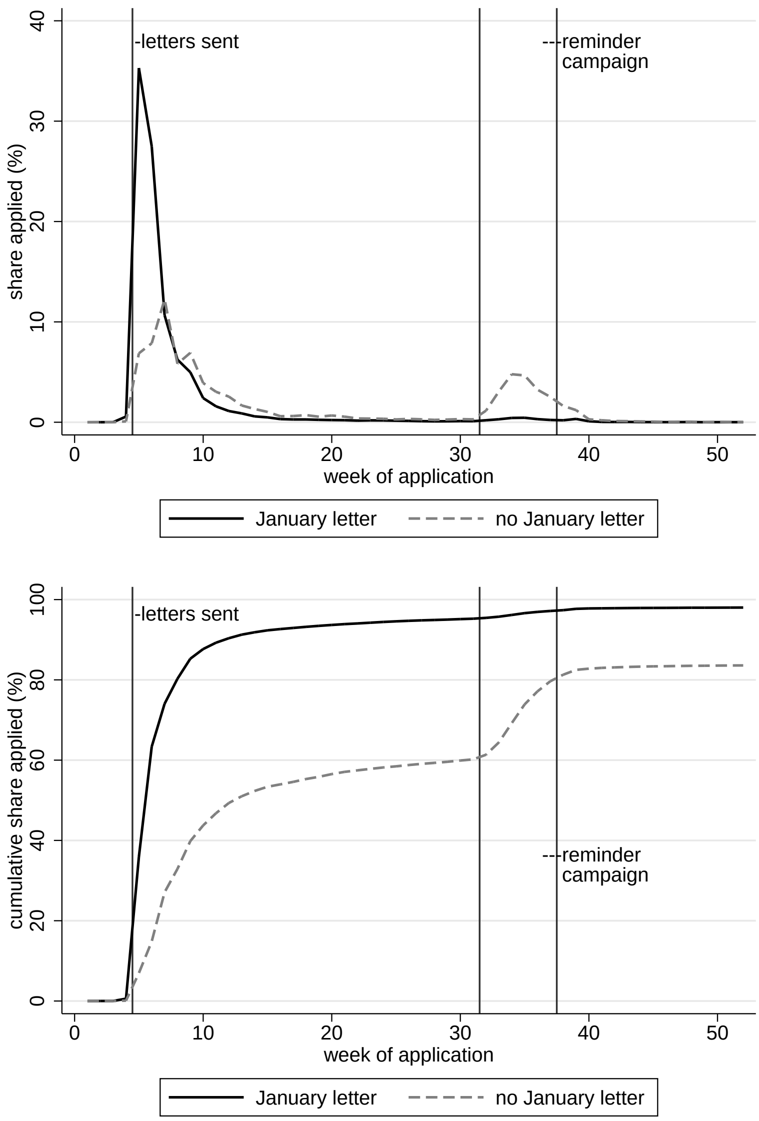
Table 2: Guarantee pension take-up rates.

	Whole sample (N=105,574)	Treatment group (N=68,655)	Control group (N=36,919)
Applied by end of March 2011	0.77	0.91	0.51
Applied by end of July 2011	0.83	0.95	0.60
Applied by end of September 2011	0.92	0.98	0.82
Applied by end of 2011	0.93	0.98	0.84
Average week of application	11 (March 14–18)	8 (February 21–25)	18 (May 2–6)

Figure 4 provides graphical evidence of the effect of the letter on the take-up rate. The upper graph shows the share of treatment and control group pensioners who applied for the benefit in different weeks in 2011. The first vertical line denotes the week when the January letters were sent (week 4, January 24–28). The second (week 32, August 8–12) and third (week 37, September 12–16) lines denote the period in which the SII contacted all remaining potential recipients of the guarantee pension. This includes both those who already received the first letter in January and those who did not.

The graph highlights that a significant share of pensioners applied for the benefit immediately after receiving the letter in late January. Over 30% of pensioners in the treatment group applied for the benefit in week 5, compared to less than 10% in the control group. This illustrates the sharp effectiveness of the January letter. Second, the share of weekly applications was very similar in both groups after week 9 and until the end of July. Within this period,

Figure 4: Above: The share of received applications in different weeks in 2011 in the treatment and control groups (January letter/no letter). Below: Cumulative share of applications in different weeks in the treatment and control groups.



Notes: The figure shows the share of eligible pensioners in the treatment and control groups who applied for the guarantee pensions in different weeks in 2011 (above), and the cumulative share of applications in different weeks (below).

no personal information or application forms were sent to any of the eligible pensioners, and there are no visible differences in take-up intensity between the groups in that period. Third, we observe a clear increase in applications in the control group starting in week 32 when the SII began to contact all remaining pensioners potentially eligible for the benefit who had not applied by then. This again shows that personally provided information on potential eligibility and details on how to apply appear to be very important for take-up. However, we do not observe a sharp increase in applications in the group that received the January letter, even though those who had not applied by then were contacted again in August–September. This finding can be at least partly explained by the fact that most of the January letter recipients had already applied before August.

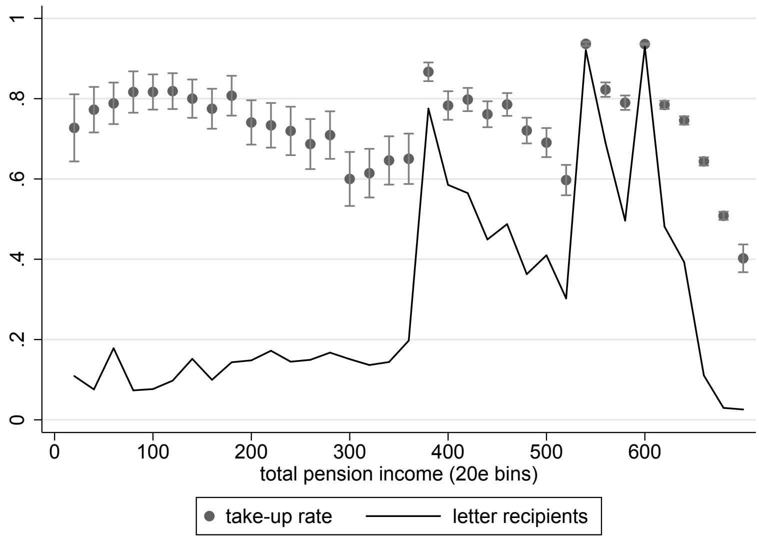
The lower graph shows the cumulative share of applications for the treatment and control groups. The take-up rate in the treatment group rises to over 80% already at the time of implementation, and reaches 91% by the end of March 2011. In comparison, the take-up rate is 51% in the control group by the end of March. This indicates that the letter significantly expedited take-up. In addition, the take-up rate is 98% in the group that received the January letter by the end of 2011, and 84% in the group that did not receive the January letter.

Figure 5 shows the average take-up rates by the end of July 2011 for eligible pensioners in 20 euro bins of monthly total pension income, together with the probability of receiving the January letter in each bin. We observe that the probability of receiving the letter is highest at three specific total pension income levels. This stems from the fact that the letters were sent to pensioners receiving the full national pension, which typically corresponds to specific levels of total pension income.[8] However, there is variation in the full national pension income level and in combinations of full national pension and employment pensions, which translates into variation in the probability of receiving the letter across different total pension income levels.[9] Therefore, we observe a positive average probability of receiving the letter in all total pension income bins in the Figure.

[8]In addition to the spikes in the probability of receiving the letter in the bins containing pension amounts of 520 and 586 euros per month, the third spike at the 380 euro bin reflects a concentration of pensioners who have taken up an early old-age pension, and therefore have a permanently lowered national pension, even though it is notionally full.

[9]Small amounts of employment pension income (below 52 euros per month) do not affect the national pension benefit, but increase total pension income and thus reduce the guarantee pension benefit. Having lived abroad for many years during one's working life can result in a downward-adjusted national pension, even though it is still notionally "full" due to sufficiently small employment pension. Also, the full national pension can be reduced due to taking up early old-age pension. In that case, the full rate is reduced by 0.4% for each month that the pension is brought forward before the age of 65 (and the maximum guarantee pension level is reduced accordingly). The pension remains at this level permanently, also after turning 65.

Figure 5: Average take-up rate (by the end of July) with 95% confidence intervals and the probability of receiving the January letter in 20 euro bins of total pension income.



Notes: The figure shows the average take-up rates of the guarantee pension in different total pension income bins of 20 euros, and the share of eligible pensioners who received the January letter within each bin.

There are clear spikes in the take-up rate exactly where the probability of receiving the letter is higher. This illustrates the significant effect of the letter on take-up. Importantly, we observe that the take-up rates are significantly higher among eligible pensioners who were more likely to receive the letter, compared to pensioners in adjacent bins who have similar pension income but were less likely to receive the letter. Similar results are obtained when plotting the letter and take-up probabilities with respect to national and employment pension incomes separately.

4 Estimation and results

4.1 Estimation model

In our main analysis, we use the following linear probability model

$$Y_i = \alpha + \beta \times letter_i + \delta \mathbf{X_i} + \varepsilon_i \tag{1}$$

where Y_i is a dummy equal to 1 if the person had applied for the guarantee pension. The coefficient β denotes the effect of receiving the letter on take-up probability. We control for the observed differences between the treatment and control groups, as well as other factors that are likely to influence take-up, such as the size of the benefit and education level, in vector $\mathbf{X_i}$, and ε_i is the error term.

We expect the size of the guarantee pension to have a positive effect on take-up probability. However, it is possible that there is a correlation between benefit size and receiving the letter, as those with a full national pension (the target group of the letters) were typically eligible for 101 or 168 euros of guarantee pension. Therefore, we also estimate the equation by adding an interaction term $letter \times benefitsize$ to the model to control for the potential interaction of the letter and the size of the guarantee pension.

We use benefit applications by July 31, 2011 as the baseline dependent variable. This shows the impact on early take-up – six months from the January information letter. In addition to the baseline analysis, we vary the time window by analyzing the effect of the January letter on take-up by the end of March, September and December. The end-of-March estimates highlight the impact of the letter on applying within one month of implementation of the program. The September and December regressions capture the effects of the letter beyond the time when all eligible pensioners who had not applied by the end of July 2011 were eventually contacted, describing the more persistent effects of the January letter.

We also conduct various subsample analyses for those with different reimbursed medicine expenses and medical diagnosis, and for old-age pensioners and disability pensioners. These estimations provide interesting information on the potential differences in responses between these groups, and characterize the potential mechanisms behind the effect of the letter. In addition, these estimations serve as robustness checks on the observed differences between the treatment and control groups.

4.2 Baseline results

Table 3 reports the results for the baseline OLS regression. Receiving the January letter increased the probability of applying for the guarantee pension by approximately 33 percentage points, which implies a 55% increase in take-up relative to the control group. This shows that the letter had an economically and statistically significant effect on take-up. Furthermore, the point estimates for receiving the letter without including other covariates (column (1)) and with different sets of control variables (columns (2)–(3)) are very close to each other. This observation highlights that different observed characteristics between the treatment and control groups do not significantly affect our results, indicating that the effect of the letter is not affected by the fact that the letters were targeted at the recipients of a full national pension.

In columns (4)–(5), the negative coefficient of the interaction term (letter×benefit amount) implies that the impact of the letter is smaller for pensioners entitled to larger benefits. Since larger entitlements typically exhibit higher take-up, the letter could therefore be of less importance at larger benefit sizes. For a typical guarantee pension level (100–170 euros), the effect of the letter in columns (4)–(5) is in line with the average impact estimated in columns (1)–(3).

In addition, we find that receiving housing allowance from the SII increases take-up, but receiving care allowance reduces it. This presumably reflects the poorer income situation of the former group, and the worse health condition of the latter group. However, other health indicators (medical expenses and diagnoses) have practically no impact. We return to the impact of health in more detail in the following subsection.

Columns (1)–(4) of Table 4 summarize the regression results when we estimate the effect of the January letter on take-up by the end of March, July (baseline), September and December of 2011, respectively. In addition, column (5) reports the impact of the letter on take-up by the end of July when we only include pensioners who eventually received the guarantee pension at some point in 2011, thus providing a lower bound for our baseline result. All regressions include the full set of controls, similarly as in column (2) of Table 3 above.

The effect of the letter on take-up is larger by the end of March, compared to the baseline model. This implies that the effect of the January letter was concentrated in the first weeks after the information was provided, as was expected. As discussed above, during August and September 2011 SII district offices contacted all potential eligible pensioners who had not applied for the benefit by then. Columns (3) and (4) show that even conditional on this other personal information provided to all eligible pensioners later on, the January letter still has a significant effect on take-up. Receiving the information letter prior to implementation increased the probability of later take-up by approximately 13 percentage points (16% relative to control group take up). The effect of the January letter on take-up by the end of September is similar to the effect by the end of the year. This indicates that

Table 3: Main results. Dependent variable: application status (0/1) by the end of July 2011.

	(1)	(2)	(3)	(4)	(5)
Letter	**0.352**	**0.326**	**0.326**	**0.385**	**0.385**
	(0.003)***	**(0.003)*****	**(0.003)*****	**(0.005)*****	**(0.005)*****
Benefit amount		0.005	0.006	0.009	0.009
(divided by 50e)		(0.001)***	(0.001)***	(0.001)***	(0.001)***
Letter * Benefit amount				-0.027	-0.027
(divided by 50e)				(0.002)***	(0.002)***
Employment pension		-0.043	-0.042	-0.049	-0.049
		(0.003)***	(0.003)***	(0.003)***	(0.003)***
Other pension income		-0.115	-0.114	-0.118	-0.117
		(0.012)***	(0.012)***	(0.011)***	(0.012)***
Disability pension		0.012	0.011	0.012	0.012
		(0.003)***	(0.003)***	(0.003)***	(0.003)***
Other pension type		-0.020	-0.021	-0.021	-0.022
		(0.004)***	(0.004)***	(0.004)***	(0.004)***
Male		-0.021	-0.020	-0.022	-0.021
		(0.002)***	(0.002)***	(0.002)***	(0.002)***
Age		-0.001	-0.001	-0.001	-0.001
		(0.000)***	(0.000)***	(0.000)***	(0.000)***
Spouse		0.081	0.082	0.091	0.093
		(0.003)***	(0.003)***	(0.003)***	(0.003)***
Pensioners' housing allowance		0.093	0.087	0.090	0.085
		(0.002)***	(0.002)***	(0.002)***	(0.002)***
Pensioners' care allowance		-0.041	-0.043	-0.043	-0.045
		(0.002)***	(0.002)***	(0.002)***	(0.002)***
Basic medicine expenses		0.006	0.005	0.006	0.006
(divided by 1,000e)		(0.001)***	(0.001)***	(0.001)***	(0.001)***
Special medicine expenses		0.001	0.000	0.001	0.001
(divided by 1,000e)		(0.000)***	(0.000)***	(0.000)***	(0.000)***
Any mental illness medicine		0.002	0.002	0.002	0.002
expenses		(0.002)	(0.002)	(0.002)	(0.002)
Net income in 2011		-0.012		-0.012	
(divided by 10,000e)		(0.003)***		(0.003)***	
Earnings in 2011			-0.072		-0.072
(divided by 10,000e)			(0.016)***		(0.016)***
Capital income in 2011			-0.007		-0.007
(divided by 10,000e)			(0.002)***		(0.002)***
Entrepreneurial income in 2011			-0.032		-0.032
(divided by 10,000e)			(0.008)***		(0.008)***
Foreign income in 2011			0.017		0.016
(divided by 10,000e)			(0.012)		(0.012)
Education dummies	no	yes	yes	yes	yes
Constant	0.599	0.639	0.629	0.635	0.625
	(0.003)***	(0.009)***	(0.008)***	(0.009)***	(0.008)***
R2	0.20	0.22	0.23	0.23	0.23
N	105,574	105,555	105,549	105,555	105,549

Notes: Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. For pension income types, the omitted category is national pension. For pension types, the omitted category is old-age pension. Control group take-up rate by the end of July is 60%.

the information provided to all eligible pensioners in August and September had a rapid impact on take-up, and thus reduced the effect of the January letter immediately.

Column (5) shows the result for the subpopulation of eligible pensioners who we observe to have received the guarantee pension at some point in 2011. Thus this group only includes pensioners who are definitely eligible, compared to our baseline analysis in which eligibility status is based on our estimation. Therefore, column (5) represents the lower-bound estimate for the effect of the letter by the end of July, excluding all eligible pensioners who did not apply for the guarantee pension at all in 2011, or whose application was rejected. This estimate, 25 percentage points, is highly significant but somewhat smaller than our baseline estimate, thus further supporting the finding that sending a simple information letter and an application form to eligible individuals can have a large impact on take-up, and in particular prompt eligible individuals to apply for the benefit more quickly.

Table 4: Impact of the January letter on take-up at different points in time.

Dependent variable: Apply by the end of...					
	March	July (baseline)	September	December	July, lower bound
	(1)	(2)	(3)	(4)	(5)
Letter	**0.382**	**0.326**	**0.128**	**0.119**	**0.246**
	(0.004)* **	**(0.003)*** **	**(0.003)*** **	**(0.002)*** **	**(0.003)*** **
Control group take-up rate	51%	60%	82%	84%	72%
R2	0.23	0.22	0.12	0.12	0.18
N	105,555	105,555	105,555	105,555	97,034

Notes: Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Regression specification as in column (2) of Table 3. Column (5) includes only pensioners who received the guarantee pension at some point in 2011, and replaces the estimated guarantee pension amount in the regression by the actual observed amount.

One potential issue in interpreting the results above is that information on eligibility could spread within various networks. This would imply that the effect of the letter is downwards-biased if the information in the letter spilled over to the control group. One feasible example of such a spillover is information spreading within a household when one of the (eligible) spouses received the letter, but the other did not. However, there are only very few cases (350) in our sample where only one of the two eligible spouses received the letter, implying that this direct spillover channel is not likely to significantly contribute to the average effect of the letter.

Finally, it is possible that pensioners receiving other means-tested benefits would be less willing to apply for the guarantee pension if they fear that applying will just reduce their other benefits by the same amount. However, in the case of pensioners' housing allowance, only 40% of income above a certain income threshold affects the benefit, so pensioners receiving this benefit would therefore "lose" only 40% of their guarantee pension income to lower housing benefits if the housing allowance thresholds are exceeded.[10] We take into account the potential confounding effect of the housing allowance by controlling for housing allowance recipient status in the regression. In the case of social assistance, all other social benefits are considered primary, and need to be claimed before social assistance can be granted. Therefore, pensioners relying on social assistance need to apply for the guarantee pension in order to receive social assistance (if they are still eligible for it). Additionally, average annual net income is very similar in both treatment and control groups, indicating that we have no clear reason to assume differences in social assistance eligibility between the groups. Therefore, we conclude that interactions with other income transfers or benefits are not likely to significantly affect the take-up rate or the impact of the letter.

4.3 Subgroup analysis and potential mechanisms

We study the potential mechanisms closer by comparing the impactfulness of the letter among different subgroups. Table 5 summarizes the results for different subsamples of eligible pensioners. The table shows the regression results for the baseline setup and for the lower bound sample (including only individuals who applied for and received the benefit at some point in 2011). The dependent variable and the control group take-up rate are defined using the application status by the end of July 2011, and all regressions include the full set of controls.

First, the table shows the results separately for old-age pensioners, disability pensioners, and those with positive or zero employment pension income. These characteristics – pension type and whether or not the pensioner received employment pension – are also the most prominent differences between the treatment and control groups. The results show that the letter appears to be more effective among old-age pensioners compared to disability pensioners, but the effect is nevertheless clearly significant in both groups.The disability pensioners - who are younger - claim the benefit more actively, which is re-

[10]In 2011, the income thresholds for the full housing allowance were 8,091 euros per year for pensioners without a spouse, and 11,860 euros and 12,996 euros per year for cohabiting pensioners depending on whether the spouse is also eligible for housing allowance or not. The size of the allowance is affected by place of residence, type of housing, and a variety of characteristics of the residence such as the heating system and construction year.

flected by the higher take-up rate in the control group (66% vs. 59% among old-age pensioners). This indicates that disability pensioners were more active in applying even without receiving the letter.

We find that eligible pensioners with (small) positive employment pension income respond to the letter more actively than those with no employment pensions. One explanation for this finding is that pensioners with employment pensions were not as certain about their eligibility as those with no employment pension. Thus the effect of the letter stating potential eligibility for the benefit had a larger effect among those who were likely to be less certain about their eligibility.

Deteriorated health can have an effect on benefit take-up. In addition, it could negatively impact the effectiveness of information provision in increasing take-up. We utilize the SII's medical register data, including both total reimbursed expenses for severe and long-term illnesses and diagnostic information, to provide new evidence on these effects. We detect no differences in the effectiveness of the letter between individuals with positive or no reimbursed medicine expenses for severe illnesses. This indicates that deteriorated health is not a driving factor behind information effects or take-up behavior, at least among pensioners with medicine for their illnesses.

In addition to total reimbursements, we focus on reimbursed medicine expenses for mental illnesses.[11] These types of illnesses could affect cognitive ability to apply for the benefit and to understand the eligibility rules and provided information. Nevertheless, we find no differences in responses among those with medicine expenses for diagnosed mental illnesses compared to those without such a diagnosis. This suggests that mental illnesses do not confound the effect of the letter, at least within the population with prescribed medicine.

Furthermore, we study pensioners receiving pensioners' care allowance. Eligibility for the care allowance typically requires a severe long-term illness or disability, such that the recipient requires constant assistance in normal daily activities such as eating, dressing, or taking medications. We find that the effect of the letter is larger for those receiving the care allowance compared to those without it, indicating that more severely ill or disabled pensioners respond to the letter more actively.[12] This finding also suggests that infor-

[11] Medicine reimbursements for mental illnesses include special reimbursements based on a medical diagnosis under the category of mental disorders, such as dementia, psychosis, paranoia and schizophrenia, and medicine prescribed for Alzheimer's disease.

[12] We also find that the impact is greater the higher the level of care allowance (results not reported in the table).

mation on eligibility is more efficient when the take-up process is potentially assisted or managed by someone else than the recipient of the letter, such as close relatives, which is more likely among those with severe illnesses or disabilities. Nevertheless, the effect of the letter is not limited to those with care allowance, as the letter also significantly affected take-up for those not receiving it. Also, as an additional result not shown in the table, we find no differences in responses within the population of care allowance recipients between those with or without prescribed medicine for mental illnesses, indicating that these types of severe illnesses do not drive the results.

In addition to information on eligibility requirements and potential eligibility, the January letter included an application form and a postage-paid return envelope. Therefore, in addition to information, the letter presumably reduced the costs of applying. As there is no variation in the content of the mailing within the treatment group, we cannot distinguish a causal difference between the effects of the different components of the mailing. Thus the average effect of the letter needs to be interpreted as containing both of these channels.

In earlier literature, Bhargava and Manoli (2015) find that providing information is more effective in increasing take-up than reducing claiming costs among the EITC eligibles in the US. Our subgroup results provide suggestive evidence that both of these channels are important for low-income pensioners. As mentioned above, information on eligibility could explain the larger response among those with positive employment pensions, who are presumably less certain about their eligibility in the absence of information compared to those with only national pensions. Also, a part of the larger response to the mailing among care allowance recipients could be due to reduced costs of applying. For these severely ill or disabled pensioners, the transaction costs are presumably larger than for those with a better ability to, for example, visit the SII field office in order to apply for the benefit. Nevertheless, it is important to note that the costs of applying for the guarantee pension are extremely low even when not receiving a personal mailing, as the benefit can be applied even with a telephone call to the SII service number. Therefore, providing information on eligibility and the application procedure is likely to increase take-up even without providing the application form and the return envelope. Finally, from the point of view of practical policy, it is relatively inexpensive and straightforward to include both the application form and information on eligibility when sending letters to the eligible population. Therefore, knowl-

Table 5: Results for different subgroups of eligible pensioners (dependent variable: application status (0/1) by July 2011).

	Old-age pensioners		Disability pensioners	
	Share received letters: 45 %		Share received letters: 82 %	
	Average take-up rate: 75 %		Average take-up rate: 90 %	
	baseline	lower bound	baseline	lower bound
Letter	**0.372**	**0.293**	**0.267**	**0.221**
	(0.004)***	**(0.004)*****	**(0.006)*****	**(0.005)*****
Control group take-up	59%	69%	66%	75%
R2	0.22	0.19	0.16	0.14
N	40,111	36,017	53,346	51,051
	Employment pension > 0		No employment pension	
	Share received letters: 26 %		Share received letters: 94 %	
	Average take-up rate: 68 %		Average take-up rate: 94 %	
	baseline	lower bound	baseline	lower bound
Letter	**0.371**	**0.255**	**0.238**	**0.142**
	(0.004)***	**(0.004)*****	**(0.011)*****	**(0.010)*****
Control group take-up	58%	70%	76%	90%
R2	0.16	0.13	0.06	0.03
N	45,346	38,789	60,209	58,245
	Special medicine expenses > 0		No special medicine expenses	
	Share received letters: 66 %		Share received letters: 64 %	
	Average take-up rate: 84 %		Average take-up rate: 81 %	
	baseline	lower bound	baseline	lower bound
Letter	**0.331**	**0.250**	**0.318**	**0.240**
	(0.004)***	**(0.004)*****	**(0.005)*****	**(0.005)*****
Control group take-up	61%	71%	59%	73%
R2	0.23	0.19	0.22	0.16
N	60,140	56,167	45,415	40,867
	Mental illness medicine expenses > 0		No mental illness medicine expenses	
	Share received letters: 82 %		Share received letters: 60 %	
	Average take-up rate: 89 %		Average take-up rate: 81 %	
	baseline	lower bound	baseline	lower bound
Letter	**0.324**	**0.261**	**0.326**	**0.244**
	(0.008)***	**(0.007)*****	**(0.004)*****	**(0.003)*****
Control group take-up	59%	67%	60%	72%
R2	0.25	0.21	0.21	0.17
N	25,121	24,156	80,434	72,878
	Care allowance > 0		No care allowance	
	Share received letters: 83 %		Share received letters: 51 %	
	Average take-up rate: 87 %		Average take-up rate: 79 %	
	baseline	lower bound	baseline	lower bound
Letter	**0.433**	**0.343**	**0.283**	**0.212**
	(0.007)***	**(0.006)*****	**(0.004)*****	**(0.003)*****
Control group take-up	49%	59%	63%	75%
R2	0.29	0.26	0.19	0.14
N	46,464	44,243	59,091	52,791

Notes: Standard errors in parenthesis. *** p<0.01, ** p<0.05, *p<0.1. Baseline regression specification as in column (2) of Table 3. Lower-bound regression specification includes only pensioners receiving the guarantee pension at some point during 2011, and replaces the estimated guarantee pension amount in the regression by the actual observed amount. Disability pensioners are restricted to those not receiving old-age pension at the same time, and vice versa. Displayed take-up rates are measured by the end of July 2011.

edge on the combined effect of these channels is of key importance when considering the effectiveness of practical means to affect the take-up rate.

4.4 Media coverage

The guarantee pension program received plenty of attention in the media, particularly in the first half of 2011. Most of the media coverage concentrated at the time of implementation around March 2011. The role of the SII was also active, as they sent out several press releases over the course of the year, and published news items in the SII's customer magazine and website. Overall, high visibility in the media implies that general awareness of the program was presumably high. However, it remains an open question whether media coverage had an effect on guarantee pension take-up, which was probably at least one of the goals of the SII's active media campaign. We study this by analyzing the effect of a SII press release on the number of guarantee pension applications.

We focus on the press release issued on May 10, 2011. With this press release, the SII informed that many eligible pensioners had not applied for the benefit, and offered simple information on applying. The headline of the release highlighted that the SII was still expecting at least 20,000 eligible pensioners to apply for the guarantee pension. In addition, the release included information on how to apply for the benefit, and what the eligibility criteria were. The press release was published in most of the largest regional newspapers and covered by the Finnish Broadcasting Company on the day it was issued.[13]

Previous press releases were concentrated around the time of implementation (late January-end of March), and it is thus difficult to separate the effects of this media coverage from other factors potentially affecting take-up at the time of implementation, including the January letter. Therefore, analyzing the May 10 press release allows us to better isolate the potential effect of media coverage. Our own media survey reveals that previous news items on the guarantee pension appeared more than three weeks before the May 10 press release, and the following ones appeared four weeks later. This suggests that no confounding coverage was taking place close to the May 10 press release.

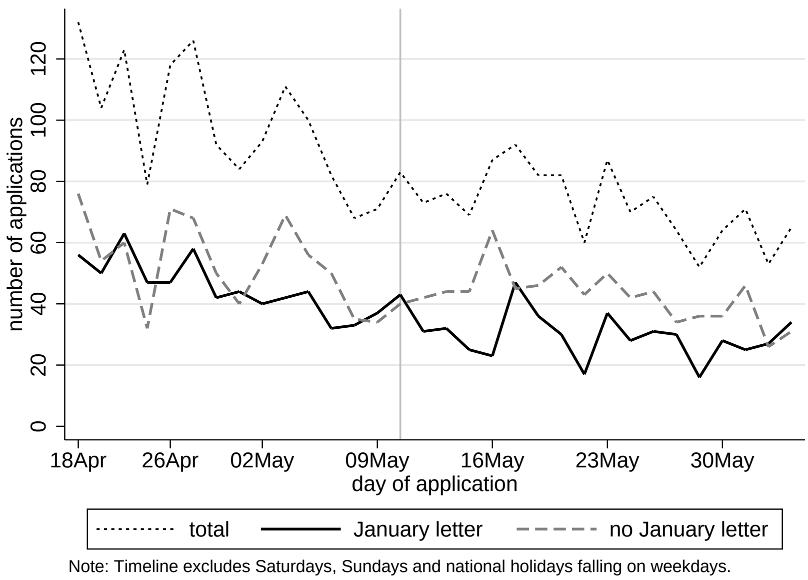
Figure 6 illustrates the number of daily guarantee pension applications around the time of the press release, denoted by the vertical line. There were

[13]Regional newspapers include e.g. Turun Sanomat, Kaleva, Etelä-Suomen Sanomat and Savon Sanomat.

23,264 eligible pensioners in our sample not having claimed by April 18, which is the first day in the figure. The date labels displayed on the horizontal axis indicate Mondays. The figure displays total daily applications, and the number of applications for the treatment and control groups, excluding weekends and bank holidays, when no applications are processed.

The figure shows that there is no jump in the number of overall applications after the press release. This suggests that publicly available information is significantly less successful than personally provided letters in affecting take-up. Supporting this visual observation, we find that the coefficient for the effect of the press release on the number of daily applications is very small, 0.32 (standard error 0.45), and not significantly different from zero. There is potentially a tiny increase in applications among the group that did not receive the January letter, compared to the group that did receive it. However, the difference in the number of daily applications between the groups before and after the press release, 0.45 (0.24), is small and not statistically significantly different from zero at 5% level.[14]

Figure 6: The number of applications per day around the SII press release on May 10, 2011.



[14]In the first regression model, the number of daily applications is regressed on the application date, day of the week, the press release dummy (0 before May 10, and 1 afterwards), and the interaction term of the press release dummy and the application date using a time interval of one month before and after the May 10 press release. For the difference between the control and treatment groups, the dependent variable is the difference in daily applications between the groups. The coefficient of the interaction term measures the effectiveness of the press release in increasing the number of applications.

Overall, these findings suggest that the impact of the media campaign on take-up is trivial, particularly when compared to the notable effect of personally provided information. This finding is underlined when comparing the large increase in applications in the control group (no January letter) in August and September, when all eligible pensioners who had not applied by then were contacted personally (see Figure 4 above). Importantly, those pensioners who responded to this second round of personal information provision had not yet applied at the time of the May 10 press release. Therefore, the low effectiveness of press coverage, particularly in the control group, cannot be explained by individuals more prone to information having already applied by the time of this press release.

5 Concluding remarks

We study the impact of an inexpensive, targeted information treatment on the take-up rate of a social benefit targeted at low-income pensioners outside the labor force. Our results clearly indicate that sending information on eligibility and an application form directly to eligible individuals increased take-up and also prompted eligible persons to apply more quickly. We find that press coverage does not affect the take-up rate in a significant manner, implying that personal mailings are much more effective in increasing take-up.

The letter had a significant effect on take-up within various subgroups. However, the results suggest that directed information is more effective, for example, among elderly individuals and in cases where uncertainty about eligibility is potentially larger (pensioners with positive employment pensions). In addition, we find that the effect of the letter is larger for those severely ill or disabled pensioners who require constant assistance in daily activities, suggesting that the letter is more effective when the application process is assisted or managed by someone other than the recipient of the letter, such as close relatives. In contrast, we find no differences in responses between pensioners with different levels of medicine reimbursements, indicating that deteriorated health does not affect the effectiveness of the letter, at least for those with medicine for their illnesses.

Our setting also relates to the take-up of a newly introduced benefit which is not yet well known among the eligible population. In such a situation there can be more confusion about eligibility rules, and information provision can be critical in order to reach eligible persons from the very beginning. In addition to increasing the overall take-up rate, the timing of applications can also be important. In cases where benefits cannot be applied for retrospectively, eligible individuals can lose part of the benefit flow they are entitled to if they do not apply in time. This is an issue for both existing benefit programs as well as new programs.

Finally, the low costs of information treatments make providing targeted information an attractive instrument for increasing the effectiveness of social policy. However, it should be borne in mind that an important requirement for the usability of information treatments is that the eligibility of individuals can be assessed easily. This is not the case for some of the more discretionary social benefits in Finland and elsewhere. Sending information to ineligible individuals can result in increased flows of applications that end up being rejected, increasing the workload of officials and creating negative publicity for social policy and the institutions implementing it, and even further increasing confusion about social benefit policies among low-income individuals.

Bibliography

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu, "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment," *Quarterly Journal of Economics*, 2012, *127* (3), 1205–1242.

Bhargava, Saurabh and Dayanand Manoli, "Psychological frictions and the incomplete take-up of social benefits: evidence from an IRS experiment," *American Economic Review*, 2015, *105* (11), 3489–3529.

Chetty, Raj and Emmanuel Saez, "Teaching the tax code: Earnings responses to an experiment with EITC recipients," *American Economic Journal: Applied Economics*, 2013, *5* (1), 1–31.

_ **, John N. Friedman, and Emmanuel Saez**, "Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings," *American Economic Review*, 2013, *103* (7), 2683–2721.

Currie, Janet, "The take-up of social benefits," in A. Auerbach, D. Card, and J.M. Quigley, eds., *Public policy and the income distribution*, Russell Sage Foundation, 2006, pp. 80–148.

Hertel-Fernandez, Alexander and Jeffrey B. Wenger, "Taking up social benefits: A cautionary tale from an unemployment insurance survey experiment," 2013. Mimeo. Available at www.ssrn.com/abstract=2341885.

Kleven, Henrik and Wojciech Kopczuk, "Transfer program complexity and the take-up of social benefits," *American Economic Journal: Economic Policy*, 2011, *3* (1), 54–90.

Liebman, Jeffrey B. and Erzo F.P. Luttmer, "Would people behave differently if they better understood social security? Evidence from a field experiment," *American Economic Journal: Economic Policy*, 2015, *7* (1), 275–299.

_ **and Richard J. Zeckhauser**, "Schmeduling," 2004. Mimeo.

Mastrobuoni, Giovanni, "The role of information for retirement behavior: Evidence based on the stepwise introduction of the Social Security Statement," *Journal of Public Economics*, 2011, *95* (7-8), 913–925.

Zantomio, Francesca, "The route to take-up: Evidence from the UK pension credit reform," *Oxford Bulletin of Economics and Statistics*, 2015, *77* (5), 719–739.

A Appendix

THE SOCIAL INSURANCE INSTITUTION OF FINLAND

P.O.Box 78
00381 Helsinki

Mailed on

27.1.2011

Date of birth: xxxxxx

APPLYING FOR GUARANTEE PENSION

Dear pensioner

The Act on guarantee pensions comes into force on March 1st, 2011. You might be entitled to a guarantee pension. A guarantee pension can be granted only if an application is made. For that purpose, we are sending you a pre-populated application form and a return envelope.

Pensioners whose national pension and other pensions before taxes amount to no more than 687.74 e/month can receive guarantee pension benefits. We ask you to clarify in the application whether you have any other pension income or compensation, from Finland or other countries, in addition to your national pension. The care allowance for pensioners, veterans' supplements, child supplements or pensioners' housing allowance are not considered pension income.

The attached application form is pre-populated with the details of the bank account to which your national pension is paid. If you wish your guarantee pension benefits to be paid to another account, please state the correct account number on the form.

The guarantee pension has a retroactive application period of six months. If you wish to get a guarantee pension starting on March 1st, 2011, you need to apply for it in September 2011, at the latest.

For additional information, call 020 692 352 (mon-fri 8am-6pm). Regular land line or mobile phone charges apply. You can also find more information about the guarantee pension from the SII's offices and web site www.kela.fi/takuuelake.

Sincerely,

THE SOCIAL INSURANCE INSTITUTION OF FINLAND

Attachments
Guarantee pension application
Return envelope

Kela|Fpa

Mailing address
P.O.Box 78
00381 Helsinki

www.kela.fi

Enquiries 020 634 11

Figure A.1: January information letter template (translated from Finnish)

Kela|Fpa

Application **GE 1**
Guarantee pension

ⓘ The guarantee pension can only be granted once you have applied for all pensions to which you are entitled, in Finland and abroad.

Application period: You can apply for the guarantee pension for a maximum of six months retroactively.

1. The applicant

Social security number Name

Phone number Email address

ⓘ The SII gets home address information from the population register.

2. Bank account

____

3. Application

I am applying for
□ guarantee pension
□ a revision to guarantee pension
□ a temporary revision guarantee pension starting on __.__.____

4. Pensions and remunerations

ⓘ You don't have to report pension income from SII or Finnish employee pension institutions.

Are you currently receiving or applying for pension or remuneration from Finland or abroad? (e.g. voluntary supplementary pension, remuneration based on an accident)

□ No

□ Yes, I am receiving; Please specify type of remuneration, starting date, country and granting institution. Please attach a decision letter.

GE 2 03.14 / CR www.kela.fi 1 (2)

□ Yes, I am applying for; Please specify type of remuneration, country and granting institution.

5. Attachments

ⓘ You can also send attachments online.

Part 4. Pensions and remunerations:

□ Pension income paid from abroad: decision letter, receipt, or other proof specifying the current gross amount per month.

6. Additional information

____

____

____

□ Additional information on a separate sheet. Please mark your name and social security number on the sheet.

7. Signature

I consent to the bank returning any unduly paid pensions to the institution that paid them. This consent applies only to such pension instalments paid unduly after the death of the pensioner.

I assure the information given here is correct and I will notify the SII should the information change.

Date Signature

If the signatory is not the same as the applicant, please specify the reason for this.

8. Person assisting with the application

Name and phone number

We might use information provided in the context of this benefit decision in other benefit decisions as well, if the law so requires. We might also use information from other benefit decisions in ruling for this benefit. The SII can provide you with a description of where we can get information about you, and where we can pass on such information.

GE 2 03.14 / CR www.kela.fi 2 (2)

Figure A.2: Application form (translated from Finnish)

Using a kinked policy rule to estimate the effect of experience rating on disability inflow*

Abstract

We study whether the experience rating of employers' disability insurance premiums affects the inflow to disability benefits in Finland. To identify the causal effect of experience rating, we exploit "kinks" in the rule that specifies the degree of experience rating as a function of firm size. Using panel data on all firms and workers in the private sector, we estimate the effect of experience rating on the inflow to sickness and disability benefits. We find no evidence that experience-rated firms would react to their incentives to prevent new disability benefit claims.

Keywords: Experience rating, disability insurance, early retirement
JEL classification codes: J14, J26,H32

*This paper is joint with Tomi Kyyrä (VATT Institute for Economic Research and IZA). Parts of this research have been published in Finnish in the report Kyyrä and Paukkeri (2015).

1 Introduction

In many countries disability benefit costs are increasing rapidly and reforming disability programs is high on the policy agenda. While several studies have analyzed the effects of disability benefits or eligibility criteria,[1] the role of employers and their incentives has attracted little attention, even though the employers may play an important role as well. As a preventive action, the employer can invest in workplace health and safety, and allocate the workload evenly between its workers in an attempt to reduce the onset of health problems at the workplace. And when a worker anyway develops a medical condition that reduces his or her working capacity, the employer has the discretion of whether to provide physical aid or retraining, and whether to modify job assignments in order to keep the worker at work. The problem is that the employer's incentives to implement disability reducing measures can be weak even when the costs of such measures to the employer are considerably less than the costs of a new disability benefit recipient to the society.

Experience rating of disability insurance (DI) premiums may help to mitigate the incentive problem. With experience rating, the employer's premium is adjusted to reflect the costs of its workers' disability benefit claims in comparison to other employers. Employers with high disability costs are penalized through a surcharge on top of the base premium, while employers with low disability costs are rewarded by giving a discount on the base premium. If successful, experience rating helps employers to internalize the societal costs of disability benefit claims and encourages them to implement cost-effective disability reducing measures, resulting in a lower disability inflow rate. In this study, we quantify the effect of experience rating on the disability inflow using data from Finland.

Although experience rating is widely used in other forms of social insurance – such as in workers' compensation and unemployment insurance – it is still rare in DI. To the best of our knowledge, DI premiums are currently experience rated only in the Netherlands and Finland. Yet experience rating has attracted considerable interest also in other countries where policymakers are considering how best to reform their DI systems to curb growth in disability caseloads. For instance, Autor (2011) and Burkhauser and Daly (2011) have proposed that the U.S. Social Security DI program should be fi-

[1]This literature includes Gruber (2000), Black, Daniel and Sanders (2002), Campolieti (2004), Autor and Duggan (2003; 2006), Karlström, Palme and Svensson (2008), Staubli (2011) and Kyyrä (2014).

nanced by an experience-rated payroll tax. They motivate their proposal with declining disability benefit enrollment in the Netherlands, which introduced experience-rated DI premiums in the late 1990s. However, due to the number of simultaneous changes confounding the effects of individual policy measures, it is not clear to what extent experience rating has reduced the disability rate in the Netherlands. Findings of Koning (2009) and van Sonsbeek and Gradus (2013) do imply that the adoption of experience rating has played an important role, but these studies use data only from post-reform years and lack a comparison group that would not have been subject to experience rating. Kyyrä and Tuomala (2013) find no effects for experience rating in Finland, but their analysis covers only a small and specific group of workers. Overall, evidence regarding the efficiency of experience rating in DI is scarce and inconclusive. Several studies have examined the effects of experience rating in U.S. and Canadian workers' compensation programs, which cover the medical cost of work-related injuries and cash payments to injured workers. But findings from this literature are suggestive only, given that employers have less control over general disabilities than workplace injuries and illnesses. By analyzing the Finnish case, we can contribute to understanding better the efficiency of experience rating in the context of DI.

In Finland, firms are subject to various degrees of experience rating depending on their size. The smallest firms are not subject to experience rating at all. The largest firms are fully experience-rated, and among the medium-sized firms the degree of experience rating increases linearly from 0 to 1 with firm size. To identify the causal effect of experience rating on the disability inflow we exploit these discontinuities or "kinks" in the experience rating rule at the threshold values for small and large firms using a regression kink design (see Nielsen, Sørensen and Taber (2010) and Card et al. (2015)). Our analysis is based on comprehensive matched employer-employee data that cover all private-sector firms and their employees over the period 2007–2013. We analyze the inflow to sick leave, which typically precedes receipt of a disability benefit, and the inflows to different types of disability benefits (temporary vs. indefinite duration, and partial vs. full benefit), which all affect the employer's DI premium differently. Since our data include medical diagnoses for those who were awarded a disability benefit, we can also analyze the disability inflow by main diagnosis category.

In the first step, we construct firm-year disability risk measures that are adjusted for differences in the characteristics of the workforce across firms and over time. In the second step, we examine to what extent differences in these adjusted disability risks can be explained by differences in the degree of experience rating. Our descriptive analysis shows that the disability risks vary little by firm size. In the regression analysis, we find no robust evidence that experience rating would affect any of our disability-related outcomes. Thus, in the light of our analysis, the efficiency of experience rating in DI as a disability prevention device seems questionable.

The remainder of the paper proceeds as follows. In Section 2 we review the existing evidence on the effects of experience rating in DI and workers' compensation. In Section 3 we describe Finnish sickness and disability benefit schemes, explain how the experience-rated DI premiums are determined, and discuss the financial incentives they impose on employers. In Section 4 we describe our data and report some descriptive statistics. In Sections 5 and 6 we discuss the statistical method used and provide empirical evidence that the underlying assumptions of the method are likely to hold in our case. The main results and robustness checks are reported in Sections 7 and 8, respectively. Section 9 contains concluding remarks.

2 A review of experience rating literature

In the Netherlands, experience-rated DI premiums were introduced in 1998. This change applied to all firms, and it was part of a series of disability program reforms implemented over the past two decades. Following these reforms, both the disability inflow and the share of the Dutch population on disability benefits have declined considerably (see e.g. García-Gómez, Gaudecker and Lindeboom (2011) and Koning and Lindeboom (2015)). Koning (2009) exploits variation in the DI premiums triggered by past changes in the disability benefit claims made by the firm's own employees, and finds that disability inflow decreased in those firms that experienced a premium change, compared to the firms with unchanged premiums. He interprets this as evidence that employers were not completely aware of experience rating and therefore the premium change served as a "wake-up call", which induced preventative measures that reduced the disability events in subsequent years. Using quarterly data, van Sonsbeek and Gradus (2013) regress the aggregate disability inflow rate against a set of policy-relevant variables, including the gradually increasing degree of experience rating, and find that experience rating has reduced the disability inflow by 13%. This conclusion however hinges on the assumption that their business cycle proxy (the unemployment rate or a business cycle indicator based on unemployment, and producer and consumer confidence) is a sufficient control for the time trend in the disability inflow. This is a strong assumption because it is not obvious that the disability inflow and business cycle have identical trends. de Groot and Koning (2016) estimate that the removal of experience rating from part of the firms in 2003 reduced disability inflow by 7% among these firms, slightly less than the estimate of van Sonsbeek and Gradus (2013). However, due to confounding reforms they can only use 2003–2004 as the treatment period. It is unclear how strong behavioral impacts we can expect to see in such a short time period as disability prevention actions potentially have long-lasting impacts. The number of reforms in the early 2000s also poses challenges for the identification of the impact of any single reform.

In Finland, firms have been partially responsible for the disability benefit costs of their employees since the 1960s. Until 2005, the system was based on lump-sum contributions. Firms employing more than 50 workers (300 workers before 1996) were required to pay a given share of the present value of a new disability benefit claim as a lump sum payment to the insurance provider at the time when the benefit was awarded to their former employee. Medium-sized firms paid only a small share of this present value, but large ones paid the full amount. Korkeamäki and Kyyrä (2012) exploit the 1996 change in the relationship between the cost share and firm size for identification, and find that the disability cost liability reduced transitions into sickness benefits and further transitions from sickness benefits to disability benefits. The former effect implies that a higher share of disability benefit costs encouraged the employers to invest in preventive measures, whereas the latter suggests that the greater cost share also motivated the employers to make accommodations for their workers with health problems.

In 2006, as a consequence of the adoption of the International Financial Reporting Standards in Finland, which was required by the European Union, the lump sum liabilities were abolished and replaced with an experience-rated payroll tax. Although the new experience rating system was designed to closely mimic the incentive structure of the lump sum payment system in terms of average costs and the allocation of costs across individual employers, it is not obvious that the desired effects of the lump sum liabilities documented in Korkeamäki and Kyyrä (2012) did transfer into the new experience rating system. Compared to the experience-rated premiums, the lump sum liabilities were more transparent from the employer's viewpoint because the costs of a new disability benefit claim realized immediately and were directly attributable to the disability of a given worker.

In the present study, we are interested in the causal effect of experience rating compared to the counterfactual case of flat-rate premiums, which is the status quo in most countries. For this purpose the 2006 reform is not useful for identification because it only converted the lump sum contributions into experience-rated contributions. However, a large pension reform one year later provides some useful quasi-experimental variation in the degree of experience rating for the employers of a particular worker group. The reform in 2007 unified the major pension Acts in the private sector, which coincidentally extended experience rating to cover a certain group of workers that was not subject to the lump sum liabilities before 2006 (workers who were insured under the Temporary Employees' Pension Act). As a result of the pension reform, large employers became liable for the costs of disability benefit claims made by this group for the first time in 2007 through experience-rated DI premiums. Kyyrä and Tuomala (2013) exploit this variation, and find no evidence that the introduction of experience rating would have affected disability outcomes within the affected group. However, the treatment group in their difference-in-differences analysis was small and rather specific, including only manual workers in construction, forestry, agriculture and dock work, who typically work on a temporary basis, and therefore it is not clear to what extent the results can be generalized to other worker groups and their employers. In this study we consider all other worker groups that account for over 80% of private-sector employment. The employers may have more opportunities to affect the health and working conditions of such workers, and therefore the incentives should be clearer for them as well.

The effects of experience rating have been more extensively studied in the context of other forms of social insurance than DI. The studies most relevant for our analysis are those that have evaluated the effects of experience rating in workers' compensation (WC) insurance in the United States and Canada, which provides coverage for employees' medical costs and wage losses resulting from on-the-job injuries. The WC premium is determined as a weighted average of a base rate, which is a mixture of industry and occupation rates, and the firm's incurred loss rate. The weight of the firm's incurred loss rate rises with firm size, as in the Finnish DI system. Bruce and Atkins (1993) find that the fatality rate went down significantly in the forestry and construction industries after experience rating was introduced in those sectors in the province of Ontario in 1984. Ruser (1985; 1991) exploits variation in benefit levels across U.S. states and finds that higher benefits increase benefit claims, but this effect is much smaller in larger firms that are subject to a higher degree of experience rating. This implies that greater experience rating leads to higher investments in workplace safety in response to benefit increases. As another example, by comparing the injury duration of employees of self-insured firms (fully experience-rated) and privately insured firms (imperfectly experience-rated) in Minnesota, Krueger (1990) finds that workers return to work after an injury more quickly if their employer bears the full cost of WC claims instead of being only partially experience-rated. It is noteworthy that in addition to positive impacts on workplace safety, unwanted behavior has also been documented. For example, Thomason and Pozzebon (2002) find evidence that experience rating induces "claims management", where firms attempt to reduce their WC costs using legal measures, such as disputing workers' claims for benefits, rather than by investing in proactive health and safety measures.

To sum up, several studies have found that experience rating in WC reduces on-the-job injuries and the duration of injury spells, and that most of these effects are likely to be due to actual improvements in workplace safety, not just due to benefit claims suppression by employers. However, evidence on the incentive effects of experience rating in DI is much more limited and inconclusive.

3 Institutional framework

3.1 Sickness and disability benefits

When a worker falls ill and receives a doctor's statement certifying that he or she is not capable of work, he or she is entitled to a compensation for wage loss. For the first weeks (typically one to three months depending on the collective agreement), the worker is fully compensated and receives payment from the employer, after which he or she can claim a *sickness benefit* from the Social Insurance Institution.[2] The sickness benefit can be received for a maximum of about one year (300 working days, Saturdays included).

[2]For part of the fully compensated period that exceeds 9 working days, the Social Security Institution pays the sickness benefit to the employer, so the employer's direct cost for this period is the difference between the wage rate and sickness benefit.

Depending on the medical condition, the applicant's rehabilitation needs are assessed in a more extensive medical examination during the sickness benefit period. In case of prolonged disability, the individual may qualify for one of four possible disability benefits: (*i*) a *partial rehabilitation benefit*, (*ii*) a *full rehabilitation benefit*, (*iii*) a *partial disability pension*, or (*iv*) a *full disability pension*. When it is probable that the applicant will return to work, he or she is awarded a rehabilitation benefit for a specific period (previously known as a temporary disability pension) provided that a rehabilitation plan has been drafted. If the return to work is unlikely, the applicant may qualify for a disability pension, which is awarded for an indefinite period of time. For both benefits, a full benefit is conditional on a loss in the working capacity of at least 60% and a partial benefit for a loss of at least 40% but below 60%. The disability evaluations are always made by trained professionals.

When determining eligibility, the individual's age, education, occupation, place of residence and capability to support himself or herself by gainful employment are all taken into account along with the medical assessment. A disability pension may also be discontinued if the working capacity of the recipient improves but that rarely happens among older recipients. There is no automatic retesting of the disability status, except for new periods of the rehabilitation benefit. Disability benefits can be collected until age 63, when the entitlement to old-age pension begins.

3.2 Disability insurance premiums

Since 2006 a major part of disability benefit costs have been financed by partially experience-rated premiums (or payroll taxes). A firm's DI premium rate in year t is obtained as a weighted sum of the base premium rate Q_t and the experience-rated premium rate M_tQ_t :

$$C_t = (1 - S_t)\, Q_t + S_t M_t Q_t,$$

where $S_t \in [0, 1]$ is the degree of experience rating. The base premium rate depends on the age structure of the workforce and varies over time, being around 1.5% of the payroll in our observation period. The experience-rated premium rate is obtained by multiplying the base rate with the experience multiplier $M_t = m\,(r_{t-2,t-3})$, which is an increasing function of the risk ratio $r_{t-2,t-3}$. The risk ratio is a measure of the costs of the disability pension

Table 1: Contribution categories and DI premiums

Risk ratio $r_{t-2,t-3}$	Contribution category	Experience multiplier M_t	DI premium by firm size, EUR			
			Payroll W_{t-2} (Degree of experience rating S_t)			
			1 mEUR (0.0)	5 mEUR (0.22)	15 mEUR (0.67)	25 mEUR (1.0)
≥ 5	11	5.50	16,000	160,000	960,000	2,200,000
$[4, 5)$	10	4.50	16,000	142,222	800,000	1,800,000
$[3, 4)$	9	3.50	16,000	124,444	640,000	1,400,000
$[2.5, 3)$	8	2.75	16,000	111,111	520,000	1,100,000
$[2, 2.5)$	7	2.25	16,000	102,222	440,000	900,000
$[1.5, 2)$	6	1.75	16,000	93,333	360,000	700,000
$[1.2, 1.5)$	5	1.35	16,000	86,222	296,000	540,000
$\mathbf{[0.8, 1.2)}$	**4**	**1.00**	**16,000**	**80,000**	**240,000**	**400,000**
$[0.5, 0.8)$	3	0.65	16,000	73,778	184,000	260,000
$[0.2, 0.5)$	2	0.35	16,000	68,444	136,000	140,000
< 0.2	1	0.10	16,000	64,000	96,000	40,000

claims made by the firm's former employees in years $t-2$ and $t-3$,[3] and it is constructed in a such a way that $r_{t-2,t-3} = 1$ if the firm's past disability costs were equal to the average costs in firms with the same age structure.[4] On the basis of the risk ratio, the firm is allocated to one of 11 possible contribution categories, each of which corresponds to a particular value of M_t between 0.1 and 5.5 (see the first three columns in Table 1). The experience-rated premium rate $M_t Q_t$ can thus differ substantially from the base rate Q_t. Namely, a firm can earn a 90% discount on the base premium or be obligated to pay a 450% surcharge on top of the base premium.

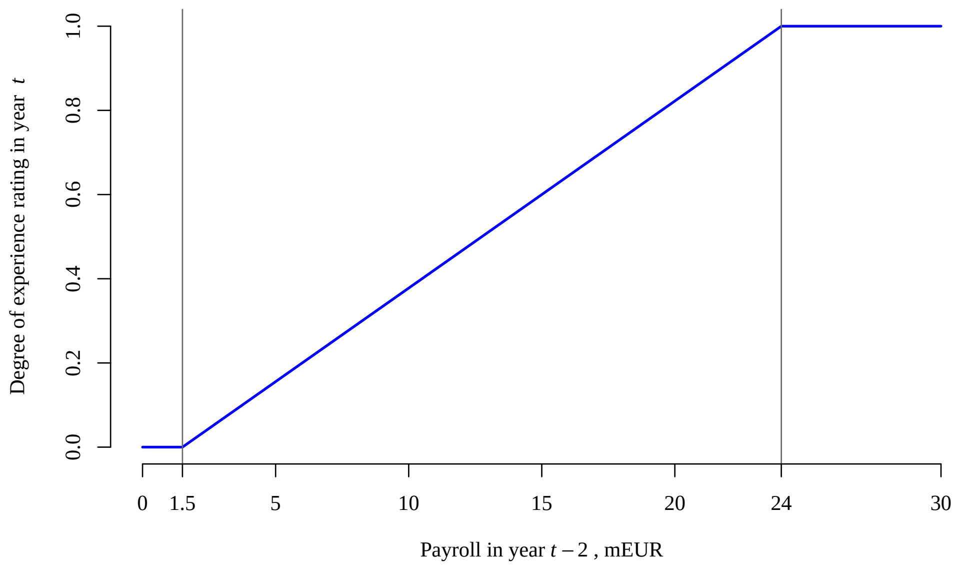
The degree of experience rating $S_t = s(W_{t-2})$ is a function of the firm's payroll two years earlier, W_{t-2}. Throughout the paper, we measure the payroll in 2004 euros. Firms with $W_{t-2} \leq 1.5$ mEUR ("small firms") are not subject to experience rating and pay the base rate because for them $S_t = 0$. For firms with $W_{t-2} \geq 24$ mEUR ("large firms") $S_t = 1$, so they pay only the experience-rated premium rate.[5] Other firms ("medium-sized firms") pay a premium rate equal to a weighted sum of the base and experienced-rated rates, and are thus only partially covered by experience rating. Within this

[3]The cost of a new pension claim equals the expected amount of disability pension benefits until age 63.

[4]The risk ratio is adjusted for the age structure in order to eliminate incentives to discriminate against older applicants in hiring.

[5]The threshold values of the 2-year lagged payroll for small and large firms were set in 2006, when the reform came into effect. These threshold values are updated annually using a payroll index. With an average salary level, the thresholds correspond approximately to firm sizes of 50 and 800 employees.

Figure 1: Degree of experience rating S_t as a function of payroll W_{t-2}



group, S_t increases linearly with W_{t-2} from 0 to 1 with the slope $\frac{1}{22.5}$. This key relationship is plotted in Figure 1.

The last four columns of Table 1 show the DI premium in each possible contribution category for four selected firm sizes. The smallest firm (1 mEUR payroll) cannot affect its DI premium which is completely determined by the base premium rate, assumed to be 1.6% of the payroll in this example (the average base premium rate in 2009). The difference between the largest and smallest possible premium is 96,000 EUR or 1.9% of the payroll for the firm with total payroll equal to 5 mEUR. This difference is as much as 5.8% and 8.6% of the payroll for the two larger firms with payrolls equal to 15 and 25 mEUR, respectively.

As another example, consider a firm whose disability costs are close to the adjusted average, so that its risk ratio lies on the interval $[0.8, 1.2)$. For this firm, regardless of its size, $M_t = 1$ and thus $C_t = Q_t$. Suppose that the firm adopts a successful health and safety program that reduces its risk ratio to the interval $[0.5, 0.8)$. As a result, M_t drops to 0.65 with a delay of two to three years and therefore the DI premium rate will decline by $0.35 S_t Q_t$. The size of this reduction is fully determined by firm size, being 35% for a large firm ($S_t = 1$), 0% for a small firm ($S_t = 0$), and something between 0% and 35% for a medium-sized firm. For this reason, we treat S_t as a measure of the firm's incentives to invest in disability-reducing measures due to experience rating, and focus on estimating the effect of S_t on the inflow to sickness and disability benefits.

The likely effects of experience rating on disability outcomes are not as obvious as one might expect at first glance. First, notice that the risk ratio depends only on disability pension claims, not on rehabilitation benefit claims. This may induce large and medium-sized employers to encourage newly disabled workers to apply for a rehabilitation benefit rather than for a disability pension. Another important point is that only the first disability pension claim of each worker is taken into account when determining the risk ratio. If a worker is first entitled to a partial disability pension but then qualifies for a full disability pension in the next year or later, only the cost of the partial pension has an effect on the firm's risk ratio in the year when that pension was awarded. Large and medium-sized employers may thus encourage their workers with health problems to apply for a partial disability pension by providing part-time work for at least a short period of time. These two features of the risk ratio calculations suggest that the effect of experience rating on partial disability pension claims and on rehabilitation benefit claims is ambiguous. To the extent that greater experience rating induces preventive measures, it should reduce transitions to sickness benefits and to all types of disability benefits. However, for a given overall inflow to disability benefits, a higher degree of experience rating may increase the shares of partial disability pension and rehabilitation benefits, and therefore the overall effect of experience rating on these benefits is a priori ambiguous.[6]

4 Data and descriptive statistics

Our data was compiled by merging administrative registers of the Finnish Centre for Pensions (ETK) and Statistics Finland. ETK is a semi-governmental body that co-ordinates the entire pension system and collects data from all pension insurance providers for statistics and research purposes. Its databases include comprehensive records on job spells and earnings for all people with some work history, as well as detailed information on disability benefit spells and the spells of sickness benefits paid directly to the worker by the Social

[6]The Finnish Centre for Pensions (ETK) performed a survey among Finnish employers in 2016 regarding firms' awareness on and attitutes towards the experience rating system. They find that the larger the firm (i.e. the higher the rate of experience rating), the more these employers considered the system to provide incentives to take care of employees' capacity to work (Liukko et al., 2017). It is possible, however, that this awareness has not always been so high since the start of the system in 2007, and that the response sample is biased towards more extreme views on the system.

Insurance Institution.[7] The records also include the "retirement events", that is, the dates when a diagnosis was made for the disability that eventually led to a rehabilitation benefit or disability pension. This is important as the disability pension costs are assigned to the employers on the basis of the year of the retirement event. Namely, the cost of a new disability pension claim increases the risk ratios of the firms where the claimant worked one and two years prior to the year of the retirement event.

The ETK data was supplemented by merging background information on workers' education, family status and living region from the Finnish Longitudinal Employer-Employee Database of Statistics Finland, which covers all people who live in Finland. Additional information on firms' ownership, turnover, size of personnel, number of establishments, industry and import/export status was obtained from the Business Register of Statistics Finland, which includes all firms subject to value added taxation or that have at least one paid employee. Together these databases allow us to follow the entire Finnish population and the universe of all firms over several decades until 2013. However, we restrict our analysis to private-sector firms and their workers in the years 2007–2013 when the experience rating system has been in effect.[8]

The outcome of interest is the probability that the worker develops a medical condition that reduces his or her working capacity, temporarily or permanently, by the extent that he or she qualifies for a sickness or disability benefit. When analyzing the incidence of sick leave, we model the probability that a new benefit period begins within the year. In the case of disability benefits, we model the probability of the onset of a disability that leads to receipt of a rehabilitation benefit or disability pension, typically with the lag of one or two years. That is, we do not consider the year when the disability benefit is granted, but the year when the underlying medical condition was diagnosed. We focus on disability events until 2011, as this leaves enough time to observe receipt of any disability-related benefits by 2013 (the disability event is only recorded once a disability benefit is granted). Notice that our outcome variables are not mutually exclusive. As an example, consider a worker who first collects a sickness benefit for one year, then a rehabilitation benefit for the next two years and finally transfers into a disability pension. Provided that all these benefits were awarded for the same medical condition

[7] Since for the first weeks of sickness (typically one to three months depending on the collective agreement) the applicant is paid by the employer, we only observe relatively long spells of sickness.

[8] We exclude 2006 because in that year the degree of experience rating was determined by the number of workers, not by the payroll.

diagnosed at the beginning of sick leave, the worker became a recipient of the sickness benefit, rehabilitation benefit and disability pension in the same year in our analysis.

To be at risk of becoming disabled in year $t \in \{2007, 2008, ..., 2011\}$, we require that the worker (*i*) is 20–62 years old, (*ii*) worked in the same private firm from year $t-2$ to year t, and (*iii*) received a certain minimum amount of wages from that firm in years $t-1$ and $t-2$, and that these wages accounted for over 50% of the worker's all wages in both years. These conditions imply that a major part of disability pension costs will be assigned to this primary employer in the case the worker becomes disabled and receives a medical diagnosis in year t that eventually leads to receipt of a disability pension.[9] Finally, we drop workers whose primary employer is very small by requiring that the firm's payroll was no less than 100,000 euros in year $t-2$ and that at least 10 employees of the firm belong to the risk set in year t .

There are 957,364 workers who satisfied these conditions at least once between the years 2007 and 2011. They worked in 14,154 different firms, amounting to over 3.1 million worker-year observations and 52,069 firm-year observations; see Table 2. 83% of these firms are classified as small according to their past payroll, and hence are not subject to experience rating. Only 172 belong to the group of large firms that is fully covered by experience rating. In terms of workers, differences in the number of observations between the size categories are much smaller, and most workers are employed by medium-sized firms. This is further illustrated in Figure 2, which displays log payroll density functions for workers who were at risk in 2011, and for their employers (densities for other cross sections are very similar).[10] Around the cutoffs of 1.5 and 24 mEUR for small and large employers, there are roughly an equal number of workers in the risk set but much less firms at the upper cutoff.

On average, workers of large firms have a longer job tenure and a higher education compared to those employed in small and medium-sized firms. Larger firms are more often located in the capital region, have a higher share of foreign owners, and quite often operate in the manufacturing sector. Some of

[9]To be specific, if a worker is awarded a disability pension in year $s \geq t$ based on the disability diagnosed in year t, the cost of this pension is assigned to the firms in which the individual worked in years $t-1$ and $t-2$, when determining the risk ratios for year s. If there were more than one employer, the pension cost is divided between the employers in proportion to the wages they paid to the worker in years $t-1$ and $t-2$.

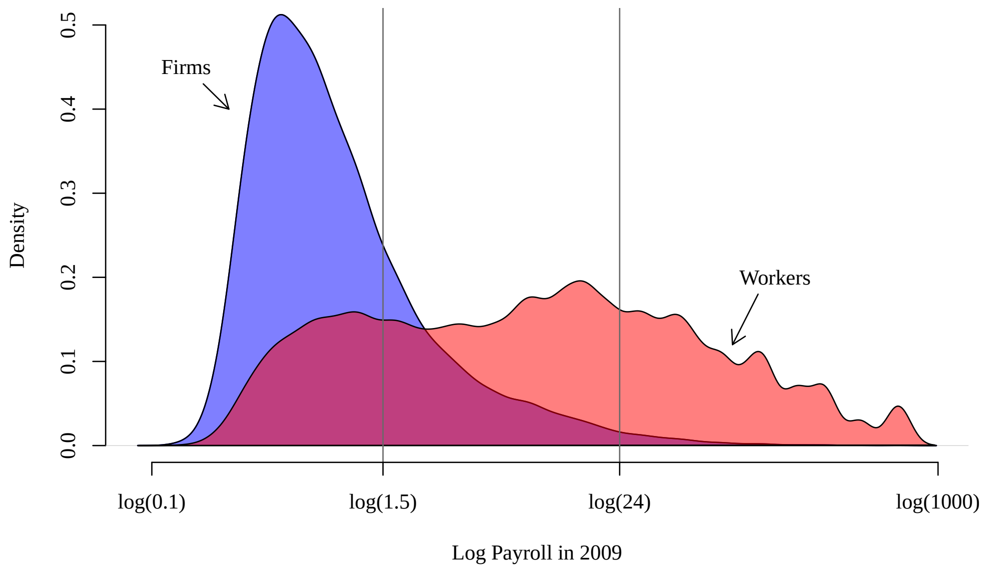
[10]Because the firm distribution is heavily skewed towards small firms and because relative differences in firm size are more relevant than absolute differences, we shall use log payrolls throughout our analysis.

Table 2: Sample means by firm size

	All firms (1)	Small (2)	Med-sized (3)	Large (4)
A. Means across workers				
Age	42.6	42.3	42.6	42.6
Tenure	11.6	9.9	11.8	12.5
Female,%	41.4	41.8	40.5	42.3
Married,%	52.9	51.3	52.8	54.1
Education,%				
Basic	14.9	16.8	14.7	13.6
Upper secondary	45.7	49.2	46.4	42.3
Lower tertiary	29.0	26.5	28.9	30.7
Upper tertiary	10.5	7.5	9.9	13.3
Number of observations	3,119,133	719,411	1,368,548	1,031,174
Number of workers	957,364	269,559	449,990	308,819
B. Means across firms				
Payroll, mEUR	2.4	0.6	4.8	62.5
Firm's age	24.4	23.5	26.6	30.4
Number of plants	3.3	1.8	5.9	46.5
Capital region,%	37.6	32.4	52.2	77.5
Incorporated company,%	82.2	80.5	87.2	89.3
Industry,%				
Manufacturing	27.2	24.9	34.2	40.6
Construction	6.3	6.9	4.0	6.3
Wholesale and retail trade	19.9	20.6	17.7	17.1
Transportation and storage	6.4	6.7	5.6	2.0
Information and communication	5.4	4.5	8.3	8.7
Finance and insurance	2.9	2.4	4.1	6.5
Health and social work	7.4	8.2	5.1	2.7
Other	24.5	25.7	21.2	16.2
Foreign ownership,%	11.4	6.1	27.2	37.0
Exporter,%	40.2	33.4	59.7	85.3
Importer,%	27.7	21.7	45.2	64.9
Year,%				
2007	18.6	18.7	18.4	18.7
2008	18.9	18.9	19.1	19.7
2009	20.3	20.3	20.3	20.6
2010	21.2	21.0	21.8	21.6
2011	21.0	21.2	20.4	19.4
Number of observations	52,069	39,324	12,033	712
Number of firms	14,154	11,728	3,187	172

Notes: Small firms had a payroll of 1.5 mEUR or less in year $t-2$ and large firms no less than 24 mEUR, while the two-year lagged payroll for medium-sized firms is between these thresholds. The firm and its workers can change size category between years. Number of observations refers to worker-year observations in Panel A and firm-year observations in Panel B.

Figure 2: Kernel density functions of the logarithm of the firm's 2-year lagged payroll across workers who were at risk in 2011 and across their employers



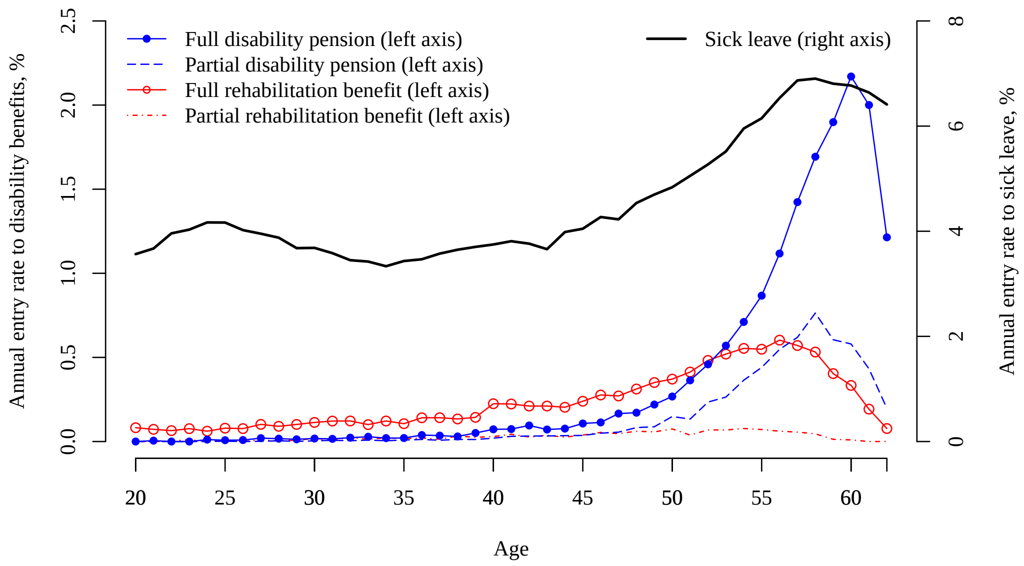
these differences between firms of different size, such as those in education and industry, are likely to be correlated with the disability risk.

Figure 3 shows the incidence of sick leave and different disability benefits by age. Not surprisingly, both sickness benefits and all kinds of disability benefits are much more common among old than young workers. The age differences are particularly pronounced in the case of full disability pensions. Workers below age 53 have been granted a full rehabilitation benefit more often than a full disability pension. The entry rate to partial rehabilitation benefits is very low at all ages, and therefore we shall merge partial and full rehabilitation benefits into one measure of rehabilitation benefits.

Table 3 reports some descriptive numbers for our disability measures by firm size. Despite the large number of workers in our data, the aggregate numbers of transitions to disability benefits are not overwhelmingly large in Panel A. Most workers in the data are relatively young and thereby have a very small risk of disability. Another reason is that a notable fraction of all disability benefit recipients have been out of work for a few years before being diagnosed as disabled, and these cases do not belong to our risk set.

As seen in Panel B, the average duration of partial and full disability pension spells is much longer than that of rehabilitation benefits (spell duration is measured until the end of 2013). This is not surprising because the rehabilitation benefits are awarded for a fixed period of time due to the expected

Figure 3: Hazard rates to sickness and disability benefits



recovery and because the recipients of such benefits are much younger on average. The periods of rehabilitation benefits are not longer compared to those of disability pension for the medium-sized and large firms than for the small firms, which we might expect to find, had the experience-rated firms encouraged their disabled employees to stay on rehabilitation benefits as long as possible in an attempt to minimize their DI premiums. An alternative check of the same hypothesis is to look at the share of a given type of benefits in all benefit days received for the same medical condition (i.e. different benefit periods that have the same retirement event). From Panel C we see that in all firm size categories around 54% of all benefit days associated with the same diagnosis are full disability pension benefit days. The relative importance of rehabilitation benefits is smaller while that of partial disability pension is larger in the medium-sized and large firms than in the small firms. One might have expected the opposite for the relative importance of rehabilitation benefits, given that the large employers have an incentive to favor rehabilitation benefits. On the other hand, the higher frequency of partial disability pensions in experience-rated firms is in accordance with the likely effect of experience rating.

Taken together these numbers do not point to clear differences between the firm size categories that we could interpret as being indicative of the behavioral effects of experience rating. Of course, these findings should be treated with caution because firms of different sizes are not directly comparable due

Table 3: Sickness and disability outcomes

	All firms	Small	Med-sized	Large
A. Number of events				
Sick leave	141,164	31,167	62,092	47,905
Rehabilitation benefits	8,567	2,247	3,664	2,656
Partial disability pension	4,383	922	1,874	1,587
Full disability pension	11,570	2,790	4,947	3,833
B. Average duration in days				
Sick leave	61	67	60	57
Rehabilitation benefits	489	508	482	481
Partial disability pension	1,028	1,033	1,045	1,005
Full disability pension	892	863	898	907
C. Share of days associated with the same retirement event,%				
Rehabilitation benefits	22.0	25.4	21.6	20.1
Partial disability pension	23.7	21.2	24.0	25.1
Full disability pension	54.3	53.5	54.4	54.8

to a large degree of heterogeneity among them, which is evident in Table 2. In the next section, we discuss statistical methods to control for heterogeneity and to conduct causal inference.

5 Econometric methods

We consider the following linear probability model for the onset of disability:

$$Y_{ijt} = \varphi_t + \tau S_{jt} + \mathbf{X}_{ijt}\boldsymbol{\beta} + v_{jt} + \eta_{ijt}, \tag{1}$$

where i indexes worker, j indexes firm and t indexes time. Y_{ijt} is a dummy variable which equals 1 if worker i in firm j becomes disabled in year t, and 0 otherwise. Since we analyze different disability outcomes, the disability event may refer to the beginning of a sickness benefit period in year t or to the retirement event in year t associated with receipt of a given type of disability benefits in year $s \geq t$. $\mathbf{X}_{ijt}$ is a vector of worker characteristics, φ_t is the calendar time effect, and v_{jt} and η_{ijt} are error terms. The degree of experience rating is given by

$$S_{jt} = s\left(W_{j(t-2)}\right) = \begin{cases} 1, & W_{j(t-2)} \geq \log(24) \\ \left(e^{W_{j(t-2)}} - 1.5\right)/22.5, & \log(1.5) < W_{j(t-2)} < \log(24) \\ 0, & W_{j(t-2)} \leq \log(1.5), \end{cases}$$

where $W_{j(t-2)}$ is the logarithm of the payroll in millions in 2004 euros. The parameter of interest is τ, the effect of the employer's degree of experience rating on the worker-specific disability risk.

Since workers in the same firm are affected by the same health and safety policy, and also share other working conditions, the disability outcomes within firms are likely to be correlated. This correlation is captured by v_{jt}. Moreover, the firm-specific working environment on the one hand, and unobserved determinants of individual health on the other hand are likely to be persistent over time, suggesting that both v_{jt} and η_{ijt} are potentially serially correlated. To deal with multilevel clustering and serial correlation we estimate the model using a simple two-step procedure.

In the first step, we construct covariate-adjusted firm-year effects by estimating

$$Y_{ijt} = \mu_{jt} + \mathbf{X}_{ijt}\boldsymbol{\beta} + \eta_{ijt}, \tag{2}$$

where $\mu_{jt} = \varphi_t + \tau S_{jt} + v_{jt}$ are firm-year fixed effects. The estimated $\hat{\mu}_{jt}$ are firm-year disability inflow rates that are adjusted for differences in worker characteristics across firms and over time. In the second step, we consider the firm-level model

$$\hat{\mu}_{jt} = \varphi_t + \tau S_{jt} + \varepsilon_{jt}, \tag{3}$$

where the error term is given by $\varepsilon_{jt} = v_{jt} + (\hat{\mu}_{jt} - \mu_{jt})$. The problem for inference is that the adjusted disability risk may vary with firm's payroll also for reasons not related to the degree of experience rating. There may be economies of scale in preventive health and safety measures; it may be easier for larger firms to accommodate and rehabilitate employees with impairments due to a larger pool of job slots; and firms with risky working environments may pay higher wages to compensate for the risk level, which would then inflate their payrolls compared to safer firms with the same number of workers. For these reasons, v_{jt} may be correlated with $W_{j(t-2)}$ and, consequently, with S_{jt}, in which case the OLS estimate of τ from (3) would be biased. This is a standard omitted variable problem, suggesting that we should also control for the direct effect of $W_{j(t-2)}$. But, because S_{jt} is a deterministic function of $W_{j(t-2)}$, it is not obvious how to separate the effects of these two variables without imposing strong functional assumptions. Nielsen, Sørensen and Taber (2010) however show that we can exploit the kinks in the policy rule for identification.

As seen in Figure 1, the relationship between the degree of experience rating and the payroll is not smooth but has kinks at the size thresholds of small and large firms. By contrast, it is quite likely that $E\left(\varepsilon_{jt}\middle| W_{j(t-2)}\right)$ is smooth at these points in the sense that

$$\lim_{w\downarrow w^*} \frac{\partial E\left(\varepsilon_{jt}\middle| W_{j(t-2)}=w, T=t\right)}{\partial w} = \lim_{w\uparrow w^*} \frac{\partial E\left(\varepsilon_{jt}\middle| W_{j(t-2)}=w, T=t\right)}{\partial w}$$

at $w^* \in \{\log(1.5), \log(24)\}$. If so, we can augment the second-stage equation with a control function $g\left(W_{j(t-2)}\right) \equiv E\left(\varepsilon_{jt}\middle| W_{j(t-2)}, T=t\right)$ to obtain

$$\hat{\mu}_{jt} = \varphi_t + \tau S_{jt} + g\left(W_{j(t-2)}\right) + \xi_{jt}, \tag{4}$$

where S_{jt} and $W_{j(t-2)}$ are mean-independent of the new error term $\xi_{jt} \equiv \varepsilon_{jt} - E\left(\varepsilon_{jt}\middle| W_{j(t-2)}, T=t\right)$ by construction. It follows that the kink at the payroll cutoff identifies τ. Specifically, we have

$$\tau = \frac{\lim_{w\downarrow w^*} \frac{\partial E\left(\hat{\mu}_{jt}\middle| W_{j(t-2)}=w, T=t\right)}{\partial w} - \lim_{w\uparrow w^*} \frac{\partial E\left(\hat{\mu}_{jt}\middle| W_{j(t-2)}=w, T=t\right)}{\partial w}}{\lim_{w\downarrow w^*} s'(w) - \lim_{w\uparrow w^*} s'(w)}, \tag{5}$$

because $g(\cdot)$ is smooth at $w^* \in \{\log(1.5), \log(24)\}$ by assumption. The right-hand side of the equation equals the ratio of the change in the slope of the conditional expectation of the adjusted risk to the change in the slope of the deterministic experience rating rule at the payroll cutoff w^*. In other words, the causal effect of S_{jt} is identified from a kink in the average outcome associated with a kink in the experience rating rule without any assumptions about $g(\cdot)$ except the smoothness. This identification strategy that hinges on the kinks in the policy rule was coined "regression kink design" (RKD) by Nielsen, Sørensen and Taber (2010).[11]

[11] Earlier applications using similar strategies include Guryan (2001), Rothstein and Rouse (2007) and Dahlberg et al. (2008). Card et al. (2015) develop a formal statistical theory for RKD and provide conditions under which causal effects in even more general nonseparable models are identified. They also discuss nonparametric inference using local linear and local quadratic regression models. Böckerman, Kanninen and Suoniemi (2015) apply the RKD approach to study the effects of sickness benefits on the duration of sick leave using Finnish data.

In practice, the control function $g(\cdot)$ is unknown. One possibility is to adopt some flexible function for $g(\cdot)$, such as a polynomial function, and estimate τ from (3) by ordinary least squares (OLS) or by weighted least squares (WLS) using the number of workers at risk as weights. WLS is likely to be more efficient because the estimation error $\hat{\mu}_{jt} - \mu_{jt}$ is smaller for large firms on average. Yet this cannot be taken for granted; if $\hat{\mu}_{jt} - \mu_{jt} \approx 0$ and v_{jt} is homoskedastic or its variance is not related to the size of risk group, WLS may also be less efficient. Moreover, if the effect of S_{jt} varies with firm size, OLS and WLS will identify different parameters, i.e. different weighted averages of the heterogeneous effects. For these reasons, we shall report both the OLS and WLS estimates. Using the cluster-robust covariance matrix estimator for the second-stage results, we make our statistical inference robust against any type of intragroup heteroskedasticity and serial correlation.

Another possibility is to invoke the relationship in (5) directly. Since the denominator is known, we only have to estimate the numerator at the two cutoffs. This can be done in a nonparametric fashion using only observations in the neighborhood of the cutoff $w^* \in \{\log(1.5), \log(24)\}$. In this case, we estimate a local polynomial model of the form

$$\hat{\mu}_{jt} = \varphi_t + \sum_{p=1}^{P} \left[\gamma_p \left(W_{j(t-2)} - w^* \right)^p + \varphi_p 1 \left\{ W_{j(t-2)} > w^* \right\} \left(W_{j(t-2)} - w^* \right)^p \right] + \vartheta_{jt} \tag{6}$$

where the indicator function $1\{A\}$ equals 1 if A is true and 0 otherwise, and all observations included in the sample satisfy the condition $\left| W_{j(t-2)} - w^* \right| \leq h$, where h is the bandwidth. As φ_1 is the change in the slope of the conditional expectation of $\hat{\mu}_{jt}$ at w^*, we can obtain an estimate of τ by dividing $\hat{\varphi}_1$ with the change in the slope of the experience rating rule at w^*, which is equal to $s'(\log(1.5)) = 1.5/22.5$ at the lower cutoff and $-s'(\log(24)) = -24/22.5$ at the upper cutoff.

The former approach ("global RKD") is more efficient as it uses all available data, but the latter ("local RKD") is more robust by utilizing only data around the payroll cutoff at which the parameter of interest is identified. If the effect of experience rating is heterogeneous, the two approaches identify different parameters. In practice, the local RKD estimates are rather imprecise and not very informative due to the limited number of firms around the payroll cutoffs, especially around the upper cutoff. We thus apply the global approach in our main analysis but discuss the local estimates as part of our robustness analysis.

The smoothness assumption for $g(\cdot)$ rules out discrete changes in other policy parameters at the size thresholds of small and large firms. If there are other discrete changes at the same size thresholds, $g(\cdot)$ may not fully capture their confounding effects and thereby the estimate of τ will be biased. This is a matter of concern in our case because large employers are partially liable also for the costs of extended unemployment benefits received by their former employees, and the firm size thresholds for these liabilities coincide with those for DI premiums. In Finland, the entitlement period of unemployment insurance (UI) benefits is about two years, but those who are 59 or older on the day the regular benefits expire are entitled to extended benefits until age 63. That is, workers aged 57 or above at the time of dismissal can collect UI benefits until old-age retirement. This scheme is known as "unemployment tunnel" (UT). Large employers in particular often target dismissals at those employees who can qualify for the extended benefits after two years of unemployment (Kyyrä and Wilke, 2007). When an extended benefit is granted to the worker, the former employer may have to pay a given share of the extended benefit costs as a lump sum payment to the Unemployment Insurance Fund.[12] This cost share increases linearly from 0% to 80% as a function of the employer's payroll in the year preceding the dismissal. The payroll thresholds for these minimum and maximum cost shares are the same as those used in DI, and thereby any two firms that differ in the degree of experience rating in DI also differ in the degree they are responsible for the extended UI benefit costs. By implication, the estimate of τ from (3) may also capture the effect of employer's liabilities for the extended UI benefits.

Nevertheless, we can still identify the effect of experience rating in DI because the disability pension benefits of workers of all ages affect the DI premium, whereas only the extended UI benefits received by workers dismissed at age 57 or later affect the employer's UI liabilities. To separate the two effects we augment our model by allowing the effect of experience rating to differ between workers who would be eligible for the UT scheme in the case of layoff and those who would not be; that is, we replace the first-stage estimating equation (2) with

$$Y_{ijt} = \mu_{jt} + \delta\left(UT_{ijt} \cdot S_{jt}\right) + \mathbf{X}_{ijt}\boldsymbol{\beta} + \eta_{ijt}, \tag{7}$$

[12] The cost of extended benefits is calculated assuming the worker will collect them until age 63 irrespective of the actual behavior. In the case of a worker who qualifies for an old-age pension before the regular benefits expire, the former employer is liable for a share of the costs of regular benefits actually paid to the worker.

where UT_{ijt} equals 1 if the worker is aged 57 or more in year t, and 0 otherwise. The second-stage equation (3) remains unchanged. For UT-eligible workers the effect of S_{jt} is $\tau + \delta$, where δ is the confounding effect of employers' liabilities for extended UI benefits, and it is identified from differences in the disability risk between workers under and above the UT age threshold within medium-sized and large firms. Without the inclusion of the interaction term $UT_{ijt} \cdot S_{jt}$, the estimates of μ_{jt} may also capture the potential spillover effect of UI liabilities on the disability outcomes.

6 Validity of identifying assumptions

The key condition for identification is that the payroll density is sufficiently smooth at the payroll cutoffs. This implies that firms cannot perfectly control which side of the kink they end up on, which plausibly holds in our case. Even though firms can decide the number of workers they hire and they are free to raise wages, it would be difficult for them to precisely determine the sum of all wages. This is because the firms cannot (at least easily) cut wages, because wages can increase during the year due to the collective agreement, and because the size and composition of the workforce can change unexpectedly due to quits, parental leaves etc. And even if the firms were able to manipulate freely their payroll, it is unlikely they would do so in order to choose a particular degree of experience rating because the DI premiums are only a small fraction of the overall labor costs. Nonetheless, we test formally the smoothness of the density around the payroll cutoffs.

The density function across firms in Figure 2 may mask spikes around the kink points, as it was obtained by Kernel smoothing. To analyze the densities around the kinks more closely we apply the test proposed by McCrary (2008) separately to two samples of observations.[13] The "bottom kink sample" includes firms with past payroll between 0.18 and 12.75 mEUR (the bandwidth $h \approx 2.14$ for log payroll), while the "top kink sample" includes those firms with past payroll between 12.75 and 45.18 mEUR (the bandwidth $h \approx 0.63$). These bandwidths were chosen in such a way that the degree of experience rating varies on the interval $[0, 0.5)$ in the bottom kink sample and on the interval $[0.5, 1]$ in the top kink sample. In the first step of the test procedure, a finely gridded density histogram is computed. This histogram is then smoothed using local linear regression on each side of the cutoff, and, finally,

[13]The test was performed using Drew Dimmery's rdd package for R.

a Wald test for the null hypothesis that the discontinuity at the cutoff is zero is computed. The results are graphically illustrated in Figure 4, where we plot the density estimates and the smoothed regression lines along with the 95% confidence intervals for both samples. The density looks very smooth at both cutoffs, and we cannot reject the null hypothesis that the discontinuity at the cutoff is zero.[14]

Another assumption is that the distributions of the predetermined variables should evolve smoothly across firms near the cutoffs. In Figures 5 and 6 we plot the local averages of selected firm covariates against log payroll, measured in deviation from the cutoff value. On each side of the cutoff, firms are divided into 20 equally-sized bins, and the dots in the graphs are local averages plotted against the bin midpoints. In the bottom kink sample, some of the covariates, such as foreign ownership, importer and exporter status, are clearly correlated with firm size, but their values evolve smoothly through the cutoff value as required. So there is no evidence of notable discontinuities in the covariate values at the payroll cutoffs. We thus conclude that the identifying assumptions of the RKD model are satisfied.

7 The effects of experience rating

In the first step we construct the covariate-adjusted risk measures $\hat{\mu}_{jt}$ by estimating (7) for various disability outcomes. Due to the large number of firm-year fixed effects, we estimate $\boldsymbol{\beta}$ and δ using the within estimator, and then compute $\hat{\mu}_{jt}$ from the residuals of this regression. The estimates of $\boldsymbol{\beta}$ and δ for transitions to sickness and rehabilitation benefits, and to partial and full disability pensions, are reported in Table A.1 in the Appendix. In addition, we also consider receipt of disability and rehabilitation benefits that were granted for a given medical condition. In these cases, we make no distinction between partial and full disability pensions, and only consider the two most common diagnosis categories – mental and behavioral disorders and diseases of musculoskeletal system and connective tissue – while pooling all other medical diagnoses into a single category. The first-stage estimates for these models are shown in Table A.2.

Because disability risk differences between workers are not the focal point in our analysis, we do not discuss those results in detail. The interaction

[14] The graph and tests were performed using the data pooled over the years, ignoring the likely correlation between observations on the same firm. If the test is applied separately to each cross section, the null hypothesis is never rejected.

Figure 4: Densities for log payroll around the cutoffs of 1.5 and 24 mEUR from the firm-level data pooled over the years

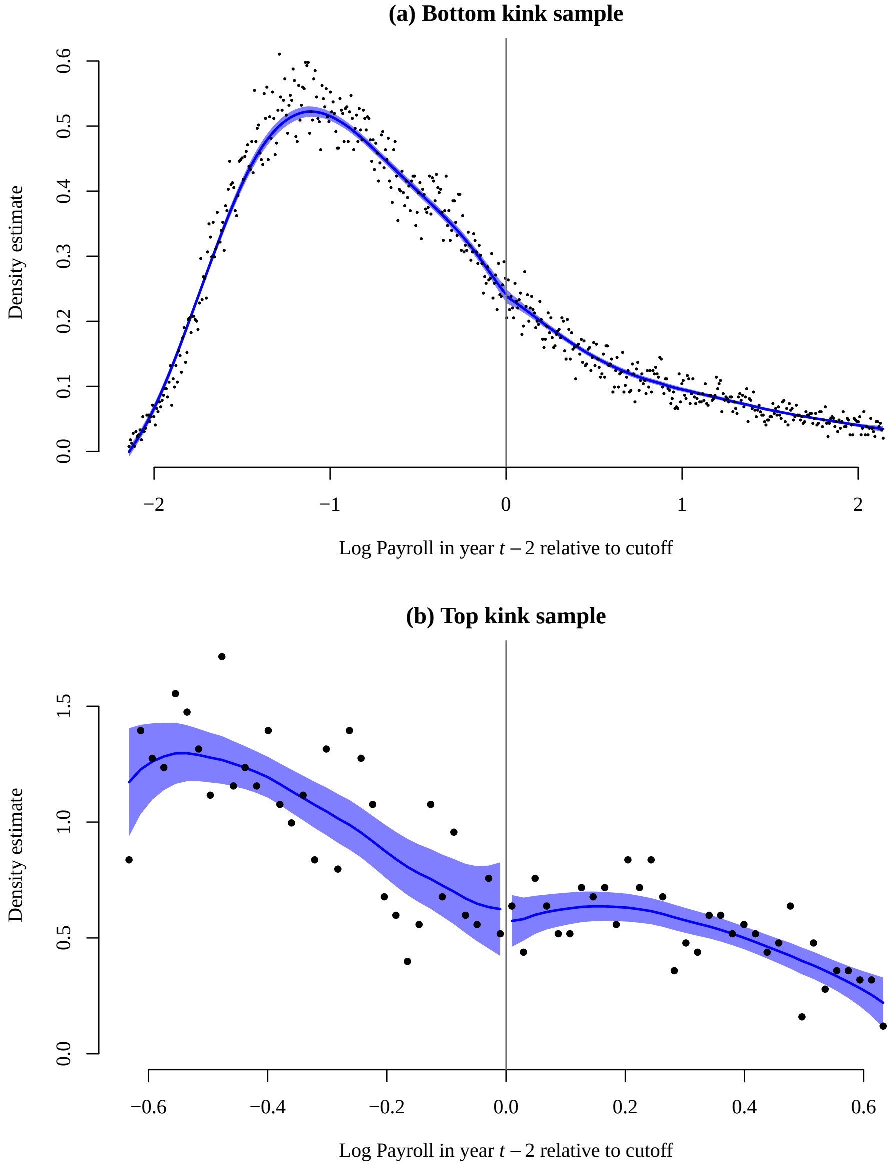


Figure 5: Local averages of selected firm covariates around the bottom kink

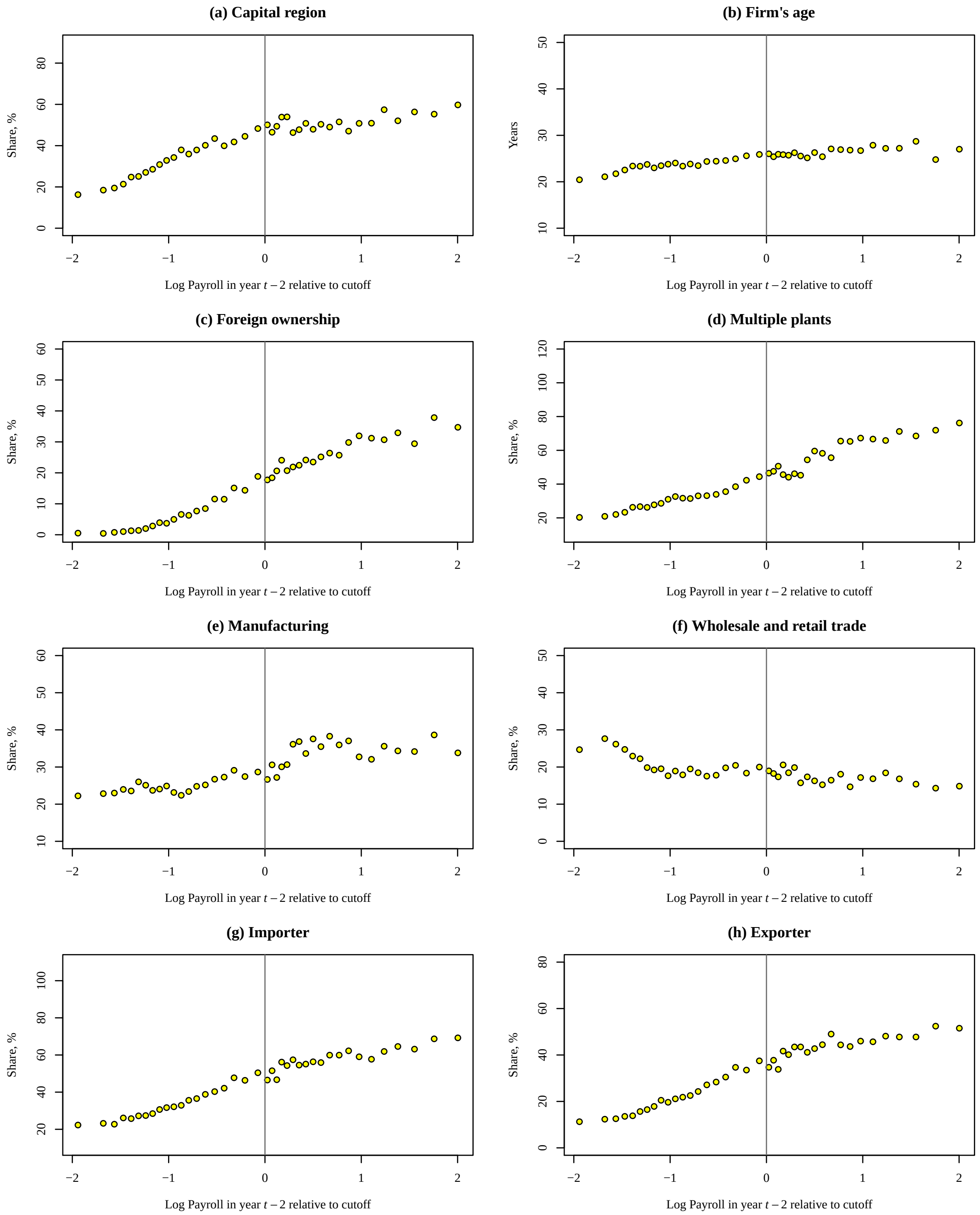
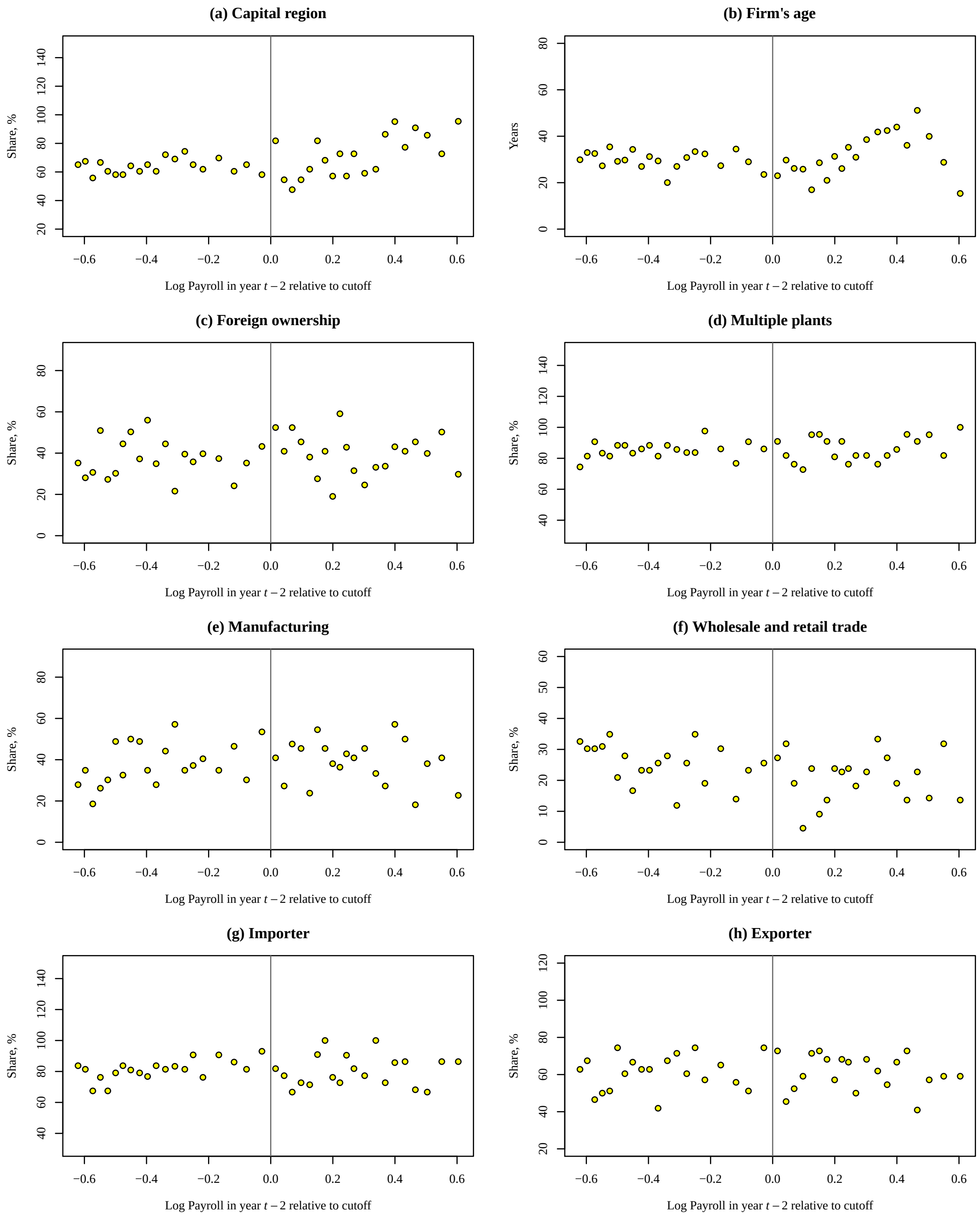


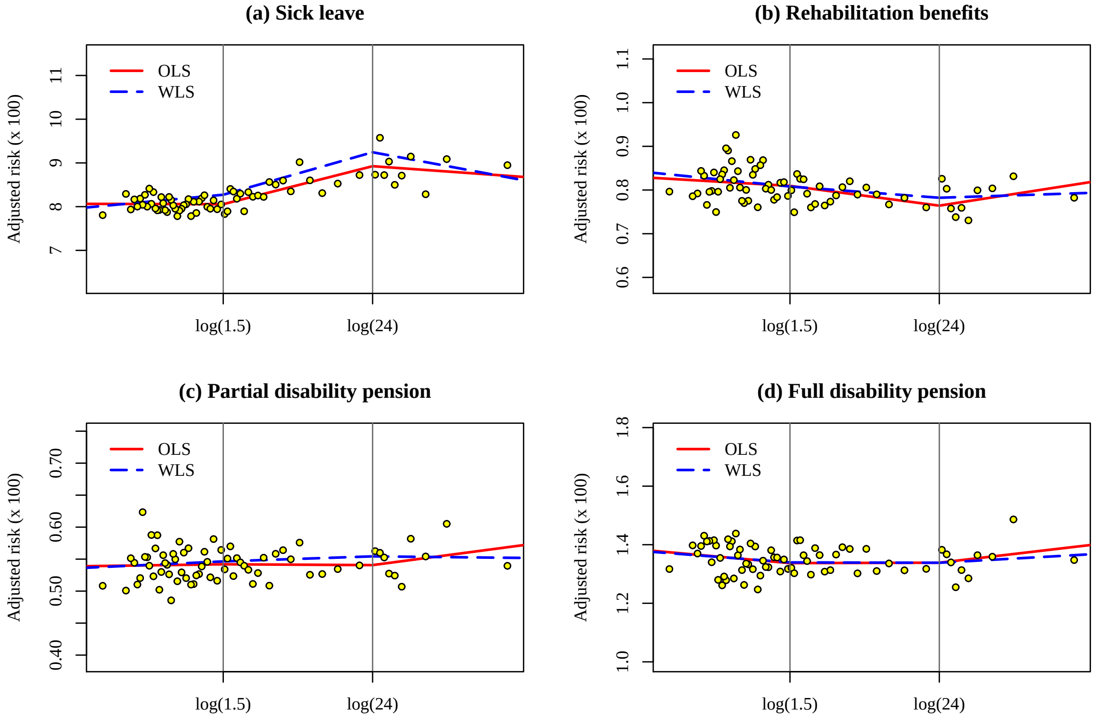
Figure 6: Local averages of selected firm covariates around the top kink



term for the degree of experience rating and the UT-eligibility dummy is of particular interest, however, because its coefficient captures the possible confounding effect due to employers' liabilities in extended UI benefits. This effect is statistically significant only for the receipt of a partial disability pension, in which case the effect is positive and significant at the 5% level (Table A.1). The point estimate implies that the likelihood of being awarded a partial disability pension increases by one-fifth (from about 0.5% to 0.6% a year for the 57-year old worker) when the degree of experience rating increases from 0 to 1. This may indicate that laying off a worker who becomes partially disabled is a worthy alternative to offering a part-time job combined with a partial disability pension. For an experience-rated firm, laying off a partially disabled worker aged 57 or older is probably a more costly choice due to the expected costs from extended UI benefits. If laid off, the worker may decide not to apply for a partial disability pension but just collect full-time UI benefits until old-age pension, in which case the employer has to pay its share of the extended UI benefits. On the other hand, if the laid-off worker chooses to claim a partial disability pension, he or she may also receive partial UI benefits on top of the partial pension, in which case the employer will incur costs from both types of benefits. Recall that the employer cannot avoid disability costs by laying off workers with health problems because the disability pension costs are assigned to the firms in which the individual worked one and two years prior to the medical diagnosis. In the case of a younger worker who becomes partially disabled there is no obvious cost difference between the two alternatives.

Turning to the main outcomes, we plot the local averages of $\hat{\mu}_{jt}$ against $W_{j(t-2)}$ in Figure 7. Within size categories, which are separated by vertical lines, the firms are divided into equally-sized subsets on the basis of the number of observations: small firms were divided into 40, medium-sized firms into 20 and large ones into 10 groups. These bins include 984, 601–602 and 71–72 observations for small, medium-sized and large firms, respectively. The different group sizes also reflect the fact that $\hat{\mu}_{jt}$ are more accurately estimated for larger firms with larger numbers of workers at risk. The dots in the graph correspond to the average disability risk of a reference worker (i.e. one with $\mathbf{X}_{ijt} = \mathbf{0}$ and $UI_{ijt} = 0$) within firm groups, and differences between them describe how this risk varies across firms of different sizes that are subject to different degrees of experience rating. The reference worker is a 56-year-old man who is not married, has a lower than upper secondary education and

Figure 7: Local averages of adjusted disability risks $\hat{\mu}_{jt}$ versus past payrolls $W_{j(t-2)}$ along with piecewise regression lines.



has worked in the current firm for the past 11 years. He is clearly a high-risk person on the basis of his age and education, which should be kept in mind when looking at the average risk levels across firms. In the graph we also show piecewise linear OLS and WLS regression lines. These were estimated from the panel data on firms, and hence independent of the chosen bin sizes for the local averages.[15] The slope of the regression line for the small and large firms describes only the effect of firm size, because the degree of experience rating is constant within these groups. Among the medium-sized firms, the degree of experience rating increases with the payroll, and therefore the slope for these firms captures the joint effect of firm size and experience rating. Thus, under the assumption that the firm size effect is approximately constant across the firm size distribution, a smaller slope for the medium-sized firms than for the other two firm groups would indicate a negative effect for the degree of experience rating.

Somewhat surprisingly, the slope is largest for the medium-sized firms in

[15]In the WLS regressions the firm-year observations are weighted by the number of workers at risk during the year. Allowing level shifts in addition to the slope changes at the size thresholds has very little effect on the size of the estimated slope changes, nor does it affect the statistical significance of these estimates.

Figure 7a which suggests that the degree of experience rating *increases* the likelihood of sick leave. Although the slopes for the different firm groups are quite different visually in Figure 7a, statistical significance of the differences is less clear. In the OLS regression the slope change is statistically significant at the bottom kink but not at the top kink, whereas in the WLS regression the slope change is significant at the top kink but not at the bottom kink.

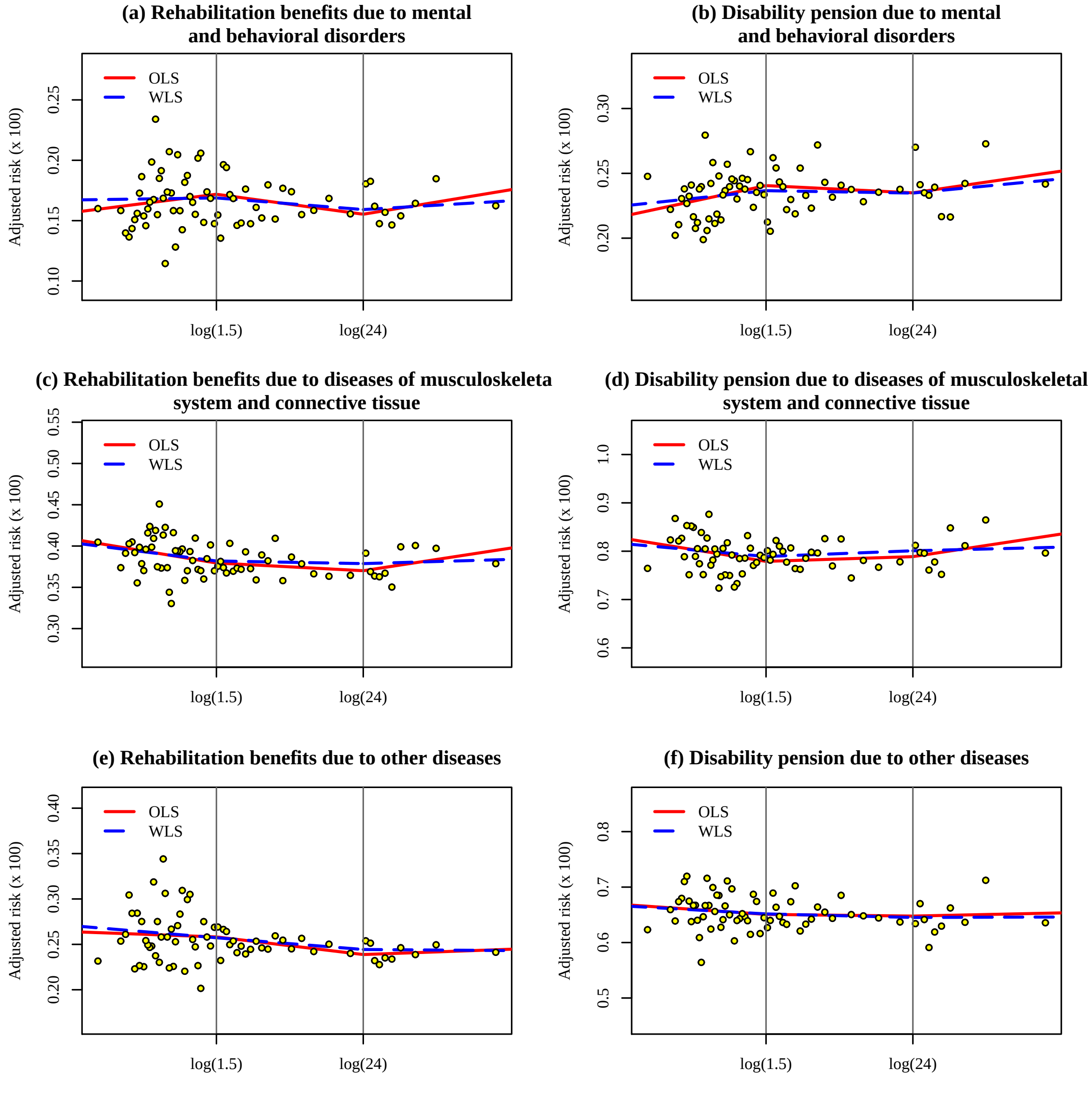
In Figure 7b, the slope differences are small but consistent with a negative effect of experience rating on transitions to rehabilitation benefits. However, the slope change is marginally significant (at the 10% level) only at the top kink in the OLS regression. In Figures 7c and 7d, we see no evidence of notable slope differences for the incidence of a partial or full disability pension, and none of the slope changes are different from zero at the conventional risk levels.

In practice, the employers can only influence the onset of certain health problems. For this reason, we replicate the analysis by looking at the incidence of disability benefits due to a given medical condition. We differentiate between mental and behavioral disorders and diseases of the musculoskeletal system and connective tissue, which are the most common diagnosis categories, and other diseases, which include all other diagnoses. These results are shown in Figure 8. Compared to Figure 7, the local averages are more noisy but the slope differences between the firm size groups remain very small and none of the slope changes are statistically significant.

In summary, the adjusted disability risks vary very little with firm size, indicating no effects for the degree of experience rating. This finding should, however, be treated with caution as we have only controlled for differences in the composition of the workforce across firms and over time. If uncontrolled differences in observed or unobserved characteristics of firms are correlated with firm size, the relationships observed in Figures 7 and 8 may describe poorly the effect of experience rating.

In Tables 4 and 5, we report the OLS and WLS estimates of τ obtained by estimating equation (3) when $g(\cdot)$ is specified as a linear, quadratic, cubic or quartic function. All numbers are multiplied by 100, so the point estimates give the change in the outcome probability measured in percentages when the degree of experience rating jumps from 0 to 1. The average outcome probability for the reference worker is shown in the first column. The specifications preferred on the basis of the Akaike information criterion (AIC) are marked in bold. The estimated effects in Table 4 are in general relatively small, and

Figure 8: Local averages of adjusted disability risks$\hat{\mu}_{jt}$ versus past payrolls $W_{j(t-2)}$ by diagnosis along with piecewise regression lines



almost all of them are also statistically insignificant. The results from the linear specification are somewhat sensitive to the inclusion of firm controls (columns 1 and 5), but the effects are very small and statistically insignificant at the 5% level except for the inflow rate to sickness benefits in column 1 in Panel A. It is reassuring that, except for the receipt of sickness benefits, the results from polynomial orders 2 to 4 are insensitive with respect to the inclusion of firm controls. There is a statistically significant negative effect on the incidence of a full disability pension in the quartic model in column 4 in Panel A, and this effect is driven by diseases other than mental and behavioral disorders (see Panel C). On the basis of the point estimate from the quartic model in Panel A, the maximum increase in the degree of experience rating from 0 to 1 could reduce the incidence of a disability pension by one-fifth (from 1.35% to 1.10% for the reference worker). However, the quartic model is not the preferred specification according to the AIC. In fact, it has the worst AIC among the competing polynomial specifications, suggesting the model is likely to be over-parametrized. In Table 5, the quartic model for the disability pension outcomes is the preferred specification when using WLS, but the point estimates are smaller and do not differ from zero at the conventional risk levels.

Compared to the OLS estimates in Table 4, the WLS estimates in Table 5 are rather similar, albeit closer to zero in most cases, and somewhat more precise (expect for sick leave) as one might have expected. Unlike in the OLS regressions, the preferred WLS specification is typically the quartic model. This does not affect our conclusions, however. The results from the AIC-preferred specifications in Tables 4 and 5 uniformly suggest that the degree of experience rating has no effect on any of the considered outcomes.

8 Robustness checks

The model outlined in equation (1) assumes an instant effect for the degree of experience rating. In practice, the disability reducing programs are likely to last for some years and their effects may be realized with a lag. The firms also have forecasts for their growth rates and can therefore anticipate their degrees of experience rating in the next few years. If so, annual variation in the degree of experience rating around the average level may not be very important for the firm's health and safety policy. As a robustness check, we re-estimated the models without using the annual variation for identification.

Table 4: OLS estimates for the effects of the time-varying degree of experience rating τ ($\times 100$)

	Average risk ($\times$100)	No firm controls				With firm controls			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Baseline results									
Sick leave	8.133	0.592**	0.328	**-0.048**	-0.718	**-0.404**	0.132	0.082	-0.361
		(0.264)	(0.547)	**(0.571)**	(1.011)	**(0.250)**	(0.495)	(0.537)	(0.933)
Rehabilitation benefits	0.811	**-0.004**	0.041	0.090	-0.072	-0.068*	0.031	**0.097**	-0.039
		(0.037)	(0.051)	(0.075)	(0.093)	(0.038)	(0.053)	**(0.076)**	(0.090)
Partial disability pension	0.541	**0.001**	0.006	0.019	-0.097	**-0.023**	0.007	0.027	-0.095
		(0.025)	(0.033)	(0.048)	(0.063)	**(0.025)**	(0.033)	(0.047)	(0.063)
Full disability pension	1.350	**0.033**	-0.091	-0.115	-0.253**	**-0.023**	-0.089	-0.086	-0.204*
		(0.045)	(0.066)	(0.089)	(0.113)	**(0.046)**	(0.063)	(0.088)	(0.111)
B. Rehabilitation benefits by diagnosis									
Mental and behavioral disorders	0.166	**-0.020**	0.024	0.050	0.010	**-0.030**	0.020	0.050	0.004
		(0.019)	(0.027)	(0.039)	(0.047)	**(0.020)**	(0.028)	(0.041)	(0.047)
Diseases of musculoskeletal system and connective tissue	0.387	**0.021**	-0.010	-0.010	-0.042	**-0.015**	-0.018	-0.010	-0.019
		(0.023)	(0.031)	(0.048)	(0.059)	**(0.023)**	(0.031)	(0.048)	(0.058)
Other diseases	0.258	**-0.006**	0.028	0.050	-0.040	**-0.023**	0.028	0.058	-0.024
		(0.021)	(0.026)	(0.038)	(0.043)	**(0.021)**	(0.026)	(0.039)	(0.042)
C. Disability pension by diagnosis									
Mental and behavioral disorders	0.234	**-0.023**	-0.002	0.009	0.044	**-0.023**	0.002	0.015	0.051
		(0.018)	(0.025)	(0.037)	(0.047)	**(0.018)**	(0.024)	(0.036)	(0.046)
Diseases of musculoskeletal system and connective tissue	0.793	**0.052**	-0.052	-0.077	-0.190**	**0.000**	-0.048	-0.051	-0.151**
		(0.033)	(0.047)	(0.063)	(0.079)	**(0.034)**	(0.044)	(0.062)	(0.075)
Other diseases	0.655	**0.005**	-0.014	-0.012	-0.138*	**-0.020**	-0.025	-0.014	-0.132*
		(0.030)	(0.040)	(0.059)	(0.073)	**(0.031)**	(0.039)	(0.058)	(0.072)
Effect of past payroll		Linear	Quadratic	Cubic	Quartic	Linear	Quadratic	Cubic	Quartic

Notes: Average risk is the unweighted average of $\hat{\mu}_{jt}$ across firms and over time. Firm controls include indicators for region, industry, multiple establishments, export/import status and legal form, as well as the share of foreign ownership and log firm age. The AIC-preferred polynomial in bold. Data include 52,069 firm-years observations from 14,154 unique firms. The robust standard errors clustered at the firm level in parenthesis. Significance levels: *** 1%, ** 5% and * 10%.

That is, we replaced S_{jt} and $W_{j(t-2)}$ in the first-stage equation with their averages over the years and the firm-year fixed effect in the second-stage equation with the firm fixed effect. In the second stage, we then regressed the estimated firm fixed effects on the average degree of experience rating using various polynomials of the average payroll. The OLS results of this exercise are reported in Table 6. The results are in accordance with the OLS findings in Table 4. In particular, the estimates are robust with respect to the inclusion of the covariates while the AIC-preferred effects are close to zero and statistically insignificant, albeit less precisely estimated than the effects in Table 4. We do not report the corresponding WLS results but also those are in line with the WLS estimates in Table 5.

We have so far considered the global versions of the RKD model, which we have estimated using all data that also include observations that are quite far away from the payroll cutoffs. One may argue this approach is somewhat dubious given that the parameter of interest is semiparametrically identified at the cutoff value. As a more robust approach, we can use only observations

Table 5: WLS estimates for the effects of the time-varying degree of experience rating τ ($\times 100$)

	Average risk ($\times$100)	No firm controls (1)	(2)	(3)	(4)	With firm controls (5)	(6)	(7)	(8)
A. Baseline results									
Sick leave	8.703	1.016*	0.973**	-0.274	**-0.185**	0.393	0.372	-1.131	**-0.995**
		(0.546)	(0.445)	(1.377)	**(1.261)**	(0.541)	(0.438)	(1.316)	**(1.240)**
Rehabilitation benefits	0.797	0.008	0.009	-0.017	**-0.007**	-0.010	-0.010	-0.018	**-0.010**
		(0.024)	(0.025)	(0.053)	**(0.052)**	(0.023)	(0.023)	(0.049)	**(0.049)**
Partial disability pension	0.550	0.018	0.018	-0.001	**0.003**	0.005	0.004	-0.015	**-0.011**
		(0.019)	(0.019)	(0.037)	**(0.037)**	(0.019)	(0.018)	(0.036)	**(0.036)**
Full disability pension	1.346	0.033	0.034	-0.084	**-0.070**	0.019	0.019	-0.127	**-0.113**
		(0.040)	(0.040)	(0.084)	**(0.080)**	(0.038)	(0.039)	(0.079)	**(0.076)**
B. Rehabilitation benefits by diagnosis									
Mental and behavioral disorders	0.164	**-0.005**	-0.005	0.002	0.005	**-0.007**	-0.007	0.015	0.016
		(0.010)	(0.010)	(0.020)	(0.020)	**(0.011)**	(0.011)	(0.020)	(0.020)
Diseases of musculoskeletal system and connective tissue	0.382	0.013	0.013	-0.020	**-0.017**	-0.001	-0.001	-0.033	**-0.029**
		(0.015)	(0.015)	(0.032)	**(0.032)**	(0.014)	(0.014)	(0.030)	**(0.030)**
Other diseases	0.251	**0.000**	0.001	0.001	0.004	**-0.002**	-0.002	0.000	0.003
		(0.011)	(0.011)	(0.024)	(0.023)	**(0.011)**	(0.011)	(0.021)	(0.021)
C. Disability pension by diagnosis									
Mental and behavioral disorders	0.236	0.000	0.000	-0.006	**-0.001**	**0.000**	0.000	0.004	0.006
		(0.011)	(0.011)	(0.021)	**(0.020)**	**(0.010)**	(0.010)	(0.020)	(0.020)
Diseases of musculoskeletal system and connective tissue	0.799	0.039	0.040	-0.023	**-0.017**	0.017	0.017	-0.071	**-0.062**
		(0.031)	(0.030)	(0.064)	**(0.061)**	(0.026)	(0.026)	(0.053)	**(0.052)**
Other diseases	0.649	0.004	0.004	-0.047	**-0.041**	0.001	0.001	-0.061	**-0.056**
		(0.020)	(0.020)	(0.038)	**(0.038)**	(0.020)	(0.020)	(0.038)	**(0.038)**
Effect of past payroll		Linear	Quadratic	Cubic	Quartic	Linear	Quadratic	Cubic	Quartic

Notes: Average risk is the weighted average of $\hat{\mu}_{jt}$ across firms and over time. Firm controls include indicators for region, industry, multiple establishments, export/import status and legal form, as well as the share of foreign ownership and log firm age. The AIC-preferred polynomial in bold. The number of workers at risk used as weights. Data include 52,069 firm-years observations from 14,154 unique firms. The robust standard errors clustered at the firm level in parenthesis. Significance levels: *** 1%, ** 5% and * 10%.

near the payroll cutoff and estimate the local polynomial model (6) in the second step. To implement this approach we have to choose the polynomial order P and bandwidth h. In practice, P is typically set to either 1 or 2 as higher order polynomials become unstable when data from the narrow interval is used. The bandwidth is a trade-off between the precision of the estimates and accuracy of the polynomial approximation to the unknown underlying expectation function. In applied work, some ad hoc value is often chosen and the robustness of the results is verified by re-estimating the model with alternative bandwidths. We follow this practice, and estimate both local linear and local quadratic specifications using a range of bandwidths. The maximum bandwidths considered are the same we used to assess the smoothness assumptions in Section 6: 2.14 for the bottom kink sample, which corresponds to the interval of $[0, 0.5)$ for the degree of experience rating, and 0.63 for the top kink sample, which corresponds to the interval of $[0.5, 1]$. Starting with the maximum bandwidth, we have re-estimated the models by

Table 6: OLS estimates for the effects of the average degree of experience rating τ ($\times 100$)

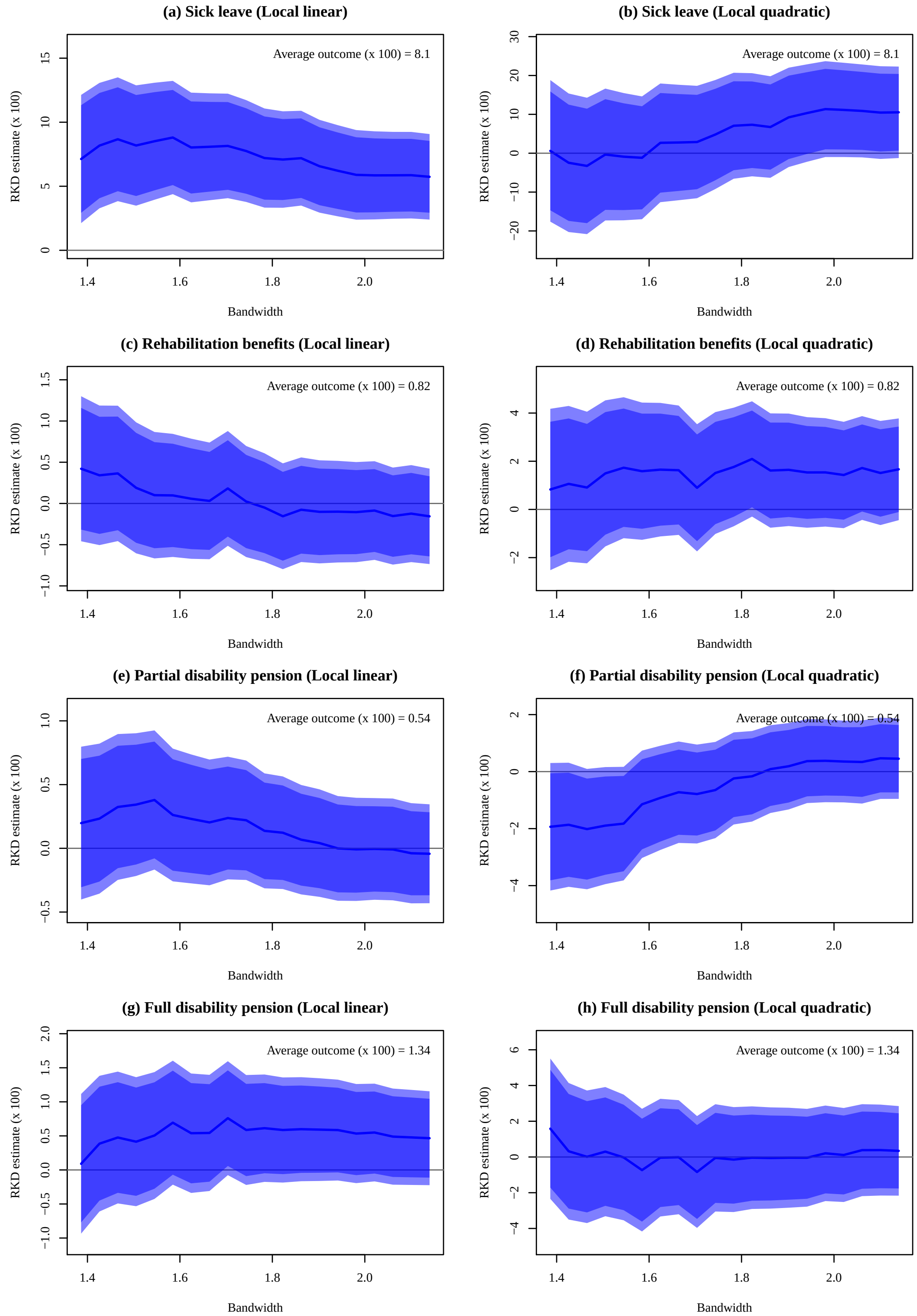
	Average risk	No firm controls				With firm controls			
	($\times$100)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Baseline results									
Sick leave	7.957	0.568*	-0.056	**-0.477**	-0.413	**-0.374**	-0.164	-0.351	0.160
		(0.308)	(0.625)	**(0.667)**	(1.160)	**(0.302)**	(0.572)	(0.643)	(1.115)
Rehabilitation benefits	0.830	**0.020**	-0.002	0.006	0.092	**-0.040**	-0.013	0.003	0.168
		(0.061)	(0.090)	(0.125)	(0.171)	**(0.062)**	(0.089)	(0.126)	(0.175)
Partial disability pension	0.524	**-0.015**	0.009	0.023	-0.100	**-0.035**	0.008	0.026	-0.081
		(0.036)	(0.050)	(0.067)	(0.086)	**(0.035)**	(0.048)	(0.065)	(0.084)
Full disability pension	1.437	**0.049**	-0.177	-0.221	-0.263	**0.023**	-0.172	-0.208	-0.167
		(0.073)	(0.111)	(0.145)	(0.199)	**(0.075)**	(0.109)	(0.144)	(0.197)
B. Rehabilitation benefits by diagnosis									
Mental and behavioral disorders	0.172	**-0.015**	0.041	0.063	0.013	**-0.022**	0.032	0.056	0.004
		(0.027)	(0.042)	(0.058)	(0.075)	**(0.030)**	(0.042)	(0.059)	(0.077)
Diseases of musculoskeletal system and connective tissue	0.396	**0.008**	-0.024	-0.026	0.118	**-0.018**	-0.034	-0.034	0.163
		(0.038)	(0.061)	(0.087)	(0.134)	**(0.039)**	(0.060)	(0.087)	(0.137)
Other diseases	0.263	**0.028**	-0.019	-0.031	-0.039	**0.000**	-0.012	-0.019	0.001
		(0.037)	(0.049)	(0.069)	(0.077)	**(0.037)**	(0.048)	(0.069)	(0.076)
C. Disability pension by diagnosis									
Mental and behavioral disorders	0.257	**-0.004**	-0.002	0.001	0.025	**0.000**	0.000	0.002	0.027
		(0.026)	(0.041)	(0.057)	(0.075)	**(0.028)**	(0.039)	(0.056)	(0.072)
Diseases of musculoskeletal system and connective tissue	0.806	**0.010**	-0.067	-0.072	-0.211*	**-0.019**	-0.065	-0.062	-0.129
		(0.050)	(0.068)	(0.090)	(0.113)	**(0.051)**	(0.067)	(0.090)	(0.110)
Other diseases	0.678	**0.037**	-0.084	-0.116	-0.109	**0.019**	-0.090	-0.119	-0.078
		(0.052)	(0.079)	(0.108)	(0.152)	**(0.053)**	(0.077)	(0.106)	(0.150)
Effect of past payroll		Linear	Quadratic	Cubic	Quartic	Linear	Quadratic	Cubic	Quartic

Notes: Average risk is the unweighted average of $\hat{\mu}_j$ across firms. Firm controls include indicators for region, industry, multiple establishments, export/import status and legal form, as well as the share of foreign ownership and log firm age. The AIC-preferred polynomial in bold. Data include 14,154 unique firms. The robust standard errors clustered at the firm level in parenthesis. Significance levels: *** 1%, ** 5% and * 10%.

reducing the bandwidth marginally until $h \approx 1.39$ in the bottom kink sample, and $h \approx 0.21$ in the top kink sample. With these minimum bandwidths, the degree of experience rating varies on the intervals $[0, 0.2)$ and $[0.8, 1]$ in the bottom and top kink samples, respectively.

Figure 9 shows the results for the bottom kink sample. The local RKD estimate of τ equals the change in the disability risk measured in percentages when the degree of experience rating jumps from 0 to 1. The average value of the disability risk in the sample with the minimum bandwidth is shown at the top-right corner. Compared to our baseline global estimates in Table 4, the local RKD estimates are often much larger in absolute value. In addition, they are rather sensitive with respect to the polynomial order and bandwidth, and they are very imprecisely estimated, especially those from the local quadratic model. Except for the local linear estimate for the incidence of sick leave, none of the estimates differ from zero at the 5% level. The local linear model implies that greater experience rating increases sickness bene-

Figure 9: RKD estimates for the effect of the time-varying degree of experience rating τ ($\times 100$) along with the cluster-robust 90% and 95% confidence intervals at the bottom kink. OLS estimation without firm controls.

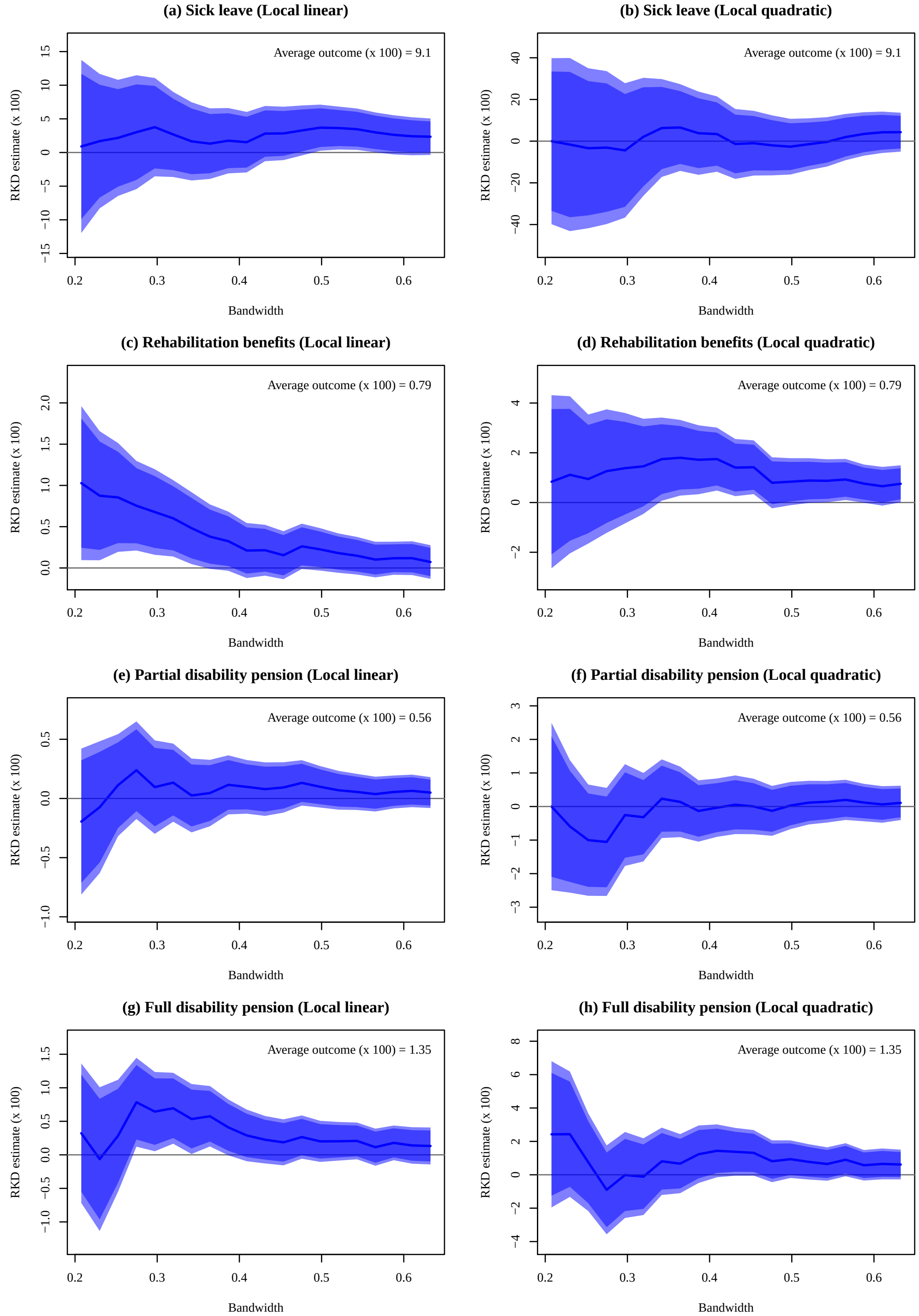


fit claims notably; the maximum increase in the degree of experience rating from 0 to 1 is estimated to roughly double the incidence of sick leave. Though much larger in size, the sign of the effect is consistent with the evidence in Figure 7a and with our estimates from the baseline linear models without firm controls (column 1, Panel A in Tables 4 and 5). However, just like the corresponding global estimates before, the local effect on sick leave appears to be particularly sensitive with respect to the polynomial order and to the inclusion of the firm covariates. As seen in Figure 9b, the effect does not differ from zero in the local quadratic model. Furthermore, also in the local linear model the effect disappears if we control for firm characteristics (Figure A.1a in the Appendix). It is evident that the local linear model without firm controls generally works poorly for the incidence of sick leave, producing the spurious positive effect for experience rating.

The local RKD estimates for the top kink sample are depicted in Figure 10. These describe the change in the disability risk due to doubling the degree of experience rating from 1 to 2. By reversing the sign of the effect, we get the change in the risk that would result from the decline in the degree of experience rating from 1 to 0, which is a more interesting and realistic case. In Figure 10, the point estimates are mostly statistically insignificant but the estimates for the receipt of rehabilitation benefits and disability pension are significantly positive for certain bandwidths. In particular, for the receipt of rehabilitation benefits the local linear model implies a positive effect of 0.5 to 1 when the bandwidth of 0.35 or less is used. The quadratic model produces somewhat larger effects, though these are significant over a narrower range of the bandwidths around the value of 0.4. If we add the firm covariates as controls, the local linear estimates loose their statistical significance while the local quadratic estimates do not (Figures A.2c and A.2d in the Appendix), so that the evidence of the positive effect on the receipt of rehabilitation benefits is not robust. Likewise, the positive effect on the disability pension inflow over a narrow range of the bandwidths in the local linear model in Figure 10g is sensitive with respect to the inclusion of firm covariates (Figure A.2g).

In summary, the local RKD results in Figures 9 and 10 do not provide any robust evidence that the degree of experience rating would affect our main outcomes. Nor does an analogous analysis for the receipt of rehabilitation benefits and disability pensions granted for a given medical condition produce significant effects for experience rating (Figures A.3–A.6 in the Appendix). An obvious problem with these results is that the estimates are very imprecise

Figure 10: RKD estimates for the effect of the time-varying degree of experience rating τ ($\times$100) along with the cluster-robust 90% and 95% confidence intervals at the top kink. OLS estimation without firm controls.



and therefore do not provide a strong test for our baseline results from the global RKD model.

Finally, it is worth noting that we have analyzed individual firms and ignored the fact that many of them are part of consolidated corporations. The experience rating status of such firms is not clear because the consolidated corporations can choose whether their DI premiums are determined at the level of individual firms – in which case the degree of experience rating varies with the payroll of individual firms – or collectively at the corporation level – in which case the degree of experience rating is based on the sum of payrolls across all member firms. Which one is less costly depends on whether the corporation's realized disability cost is below or above the average level, and thereby the choice between the two alternatives can be endogenous. However, according to our survey to the pension insurance companies, the use of collective DI premiums at corporation level is extremely rare, and hence our approach of analyzing individual firms and premiums is justified.

9 Concluding remarks

In the Finnish system, a new disability pension claimant can cause substantial costs to the former employer through an increase in the DI premium. Given the size of the potential costs, experience rating should promote preventive health and safety practices and encourage employers to accommodate their employees with work limitations. However, our findings suggest that experience rating is not succeeding in reducing sick leaves or disability benefit claims. We found no notable differences in sickness or disability inflow rates between small, medium-sized and large firms, which are subject to different degrees of experience rating and thus have differential incentives. In our regression models the degree of experience rating did not have a desired effect on the disability inflow. This finding proved to be robust in terms of the model specification and the outcome variable considered. In particular, we found no effects on the incidence of sick leave or receipt of three different types of disability benefits, nor did we find effects on the receipt of disability benefits associated with different types of medical conditions.

From the description of the institutional framework it is evident that the design of the Finnish experience rating scheme is rather complex. A long delay, of possibly several years, between a medical diagnosis of disability and a possible, but not inevitable, increase in the experience-rated premium may hinder employers from recognizing the causes for premium changes. If so, a simpler and more transparent procedure to determine the DI premiums could work better. However, even without the desired behavioral effects, the experience rating system provides a means of allocating the overall costs of the disability benefits more equitably among individual employers.

Bibliography

Autor, David H., "The unsustainable rise of the disability rolls in the United States: Causes, consequences, and policy options," NBER Working Paper 17697, National Bureau of Economic Research 2011.

_ **and Mark G. Duggan**, "The rise in the disability rolls and the decline in unemployment," *Quarterly Journal of Economics*, 2003, *118* (1), 157–205.

_ **and** _ , "The growth in the social security disability rolls: A fiscal crisis unfolding," *Journal of Economic Perspectives*, 2006, *20* (3), 71–96.

Black, Dan, Kermit Daniel, and Seth Sanders, "The impact of economic conditions on participation in disability programs: Evidence from the coal boom and bust," *American Economic Review*, 2002, *92* (1), 27–50.

Böckerman, Petri, Ohto Kanninen, and Ilpo Suoniemi, "A kink that makes you sick: The effect of sick pay on absence in a social insurance system," Working Paper 297, Palkansaajien tutkimuslaitos 2015.

Bruce, Christopher J. and Frank J. Atkins, "Efficiency effects of premium-setting regimes under workers' compensations: Canada and the United States," *Journal of Labor Economics*, 1993, *11* (1), S38–S69.

Burkhauser, Richard and Mary C. Daly, *The declining work and welfare of people with disabilities: What went wrong and a strategy for change*, American Enterprise Institute, 2011.

Campolieti, Michele, "Disability insurance benefits and labor supply: Some additional evidence," *Journal of Labor Economics*, 2004, *22* (4), 863–890.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber, "Inference on causal effects in a generalized regression kink design," *Econometrica*, 2015, *83* (6), 2453–2483.

Dahlberg, Matz, Eva Mörk, Jørn Rattsø, and Hanna Ågren, "Using a discontinuous grant rule to identify the effect of grants on local taxes and spending," *Journal of Public Economics*, 2008, *92* (12), 2320–2335.

de Groot, Nynke and Pierre Koning, "Assessing the effects of disability insurance experience rating. The case of The Netherlands," *Labour Economics*, 2016, *41*, 304–317.

García-Gómez, Pilar, Hans-Martin Gaudecker, and Maarten Lindeboom, "Health, disability and work: patterns for the working age population," *International Tax and Public Finance*, 2011, *18* (2), 146–165.

Gruber, Jonathan, "Disability insurance benefits and labor supply," *Journal of Political Economy*, 2000, *108* (6), 1162–1183.

Guryan, Jonathan, "Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts," NBER Working Paper 8269, National Bureau of Economic Research 2001.

Karlström, Anders, Mårten Palme, and Ingemar Svensson, "The employment effect of stricter rules for eligibility for DI: Evidence from a natural experiment in Sweden," *Journal of Public Economics*, 2008, *92* (10-11), 2071–2082.

Koning, Pierre, "Experience rating and the inflow into disability insurance," *De Economist*, 2009, *157* (3), 315–335.

_ **and Maarten Lindeboom**, "The rise and fall of disability insurance enrollment in the Netherlands," *Journal of Economic Perspectives*, 2015, *29* (2), 151–172.

Korkeamäki, Ossi and Tomi Kyyrä, "Institutional rules, labor demand and retirement through disability programme participation," *Journal of Population Economics*, 2012, *25* (2), 439–468.

Krueger, Alan B., "Workers' Compensation Insurance and the Duration of Workplace Injuries," NBER Working Papers 3253, National Bureau of Economic Research 1990.

Kyyrä, Tomi, "Early retirement policy in the presence of competing exit pathways: Evidence from pension reforms in Finland," *Economica*, 2014, *82* (325), 46–78.

_ **and Juha Tuomala**, "Does experience rating reduce disability inflow?," VATT Working Paper 46, VATT Institute for Economic Research 2013.

_ **and Ralf A. Wilke**, "Reduction in the long-term unemployment of the elderly: A success story from Finland," *Journal of the European Economic Association*, 2007, *5* (1), 154–182.

_ **and Tuuli Paukkeri**, "Työkyvyttömyyseläkkeiden maksuluokkamallin kannustinvaikutukset," Reports 07/2015, Finnish Centre for Pensions 2015.

Liukko, Jyri, Anu Polvinen, Meeri Kesälä, and Jarno Varis, "Työkyvyttömyyseläkkeiden maksuluokkamallin toimivuus ja kannustinvaikutukset. Kyselytutkimus työnantajien näkemyksistä (Effectiveness and incentive effects of experience rating in disability insurance. A questionnaire study on employers' views)," Studies 03/2017, Finnish Centre for Pensions (ETK) 2017.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Nielsen, Helena Skyt, Torben Sørensen, and Christopher R. Taber, "Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform," *American Economic Journal: Economic Policy*, 2010, *2* (2), 185–215.

Rothstein, Jesse and Cecilia Elena Rouse, "Constrained after college: Student loans and early career occupational choices," NBER Working Paper 13117, National Bureau of Economic Research 2007.

Ruser, John W., "Workers' compensation insurance, experience-rating, and occupational injuries," *The RAND Journal of Economics*, 1985, *16* (4), 487–503.

_ , "Workers' compensation and occupational injuries and illness," *Journal of Labor Economics*, 1991, *9* (4), 325–350.

Staubli, Stefan, "The impact of stricter criteria for disability insurance on labor force participation," *Journal of Public Economics*, 2011, *95* (9-10), 1223–1235.

Thomason, Terry and Silvana Pozzebon, "Determinants of firm workplace health and safety claims management practices," *Industrial and Labor Relations Review*, 2002, *55* (2), 286–307.

van Sonsbeek, Jan-Maarten and Raymond H.J.M. Gradus, "Estimating the effects of recent disability reforms in the Netherlands," *Oxford Economic Papers*, 2013, *65* (4), 832–855.

A Appendix

Table A.1: Within estimates of coefficients on worker-specific covariates for main outcomes

Covariate (× 100)	Sick leave (1)	Rehabilitation benefit (2)	Partial disability pension (3)	Full disability pension (4)
Female	1.216***	0.050***	0.058***	-0.018*
	(0.073)	(0.008)	(0.006)	(0.011)
Married	-0.156***	-0.032***	0.020***	-0.013*
	(0.035)	(0.007)	(0.004)	(0.008)
Upper secondary education	-1.393***	-0.113***	-0.046***	-0.244***
	(0.081)	(0.013)	(0.009)	(0.018)
Lower tertiary education	-3.793***	-0.271***	-0.118***	-0.391***
	(0.169)	(0.014)	(0.010)	(0.020)
Upper tertiary education	-4.466***	-0.302***	-0.121***	-0.381***
	(0.214)	(0.015)	(0.011)	(0.023)
Tenure - 11, years	-0.114***	-0.003***	0.002***	0.001
	(0.005)	(0.001)	(0.001)	(0.001)
Age 20–30	-4.413***	-0.611***	-0.508***	-1.083***
	(0.170)	(0.032)	(0.031)	(0.046)
Age 31–40	-3.216***	-0.472***	-0.488***	-0.998***
	(0.115)	(0.030)	(0.030)	(0.045)
Age 41–45	-2.643***	-0.379***	-0.476***	-0.951***
	(0.110)	(0.030)	(0.030)	(0.045)
Age 46–50	-1.865***	-0.266***	-0.434***	-0.858***
	(0.101)	(0.030)	(0.030)	(0.045)
Age 51	-1.335***	-0.196***	-0.395***	-0.707***
	(0.118)	(0.037)	(0.032)	(0.048)
Age 52	-1.136***	-0.100***	-0.296***	-0.622***
	(0.123)	(0.037)	(0.034)	(0.050)
Age 53	-0.929***	-0.066*	-0.270***	-0.517***
	(0.119)	(0.038)	(0.034)	(0.049)
Age 54	-0.499***	-0.034	-0.173***	-0.378***
	(0.114)	(0.040)	(0.035)	(0.048)
Age 55	-0.331***	-0.034	-0.102***	-0.235***
	(0.111)	(0.039)	(0.034)	(0.050)
Age 57	0.102	-0.033	0.016	0.339***
	(0.148)	(0.042)	(0.043)	(0.074)
Age 58	0.168	-0.076*	0.156***	0.614***
	(0.156)	(0.041)	(0.046)	(0.074)
Age 59	0.138	-0.232***	-0.003	0.813***
	(0.158)	(0.042)	(0.045)	(0.077)
Age 60	0.120	-0.304***	-0.030	1.079***
	(0.170)	(0.041)	(0.045)	(0.085)
Age 61-62	-0.134	-0.509***	-0.287***	0.546***
	(0.153)	(0.037)	(0.041)	(0.072)
$UT \times S$	0.349	-0.006	0.105**	-0.102
	(0.239)	(0.033)	(0.047)	(0.116)
R-squared	0.044	0.021	0.020	0.029

Notes: The reference worker is a 56-year-old unmarried man with a lower than upper secondary education and 11 years of job tenure. Number of observations is 3,119,133. The robust two-way clustered standard errors in parenthesis. Significance levels: *** 1%, ** 5% and * 10%.

Table A.2: Within estimates of coefficients on worker-specific covariates for disability benefits by diagnosis

	Rehabilitation benefit			Partial or full disability pension		
	Mental and behavioral disorder	Diseases of muscu- loskeletal system and connective tissue	Other diseases	Mental and behavioral disorder	Diseases of muscu- loskeletal system and connective tissue	Other diseases
Covariate (× 100)	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.024***	0.024***	0.002	0.007	0.057***	-0.040***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.007)	(0.007)
Married	-0.017***	-0.004	-0.011***	-0.020***	0.029***	-0.012**
	(0.004)	(0.005)	(0.004)	(0.003)	(0.006)	(0.005)
Upper secondary education	-0.013**	-0.064***	-0.036***	-0.019***	-0.144***	-0.102***
	(0.006)	(0.008)	(0.007)	(0.006)	(0.013)	(0.011)
Lower tertiary education	-0.044***	-0.141***	-0.086***	-0.039***	-0.267***	-0.154***
	(0.006)	(0.009)	(0.007)	(0.007)	(0.016)	(0.011)
Upper tertiary education	-0.065***	-0.142***	-0.095***	-0.053***	-0.246***	-0.153***
	(0.007)	(0.008)	(0.007)	(0.008)	(0.016)	(0.012)
Tenure - 11, years	-0.001***	-0.001***	-0.001	0.001	0.001	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Age 20–30	-0.110***	-0.317***	-0.184***	-0.201***	-0.682***	-0.519***
	(0.014)	(0.024)	(0.017)	(0.016)	(0.040)	(0.028)
Age 31–40	-0.077***	-0.256***	-0.139***	-0.184***	-0.639***	-0.480***
	(0.014)	(0.023)	(0.017)	(0.016)	(0.038)	(0.028)
Age 41–45	-0.058***	-0.222***	-0.099***	-0.170***	-0.630***	-0.442***
	(0.014)	(0.022)	(0.016)	(0.016)	(0.038)	(0.028)
Age 46–50	-0.038***	-0.161***	-0.066***	-0.142***	-0.590***	-0.390***
	(0.014)	(0.021)	(0.016)	(0.016)	(0.038)	(0.028)
Age 51	-0.023	-0.125***	-0.049**	-0.113***	-0.513***	-0.312***
	(0.018)	(0.027)	(0.022)	(0.018)	(0.039)	(0.031)
Age 52	0.018	-0.094***	-0.024	-0.069***	-0.450***	-0.272***
	(0.019)	(0.026)	(0.021)	(0.020)	(0.039)	(0.032)
Age 53	0.013	-0.053**	-0.026	-0.062***	-0.389***	-0.220***
	(0.018)	(0.026)	(0.023)	(0.020)	(0.039)	(0.034)
Age 54	0.009	-0.035	-0.009	-0.041**	-0.275***	-0.140***
	(0.018)	(0.029)	(0.022)	(0.021)	(0.042)	(0.033)
Age 55	-0.001	-0.039	0.006	-0.038*	-0.177***	-0.078**
	(0.019)	(0.026)	(0.023)	(0.021)	(0.040)	(0.037)
Age 57	-0.017	-0.025	0.008	0.047*	0.160***	0.107**
	(0.019)	(0.030)	(0.024)	(0.026)	(0.061)	(0.045)
Age 58	-0.028	-0.012	-0.036	0.044*	0.361***	0.264***
	(0.019)	(0.030)	(0.024)	(0.026)	(0.062)	(0.046)
Age 59	-0.068***	-0.151***	-0.013	0.049*	0.363***	0.399***
	(0.018)	(0.028)	(0.024)	(0.026)	(0.059)	(0.050)
Age 60	-0.095***	-0.147***	-0.062**	0.067**	0.617***	0.441***
	(0.016)	(0.030)	(0.024)	(0.028)	(0.064)	(0.053)
Age 61-62	-0.113***	-0.244***	-0.151***	-0.031	0.205***	0.281***
	(0.015)	(0.027)	(0.019)	(0.022)	(0.060)	(0.044)
$UT \times S$	-0.010	-0.013	0.017	0.001	-0.024	-0.035
	(0.010)	(0.021)	(0.017)	(0.022)	(0.092)	(0.050)
R-squared	0.019	0.020	0.019	0.018	0.025	0.022

Notes: The reference worker is a 56-year-old unmarried man with a lower than upper secondary education and 11 years of job tenure. Number of observations is 3,119,133. The robust two-way clustered standard errors in parenthesis. Significance levels: *** 1%, ** 5% and * 10%.

Figure A.1: RKD estimates for the effect of the time-varying degree of experience rating τ ($\times$100) along with the cluster-robust 90% and 95% confidence intervals at the bottom kink. OLS estimation with firm controls.

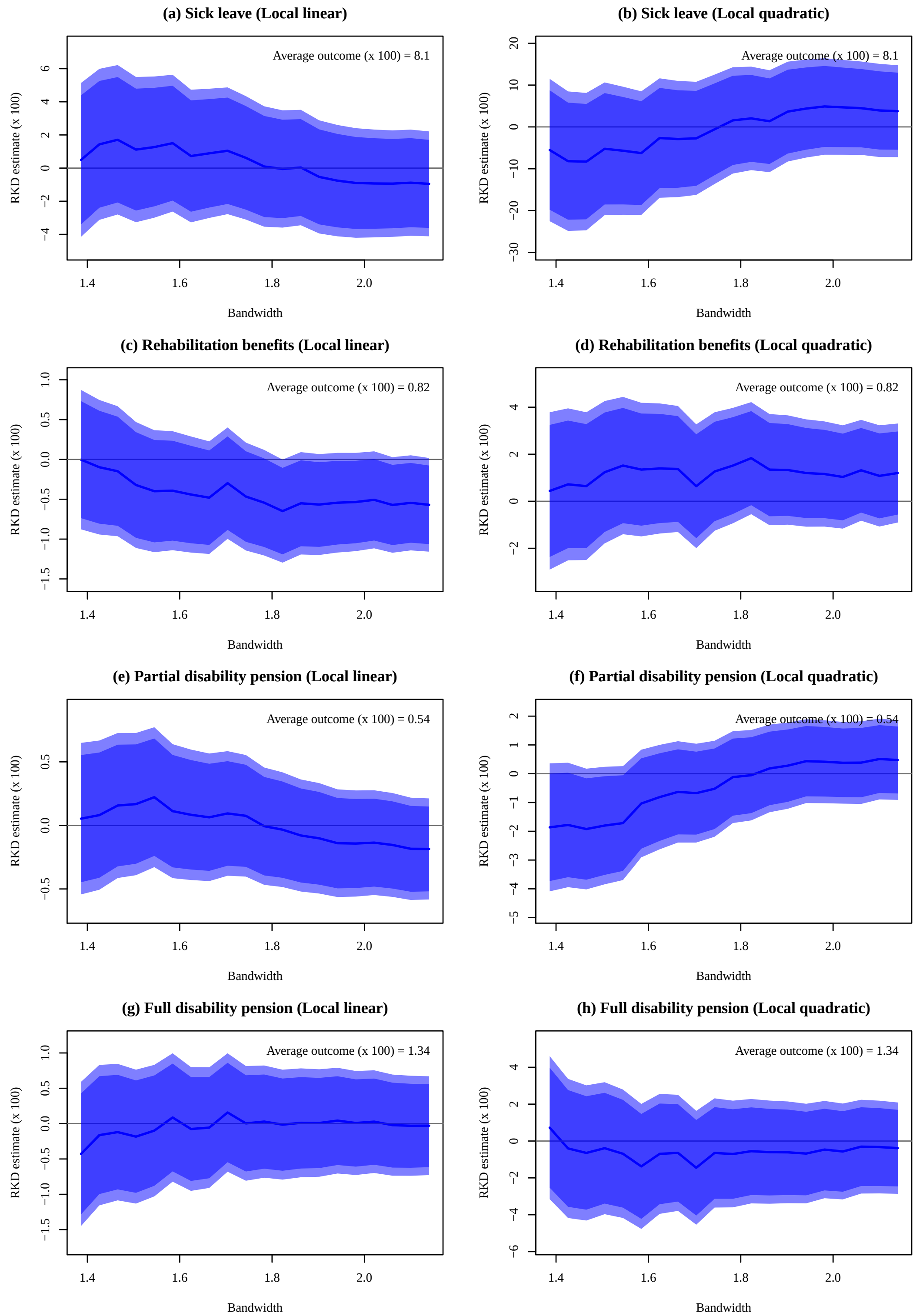


Figure A.2: RKD estimates for the effect of the time-varying degree of experience rating τ ($\times 100$) along with the cluster-robust 90% and 95% confidence intervals at the top kink. OLS estimation with firm controls.

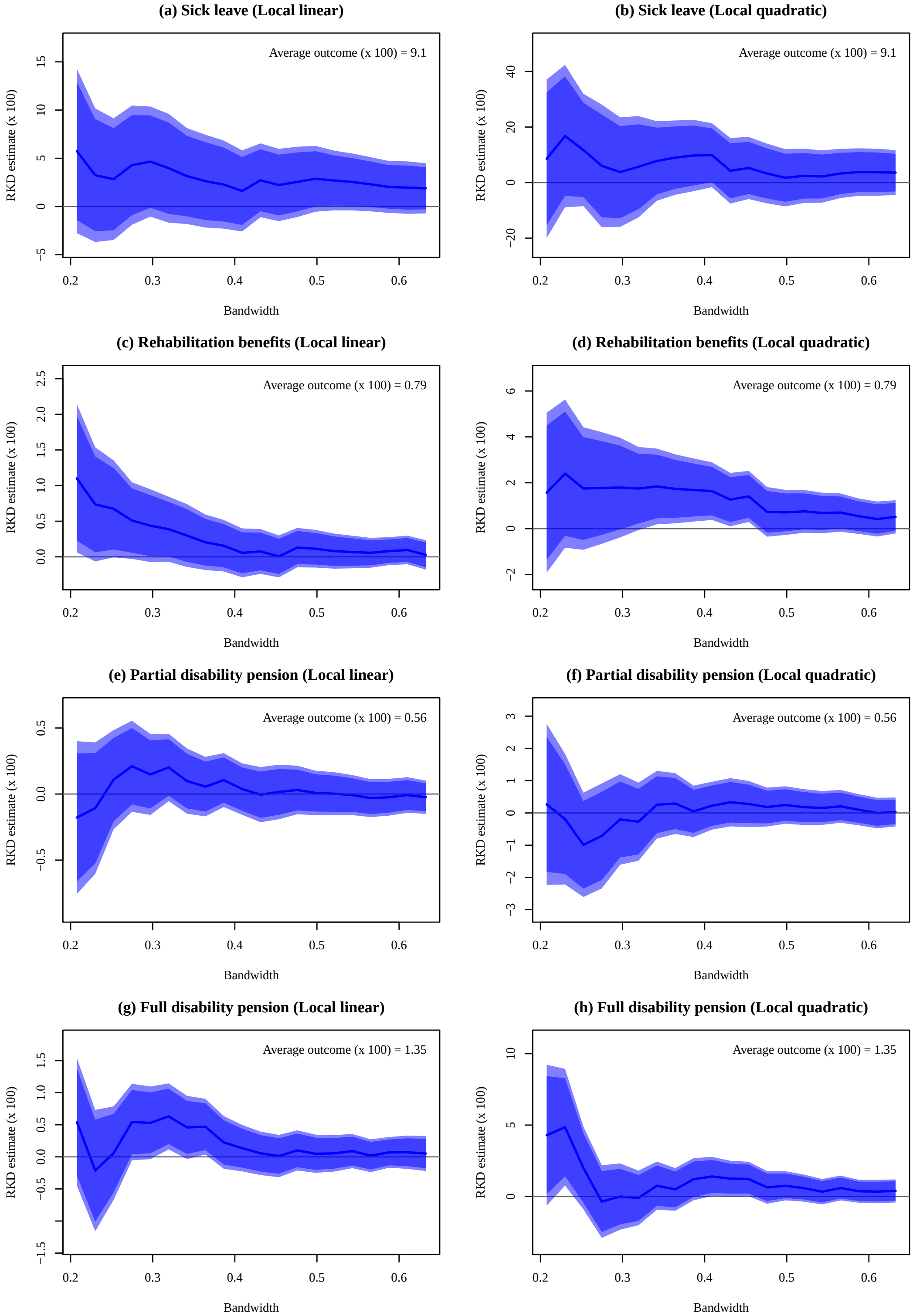


Figure A.3: RKD estimates for the effect of the time-varying degree of experience rating τ ($\times 100$) on the incidence of rehabilitation benefits by diagnosis along with the cluster-robust 90% and 95% confidence intervals at the bottom kink. OLS estimation without firm controls.

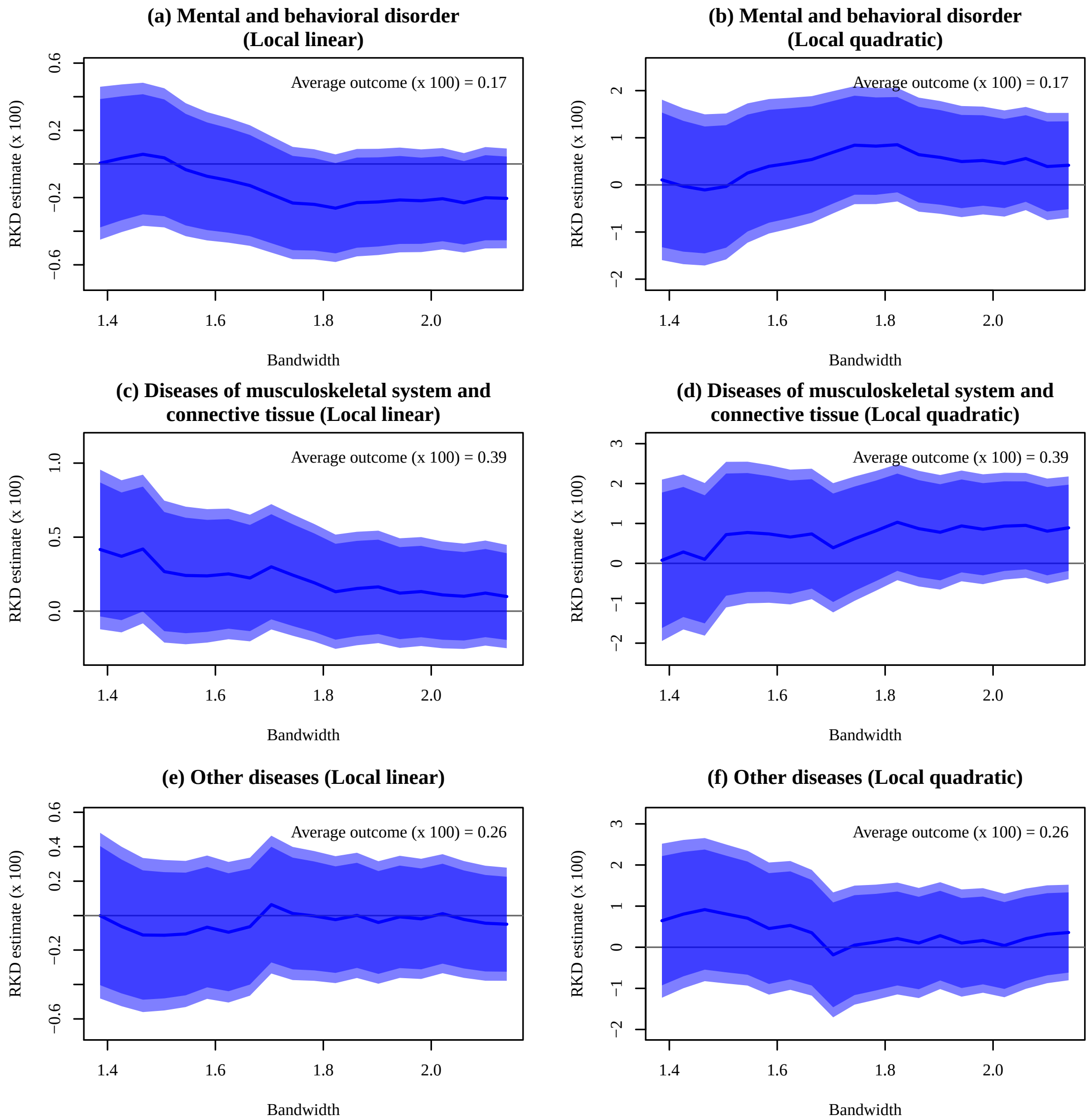


Figure A.4: RKD estimates for the effect of the time-varying degree of experience rating τ ($\times$100) on the incidence of disability pension by diagnosis along with the cluster-robust 90% and 95% confidence intervals at the bottom kink. OLS estimation without firm controls.

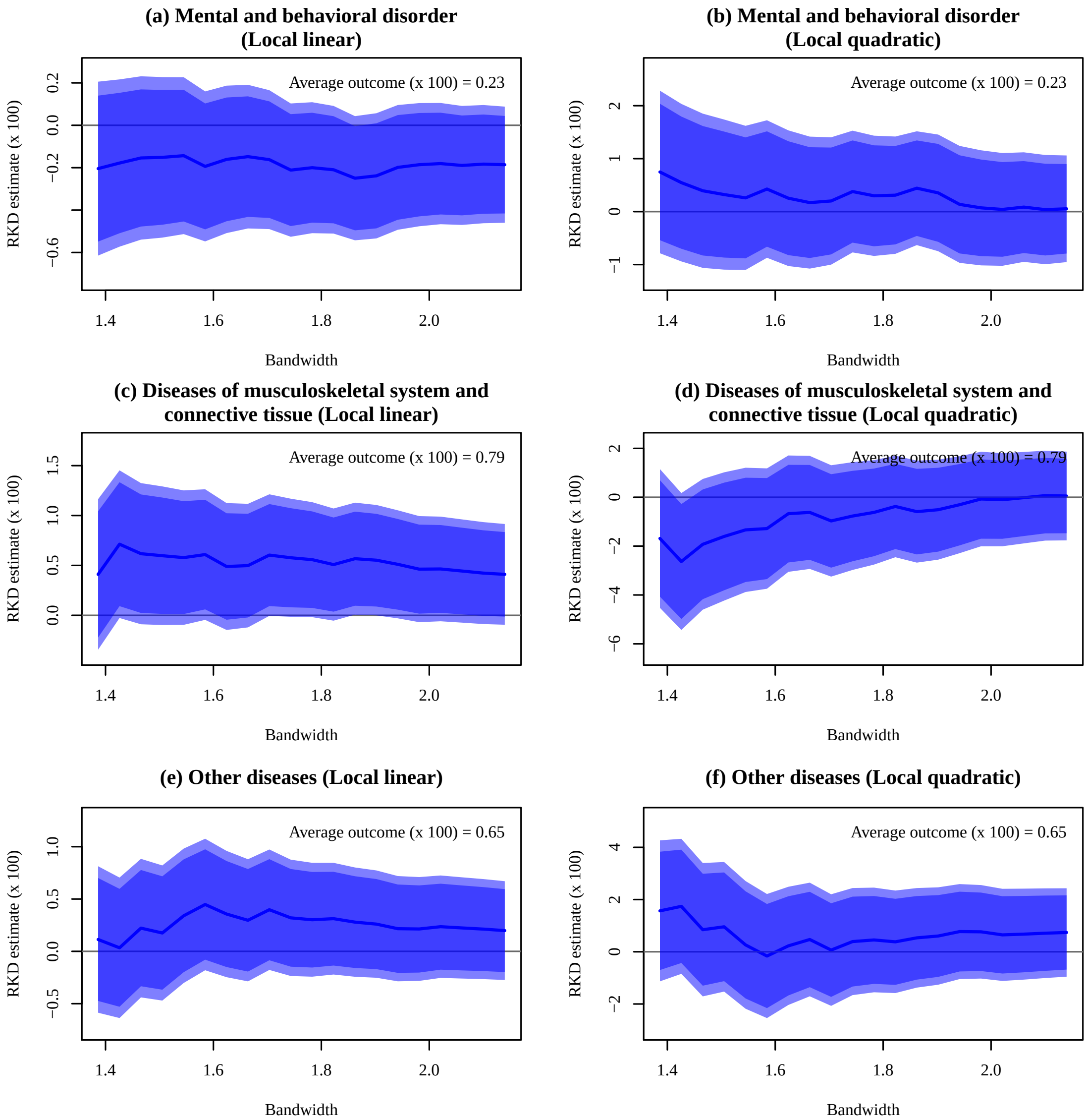


Figure A.5: RKD estimates for the effect of the time-varying degree of experience rating τ ($\times 100$) on the incidence of rehabilitation benefits by diagnosis along with the cluster-robust 90% and 95% confidence intervals at the top kink. OLS estimation without firm controls.

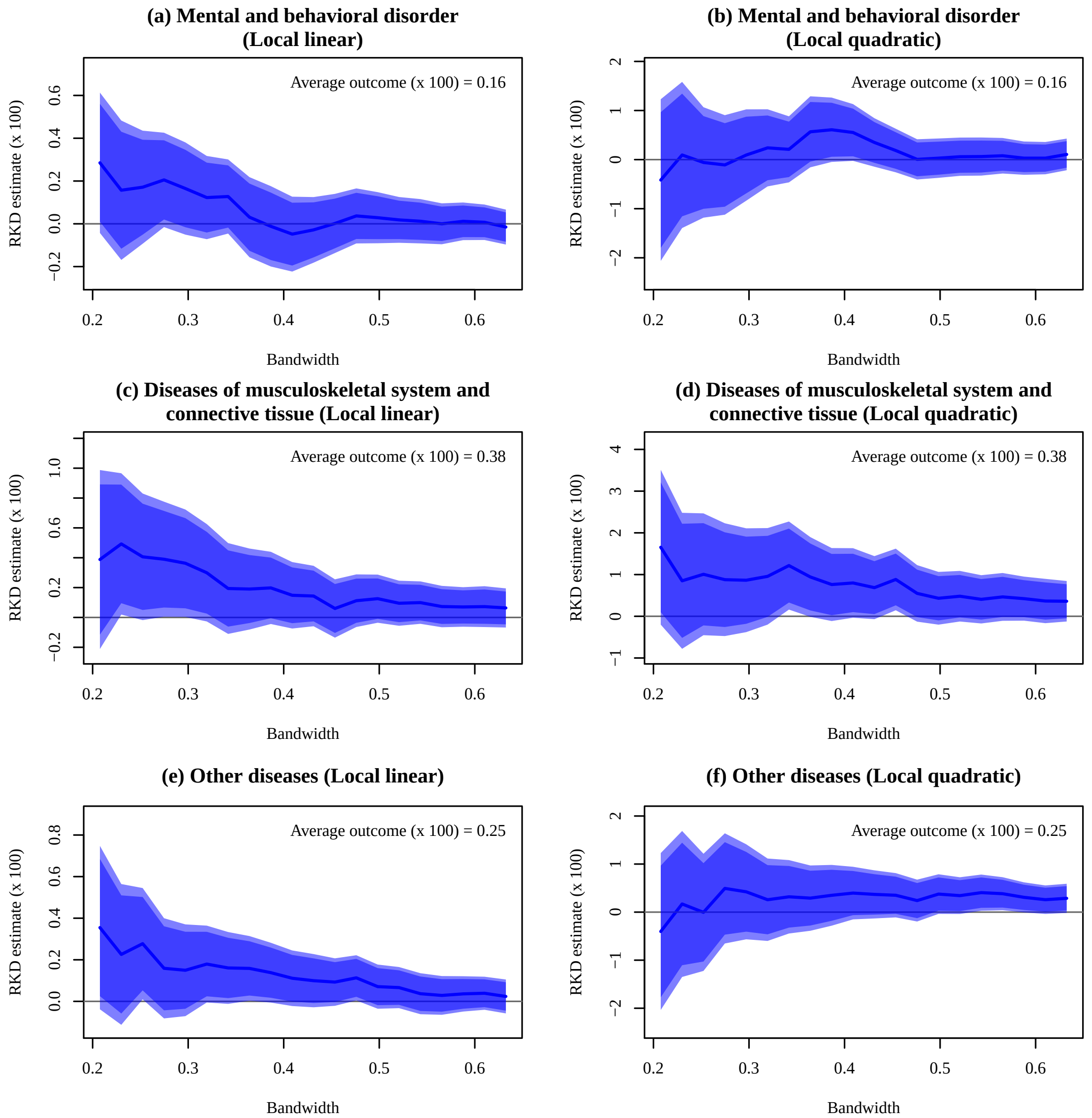
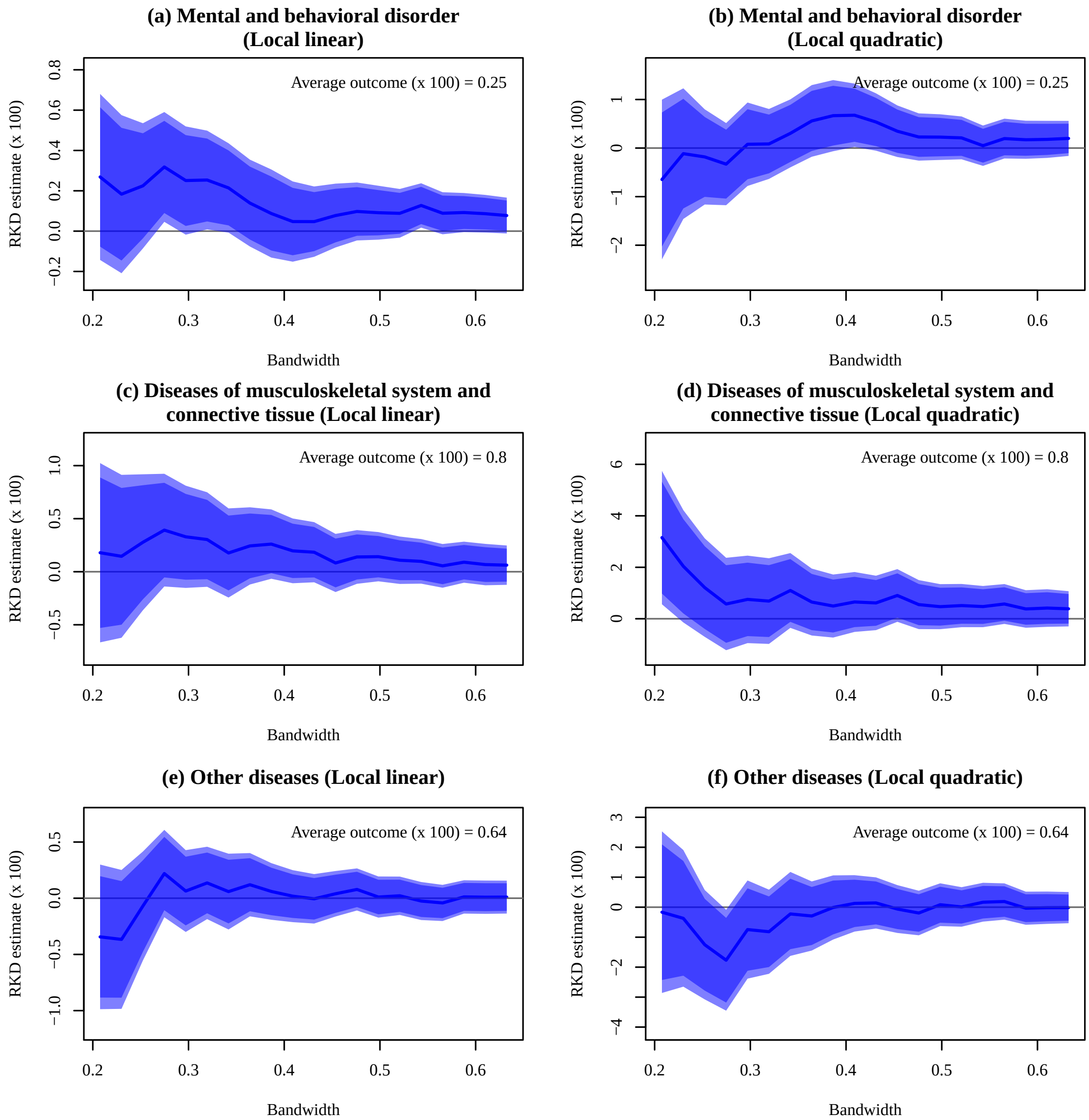


Figure A.6: RKD estimates for the effect of the time-varying degree of experience rating τ ($\times$100) on the incidence of disability pension by diagnosis along with the cluster-robust 90% and 95% confidence intervals at the top kink. OLS estimation without firm controls.



Int Tax Public Finance (2018) 25:64–98
https://doi.org/10.1007/s10797-017-9443-6

Optimal taxation and public provision for poverty reduction

Ravi Kanbur[1] · **Tuuli Paukkeri**[2,3] ·
Jukka Pirttilä[4,5] · **Matti Tuomala**[5]

Published online: 27 April 2017

Abstract The existing literature on optimal taxation typically assumes there exists a capacity to implement complex tax schemes, which is not necessarily the case for many developing countries. We examine the determinants of optimal redistributive policies in the context of a developing country that can only implement linear tax policies due to administrative reasons. Further, the reduction of poverty is typically the expressed goal of such countries, and this feature is also taken into account in our model. We derive the optimality conditions for linear income taxation, commodity taxation, and public provision of private and public goods for the poverty minimization case and compare the results to those derived under a general welfarist objective function. We also study the implications of informality on optimal redistributive policies for such countries. The exercise reveals non-trivial differences in optimal tax rules under the different assumptions.

✉ Tuuli Paukkeri
tuuli.paukkeri@vatt.fi

Ravi Kanbur
sk145@cornell.edu

Jukka Pirttilä
jukka@wider.unu.edu

Matti Tuomala
matti.tuomala@uta.fi

1 Cornell University, Ithaca, NY, USA

2 VATT Institute for Economic Research, Helsinki, Finland

3 Aalto University, Helsinki, Finland

4 UNU-WIDER, Helsinki, Finland

5 University of Tampere, Tampere, Finland

Keywords Redistribution · Income taxation · Commodity taxation · Public good provision · Poverty

JEL Classification H21 · H40 · O12

1 Introduction

High levels of within-country inequality in many otherwise successful developing countries have become a key policy concern in global development debate. While some countries have very unequal inherent distributions (e.g., due to historical land ownership arrangements), in others the fruits of economic growth have been unequally shared. No matter what the underlying reason for the high inequality, often the only direct way for governments to affect the distribution of income is via redistributive tax and transfer systems. Clearly, public spending on social services also has an impact on the distribution of well-being, although some of the effects (such as skill-enhancing impacts from educational investment) only materialize over a longer time horizon.

Reflecting the desire to reduce poverty and inequality, redistributive transfer systems have, indeed, proliferated in many developing countries. Starting from Latin America, they are now spreading to low-income countries, including those in Sub-Saharan Africa.[1] In low-income countries, in particular, redistributive arrangements via transfers are still at an early stage, and they often consist of isolated, donor-driven programs. There is an urgent and well-recognized need to move away from scattered programs to more comprehensive tax-benefit systems.

This paper examines the optimal design of cash transfers, commodity taxes (or subsidies), the provision of public and private goods (such as education and housing), and financing them by a linear income tax. The paper also includes an analysis of optimal income taxation in the presence of an informal sector. The paper therefore provides an overview of many of the most relevant instruments for redistributive policies that are needed for a system-wide analysis of social protection. We build on the optimal income tax approach, which is extensively used in the developed country context[2], but much less applied for the design of redistributive systems in developing country circumstances. This approach, initiated by Mirrlees (1971), allows for a rigorous treatment of efficiency concerns (e.g., the potentially harmful effect of distortionary taxation on employment) and redistributive objectives. Achieving the government's redistributive objectives is constrained by limited information: the social planner cannot directly observe individuals' income-earning capacity, and therefore it needs to base its tax and transfer policies on observable variables, such as gross income. The most general formulations of optimal tax models apply nonlinear tax schedules, but in a developing country context, using fully nonlinear taxes is rarely feasible. In this paper, we therefore limit the analysis to redistributive linear income taxes, which combine a

[1] For a recent treatment and survey, see Barrientos (2013).

[2] See IFS and Mirrlees (2011) for an influential application of optimal tax theory to policy analysis for rich countries.

lump-sum transfer with a proportional income tax, and which can be implemented by withholding at source if necessary.

Linear income taxes are not very common in practice: less than 30 countries had flat tax rates for personal income in 2012, with some concentration in ex-Soviet Eastern Europe (Peichl 2014). It is noteworthy that even though flat taxes are not particularly common in low-income countries, in many instances in such countries the progressive income tax reaches only a small share of the population. This would indicate that despite the existence of a progressive income tax, these countries do not yet possess enough tax capacity to implement well-functioning progressive income taxes. This is one motivation for our interest of modeling optimal linear taxes. Peichl (2014) suggests that simplification benefits can be especially relevant for developing countries.[3]

In conventional optimal taxation models, the government's objective function is modeled as a social welfare function, which depends directly on individual utilities. We depart from this welfarist approach by presenting general non-welfarist tax rules, as in Kanbur et al. (2006), and, in particular, optimal tax and public good provision rules when the government is assumed to minimize poverty. We have chosen this approach as it resembles well the tone of much of the policy discussion in developing countries, including the Millennium Development Goals (MDGs) and the new Sustainable Development Goals (SDGs), where the objective is explicitly to reduce poverty rather than maximize well-being.[4] Similarly, the discussion regarding cash transfer systems is often couched especially in terms of poverty alleviation. While we do not necessarily want to advocate poverty minimization over other social objectives, we regard examining its implications, and contrasting them with traditional welfaristic approaches, useful. Using non-welfarist objectives is, as such, nothing new in economics. In fact, as Sen (1985) has argued, one can be critical of utilitarianism for many reasons. Note also that the objective of poverty minimization is not at odds with the restriction of a linear tax scheme that we impose: a flat tax regime together with a lump-sum income transfer component can achieve similar amounts of redistribution toward the poor as a progressive tax system, if specified suitably (Keen et al. 2008; Peichl 2014). In all our analysis, we first present welfarist tax rules (which are mostly already available in the literature) to provide a benchmark to examine how applying poverty minimization as an objective changes the optimal tax and public service provision rules.

We also deal with some extensions to existing models, which are motivated by the developing country context, such as the case where public provision affects the individuals' income-earning capacity, thus capturing (albeit in a very stylized way) possibilities to affect their capabilities. An important feature to take into account in tax analysis of developing countries is the presence of a large informal sector, and we also examine the implications of this for optimal redistributive policies.

Our paper is related to various strands of earlier literature. First, Kanbur et al. (1994) and Pirttilä and Tuomala (2004) study optimal income tax and commodity tax rules, respectively, from the poverty alleviation point of view, but their papers build on the nonlinear tax approach which is not well suited to developing countries. Kanbur and

[3] Note that it might be reasonable for some countries to move to a progressive income tax system as their tax capacity increases with development; the study of such dynamics is beyond the scope of this paper.

[4] In fact, the first SDG is simply "End poverty in all of its forms everywhere."

Keen (1989) do consider linear income taxation together with poverty minimization, but they do not produce optimal tax rules but focus on a tax reform perspective, and provide tax rate simulations. Others have considered different departures from the welfarist standard. For example, Fleurbaey and Maniquet (2007) consider fairness as an objective of the tax-transfer system and its implications on optimal taxation. Roemer et al. (2003) employ a maximin type of social goal and characterize how well tax and transfer systems achieve the goal of equality of opportunity. Second, our work is related to new contributions in behavioral public finance, which address the situation where the behavioral biases of the individuals lead the social planner to adopt a different objective function than the individuals have; see Chetty (2015), Gerritsen (2016), Farhi and Gabaix (2015). A third strand of literature considers taxation and development more generally, such as Gordon and Li (2009), Keen (2009, 2012), Bird and Gendron (2007) and Besley and Persson (2013).[5] This field, while clearly very relevant, has not concentrated much on the design of optimal redistributive systems. Finally, optimal linear income taxation has been studied from the standard welfarist perspective. We describe these models in Sect. 2.1. The most recent description of linear income tax models can be found in Piketty and Saez (2013). They also emphasize how linear tax rules, while analytically more feasible, provide the same intuition as the more complicated nonlinear models. The linear tax rules, they argue, are robust to alternative specifications[6], and examining this forms part of our motivation: we study optimal linear tax policies, in our understanding for the first time, from the poverty minimization perspective.

The paper proceeds as follows. Section 2 examines optimal linear income taxation, while Sect. 3 turns to optimal provision rules for publicly provided private and public goods that are financed by such a linear income tax. Section 4 analyzes the combination of optimal linear income taxes and commodity taxation and asks under which conditions one should use differentiated commodity taxation if the government is interested in poverty minimization and also has optimal cash transfers at its disposal. The question of how optimal poverty-minimizing income tax policies are altered in the presence of an informal sector is examined in Sect. 5, whereas Sect. 6 presents a numerical illustration of optimal income taxation for poverty minimization. Finally, conclusions are provided in Sect. 7.

2 Linear income taxation

2.1 Optimal linear income taxation under the welfarist objective

In this section, we give an overview of some of the models and results for optimal linear income taxation as they have been presented in the literature. Many formulae for optimal taxation were developed in the 1970s and 1980s (see Dixit and Sandmo 1977;

[5] Besley and Persson (2013) use a model with *groups* that can differ in their income-earning abilities. Their analysis focuses, however, on explaining how economic development and tax capacity are interrelated, and not on redistribution between *individuals*.

[6] They also describe some implications of departures from the welfarist standard in the optimal nonlinear tax model.

Tuomala 1985 and the survey by Tuomala 1990), and they are still being used, whereas Piketty and Saez (2013) offer fresh expressions of the tax rules. Our exposition mainly follows that of Tuomala (1985), but Appendix 1 shows how the results relate to those in Piketty and Saez (2013).

The government collects a linear income tax τ, which it uses to finance a lump-sum transfer b, along with other exogenous public spending R. The individuals differ in their income-earning capacity (w^i), and z^i denotes individual labor income ($w^i L^i$, where L^i represents hours worked). Consumption equals $c^i = (1-\tau)z^i + b$, where the superscript-i refers to individuals.[7] There is a discrete distribution of N individuals, whose heterogeneous preferences over consumption and labor are captured by the utility function $u^i(c^i, z^i)$. The maximized (subject to the individual budget constraint) value of this utility function is captured by the indirect utility function, which is denoted by $V^i(1-\tau, b)$, and we refer to the net-of-tax rate as $1-\tau = a$. To simplify notation, subscript-a refers to the derivative with respect to the net-of-tax rate.

The government has redistributive objectives represented by a Bergson–Samuelson function $W\left(V^1, \ldots, V^N\right)$ with $W' > 0$, $W'' < 0$. The government's problem is to choose the tax rate τ and transfer b so as to maximize the social welfare function $\sum W\left(V^i(a, b)\right)$ under the budget constraint $(1-a)\sum z^i = Nb + R$.[8] We denote the social marginal utility of income by $\beta^i = W_V V_b^i$.

All the mathematical details are presented in Appendix 1. There it is shown that the optimal tax rule is given by

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\varepsilon}\left(1 - \frac{z(\beta)}{\bar{z}}\right), \tag{1}$$

where $\varepsilon = \frac{\mathrm{d}\bar{z}}{\mathrm{d}(1-\tau)}\frac{(1-\tau)}{\bar{z}}$ is the elasticity of total income with respect to the net-of-tax rate, $\bar{z}$ is average income and $z(\beta) = \frac{\sum \beta^i z^i}{\sum \beta^i}$ welfare-weighted average income. Define $\Omega = \frac{z(\beta)}{\bar{z}}$, so that $I = 1 - \Omega$ is a normative measure of inequality or, equivalently, of the relative distortion arising from the second-best tax system. Clearly Ω should vary between zero and unity. One would expect it to be a decreasing function of τ (given the per capita revenue requirement $g = \frac{R}{N}$). There is a minimum feasible level of τ for any given positive g, and of course g must not be too large, or no equilibrium is possible. Hence any solution must also satisfy $\tau > \tau_{\min}$ if the tax system is to be progressive. That is, if the tax does not raise sufficient revenue to finance the non-transfer expenditure, R, the shortfall must be made up by imposing a poll tax ($b < 0$) on each individual. One would also expect the elasticity of labor supply with respect to the net-of-tax rate to be an increasing function of τ (it need not be).

We can rewrite (1) as $\tau^* = \frac{1-\Omega}{1-\Omega+\varepsilon}$ to illustrate the basic properties of the optimal tax rate. Because $\varepsilon \geq 0$ and $0 \leq \Omega < 1$, both the numerator and denominator are nonnegative. The optimal tax rate is thus between zero and one. The formula captures

7 We consider "income" here as the labor income of individuals, but considering that our model is intended especially for the poorer countries, agricultural income could as well be included in the concept of income. In Sect. 5 we discuss the implications of untaxed home consumption in agricultural production.

8 Summation is always over all individuals i, which is suppressed for simplification.

neatly the efficiency-equity trade-off. τ decreases with ε and Ω, and we have the following general results: (1) In the extreme case where $\Omega = 1$, i.e., the government does not value redistribution at all, $\tau = 0$ is optimal. We can call this case libertarian. According to the libertarian view, the level of disposable income is irrelevant (ruling out both basic income b, and other public expenditures, g, funded by the government). (2) If there is no inequality, then again $\Omega = 1$ and $\tau = 0$. There is no intervention by the government. The inherent inequality will be fully reflected in the disposable income. Furthermore, lump-sum taxation is optimal; $b = -g$ or $T = -b$. (3) We can call the case where $\Omega = 0$ as "Rawlsian" or maximin preferences. The government maximizes tax revenue (optimal $\tau = \frac{1}{1+\varepsilon}$) as it maximizes the basic income b (assuming the worst off individual has zero labor income). In fact, maximizing b can be regarded as a non-welfarist case, which is the focus in the next subsection.

2.2 Optimal linear income taxation under non-welfarist objectives

A non-welfarist government is one that follows a different set of preferences than those employed by individuals themselves (Kanbur et al. 2006). Thus, instead of maximizing a function of individual utilities, the government has other, paternalistic objectives that go beyond utilities. A special case taken up in more detail below is the objective of minimizing poverty in the society. To be as general as possible, let us define a "social evaluation function" (as in, e.g., Kanbur et al. 2006) as $S = \sum F(c^i, z^i)$, which the government maximizes instead of the social welfare function. $F(c^i, z^i)$ measures the social value of consumption c^i for a person with income z^i and can be related to $u(c^i, z^i)$ but is not restricted to it. Following Tuomala's model as above, given the instruments available, linear income tax τ, lump-sum grant b and other expenditure R the government thus maximizes $\sum F(az^i + b, z^i)$ subject to the budget constraint $(1-a)\sum z^i - Nb = R$. Define

$$\frac{\sum \left(F_c(z^i + az^i_a) + F_z z^i_a\right)}{\sum \left(F_c(1 + az^i_b) + F_z z^i_b\right)} \equiv \tilde{F}, \tag{2}$$

which reflects the relative impact of taxes and transfers on the social evaluation function. Using this definition, and following the same steps as in the previous section (see Appendix), the optimal tax rate becomes:

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\varepsilon}\left(1 - \frac{\tilde{F}}{\bar{z}}\right). \tag{3}$$

The result resembles the welfarist tax rule in (1). In addition to labor supply considerations via the term $\frac{1}{\varepsilon}$, they both entail a term that measures the relative benefits of taxes and transfers, in the welfarist case via welfare-weighted income, in the non-welfarist case via $\tilde{F}$, the relative impact on the social evaluation function. Note that since under non-welfarism individuals are not necessarily at their utility optimum, the envelope condition does not apply and thus the behavioral responses z^i_a and z^i_b are not cancelled out in $\tilde{F}$. That is, the impacts of tax changes on labor supply are not trivial under non-welfarism. The terms $\sum z_a^i (F_c a + F_z)$ in the numerator and $\sum z_b^i (F_c a + F_z)$ in the denominator of (2) capture these effects on the social evaluation function. If taxation had no behavioral impacts ($z_a^i = z_b^i = 0$), it would affect the value of the social evaluation function only by mechanically altering individual after-tax income. Note that in this case, $\tilde{F} = \frac{\sum F_c z^i}{\sum F_c}$ would be a more direct equivalent to $z(\beta) = \frac{\sum \beta^i z^i}{\sum \beta^i}$. The same equivalence would be achieved also when $F_c a + F_z = 0$, that is, the social marginal rate of substitution between income and consumption equals the private rate: $-\frac{F_z}{F_c} = a = -\frac{u_z^i}{u_c^i}$ (the latter is obtained from the individual's first-order condition). In these cases, $\tilde{F}$ would be a purely redistributive term, albeit a non-welfaristic one. Paternalistic concerns additionally enter the optimal tax rule via labor supply changes, captured by the response of z. In this way, the tax rule in (3) can be decomposed, and this decomposition is similar in spirit to the corrective parts of the tax formulae in the new optimal tax literature with behavioral agents, such as Farhi and Gabaix (2015) and Gerritsen (2016).

The signs and magnitudes of F_c and F_z and thus of $\tilde{F}$ depend on the specific objective of the government, that is, on the shape of F. Let us consider the specific case of poverty minimization below.

2.2.1 Special case: poverty minimization

Now let us derive the optimal linear tax results for a government whose objective is to minimize poverty in society. The instruments available to the government are the same, τ and b, and other exogenous expenditure is R. Note first that the revenue-maximizing tax rate is in fact equivalent to the tax rate obtained from a maximin objective function, since when the government only cares about the poverty (consumption) of the poorest individual, its only goal is to maximize redistribution to this individual, i.e., maximize tax revenue.

Let us first define the objective function of the government explicitly. Poverty is defined as deprivation of individual consumption c^i relative to some desired level $\bar{c}$ and measured with a deprivation index $D\left(c^i, \bar{c}\right)$, such that $D > 0\ \forall\, c \in [0, \bar{c})$ and $D = 0$ otherwise, and $D_c < 0,\ D_{cc} \geq 0\ \forall\, c \in [0, \bar{c})$, as in Pirttilä and Tuomala (2004). A typical example of such an index would be the P_α family of Foster–Greer–Thorbecke (FGT) poverty indices. We discuss the application of FGT indices in our model in Appendix 2. Note, however, that the choice of poverty index depends on the preferences of the government, whether they wish to minimize the total amount of deprivation in the society, or are for instance concerned especially about the incomes of the poorest of the poor. The social evaluation function $F(c^i, z^i)$ becomes $D\left(c^i, \bar{c}\right)$ and the objective function is min $P = \sum D\left(c^i, \bar{c}\right)$. Now $F_c = D_c$ and $F_z = 0$, so

$$\tilde{F} = \tilde{D} = \frac{\sum D_c \left(z^i + a z_a^i\right)}{\sum D_c \left(1 + a z_b^i\right)}, \tag{4}$$

and the optimal tax rule becomes:

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\varepsilon}\left(1 - \frac{\tilde{D}}{\bar{z}}\right). \tag{5}$$

Since now $F_z = 0$, the result is closer to (1) than (3) was, although part of the labor supply impacts still remain. Here $\tilde{D}$ describes the relative efficiency of taxes and transfers in reducing deprivation. Both the numerator and denominator of $\tilde{D}$ depend on D_c, so the difference in the relative efficiency of the two depends on z_a^i and z_b^i. The more people react to taxes (relative to transfers) by earning less, the higher is $\tilde{D}$ and the lower should the tax rate be. In (1), the higher is the social value of income, the higher is $z(\beta)$ and the lower should the tax rate be.

Since the form of the result is similar in the welfarist and the poverty minimization cases, the analysis could be also seen as a special case of the argument in Saez and Stantcheva (2016), who derived generalized social welfare weights and express the tax formulae in terms of those.[9] Here, the generalized social welfare weight would thus be derived from a poverty minimization objective. It could be close to a suitably defined welfarist criterion, and clearly it would be exactly the same only if the welfarist criterion would correspond to the chosen poverty minimization objective.

We can also rewrite $\tilde{D}$, using $a = 1 - \tau$, as: $\frac{\sum D_c\left(z^i + (1-\tau)\frac{\partial z^i}{\partial(1-\tau)}\right)}{\sum D_c\left(1+(1-\tau)z_b^i\right)} = \frac{\sum D_c\left(1+\frac{(1-\tau)}{z^i}\frac{\partial z^i}{\partial(1-\tau)}\right)z^i}{\sum D_c\left(1+(1-\tau)z_b^i\right)} = \frac{\sum D_c\left(1+\varepsilon^i\right)z^i}{\sum D_c\left(1+(1-\tau)z_b^i\right)}$. Thus the $\tilde{D}$ in the optimal tax result (5) entails a further consideration that depends on labor supply responses. It combines paternalistic preferences—how much poverty is reduced—with the behavioral responses to a tax system—how much labor income increases when the take-home pay goes up. The latter effect tends to lower the optimal tax rate to induce the poor to work more. Kanbur et al. (1994) find a similar result in their nonlinear poverty-minimizing tax model. Here, however, we are restricted to lower the tax on everyone instead of only the poorest individuals.

To summarize, the non-welfarist tax rules differ from the welfarist ones, depending on the definition of non-welfarism in question (the F_c and F_z terms). However, when we take poverty minimization as the specific case of non-welfarism, the tax rules are quite similar to welfarist ones. The basic difference is that equity is not considered in welfare terms but in terms of poverty reduction effectiveness. A more notable difference arises from efficiency considerations. With linear taxation, taking into account labor supply responses means that everybody's tax rate is affected, instead of just the target group's. If we want to induce the poor to work more to reduce their poverty, we need to lower everyone's tax rate. The welfarist linear tax rule does not take this into account. It is not, however, possible to state that under poverty minimization tax rates are optimally lower than under welfare maximization, since we cannot directly compare the welfare and deprivation terms. However, there is an additional efficiency consideration involved under poverty minimization. Nonlinear tax rules of course make it possible to target lower tax rates on the poorer individuals, but in a developing country context with

[9] We are grateful to a referee for this point.

lower administrative capacity this is not necessarily possible, and such considerations affect everyone's tax rate.

3 Public good provision with linear income taxes

3.1 Optimal public provision under the welfarist objective

Let us first extend the welfarist model of linear taxation to include the provision of pure public goods. The government offers a universal pure public good G, which enters individual utilities in addition to the consumption of private goods. The government's objective function is now $\sum W\left(V^i(a,b,G)\right)$, whereas the budget constraint becomes $(1-a)\sum z^i - Nb - N\pi G = R$ where π is the producer price of the public good. The consumer price of private consumption is normalized to 1. Let us now define the marginal willingness to pay for the public good by the expression $\sigma = \frac{V_G}{V_b}$ and $\sigma^* = \frac{\sum \beta^i \sigma^i}{\sum \beta^i}$ as the welfare-weighted average marginal rate of substitution between public good and income for individual i. The rule for public provision can then be written as

$$\pi = \sigma^* - \tau\left(\sigma^* \bar{z}_b - \bar{z}_G\right). \tag{6}$$

This public good provision rule is a version of a modified Samuelson rule. It equates the relative cost of providing the public good to the welfare-weighted sum of marginal rates of substitution (MRS). It also includes a revenue term, which takes into account the impacts of public good provision and income transfers on labor supply and thus tax revenue.

Consider first the case when labor supply does not depend on public good provision and there are no income effects, i.e., $\bar{z}_G = \bar{z}_b = 0$. Then we are left with a more familiar rule that welfare-weighted aggregate MRS must equal the cost of the public good. When we add income effects so that $\bar{z}_b < 0$, and since σ^* is positive, then because of the second term in (6), the financing costs of the public good are reduced. Likewise, if labor supply and public provision are positively related, the financing costs of the public good are reduced.

3.2 Optimal provision of public goods under poverty minimization

Now consider a non-welfarist government interested in minimizing poverty. The public good G which it offers enters the deprivation index separately from other, private consumption x: $D\left(x, G, \bar{x}, \bar{G}\right)$. The government still offers a lump-sum cash transfer b as well and finances its expenses with the linear income tax τ.

Again alternative formulations of the public good provision rule can be written. The first is

$$\pi = D^* - \tau\left(D^* \bar{z}_b - \bar{z}_G\right), \tag{7}$$

which can be compared with Eq. (6). Here, $D^* = \frac{\sum D_G + \sum D_x a z_G^i}{\sum D_x\left(1 + a z_b^i\right)}$ captures the efficiency of the public good in reducing deprivation relative to the income transfer (because $D_G, D_x < 0$, $D^* > 0$). Again, if $\bar{z}_G = \bar{z}_b = 0$, the equation reduces to $\pi = D^* = \frac{\sum D_G}{\sum D_x}$. This rule highlights a considerable difference to the standard modified Samuelson rules, reflecting instead of a welfare-based MRS the direct poverty reduction impact of the public good. With $\bar{z}_G \neq 0$ and $\bar{z}_b \neq 0$, D^* also depends on the indirect impacts of the public good via labor supply on consumption. As previously, the right-hand side includes a tax revenue term. Using the same example as in the context of (6), if $\bar{z}_G = 0$ and $\bar{z}_b < 0$, the price π of the public good would be higher than its relative efficiency in eliminating deprivation.

Here we have allowed the government to be directly interested in the consumption of some pure public good. But if the government is solely interested in reducing income poverty, it might not include such goods in the deprivation measure.[10] However, suppose that individual welfare does not directly depend on the public good provided but the public good can have a productivity increasing impact. An example could be publicly provided education services that affect individuals' productivity via the wage rate. We therefore suppose that the direct impact of the public good on deprivation cancels out (i.e., $D_G = 0$), whereas the wage rate becomes an increasing function of G, i.e., $w'(G) > 0$ (denoting $z = w(G)L$). This means that the expression for D^* is rewritten as

$$D^* = \frac{\sum D_x a \left(w \frac{\partial L}{\partial G} + w' L \right)}{\sum D_x \left(1 + a w \frac{\partial L}{\partial b} \right)}. \tag{8}$$

This means that even if labor supply would not react to changes in public good provision, such provision would still be potentially desirable through its impact on the wage rate. In this way, public good provision can be interpreted as increasing the capability of the individuals to earn a living wage, which serves as a poverty reducing tool, and which can in some cases be a more effective way to reduce poverty rather than direct cash transfers. The optimality depends on the relative strength of $w'(G) > 0$ versus the direct impact of the transfers.

An alternative provision rule for the public good, which results from extending the Piketty–Saez approach, in the usual case where it also enters individuals' utility function is

$$\frac{\int \left(D_G + D_x (1 - \tau) \frac{\partial z^i}{\partial G} \right) \mathrm{d}\nu(i)}{\int D_x \, \mathrm{d}\nu(i)} = \pi - \tau \frac{\mathrm{d}Z}{\mathrm{d}G}. \tag{9}$$

In the numerator of the left-hand side, the first term is the direct deprivation effect of G and the second term captures the indirect deprivation effect, operating via the labor supply impacts of the public good, which affect the level of private consumption, x. These impacts are scaled by the poverty alleviation impact of private consumption itself (the impact of a cash transfer). The right-hand side reflects the costs of public good provision: besides the direct cost of the good there is an indirect tax revenue effect

[10] See also Appendix 2 for multidimensional considerations in poverty measurement.

operating through labor supply. The condition is directly comparable to the welfarist rule, given in (39) in the Appendix, because even though the welfarist case relies on utilities, in the FOC for G no envelope condition is evoked. The only difference between Eqs. (39) and (9) is that the utility and welfare weight terms are exchanged for deprivation terms.

Consider finally the provision of a quasi-private good, such that in addition to the publicly provided amount, individuals can purchase ("top-up") the good themselves as well. The good is denoted by s and its total amount consists of private purchases h and public provision G: $s = G + h$. In addition to good s, individuals consume other private goods, denoted by x. The individual budget constraint is thus $c^i = x^i + ph^i = (1-\tau)z^i + \tau Z(1-\tau) - R - \pi G$, where p is the consumer price of private purchases of the quasi-private good. The producer price of education in the private sector (p) or in the public sector (π) can be equal, or one sector could have access to cheaper technology. Deprivation is determined in terms of consumption of x and s, so the objective function is min $P = \int D\left(x^i, s^i, \bar{x}, \bar{s}\right) \mathrm{d}\nu(i)$. In this case, the provision rule is

$$\frac{\int \left[D_x \left((1-\tau)\frac{\partial z^i}{\partial s}\frac{\partial s}{\partial G} - p\frac{\partial h^i}{\partial G}\right) + D_s \frac{\partial s^i}{\partial G}\right] \mathrm{d}\nu(i)}{\int D_x \,\mathrm{d}\nu(i)} = \pi - \tau \frac{\mathrm{d}Z}{\mathrm{d}G}. \tag{10}$$

The result is analogous to the pure public good result in (9), with the difference that now the impact G has on poverty depends on whether public provision fully crowds out private purchases of the good (i.e., $\frac{\mathrm{d}h}{\mathrm{d}G} = -1 \Leftrightarrow \frac{\mathrm{d}s}{\mathrm{d}G} = 0$) or not (i.e., $\frac{\mathrm{d}h}{\mathrm{d}G} = 0 \Leftrightarrow \frac{\mathrm{d}s}{\mathrm{d}G} = 1$). If there is full crowding out, an increase in public provision of G that is fully funded via a corresponding increase in the tax rate has no impact on the consumption of s and consequently no impact on poverty. If there is no crowding out, however, the FOC becomes

$$\frac{\int \left[D_x \left((1-\tau)\frac{\partial z^i}{\partial s}\right) + D_s\right] \mathrm{d}\nu(i)}{\int D_x \,\mathrm{d}\nu(i)} = \pi - \tau \frac{\mathrm{d}Z}{\mathrm{d}G}, \tag{11}$$

which is the same as in the case of a pure public good in Eq. (9).

To summarize, the welfarist public provision rule, when public goods are financed with linear income taxes and supplemented with lump-sum transfers, differs from the standard modified Samuelson rule. It equates a welfare-weighted sum of MRS to the marginal cost where tax revenue impacts are taken into account. Indirect effects of public provision (through labor supply decisions and thus private consumption) are incorporated. The poverty-minimizing public provision rule, however, replaces the welfare-weighted sum of MRS with the relative marginal returns to deprivation reduction. Here the "MRS" term measures how well public good is translated to reduced poverty (incorporating indirect effects as well), relative to private consumption. Finally, when the public good has positive effects on productivity, its provision can be desirable even if it would not have any direct impact on poverty.

4 Commodity taxation with linear income taxes

4.1 Optimal commodity taxation with linear income tax under the welfarist objective

This section considers the possibility that the government also uses commodity taxation (subsidies) to influence consumers' welfare. We follow the modeling of Diamond (1975). Unlike the analysis above, there are J consumer goods x_j instead of just two. Working with many goods is used to be able to more clearly describe the conditions under which uniform commodity taxation occurs at the optimum. The government levies a tax t_j on the consumption of good x_j, so that its consumer price is $q_j = p_j + t_j$, where p_j represents the producer price (a commodity subsidy would be reflected by $t_j < 0$). Let q denote the vector of all consumer prices. In addition, the government can use a lump-sum transfer, b. Note that in this exposition, leisure is the untaxed numeraire commodity. Alternatively, one could also imply a linear tax on labor supply as above and treat one of the consumption goods as the untaxed numeraire. However, choosing leisure as the numeraire makes the exposition easier. Thus, the consumer's budget constraint is $\sum_j q_j x_j^i = z^i + b$.

The government maximizes $\sum_i W\left(V^i(b, q)\right)$ subject to its budget constraint $\sum_i \sum_j t_j x_j^i - Nb = R$. It is useful to define, following Diamond (1975),

$$\gamma^i = \beta^i + \lambda \sum_j t_j \frac{\partial x_j^i}{\partial b} \tag{12}$$

as the net social marginal utility of income for person i. This notion takes into account the direct marginal social gain, β^i , and the tax revenue impact arising from commodity demand changes. The rule for optimal commodity taxation for good k is shown to be

$$\frac{1}{N} \sum_i \sum_j t_j \frac{\partial \tilde{x}_k^i}{\partial q_j} = \frac{1}{\lambda} \mathrm{cov}(\gamma^i, x_k^i). \tag{13}$$

The left-hand side of the rule is the aggregate compensated change (weighted by commodity taxes) of good k when commodity prices are changed. The right-hand side refers to the covariance of the net marginal social welfare of income and consumption of the good in question. The rule says that the consumption of those goods whose demand is the greatest for people with low net social marginal value of income (presumably, the rich) should be discouraged by the tax system. Likewise the consumption of goods such as necessities should be encouraged by the tax system.

The key policy question is whether or when uniform commodity taxes are optimal, or, in other words, when would a linear income tax combined with an optimal demogrant be sufficient to reach the society's distributional goals at the smallest cost. Deaton (1979) shows that weakly separable consumption and leisure and linear Engel curves are sufficient conditions for the optimality of uniform commodity taxes. These requirements are quite stringent and unlikely to hold in practice; however, the economic importance they imply is unclear. If implementing differentiated commodity taxation entails significant administrative costs, they may easily outweigh the potential benefits of distributional goals and that is why economists have typically been quite skeptical about non-uniform commodity taxation when applied to practical tax policy.

4.2 Optimal commodity taxation with linear income tax under poverty minimization

Poverty could be measured in many ways when there are multiple commodities: the government may care about overall consumption, the consumption of some of the goods (those that are in the basket used to measure poverty) or then it cares about both the overall consumption and the relative share of different kinds of consumption goods (such as merit goods). We discuss these measurement issues in Appendix 2, but here we examine the simplest set-up where deprivation only depends on disposable income, $c^i = z^i + b$. Using the consumer's budget constraint, this is equal to the overall consumption level, $\sum_j q_j x_j^i$.

The government thus minimizes the sum of the poverty index $D\left(\sum_j q_j x_j^i, \bar{c}\right)$, and the budget constraint is the same as before. It is again useful to define

$$\gamma_P^i = D_c \sum_j q_j \frac{\partial x_j^i}{\partial b} + \lambda \sum_j t_j \frac{\partial x_j^i}{\partial b} \tag{14}$$

as the net poverty impact of additional income for person i. This notion takes into account the direct impact on poverty and the tax revenue impact arising from commodity demand changes.

As shown in Appendix 1 section "Commodity taxation", this leads to an optimal tax rule as below:

$$\frac{1}{N}\sum_i \sum_j t_j \frac{\partial \tilde{x}_k^i}{\partial q_j} = -\frac{1}{\lambda}\left[\frac{1}{N}\sum_i D_c x_k^i + \frac{1}{N}\sum_i \sum_j D_c q_j \frac{\partial \tilde{x}_k^i}{\partial q_j}\right] + \frac{1}{\lambda}\mathrm{cov}\left(\gamma_P^i, x_k^i\right). \tag{15}$$

In this formulation, the left-hand side is the same as in the welfarist case and it reflects the aggregate compensated change in the demand of good k. The first two terms in the square brackets at the right-hand side capture the impacts of tax changes on poverty: the first term is the direct impact of the price change (keeping consumption unaffected) on measured poverty, whereas the second depends on the behavioral shift in consumption. Multiplied by the minus sign, the former term implies that the consumption of the good should be encouraged, whereas if demand decreases when the prices increase, the latter term actually serves to discourage consumption. The last term on the right reflects the same principles as the covariance rule in Eq. (13), the correlation of the net poverty impact of income and the consumption of the good in question. That is, the covariance part of the tax rule moves the tax rule in the direction of favoring goods that have high poverty reduction impact on the poor (i.e., that the poor consume more).

The key lesson to note from the optimal commodity tax rule in the poverty minimization case is that the conventional conditions for uniform commodity tax to be optimal are not valid anymore. The reason is that even if demand was separable from labor supply, the first term on the right still remains in the rule, and its magnitude clearly varies depending on the quantity of good consumed. Thus, income transfers are not sufficient to alleviate poverty when the government aims to minimize poverty that depends on disposable income. The intuition is very simple: commodity tax changes have a direct effect on the purchasing power of the consumer, and these depend on the amount consumed. The extent of encouraging the consumption of the goods is the greater, the larger is their share of consumption among the consumption bundles of the poor. The result resembles that of Pirttilä and Tuomala (2004), meaning that the intuition from optimal nonlinear income taxation under poverty minimization carries over to linear income taxation. A formal proof is provided in Appendix 1.

In sum, the rule for optimal commodity taxation is changed when we shift from welfare maximization to poverty minimization. The welfarist rule reflects a fairly straightforward trade-off between efficiency (tax revenue) and equity (distributional impacts). The poverty-minimizing commodity tax rule brings new terms; the interrelations of which are not easy to disentangle. It, however, also takes into account efficiency considerations (tax revenue through indirect labor supply effects) and equity (direct impact of the taxed good on poverty and indirect impact via labor supply effects). Most importantly, the conventional wisdom of when uniform commodity taxation is sufficient fails to hold in the poverty minimization case. Thus, observed commodity subsidies in developing countries, such as fuel or food subsidies, can be considered optimal given the preference for poverty minimization.[11] In practice, it would be wise to limit the number of differentiated commodity tax rates to a few essential categories such as fuel and food, in order to keep the administrative complexity at a minimum.

5 Poverty minimization in the presence of an informal sector

An important issue for a developing country attempting to collect taxes is the issue of a large informal sector. If part of tax revenue is lost due to tax evasion in the informal sector, which is likely to be the case in the less developed economies, then the income transfer is reduced and redistributive targets may not be met. In this section, we discuss the implications of informality for optimal redistributive policies for a government wishing to minimize poverty.[12] The results can thus be contrasted to those obtained in previous sections.

[11] Keen (2014) uses a tax reform approach and examines how much more effective transfers need to be than differentiated commodity subsidies in reaching the poor to achieve the same poverty reduction with lower government outlays.

[12] Such a society might also reflect poor administrative power and corruption in the tax collecting authority. Notice, however, that considering only the "leakage" of tax revenue in the model would only reduce the extent of poverty reduction achieved with taxation by lowering the income transfer for everyone. The poverty reduction efficiency of taxation would thus be lowered, but there would be no differential effects across individuals.

Following Kanbur (2015) and Kanbur and Keen (2014), informal operators can be categorized as those who should comply with regulations but illegally choose not to, and those who legally remain outside regulation, e.g., due to the smaller size of operations (either naturally or by adjusting size as a response to regulation). For our purposes, however, it is enough to lump these categories into one "informal sector," where it is possible to avoid taxes at least to some extent. It is also possible for workers to work in both sectors, such that part of total income is declared for taxation and part is evaded (consider, e.g., supplementing official employment income with street vendoring). Note also that especially in the case of agriculture, evasion can also consist of home production. In this case, the reason for "informality" would be the small size of the producing entity, such that they are naturally not liable for taxes. Production for own consumption is, however, still relevant for the well-being and measured poverty of the family.

In this application, we follow the approach pioneered in Besley and Persson (2013). They work with a model that fits into the description above, where part of the tax base evades taxes. We thus take informality as given, and do not consider whether informality is "natural," illegal or a response to taxation. Furthermore, this intensive margin model (what extent of income is earned in the informal sector), they argue, yields essentially similar results as an extensive margin model (whether to participate in the formal job market).

Consider the case of income taxation. We can incorporate informality into the model by noting that people can shelter part e of their labor income from taxation. The extent of evasion is assumed to increase when the tax rate goes up, and thus $\frac{\partial e}{\partial a} < 0$. Income taxes are only paid from income $z^i - e^i$. It is noteworthy that for a government wishing to minimize income poverty, this is in fact beneficial: disposable incomes rise. The more this effect is concentrated among the poor who enter the deprivation index, the better. Individual consumption is now $z^i - \tau(z^i - e^i) + b = e^i + a(z^i - e^i) + b$. On the other hand, tax collections are reduced: the budget constraint becomes $(1-a)\sum(z^i - e^i) = Nb + R$. Our formulation follows that of Besley and Persson (2013), but we simplify it in order to explicitly consider the problem of optimal taxation, whereas they focus on the issue of investments in the state's fiscal capacity (we abstract from this issue here and take evasion as given).[13] The framework, however, nicely captures the essential trade-offs a government faces when there is tax evasion.

The government now minimizes the Lagrangian $L = \sum D\left(e^i + a(z^i - e^i) + b, \bar{c}\right) + \lambda((1-a)\sum(z^i - e^i) - Nb - R)$. The first-order condition with respect to the net-of-tax rate is:

$$\sum D_c\left(\frac{\partial e^i}{\partial a} + z^i - e^i + a\left(\frac{\partial z^i}{\partial a} - \frac{\partial e^i}{\partial a}\right)\right) = \lambda\left(\sum(z^i - e^i) - (1-a)\sum\left(\frac{\partial z^i}{\partial a} - \frac{\partial e^i}{\partial a}\right)\right), \tag{16}$$

[13] Another difference is that in their original formulation, people face costs of evasion. When the tax rate goes up, the relative attractiveness of tax evasion increases, producing the same kind of effect $\left(\frac{\partial e}{\partial a} < 0\right)$ we assume directly here for brevity. (These costs could be related to, e.g., Allingham–Sandmo-type risk of being caught and facing sanctions.) Also Slemrod's (1990) review suggests that higher tax rates tend to increase the supply of labor to the informal sector.

whereas, under the assumption that there are no income effects in evasion, the first-order condition with respect to b stays the same. From here, we can derive a rule for the optimal tax following the same steps as in Sect. 2.2:

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\varepsilon^e}\left(1 - \frac{\tilde{D}^e}{\bar{z}^e}\right), \tag{17}$$

where now ε^e is a tax elasticity of the net-of-evasion tax base $\bar{z}^e = \bar{z} - \bar{e}$ and $\tilde{D}^e$ represents the relative impact of taxes and transfers on the deprivation index (see Appendix 1 for further detail). The rule represents a trade-off between poverty reduction and efficiency, both of which are now altered by evasion. There is a pressure toward lower tax rates, as now distortions of taxation are increased by evasion behavior, so $\varepsilon^e > \varepsilon$. Contrary to this effect, $\tilde{D}^e$ is reduced compared to $\tilde{D}$ because reducing taxes (increasing a) is now a less useful instrument for poverty reduction, as part of the taxes have been evaded. As $\frac{\partial e}{\partial a} < 0$, people pay more taxes when tax rates are reduced, and therefore poverty in fact increases. $\tilde{D}^e$ thus works to increase tax rates.

Therefore, an interesting trade-off arises: informality increases the cost of raising taxes, but it also means that higher taxes are less harmful as those in the informal sector do not need to pay them (and they are still entitled to the lump-sum transfer).[14] These countervailing forces have not been noted by the literature before. The presence of informality therefore seems to give rise to tax policy rules that are far from trivial. Future work could also look more deeply into the issue of the tax mix in the presence of informality. If income tax is more easily evaded than commodity taxation, as Boadway et al. (1994) suggest, this could give rise to policies that focus taxation and redistribution on commodity taxes and subsidies, instead of income taxes and lump-sum transfers. Slemrod and Gillitzer (2014) have also suggested focusing on a "tax systems approach" and including, among other things, evasion behavior into optimal taxation analysis to obtain more useful prescriptions for actual tax policy. This topic certainly deserves a more detailed analysis.

6 A numerical illustration

To further illustrate the differences of tax rates under poverty minimization and welfarism, we provide a simple numerical simulation. Here we concentrate on the special case where there are no income effects on labor supply and the elasticity of labor supply with respect to the net-of-tax wage rate is constant. If ε denotes this elasticity, the quasi-linear indirect utility function is given by $v(w(1-\tau), b) = b + \frac{[w(1-\tau)]^{1+\varepsilon}}{1+\varepsilon}$, so that ε is constant. Like most work on optimal nonlinear and linear income taxation, we use the lognormal distribution $\ln(n, m\sigma^2)$ to describe the distribution of productivities with support $[0, \infty)$ and parameters m and σ (see Aitchison and Brown 1957). The first parameter, m, is the log of the median wage. The second parameter, the variance

[14] The idea that those in the informal sector can still receive transfers matches well with reality: many of the cash transfer systems reach those with little or no connection to the formal sector.

of log wage σ^2, is itself an inequality measure. As is well known, the lognormal distribution fits reasonably well over a large part of the income range but diverges markedly at both tails. The Pareto distribution in turn fits well at the upper tail. We also use the two-parameter version of the Champernowne distribution (known also as the Fisk distribution). This distribution approaches asymptotically a form of Pareto distribution for large values of wages but it also has an interior maximum. In our simulations, the revenue requirement is set to zero; thus, the system is purely redistributive.

To illustrate the poverty-minimizing tax formula in (3), we also need to specify a measure of poverty. Typically, poverty indices consist of computing some average measure of deprivation by setting individual needs as defined above at the agreed upon poverty line $\bar{c}$. For this purpose, we take a poverty index of the form developed by Foster et al. (1984). They have proposed defining a poverty index as the average of these poverty gaps across individuals raised to some power α. When $\alpha = 1$, it is just the proportion of units below the poverty line multiplied by the average poverty gap. (See Appendix 2 for more details.) We consider the cases where either 30 or 40% of the population lie below the poverty line.

The results from the simulation of the optimal tax when the government minimizes the poverty gap for the lognormal case are presented in Table 1. Results are shown for two different values of labor supply elasticity ε, two different values regarding income dispersion σ, and two values of the share of population below the poverty line $F(\bar{w})$. The tax rates are high, above 60%, for all the combinations of parameter values.[15]

Comparing these results to the welfarist case is not straightforward, as those depend on the chosen welfare function. We adopt a constant relative inequality aversion form of the welfare function: the contribution to social welfare of the ith individual is $\frac{w_i^{1-\eta}}{1-\eta}$, where η is the constant relative inequality aversion coefficient. Hence, the social marginal value of income to an individual with wage rate w is proportional to $w^{-\eta}$. Using the property of the lognormal distribution $\ln(E(w^s)) = sm + s^2\frac{\sigma^2}{2}$, we can calculate the optimal tax rate from the following formula: $\frac{\tau}{1-\tau} = \frac{1}{\varepsilon}[1 - e^{-\eta(1+\varepsilon)\sigma^2}]$. Or, using the property of the lognormal distribution that $\ln(1 + cv^2) = \sigma^2$, where cv is the coefficient of variation, we can rewrite $\tau = \frac{1}{1+\varepsilon/[1+cv^2]^{-\eta(1+\varepsilon)}}$.

A wide range of values for the inequality aversion parameter η have been employed in the literature, varying typically from 0.5 to 2. Note that, as discussed in Sect. 2.1, as $\eta \to \infty$, social preferences approach "maximin" preferences, where the optimal tax rate is the same as the revenue-maximizing tax rate, $\tau = \frac{1}{1+\varepsilon}$, which does not depend on the original income distribution. Naturally, if there is no regard for inequality in the society, $\eta = 0$ and $\tau = 0$. Table 2 displays the welfaristic tax simulation results for two different values of labor supply elasticity ε, for two different values of income dispersion σ, and for five different values of inequality aversion η.

The simulation results illustrate clearly that at conventional inequality aversion levels, optimal welfaristic tax rates lie well below the poverty-minimizing rates. Only as inequality aversion becomes extremely high do the welfaristic rates approach the

[15] The results are very similar using the Champernowne distribution (with income dispersion parameters chosen so that inequality is similar in both cases), which is not very surprising as the distributions only differ at the top of the income schedule. These results are available upon request.

Table 1 Simulated tax rates for poverty minimization under different values of ε, σ, and $F(\bar{w})$

ε	$\sigma = 0.7$		$\sigma = 1.0$	
	$F(\bar{w}) = 0.3$	$F(\bar{w}) = 0.4$	$F(\bar{w}) = 0.3$	$F(\bar{w}) = 0.4$
0.25	79	77	79	78
0.5	65	63	66	64

Table 2 Simulated tax rates in the welfaristic case under different values of ε, σ, and η

ε	$\sigma = 0.7$			
	$\eta = 0.5$	$\eta = 1$	$\eta = 2$	$\eta \to \infty$
0.25	43	58	69	80
0.5	31	44	56	67
ε	$\sigma = 1.0$			
	$\eta = 0.5$	$\eta = 1$	$\eta = 2$	$\eta \to \infty$
0.25	52	65	74	80
0.5	38	51	61	67

poverty-minimizing ones. With poverty minimization as the social objective, optimal tax rates are close to the revenue-maximizing "maximin" rate.

Another point of comparison could be the welfaristic linear tax simulations of Stern (1976). His calculations differ from ours as he incorporates income effects and a non-constant elasticity of labor supply with respect to the tax rate.[16] With the elasticity of substitution between consumption and leisure at 0.5 and income dispersion described by $\sigma = 0.39$, as concern for inequality rises from low to medium and high, he finds tax rates rising from 19 to 43 and 48%. The extreme "maximin" result is 80%. These tax rates are also clearly lower than the poverty-minimizing rates, except at very extreme values of inequality aversion.

These numerical examples and Stern's (1976) results tend to suggest that the tax rates for the poverty minimization case are likely to be higher than for many welfarist examples. The results compare to Kanbur et al. (1994), who also found that the (non-linear) marginal tax rates on the poor are fairly high under the poverty minimization objective. Both their and our results are interesting from the point of view that the analytical formulae for the optimal tax rate include a term that, ceteris paribus, encourages labor supply, but in computational results its influence is offset, most likely, by the need to minimize the poverty gap. The higher the poverty rate, the higher the lump-sum grant financed by these taxes needs to be, in order to raise more people out of poverty.

[16] Our simplifying assumptions allow us to provide tax rates with respect to the three parameters in Table 2.

7 Conclusion

This paper examined optimal linear income taxation, public provision of public and private goods and the optimal combination of linear income tax and commodity taxes when the government's aim is to minimize poverty. The linear tax environment was chosen because such taxes are more easily implementable in a developing country context and since optimal linear tax rules are seen to provide similar intuition as the more complex nonlinear tax formulas.

The results show that the linear income tax includes additional components that work toward lowering the marginal tax rate. This result arises from the goal to boost earnings to reduce income poverty. Unlike in the optimal nonlinear income tax framework, this lower marginal tax affects all taxpayers in the society. However, the numerical simulations offered suggest that this mechanism is offset by the distributive concerns and in practice the optimal tax rates for poverty minimization appear high. Public good provision in the optimal tax framework under poverty minimization was shown to depend on the relative efficiency of public provision versus income transfers in generating poverty reductions. One particular avenue where public provision is useful is via its potentially beneficial impact on individuals' earnings capacity. Thus, public provision can be desirable even if its direct welfare effects were non-existent.

Perhaps more importantly, poverty minimization as an objective changes completely the conditions under which uniform commodity taxation is optimal. When the government's objective is to minimize poverty that depends on disposable income, uniform commodity taxation is unlikely to be ever optimal: this is because the commodity tax changes have first-order effects on consumers' budget via the direct impact on the cost of living, and this direct effect depends on the relative importance of different goods in the overall consumption bundle. Separability in demand coupled with linear Engel curves is not sufficient to guarantee optimality of uniform commodity taxes. In reality, the administrative difficulties of implementing commodity taxation with many tax rates must, of course, be taken into account, as well.

We also examined the implications of the presence of an informal sector for optimal tax and transfer policies. The results revealed that when the government is concerned about income poverty, the presence of the informal sector is, on the one hand, useful, as it reduces the poverty-increasing effect of higher taxes but, on the other hand, it is also costly since it is likely to increase the elasticity of the tax base. Examining the implications of informality on the role of other instruments of government policies is an important avenue for future work.

Another strand of follow-up work should address the question of complementary policies for redistribution, such as minimum wages. It should be borne in mind that different policies impose different requirements on administrative capacity,[17] and

[17] For example, Lee and Saez (2012) show how a minimum wage policy can usefully complement an optimal nonlinear income tax and transfer policy under welfarist objectives. However, imposing minimum wage regulation implies that the government needs to be either able to observe individual wage rates, or has sufficient institutional strength to rely on whistleblowers to denounce non-complying employers, in order to enforce the legislation.

examining which poverty reduction instruments become available only as the societies advance on their development path is an interesting avenue for further work.

Acknowledgements We are grateful to two anonymous referees and seminar audiences at the UNU-WIDER Conference on 'Inequality: Measurement, Trends, Impacts and Policies', the Annual Meeting of the Finnish Economic Association, the HECER Development Economics seminar and the IIPF conference in Dublin for useful comments. This research originates from the UNU-WIDER project *The economics and politics of taxation and social protection*. Funding from the Academy of Finland (Grant No. 268082) is gratefully acknowledged.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflicts of interest.

Appendix 1: Mathematical appendix

Linear income taxation

Welfarism

Consider first the welfarist case. Using λ to denote the multiplier associated with the budget constraint, the government's Lagrangian is $L = \sum W\left(V^i(a,b)\right) + \lambda\left((1-a)\sum z^i - Nb - R\right)$. The social marginal utility of income is $\beta^i = W_V V_b^i$. Using Roy's theorem, $V_a^i = V_b^i z^i$, we have $W_V V_a^i = \beta^i z^i$. The first-order conditions with respect to a and b, respectively, are then:

$$\sum \beta^i z^i = \lambda\left(\sum z^i - (1-a)\sum z_a^i\right) \tag{18}$$

$$\sum \beta^i = \lambda\left(N - (1-a)\sum z_b^i\right). \tag{19}$$

Divide (18) by (19) to get:

$$\frac{\sum \beta^i z^i}{\sum \beta^i} = \frac{\sum z^i - (1-a)\sum z_a^i}{N - (1-a)\sum z_b^i}. \tag{20}$$

Denote average income $\bar{z} = \frac{\sum z^i}{N}$ and welfare-weighted average income $z(\beta) = \frac{\sum \beta^i z^i}{\sum \beta^i}$ to get:

$$z(\beta) = \frac{\bar{z} - (1-a)\bar{z}_a}{1 - (1-a)\bar{z}_b}. \tag{21}$$

Multiply the government's revenue constraint by $\frac{1}{N}$ and define $g = \frac{R}{N}$ to get $(1-a)\bar{z} - b = g$, and totally differentiate, keeping g constant:

$$\frac{\mathrm{d}b}{\mathrm{d}a}|_{g\text{const}} = \frac{\bar{z} - (1-a)\bar{z}_a}{-1 + (1-a)\bar{z}_b} = -z(\beta). \tag{22}$$

The fact that $z(\beta) = -\frac{\mathrm{d}b}{\mathrm{d}a}|_{g\text{const}}$ tells us that welfare-weighted labor supply should be equal to the constant-revenue effect of tax rate changes in b.

By totally differentiating average labor income $\bar{z}$ and using (22), we have

$$\frac{\mathrm{d}\bar{z}}{\mathrm{d}a}|_{g\text{const}} = \bar{z}_a + \bar{z}_b \frac{\mathrm{d}b}{\mathrm{d}a}|_{g\text{const}} = \bar{z}_a - \bar{z}_b z(\beta). \tag{23}$$

When we impose g as a constant we have to give up one of our degrees of freedom. Now the interpretation of $\frac{\mathrm{d}\bar{z}}{\mathrm{d}a}|_{g\text{const}}$ is then the effect on labor supply when a is changed, as is b, in order to keep tax revenue constant. Using (23) we can write (21):

$$z(\beta) - \bar{z} = -(1-a)\frac{\mathrm{d}\bar{z}}{\mathrm{d}a}|_{g\text{const}} = -\tau \frac{\mathrm{d}\bar{z}(1-\tau)\bar{z}}{\mathrm{d}(1-\tau)(1-\tau)\bar{z}}, \tag{24}$$

from which we get the optimal tax rate of Eq. (1).

We now derive the results in the form of the Piketty and Saez (2013) model. In their model, there is a continuum of individuals, whose distribution is $\nu(i)$ (population size is normalized to one). Individuals maximize their utility $u^i((1-\tau)z^i + b, z^i)$, and their FOC implicitly defines the Marshallian earnings function $z_u^i(1-\tau, b)$. Using this, aggregate earnings are $Z_u(1-\tau, b)$. The government's budget constraint $b + R = \tau Z_u(1-\tau, b)$ implicitly defines b as a function of τ, and consequently Z_u can also be defined solely as a function of τ: $Z(1-\tau) = Z_u(1-\tau, b(\tau))$. Z has elasticity $\varepsilon = \frac{1-\tau}{Z}\frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)}$.

To start, note that if the government only cared about maximizing tax revenue $\tau Z(1-\tau)$, it would set τ such that $\frac{\partial(\tau Z(1-\tau))}{\partial\tau} = 0$: $Z(1-\tau) - \tau\frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)} = 0$. Using $\frac{\tau}{Z}\frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)} = \frac{\tau}{1-\tau}\varepsilon$, this gives

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\varepsilon}$$
$$\Leftrightarrow \tau^* = \frac{1}{1+\varepsilon}. \tag{25}$$

When the government is concerned about social welfare, its problem is to max $\text{SWF} = \int \omega^i W(u^i((1-\tau)z^i + \tau Z(1-\tau) - R, z^i))\,\mathrm{d}\nu(i)$, where use has been made of the individual consumption $c^i = (1-\tau)z^i + b = (1-\tau)z^i + \tau Z(1-\tau) - R$. Here ω is a Pareto weight and W is an increasing and concave transformation of utilities. The FOC $\frac{\partial \text{SWF}}{\partial\tau} = 0$ is:

$$\int \omega^i W_u \left[u_c^i \left(-z^i + (1-\tau)\frac{\partial z^i}{\partial\tau} + Z + \tau\frac{\mathrm{d}Z}{\mathrm{d}\tau}\right) + u_z^i \frac{\partial z^i}{\partial\tau}\right] \mathrm{d}\nu(i) = 0,$$

which, using the individual's envelope condition, becomes:

$$\int \omega^i W_u u_c^i \left(-z^i + Z - \tau\frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)}\right) \mathrm{d}\nu(i) = 0.$$

Taking $Z - \tau \frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)}$ out of the integrand and leaving it to the left-hand side, we have on the right-hand side $\frac{\int \omega^i W_u u_c^i z^i \, \mathrm{d}\nu(i)}{\int \omega^i W_u u_c^i \, \mathrm{d}\nu(i)}$. Piketty and Saez define $\beta^i = \frac{\omega^i W_u u_c^i}{\int \omega^i W_u u_c^i \mathrm{d}\nu(i)}$ as a normalized social marginal welfare weight for individual i, so that the term can be simplified to:

$$Z - \tau \frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)} = \int \beta^i z^i \, \mathrm{d}\nu(i).$$

Using the definition of aggregate elasticity of earnings and defining $\bar{\beta} = \frac{\int \beta^i z^i \mathrm{d}\nu(i)}{Z}$ as the average normalized social marginal welfare weight, weighted by labor incomes z^i (it can also be interpreted as the ratio of the average income weighted by individual welfare weights β^i to the average income Z), we can rewrite this as $1 - \frac{\tau}{1-\tau}\varepsilon = \bar{\beta}$, which gives the optimal social welfare-maximizing tax rate:

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\varepsilon}\left(1 - \bar{\beta}\right). \tag{26}$$

According to Piketty and Saez, $\bar{\beta}$ "measures where social welfare weights are concentrated on average over the distribution of earnings." The welfare-maximizing tax rate is thus decreasing in both the average marginal welfare weight and the tax elasticity of aggregate earnings. A higher $\bar{\beta}$ reflects a lower taste for redistribution, and thus a lower desire to tax for redistributive reasons.

Piketty and Saez also note that (26) can be written in the form of $\tau^* = \frac{-\mathrm{cov}\left(\beta^i, \frac{z^i}{Z}\right)}{-\mathrm{cov}\left(\beta^i, \frac{z^i}{Z}\right) + \varepsilon}$. If higher incomes are valued less (lower β), then the covariances are negative and the tax rate is positive. This is a similar formulation as in Dixit and Sandmo (1977), Eq. (20), where $\tau^* = -\frac{1}{\lambda} \frac{-\mathrm{cov}(z^i, \mu^i)}{\frac{\partial \bar{z}}{\partial(1-\tau)}|_{\mathrm{comp.}}}$ (here λ represents the government's budget constraint Lagrange multiplier and μ^i the individual's marginal utility of income, s.t. $u_c = \mu^i$). Here the numerator reflects the equity element and the denominator the efficiency component, similar as in (26).

Non-welfarism

In the non-welfarist case, the Lagrangian function is $L = \sum F\left(az^i + b, z^i\right) + \lambda((1 - a)\sum z^i - Nb - R)$. The first-order conditions with respect to a and b are:

$$\sum \left(F_c(z^i + az_a^i) + F_z z_a^i\right) = \lambda \left(\sum z^i - (1-a)\sum z_a^i\right) \tag{27}$$

$$\sum \left(F_c(1 + az_b^i) + F_z z_b^i\right) = \lambda \left(N - (1-a)\sum z_b^i\right). \tag{28}$$

Dividing the first equation with the second and dividing through the right-hand side with N, we get:

$$\frac{\sum \left(F_c(z^i + az_a^i) + F_z z_a^i\right)}{\sum \left(F_c(1 + az_b^i) + F_z z_b^i\right)} = \frac{\bar{z} - (1-a)\bar{z}_a}{1 - (1-a)\bar{z}_b}, \tag{29}$$

which gives Eq. (3). Minimizing a deprivation index D is a special case of this, such that $F_c = D_c$ and $F_z = 0$. Otherwise the derivation of (5) is analogous to the above.

Let us next derive the poverty-minimizing tax rule following the formulation of Piketty and Saez. Given the government's instruments, consumption is $c^i = (1 - \tau)z^i + b = (1-\tau)z^i + \tau Z(1-\tau) - R$. The poverty minimization objective in the continuous case thus reads:

$$\begin{aligned} \min\ P &= \int D\left(c^i, \bar{c}\right) \mathrm{d}\nu(i) \\ &= \int D\left((1-\tau)z^i + \tau Z(1-\tau) - R, \bar{c}\right) \mathrm{d}\nu(i). \end{aligned} \tag{30}$$

The optimal tax rate is found from the government's FOC, $\frac{\partial P}{\partial \tau} = 0$:

$$\begin{aligned} &\int D_c \left(-z^i + (1-\tau)\frac{\partial z^i}{\partial \tau} + Z + \tau \frac{\mathrm{d}Z}{\mathrm{d}\tau}\right) \mathrm{d}\nu(i) = 0 \\ \Leftrightarrow &\int D_c \left(-z^i - (1-\tau)\frac{\partial z^i}{\partial(1-\tau)} + Z - \tau \frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)}\right) \mathrm{d}\nu(i) = 0. \end{aligned} \tag{31}$$

Define a "normalized marginal deprivation weight" as $\beta^i = \frac{D_c}{\int D_c \mathrm{d}\nu(j)}$. Using this definition, $\left(Z - \tau\frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)}\right)\int D_c\, \mathrm{d}\nu(i) = \int D_c \left(z^i + (1-\tau)\frac{\partial z^i}{\partial(1-\tau)}\right) \mathrm{d}\nu(i)$ can be written as:

$$Z - \tau \frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)} = \int \beta^i \left(z^i + (1-\tau)\frac{\partial z^i}{\partial(1-\tau)}\right) \mathrm{d}\nu(i). \tag{32}$$

Using the definition of the elasticity of individual labor earnings $\varepsilon^i = \frac{1-\tau}{z^i}\frac{\partial z^i}{\partial(1-\tau)}$, we have $(1-\tau)\frac{\partial z^i}{\partial(1-\tau)} = z^i \varepsilon^i$ and using elasticity of aggregate earnings $\varepsilon = \frac{1-\tau}{Z}\frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)}$, we have $Z - \tau\frac{\mathrm{d}Z}{\mathrm{d}(1-\tau)} = 1 - \frac{\tau}{1-\tau}\varepsilon$ and we can rewrite the above as:

$$Z\left(1 - \frac{\tau}{1-\tau}\varepsilon\right) = \int \beta^i \left(z^i + z^i \varepsilon^i\right) \mathrm{d}\nu(i). \tag{33}$$

This leads to the poverty-minimizing rule of

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\varepsilon}\left(1 - \bar{\beta} - \bar{\beta}^{\varepsilon}\right), \tag{34}$$

where analogously to Piketty–Saez, $\bar{\beta} = \frac{\int \beta^i z^i \mathrm{d}\nu(i)}{Z} \left(= \frac{\int D_c z^i \mathrm{d}\nu(i)}{Z \int D_c \mathrm{d}\nu(j)}\right)$ is an average normalized deprivation weight, weighted by labor incomes (or, analogously, average labor income weighted by individual deprivation weights). In addition, we have defined $\bar{\beta}^{\varepsilon} = \frac{\int \beta^i z^i \varepsilon^i \mathrm{d}\nu(i)}{Z} \left(= \frac{\int D_c z^i \varepsilon^i \mathrm{d}\nu(i)}{Z \int D_c \mathrm{d}\nu(j)}\right)$, which describes average labor incomes weighted by their corresponding individual elasticities and deprivation weights. This can be interpreted as a combined deprivation and efficiency effect.

As in the welfarist setting, the more elastic average earnings are to taxation, the lower is the optimal tax rate (a regular efficiency effect). The optimal poverty-minimizing tax rate is decreasing in the average deprivation weight $\bar{\beta}$, as a higher taste for redistribution toward the materially deprived implies a lower $\bar{\beta}$ and thus higher taxation for redistributive purposes. The effect is analogous to the welfarist tax rate, of course with slightly different definitions for $\bar{\beta}$.

The new term $\bar{\beta}^{\varepsilon}$ can be interpreted as a combined deprivation weight and efficiency effect. The elasticity term implicit in $\bar{\beta}^{\varepsilon}$ takes into account the incentive effects of taxation on working and works to reduce τ^*. To avoid discouraging the poor from working, their tax rates should be lower. But because the tax instrument is forced to be linear, tax rates are then lowered for everyone, as we found in the Tuomala model in Eq. (5). The value of $\bar{\beta}^{\varepsilon}$ depends on the relationship of the individual earnings elasticities and income: if the elasticity is the same across income levels, there is just a level effect moving from $\bar{\beta}$ to $\bar{\beta}^{\varepsilon}$; however, if the elasticity were higher for more deprived individuals, for example, $\bar{\beta}^{\varepsilon}$ would most likely be higher than under a flat elasticity. This works toward a lower tax rate in order to avoid discouraging the poorest from working. However, whether $\bar{\beta}^{\varepsilon}$ is high or low does not depend only on the shape of the elasticity but also on the shape of the deprivation weights, which also affect $\bar{\beta}$.

Finally, the third way for expressing the optimal tax rule in the case of poverty minimization is one following the Dixit and Sandmo (1977) formulation and it can be written as

$$\tau^* = -\frac{1}{\lambda} \frac{\operatorname{cov}\left(D_c, z^i\right) + \frac{1}{N} \sum D_c a \tilde{z}_a^i + \operatorname{cov}\left(D_c a z_b^i, z^i\right)}{\frac{1}{N} \sum \tilde{z}_a^i}. \tag{35}$$

In this expression, the denominator is the same as in Eq. (20) of Dixit and Sandmo (1977) presented before, that is, the average derivative of compensated labor supply with respect to the net-of-tax rate. In the numerator, the first term measures the strength of the association between income and poverty impact: when the association between overall poverty and small income is strong (this would be the case with the squared poverty gap), the tax should be high so that it will finance a sizable lump-sum transfer. If the association is weaker (as with the headcount rate), the tax rate is optimally smaller. The second and the third terms in the numerator are new. They measure the indirect effects from changes in the tax rate on labor supply. Here $\tilde{z}$ is the compensated (Hicksian) labor supply. The greater is the reduction in the labor supply following an increase in the tax rate (it is the compensated change as the tax increase is linked with a simultaneous increase in the lump-sum transfer), the smaller should the tax rate be in order to avoid increases in deprivation arising from lower earned income. The last two terms in the numerator are closely linked with a formulation $D_c(1-\tau)\frac{\partial \tilde{z}}{\partial q}|_{\text{comp}}$, where the idea is that the last covariance term serves as a corrective device for the mean impact of taxes on labor supply (similarly as in the denominator in the original Dixit–Sandmo formulation).

Public good provision

Welfarism

The Lagrangian is $L = \sum W\left(V^i(a,b,G)\right) + \lambda\left((1-a)\sum z^i - Nb - N\pi G - R\right)$. Maximizing the Lagrangian with respect to b and G gives:

$$\sum \beta^i = \lambda\left(N - (1-a)\sum z_b^i\right) \tag{36}$$

$$\sum W_V V_G^i = \lambda\left(N\pi - (1-a)\sum z_G^i\right). \tag{37}$$

Dividing (37) by (36) we obtain

$$\frac{\sum \beta^i \sigma^i}{\sum \beta^i} = \frac{\pi - (1-a)\bar{z}_G}{1-(1-a)\bar{z}_b}, \tag{38}$$

where we define $\sigma^* = \frac{\sum \beta^i \sigma^i}{\sum \beta^i}$ to be the welfare-weighted average marginal rate of substitution between public good and income for individual i. Rewriting this rule gives Eq. (6) in the main text.

Extending the Piketty and Saez approach to include public provision, the government's goal function is

$$\mathrm{SWF} = \int \omega^i W\left(u^i\left((1-\tau)z^i + \tau Z((1-\tau),G) - R - \pi G, G, z^i\right)\right) \mathrm{d}\nu(i).$$

The FOC for τ is as before, and the FOC for public good provision G is

$$\int \omega^i W_u\left(u_G^i + u_x^i\left((1-\tau)\frac{\partial z^i}{\partial G} + \tau\frac{\mathrm{d}Z}{\mathrm{d}G} - \pi\right)\right)\mathrm{d}\nu(i) = 0,$$

which produces the following public good provision rule:

$$\frac{\int \omega^i W_u\left(u_G^i + u_x^i(1-\tau)\frac{\partial z^i}{\partial G}\right)\mathrm{d}\nu(i)}{\int \omega^i W_u u_x^i\,\mathrm{d}\nu(i)} = \pi - \tau\frac{\mathrm{d}Z}{\mathrm{d}G}. \tag{39}$$

The left-hand side relates the welfare gains of public good provision (a direct (u_G) and indirect effect ($u_x(1-\tau)\frac{\partial z^i}{\partial G}$ via labor supply reactions)) to the welfare gains of directly increasing consumption (cash transfers) and the right-hand side relates the costs of providing the public good (both its price and the effect it has on tax revenue) to the costs of directly increasing consumption (equal to 1 in this model).[18]

Poverty minimization

Using Tuomala's model, and the deprivation index $D\left(x, G, \bar{x}, \bar{G}\right)$ defined over consumption of the public good G and other private consumption x, we can divide the government's first-order condition for G (analogous to Eq. 37) with that of b (analogous to Eq. 36) to get the following relationship:

$$D^* = \frac{\pi - (1-a)\bar{z}_G}{1-(1-a)\bar{z}_b}, \tag{40}$$

where $D^* = \frac{\sum D_G + \sum D_x a z_G^i}{\sum D_x \left(1 + a z_b^i\right)}$. This can be rewritten to get Eq. (7).

In the Piketty–Saez type of model, individual private consumption is $x = (1-\tau)z^i + b = (1-\tau)z^i + \tau Z((1-\tau), G) - R - \pi G$. The government's problem is then:

$$\begin{aligned}\min\ P &= \int D\left(x^i, G, \bar{x}, \bar{G}\right) \mathrm{d}\nu(i)\\ &= \int D\left((1-\tau)z^i + \tau Z((1-\tau), G) - R - \pi G, G, \bar{x}, \bar{G}\right) \mathrm{d}\nu(i).\end{aligned} \tag{41}$$

The first-order condition for optimal tax τ is unchanged, and the FOC for public good provision is $\int \left[D_G + D_x \left((1-\tau)\frac{\partial z^i}{\partial G} + \tau \frac{\mathrm{d}Z}{\mathrm{d}G} - \pi\right)\right] \mathrm{d}\nu(i) = 0$, which gives the public provision rule of (9).

The poverty minimization problem in the case of provision of a quasi-private good is

$$\begin{aligned}\min\ P &= \int D\left(x^i, s^i, \bar{x}, \bar{s}\right) \mathrm{d}\nu(i)\\ &= \int D\left((1-\tau)z^i + \tau Z((1-\tau), G) - R - \pi G - ph^i, s^i, \bar{x}, \bar{s}\right) \mathrm{d}\nu(i).\end{aligned} \tag{42}$$

The FOC for public good provision G is $\int \left[D_x \left((1-\tau)\frac{\partial z^i}{\partial s}\frac{\partial s}{\partial G} + \tau \frac{\mathrm{d}Z}{\mathrm{d}G} - \pi - p\frac{\partial h^i}{\partial G}\right) + D_s \frac{\partial s^i}{\partial G}\right] \mathrm{d}\nu(i) = 0$, which gives the public provision rule (10).

[18] In Eq. (39), we could define a normalized marginal social welfare weight, similar as before, $\beta^i = \frac{\omega^i W_u u_x^i}{\int \omega^i W_u u_x^i \mathrm{d}\nu(i)}$ to get $\frac{\int \omega^i W_u u_G^i \,\mathrm{d}\nu(i)}{\int \omega^i W_u u_x^i \,\mathrm{d}\nu(i)} + \int \beta^i (1-\tau)\frac{\partial z^i}{\partial G} \mathrm{d}\nu(i) = \pi - \tau \frac{\mathrm{d}Z}{\mathrm{d}G}$.

Commodity taxation

Welfarism

The Lagrangian of the government's optimization problem is the following:

$$L = \sum_i W\left(V^i(b,q)\right) + \lambda\Big(\sum_i\sum_j t_j x_j^i - Nb - R\Big). \tag{43}$$

The first-order conditions with respect to b and q_k are:

$$\sum_i \beta^i + \lambda \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial b} - \lambda N = 0 \tag{44}$$

$$-\sum_i \beta^i x_k^i + \lambda \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial q_k} + \lambda \sum_i x_k^i = 0, \tag{45}$$

where Roy's identity has been used in (45), i.e., $\frac{\partial V^i}{\partial q_k} = -\frac{\partial V^i}{\partial b} x_k^i$. Using the definition of γ^i, this means that (44) can be rewritten as

$$\frac{\sum_i \gamma^i}{N} = \lambda, \tag{46}$$

implying that the average net social marginal utility of income must equal the shadow price of budget revenues at the optimum. Next use the definition of γ^i and the Slutsky equation for the commodity demand

$$\frac{\partial x_j^i}{\partial q_k} = \frac{\partial \tilde{x}_j^i}{\partial q_k} - x_k^i \frac{\partial x_j^i}{\partial b},$$

where $\tilde{x}_j^i$ denotes the compensated (Hicksian) demand for good x_j^i, in (45), to get

$$\sum_i \sum_j t_j \frac{\partial \tilde{x}_j^i}{\partial q_k} = \frac{1}{\lambda} \sum \left(\gamma^i - \lambda\right) x_k^i. \tag{47}$$

The covariance between γ^i and the demand of the good x_k can be written as (using (46))

$$\operatorname{cov}\left(\gamma^i, x_k^i\right) = \frac{\sum_i \gamma^i x_k^i}{N} - \frac{\sum_i \gamma^i}{N}\frac{\sum_i x_k^i}{N} = \frac{\sum_i \gamma^i x_k^i}{N} - \lambda \frac{\sum_i x_k^i}{N}.$$

Using Slutsky symmetry, Eq. (47) can therefore be written as a covariance rule (13).

Poverty minimization

The deprivation index to be minimized is $D\left(\sum_j q_j x_j^i, \bar{c}\right)$. The first-order conditions with respect to b and q_k are:

$$\sum_i D_c \sum_j q_j \frac{\partial x_j^i}{\partial b} + \lambda \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial b} - \lambda N = 0 \tag{48}$$

$$\sum_i D_c x_k^i + \sum_i D_c \sum_j q_j \frac{\partial x_j^i}{\partial q_k} + \lambda \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial q_k} + \lambda \sum_i x_k^i = 0. \tag{49}$$

Using the Slutsky equation in Eq. (49) and dividing by N leads to

$$\frac{1}{N} \sum_i D_c x_k^i + \frac{1}{N} \sum_i D_c \sum_j q_j \left(\frac{\partial \tilde{x}_j^i}{\partial q_k} - x_k^i \frac{\partial x_j^i}{\partial b} \right) + \frac{\lambda}{N} \sum_i \sum_j t_j \left(\frac{\partial \tilde{x}_j^i}{\partial q_k} - x_k^i \frac{\partial x_j^i}{\partial b} \right) + \frac{\lambda}{N} \sum_i x_k^i = 0. \tag{50}$$

Multiplying Eq. (48) by $\frac{\sum_i x_k^i}{N^2}$ and adding it with Eq. (50) gives

$$\frac{1}{N} \sum_i D_c x_k^i + \frac{1}{N} \sum_i D_c \sum_j q_j \frac{\partial \tilde{x}_j^i}{\partial q_k} - \frac{1}{N} \sum_i \sum_j D_c q_j x_k^i \frac{\partial x_j^i}{\partial b} + \frac{1}{N} \frac{\sum_i D_c}{N} \sum_j q_j \frac{\partial x_j^i}{\partial b} \sum_i x_k^i + \frac{\lambda}{N} \sum_i \sum_j t_j \frac{\partial \tilde{x}_j^i}{\partial q_k} - \frac{\lambda}{N} \sum_i \sum_j t_j x_k^i \frac{\partial x_j^i}{\partial b} + \frac{1}{N} \frac{\lambda}{N} \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial b} \sum_i x_k^i = 0. \tag{51}$$

Noticing that the covariance of γ_P^i and x_k^i can be written as $\frac{1}{N} \sum_i \sum_j D_c q_j x_k^i \frac{\partial x_j^i}{\partial b} + \frac{\lambda}{N} \sum_i \sum_j t_j x_k^i \frac{\partial x_j^i}{\partial b} - \frac{1}{N} \frac{\sum_i D_c}{N} \sum_j q_j \frac{\partial x_j^i}{\partial b} \sum_i x_k^i - \frac{1}{N} \frac{\lambda}{N} \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial b} \sum_i x_k^i$, the rule above can be written as Eq. (15) in the main text.

Non-optimality of uniform commodity taxation

We demonstrate formally how uniform commodity taxation is not optimal in the case of poverty minimization. To see this, rewrite first the FOC with respect to b (Eq. 48) as

$$\frac{1-\frac{1}{N}\sum_i\sum_j t_j\frac{\partial x_j^i}{\partial b}}{\frac{1}{N}\sum_i D_c\sum_j q_j\frac{\partial x_j^i}{\partial b}}=\frac{1}{\lambda}. \tag{52}$$

Next, rewriting the FOC for q_k (Eq. 50) yields

$$\frac{1}{N}\sum_i\sum_j t_j\frac{\partial\tilde{x}_k^i}{\partial q_j}=-\frac{1}{\lambda}\frac{1}{N}\sum_i D_c x_k^i-\frac{1}{\lambda}\frac{1}{N}\sum_i D_c\sum_j q_j\frac{\partial\tilde{x}_k^i}{\partial q_j}$$
$$+\frac{1}{\lambda}\frac{1}{N}\sum_i D_c\sum_j q_j\frac{\partial x_j^i}{\partial b}x_k^i+\frac{1}{N}\sum_i\left(\sum_j t_j\frac{\partial x_j^i}{\partial b}-1\right)x_k^i. \tag{53}$$

Here we can substitute for $\frac{1}{\lambda}$ from Eq. (52) in the first term at the lower row of Eq. (53). Following Deaton (1979, pp. 359–360), when preferences are separable and Engel curves are linear, demand is written as $x_j^i=\delta_j^i(q)+\theta_j(q)c^i$; hence, the derivative of demand with respect to disposable income c or transfer b is $\theta_j(q)$, i.e., independent of the person i. By writing out explicitly the solution that the derivative of demand w.r.t b is independent of i and write $\frac{\partial x_j^i}{\partial b}=\theta_j(q)$, we have:

$$\frac{1}{N}\sum_i\sum_j t_j\frac{\partial\tilde{x}_k^i}{\partial q_j}=-\frac{1}{\lambda}\frac{1}{N}\sum_i D_c x_k^i-\frac{1}{\lambda}\frac{1}{N}\sum_i D_c\sum_j q_j\frac{\partial\tilde{x}_k^i}{\partial q_j}$$
$$+\frac{1-\frac{1}{N}\sum_i\sum_j t_j\theta_j(q)}{\sum_i D_c\left(\sum_j q_j\theta_j(q)\right)}\sum_i D_c x_k^i\left(\sum_j q_j\theta_j(q)\right)$$
$$+\frac{1}{N}\sum_i x_k^i\left(\sum_j t_j\theta_j(q)-1\right), \tag{54}$$

where in the second row we can cancel out the $\sum_j q_j\theta_j(q)$ terms and rewrite $\sum_i\sum_j t_j\theta_j(q)=N\sum_j t_j\theta_j(q)$ in the numerator because the term is independent over i:

$$\frac{1}{N}\sum_i\sum_j t_j\frac{\partial\tilde{x}_k^i}{\partial q_j}=-\frac{1}{\lambda}\frac{1}{N}\sum_i D_c x_k^i-\frac{1}{\lambda}\frac{1}{N}\sum_i D_c\sum_j q_j\frac{\partial\tilde{x}_k^i}{\partial q_j}$$
$$+\frac{1-\sum_j t_j\theta_j(q)}{\sum_i D_c}\sum_i D_c x_k^i+\frac{1}{N}\sum_i x_k^i\left(\sum_j t_j\theta_j(q)-1\right). \tag{55}$$

Note next that due to homogeneity of degree 0 of compensated demand, $\sum_j q_j \frac{\partial \tilde{x}_k^i}{\partial q_j} + w_i \frac{\partial \tilde{x}_k^i}{\partial w_i} = 0$. This, together with the observation that if a uniform commodity tax t was a solution to the problem at hand, this would mean that the left-hand side of (53) could be written as $-\frac{t}{N} \sum_i w_i \frac{\partial \tilde{x}_k^i}{\partial w}$. Because of separability, the substitution response is linked to the full income derivative, so that $\frac{\partial \tilde{x}_k^i}{\partial w} = \phi^i \theta_j(q)$. Because of these arguments, (53) becomes

$$-\frac{t}{N}\theta_j(q) \sum_i w_i \phi^i = -\frac{1}{\lambda}\frac{1}{N} \sum_i D_c x_k^i - \frac{1}{\lambda}\frac{1}{N}\theta_j(q) \sum_i D_c w_i \phi^i + \frac{1 - t\sum_j \theta_j(q)}{\sum_i D_c} \sum_i D_c x_k^i + \frac{1}{N} \sum_i x_k^i \left(t \sum_j \theta_j(q) - 1 \right). \quad (56)$$

Note that terms incorporating $\theta_j(q)$ cannot be canceled out from the equation so the result remains dependent on j. In addition, even if the terms were canceled, the term $\sum_i D_c \frac{x_k^i}{N}$ still depends on j. This shows that uniform commodity taxation is not optimal when the objective function of the government is to minimize poverty.

Optimal income taxation with an informal sector

Welfarism

The welfarist Lagrangian, in the presence of informality, is $L = \sum W\left(V^i(a, b, e)\right) + \lambda((1-a)\sum(z^i - e^i) - Nb - R)$. We can denote the effective tax base as $z^e = z - e$. The derivative of this tax base with respect to tax rate a is denoted $z_a^e = z_a - \frac{\partial e}{\partial a}$, where we assume $\frac{\partial e}{\partial a} < 0$ (whereas $\frac{\partial e}{\partial b} = 0$). The first-order conditions with respect to a and b are:

$$\sum W_V V_a^e = \lambda \left(\sum z^e - (1-a) \sum z_a^e \right)$$
$$\sum W_V V_b = \lambda \left(N - (1-a) \sum z_b \right),$$

where V_a^e is a shorthand for the derivative of the indirect utility function that takes individual evasion behavior into account. Should there be no evasion, the individual would maximize her utility over income $az + b$ and $V_a = \lambda z$. Under evasion, consumption is $a(z-e)+e+b$ and, by the envelope theorem, $V_a^e = \lambda(z-e) = \lambda z^e$. Roy's theorem adapts in this case to: $V_a^e = V_b z^e$, and welfare-weighted average income can be denoted as $z^e(\beta) = \frac{\sum \beta^i z^{e,i}}{\sum \beta}$. The ratio of the first-order conditions is:

$$z^e(\beta) = \frac{\bar{z}^e - (1-a)\bar{z}_a^e}{1 - (1-a)\bar{z}_b},$$

and we can derive the optimal tax rate by following the same steps as in the model without evasion, by considering the evasion-modified tax base z^e instead of z:

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\varepsilon^e}\left(1 - \frac{z^e(\beta)}{\bar{z}^e}\right).$$

The intuition behind the derivation and the tax rule is the same as before, but we must consider the relevant tax base in the context of evasion. Both the elasticity of labor income with respect to the tax rate and the relevant welfare concepts change when part of the income base evades taxation.

Poverty minimization

The derivation of Eq. (17) follows the same steps as presented above and in the poverty-minimization model without evasion. The first-order conditions with respect to a and b are:

$$\sum D_c\left(\frac{\partial e^i}{\partial a} + z^e + az_a^e\right) = \lambda\left(\sum z^e - (1-a)\sum z_a^e\right)$$
$$\sum D_c(1 + az_b^i) = \lambda\left(N - (1-a)\sum z_b^i\right).$$

From the ratio of the two conditions we get the measure of relative deprivation impact under tax evasion, $\tilde{D}^e$:

$$\tilde{D}^e \equiv \frac{\sum D_c\left(z^e + az_a^e + \frac{\partial e}{\partial a}\right)}{D_c(1+az_b^i)} = \frac{\bar{z}^e - (1-a)\bar{z}_a^e}{1-(1-a)\bar{z}_b},$$

which gives us Eq. (17) in the text. $\tilde{D}^e$ measures the relative efficiency of taxes and transfers. The latter impact (the denominator) is the same as before, but the impact of taxation (numerator) is different in the presence of tax evasion.

Appendix 2: Measuring poverty

One of the most popular poverty measures is the P_α category developed by Foster et al. (1984). It is usually written in the form of $P_\alpha = \int_0^z \left(\frac{z-y}{z}\right)^\alpha f(y)\,\mathrm{d}(y)$ where z is the poverty line and y is income. Defining the poverty index in terms of disposable income (as in Kanbur and Keen 1989 for example), the measure becomes: $P_\alpha = \int_0^{\bar{c}} \left(\frac{\bar{c}-c^i}{\bar{c}}\right)^\alpha \mathrm{d}\nu(i)$. c^i is disposable income, which is defined in the Piketty–Saez model as $c^i = (1-\tau)z^i + b = (1-\tau)z^i + \tau Z(1-\tau) - R$. We can use this specification of the functional form of deprivation to define the derivative $D_c = -\frac{\alpha}{\bar{c}}\left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1}$ (note that $D_c < 0$ as long as $c^i < \bar{c}$). We can follow the same steps to arrive at the optimal tax rate $\tau^* = \frac{1-\bar{\beta}-\bar{\beta}^\varepsilon}{1-\bar{\beta}-\bar{\beta}^\varepsilon+\varepsilon}$ where now

$$\beta^i = \frac{D_c}{\int_0^{\bar{c}} D_c \, \mathrm{d}\nu(i)} = \frac{-\frac{\alpha}{\bar{c}}\left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1}}{\int_0^{\bar{c}} -\frac{\alpha}{\bar{c}}\left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1} \mathrm{d}\nu(i)} = \frac{\left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1}}{\int_0^{\bar{c}}\left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1} \mathrm{d}\nu(i)}$$

and consequently $\bar{\beta} = \frac{\int_0^{\bar{c}} \beta^i z^i \, \mathrm{d}\nu(i)}{Z}$ and $\bar{\beta}^{\varepsilon} = \frac{\int_0^{\bar{c}} \beta^i z^i \varepsilon^i \, \mathrm{d}\nu(i)}{Z}$, as before. Everything else stays exactly the same as in the calculations of Appendix 1. Also in the case of Tuomala's and Dixit and Sandmo's models, the results stay the same, and we can plug in the explicit definition for D_c, the derivative of the poverty measure with respect to disposable income, into the results.

Poverty measurement in the context of public good provision

Employing the FGT poverty measure in the context of public good provision for poverty reduction is more complicated than in the case of just disposable income. In Sect. 3.2 the government's objective function was defined as min $P = \int D\left(x^i, G, \bar{x}, \bar{G}\right) \mathrm{d}\nu(i)$, that is, deprivation was measured both as deprivation in private consumption (i.e., disposable income) as well as with respect to the public good. But the FGT index is a uni-dimensional measure, measuring deprivation with respect to one dimension only (e.g., disposable income). If one wants to consider publicly offered goods such as education as separate from private consumption, a multidimensional FGT measure is needed. Multidimensionality, however, entails a difficult question of determining when a person should be determined as deprived.

There are several approaches to multidimensionality of FGT-type poverty measures.[19] For example, Besley and Kanbur (1988), who consider the poverty impacts of food subsidies, employ the uni-dimensional FGT measure but define deprivation in terms of equivalent income: $P_\alpha = \int_0^z \left(\frac{z_E - y_E}{z_E}\right)^\alpha f(y)\,\mathrm{d}(y)$, where y_E is equivalent income, defined implicitly from $V(p, y_E) = V(q, y)$, and z_E is the poverty line corresponding to equivalent income. But given our aim of defining optimal policy in terms of poverty reduction, irrespective of individual welfare, the use of equivalent income is problematic as it forces the solution to be such that, by definition, individuals are kept as well off as before. Pirttilä and Tuomala (2004) employ shadow prices in a poverty-minimizing context to allow for several goods in the poverty measure. For them, deprivation is measured as $D(z, y(q, w))$ where $z^h = s_x x^* - s_L^h L^*$and $y^h(q, w^h) = s_x x(q, w^h) - s_L^h L(q, w^h)$. This approach requires determining shadow prices s_x, s_L for consumption and leisure in order to construct a reference bundle respective to which deprivation can be measured, but there is no clear guideline to the choice of the shadow prices.

The approach in Bourguignon and Chakravarty (2003) is more suitable for our purposes. They provide a multidimensional extension of the FGT measure, according to which a person is poor if she is deprived in at least one dimension. A simple example of such an extension of the FGT is

19 See Foster et al. (2010, pp. 504–5) for a brief overview of multidimensional FGT extensions that allow the inclusion of dimensions such as health, education, and nutrition in addition to other consumption.

$$P_\theta = \frac{1}{n}\sum_{j=1}^{m}\sum_{i\in S_j} a_j \left(\frac{z_j - x_{ij}}{z_j}\right)^{\theta_j},$$

where θ_j and a_j are weights given to dimension j, and S_j is the group of people who are poor in dimension j. Alkire and Foster (2011) for their part provide a similar measure which uses a weighted count of dimensions in which the person is deprived to determine whether she is poor. An aspect of this is also whether the goods under consideration are complements or substitutes. Following the Bourguignon–Chakravarty approach and defining $x_{i1} = x_i$ as private consumption, $z_1 = \bar{x}$, $x_{i2} = G$ as the amount of public good, and $z_2 = \bar{G}$ would give us $P_\theta = \frac{1}{n}\sum_{i\in S_j}\left(a_1\left(\frac{\bar{x}-x^i}{\bar{x}}\right)^{\theta_1} + a_2\left(\frac{\bar{G}-G}{\bar{G}}\right)^{\theta_2}\right)$. Using this measure, $D_x = -\frac{\theta_1 a_1}{\bar{x}}\left(\frac{\bar{x}-x^i}{\bar{x}}\right)^{\theta_1-1}$ and $D_G = -\frac{\theta_2 a_2}{\bar{G}}\left(\frac{\bar{G}-G}{\bar{G}}\right)^{\theta_2-1}$. These can then be inserted to the public provision rules. For example, (9) becomes

$$\frac{\int\left(\frac{\theta_2 a_2}{\bar{G}}\left(\frac{\bar{G}-G}{\bar{G}}\right)^{\theta_2-1} + \frac{\theta_1 a_1}{\bar{x}}\left(\frac{\bar{x}-x^i}{\bar{x}}\right)^{\theta_1-1}(1-\tau)\frac{\partial z^i}{\partial G}\right)\mathrm{d}\nu(i)}{\int \frac{\theta_1 a_1}{\bar{x}}\left(\frac{\bar{x}-x^i}{\bar{x}}\right)^{\theta_1-1}\mathrm{d}\nu(i)} = p - \tau\frac{\mathrm{d}Z}{\mathrm{d}G}$$

and (40) becomes

$$\frac{\sum \frac{\theta_2 a_2}{\bar{G}}\left(\frac{\bar{G}-G}{\bar{G}}\right)^{\theta_2-1} + \sum \frac{\theta_1 a_1}{\bar{x}}\left(\frac{\bar{x}-x^i}{\bar{x}}\right)^{\theta_1-1} a z^i_G}{\sum \frac{\theta_1 a_1}{\bar{x}}\left(\frac{\bar{x}-x^i}{\bar{x}}\right)^{\theta_1-1}\left(1 + a z^i_b\right)} = \frac{p - (1-a)\sum z^i_G}{1-(1-a)\sum z^i_b},$$

from where it can be seen that the relative efficiency of the public good versus cash transfers on reducing poverty can be directly traced back to the magnitudes of θ_1 and θ_2.

Poverty measurement in the context of commodity taxation

In the case of commodity taxes, we run into the same issues regarding deprivation measurement as with public goods. However, in Sect. 4.2 deprivation was measured only in terms of disposable income, c. We thus escape the multidimensionality issue and employing the FGT poverty measure is thus as simple as in the linear income tax case: we simply need to define $D = P^\alpha$ and thus $D_c = -\frac{\alpha}{\bar{c}}\left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1}$ in Eq. (15). Potentially the government might also consider weighting different goods according to their importance to measured poverty.

References

Aitchison, J., & Brown, J. A. C. (1957). *The lognormal distribution*. Cambridge: Cambridge University Press.

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, *95*, 476–487.

Barrientos, A. (2013). *Social assistance in developing countries*. Cambridge: Cambridge University Press.

Besley, T., & Kanbur, R. (1988). Food subsidies and poverty alleviation. *The Economic Journal*, *98*, 701–719.

Besley, T., & Persson, T. (2013). Taxation and development. In A. J. Auerbach, R. Chetty, M. Feldstein, & E. Saez (Eds.), *Handbook of public economics* (Vol. 5, pp. 51–110). Amsterdam: Elsevier.

Bird R.M., & Gendron, P. (2007). The VAT in developing and transitional countries. Cambridge: Cambridge University Press.

Boadway, R., Marchand, M., & Pestieau, P. (1994). Towards a theory of the direct–indirect tax mix. *Journal of Public Economics*, *55*, 71–88.

Bourguignon, F., & Chakravarty, S. (2003). The measurement of multidimensional poverty. *Journal of Economic Inequality*, *1*, 25–49.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, *105*, 1–33.

Deaton, A. (1979). Optimally uniform commodity taxes. *Economics Letters*, *2*, 357–361.

Diamond, P. (1975). A many-person Ramsey tax rule. *Journal of Public Economics*, *4*, 335–342.

Dixit, A., & Sandmo, A. (1977). Some simplified formulae for optimal income taxation. *Scandinavian Journal of Economics*, *79*, 417–423.

Farhi, E., & Gabaix, X. (2015). Optimal taxation with behavioral agents. Working Paper 21524, National Bureau of Economic Research.

Fleurbaey, M., & Maniquet, F. (2007). Help the low skilled or let the hardworking thrive? A study of fairness in optimal income taxation. *Journal of Public Economic Theory*, *9*, 467–500.

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, *52*, 761–766.

Foster, J., Greer, J., & Thorbecke, E. (2010). The Foster-Greer-Thorbecke (FGT) poverty measures: 25 years later. *Journal of Economic Inequality*, *8*, 491–524.

Gerritsen, A. (2016). Optimal taxation when people do not maximize well-being. *Journal of Public Economics*, *144*, 122–139.

Gordon, R., & Li, W. (2009). Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics*, *93*, 855–866.

IFS & Mirrlees, J. (2011). *Tax by design*. Oxford: Oxford University Press.

Kanbur, R. (2015). *Informality: Causes, consequences and policy responses*. Discussion Paper 10509, Centre for Economic Policy Research.

Kanbur, R., & Keen, M. (1989). Poverty, incentives and linear income taxation. In A. Dilnot & I. Walker (Eds.), *The economics of social security* (pp. 99–115). Oxford: Clarendon Press.

Kanbur, R., & Keen, M. (2014). Thresholds, informality, and partitions of compliance. *International Tax and Public Finance*, *21*, 536–559.

Kanbur, R., Keen, M., & Tuomala, M. (1994). Optimal non-linear income taxation for the alleviation of income-poverty. *European Economic Review*, *98*, 701–719.

Kanbur, R., Pirttilä, J., & Tuomala, M. (2006). Non-welfarist optimal taxation and behavioural public economics. *Journal of Economic Surveys*, *20*, 849–868.

Keen, M. (2009). What do (and don't) we know about the value added tax? A review of Richard M. Bird and Pierre-Pascal Gendron's *The VAT in developing and transitional countries*. *Journal of Economic Literature*, *47*, 159–170.

Keen, M. (2012). Taxation and development: Again. Working Paper 12/220, International Monetary Fund.

Keen, M. (2014). Targeting, cascading and indirect tax design. *Indian Growth and Development Review*, *7*, 181–201.

Keen, M., Kim, Y., & Varsano, R. (2008). The "flat tax(es)": Principles and experience. *International Tax and Public Finance*, *15*, 712–751.

Lee, D., & Saez, E. (2012). Optimal minimum wage policy in competitive labor markets. *Journal of Public Economics*, *96*, 739–749.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, *38*, 175–208.

Peichl, A. (2014). Flat-rate tax systems and their effect on labor markets. IZA World of Labor, 61.

Piketty, T., & Saez, E. (2013). Optimal labor income taxation. In A. J. Auerbach, R. Chetty, M. Feldstein, & E. Saez (Eds.), *Handbook of Public Economics* (Vol. 5, pp. 391–474). Amsterdam: Elsevier.

Pirttilä, J., & Tuomala, M. (2004). Poverty alleviation and tax policy. *European Economic Review*, *48*, 1075–1090.

Roemer, J. E., Aaberge, R., Colombino, U., Fritzell, J., Jenkins, S. P., Lefranc, A., et al. (2003). To what extent do fiscal regimes equalize opportunities for income acquisition among citizens? *Journal of Public Economics*, *87*, 539–565.

Saez, E., & Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, *106*, 24–45.

Sen, A. (1985). *Commodities and capabilities*. Amsterdam: North Holland.

Slemrod, J. (1990). Optimal taxation and optimal tax systems. *Journal of Economic Perspectives*, *4*, 157–178.

Slemrod, J., & Gillitzer, C. (2014). Insights from a tax-systems perspective. *CESifo Economic Studies*, *60*, 1–31.

Stern, N. (1976). On the specification of models of optimum income taxation. *Journal of Public Economics*, *6*, 123–62.

Tuomala, M. (1985). Simplified formulae for optimal linear income taxation. *Scandinavian Journal of Economics*, *87*, 668–672.

Tuomala, M. (1990). *Optimal income tax and redistribution*. Oxford: Oxford University Press.