FINNISH CENTRE FOR PENSIONS,
STUDIES

02
2020

New methods in pension evaluation

Applications of trajectory analysis and dynamic microsimulation

JANNE SALONEN

FINNISH CENTRE FOR PENSIONS, STUDIES 02/2020

New methods in pension evaluation

Applications of trajectory analysis and dynamic microsimulation

JANNE SALONEN

Finnish Centre for Pensions
FI-00065 ELÄKETURVAKESKUS, FINLAND
Phone: +358 29 411 20
E-mail: firstname.surname@etk.fi

Eläketurvakeskus
00065 ELÄKETURVAKESKUS
Puhelin: 029 411 20
Sähköposti: etunimi.sukunimi@etk.fi

Pensionsskyddscentralen
00065 PENSIONSSKYDDSCENTRALEN
Telefon: 029 411 20
E-post: förnamn.efternamn@etk.fi

Hansaprint Oy | Turenki 2020

ISBN 978-951-691-313-4 (s-b.)
ISBN 978-951-691-314-1 (PDF)

ISSN 1236-3049 (printed)
ISSN 1798-7482 (online)

MADE IN FINLAND

Supervisors
Professor Lasse Koskinen
Insurance Science
Business Studies, Faculty of Management and Business, Tampere University, Finland

Research Director, Docent Tapio Nummi
Statistics
Computing Sciences, Faculty of Information Technology and Communication Sciences, Tampere University, Finland

Reviewers
Research Director, Docent Niku Määttänen
Research Institute of the Finnish Economy (ETLA) and Aalto University, Finland

Professor Asko Tolvanen
Psychology
Methodology Centre for Human Sciences, Faculty of Education and Psychology, University of Jyväskylä, Finland

Dissertation Opponent
Research Director, Docent Niku Määttänen
Research Institute of the Finnish Economy (ETLA) and Aalto University, Finland

Author's Address
Janne Salonen
Finnish Centre for Pensions
FI-00065 Eläketurvakeskus, Finland

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the Tampere University.

New methods in pension evaluation

Applications of trajectory analysis and dynamic microsimulation

JANNE SALONEN

ACADEMIC DISSERTATION To be presented,
with the permission of the Faculty of Management
and Business of Tampere University, for public
discussion in Pinni Building / Paavo Koli Auditorium
on 28 August 2020, at 12 o'clock.

ACKNOWLEDGEMENTS

I am highly impressed about the way that Lasse Koskinen and Tapio Nummi have supervised me through this thesis. Your enthusiasm, talent and wisdom have challenged and inspired me. You have encouraged me to bring this thesis into completion. I cannot enough thank you for your help and support. Writing this thesis and sub-studies has been a great adventure in science.

I want to thank pre-examiners Asko Tolvanen and Niku Määttänen for reviewing my thesis. I wish to thank Niku Määttänen for accepting the invitation to be the opponent at my disputation. I extend my warm thanks to Vilma Itäniemi and Merja Raunis for the layout of the dissertation and David Kivinen and Lena Koski for revising the language of this thesis and the sub-studies – now wider audiences can take part of my research.

In the last few years, working with the sub-studies, I have been honored to work and write with great colleagues and scholars: Heikki Tikanmäki, Hannu Sihvonen, Antti Saloniemi, Pekka Virtanen, Timothy O'Brien, Sampo Lappo and Sanna Tenhunen. Many other people at the Finnish Centre for Pensions and in the international microsimulation community have contributed, as well. All of you have forced me to write correctly, think broadly and be scientific. It has been great working with you, thank you for sharing your time and talent with me. I will never forget our discussions.

I want to thank the Finnish Centre for Pensions for giving me free hands to create the Research Register for the benefit of so many scholars. Thank you Mikko Kautto and Susan Kuivalainen for your support and believe in my stubborn mission with the registers. Moreover, I have many good friends and colleagues working with the administrative registers – you are all special to me.

During my career I have been privileged to work with former colleagues such as Bo Lundqvist, Mikael Forss and Mervi Takala. I would not be in this situation without you. Thank you friends, I will never forget your kindness and support.

This thesis was written during a challenging time in my personal life. My children, Hiljatuuli and Edwin, you have given me trust in life. You are

examples of hope and a candid belief in the future, despite everything, and I am proud to be your father. I want to thank my mother Terttu and my aunt Anneli for your endless support and belief in me. I also want to thank Mika Lindfors for giving me new perspectives in life during these past difficult years. I address my deepest gratitude to Marika for your unconditional love. Thank you all for being there.

In Espoo, June 2020

Janne Salonen

ABSTRACT

This thesis has three aims. First, to introduce and apply two new techniques in pension evaluation: trajectory analysis and dynamic microsimulation. Trajectory analysis has been used in many fields of science, and it has proven to be a useful technique in longitudinal analysis. Trajectory analysis can be used to reveal latent sub-groups of population that might otherwise remain hidden. Early adulthood labor market attachment paths of the Finnish male cohort born in 1987 is the focus of one sub-study, and pension contribution behavior of the self-employed is the focus of another. The Finnish Centre for Pensions has developed a dynamic microsimulation model, ELSI. The model can be used to analyze the effects of changes to pension rules on pensions. The model is presented in one sub-study. Also, a study that merges dynamic microsimulation and trajectory analysis yields new insights into microsimulation results and advances the validation of the microsimulation model.

The second aim of this thesis is to discuss the substantial contents of two central topics. First, how to detect those young people who, after completing compulsory education, experience difficulties with labor market attachment, as well as how to measure labor market attachment in general. The study shows that about 12 per cent of each cohort is at risk of becoming a NEET. Second, a prolonged problem in the pension scheme for the Finnish self-employed has been that they choose to pay pension contributions that are too low (based on a technically confirmed income), which has led to deficits in pension security. Trajectory analysis is used to reveal the sub-groups of the self-employed with different levels of confirmed income. A majority, nearly 84 per cent of the self-employed, is financially ill-prepared for old age.

The third aim of this dissertation is to introduce new concepts and measures to labor market research. The concept of a cross-sectional NEET is challenged by introducing a longitudinal measure that includes the labor market statuses of the concept. The traditional measures of life course studies – duration of working life and working life expectancy – can be complemented with a new measure: partition of the life course. For pension evaluation, the partition might be enlightening as it reveals the shares of active stages (employment) and inactive stages (e.g., retirement, spells on social security benefits) in a lifespan.

Key terms:

mixture regression, trajectory analysis, dynamic microsimulation, NEET, self-employed, statutory pension

TIIVISTELMÄ

Tällä tutkimuksella on kolme tavoitetta. Ensimmäinen osa tutkimuksen tavoite on esitellä trajektorianalyysi ja dynaamisen mikrosimuloinnin menetelmä, sekä osoittaa niiden käytännöllisyys eläketurvan arvioinnissa. Trajektorianalyysi soveltuu pitkittäisaineiston analysointiin ja sitä onkin käytetty monilla tieteenaloilla, nykyään myös eläketutkimuksessa. Trajektorianalyysin keskeinen idea on löytää tutkittavasta jakaumasta mahdolliset piilevät osajakaumat. Ensimmäisen osatutkimuksen aihe on löytää vuonna 1987 syntyneen mieskohortin työmarkkinapolut oppivelvollisuuden jälkeen, elinkaaren alkuvaiheessa. Toisen osatutkimuksen aiheena on yrittäjien YEL-vakuutusmaksuun liittyvät valinnat. Kolmannessa osatutkimuksessa esitellään ELSI-malli ja sen toiminnallisuudet. Neljännessä osatutkimuksessa sovelletaan trajektorianalyysiä ELSI-mallin tulosaineistoon.

Tutkimuksen toinen tavoite on analysoida kahta sisällöllisesti keskeistä aihetta. Ensimmäinen tutkimusaihe on, tunnistaa nuoren kohortin työmarkkinakiinnittymisen polut ja erityisesti ne, joilla kiinnittymiseen liittyy ongelmia. NEET-käsitteeseen liittyvä mittaaminen on lisäksi tärkeä kysymys. Tulosten mukaan noin 12 prosenttia mieskohortista on pitkäaikaisesti vaarassa ajautua heikon kiinnittymisen poluille. Toinen tutkimusaihe liittyy yrittäjien valitsemaan, liian pieneksi osoittautuneeseen, YEL-työtuloon. Trajektorianalyysin perusteella 84 prosenttia yrittäjistä valitsee liian pienen työtulon, mikä voi johtaa vajaaseen eläketurvaan.

Tutkimuksen kolmas tavoite on esitellä työmarkkinatutkijoille uusia mittareita. Tilastollinen NEET-mittari perustuu yhden vuoden havaintoon, vaikka työmarkkinakiinnittyminen ilmiönä on ajassa muuttuva. Trajektorianalyysin avulla voidaan osoittaa, että mittarin työmarkkinatila on hyödyllistä mallintaa samanaikaisesti. Elinkaaren eri vaiheiden työllisyyttä kuvaavia mittareita voidaan täydentää mikrosimulaation avulla saatavalla elinajan ositteella. Uusi mittari voi olla hyödyllinen eläkepolitiikan arvioinnissa, koska sillä voidaan yksinkertaisesti kuvata työllisen ajan ja eläkeajan osuuksia elinkaarella.

CONTENTS

ORIGINAL PUBLICATIONS

The thesis consists of the following publications, which are referred to in the text by Roman numerals.

I Nummi, Tapio & Salonen, Janne & O'Brien, Timothy (2017) Statistical analysis of labor market integration: A mixture regression approach. In D.-G. Chen et al. eds. New advances in statistics and data science, ICSA book series in statistics, Springer International Publishing AG 2017. https://doi.org/10.1007/978-3-319-69416-0_18S

II Salonen, Janne & Koskinen, Lasse & Nummi, Tapio. Underinsurance in the statutory pension scheme: A trajectory analysis of self-employed earnings in Finland. Manuscript accepted for publication into the International Social Security Review.

III Tikanmäki, Heikki & Sihvonen, Hannu & Salonen, Janne (2015) Distributional effects of the forthcoming Finnish pension reform – A dynamic microsimulation approach. International journal of microsimulation (2015) 8(3) 75–98.

IV Salonen, Janne & Tikanmäki, Heikki & Nummi, Tapio (2019) Using trajectory analysis to test and illustrate microsimulation outcomes. International journal of microsimulation 2019; 12(2). DOI: https://doi.org/10.34196/ijm.00198

1 Introduction

Statutory pensions play a large role in the Finnish social security system in terms of maintaining livelihood in old age, and in case of disability or widowhood. The scope and role of pensions has also guaranteed a continuous interest of policy makers and citizens in pension system outcomes and pension policy.

Pension policy has been evaluated from different angles with the help of theoretical frameworks of social policy, sociology, insurance science and economics. The mainstream in pension evaluation has been empirical analysis with a mixture of descriptive and confirmatory approaches that use statistical techniques and rule-based predictions.

One aim of this thesis is to introduce new techniques into pension evaluation. A statistical technique called trajectory analysis (e.g., Nagin, 2005) is an application of finite mixture modeling. It has been used in many fields of science, but now increasingly also in life course analysis, making it useful in pension evaluation. Another technique developed in recent years is dynamic microsimulation (e.g., O'Donoghue, 2014). The ELSI model of the Finnish Centre for Pensions is now in a stage where it can be used in pension evaluation on a regular basis.

These techniques allow research to focus on longitudinal study-designs with register data. In general, register-based pension evaluation focuses on three research areas: the livelihood of pensioners, work ability and working lives. Both current and future pensioners are of great interest.

For decades, pension evaluation has been a task mainly of the Finnish Centre for Pensions, although other research institutes and academia have also carried out pension evaluation. As an example of the Finnish Centre for Pensions' investment in research is the ELSI microsimulation model. It is unlikely that another institution could have built a dynamic microsimulation model to study pension policy, including earnings-related pensions.

In this context, pension evaluation refers to pension-related research, mainly research where the individual and their current and future livelihood is at the center of focus. The range of pension evaluation also covers other useful studies which do not necessarily meet scientific publication criteria. Pension relevancy is a broad definition, as most labor market behavior over the life course is at some stage relevant in terms of pensions. Recognizing that broader macroeconomic topics and features relating to pension financing also affect individuals, at least indirectly, is also relevant. From a life course perspective, the whole individual lifespan is relevant in pension evaluation, which again draws a range of human issues into the focal point, whether they are socially insurable or not.

The sub-studies of this thesis constitute a collection of two scientific traditions, namely finite mixture analysis for longitudinal data (also called trajectory analysis) and dynamic microsimulation. Trajectory analysis has been used in labor market studies for more than a decade. The applications of the technique first appeared in scientific literature around the year 2005. The available software was naturally noticed by labor market researchers. The technique has proved to be invaluable also in pension evaluation. Individual-level microsimulation, on the other hand, has long traditions globally, especially in studies focusing on labor markets, taxation, life courses and pensions. Without understating the tradition of static microsimulation, dynamic microsimulation is a natural approach to pension evaluation, particularly with an ageing population.

In this thesis and the sub-studies, both scientific traditions are introduced independently and merged.

The second aim of this thesis is to discuss the substantial contents of three topics: how to detect those young people who, after completing compulsory education, experience difficulties with labor market attachment; what is the prevalence of underinsurance among the self-employed; and what is the effect of the 2017 pension reform on pension distribution in Finland?

The sub-study I on young people seeks to be intuitive in measurement and in revealing latent development trajectories of labor market attachment. The applied measurement of labor market attachment consists of four separate measures: employment, education, unemployment and spells on other social security benefits. The applied measure resembles the statistical concept of the NEET (acronym for ‘Not in Employment, Education, or Training’). The widely used concept has been studied from many perspectives and in many disciplines (e.g., Furlong, 2006). The sub-study shows the share of a male cohort which faces prolonged difficulties with labor market attachment. The analysis reveals that there are several distinctive sub-groups with weak attachment trajectories.

The sub-study II on the self-employed focuses on the outcomes of choices made. A prolonged problem in relation to the pension scheme for the Finnish self-employed has been that they choose to pay pension contributions that are too small (based on a technically confirmed income), which has led to deficits in their pension security and other basic social security benefits. In the Finnish context, the topic has been evaluated in detail in Benzarti et al. (2019), and also in a Government paper STM (2019). The sub-study uses trajectory analysis to reveal sub-groups of self-employed with different levels of confirmed income. The grim fact is that a vast majority of the self-employed select a confirmed income that is too low compared to their actual salary earnings (declared income in tax registers). This is an indication of underinsurance. The analysis also revealed a group with insurance-in-place (the confirmed income equals the declared income) and a small share of self-employed who have overinsured themselves. There seem to be information failures in the scheme although public information has been available. The sub-study includes further statistical analysis of the composition of the sub-groups and of risk-factors that explain group membership. Statistical analysis and auxiliary information on actual pension outcomes confirm the group-specific results.

The sub-study III introducing the dynamic microsimulation model ELSI to the international microsimulation community focuses on analyzing the effects of the 2017 pension reform in Finland from many perspectives.

The third aim of this thesis is to introduce a few new concepts and measures for discussion. One, the statistical concept NEET (e.g., OECD 2019a) is challenged in sub-study I on young people integrating into the labor market. Clearly the problem of weak labor market attachment is longitudinal in nature, which calls for longitudinal statistical techniques. These young people are not well recognized by established (cross-sectional) statistical indicators like the NEET rate. The rate presents the share of young people who are not in employment, education or training as a percentage of the total number of young people in the corresponding age group, by gender (ibid.). Using a cohort follow-up and trajectory analysis, it was possible to identify the young people with longstanding NEETs. The key issue is the measurement of the NEET statuses themselves. The sub-study demonstrates the use of multi-trajectory modeling, where an individual can occupy several labor market statuses simultaneously. This is sensible as interleaving labor market statuses are a reality.

Two, there is a strong tradition in life course studies to analyze mainly employment, often with an emphasis on some specific stage of life. It is rarely seen to illustrate the entire adult life course. Merging empirical observations from research registers with decades of historical data with individual-level microsimulation projections brings new possibilities to life course measuring. Sub-study III on pension reform evaluation using ELSI introduces a new measure: partitioning of the life course. It illuminates the key stages of the lifespan of different sub-populations.

Three, the essence of trajectory modeling is to reveal latent classes in outcome variable(s). Latent classes are hypothetical constructs that cannot be directly measured. But when revealed, these classes could be used to illustrate the research questions at hand and, furthermore, as complements to ex ante classifiers such as age, gender, or education. This topic is discussed in sub-study IV.

Research ethics cannot be overstated in this context. Ethics are a part of every stage of the research process: when forming the hypothesis, making the study-design, collecting data, using different methods, writing the results and publishing the results. The entire research process should meet different ethical concepts. In the context of this thesis and its articles, the research ethics follows the utilitarian principles, which emphasize the usefulness of the results in an increasing general awareness of social risks and life-course-related issues in general.

Self-evaluating this thesis and the sub-studies, we (I and the co-authors) have committed ourselves to the guidelines of the Finnish Advisory Board on Research Integrity, TENK and European Commission (2010). In my opinion, the key ethical principles are to study minority groups and bring new insights and introduce new statistical approaches into topics that have been studied for decades.

Consequently, I hope this thesis meets the good research writing principles defined by, for example, Clarkeburn and Mustajoki (2007). According to them, it is ethically sound to publish research if a writer: i) participates in the research process (including the data collection), ii) contributes in the writing of the research, and iii) permits the publication process.

The outline of this thesis is the following. Chapter 2 presents the sub-studies, study-designs, research questions and main results. Chapter 3 sets the theoretical framework of social insurance in economics and insurance science and compiles the central concepts of social insurance. Chapter 4 gives an overview of pension-related research in Finland over the past years (with literature being discussed in other chapters, as well). Chapter 5 is an independent discussion of measurements in pension registers, giving a perspective on the measures used in the sub-studies. Chapter 6 provides an overview of finite mixture modeling and lines of thinking in trajectory analysis and collects some lessons from the sub-studies. Chapter 7 presents lines of thinking in microsimulation, painting a picture of the landscape in which the ELSI model can be applied. Chapter 8 introduces the ELSI model in more detail, providing information on the functionalities in which behavioral heterogeneity is generated. Chapter 9 collects the key-results of the sub-studies and presents policy recommendations.

2 Scientific contribution of the articles

2.1 Author’s contribution to the articles

The articles of this thesis are based on scientific work of the author over the past ten years. Research topics are easy to brainstorm, but the data is difficult to collect. Most of the time has been spent on collecting data and material, pondering measurements, constructing indicators, summarizing information and writing. With experienced co-authors, the modeling or statistical analysis is the pleasant part of the research process. The majority of the articles is based on long-time research collaboration between the Finnish Centre for Pensions and Tampere University.

In scientific papers written together with co-authors, it is ethically important to acknowledge the responsibilities of the authors. The ethics of the author’s contribution can be summarized in the following way.

Sub-study I is a result of an ongoing collaboration with Tampere University. The authors of the article came up with the idea to study a particular cohort to find any outcast population sub-groups. The Finnish male cohort born in 1987 was chosen because there are other peer studies of the same cohort. The article is written by Tapio Nummi, the author, and Timothy O’Brien. The statistical analysis is done by the author. The data of the study is drawn from the Finnish Centre for Pensions’ research register by the author. A further article on this topic has also been published (see Saloniemi et al. 2020).

Sub-study II is another collaboration with Tampere University. The article aims to participate the public debate about pension security of the self-employed. Therefore, the article focuses on insurance behavior of the self-employed. The results show the sub-groups of self-employed and their incomes. The statistical analysis is trajectory analysis, done by the author. The article is written by the author, Lasse Koskinen and Tapio Nummi. The data of the study is drawn from Finnish Centre for Pensions’ research register by the author.

Sub-study III is the result of early work on microsimulation. The article includes the author’s main interests in microsimulation, namely income taxation, the income distribution measure or the Gini-coefficient (e.g., Myles 1995, 74–75), and a new measure on life-course partitioning. The article is written by Heikki Tikanmäki, the author, and Hannu Sihvonen. The data was drawn from the Finnish Centre for Pensions' research register by the author and Hannu Sihvonen.

Sub-study IV is the result of new work introducing trajectory analysis into a microsimulation context. The article presents the author's original idea to combine dynamic microsimulation and trajectory analysis, and to demonstrate this to the international microsimulation community. The statistical analysis is trajectory analysis, done by the author. The article is written by the author, Heikki Tikanmäki, and Tapio Nummi. The data is drawn from the ELSI model by the author.

Summary of studies

The leading journal in the fields of insurance economics and risk management, The Journal of Risk and Insurance, outlines eight key areas of research. The sub-studies of this thesis increase the body of knowledge and literature in three areas of specialization: economics of employee benefits, pension plans, and social insurance (see ibid. 1999). Table 1 summarizes the purpose of the sub-studies.

Table 1.
Purpose of sub-studies I–IV.

Name of sub-study	Aim	Research questions/ hypotheses	Main contribution to research literature
I: Statistical Analysis of Labor Market Integration: A Mixture Regression Approach	To identify young individuals with a prolonged risk of weak labor market attachment To introduce longitudinal measures and indicator in studying NEETs	H1: The cross-sectional NEET indicator is limited in depicting longitudinal real-life problems in labor market attachment	With several labor market outcome variables that are measured longitudinally, a multi-trajectory modeling technique is recommended It is useful to consider longitudinal methods, such as trajectory analysis alongside cross-sectional techniques when studying NEETs Trajectory analysis revealed that nearly 12 per cent of young males aged 18 to 26 years are at risk of entering dismal labor market tracks

Name of sub-study	Aim	Research questions/ hypotheses	Main contribution to research literature
II Underinsurance in the statutory pension scheme: A trajectory analysis of self-employed earnings in Finland	To identify and describe groups of self-employed with different pension contribution incentives To evaluate the share of self-employed with insufficient pension contribution i.e. underinsurance	Q1: How common and persistent is under/overinsurance? Q2: Which factors explain under/overinsurance? Q3: What kind of factor-based under/overinsured subgroups are there among the self-employed?	A subjective choice of pension contribution in a compulsory pension system leads to insufficient pension security for the self-employed About 84 per cent of the self-employed persistently select to pay pension contributions that are too low Short-sighted behavior appears among the self-employed relating to the subjective choice in a compulsory pay-as-you-go pension system The contribution ceiling is a necessity as some groups of self-employed would otherwise choose to exploit the compulsory pay-as-you-go pension system

Name of sub-study	Aim	Research questions/ hypotheses	Main contribution to research literature
III: Distributional Effects of the Forthcoming Finnish Pension Reform – A Dynamic Microsimulation Approach	To introduce the ELSI model to the international microsimulation community To analyze the effects of the 2017 pension reform on pension distribution and other measures To demonstrate the developed measures and indicators	Q1: What is the effect of the 2017 pension reform on working lives? Q2: What are the substantial implications of the simulated pension distribution? Q3: How does the level of education affect pensions?	Microsimulation demonstrates how the 2017 pension reform extends working lives by 1–2 years on average Pension rules and labor market outcomes implicate some polarization of insured people within pension distribution Education deeply affects labor market outcomes and, later on, pensions A new measure of life course – partition of the life course – has been introduced and demonstrated
IV: Using Trajectory Analysis to Test and Illustrate Microsimulation Outcomes	To introduce the trajectory analysis technique to the microsimulation community To demonstrate how to apply trajectory analysis together with microsimulation To promote group-based classification rules	Q1: Can trajectory analysis be used to validate microsimulation outcomes? Q2: How is a statistical technique like trajectory analysis applied to simulated data?	The steps and options of trajectory analysis have been introduced to microsimulation practitioners, providing them with examples A discussion of the benefits of data-driven group-based classification rules has been initiated Misspecifications of the microsimulation model have been demonstrated, using trajectory analysis helping model validation

2.2 Statistical analysis of labor market integration: A mixture regression approach

This sub-study is content-, method- and data-driven. First, the content of the article focuses on the males in the 1987 cohort who struggle to find their place in the labor market after completing basic education. The article is the first study of the male cohort born in 1987. Using different data, the same cohort, (called the Finnish Birth Cohort) has been extensively studied in Paananen and Gissler (2013). The Finnish Birth Cohort has also been used to analyze NEETs (Larja et al. 2016). However, in our studies and articles, the point has been to identify those young males who experience difficulties with labor market attachment, that is, the young males who cannot find jobs or get the further education they need. In youth studies, such young people are covered by the concept of NEET. There is a range of scientific literature, especially in the field of sociology, focusing on the topic of NEET (e.g., Furlong 2006; Simmons and Thompson 2011; Serracant 2014).

The idea of our paper is to add a longitudinal perspective to an otherwise cross-sectional measure of the NEET (e.g., Eurostat 2018; Official Statistics of Finland 2019a). The methodology that we apply can reveal latent groups of young people with different labor market attachments. Our aim is to use the method to reveal those young people who experience long-term or persistent difficulties in labor market attachment. The measure or indicator that we model combines the essential contents of the NEET concept.

In this sub-study, we selected to use several outcome variables to measure longitudinal labor market attachment. The idea is to measure the attachment using the data available for measuring labor market attachment in the NEET concept. We selected the indicators employment, time spent in education leading to graduation/a degree, and unemployment and other social security benefits. This indicator is a result of the new data possibilities covering these measures of labor market attachment in a reliable way and spanning a reasonably long time-period. Based on our findings and those of others, it would be useful to postulate a new term for longitudinal NEETs. Our suggestion is deep NEET, which would put the focus on the longitudinal nature of this topic.

The approach of this sub-study is empirical. We have no clearly stated research question, but we seek to identify latent labor market attachment groups of the selected cohort. Our data consists of Finnish males born in

1987 (N=29,383) who lived in Finland over the study period 2005–2013 (age 18 to 26 yrs.). Our study-design is a cohort follow-up. In addition, we aim to illustrate latent groups with different labor market attachment trajectories, especially those with a weak labor market attachment.

Our study uses a multivariate trajectory analysis with the above-mentioned four outcomes over the study period. It is a novelty to use such a composition of measures to analyze, among other things, labor market attachment. We selected a multi-trajectory approach to gain new insight into labor market attachment, using cohort follow-up as the study-design. The statistical foundation of the multi-trajectory analysis is presented in Nagin et al. (2016). The selected technique is a new approach applied to youth labor market issues. We use multivariate binary mixture modeling as we model mixture groups of four binary outcomes simultaneously. To my knowledge, this technique has not described much in previous literature. The novel feature of multiple trajectory modeling is that each individual is evaluated by the different outcomes simultaneously.

Model selection must be clarified a bit further at this point. Mixture analysis is sensitive to the number of sub-groups. The number of groups can be decided by information criteria. For example, BIC or AIC are measures that can help decide the number of groups. During the analysis, we tried different group compositions. BIC values indicated a high number of groups. A subjective evaluation was in favor of ten groups. Plenty has been written on the number of sub-groups (e.g., McLachlan and Rathnayake 2014). Within the ten groups we selected, the outcomes illuminate interesting joint progressions of the labor market outcomes. The solution reveals three weak labor market attachment sub-groups.

The key result is that about 12 per cent (three groups together) of the selected cohort is facing difficulties in labor market attachment. The first group is labelled the (aggregated) LES (Low Employment Status) group. The other side of the coin is that about 88 per cent of the young cohort face a fluent transition from adolescence to adulthood. This group is called the HES (High Employment Status) group. The results are in accordance with the findings in Larja et al. (2016).

As an auxiliary result, the analysis and deep NEET have profound implications for future pension security. The article does not cover the topic of pension accrual in detail. However, the implied measures of the accumulation of working life may have implications on future career opportunities. At least scar in early career employment raise a question mark for old age security. Longitudinal problems in labor market attachment reflect on the accumulation of working life and future earnings possibilities. It is possible, of course, that some of the young who deal with problems relating to labor market attachment find new opportunities in the labor market later on during their life course. However, this is doubtful on a grand scale. We think the young people with a weak labor market attachment may have difficulties in finding employment also in the future. This could lead to a polarization of the cohort in many respects as the deep NEETs differentiate from the majority. One final concern is that the possible scar on working life reflects on old age security as pension accumulation falls short.

2.3 Underinsurance in the statutory pension scheme: A trajectory analysis of self-employed earnings in Finland

This sub-study was possible because a unique dataset was available. The panel data on the self-employed was collected from the registers of the Finnish Centre for Pensions, pension providers and the Tax Administration, covering the years 2012 to 2016. The dataset includes 144,014 individuals who can be described as traditional self-employed persons, that is, persons who work in their own firms, including freelancers and independent professionals, and persons who participate in firms as business partners. The data covers 46 per cent of all self-employed in Finland over the study period.

Originally the dataset was collected for the advisory group whose task was to promote a reform of the pensions system of the self-employed. Negotiations have been going on for years between employer and employee organizations, the Ministry of Social Affairs and Health, and some other interest groups to reform the system. Thus far, the collection of analyzes and policy options have been published in a Ministry report (STM 2019).

The ongoing policy aim is to ease the financial burden of the system and improve earnings-related pension security. The central problem in the pension system for the self-employed is the subjective setting of the pension contribution (within certain limits and in co-operation with the pension provider). Because of the subjective choice of pension contribution, there is a persistent problem with contributions that are too low which, in turn, yield inadequate earnings-related pension security.

The sub-study focuses on an alternative policy in which the pension contribution would be based on the taxable earnings rather than on the confirmed income. The analysis proceeds as follows. First, trajectory analysis is used to find sub-groups of self-employed, using confirmed income as the outcome or analysis variable. It appears that there are six distinctive groups with different levels of confirmed incomes. Within these groups, comparisons between the confirmed income and taxable earnings are presented. Our results confirm earlier findings (e.g., Knuuti and Palomurto 2015, 120–123) that the majority, or 84 per cent of the self-employed, select a confirmed income that is too low (i.e., a case of underinsurance). There is a group of self-employed whose confirmed income and taxable income are about equal. This group (12 per cent of all self-employed) is the insurance-in-place group. Earlier administrative studies have shown that there is a small portion of the self-employed who place their confirmed income at the maximum level allowed (a de facto contribution ceiling). We discovered this group in our analysis, as well. In fact, there are three small groups (accounting for 4 per cent of the self-employed) whose confirmed income is higher than their taxable income. This makes them the overinsurance group. The terms underinsurance, insurance-in-place and overinsurance are not yet established terms and should be understood here as classifiers, lacking moral connotations.

Further analysis showed the composition of the groups in more detail. One auxiliary result is that the groups are highly polarized by profession. Another in many ways illuminating result concerns pensions. As a share (8%) of the study population retired during the study period, the pension levels confirm what the earnings trajectories would predict.

The results are interesting in many ways. First of all, the switch from confirmed income – a more-or-less subjective setting of the pension contribution – to tax registers and true income would allow for a solid connection between earnings and contributions and earnings-related pension benefits, a desired property of a statutory pension system. Second, a subjective setting of the confirmed income (i.e., the pension contribution) would require a maximum contribution ceiling. A small share of the self-employed systematically pay the maximum contribution. In a statutory pay-as-you-go scheme, the ceiling is a necessity, and our results emphasize it. Third, as underinsurance and overinsurance are common phenomena, there are probably both moral hazard and adverse selection issues (see Myles 1995, 487–484) involved in the incentives of the self-employed which would need further economic investigation. Fourth, we must not forget the biggest problem, which is unequivocally the low contributions as the majority of the self-employed select to pay the lowest possible pension contribution (based on the lowest possible confirmed income). In the public debate, many self-employed willingly select to pay the lowest possible earnings-related pension contributions and rely on national pension for old age security. This provokes the suspicion of free-riding (Atkinson and Stiglitz, 1980, 513–518). A further economic debate on asymmetric information is beyond this thesis, but our study offers some interesting observations and call for further study.

2.4 Distributional effects of the forthcoming Finnish pension reform – A dynamic microsimulation approach

This study introduces the ELSI model to the international microsimulation community. In addition, the article studies the impact of the rules of the 2017 pension reform in relation to those of the 2005 reform on future pensions and pension distribution. The aim of the pension reform 2017 was to extend working life and ease the financial burden of the earnings-related pension system. The finances of the pension system cannot be analyzed with the ELSI model, but the impact of the rules on individuals can be studied in detail.

This sub-study shows the results of the ‘what if’ question of change from the 2005 rules (baseline) to the 2017 rules (reform). From several possible microsimulation outcomes, the results of changes to the rules concerning working life, pension level, replacement rate and Gini-coefficient are presented in this article. In addition, some of the results are shown by education and gender. This article also shows the impact of income taxation on the pensions of the retired. The article also presents a novel use of data by constructing a new lifespan measure: the partitioning of the life course. A further example of the new measure is given in Chapter 5.4, and a discussion on the concepts of employment in Chapter 5.2 of this thesis. The ELSI model generates numerous results which were beyond the scope of the sub-study, but which can be reported in later studies.

The key result is that the recent pension reform leads to longer working lives. The working lives of cohorts born after 1970 have been extended by 1 to2 years on average. The effective retirement age has been postponed even more, resulting in shorter times in retirement. Future pensions will be higher and, for the first time, it has been shown that the increase will be larger at the lower end of the income distribution.

Although not explicitly shown in the article, the simulation results emphasize a common feature of pension reforms: they are not retroactive. However, through microsimulation it is possible to analyze the existing pension stock and new retirees separately. The results indicate that the gaps in pensions will increase, as old pension rules are applied to the large pension stock and new rules only to new retirees. The development of the total pension stock is primarily ruled by indexation, and new pensions are governed by the new rules, as well.

Microsimulation provides new possibilities for research. The output data generated by the microsimulation model is actually a panel of the Finnish population over the simulation period (2012–2085), which allows for several possible study designs with the individual level outcome data. Empirical data-driven research must focus on a population, and a cohort follow-up is the natural design with dynamic microsimulation output data.

The sub-study was published in 2015, when the ELSI model simulated a 30 per cent random sample of the population. Since then, ELSI has been updated and the current model population covers the total population, allowing for even more reliable and stable results and representative study-designs.

In pension evaluation, microsimulation and macro modeling complement each other. For example, the recent long-term projections of the Finnish Centre for Pensions (Tikanmäki et al. 2019) is an example of a coherent work covering both benefits and finances of the Finnish statutory (earnings-related and national) pension systems. Microsimulation is all about heterogeneity – maintaining, creating and controlling it. The results show the dynamics of the pension distribution (among the 75-year-olds) in the sample, which illustrates the heterogeneity in the model-generated population. In real terms, with the total population, the distribution measures are definitely smoother. In fact, this was the main reason for updating the model to handle the total population (high-frequency data).

2.5 Using trajectory analysis to test and illustrate microsimulation outcomes

This sub-study shows how to combine a statistical technique and model-generated data in a way that helps test simulation outcomes and illustrate results in unconventional ways. A natural study-design with trajectory analysis is a cohort follow-up, which is common output in dynamic microsimulation models. Many practical study-designs include a single cohort, but it could also include several cohorts.

In this sub-study we propose a classification rule consisting of a group-based trajectory approach as a supplement to the ex ante classification rules. These two approaches could complement each other. The ex ante rules mean that the study results are presented by a priori known factors such as gender, education, or age. With trajectory analysis, the classification can be done in two ways. First, the data can be split by various background factors (e.g., by gender). In that case, the developmental trajectories of the outcome (e.g., wages) within these ex ante groups are modelled. This would yield information on how common the earnings trajectories (by gender) are in a population. Second, a new way would be to classify the trajectory groups by some ex ante factor afterwards. This would yield information on how the ex ante classifier (e.g. gender) is divided into trajectory groups. Checking the composition of the sub-groups is also a good way to validate trajectory results. Oftentimes trajectory-based rules may shed new light on a study topic when used alongside ex ante rules.

This sub-study introduces the basic steps to trajectory analysis with examples from life course analysis. The examples cover dynamics in wages, education and pension level, each of which are typical outcomes of dynamic microsimulation. Further, each outcome is transformed to give examples of three types of data: binary variable, class variable and continuous variable. An analysis of the number of sub-groups is given as statistical information criteria (BIC) and solution figures themselves. The figures illuminate the significance of the number of groups in trajectory analysis and give insight into possible artefact groups.

Trajectory mean curves can be illustrated in two ways. First, the group-specific regression model means (curve fit) can be drawn. Second, the group-specific yearly averages of outcome can be drawn. In the article we select the last approach because the group averages are visually more sensitive to heteroskedasticity of the outcome, which is valuable information in detecting possible oddities such as microsimulation misspecification.

This sub-study is indeed informative in that it gives an example of how to test microsimulation outcomes. In this case it is not a question of pure statistical testing but merely a detailed explorative analysis of the outcome that may reveal possible oddities in the distribution of interest. The trajectory analysis did reveal a slight misspecification of the ELSI model that was not detectible with other techniques used.

With the information and introduction given in the sub-study, microsimulation practitioners should be equipped with tools to experiment or apply trajectory analysis to their research theme at hand. The study also includes an example code for SAS.

3 Statutory pensions as social insurance

3.1 Social insurance as part of social security

According to Castles (2004), a welfare state or, specifically, social expenditure can be classified into three categories: social insurance, means-tested social security and social services. The share of social insurance varies in different countries. In the Finnish case, the share is about 50 per cent. (See Hagfors et al 2004 for more information on the scale of a welfare state and the composition of social expenditure).

The purpose of social security is to guarantee individuals an income during uncertain life situations. There are different approaches to social security and its scope. One useful social security classification goes as follows. Broadly speaking, there are three layers or pillars of social security. The first layer is the compulsory safety net, which provides minimum security for all citizens. The provider is usually a public sector institution, the State or a municipality. The second layer is usually compulsory and provides earnings-related benefits. The provider is usually a non-profit provider. The third layer provides additional security for old age and for various other contingent mishaps. This layer is oftentimes voluntary and combines pooled risk insurance or actuarially based savings arrangements. The provider is usually a private sector insurance company (e.g., Tuomala 2004, 276–278). In a Nordic country like Finland, social security is largely considered the responsibility of the State and the municipalities, but there is a market for individual insurance, as well, for example in the form of life-insurance and non-life insurance.

Public social security is also guarded by law. For example, in the Finnish Constitution, the individual is entitled to necessary social security (Constitution of Finland 731/1999, Chapter 19). The Constitution itself promises insurance for many risks (Forss et al. 2004, 334). There are numerous other laws, as well, that determine the total range of social security. For example, the laws regarding compulsory pensions are numerous. The basic pensions or the pension safety net is the responsibility of Kela and it is guided by the National Pensions Act (568/2007) and the Act on Guarantee Pensions (703/2010). Earnings-related pensions are the responsibility of some 30 pension providers (see Figure 1) which are regulated by several Employees Pensions Acts (see, e.g., Työeläkelainsäädäntö 2017; saadospalvelu.fi). Especially regarding earnings-related pensions, the laws are only the tip of the iceberg regarding practical insurance arrangements in various individual life situations. There are a range of other practices which seek to guarantee equal and fair individual rights (see telp.fi). The strong insurance implicated by law and practices has led to the conclusions that individuals have legal ownership of pensions (e.g., Arajärvi 2011).

The western history of universal social insurance dates back to the 1880s in Germany, where sickness, accident, old-age and disability insurance arrangements were available in certain occupations. Heavy labor was a norm in those days, underlining the need for insurance. From Germany, the idea of social insurance spread across the world and was transformed to specific situations in the different countries. Another major event was the idea and, eventually, the actual program to provide a universal minimum benefit for all citizens, which was the mission of Lord Beveridge (1942) in Britain. The Beveridge report suggested that social insurance, with its services, is the central tool to alleviate or control certain social risks. These two programs included the founding ideas of social security and social insurance, which were adopted in Finland, as well. The Finnish system is strongly earnings-related and thus closer to the Bismarchian tradition. From a historical perspective, social insurance has grown alongside changes in employment contracts. Social insurance with protection against core risks lowered the step to the labor markets for a greater shares of people. (See Hinrichs, 2000 for additional discussions and reflections of Bismarck and Beweridge on social policy and contemporary pensions).

In Finland, there were mainly poverty relief programs in the late 1800s. Some forms of accident and pension insurance, mainly for a limited group of occupations (i.e., civil servants), were available in the early 1900s. Universal social insurance was strongly advanced in the 1930s when early forms of national pensions were adopted and Kela was established (in 1937). In 1956 the national pension was reformed, and flat-rate pension became universal benefit and a safety net for pensioners. Later in the 1960's the foundations of earnings-related pensions for the employees were formed. Since then social insurance legislation, especially regarding earnings-related pensions, has been updated and reformed greatly. There are excellent introductions of history of Finnish pension systems (Niemelä

2011; Hannikainen and Vauhkonen 2012a; b) and history of insurance (Rantala and Pentikäinen 2009, 18–49), for example.

In practice the income guarantee by the state is tied to a situation in life when individual is unable to gain incomes via employment. Risk of unemployment, accident at work, long term sickness, disability or old age for example faces every working age individual. Social security is all about sharing these risks with a large pool of individuals. Some of the risks, such as old age or parenthood, are predictable while others, such as disability or unemployment, are unpredictable. Social security can be arranged by private or public institutions or both. Public provision of social security has proven to be the long-term solution for most developed countries, because insurance markets in practice fail to provide all necessary insurance products or the price is too high.

Before discussing the concept of risk in insurance science and economics, the characteristics of risk in sociology must be mentioned. In short, in sociology the discussion of risks relates to negative situations in individual life in changing society. Change – slow or fast – happens in many areas of life including labor markets. In sociological discourse change relates to the megatrend in transition to a post-industrial society. Peter Taylor-Gooby defines two classes of risks. Old social risks are the same as mentioned earlier: old age, disability and widowhood. New social risks emerged after industrialization. They include implications on sub-groups of population and reflections of different stages of the life course. Examples of new risks are low skilled workers, single parents and immigrants. (see Taylor-Gooby 2005). Beside old risks, the new risks are essential part of pension evaluation like in the sub-studies of this thesis.

Even broader views on risks and trust in modern society are discussed (e.g., Giddens 1990, 29–36; Beck 1992; Beck 2000, 67–88). These broad views are important in holistic understanding the discourse of risk and are in practice embedded in many pension studies. However, further discussion on the concept of risk in sociology in more detail is beyond this thesis.

3.2 Social insurance in economics

There is a body of international economic literature of the role of social insurance in an economy (e.g., Diamond and Mirrlees 1978; Atkinson and

Stiglitz 1980; Barr 1994; Tuomala 2016; Sen 2017). There are also some introductions on the economics rationale of social insurance in the Finnish context to (e.g. Kiander and Lönnqvist 2002; Tuomala 2009). In literature, the discussions on insurance center around the question of who should provide insurance in general? Can private markets provide acceptable insurance solutions for the majority of the population and for risks that are generally considered social risks?

One view is that public intervention is unnecessary and that well-informed and rational individuals in perfect market conditions should be able to decide for themselves when to save or borrow and when to spend savings as consumption, or when to transfer bequests. There is a long tradition of life cycle and overlapping generations modeling, originating from seminal papers of Samuelson and Diamond (Samuelson 1958; Diamond 1965, also Ando and Modigliani 1963), where the ideal individual resource allocation over the life course can be analyzed within and between cohorts. A conscious introduction to overlapping modeling traditions and generational accounting can be found in Blake (2006). In Finland, there are examples of evaluating pension systems and reform options with overlapping generation models (e.g., Broer and Lassila 1997; Lassila and Valkonen 2000; 2019; Määttänen and Valkonen 2008; Lassila et al. 2014).

However, today there are arguments against publicly provided insurance in economic literature because of market failures. Private social insurance, or even pension insurance, may not be profitable on a large scale under market failures and incomplete information. The key reason why private insurance is difficult or impossible in general is that there is not enough information available to tailor insurance for all individual situations, i.e., risks. Another way of saying the same thing is that there may not be markets for all risks, which implies selection bias regarding possible clients. This is not to say that private insurance markets fail with all risks. There is certainly a market for private insurance and private saving products, as well, as complements to social insurance. A bleak practical issue with private social insurance is the real possibility of bankruptcy of a benefit provider.

The rationale for public statutory social insurance is anchored around the issue of asymmetric information and much studied topics of adverse selection and moral hazard. These information-failure-related topics, which are valid today, have been discussed among economists since Akerlof (1970) and Pauly (1974). There is a body of economic literature on these topics, which are beyond this thesis, but basic facts of information failure must be repeated here.

Adverse selection refers to a situation in which individuals have the best knowledge of their physical condition or driving habits, for example, but insurance providers cannot distinguish individuals with different risks, and thus place the price or premium for the insurance improperly. Insurance market equilibrium will not be stable. Adverse selection causes two sets of problems: a pooling equilibrium is unstable, either because low-risk individuals drop out or because competitive behavior of insurance providers, and a separating equilibrium, if it exists at all, is inefficient because it prevents low-risk individuals from buying complete cover. In the face of adverse selection, the market is either inefficient (too little supply) or fails entirely, the ultimate outcome depending on the precise behavior of the insurance provider or the insured (Barr 1994, 120; Hellwig 1987).

Moral hazard refers to a range of situations in which an individual can affect the probability of some risks and losses of the insurance provider. In fact, moral hazard causes a fundamental problem: the more complete the cover and the lower the psychological loss from the insured event, the less individuals have to carry the consequences of their actions and the less, therefore, the incentives to behave as they would if they had to bear for their losses themselves. (Barr 1994, 122.)

Because of asymmetric information, public provision of insurance products is therefore feasible or insurance providers in the markets must receive some kind of subsidy or at least guarantee from the state for arranging the required insurance to large shares of the population. Public provision does not solve the asymmetric information problem completely, but it can assure that there is insurance available for everyone (see Tuomala 2004, 273–275). For example, regarding the Finnish statutory pension system for the self-employed, there is some evidence that suggests that some professionals pay maximum premiums. This a profitable strategy if the life expectancy of these individuals is high. This might be problematic in a defined-benefit pay-as-you-go pension system. The sub-study II of this thesis illuminates the insurance behavior of these professions.

In addition to asymmetric information, there is also a totally different discussion regarding how well individuals can and will plan for their future. Individuals in real life might not plan for their distant future (via savings) but focus merely on and adjust their behavior based on their current situation in life. There is an evidence-based framework that considers myopic behavior of individuals. (e.g., Feldstein 1987; Myles 1995, 468–470). Another line of thought and empirical practice considers partial myopic behavior or time-inconsistent discounting (e.g., Rubinstein 2003). Benartzi and Thaler (2007) bring the lessons from behavioral economics to the center of economic debate when they discuss heuristic savings and retirement decisions.

There is even a view according to which public goods, welfare services and social transfers are essentially social insurance because, as incomes increase, the use of public goods increases significantly less than the tax revenues by which the goods are financed. This is a question of how the most part of public social insurance consists of redistribution of taxation (Tuomala 2004, 274). This view can also be expanded to a life cycle context: over the life course, all redistributive mechanisms could be considered insurance.

A contribution to the economic debate on pension evaluation is the discussion of taxation, namely its incentives on businesses' retirement behavior, for example (see Uusitalo and Nivalainen 2013; Benzarti et al., 2019 for a Finnish case).

One less-discussed issue is that social insurance can provide an individual with more possibilities in life through increased risk taking. Unemployment insurance or pension insurance provides possibilities to change employer or spend time studying, i.e., investing in human capital. This is closely related to the view that insurance and redistribution by a welfare state in general can and probably will lead to more efficient labor markets.

Today more than for a long time, administrative costs are at the center of public discussion. There is a large consensus that public social security or even the welfare state is more efficient than what private markets could provide via several insurance providers. One large, compulsory, public sector institution with no consumer choice has scale-advantage over several competing providers with some market power in the price setting of insurance premiums. (Tuomala 2004, 278.)

Another view is that public social insurance can affect individual incentives directly. There seem to be some examples of this in the Finnish pension system with long-standing mechanisms of tradeoff between better benefits and contributions, with the aim to extend working lives. In a private pension system, this would probably not be possible as the public sector could only regulate pension insurance companies.

There is a long-standing debate of political cycles on, for example, pension rules or investment policies of pension funds in the public social security system. Kela is supervised by Parliament. The earnings-related pension system is governed by labor parties (employers and labor unions) and the Government. Political risks can be a reality in the public pension system. For example, Diamond (1997) discusses the role of the Constitution with respect to other laws and political culture (visibility of political decision-making). It can be speculated that private pension insurance could avoid political risks. This is, however probable, not true, as private pension insurance usually relies on public sector guarantees and tax deductions. In Finnish voluntary life-insurance markets, there is one sad example (PS-tili) of how changes to eligibility rules and tax-deduction rules can negatively affect an emerging insurance product and markets. (See Harju 2012 for a more detailed discussion on the tax reform on voluntary pension plans.)

Economists have presented broad ideas on how to turn social security on a larger scale into insurance, that is, consider all social risks within social insurance in a way that could benefit citizens and the economy as a whole (e.g., Forss et al. 2001; 2004). These arguments and range of risks are close to sociological thinking.

Economists do discuss insurance in practical ways, as well, and use the same language as actuaries. One simple way of defining insurance is as follows: “Insurance can be considered a device which offers individuals protection against risks and/or as an actuarial mechanism” (Barr 1994, 111). This is a great simplification, but further discourse on the fusion of economics and insurance science dates back to Karl Borch (e.g., Borch 1974; 1990).

3.3 Social insurance in insurance science – the concepts and practice

The world is a risky place. There is a range of risks which can broadly be classified as individual risks, property risks, and business risks. All of these classes contain dozens of risky events, and they can affect both individuals and businesses. Individual risks are the center of focus in social insurance, so we concentrate on them and, especially, compulsory insurance. Some examples of risky events are traffic accidents (property risk), accidents at work, disability, unemployment, old age and death (all personal risks). There are other classes of risk events, but the aforementioned relate to pension insurance and are therefore important in this context. Social insurance is also about risk management on a grander scale (see Liukko 2005; 2013). But what is a risk?

Risk is the chance that some uncertain and unwanted event happens. The concept of risk has certain properties. First, risk is arbitrary, like accidents at work or unemployment. For example, old age (e.g., 63 years) is not arbitrary, yet it is uncertain for an individual to reach that age. Second, the law of large numbers gives the assurance that, although an individual faces uncertainty, a large enough pool of individuals face approximate certainty. For example, it is uncertain if a male born in 1970 will die in 2020, yet the death rate of males at age 50 is known. Third, risk can be measured, in a monetary sense, which guarantees the price setting (see Barr 1994, 114; Rantala and Pentikäinen 2009, 53–59).

There is uncertainty in life, and everybody, even the luckiest ones, face at least one or two of the risks discussed here: old age and death. The length of a lifetime, and especially time in retirement, includes the risk on livelihood, but the same applies to other unpaid periods over the early life course, like parenthood. The modern answer to alleviate the negative unwanted event is insurance. Insurance is, by definition, an arrangement where a pool of individuals make an agreement about compensation for some risky event with the insurance provider. Further, the agreement presupposes that individuals pay an insurance contribution to the provider. In practice, insurance agreements are far more complicated.

Insurance as viable business requires some basic boundary conditions. First, as insurance is targeted to a certain risky event, risk must be arbitrary, i.e., it must not be seen ex ante that the event will occur. Second, there must be a possibility for economic loss for the individual if the risky event happens. Third, in a financially stable insurance agreement, the insurance contribution and the corresponding amount of risk are equal (the insurance provider can charge an additional administrative fee). People with different risk-profiles (e.g., age and gender) can have different contributions. For example, in life-insurance, the contribution of older individuals is higher than that of younger individuals because the risk of death is higher for the older individual. Fourth, the law of large numbers indicates that the risk pool of individuals should be large. Fifth, the insurance provider should be legally and economically independent from the insured individuals. Statutory pension insurance is an example of a situation where the employer takes out insurance for its employees, and the employees get the benefit if the risky event occurs. (Rantala and Pentikäinen 2009, 62–67.)

Finally, there are some practical conditions which define what risks can be insured technically. These conditions are the foundation for setting the price for the insurance, managing the risk and finding customers for the insurance product. First, the risk must be predictable with some accuracy. Second, the risk must be independent from the individual. In economics, this is closely related to previously discussed problems of asymmetric information. Third, the risk must be stable in time. Fourth, the risk must be rare. (Rantala and Pentikäinen 2009, 67–69.)

The general classification of risks presented above is relatively broad. One way to be more specific is to consider individual risks, as they constitute the concept of social risks. Social risks are those which are somehow managed by statutory social security systems, whether they are privately or publicly provided. What risks are expected to be covered in a statutory social security system? According to ILO, there are nine specific social risks:

- Sickness
- Accident at work
- Disability
- Invalidity
- Unemployment
- Motherhood
- Parenthood
- Old age
- Widowhood

These risks are monetary in nature. By and large they can be transformed into benefits or compensation to the individual (or family) which faces them. There are other important classes of social security, namely services and social assistance, which are not discussed here. In a European context, Finnish or Nordic social security centers around the concept of social protection, which includes monetary compensation and services.

Social insurance means, by definition, publicly organized statutory management of social risks. Benefits or entitlements are paid to an individual when earning a wage is impossible for some reason. The focus of the benefits centers around an individual facing a risk but, in some cases, the benefits (i.e., survivors' pensions or family benefits) cover family members, as well.

The institutional arrangement of Finnish social insurance centers around social risks. There are specific institutions to cover different risks, namely: health insurance, worker's compensation insurance, pension insurance, employees' group life insurance and unemployment insurance. In monetary terms, pension insurance is by far the largest part of social insurance. (For a comprehensive introduction on institutional arrangements of earnings-related pension systems, see Rissanen et al. 2017).

There are markets in place for, for example, private life-insurance and private saving for old age. Yet public social insurance in Finland has such wide coverage that private insurance is not so important for individual income. Of course, there are other classes of non-life insurance, such as housing insurance or traffic insurance, which are invaluable but beyond the scope of this thesis.

Oftentimes in international comparisons, pension systems are classified in the context of pillars (e.g., OECD 2005). According to this classification, the first pillar consists of statutory pensions provided by the government itself or its institutions. First pillar pensions can include means testing, as well. The second pillar includes pensions which are the responsibility of employers. These pensions are occupation-specific or otherwise dependent on the agreement between employers and employees. Second pillar pension agreements can be either statutory or voluntary, but they cannot include earnings or means testing. The third pillar consists of voluntary pension arrangements by the individual or the employee. The Finnish statutory pension system, including national pensions and earnings-related pensions, is classified as a first pillar system.

Earnings-related pensions

The aim of the earnings-related pension scheme is to guarantee a reasonable consumption level in retirement. The pension accrual is based on earnings (employment) and working life (employed time). For this dualism of earnings and working life together, individual pension accrues. It should be enough to provide for livelihood in retirement.

In Finland, the first elementary pension arrangements were a reality already in the late 1800s. During the emerging years of the current earnings-related pensions, the insurance covered workers and specific occupations only. The first modern earnings-related pension system was for mariners. The Seafarer's Pensions Act (MEL) took effect in 1956. In 1962, the Employees' Pensions Act (TEL) and (for short-term employment contracts) the Temporary Employees' Pensions Act (LEL) took effect. In 1964, the Local Government Employees' Pensions Act (KVTEL, since 2003 KuEL) was introduced. In 1966, The Evangelical-Lutheran Church Pensions Act (KiEL) took effect. In 1967, the State Employees' Pensions Act (VEL, since 2007 VaEL) took effect. In 1986, the Freelance Employees' Pensions Act (TaEL) took effect. TaEL was originally the pension act for certain artists and reporters in employment. The Self-employed Persons' Pension Act (YEL) and the Farmers' Pensions Act (MyEL) came into the pension system in 1970. In recent years, the pension acts have merged further. In the private sector, the Employees Pensions Act (TyEL) took effect in 2007, merging TEL, LEL and TaEL into one single act. In 2017, the rules of the public sector pension acts (KuEL, VaEL, KiEL and KelaL, the regulations concerning the employees of Kela) merge into one act, the Public Sector Pensions Act (JuEL). The employees of the Orthodox Church, the Bank of Finland and the regional government of Åland have their own pension regulations. The abovementioned acts are collected in Työeläkelainsäädäntö 2017. A comprehensive introduction of the political history of Finnish private sector earnings-related pensions is given in Hannikainen and Vauhkonen (2012a; b). A comprehensive collection of historical background information of central topics in earnings-related pensions is given in Ahtokari (1988), Salminen (1993), Pentikäinen (1997) and Pietiläinen (2005).

The coverage of the earnings-related pension system is wide in terms of risks covered. It is also fair to evaluate that the Finnish statutory pension system is a universal system with nearly 100 per cent coverage of the employed. Figure 1 shows the major stocks and flows of the Finnish statutory pension system. Currently, the system provides the main income for most retirees (1.49 million pensioners).

The scope of the statutory earnings-related pension system is narrower than all social insurance. Social insurance should cover the nine risks listed in the previous chapter. The aim of the pension system is to cover risks such as permanent disability, old age and widowhood (universal survivors' benefits since 1990). Currently the following benefits are available for new retirees: old-age pension, disability pension, partial old-age pension, years-of-service pension, occupational rehabilitation benefits and survivors' pensions.

However, looking at the range of benefits in the Finnish pensions system since the 2005 pension reform, the scope of risks is wider. Especially in the times of early retirement benefits, the tasks of pension systems expanded from the three core risks. Today, the pension accrual from various social security benefits (albeit small) is another example of the wide scope of the pension system.

There are some fundamental principles in the earnings-related pension system which have been at the center of focus when reforming pensions. First, there is a vesting principle applied to pension accrual, which means that the pension is owned solely by the employee. The accrual is not affected by, for example, a change of employer. Second, earnings-related pensions are based on earnings, without a ceiling. This is true for all employees, but there are some rules in the pension schemes of the self-employed which postulate an earnings ceiling on their earnings. For workers, there are rules which determine the income classes for which pension accrues. For example, options or shares are not pensionable earnings. Third, the value of pension accrual and the actual pension, as well, are protected by indexation (mixture of wages and prices). Pensions are annually adjusted at least for inflation. Fourth, the pension expenditure is shared by the employer and the employee. The employer and employee contributions are used to finance the expenditure of the earnings-related pension system, though a share of the profits of the pension funds is also used to finance current pensions in payment.

In addition to the abovementioned principles, which have direct consequences on employees and retirees, there are some general system-level principles which need to be clarified here. First, the earnings-related pension system is simply a defined benefit scheme. Second, the finances of earnings-related pensions are based on partial funding. The Finnish system collects a reserve fund to ease future increases in the pension insurance contribution. In 2017, the pension contribution is 29.0 per cent of the wage sum. Around 30 per cent of the contribution is funded and 70 per cent is used to finance current pension expenditure. There are some exceptions in the funding within different schemes. While private and public sector employee pension schemes collect reserve funds (83% of GDP), the schemes for the self-employed (Self-employed Persons' Pension Act, YEL), farmers (Farmers' Pensions Act, MyEL) and sailors (Seafarer's Pensions Act, MEL), are financed by pension contributions and the state budget.

There is a total of 30 pension providers that manage earnings-related pension insurance. The majority of employers are insured in pension insurance companies, yet there are some small occupation-specific providers, as well. A unique feature of the Finnish system is that private sector insurance companies manage social insurance. The management of pension funds is regulated so that the companies cannot share profits externally. It is agreed with the European Union that the Finnish earnings-related pension system is classified as a first pillar system, and the pension funds are part of the public sector in national accounting – a significant issue in terms of EU-wide public sector sustainability calculations. In the context of statutory pension insurance, the providers take a different approach to insurance compared to business-driven life-insurance. Pension providers apply standard actuarial mechanisms to manage risk pools. There are a number of Acts and other general practices and codes which must be taken into account when managing the stock of insured and the funds.

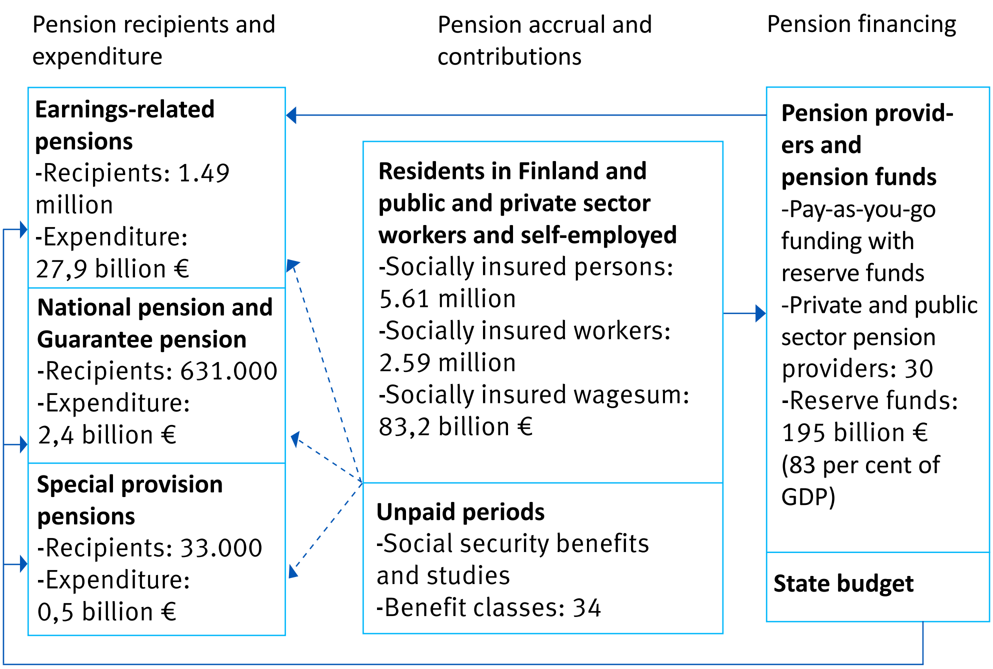
National pension

The aim of the national pension and the guarantee pension is to guarantee a minimum income in retirement. The earnings-related pension scheme and the national pension scheme affect each other so that, when the earnings-related pension exceeds a certain amount, the national pension is reduced. There are about 631,000 individuals receiving a national or a guarantee pension.

As the coverage of the earnings-related pension scheme is restricted to employees and self-employed persons, large shares of citizens are excluded from insurance coverage. The national and guarantee pensions are designed to cover citizen with little or no working life. The obvious difference to the earnings-related system is the amount of pension, which is fixed to a certain degree. In the Finnish pension system, there are currently two classes of national pension: the national pension and the guarantee pension, both of which are noncontributory universal residence-based pensions. This high level of coverage could not be possible with any other institution than the statutory pension system.

Not every resident is entitled to a national pension or a guarantee pension. As the earnings-related pension system has matured (since 1998, see Helsingin Sanomat) to a point where more and more individuals draw a full pension based on a whole working life, the pension level has reached its designed level. There is a technique applied where the national pension is fully deductible from the earnings-related pension.

Figure 1.
The scope of Finnish statutory pension system in 2018



*Dotted line indicates pension accrual, and solid line contribution and expenditure flows.

Source: Tela, Finnish Centre for Pensions research register, own calculations.

Special provision pensions

Special provision pensions (SOLITA-pensions) are accident-related benefits, but they are part of the statutory pension system. They are governed by the Occupational Accidents, Injuries and Diseases Act, the Motor Liability Insurance Act, the Act on Compensation for Military Accidents and Service Related Illnesses, and the Act on Compensation for

Accidents, Service Related Illnesses in Crisis Management Duties and the Military Injuries Act.

Supplementary pensions

Private or employer-specific pension insurance plays a small role in the Finnish pension system or in the livelihood of Finnish pensioners. There are, of course, individual exceptions with both of these benefit classes. First, the employer-specific TEL supplementary pension played a role in statutory pension security until the year 2000, when the system was closed from new employees. At 2016 the arrangement was closed for the already insured, and entitlements were capitalized. The supplementary pension is paid currently and also into the near future. Typically, supplementary pensions are time-limited with the idea that benefits are paid until a certain maximum age (e.g., until the age of 65), which means they are not intended for permanent old-age security. Second, private pensions play a minor role in Finnish pension security. In their early history – from 1910 to 1930 – private pensions were popular arrangements, but during the improvement of the national and earnings-related pension systems, the role of private pensions gradually diminished.

There are a range of voluntary private pension insurance and long-term savings products available on the markets with the intention to postpone consumption to old age (e.g., Valtiovarainministeriö 2018), but on a large scale, their role has remained small compared to compulsory pensions. Voluntary pension insurance and saving is beyond this thesis, although they can be important in individual life situations.

4 Previous evaluation of statutory pensions

The history of statutory pensions in Finland is also a story of pension reforms. The evaluation of earnings-related pensions since the 1960s has been orderly. The basic principles of pensions have remained the same, yet the system has evolved greatly in terms of coverage and pension benefits. Each reform has been carefully designed and evaluated using the best knowledge of that time. The evaluation techniques and methods have naturally developed over time. In brief, major reforms took place in the 1980s and 1990s, when new early retirement benefits were introduced and pension financing was reformed. I will focus on the evaluation of the 2005 and 2017 reforms which have been studied from many perspectives (e.g., Uusitalo and Nivalainen 2013; Tenhunen 2017; Tenhunen et al. 2017).

Reforming pensions is not always straightforward process. For example, survivors' pensions (see Takala ed. 2013; Takala et al. 2015) are benefits that would need more than a fine-tuning of the details under current law. In addition, there is currently an evaluation under way to reform the Self-employed Persons' Pension Act (YEL) (see STM 2019).

Long-term research on pensions is part of the evaluation process. For example, the research at the Finnish Centre for Pensions has broadened in depth and content over the last two decades. Better in-house and external data resources has fostered improved, peer-reviewed research. Of course, there are also other research institutes and universities which evaluate pension policy, but the role of the Finnish Centre for Pensions is broad. International research activities on pensions is extensive. Examples of pension analysis from a comparative perspective are Ebbinghaus (2006) and Barr and Diamond (2008). The embracing Pensions at a Glance (OECD 2019b) gives examples of several pension indicators with special emphasis on old-age poverty and inequality. A thorough discussion on international research traditions is beyond this thesis as its focus is on Finnish realities.

There are some recurring topics on the research agenda of the Finnish Centre for Pensions. For example, the cause of disability and occupational rehabilitation are constantly studied (Rantala and Gould 2015; Polvinen 2016; Polvinen et al. 2018; Laaksonen et al. 2016a and b). Another established research topic centers around pension benefits (Salonen ed. 2015; Tenhunen and Salonen 2016; Tenhunen et al. 2018). Over the last decade working life in different population groups (e.g., Järnefelt et al. 2013; 2014; Laaksonen et al. 2016a) and statistics (e.g., Pension indicators 2019) have become constant area of study. Another long-term area of study centers around the income and income distribution of different population groups, especially in the form of comparisons between pensioners and other groups and also from the point of view of poverty (Rantala and Suoniemi 2011; Rantala 2014; Kuivalainen et al. 2017). There is also a growing amount of research using survey data and analysis. Research topics include (institutional) trust (Takala 2015), employer views on pension incentives (Liukko et al. 2017) and insurance technique in terms of disability pension (Kyyrä and Paukkeri 2015). There is also a research tradition of making comparative analysis of pensions and pension systems across different countries. (Ahonen 2018; Palomäki 2018). Another type of evaluation are the cohort or population level macro projections. Since the 1970s, the Finnish Centre for Pensions has published long-term projections (PTS) on pension benefits and the financing of pensions. Financing pensions is an emerging area of study on the research agenda (Sankala et al. 2018; Kautto ed. 2019). Another area of research is generational accounting, which is an area of study, where the focus is on inter-generational transfers (e.g.,Vaittinen and Vanne 2011; 2017). Closely related to the generational aspects are studies on the profitability of pension contributions in pay-as-you-go pension schemes (e.g., Korkman et al. 2007; Gröhn 2008; Risku 2015).

Basically, research on and the evaluation of pensions centers around the concept of risk (cf. Chapter 3.3). In practice, risk is sometimes understood broadly as an event in the life course which can be positive (promotion in career, having a child) or negative (being disabled, death). An analysis of the outcomes of a pension system is an analysis of the risk (negative life event), but there is more room for negative connotations, as well. Another point worth noting is that pension research aims at confirmatory analysis. Yet, there is a solid tradition to keep analysis at an exploratory level.

The entire life course is increasingly under investigation. New studies focusing on young adults are emerging. Studies on the labor market attachment of post basic-graduates (Saloniemi et al. 2013; Salonen et al. 2014), NEET young (sub-study I, Saloniemi et al. 2020) and labor market risks of parenthood of the young (Kuitto et al. 2019a; 2019b) are examples of an indirect evaluation of pension accrual, using register material and modern statistical methods, namely trajectory analysis.

There is a wide range of study-designs and statistical methods applied in the research mentioned above. The majority of modern pension research is evaluative by nature and based on high quality up-to-date register material. Modern registers offer possibilities to analyze the effects of pension policy, or economic or demographic changes on pensioners and pensions. The evaluation of pension finances is macroeconomic by nature, yet individual-level material often provides a base for such analysis.

A majority of the pension research and evaluation is interested in ex ante classified groups. It is common to report study results by some group like gender, cohort, age group, income bracket, educational level, socioeconomic status, or sub-population (retired people). A Finite mixture analysis and other clustering methods are becoming more popular. This is for a good reason, as ex post classified groups can reveal new insight into pension-related research.

Several research methods are used in pension evaluation. Regression methods such as logistic regression is frequently used. Survival analysis is analysis with duration data (e.g., Salonen and Möttönen 2019). Survey methods with survey data. Over the last years, mixture methods have become more popular. Recent trends include sequence analysis (e.g., Riekhoff 2018; Perhoniemi et al. 2019) of late-life labor market transitions (Kurvinen et al. 2018). In a completely new type of analysis, machine learning techniques are used to predict new disability retirees (Varis 2018). Both techniques will certainly be broadly and rigorously used in pension evaluation in the near future.

5 State-of-the-art data

5.1 From data to intelligence

Raw data has little value unless analyzed and transformed to useful information. It is not produced strictly according to the research problembut is often collected for some other purpose, such as for statistics. Administrative register material is basically raw data, which must be processed in order to be useful for research. Processing register material involves summarizing data, merging datasets from different sources and counting new information. This leads to the question of secondary data. The research register of the Finnish Centre for Pensions includes pre-processed secondary data.. The pre-processing involves carefully planned and executed data manipulations. The end result is reasonably coherent and consistent research data that is flexible enough to be used with different study-designs.

Contemporary data possibilities of the statutory pension system are better than before in terms of measuring, for instance, employment, earnings, social security benefits, pensions, pension eligibility and working life. One advantage with administrative registers is that they are usually up to date compared to the research material of statistical institutions. For example, Statistics Finland offers research data that is usually a few years old. When evaluating pensions, the need for the most recent data is usually high.

In quantitative analysis, the available data defines the possible study designs. The study design is a framework of methods and techniques used to merge elements of research in a reasonably logical way in order to handle the research problem effectively. In pension evaluation, a common design is follow-up studies. Other frequent designs include cross-sectional studies and inverse follow-up case-control studies.

Study-design includes the following three sections: data collection, measurement and (statistical) analysis. In addition, it consists of four key characteristics: neutrality, reliability, validity and generalization. Neutrality means that the results implied in study-design should be free from bias. Reliability means that the study-design should indicate how the research questions were formed and how the results could be replicated. Validity means that, among many possible analysis strategies, the valid measuring tools are those which help the researcher assess the results according to the objective of research and nothing else. Generalization means that the results of the study (study-design) must be applicable to the total population and not just to a sample.

Another way of looking at the research process is to take a broader view, that is, to see the process from raw data to intellectual ideas. On a timeline, this would include the following steps: data (properties of objects and events), information (processed data such as statistics), knowledge (answers to how-to questions), understanding (answers to why questions), wisdom (values and exercise of judgement) and intelligence. These steps lead to two desired properties: efficiency and effectiveness. Information, knowledge, and understanding enable us to increase efficiency, and effectiveness is evaluated efficiency. Intelligence is the ability to increase efficiency; wisdom is the ability to increase effectiveness. (Ackoff 1999.) I would argue that these ideas are important also in pension evaluation. Good practices are needed to proceed from data to wisdom, effectively and efficiently.

The research register of the Finnish Centre for Pensions contains data and information points that are intended to be easy and flexible to use down the line of the research process.

The goal of pension evaluation is to search for appropriate answers to relevant problems. There are numerous relevant problems in the field of pensions, but focusing on the essential ones requires an understanding of the problem, the data, the analysis and the results. Pension researchers must have some latent background knowledge of data, metadata, problems and statistical methods.

The principles of measurement of secondary data can be summarized as follows: reality can be confronted by recording observations that reflect the phenomenon of interest. Information is a product of data and pre-knowledge. Infological equation $I=i(D, S, t)$. Information (I) is produced from the data (D), semantic background (S) and pre-knowledge using interpretation process (i) at a point of time/interval of time (t). (Langefors 1973.) In short, the only accepted criterion for useful information is that it is produced using the scientific method.

Pension evaluation is not based solely on administrative register material. Specific research data from Statistics Finland (e.g., Total statistics of income distribution,Official Statistics of Finland 2019a), and Eurostat are frequently used in research focusing on different income sources and taxation. Another form of data is a combination of register material and survey responses (e.g., Takala 2015; Tenhunen 2017; Liukko et al. (2017). The author has been involved in the study-design of the latest surveys of the Finnish Centre for Pensions and has collected and combined register material of the survey population.

Pension evaluation has advanced over the last 15 years with more and better register material. Especially advances in information technology have made it possible to handle large datasets and perform statistical computing.

Nonetheless, there are certain limitations relating to secondary data. First of all, the content of register material focuses on the initially recorded information. Data can be converted to information within certain limits, but individual opinions are usually not recorded into registers. This limits the range of interesting research questions. Second, relating to the first point is a question of whether registers contain all necessary information. Often there is a need to get a clear picture of an individual life course, but there is not enough clear information available (e.g., received social security benefit, income or health information). Third, missing information can rarely be easily corrected or supplemented. There is no simple trick to overcome problems arising from the fundamental limitations of register data.

An individual life course is typically divided into three consecutive life or labor market phases: childhood, and especially adolescence, provides young people with skills and prepares them for the phase of active workers on the labour market, and eventually, with age, people retire (e.g., Meyer 2009). Although a rather schematic model of a life course, which can be criticized, this is a rather useful model from the perspective of pension eligibility.

5.2 Challenges in measuring working life and life course

The Finnish Centre for Pensions is one of the key providers of employment or working life information for research purposes. As the central body of earnings-related pension system, the Finnish Centre for Pensions collects employment data from pension providers. An individual's employment contract information is stored in central registers. Information on employment and self-employment spells is collected for pension calculation purposes. The information about pension accrual (contracts of employment and social security benefits) must be available at the time of retirement and is therefore collected and stored.

Information on employment has proven to be valuable for both research and statistics. Working life statistics is based on information about employment contracts. Since the early work in Järnefelt et al. (2013), research on working life has advanced greatly. Nowadays it is part of many labor market studies conducted by the Finnish Centre for Pensions and other research institutions and universities. Since 2013,working life has been part of statistical reporting (e.g., Pension Indicators 2019). The register information about employment in, for instance, research datasets by Statistics Finland often originates from the registers of the Finnish Centre for Pensions.

Converting administrative register data on employment and earnings into research material is challenging. There are six major problems associated with employment contract information that researchers and readers of research results should be aware of (see, for example, Laesvuori and Tuominen [1996], and Kautto and Salonen [2013] for discussions on the limitations of employment information).

First, the eligibility for pension coverage in Finland has changed over the course of decades. In the early years, earnings-related pension systems were based on occupation. The first earnings-related pension system was for mariners. The Seafarer's Pensions Act (MEL) came into force in 1956. In 1962, the Employees' Pensions Act (TEL) and the Temporary Employees' Pensions Act (LEL; for short-term employment contracts)were introduced. In 1964, the Local Government Employees' Pensions Act (KVTEL/KuEL since 2003) and in 1966, the Evangelical-Lutheran Church Pensions Act (KiEL) took effect. In 1967, the State Employees' Pensions Act (VEL/VaEL since 2007) came into force. In 1986, the Freelance Employees' Pensions Act (TaEL) took effect. TaEL was originally the pension act for certain artists and reporters with an employment contract. The Self-employed Persons' Pensions Act (YEL) and the Farmers' Pensions Act (MyEL) became part of the pension system in 1970. In recent years, the pension acts have merged further. In the private sector, the Employees Pensions Act (TyEL) took effect in 2007, incorporating TEL, LEL and TaEL into one single act. In 2017, the public sector pension acts (KuEL, VaEL, KiEL, KelaL [regulations concerning the employees of Kela] merged into one act, the Public Sector Pensions Act (JuEL). The employees of the Orthodox Church, the Bank of Finland and the regional government of Åland have their own pension regulations. The Finnish earnings-related pension system covers practically all employed. About 2.38 million persons were employed and insured under earnings-related pension acts at the end of 2017.

Second, earnings limits and accrual time requirements have affected the pension accrual of workers and the registration of private sector employment. The accrual conditions have changed over time. In the early years (1962–1963), the minimum duration of employment under the Employees' Pensions Act was six months. In 1964–1971, the minimum duration was four months. In 1971–1988, the minimum duration of employment under the Employees' Pensions Act was one month. After 1997, short employment spells (less than one month) were insured under the Freelance Employees' Pensions Act. Since 1998, all short or low-wage employments have been incured. In the early decades, earnings limits under the Employees' Pensions Act restricted insurance of low-paid employment. The earnings limits have been reduced steadily. Since 2005, pension accrues for virtually all gainful employment.

Third, pension accrual age limits have changed over the years. Initially pension accrued as of age 23. The 2005 pension eform lowered the age limit to 18 years. With the 2017 pension reform, private-sector employees earn a pension already from the age of 17 (the self-employed as of age 18) until the age at which their insurance obligation ends (68 years). In essence, pension accrual does not affect the measuring of the early life course. The latter part of the life course, however, influences what is considered the final stage of working life.

Fourth, pension also accrues from certain social security benefits. Before 2005, the pension accrual based on employees' social security benefits over their whole career was aggregated into measure: one sum of days called increment for credited periods. Earnings-related pension accrued for three types of benefits: the unemployment benefit 1969–2004, vocational rehabilitation (1994–2004) and occupational training (1994–2004). As of 2005, pension accrues for several unpaid periods of earnings-related social security benefits which are payed directly to the employee (e.g., Turunen 2015). In addition, pension accrues for periods of home care of one's own children under the age of three and studies leading to a degree. The aforementioned periods were recorded into administrative registers.

Fifth, it is compulsory for employers to arrange statutory pension insurance for their workers beyond the pension accrual age. Initially, the obligation was to arrange pension insurance for employees aged 16 or more. Since 1998, the age limit was lowered to 14 years. The insurance practices changed in 2005 as the obligation to take out earnings-related insurance was increased to 18 years. In 2017, the age was lowered to 17 years. This is important as it means that working life can be calculated from ages below the worker's pension accrual age. In practice, working life and employment can be calculated reliably from the age of 18 onwards.

Finally, a projected pension component is calculated mainly for the disability pension. The objective of the projected pension component is to compensate the worker for the loss of income from the time of retirement on a disability pension to reaching the retirement age. When calculating the pension, the projected pension component is considered a contract of employment although it is not registered as such. The projected pension components are not part of employment or working life.

Overall, we can conclude that there are some reservations with respect to measureing employment and working life via administrative registers. Changes in pension rules over time distort the measurements of working life for some cohorts. On the other hand, the rules have been stable and reasonable from the point of view of research from 2005 onwards. This allows us to study employment and earnings longitudinally.

5.3 Established life course measures

There is a lively tradition of life course studies with many disciplines involved. However, this is not the object of investigation in this thesis. For pension evaluation, the relevant analysis traditions focus on active working life. The measures are relatively simple compared to empirical studies published in dedicated life course journals. The measures used in pension evaluation and policy analysis are equivalent to statistics.

Uusitalo ed. (2011) gives examples of different relevant measures in pension evaluation. The working life indicators must be policy relevant. Good indicators should react to changes in labor input over the life course, the indicators should be constructed for relevant population groups, and they should be used on reliable and up-to-date material (ibid.). In addition, the indicators should provide new insight into policy changes. In practice, pension evaluation focuses on changes in labor market status by gender, age and level of education. Expectation-based techniques are good indicators, but dynamic microsimulation provides an even better means for analyzing future life courses. There are at least four measures or indices to evaluate employment and working life in a pension context.

The expectation-based life table is an established tool in actuarial science (Hoem 1970). The technique is used to study labor force status and mobility by constructing a working life table and computing estimates of working life expectancy, which is the future time an individual is expected to spend in employment. There are different types of analytical approaches (see Hytti and Nio 2004).

One, the duration of employment depicts the average years that a 15-year-old person can be expected to be in employment or self-employment during the remaining years of life, if the rates of employment during the year in question would prove to be permanent. This is in essence an application of the Sullivan method of estimating working-life expectancy (Sullivan 1971). The material used in this dissertation are workforce share and employment rate. The data is drawn from the Labor Force Survey of Finland. The first measure, the duration of active working life for a 15-year-old, was estimated to be 37.8 years in 2017. The second measure, duration of working life, was 34.5 years. The difference is mainly due to unemployment. (Pension Indicators 2019, 10; Official Statistics of Finland 2019b.)

Two, working-life expectancy is the expected number of working years remaining in an individual’s life. In practice, the expected times can be time spent in employment, unemployment and/or outside the labor force. This is an application of a multistate life table regression method for aggregate data by Brett Davis (Davis et al. 2001). The technique uses a multivariate logistic regression model to estimate age- and year-specific probabilities of being in the labor market state of interest. Using Finnish Labor Force Survey data, the method provides working-life expectancies for 15–64-year-olds over empirical data. Further, model-based predictions can be extended for some additional years. (Nurminen 2012; Järnefelt and Nurminen 2013.)

Three, length of working life is the duration of the time, in months or years, of coverage by the earnings-related pension scheme. In such cases, working life includes only employment or self-employment insured for earnings-related pensions. In this review, a person is considered to have been at work during a specific month if they have been employed or self-employed and insured for earnings-related pensions during said month according to register information. According to statistics, the average working life of those retiring on an old-age pension in 2017 was 35.4 years and median 38.8 years (Pension Indicators 2019, 12).

Four, in addition to the above, the much used estimate of active working life, that is, the employment rate, is the percentage share of employed persons of the population of the same age. The review is based on the annual average values of the Labor Force Survey by Statistics Finland. Normally, the employment rate is calculated as a percentage share of the same-age population among the employed between 15 and 64 years of age. This being the case, 65–69-year-olds do not impact the employment rate of the overall population. The employed are considered to be persons who, during the survey week, were in gainful employment and received either a monetary salary for at least an hour's work, fringe benefits for work, profit from self-employment, or who were temporarily off work. In 2017, the employment rate of 15–64-year-olds was 69.6 per cent. (For more detailed information, see Pension Indicators [2019, 11] and Official Statistics of Finland [2019b]).

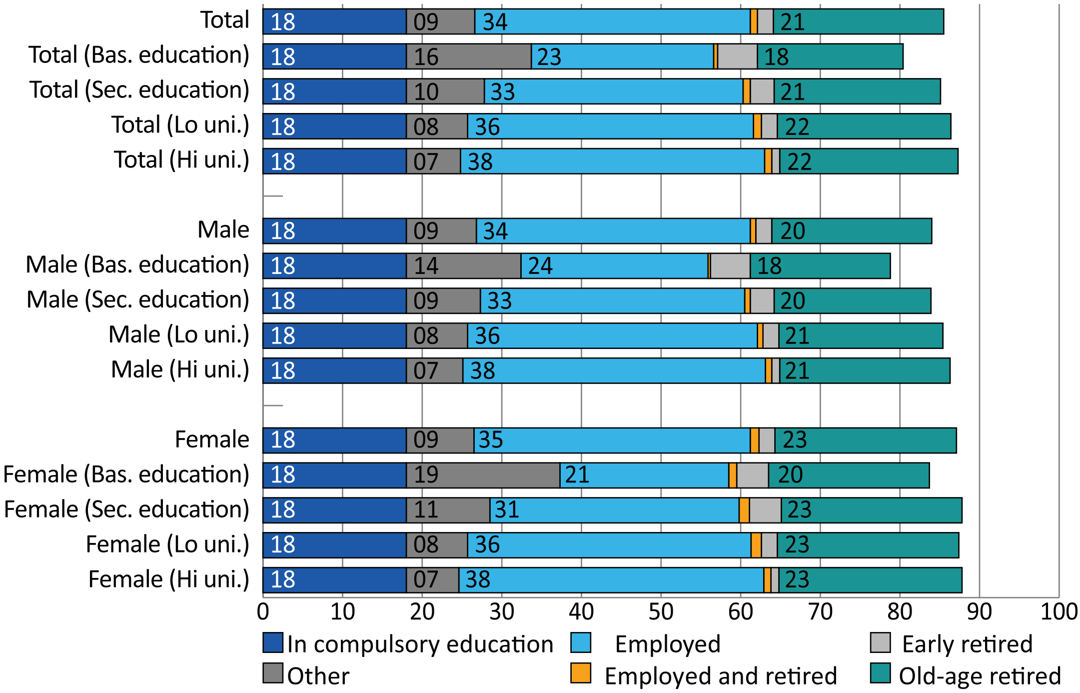
5.4 New measure: partition of the life course

The dynamic microsimulation model offers new possibilities to illustrate the lifespan over a simulation horizon. With the ELSI model we can illustrate the lifespan via the model's labor market and population states and use other individual level information, such as gender and level of education, to go further into details of the life course. Beyond illustrating the lifespan in general, it is also possible to evaluate policy reforms from a life course perspective. For example, in the second article of the thesis, we presented the lifespan effects of the 2017 pension reform on the cohort born in 1988. We call this analysis a partition of the life course.

I can elaborate further on the analysis of the same cohort. The ELSI model includes empirical and simulated labor market states for the years 2009 to 2085. We can reasonably well track the labor market and population states from compulsory education until death for the aforementioned cohort. Figure 2 gives an overview of the lifespan by gender and highest educational level. The education is the highest educational level at the end of the simulation period, which means that it includes the individual education dynamics of the microsimulation model. The lifespan consists of six phases (in education, other, employed, employed and retired, in early retirement and in old-age retirement), which are aggregated from the ELSI model labor market and population states. The population includes those who reside permanently in Finland. Immigrants and emigrants arriving to or leaving Finland during the simulation period are excluded from this example.

The picture shows that, on average, this cohort (Total) spends 18 years in compulsory education, 8.6 years outside labor markets (unemployed, on parental leave), 34.6 years employed, 2.3 years in early retirement and 21.4 year in old-age retirement. The small section without an estimate of duration is time spent in employment while drawing a pension, which is 0.9 years in total. The total lifespan consists of these shares, which add up to 85.8 years. When comparing men and women, we can see that women spend more time in retirement than men (22.8 yrs. vs. 20.1 yrs.), while the time spent in employment is fairly equal (34.7 yrs. vs. 34.4 yrs.) for both men and women. The women are expected to live 87.4 years and men 84.3 years, a difference of 3.1 years. The figure shows profound differences relating to the level of education.

Figure 2.
Partition of the life course of the 1988 birth cohort, years on average



Source: Own calculations based on ELSI model.

As shown in Figure 2, we can also focus on individual life stages such as employment or time spent in retirement. Concentrating on specific parts of the life course makes it possible to evaluate a wider range of cohorts. Currently, we can estimate employment for cohorts born between 1988 and 2015. The same applies to unpaid periods. The estimate on time spent in old-age retirement can be given to cohorts born between 1945 and 1988.

6 Statistical foundation of trajectory analysis in finite mixture modeling

6.1 Finite mixture modeling basics

The statistical foundation of trajectory analysis is in finite mixture modeling. The finite mixture model provides a representation of the population heterogeneity in a finite number of latent sub-population classes. Well-established in the field of statistics, the mixture modeling method concerns modeling observed probability distribution by a mixture (or weighted sum) of distributions (see Titterington et al. 1985; McLachlan and Peel 2000). Böhning et al. (2007) give examples of the use of finite mixture modeling in various statistical applications.

The fundamental idea of mixture modeling is that the observed distribution is assumed to be a mixture of K different distributions. In other words, the data is supposed to be composed of K sub-populations (or alternatively latent groups). The finite mixture model is semi-parametric, as it combines parametric estimates with a nonparametric estimate of the population structure. In its simplest form, the idea can be formulated in the following way. First, the density function for K-component finite mixture is

$$f(y|\boldsymbol{x};\theta_1,\theta_2,\dots,\theta_K;\pi_1,\pi_2,\dots,\pi_K)=\sum_{j=1}^{K}\pi_j f_j(y|\boldsymbol{x};\theta_j),$$

where $0<\pi_j<1$ and $\sum_{j=1}^{K}\pi_j=1$. π_k refers to the probability of mixture component k, f refers to the conditional probability density function, θ_k refers to parameters of distribution in component k. and y refers to the observed variable or outcome. The π's are the shares or proportions of components in the whole population or sample. The θ's are the component specific parameters of the distribution. The population density function can be, for example binomial, multinomial, poisson or normal.

When estimating the mixture model, various covariates are commonly included into the model, as well. Therefore, the covariates may contain important auxiliary information and thus help to find the right model solution.

The mixture model is usually estimated with EM algorithm (Dempster et al. 1977). The estimation yields component shares and other parameter estimates. The EM algorithm yields maximum likelihood estimates for model parameters. One drawback of the EM technique is that the algorithm may converge to local maximum, which is why the model should be estimated several times with different starting values. The maximum likelihood estimation problem is defined by the equation

$$\max_{\pi,\theta} \ln L = \sum_{i=1}^{N} \left(log \sum_{j=1}^{K} \pi_j f_j(y|\theta_j) \right).$$

The final assignment of individuals to specific groups is based on the posterior probability

$$Pr[y_i \in population\ k|x_i, y_i; \theta] = \frac{\pi_k f_k(y_i|x_i, \theta_k)}{\sum_{j=1}^{K} \pi_j f_j(y_i|x_i, \theta_j)}, k = 1,2, \dots, K,$$

where each parameter is substituted by their final estimates. Note that the identified groups are not fixed constructs but are based on probabilities. Therefore, the groups should be used with caution in further analysis.

The crucial step in finite mixture analysis is to estimate model parameters with an increasing number of groups or components (1, 2,…, K). The decision on the number of components is usually based on some information criteria of which the most common are Bayesian Information Criteria (BIC) and Akaike Information Criteria (AIC). Beside information criteria, looking at the quality of the estimated model and its interpretation is important as it is always good to have clear discrimination between components.

In practical applications, the stages of modeling are often the following:

1) specification of the model,
2) checking the identifiability of the model,
3) estimation of the model,
4) testing and evaluating the overall fit of the model, and
5) doing diagnostics of the parameters of the model.

Oftentimes the model needs some modification or re-specification, in which case steps 1 to 5 are repeated. In principle, the same type of operation can also be applied to mixture modeling where applicable.

An example: mixture components of pension distribution

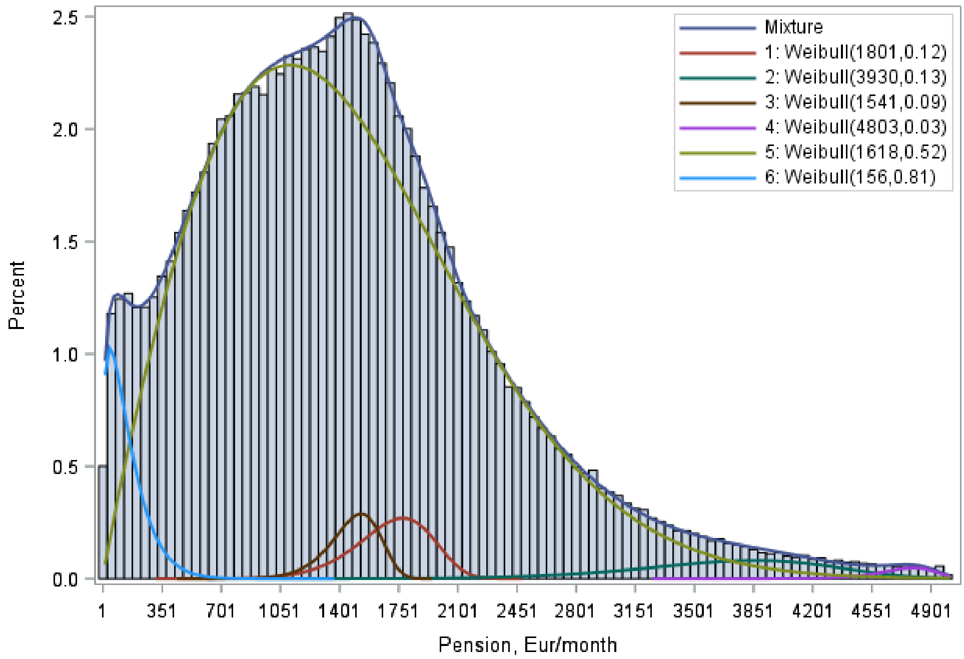
As an example of finite mixture modeling, the distribution of statutory pensions is considered next. In the example, a 25-per-cent random sample (N=359,708) of earnings-related pensions (paid on 31 Dec. 2018) is selected from a research register of the Finnish Centre for Pensions. The sample consists of pension in own's right (i.e., excluding survivors' pensions). It can be expected that the stock of pensions is not homogenous but consists of a large body of old pensions and some new pensions). The pension stock further reflects the working lives and income distribution of the retirees over a long time-period.

Figure 3 shows the distribution of pensions indicated by the histogram bars (SAS, FMM Procedure implementation). The distribution is not symmetric like a normal distribution but slightly positively skewed (g_1= 0.8795). The stock of pensions centers around 1,600 euros (Mode = 1,520 and Mean = 1,453). From the perspective of mixture modeling, the question is: does the distribution consist of latent sub-groups, and of how many? The example includes the analysis of the number of mixture components (sub-groups) and is indicated by the BIC-values. According to BIC, a six-group solution yields the lowest value and thus provides the best fit for the data. It is expected that the population mixture follows the Weibull distribution. Based on the analysis it appears that the distribution of pensions consists of a mixture of six Weibull distributions.

The estimated sub-populations are of various sizes. The largest group ($\hat{\pi}_5$) includes 87.9 per cent of the sample while the smallest group ($\hat{\pi}_4$) includes only 0.4 per cent of the sample. In the range of pensions, group $\hat{\pi}_6$ is located at the low end (average pension 156 euros) and group ($\hat{\pi}_4$) at the high end (average pension 4,803 euros) of the distribution.

The example shows that the pension distribution consists of a mixture of six Weibull distributions, which was not evident from the original histogram. The groups revealed by finite mixture modeling can be studied in more detail in further statistical analysis; for example, by linear regression models with covariates.

Figure 3.
Distribution and estimated density of earnings-related pension with estimated component densities



Source: Own calculations based on Finnish Centre for Pensions pension register.

6.2 Trajectory modeling

Trajectory modeling is an application of finite mixture modeling with a focus on longitudinal analysis, that is, capturing developmental changes over time. The early origins of latent modeling relate to the widely used tradition of structural equations modeling (SEM), dating back to the pioneering works of Tucker (1958) and Rao (1958). The structural equations modeling tradition is discussed in depth in, for example, MacCallum and Austin (2000). A further discussion on the SEM approach or other similar approaches (e.g., Warren et al. 2015) are beyond the scope of this thesis.

Trajectory modeling includes three modeling approaches, which share the same task: to capture developmental trajectories in a given population. These techniques are applied in many fields of science. Part of the success of these techniques lies inthe increasingly improved computational capacity. Mixture modeling techniques are readily available in many statistical program packages such as Mplus, Stata, SAS and R.

The homogenous example (eq 1) is now extended to cover longitudinal measurements of individual behavior over T periods ($\boldsymbol{Y}= (y_1, y_2, \ldots, y_T)$. The model parameter(s) θ also depend on time.

In the context of trajectory analysis, the terminology from finite mixture modeling a bit unsettled. Terms class, group, or component refer to an unobserved or latent sub-population. In addition, the terms class probability, component probability and mixing weights refer to the probability to belong to a given sub-group. Furthermore, the term trajectory is used quite liberally in the literature. The techniques described next are about group-specific (latent or ex ante classified) mean modeling and thus in essence trajectory analysis. In social sciences, the term trajectory has also been used in connection with sequence analysis, where the term is often used to describe life course transitions without implications to mean modeling.

Trajectory analysis or latent class analysis includes two basic main modeling approaches: the growth mixture modeling and the group-based trajectory modeling. They share the same analytical objective of measuring and explaining differences across population members in their developmental course. The difference between approaches lies in the way they model the individual-level heterogeneity in developmental trajectories. These approaches make a slightly different technical assumptions about the distributions of parameters and error terms within the trajectory groups (see Nagin and Odgers [2010] and Nathalie et al. [2017] for a discussion of different trajectory analysis approaches). Next, I will briefly introduce the techniques.

The first trajectory technique is called growth curve modeling (GCM), which is part of a hierarchical linear modeling approach and specifically designed for longitudinal analysis (e.g., Duncan et al. 1999; Singer and Willett 2003; Curran et al. 2010). Growth curve modeling is widely applied in youth studies, psychology, medicine and sociology (e.g., Lyytinen et al. 2006; Onogi et al. 2019; Meuleman et al. 2017). The analysis centers around inter-individual growth-patterns and possible underlying predictors. In practice, the technique is advanced in measuring different aspects of development or growth, statistical predictions (e.g., model fit), handling missing observations and different data structures or study-designs. The growth curve modeling deals with a fundamental issue – developmental trajectory – but it does not intend to reveal latent groups with different trajectories. Growth curve modeling can be used to analyze mean trajectories by ex ante classifiers (e.g., gender).

The second trajectory technique is called growth mixture modeling (GMM), in which the idea is to apply a finite mixture modeling so that two or more growth curve models are used to model the population variability within developmental trajectories (Muthén and Shedden 1999; Muthén et al. 2006; Muthén and Muthén 2010, 186–187). In practice, growth mixture models imply that separate sub-populations follow a different growth curve. Growth mixture models depict the average trend of outcome and individual-specific variation around the average trend (random effects), using the same parameters of change for the population (Nagin and Odgers 2010). See Tolvanen (2007) for advances in growth mixture modeling, also important in the context of this thesis.

There is a wide range of applications of growth mixture modeling. Adolescence delinquent behavior is a long-standing topic in psychology where growth curve modeling has been applied (e.g., Wiesner and Wiendle 2014). Gomez et al. (2017) is another example of psychological application. The technique is also applied in medical sciences (e.g., Stull et al. 2011; Espy et al. 2009) and in youth studies (e.g., Salmela-Aro and Tynkkynen 2009], Salmela-Aro and Upadysya (2014).

The third trajectory technique is called group-based trajectory modeling (GBTM), which is also an application of finite mixtures (Nagin, 1999; 2005). The technique has been applied in many fields of science, especially in psychology, criminology, health sciences and sociology. Nagin and Odgers (2010) give an overview of the applications in psychology and criminology. Tthe next section includes more examples. Group-based trajectory modeling assumes that the population is composed of a mixture of distinct groups defined by their developmental trajectories. It takes no stand on the population distribution of trajectories but, instead, uses the trajectory groups as a statistical device for approximating the unknown distribution of trajectories across population members (Nagin and Odgers 2010).

Group-based trajectory modeling can adapt to longitudinal analysis with different study-designs. The method is flexible an adapts to different distributions of the outcome. The group-specific mean model can vary between groups, and contain time-stable or time-varying covariates.

The two trajectory methods have one difference. Nagin's original idea assumes that, in a given sub-population, longitudinal measurements are independent (conditional independence). This assumption implies that, for each individual or object within a sub-population, the distribution of outcome in period t is independent of the outcomes in prior periods. However, on a population level, measurements can still be dependent. The assumption of conditional independence yields a simple likelihood function, and the model can be easily estimated with readily available software. In growth mixture modeling, the longitudinal measurements within trajectory groups may be dependent. Unfortunately, in some cases that may yield a proliferation of the model parameters. Therefore one weakness of the growth mixture model is its increased computational complexity and instability. A further discussion on the nature of trajectory groups in growth mixture modeling and group-based trajectory analysis is given in Nagin (2005, 54–56), Nagin (2016) and Nathalie et al. (2017).

The two trajectory methods (GMM and GBTM) have similarities, but which one is most useful in pension evaluation? Both methods are very useful and bring new insights into many research questions. Maybe the selection is best based on available software and the nature of the data at hand (e.g., Jung and Wickrama 2008). In terms of explorative vs. confirmatory analysis, both techniques are equally valuable as they can flexibly include multinomial logistic regression. The trajectory software is available in several environments. The datasets in pension evaluation are often large, consisting of high-frequency data, so any computational bottleneck cannot be tolerated. This thesis focuses on the group-based trajectory modeling approach.

Indeed, trajectory analysis, or group-based trajectory modeling, has become popular in many fields of science, including psychology, criminology (e.g., Nagin and Odgers 2010; Nagin 2016; van der Geest et al. 2016, Nathalie et al. 2017), sociology (e.g., Hynes and Clarkberg 2005; Don and Mickelson 2014), education (e.g., Kokko et al. 2008), marketing (e.g., Mani and Nandkumar 2016), health sciences (e.g., Hilliard et al. 2013; Nummi et al. 2014; 2017a), youth studies (e.g., Edgerton et al. 2015) and gerontology (e.g., Hsu 2015), just to give some examples. Disability and sickness benefits are much-studied topics in pension evaluation, involving trajectory analysis (e.g., Murphy et al. 2016; Laaksonen et al. 2016b). Trajectory analysis has increasingly been used in studies investigating labor market attachment (Peutere et al. 2015;

sub-study I; Saloniemi et al. 2020). Sub-study I of this thesis is another example of an analysis of labor market attachment in the context of youth studies. Sub-study IV, on the other hand, is a novel, first-line introduction of trajectory analysis in the field of microsimulation.

Trajectory analysis can easily be extended further. For example, the dual-trajectory modeling approach outlined in Nagin (2005, 141–170) makes it possible to model two outcomes simultaneously. An example of dual-trajectory analysis with application in health sciences is given in Barker et al. (2007) and Li et al. (2012). The modeling has been developed further to enable more than two outcomes (see Nagin et al. 2016). Multi-trajectory modeling in the form of multivariate trajectory modeling is potentially very interesting in pension evaluation. More and more applications of multi-trajectory modeling appear daily (e.g., Hsu 2015 and sub-study I of this thesis). There are naturally many places in modeling and estimation where statistical advances are possible. One is related to more flexible mean modeling within sub-groups, where a wide range of curves are available for mean modeling. The novel thing about this approach is that it can be done separately within sub-groups (see Nummi et al. 2017b). Another relates to handling different missing data patterns (e.g., Haviland et al. 2011). A further topic for development would relate to stability of trajectory groupings with respect to transformations of outcome variables.

6.3 Trajectory analysis in practice

Tips for practitioners

The central part of trajectory analysis is group-assignment of individuals. Individuals are assigned to sub-groups according to posterior probabilities. An individual has a probability to belong to any of the groups requested by the researcher. The individual is assigned to a specific group based on the highest posterior probability. The assignment must be understood right, as the groups are not fixed constructs but are based on probabilities. The groups revealed by trajectory analysis are not necessarily real entities. They are mere approximations about groups, as real entities are described in Nagin and Odgers (2010) and Nagin (2016).

Operationalization of trajectory analysis requires careful consideration of study-design. Simple events without specifying time or age are not possible designs in trajectory analysis (they can be analyzed with different finite mixture techniques, like discussed in Chapter 6.1. of this thesis). Trajectory analysis is essentially longitudinal analysis aiming to reveal latent groups. Therefore, a follow-up design with some (usually at least four) repeated measures is suitable for analysis. In labor market or life course studies, a cohort or group of cohorts is a natural subject for analysis. The study period defines the groups and trajectories. A change in study period can also change groups and trajectory model estimates. This is something a researcher must keep in mind if the longitudinal nature of the data can or must be adjusted.

There are several possibilities with trajectory analysis. A simple case would be a univariate trajectory analysis with only one outcome variable. Oftentimes a more interesting possibility is a multivariate trajectory analysis where two or more outcomes are modeled simultaneously. This analysis yields a rich and easily comprehensible statistical view of the developmental linkages between outcomes. Another option would be dual-trajectory analysis with two outcomes, which is analytically different from the aforementioned approaches. In dual-trajectory analysis, it is possible to have, for example, three latent groups of one outcome, and the same individuals are positioned into two groups of the other outcome (e.g., Nagin and Tremblay 2001; Nagin 2005, 141–170).

When applying trajectory analysis to economic data like pensions or earnings, it is good practice to consider the scale of the outcome variables, should original scale be used, or should the variable be standardized or otherwise scaled. When modeling pensions or earnings, one approach is to divide outcome by 100 for instance. Another approach, which is commonly used in econometrics is a natural logarithmic transformation. If outcome(s) and possible covariates are transformed to logarithms, there are nice properties with the model estimates, as the effects of covariates on outcome can approximately be interpreted as percentage changes. Furthermore, in econometrics one transformation is especially common, namely inverse hyperbolic sine transformation (Burbridge and Robb 1988; Friedline et al. 2015), which is approximately the same as logarithmic transformation. The IHS-transformation has nice properties too. First, when applied in software packages, it deals with zeroes. Zeroes are usually valuable information and should be included in the analysis. Second, the transformation usually affects the distributional properties of an outcome in way that adjusts for skewness. There is further discussion on data transformation in Ruppert (2001), for example. It is important to consider data transformations in trajectory analysis too, because the sub-groups may be affected by the transformation.

Scaling the covariates of the mean model is to be advised. Time or age and their polynomials are required factors in trajectory modeling. Scaling the successively higher terms in the polynomial that defines the trajectory into the same scale is good practice for computational reasons (see Nagin 2005, 44).

Identifiability is an important required property of any statistical model. In trajectory analysis, identifiability is challenging as it appears in at least two stages: identifying true number of sub-groups from an underlying population, and identifying the true number within group models. Statistical testing and experimenting with the model specification gives the assurance of identifiability, keeping in mind that the simple model is often the preferred model, because it is more generalizable to the total population A good starting point is to make the parameters equal between sub-groups to get the most parsimonious model. To ensure the identifiability of the model empirically, one must start from a model that is known to be identified and check if the observed data log likelihood changes when adding a parameter (see Muthén and Shedden 1999). Oftentimes this gives a reasonable model specification. In practice, making a visual inspection of the trajectory curves, and subjectively (or theoretically) evaluating the group shares with respect to reality, is good advice. Checking the composition of the groups is also a good way to validate trajectory results. (See Nylund et al. 2007 and Tein et al. 2013 for a further discussion about the correct testing measures to reveal correct groups.)

The outcome of trajectory analysis includes both group assignments and estimated model parameters within each group. This provides possibilities for illustrating the developmental trajectory mean estimates. Readily available software usually produce graphs of developmental trajectories based on model estimates or model fit. The model estimates are usually rather smooth curves, depending on the underlying group-specific regression models. Another possibility is to illustrate trajectories based on averages or medians calculated from the outcome data. The latter is a good approach in my mind. It is used in the sub-studies of this thesis.

In finite mixture modeling, as in trajectory analysis, the number of sub-groups is the central question. The ordinary $\chi 2$ likelihood ratio test is oftentimes used to evaluate the fit of a statistical model in general. However, this is not directly appropriate when testing the number of latent sub-groups. The information criteria AIC or BIC, or statistical tests like VLMR, LMR (see Lo et al. 2001) or BLRT, are used instead. Nagin gives examples of using BIC to evaluate the question (Nagin 2005, 63–77) and Don and Mickelson (2014) discuss the good practices of trajectory analysis to some degree, especially in terms of model selection. There are also other and more sophisticated statistical criteria that may be useful (e.g., McLachlan and Rathnayake 2014). Trajectory analysis practitioners should also use good judgement and expert elicitation when evaluating the number of groups with respect to the research question at hand. There should be a clear discrimination between sub-groups so that the groups represent real sub-populations. An inspection of the posterior probabilities of individual trajectory assignment is also recommended. The maximum probability of belonging to a certain group should be near to one. One way of evaluating the group assignment is to count averages of group posterior probabilities. The group specific average should be at least 0.7 for all groups (see Nagin 2005, 88).

Finally, computational limitations should be taken into consideration when applying trajectory analysis. It is advisable to consider sub-samples if the original population exceeds, for example, 50,000 individuals. As always with a statistical method using EM algorithm, it is good practice to run the trajectory model up to ten times to find stable model estimates and group assignments for the study population. Discovering the best model fit requires several starting values. As different starting values produce equal results in terms of equal log likelihood, a best fitting model is probably obtained.

Possibilities in pension evaluation

This thesis, and especially sub-study IV, promotes the idea of group-based classification rules. In general, most pension evaluation and life course research apply ex ante classification rules. Typical ex ante classifiers or pre-determined groups in pension research are gender, level of education, age and recipients of a certain pension benefit. These classification rules can be flexibly used in various study-designs. This kind of research yields vivid information on progress in and within those ex ante groups. Readers, fellow researchers and policy makers are used to these classification results, and there is certainly a permanent need for them. Trajectory analysis (and group-based techniques in general), on the other hand, provides new possibilities in life course research and pension evaluation. The data-driven methodology reveals latent groups in a study population. Groups and their relative sizes give a clear indication of which progressions are dominant or marginal in a given population. These group-based rules can bring new insights. In my view, traditional ex ante classification rules and group-based rules could complement each other.

New data possibilities yield better opportunities to increase the frequency of longitudinal measurements. In labor market studies, outcomes have usually been measured on a yearly basis. For example, labor market state (e.g., employed, unemployed, sick or retired) is traditionally measured either during a calendar year or more often at the end of the year (31 Dec). Currently the register material of the Finnish pension system allows for more frequent measuring of labor market states. Doing it on a monthly basis is an interesting possibility for the future. Trajectory analysis is promising in this landscape.

The question of executing pension evaluation with a descriptive or a confirmatory approach can also be answered within trajectory analysis. Some scientific research and the majority of administrative studies of pensions or life course in general are purely descriptive. The research and substantial contents are based on groups (ex ante or group-based classification rules) and their qualities. The trajectory methodology is in place to make a more advanced research with confirmatory analysis within the trajectory groups. For example, multinomial logistic analysis is possible in trajectory analysis (e.g., Kuitto et al. 2019b).

Register-based individual analysis, or high-frequency data, is not the only possibility with finite mixture techniques. Trajectory analysis can equally well be used with low-frequency data. For example, a group of countries can be analyzed if there are enough longitudinal measurements per country. Latent groups of countries would be the result in this case (e.g., Jahn and Helmdag 2019).

Limitations

Group-based methods, especially trajectory analysis, has also been criticized (e.g., Warren et al. 2015; Berlin et al. 2018). Concerning pension evaluation, and especially microsimulation, there are at least seven issues that need to be kept in mind.

One, group-based methods can be applied directly with a discrete-time outcome. The time-wise measurements of outcome must be the same for every individual in the population. In the context of microsimulation, this means that group-based modeling can be applied with the state microsimulation modeling, that is, with a model that operates with discrete time. Some microsimulation models operate with continuous-time events, and such data cannot directly be analyzed with group-based modeling techniques. The study-design with event data should include converting outcome into discrete-time context.

Two, the aim of trajectory modeling is to reveal sub-groups of the population. The modeling is merely a technique, which can reveal irrelevant sub-populations. It is possible to find artefact groups despite using statistical information criteria. Consequently, it is important to use expert opinion in selecting the number of groups which have meaningful real-life interpretation.

Three, mean-based modeling can always be criticized by cancelling or offsetting the variation in the population. Though group-based modeling reveals sub-populations with different mean trajectories, it is slightly insensitive to within-group variation. The yearly variation of outcome or heteroskedasticity is somewhat overlooked by mean-based modeling, yet group-specific regression models include confidence limits.

Four, group-based modeling is often used as explorative analysis with an emphasis on revealing latent groups in data. However, the confirmative analysis,as with any regression analysis, is possible, and it provides enough information content on the topic of interest. Latent trajectories may also reveal interesting causalities.

Five, the assumption of conditional independence was mentioned earlier (see Nagin 2005, 26–27 for a further discussion). However, conditional independence is an assumption applied with many other statistical methods. From a practical point of view, it can probably be seen as a slightly broken assumption with no real significance in practical applications (see Nagin 2005; Mani and Nandkumar 2015, 194).

Six, the optimal number and composition of sub-groups is probably affected when changing the period of observation. Such a change would naturally alter the results of the model, and the groupings might also be affected to a certain extent. This is true with other statistical analyzes, as well. However, according to my experience, with trajectory analysis, the basic main grouping structure would remain quite similar.

Seven, the model estimation is oftentimes done using the EM algorithm, which is inclined to also find possible local solutions (e.g., Goodman 1974). In practice, non-convergence issues are somewhat common. (See Hipp and Bauer 2006 for a further discussion.)

7 Microsimulation

7.1 The origins of microsimulation

In general, microsimulation is an attempt to model complex reality in an analytical framework. In the context of policy evaluation, the complexity arises in three areas: structure of the population, behavioral responses to policy, and policy structure. (O'Donoghue 2014, 3.) From a conceptual point of view, these areas are (to some degree) the focus in all models. From a practical point of view, microsimulation is a tool to answer what-if questions using model generated synthetic units, usually households or individuals.

The conceptual framework of microsimulation was outlined in groundbreaking work by Guy Orcutt (1957). He proposed a model with demographic change, labor market dynamics and family formation to simulate sample data. The foundation of dynamic microsimulation was laid 60 years ago (Orcutt 1960; Orcutt et al. 1961). However, the origins of microsimulation were established together with the construction of the system of national accounts (SNA) in the 1940s. The SNA soon became the successful framework for describing financial activity in an economy. Guy Orcutt was one of the pioneers, who applied the empirically-based macroeconomic simulation model to SNA. (Wolfson 2009.)

The first proper microsimulation model, DYANSIM, was developed in the late 1970s (Orcutt et al. 1976). Since then, both static and dynamic microsimulation has spread to a range of applications. There appears to be three parallel lines of work going on in the field of microsimulation. The early work was centered around static modeling which focused on tax-benefit analysis and related distributional assessments. This was essentially analysis at a given point in time, yet it has developed and spread around the world tremendously over recent decades. Essentially, this is still the core of static microsimulation. The second line of application involved time and populational characteristics over time. This is essentially dynamic microsimulation, with a more-or-less ambitious focus on modeling individual behavior within a changing population. The third line, which saw a need to merge microeconomic behavior with macroeconomic modeling and vice versa, has recently become more popular. As microsimulation and macroeconomic modeling need to measure the same fundamental aggregates, this increasingly sophisticated field of applications has benefitted both approaches. (Williamson et al. 2009.)

Since the 1970s, the field of microsimulation has broadened strongly. Currently the topics in which microsimulation has been used or has become commonplace are the following: demography, public health, social insurance, traffic analysis, farm analysis, environmental analysis, firm analysis and sociology, as well as many other aspects of public policy (e.g., Anderson and Hicks 2010; Li et al. 2015). The handbook edited by Cathal O'Donoghue (2014) gives an overview of current developments across the sub-fields of microsimulation modeling.

Next, I will briefly introduce the basic ideas of static and dynamic microsimulation with an emphasis on models developed and used in Finland or models that have otherwise influenced the ELSI model.

7.2 Static microsimulation

The static model can be defined as a tool that focuses on policy reforms and presents their resulting impact on policy reforms in a country. There are several static microsimulation models in use worldwide (see Li et al. 2014 for examples of some of the models).

In the Finnish context, there is a history of static models used by government agencies and economic research institutes (see Haataja 1994). One early taxation model for administrative use, HVS, was developed in collaboration with the Ministry of Finance and Tax administration (see Kuula 1997). TUJA, a model of the national economy , was intensively used in ministries from the 1980 to the 2000s (see Niinivaara and Viitamäki eds. 2005). SOMA is another model with a similar approach to national economy (see Parpo ed. 2006; Sallila 2009). The SOMA model is used in evaluating social policy reforms with special emphasis on poverty measures. In the 2000s, a new static model JUTTA was developed (Honkanen 2010). Its early home base was in the Kela. In 2011, Statistics Finland took over the model since it held the relevant databases. The model was significantly revised and is now called SISU. The model programming code is freely available, but the source data must be obtained from Statistics Finland. SISU is nowadays widely used by government agencies, academia and economic research institutes. Next, I will briefly describe the SISU model.

The SISU model runs on sample data that represents the total population. The model can analyze both individuals and households. SISU is a typical static model in the sense that it does not include behavioral effects or dynamic ageing. Reform analysis and immediate tax and welfare effects are at the core of SISU analysis. The model has a modular structure, with 13 modules in the base model. The modules simulate various parts of the income transfer system. Detailed simulation includes the following benefits: sickness benefits, unemployment benefits, pensions paid by Kela, child- and parental benefits, financial aid for students, housing benefits and social assistance. The model includes taxation, namely personal income taxation and personal property taxation.

The model is based on a large register-based sample data on the service data of income distribution. It is common to run the model with a sample of 800,000 individuals. The SISU model includes static data ageing tools that can be used together with macro-level predictions in order for the source data to better represent the current year or a year of interest a few years into the future.

In a short time, the SISU model has acquired a strong group of users. Knowledge-based decision making has advanced greatly with SISU being used in ministries and parliament. The excellent webpages of Statistics Finland provide an up-to-date view of the policy areas where the model has been used so far. Some of the topics investigated with the SISU model include the national pension, the guarantee pension, pensioners and the older population.

The policy analysis of changes in unemployment insurance is a popular topic with the SISU model (e.g., Kotamäki and Ahola 2014; Ahola et al. 2015; Kotamäki et al. 2018). The evaluation of the effects of the government's yearly Budgetary Plan on basic social security (including the national pension and the guarantee pension) has been established in recent years (e.g., Lehtelä et al. 2016; Mukkila et al. 2017; Mukkila et al. 2018). Another established analysis with the SISU model is the evaluation of basic social assistance, with an emphasis on poverty analysis among the older population (e.g., THL 2015; Mukkila et al. 2015; Moisio et al. 2016; Mukkila et al. 2017). Pensioner poverty has been analyzed further in Kotamäki and Kärkkäinen (2017). A detailed analysis of the use of social assistance among ageing adults (Ahola et al. 2014) and users of the tax credit for household expenses (Pylkkänen 2015) are examples of analyzes that focus on the older population. Family policy and survivor's pensions from the national pension scheme have been studied in Koskenvuo et al. (2014) and Haataja ed. (2016). In recent years, some methodological developments in terms measuring poverty and income distribution in the SISU model have also been studied (e.g., Sallila 2015; Lappo 2015). Dynamic properties of the static model have been studied in Honkanen and Tervola (2014) and Vaalavuo ed. (2018).

EUROMOD is a multi-country modeling adventure enabling a cross-country policy evaluation. It is an established tax-benefit modeling tool, incorporating Finland among other 27 European countries. The aim of EUROMOD is similar to that of the SISU model: to simulate benefits, transfers and taxes. The EUROMOD is an interesting modeling approach with similar databases (EU-SILC) across Europe (Figari et al. 2015). The model provides a detailed analytic framework to compare countries, with special emphasis on measuring the effects of policy on the income distribution. Broader substantial topics, like social inclusion or redistribution, are also analyzed by the model. However, there is a drawback with the EUROMOD in Finnish pension evaluation, as the pension content is limited to the guarantee pension (e.g., Honkanen et al. 2017).

7.3 Dynamic microsimulation

There is also a range of dynamic microsimulation models developed in the field of social sciences since Orcutt's time. Dynamic microsimulation models represent diversity in methods, yet many things have remained the same over the decades. The progression of time and changes in population are at the center of focus in dynamic microsimulation. The unit of simulation in social policy-oriented microsimulation can be the individual, the household, the firm or even the farm. In terms of pensions, the interest focuses on the individual or the household. Current models represent various approaches to life course analysis and are different in scale. Some models simulate most parts of society and some are more limited in their scope.

Over the decades, models have been developed around the world in academia, government agencies and statistics institutions. Some models have already been discontinued (e.g., Canadian LifePaths and DYANCAN), but the need to study life courses has guaranteed the development of new models. Nowadays there are a number of established models concentrating on life course analysis and pension policy evaluation: Swedish SESIM

(Klevmarken and Lindgren 2008, Flood et al. 2012), French DESTINIE (Blanchet et al. 2010), Australian APPSIM (Harding 2007), Norwegian MOSART (Fredriksen and Stølen 2017), Belgian MIDAS (Dekkers 2010; Dekkers et al. 2010), Danish SMILE (Hansen et al. 2013) and Japanese INAHSIM (Inakagi 2014). (See Zaidi and Rake 2001, Li and O'Donoghue 2013, Li et al. 2014 and Dekkers and Van den Bosch 2016 for more examples of dynamic microsimulation applications.)

Dynamic microsimulation does not necessarily require time-consuming model building as ad hoc modeling can be made with newly developed generic modeling tools such as LIAM (de Menten et al. 2014) or Modgen (Statistics Canada). (See Leombruni and Mosca 2014 for an example of how these tools have been used in analyzing redistributive aspects of the pension system.)

The ELSI model (Tikanmäki et al. 2014; sub-study I) is a new model compared to the above-mentioned established models. Experiences from other models were considered when designing and building the ELSI model. Solutions and lessons from other models are interesting and inspirational in terms of pension modeling, and also in the future development of ELSI.

There are other approaches to life course simulation, as well. One approach, which has its roots in theoretical economics and econometrics, is the stochastic life cycle modeling to simulate individual (model generated) careers. Life cycle models generate optimized, fictive individuals, in contrast to the abovementioned dynamic microsimulation models, which operate on true, empirical individuals. The dynamic models aim to model behavioral adjustments over the life course using dynamic programming techniques. They oftentimes include, on a detailed level, many essential contents of economy such as social security benefits and pension systems and taxation, making them very useful and versatile. Equivalent model have been developed in Finalnd, as well. Hakola and Määttänen (2007; 2009) give examples of evaluating the effects of 2005 pension reform in detail. A study on alternative pension reform options, and especially linking retirement age to life expectancy, is another example of using a stochastic life cycle model (Määttänen 2014).

8 ELSI model

8.1 Background

The ELSI model is designed to study pensions and evaluate pension policy reforms. It is classified among other European pension microsimulation models in Dekkers and van den Bosch (2016). In international discussions and classifications of models, the ELSI model is ascribed to semi-public institutions (such as the Finnish Centre for Pensions). In many countries, microsimulation models are ascribed to statistics institutions or government agencies.

The idea of investing in the construction of a dynamic microsimulation model of the pension system was longstanding at the Finnish Centre for Pensions as it has already developed an established semi-aggregate long-term macro model, the long-term projection (LTP) model. In fact, making pension projections is part of the statutory tasks of the Finnish Centre for Pensions. (Act on the Finnish Centre for Pensions, Chapter 2). For a long time, the decision-makers of Finnish pension policy (interest groups, labor organizations, pension providers and government) have asked for an evaluation of pension benefits and pension policy reforms from a distributional point of view.

In the early 2010s, the Finnish Centre for Pensions took preliminary steps in assess possible models. Inspiration on life course modeling was drawn from the IMA congresses and already existing dynamic models. Nordic models, especially the Swedish SESIM (Flood 2008) and the Norwegian MOSART (Fredriksen 1998), were studied. The Belgian LIAM2 modeling toolbox by Gijs Dekkers (de Menten et al. 2014) and the Dutch SADNAP model by Jan-Maarten van Sonsbeek (van Sonsbeek 2011a, b) were also studied in more detail. Eventually, in 2012, the Finnish Centre for Pensions set the task of constructing an in-house dynamic microsimulation model.

Good practices of successful model building outline the following topics emphasized by van Sonsbeek: a multidisciplinary team is perquisite, the model should me modular and simple at the beginning, and state-of-the-art data must be available. As these conditions were met, the construction of the model could start. The programming began in early 2013. The first policy analyzes were completed in 2014 (Tikanmäki et al. 2014; sub-study I). All models evolve and need some development, the ELSI model included, so it has already been updated in 2016 and 2018.

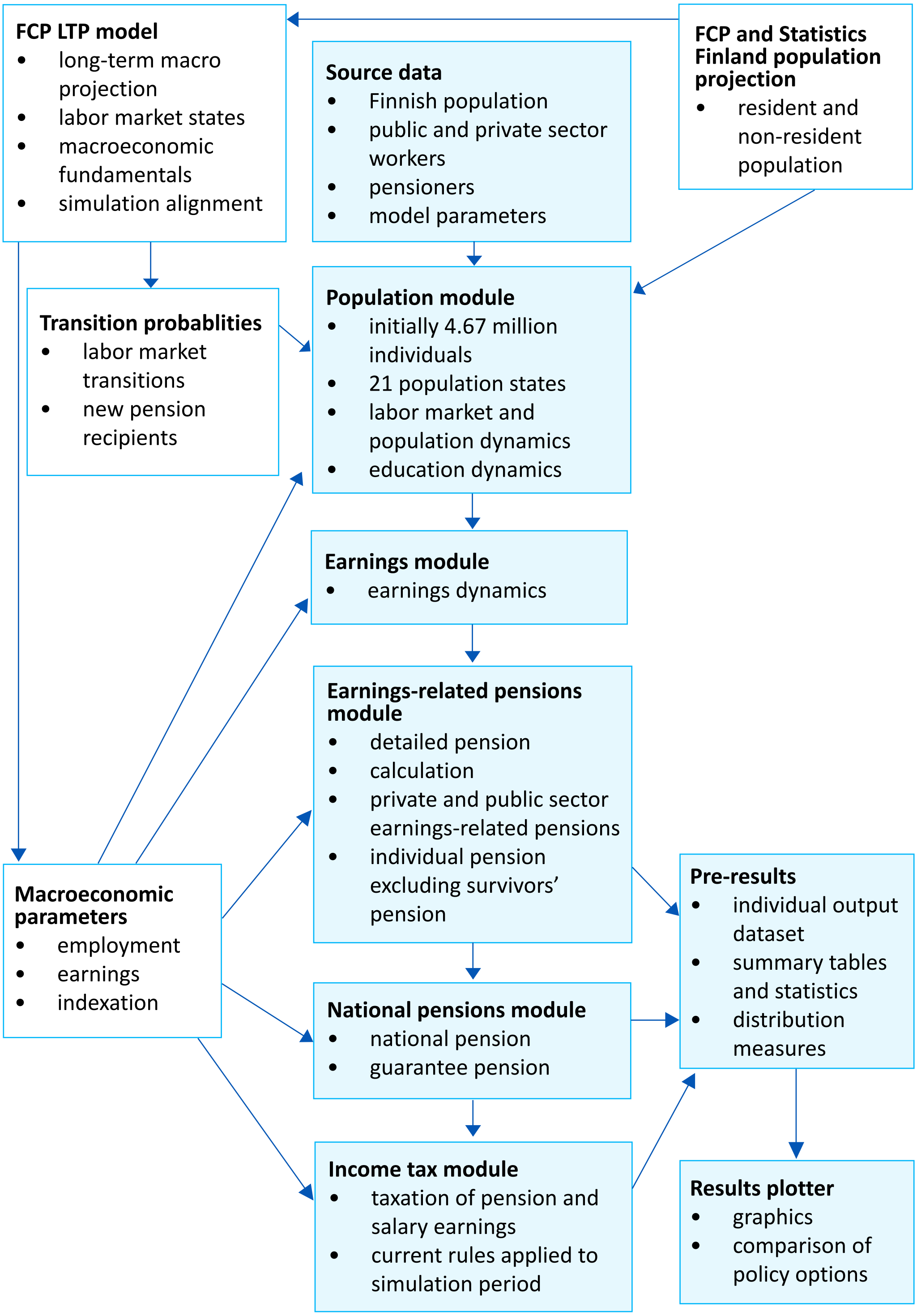
8.2 Notes on data and population

The ELSI model is based on administrative register data and research data from Statistics Finland. The Finnish Centre for Pensions' research register includes a wide range of individual-level information on the active part of the life course, as well as on time in retirement.

Information on individual properties, wages and employment contracts, certain social security benefits, statutory pensions (received in one's own right) form the core data used in parameter evaluation and simulation. The individual-level information on education is from Statistics Finland's registers. In addition, some auxiliary macroeconomic information on employment and indexation are drawn from macroeconomic projections and statistics. The population projection is from Statistics Finland, but it is re-calculated at the Finnish Centre for Pensions for various modeling tasks of the institute.

A general outline of the current (2019) ELSI model is presented in Figure 4. The dark boxes are the modules in which all simulations are located. The light boxes contain auxiliary information sources.

Figure 4.
Modular structure of the ELSI model



The yearly information covers 36 outcomes, which are the core information during the simulation and the data flow through modules. This information is stored in data files and it is the most important individual-specific information of the model. The individual-specific information in ELSI is described in Appendix 1. The information in variables 1–7 and 10–18 is defined in the early modules: source data, population and earnings. Information in variables 8, 9 and 19–23 is calculated in the earnings-related pensions module. Information in variables 24–34 is simulated and calculated in the national pensions module. Information in variables 35–36 is calculated in the income tax module. The monetary variables are defined as euros per year, except in pension calculation, where the variables are in euros per month.

As Finnish statutory pensions cover the total population, the model population should also cover all people participating in the pension systems and the labor markets. In the ELSI model, the base population is people with social insurance in the Finnish statutory pension system. The population data is drawn from the population registers of the Finnish Centre for Pensions. The register material is in a data warehouse in an SAS environment. Initially, the material is a collection of SAS tables, but the final source data is transformed into an APL environment. ELSI is made using Dyalog APL programming language. The APL programs are optimized, allowing for a large sample size (total population) and a fast runtime.

8.3 Model structure and functionalities

ELSI has a modular structure, which is based on restricted rule-based functionalities within modules. The core of the simulation is in five modules: population, earnings, earnings-related pension, national pension and income tax. During simulation, the individual is in the population and labor market state (model state). There are 21 states available (see Appendix 2). The ELSI model states are roughly similar to the Finnish Centre for Pensions' long-term model (LTP model). In terms of comparing results from the macro model and the micro model, it is sensible to have similar populations, macroeconomic parameters and labor market states.

The fundamental concept, in terms of model transitions, is cross-sectional ageing. It means that, in a given year, all individuals are simultaneously simulated to a population state. In the following year, they take another simulated step to another (or the same) population state, and so forth, until the end of simulation. Individual life events occur within the yearly model transitions. In other words, ELSI is a discrete time model (cf. Li et al. 2014, 317–318). Technically, the transitions occur at the end of June, which has implications on monetary measures when transition takes place.

The modular structure together with cross-sectional ageing has pros and cons. The model designed this way is fairly easy to implement and maintain, but there is no feedback from a module to a previous one. For example, employment and earnings are simulated in separate modules, which means that previous earnings developments cannot influence future employment (probability).

In ELSI, the individual transitions from one population or labor market state to another are defined by external transition probability matrices together with stochastic processes. The transition matrix from one simulation year to the next is 2*111*4*21*21 (by gender, age, level of education, model state of year t, year t+1, and so forth.). The model takes transition probabilities as ex ante information, but they are adjusted yearly in the model to meet the needs of ELSI, to correct the Monte Carlo error and to align results with a long-term prediction model. Defining the transition probabilities can be understood as part of the parametrization of the microsimulation model.

The population module is the first simulation module of ELSI. This module simulates population and labor market states (21 states) and the educational level of individuals. Furthermore, the module simulates residence (in Finland or abroad). The population module evaluates the simulation period on a yearly basis. All factors affecting the model state are evaluated completely before simulating states the following year.

The population module also simulates long-term unemployment benefits and study times in more detail, as accumulated pension rights from studies leading to a degree and from long-term unemployment spells are counted in the module. The duration of working life is simulated in the population and earnings modules.

During the simulation period, changes in population are driven by immigration, emigration, mortality and new cohorts entering the population. These factors are aligned with the population projection. Every year a new cohort of individuals enters the population at age 18. The cohort is then divided into population and labor market states. Emigration is based on a simple gender and age-specific Markov process. The yearly number of immigrants follows a function of age and gender. Immigration has two classes: returnees and new immigrants. The returnees are already in the model population and face the Markov probability to return to Finland from abroad. New immigrants are divided into population and labor market states as they are generated into the model population as new individuals.

In the following the abbreviations are described in Appendix 2. There are several states for employed individuals (AK, A1 and AK), various states for the retired (VE, TK, T1, TU, Ov, Ot, KV and KT) and states in which employment and retirement is combined (OV, OT and OE).

The ELSI model aims to target the total Finnish population and those emigrating from and immigrating to the country. All individuals are situated in one, and only one, population and labor market state at a time. Not all individuals, all the time, participate in the labor market, get pensions or social security benefits. Deceased people and those living abroad, who have pension entitlements in Finland are included in the model population over the whole simulation period. Unemployment (earnings-related and state-paid) is in state AT. States Ak, SY and Mu are specified to complete the population states. The flow of individuals between the inactive state Ak to labor market states is common. State SY includes those who are permanently outside the labor force. The residual state Mu typically includes young people and immigrants still searching for their place in the labor market. During the simulation, less than 1.5 per cent of the population is in state SY, and less than 2 per cent in the residual state Mu.

There are alignment and validation techniques used to guarantee that state transitions in population and labor markets in the ELSI model follow those in the LTP model.

The earnings module takes the simulated population module output data as given. The aim of the earnings module is to simulate wages and earnings rights from various short-term social security benefits. The module simulates earnings rights from spells of getting child benefits, sickness and certain other social security benefits.

The model states allocate time in one-year fragments. As that would be too coarse a measure for some social security benefits, the model states alone cannot be used to simulate all social security spells. The earnings module simulates additional short-term spells and earnings rights related to them. The main reason for this arrangement is pension calculation and information needed for that, but it also makes sense from the point of view of measuring working life.

The earnings concept in the ELSI model is the earnings for which earnings-related pensions accrue. The earnings concept centers around wage or salary earnings, as the model does not simulate new entrepreneurship. The ELSI model focuses on simulating pension distribution. That is why the occupational structure of the labor force is not that important, as long as both employment and earnings distributions evolve in a reasonable way. Earnings from self-employment (confirmed earnings) are initially included in the model, but they diminish during simulation. This means that, over the simulation period, the labor force consists increasingly more of wage earners. The earnings dynamics is described in more detail in a later in this section.

After the reform in 2005, earnings-related pensions started to accrue from several social security benefits, as well. In fact, there were a total of 34 benefits for which pension accrued in 2012. From the point of view of the final pension, the most significant 16 of these benefits are grouped into four more general categories: child care benefits, earnings-related unemployment benefits, state-paid special benefits (based on VEKL) and other social security benefits (see Appendix 3). More information about the treatment of social security spells is given in a later in this section.

Finally, an emerging feature of the Finnish labor markets is working in retirement. Working is common while drawing a partial old-age pension, a partial disability pension and a part-time pension (states OV, OT and OE) and possible while drawing an old-age pension or a full disability pension (VE, TK and T1). Small amounts of wages are simulated for a share of individuals in these states. The part-time pension will fade out from the pension stock by the 2020s as its recipients will start to receive an old-age pension.

After the earnings simulation, the next step is to calculate the pensions. The earnings-related pensions module takes the simulated results from the pervious modules as given. When the population state indicates a pension (whole range of pension benefits), the deterministic earnings-related module counts the amount of pension according to the rules of the Employees Pensions Act (TyEL). Pension calculation is fairly detailed, and it uses all the information available that affects the pension amount. If an individual switches from one pension benefit to another, the pension amount is re-calculated. There are indexation rules applied to accrued pension rights and the pension stock. New pension accrues for work in retirement. It is added to the old-age pension at a later stage.

Technically, the earnings-related pensions module includes age-specific accruals, the flexible retirement age, the life expectancy coefficient, the correct pensionable wage, indexation, the increase for late retirement, the calculation of projected pension components and accrual rules for wages and social security benefits. The pension calculation takes account of changes in pension rules, especially those of the 2005 and 2017 pension reforms.

The module counts pensions at a yearly level, considering the fact that new pensions are drawn at the end of June and pensions end on the same date. This guarantees that there is no bias in the overall pension expenditure. The module eventually saves the yearly amount of pension to the output data at a monthly level. These monthly pensions are added to the individual yearly pension expenditure, which takes account of pension months (time) in a year. The module also counts yearly pension contributions (employee and employer shares together).

The national pensions module is like the earnings-related pensions module. It counts individual national and guarantee pensions on a detailed level. The national pensions module is deterministic, with some exceptions. The most important reduction in the national pension is that 50 per cent of the earnings-related pension is deducted from the full amount of the national pension. As the national pension depends on type of housing (living alone or with a partner), the housing status is also simulated in the module. Other smaller elements reducingthe national pensions are survivors’ pensions and pensions paid from foreign countries. Certain imputation methods are applied, providing the module with information on the generality of these smaller reductions. The module calculates old-age (VE and KV) and full disability (TK, T1 and KT) pensions as national pensions. Finally, the module calculates the guarantee pension for those entitled to it. (See Sihvonen 2015 for a more detailed description of national pensions).

Technically, the module includes the following details: reduction of national pension based on a lump-sum increase of earnings-related pension, type of housing, pension paid from abroad, SOLITA benefits and supplementary pensions, time of residence in Finland (and other countries with social security arrangements), different limits, increase for late retirement (earnings-related pension), reduction for early retirement (earnings-related pension), earnings-related pension accrual after retirement.

Like the earnings-related module, the national pensions module saves the yearly amount of pension to the output data at a monthly level. These monthly pensions are added to the individual yearly pension expenditure (national pension and guarantee pension), which accounts for pension months (time) in a year.

The final module with substantial simulation is the taxation module. The module uses previously simulated data and applies simplified tax rules (e.g., average community tax and no church tax) to wage and pension earnings. The taxation is targeted to the aforementioned earnings, and taxation on social security benefits is not part of the module. The taxation is also simplified in terms of reality as the ELSI model omits other earnings and incomes (e.g., capital income) as well as a number of allowances (e.g., domestic help allowance). As a rule, the taxation of workers and pensioners changes every year depending on Government policies; thus any expectations or predictions of future tax rules are uncertain and policy-related. Therefore, over the simulation period, only current tax rules are assumed to be fixed with wage indexation of limits and parameters. (For more details on outcomes in ELSI regarding income taxation, see Knuuti and Ritola 2019, and on tax rules, see Sihvonen 2015.)

The ELSI model produces several outputs. In the first stage, the individual level simulation results are stored in an output dataset. This dataset includes all individuals in the simulation, that is, those 4.67 million individuals initially incorporated in 2012 and the additional 5.2 million individuals whose life courses are simulated during the simulation period. In total, the output dataset includes 9.87 million individuals with yearly information on a number of factors (see Appendix 1). These simulation outcomes can be collected and processed in the pre-results module. In fact, this output data is very interesting as it includes all the individual level heterogeneity generated by the microsimulation model. Based on output dataset summary tables, distributional measures and other statistics can be produced. This summary information is usually classified by ex ante classifiers such as age, gender or level of education. In the second stage, the results are plotted at a more aggregate level. Some 150 figures are drawn to illustrate different aspects of the simulation.

The functionalities described thus far are the core of ELSI. In addition to these, there are data-sensitive and parameter-sensitive stages in the simulation that require attention.

Parameter adjustment is an essential part of microsimulation modeling. There are two steps in the ELSI model parametrization. First, the ELSI source data is collected for five years (2008–2012, empirical period) before the actual simulation begins. The simulation period is 2012–2085. The population level empirical data offers possibilities to estimate transition probabilities and evaluate population shares in various model states. Second, the long-term macro model provides guidelines to model transitions over the simulation period. This is especially important regarding employment and pension transitions.

The long-term macro projection is also an important object considering the monetary outcomes of the simulation. The ELSI model simulates employment and wages, which together add up to yearly wage sums. Equally important are the simulated pensions, which add up to pension expenditure. These key monetary outcomes are aligned to macro projection via model parameters and assumptions. The alignment process aims to target the monetary outcome trajectories over the simulation period to long-term macro projection.

The yearly time-concept of the model transitions is problematic, as real-life phenomena do not have fixed time-horizons. The yearly time concept is problematic in two ways. First, short social-security spells, lasting less than one year, are easily over-estimated. Second, phenomena like employment transitions, earnings development and educational transitions, need a longer time dimension. Next, I will explain in more detail some solutions within and between modules that alleviate these evident problems so that credible heterogeneity into individual life courses and the model results align with reality.

Social security spells

Pension accrues for social security benefits under different law-based rules that are taken into account in the earnings module. The earnings that the benefits are based on is the previously simulated wage. The module simulates the part of the earnings for which pension accrues. It is important to notice that ELSI does not cover all social security benefits available to an individual, merely those for which pension accrues. Further, it is important to notice that the social security benefits in ELSI cover the amounts paid to the individual, not to the employer.

The simulation of the social security benefit differs somewhat depending on the benefit. Long-term unemployment benefits are taken into account in labor market states AL and AT, but also in the transition probabilities between model states and wages. Short-term unemployment spells and sickness spells and job alternation leaves (earnings rights) are simulated aside from population and labor market states with more simplified age and gender-specific techniques.

The population module simulates study times in more detail, as earnings rights (VEKL) from studies leading to a degree are counted in the module. Besides pension accrual, study time has no effect on, for example, the model states or working life. Education dynamics drives the progression of education, which leads to different probabilities in some parts of the model.

The child benefits are based on age-specific fertility regardless of population or labor market states. The mother earns the maximum duration or days of earnings-related maternal and parental benefits. The father also receives his share of the earnings-related family benefit. Parents are entitled to a home-care allowance for one to two years after a child is born. About 40 per cent of mothers in states A1, A2, AK and AT draw a home-care allowance. All women and half of the men in states AK and Mu draw a home-care allowance. Although different rules apply to men and women and to different model states, the overall fertility in ELSI is not dependent on model states.

The duration of working life is counted in the population and earnings modules. It is initially based on yearlong model states (A1, A2, AK, OE, OT, OV and AT), but short career brakes (less than year) are taken into account in working life.

State dynamics

The majority of model transitions are clear-cut, such as retiring from work (e.g., AK to VE), where there is no need to consider longer transition times or implications on labor market probabilities or wages. Some transitions require more time or memory for adjustment. One example is the transition from unemployment to full-time employment (AT to A1), where the path to employment flows through various states (A1, A2 and AK). The states have different transition probabilities indicating different work attachment (gradually strengthening). Another example is the transition to a full disability pension, which usually flows through from states AS (sickness benefits preceding full disability pension) to T1 and finally TK.

The fundamental reason to use gradual transitions is the fact that the Markov process used in many places of the model lacks memory. Gradual transitions present memory and reality into Markov transitions that otherwise lack memory.

Education dynamics

As part of the population module, individual changes in the level of education are simulated. The level of education is part of the model transition matrices and a key factor affecting the model results. The initial population includes information on highest education over an empirical period. The initial information on education is from Statistics Finland. In ELSI, education is classified as follows:

1. Primary school
2. High school
3. Vocational school
4. University of applied sciences and Lower university (Bachelor's degree)
5. Higher university (Masters' degree or Doctoral degree, PhD)

The empirical period is used to estimate age, gender and degree-based transition probabilities from one level to the further levels. There is a simple Markov process to model individual education transitions. These transitions can only lead to a higher level of education or degree. The education dynamics are somewhat simplified, as they do not take into account, for example, the duration of education or employment. Although education is taken into account when the accrued earnings-related pension is counted, it plays a greater role in model transitions related to, for example, wages, employment and retirement.

Earnings dynamics

The earnings module simulates wages based on individual information. Contrary to the LTP model, which includes deterministic age-, gender- and employer-sector (pension act)-specific earnings profiles, the microsimulation uses education, model state, previous wages (for two years) and individual variation to simulate wages. The wage simulation has both dynamic and static models on different populations (see Tarvainen 2017). For individuals already in the model population, a dynamic autoregressive statistical model is used. The dynamic wage model uses the following formula:

$$\ln\ Y_t = \beta_0 + \beta_1\, ln\, Y_{t-1} + \beta_2\, ln\, Y_{t-2} + \boldsymbol{\beta x_{it}} + u_i + \epsilon_{it}\ ,$$

where Y is the wages at fixed prices, β_0 is constant, β_1 and β_2 are coefficients for previous earnings, $\boldsymbol{x}$ is a vector of explanatory variables (age, yearly level of education and model state), $\boldsymbol{\beta}$ is a vector of coefficients related to explanatory variables, u_i are the individual factors and ϵ_{it} are the error components. The range of ages is 18 to 70 years, education has five levels, and the model states are those indicating full-time employment (states A1, A2 and AK). The model is also evaluated for other model states, in case individuals get employed and a wage estimate is needed. In all cases, genders are estimated separately. The individual factors and error components are expected to be independent. The individual factor is expected to follow the asymmetric Laplace distribution:

$$u_i \sim f(0, \lambda, \kappa) = \frac{\lambda}{\kappa - 1/\kappa} \begin{cases} \exp\left((\lambda/\kappa)(x-m)\right)\, if\ x < m \\ \exp\big((-\lambda\kappa)(x-m)\big)\, if\ x\ \geq m' \end{cases}$$

where the scale parameter λ and the asymmetric parameter κ are gender specific. Error components are expected to follow the normal distribution:

$$\epsilon_{it} \sim N(0, \sigma_\epsilon^2).$$

Another form of the stochastic model is used for newly generated individuals in the model population. New individuals consist of a new young cohort (18-year-olds) entering the labor markets and new immigrants. The static wage model for this group is defined as:

$$\ln Y_t = \tilde{\beta}_0 + \tilde{\boldsymbol{\beta}} \boldsymbol{x_{it}} + \tilde{u}_i + \tilde{\epsilon}_{it} ,$$

where the individual factors and error components are expected to be independent and follow the asymmetric Laplace distribution, and the normal distribution follows the dynamic model. Once individuals are generated into the model population and the initial wages are simulated with the static model, the following wages are simulated with the dynamic model.

8.4 Discussion

ELSI is a stochastic dynamic microsimulation model of the Finnish pension system, without optimization behavior of individuals (cf. Klevmarken 1997). The model has a strong link to the semi-aggregated long-term macro model (LTP model). The similar fundamental structures in the macro and the micro model are there for a reason. The strong link from the macro model to the microsimulation model ensures that both models measure the same population and transitions. This makes it easier to validate the results of the microsimulation. However, the deterministic nature of the ELSI model, with individual behavior driven by transition probabilities, makes the model cumbersome to adjust for behavioral responses.

A major concern in microsimulation is how to replicate the individual heterogeneity we see in empirical data or the real world in wages and pensions. A nice feature of ELSI is the use of stochastic processes and stepwise population and labor market states to introduce individual heterogeneity into the simulation in a realistic way, as revealed in article three of this thesis. Heterogeneity in microsimulation can be difficult to master, because dynamic microsimulation is a complex exercise, and there is a latent concern about sub-populations that do not react to model assumptions as expected. The parametrization and alignment to the macro model and statistics is always a challenging stage of modeling.
To illustrate microsimulation results and to locate possible problems in parametrization, one useful tool is trajectory analysis. This technique was introduced to the microsimulation context, using ELSI as an example, in sub-study IV of this thesis.

Thus far, ELSI has been used to analyze the Finnish pension reform of 2017 (e.g., Tikanmäki et al. 2015). Along with administrative studies (Tikanmäki et al. 2017a), microsimulation results are now also part of the long-term projection of the Finnish pension system (Tikanmäki et al. 2017b; Tikanmäki et al. 2019). The model was introduced to the international microsimulation community in sub-study III, demonstrating the core measures of the model: distribution of pensions, duration of retirement, duration of working life, replacement ratio and Gini-coefficient for the retired. The long simulation period allows for a detailed study of the life course. In fact, the article introduced a new measure called partition of the life course to illustrate the evolution of different phases of life course. The measure can be used to evaluate many what if questions in pension evaluation.

9 Discussion and conclusion

This thesis is a collection of scientific work conducted for more than a decade. Most of the time has been used preparing data and measures in a form that can be used in scientific research. The Finnish Centre for Pensions has wide administrative registers that are now available for researchers. They form a flexible research register that can be used to study pension policy in a much larger scale than what is possible in the sub-studies presented here. Today, external researchers and partners, such as economic research institutes and universities, use the research register in pension evaluation on a regular basis. The data has also inspired new research, as the Finnish register material attracts international researchers. International collaboration is an important way to learn from other countries and about other pension schemes.

Although the research register contains versatile information on working life and the life course in general, the material is not completely risk-free. As discussed in Chapter 5 of this thesis, some issues require more attention. First, in the future, more focus needs to be put on the measure or indicator of working life. Currently working life is based on employment contracts. Sometimes short-term social security benefit periods overlap with working hours. Thus, some social security benefit periods are included in working life while others are not. There is a need to define alternative indicators of working life. Second, the concept of earnings is changing, as the technical registration of employment contracts is switching from employment contracts to monthly registration of ongoing employment contracts. In administrative registers, the employment contract is fading away and is mixed in the register technique of employers. This is a vexatious thing from a research point of view – we can no longer measure employment contracts in a completely reliable way. Third, the new European-wide rules on data privacy (GDPR), issued by the European Commission, and national laws may limit the data contents and storage in the future. Researchers may find it difficult to get reliable and up-to-date material, which is a strange development in times when knowledge-based decision-making is very important.

The future of register-based pension research looks promising. There are solid processes to maintain current registers. Furthermore, new research possibilities are emerging, as the nationwide Incomes Register becomes available for pension evaluation.

Besides introducing register sources, this thesis focuses on three aims: introducing new evaluation techniques, deepening the knowledge of some substantial topics and introducing new measures. I hope the substantial results of this thesis fill some of the gaps in our knowledge of pension evaluation and the sociopolitical discussion conducted in Finland and abroad. Some lessons and policy recommendations can be drawn.

One, sub-study I emphasizes the labor market attachment of young men. The results show that a vast majority (88%) of the young switch from education to employment painlessly; in other words, they follow a normal biography in sociological terms. We named this group HES. The statistical method we used revealed that nearly 12 per cent of the males in a cohort experienced persistent problems with labor market attachment. We called this group LES.

The prolonged weak labor market attachment was mainly due to unemployment spells and other social security spells. A small but (in other studies) already detected sub-group had a background of some chronic illness and long spells of occupational rehabilitation. Today this is not surprising as the lives of young men, and reasons for permanent disability in particular, are at the center of public debate.

The study also showed that the possible scar of weak labor market attachment on working life is deep. The men in the HES group who followed a normal life course had more solid working lives. At the end of the study period, at age 26, the duration of working life for these men was 5.1 years on average, whereas for the men in the LES group, the average duration was 1.3 years. If early career problems predict later career problems, the obvious implication is that the pension accrual falls short for these young men, resulting in a further polarization of pensions in old age.

Working with clustering methods and different labor market data evokes aspirations that longitudinal statistical methods (e.g., trajectory analysis) and new data possibilities could be further used in employment administration (e.g., by the Finnish Ministry of Economic Affairs and Employment). That is why I would recommend using these methods and measures to identify policy-relevant groups of the population. I recognize that the labor administration and Kela in general, and particularly municipalities, do their best in helping the young in need, but the possibilities presented by current register data, IT technology and statistical methods could help authorities to identify the young who face persistent labor market risk. These groups of young people revealed by statistical methods could be candidates for policy experiments.

Two, following a concern in the statutory pension system about low pensions of the self-employed, sub-study II draws attention to the deficit of statutory pension security of the self-employed. They are a heterogenous group, but results focusing on the traditional self-employed and small business owners are illuminating. Sub-study II reveals the behavioral differences in different groups of self-employed – those with contributions at a proper level and those with persistently low contributions. The main result of the sub-study is that about 84 per cent of the self-employed paid too little for their pension security, i.e., their confirmed income was set below their true taxable income. A smaller share (12%) selected a confirmed income that equaled their taxable income. A small share (4%) selected a confirmed income that was persistently above their taxable income. The results are interesting in that they illustrate different behaviors in a statutory system. For many reasons, the majority of the self-employed underinsure themselves against several possible risks. On the other hand, a small share persistently pay maximum contributions, i.e., overinsure themselves, which reveals totally different incentives. More research on this topic is certainly needed.

Sub-study II, along with the reports Salonen (ed., 2015) and STM (2019), show the need for a reform of the pension scheme for the self-employed. Focusing the reform on switching the contribution base from a subjective setting of the confirmed income to true earnings would be a step forward. There are pros and cons in using the taxable income. They should be evaluated. Broadly understood, it is a question of freedom for the self-employed to select their level of statutory pension security in a compulsory system. Clearly, the self-employed cherish this subjective freedom, but there are clear problems with that. International actors like the OECD have also noted this problem with the Finnish scheme and have expressed their concern (OECD 2019b). There is another side of the coin, namely tax planning, which is easier and more relevant for the self-employed than for traditional employees. The self-employed who own a business can and most likely do shift income from wages to dividends if the system permits it (see Harju and Matikka 2012). This could imply (under current rules) that some groups of self-employed would select to draw more dividends if the pension contribution base were taxable wages. From a wider perspective, however, a sound policy reform should find a balance between decent old-age benefits and the financing of the scheme. The question about the contribution base affects certainly not only pensions: many basic social security benefits by Kela are based on the confirmed income, as are earnings-related unemployment benefits. In principle, the question is how to manage fundamental risks within statutory social security. Finally, the reform should increase the institutional trust to allow all classes of self-employed, even the modern gig workers (i.e., independent contractors, online platform workers, contract firm workers, on-call workers and temporary workers) to be covered by earnings-related pension security.

This discussion leads to a completely different issue in the Finnish statutory pension system, namely the strongly subsidized schemes of the self-employed, farmers and sailors.

In terms of the overall pension system, these schemes are not very significant financially, but in terms of the fundaments of a compulsory earnings-related pay-as-you-go system, it is questionable to foster schemes with strong financial support from the state budget. Although it is a question of securing national supply and enterprise policy, the scheme for the self-employed should still be reformed so that it follows roughly the same rules as those for conventional employees.

Three, the policy implications of sub-study III about microsimulating pension reform outcomes are that the majority of the people capable to continue working should do so for as long as they can. This is a precondition for a financially stable and fair pension system. Currently there is an apparent imbalance in the composition of the life course – the active (employed) share of the lifespan is too small compared to the inactive share (e.g., time on pension and time on unpaid periods). The pension reforms of 2005 and 2017 aimed to defer retirement and thus correct this imbalance. Microsimulation results show the evolution of pension distribution into a distant future. For the first time, the results showed the role of education for future pensions. It highlighted that many life course issues such as earnings, duration of working life and retirement choices, are education dependent.

From an actuarial point of view, postponing retirement is one obvious solution to finance the concerns. From a practical point of view, however, raising the retirement ages pose some problems. Currently, we are (all) somewhat unprepared to comprehend what the increase of the retirement age and the population projections of actuarial calculations predict. Time will show whether policies will change, that is, whether people will accept the rising retirement ages. The actuarial or macroeconomic rationale is fully understandable, but it is doubtful whether the present rules will last infinitely.

Sub-study IV, on applying trajectory analysis on simulated data, is primarily targeted at the microsimulation community but can be useful to anyone interested in life course phenomena, such as employment, retirement trajectories, education dynamics, and simulation techniques. Chapter 8 gave a thorough introduction of the ELSI model, with an emphasis on the simulated behavioral heterogeneity and its sources. This background information is probably useful in evaluating realities generated by a microsimulation model.

Some new measures and concepts were introduced in the sub-studies and chapters of this thesis. I challenge the traditional measure NEET rate. It is a cross-sectional measure whereas the problem, from the individual's perspective, is longitudinal. Based on LES groups, a new measure could be proposed – deep NEET – which would depict the naturally longitudinal nature of labor market attachment. If corresponding data is available, similar analysis and measures could be used and confirmed by others. Naturally, further analysis and confirmation are needed before a new statistical concept can be introduced.

Traditional measures of life course – such as length of working life or employment rate – could be complemented by a new measure: partition of the life course. It is a simple measure illustrating different stages of the life course. The measure was introduced in sub-study III and elaborated further in Chapter 5. With dynamic microsimulation data, the share of time in employment and time on a pension can be evaluated within different sub-populations.

Relating to sub-populations, there is an everlasting question in pension evaluation, namely: how to illustrate outcomes or what covariates or factors to use in statistical models? Gender, education or occupation are much used ex ante classifiers in pension research. Trajectory analysis is promising in providing alternative or complementary classification rules which are based on latent groups. The data-based classification rules could be useful in illustrating some topics of pension evaluation.

Finally, I hope the group-based methods (trajectory analysis, the new measures of life course, and the data-based classification rules) will establish themselves in life course studies and in pension evaluation. Finite mixture modeling applications in general may also prove to be useful and to bring new insight to classical topics of pension evaluation (e.g., survival analysis). Furthermore, I see no reason why finite mixture modeling techniques or microsimulation could not be used in the finance industry at large, beyond statutory pensions.

REFERENCES

Ackoff, R. (1999). Ackoff's Best. New York: John Wiley & Sons, p. 170–172.

Act on Guarantee Pensions 20.8.2010/703.

Ahola, E., Airio, I. and Nurminen, M. (2014). Mitä eri aineistot kertovat toimeentulotuen käytöstä täyden kansaneläkkeen saajien keskuudessa? Yhteiskuntapolitiikka 79 (2014): 4, p. 455–458. http://urn.fi/URN:NBN:fi-fe2014120852204

Ahola, E., Honkanen, P. and Sirviö, M. (2015). Hallitusohjelma ja eläkeläisten toimeentulo. Kelan työpapereita 77/2015, Helsinki.

Ahonen, K. (2018). Kotitalouden rakenteen vaikutus ikääntyneiden sukupuolten välisiin köyhyysriskieroihin EU15-maissa ja Norjassa.

Ahtokari, R. (1988). Tuntematon vaikuttaja. Työeläkejärjestelmän isä Teivo Pentikäinen. WSOY.

Akerlof, G. (1974). The Market for "Lemons": Qualitative Uncertainty and the Market Mechanism. Quarterly Journal of Economics. Vol 84, p. 488–500.

Anderson, R. and Hicks, C. (2010). Highlights of Contemporary Microsimulation. Social Science Computer Review 29(1) 3–8. DOI: 10.1177/0894439310370084

Ando, A. and Modigliani, F. (1963). The "Life Cycle" Hypothesis of Saving: Aggregate Implications and Tests. The American Economic Review, Vol. 53, No. 1, Part 1 (Mar., 1963), p. 55–84.

Arajärvi, P. (2011). Eläke ja omistusoikeus. In Johanson, J-E., Lassila, J. and Niemelä, H. eds. (2011). Eläkevalta Suomessa. ETLA, Helsinki.

Atkinson, A. and Stiglitz, J. (1980). Lectures on Public Economics. Princeton University Press.

Barker, E., Séguin, J. and Raskin White, H. (2007). Developmental Trajectories of Male Physical Violence and Theft Relations to Neurocognitive Performance. Arch Gen Psychiatry. 2007; 64(5), p. 592–599. doi:10.1001/archpsyc.64.5.592

Barr, N. (1994). Economics of the Welfare State. Second Edition. Oxford University Press.

Barr, N. and Diamond, P. (2008). Reforming Pensions: Principles and Policy Choices, Oxford University Press, Oxford.

Beck, U. (1992). Risk Society. Towards a New Modernity. SAGE Publications, Great Britain.

Beck, U. (2000). The Brave New World of Work. Polity Press, Great Britain.

Benartzi, S. and Thaler, R. (2007). Heuristics and Biases in Retirement Savings Behavior. Journal of Economic Perspectives, 21 (3), p. 81–104.

Benzarti, Y., Harju, J. and Matikka, T. (2019). Does Mandating Social Insurance Affect Entrepreneurial Activity? VATT Working Papers 119.

Berlin, K., Ankney, R. and Rybak, T. (2018) Methods Commentary: Uncovering Unobserved Data Patterns With Latent Variable Mixture Modeling. Journal of Pediatric Psychology, 43(7), 2018, 733–736 doi: 10.1093/jpepsy/jsy042

Beveridge, W. (1942). Social insurance and allied services. Stationery Office Books, London.

Blake, D. (2006). Pension Economics. Pensions Institute, Wiley.

Blanchet, D., Buffeteau, S., Crenner, E., and le Minez, S. (2010). The new Destinie 2 microsimulation model: main characteristics and illustrative results. INSEE, France. https://www.insee.fr/en/statistiques/1380834

Borch, K. (1974). The Mathematical Theory of Insurance, Lexington, Mass. Lexington Books, D.C. Heath and Co.

Borch, K. (1990). Economics of Insurance. Advanced Textbooks in Economics. Elsevier Science Publishers, The Netherlands.

Broer, D. and Lassila, J. (1997). Pension Policies and Public Debt in Dynamic CGE Models. ETLA Series A23, Helsinki.

Burbidge, J., Magee, L. and Robb. L. (1988). Alternative Transformations to Handle Extreme Values of the Dependent Variable. Journal of the American Statistical Association, 83, No. 401, 123–127.

Böhning, D., Seidel, W., Alfó, M., Garel, B., Patilea, V. and Walther, G. (2007). Advances in Mixture Models. Computational Statistics & Data Analysis, 51, p. 5205–5210.

Clarkeburn, H. and Mustajoki, A. (2007). Tutkijan arkipäivän etiikka. Vastapaino, Tampere.

Castles, F. (2004). The future of welfare state. Crisis myths and crisis realities. Oxford University Press.

Constitution of Finland 11.6.1999/731.

Curran, P., Obeidat, K. and Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. Journal of Cognition and Development. Volume 11, 2010 - Issue 2, p. 121–136.

Davis, B., Heathcote, C. and O'Neill, T. (2001). Estimating cohort health expectancies from cross-national surveys of disability. Statistics in Medicine 20, p. 1097–1111.

Dekkers G. (2010). On the impact of indexation and demographic ageing on inequality among pensioners. Validating MIDAS Belgium using a stylized model. Munich Personal RePEc Archive Paper No. 36136. https://mpra.ub.uni-muenchen.de/36136/

Dekkers, G. (2014). What are the Driving forces behind Trends in Inequality among Pensioners? Validating MIDAS Belgium Using a Stylized Model. In Dekkers, G., Keegan, M. and O'Donoghue, C. eds. (2014). New Pathways in Microsimulation. Ashgate, United Kingdom.

Dekkers, G., Buslei, H., Cozzolino, M., Desmet, R., Geyer, J., Hofmann, D., Raitano, M., Steiner, V., Tada, P., Tedeschi, S. and Verschueren, F. (2010). The Flip Side of the Coin: The Consequences of the European Budgetary Projections on the Adequacy of Social Security Pensions. European Journal of Social Security, 12(2) JUNE 2010, p. 94–120.

Dekkers, G. and Van den Bosch, K. (2016). Prospective microsimulation of pensions in European Member States. In Dekkers, G. and Mészáros, J. eds. (2016) Applications of microsimulation modelling.

de Menten, G., Dekkers, G., Bryon, G., Liègeois, P. and O'Donoghue, C. (2014). LIAM2: a New Open Source Development Tool for Discrete-Time Dynamic Microsimulation Models. Journal of Artificial Societies and Social Simulation. The 17(3):9 June.

Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood estimation for incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39, p. 1–38.

Diamond, P. (1965). National debt in a neo-classical growth model. Journal of Political Economy, 55, p. 1126–1150.

Diamond, P. (1997). Insulation of pensions from political risks. In Valdés-Prieto ed. (1997) The Economics of Pensions. Principles, Policies, and International Experience. Cambridge University Press, United States of America.

Diamond, P. and Mirrlees, J. (1978). A Model of Social Insurance with Variable Retirement. Journal of Public Economics 10(3), p. 295–336.

Don, P. and Mickelson, K. (2014). Relationship Satisfaction Trajectories Across the Transition to Parenthood Among Low-Risk Parents. Journal of Marriage and Family, 76(3), p. 677–692.

Duncan, T., Duncan S., Strycker L., Li, F. and Alpert A. (1999). An Introduction to Latent Variable Growth Curve Modeling. Mahwah, NJ: Erlbaum.

Ebbinghaus, B. (2006). Reforming Early Retirement in Europe, Japan and the USA. Oxford Univesity Press.

Edgerton, J., Melnyk, T. and Roberts, L. (2015). An exploratory study of multiple distinct gambling trajectories in emerging adults. Journal of Youth Studies. Vol. 18, No. 6, p. 743–762.

Espy, K., Fang, H., Charak, D., Minich, N. and Taylor, H. (2009). Growth mixture modeling of academic achievement in children of varying birth weight risk. Neuropsychology, 23(4), p. 460–474. https://doi.org/10.1037/a0015676

ETKL. Laki Eläketurvakeskuksesta 19.5.2006/397.

European Commission (2010). European Textbook on Ethics in Research. Directorate-General for Research Science, Economy and Society Science in society. Studies and reports. http://ec.europa.eu/research/science-society/document_library/pdf_06/textbook-on-ethics-report_en.pdf

Eurostat (2018). NEET rate at lowest point for ten years.

Feldstein, M. (1987). Should social security be means tested? Journal of Political Economy, Vol. 95, No. 3, p. 468–484.

Flood, L. (2008). SESIM: A Swedish Micro-Simulation Model. In Klevmarken, A. and Lindgren, B. (editors) Simulating an Ageing Population – A microsimulation approach applied to Sweden. Bingley, United Kingdom: Emerald Group Publishing Ltd, p. 55–84.

Figari, F., Paulus, A. and Sutherland, H. (2015). Microsimulation and Policy Analysis. In Atkinson, A. and Bourguignon, F. eds. (2015). Handbook of Income Distribution, Volume 2, Elsevier-North Holland, p. 2141–2221.

Finnish Advisory Board On Research Integrity, TENK (2012). Responsible conduct of research and procedures for handling allegations of misconduct in Finland.

Flood, L., Jansson, F., Pettersson, T., Pettersson, T., Sundberg, O. and Westerberg, A. (2012). SESIM III - A Swedish dynamic micro simulation model. sesim.org

Fredriksen, D. and Stølen, N. (2007). Model 1: MOSART (Dynamic Cross-Sectional Microsimulation Model). In Gupta, A. and Harding, A. eds. (2007). Modelling Our Future: Population Ageing, Health and Aged Care (International Symposia in Economic Theory and Econometrics, Volume 16) Emerald Group Publishing Limited, p. 433–437.

Friedline, T., Masa, R. and Chowa, G. (2015). Transforming wealth: Using the inverse hyperbolic sine (IHS) and splines to predict youth's math achievement. Social Science Research 49 (2015) 264–287. DOI: 10.1016/j.ssresearch.2014.08.018

Forss, M., Kalimo, E. and Purola, T. (2001). The Insurance Discourse and Social Security. The Geneva Papers on Risk and Insurance Vol. 26 No. 4 (October 2001), p. 517–528.

Forss, M., Kalimo, E. and Purola, T. (2004). Vakuutuksen periaate sosiaaliturvassa. In Hellsten, K. and Helne, T. eds. (2004). Vakuuttava sosiaalivakuutus? Kela.

Fredriksen, D. (1998). Projections of Population, Education, Labour Supply and Public pension benefits: Analyses with the Dynamic Simulation Model MOSART. Statistics Norway, Social and Economic Studies Paper no. 101, Oslo.

Fredriksen, D. and Stølen, N. M. (2014). Gender Aspects of the Norwegian Pension System. In Dekkers, G., Keegan, M. and O'Donoghue, C. eds. (2014). New Pathways in Microsimulation. Ashgate, United Kingdom.

Furlong, A. (2006). Not a very NEET solution: representing problematic labour market transitions among early school-leavers. Work, employment and society 20 (3), p. 553–569.

Giddens, A. (1990). The Consequences of Modernity. Polity Press, Great Britain.

Gomez, R., Skilbeck, C., Thomas, M., and Slatyer, M. (2017). Growth Mixture Modeling of Depression Symptoms Following Traumatic Brain Injury. Frontiers in Psychology. doi: 10.3389/fpsyg.2017.01320

Goodman, L. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, Vol. 61, No. 2 (Aug., 1974), p. 215–231.

Gröhn, J. (2008). Lakisääteisen eläkejärjestelmän etuudet ja rahoitus iän mukaan ja sukupolvittain. Eläketurvakeskuksen keskustelualoitteita 2008:1. http://urn.fi/URN:NBN:fi-fe2015112620062

Haataja, A. (1994). Tulonsiirtojärjestelmät ja mikrosimulaatiomallit. Kansantaloudellinen aikakauskirja – 90.vsk – 4/1994.

Haataja, A., Airio, I., Saarikallio-Torp, M. and Valaste, M. (2016). Laulu 573566 perheestä Lapsiperheet ja perhepolitiikka 2000-luvulla. Kela. Teemakirja 15. Juvenes Print, Tampere.

Hagfors, R., Kajanoja, J. and Kangas, O. (2004). Sosiaalivakuutuksen laajuudesta, taloudesta ja oikeutuksesta. In Hellsten, K. and Helne, T. eds. (2004). Vakuuttava sosiaalivakuutus? Kela.

Hakola, T. and Määttänen, N. (2007). Vuoden 2005 eläkeuudistuksen vaikutus eläkkeelle siirtymiseen ja eläkkeisiin: arviointia stokastisella elinkaarimallilla. Eläketurvakeskuksen tutkimuksia 2007:1. http://urn.fi/URN:ISBN:978-951-691-069-0

Hakola, T. and Määttänen, N. (2009). Pension system, unemployment insurance and employment at older ages in Finland, Prime Minister's Office publications, 2/2009.

Hannikainen, M. and Vauhkonen, J. (2012a). Ansioiden mukaan. Yksityisalojen työeläkkeiden historia. Suomalaisen Kirjallisuuden Seura, Helsinki.

Hannikainen, M. and Vauhkonen, J. (2012b). Earning your Keep: The History of Finnish Earnings-related Pension in the Private Sector. Finnish Literature Society, Helsinki.

Hansen, J., Stephensen, P. and Kristensen, J. (2013). Modeling Household Formation and Housing Demand in Denmark - The Dynamic Microsimulation Model SMILE. DREAM, Danish Rational Economic Agents Model, Copenhagen.

Harding, A. (2007). APPSIM: The Australian Dynamic Population and Policy Microsimulation Model. Presentation to the Technical Workshops of the 1st General Conference of the International Microsimulation Association, Vienna, 22 August 2007.

Harju, J. (2012). Voluntary Pension Savings and Tax Incentives: Evidence from Finland. FinanzArchiv Public Finance Analysis 69(1), January 2012. DOI: 10.1628/001522113X663451

Harju, J. and Matikka T. (2012). Entrepreneurs and income-shifting: Empirical evidence from a Finnish tax reform. Government Institute for Economic Research, Working Papers 43, December 2012.

Haviland, A., Jones, B. and Nagin, D. (2011). Group-based Trajectory Modeling Extended to Account for Nonrandom Participant Attrition. Sociological Methods & Research 2011 40: 367. https://doi.org/10.1177/0049124111400041

Hellwig, M. (1987). Some Recent Developments in the Theory of Competition in Markets with Adverce Selection. European Economic Review. Vol 31, p. 319–325.

Helsingin Sanomat 4.1.1998. Nainen sai ensimmäisenä täyden TEL:n työeläkkeen.

Hilliard, M., Wu, Y., Rausch, J., Dolan, L., and Hood, K. (2013). Predictors of Deteriorations in Diabetes Management and Control in Adolescents With Type 1 Diabetes. Journal of Adolescent Health 52 (2013), p. 28–34. http://dx.doi.org/10.1016/j.jadohealth.2012.05.009

Hinrichs, K. (2000) Elephants on the move. Patterns of public pension reform in OECD countries. European Review. Vol. 8, No. 3, 353–378.

Hipp, J. and Bauer, D. (2006). Local Solutions in the Estimation of Growth Mixture Models. Psychological Methods 2006, Vol. 11, No. 1, 36–53. DOI: 10.1037/1082-989X.11.1.36

Hoem, J. (1970). Point estimation of forces of transition in demographic models. Journal of the Royal Statistical Society Series B 33, p. 275–289.

Honkanen, P. ed. (2010). JUTTA-käsikirja Tulonsiirtojen ja verotuksen mikrosimulointijärjestelmä. Kelan tutkimusosasto, Helsinki.

Honkanen, P. and Tervola, J. (2014). Vero- ja tulonsiirtojärjestelmän vaikutus tulonjakoon Suomessa 1995–2013. Yhteiskuntapolitiikka 79 (2014):3. http://urn.fi/URN:NBN:fi-fe2014061026401

Honkanen, P., Räsänen, T., Simanainen, M. and Valaste, M. (2017). EUROMOD Country Report. Finland 2013–2016.

Hsu, H-C. (2015). Trajectories of multimorbidity and impacts on successful aging. Experimental Gerontology 66, (2015), p. 32–38.

Hynes, K. and Clarkberg, M. (2005). Women’s Employment Patterns During Early Parenthood: A Group-Based Trajectory Analysis. Journal of Marriage and Family, 67, p. 222–239.

Hytti, H. and Nio, I. (2004). Monitoring the employment strategy and the duration of active working life. Kela – The Social Insurance Institution, Finland. Social security and health research: working papers 38/2004.

ILO. Social Security (Minimum Standards) Convention, 1952.

Inakagi, S. (2014). Simulating Policy Alternatives for Public Pensions in Japan. In Dekkers, G., Keegan, M. and O’Donoghue, C. eds. (2014). New Pathways in Microsimulation. Ashgate, United Kingdom.

Jahn, D. and Helmdag, J. (2019). Generosität von Sozialleistungen. In: Obinger, H., and Schmidt, M. G. (eds.), Handbuch Sozialpolitik. Springer.

Jones, B., Nagin, D. and Roeder, K. (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. Sociological Research and Methods, 29, p. 374–393.

Jones, B. and Nagin, D. (2007). Advances in Group-based Trajectory Modeling and an SAS Procedure for Estimating Them. Sociological Research and Methods, 35, p. 542–571.

Jones, B. and Nagin, D. (2013). A Note on a STATA Plugin for Estimating Group-based Trajectory Models, Sociological Methods and Research, 42, p. 608–613.

Jung, T. and Wickrama, K. (2008). An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. Social and Personality Psychology Compass, January 2008. DOI: 10.1111/j.1751-9004.2007.00054.x

Järnefelt, N., Kautto, M., Nurminen, M. and Salonen, J. (2013). Development of length of working lives in the 2000s. Finnish Centre for Pensions, Reports 01/2013. http://urn.fi/URN:ISBN:978-951-691-179-6

Järnefelt, N., and Nurminen, M. (2013). Työllisen ajan odotteet Suomessa 2000-luvulla. In Järnefelt, N., Kautto, M., Nurminen, M. and Salonen, J. (2013). Työurien pituuden kehitys 2000-luvulla. Eläketurvakeskuksen raportteja 01/2013.

Järnefelt, N., Nivalainen, S., Salokangas, S. and Uusitalo, H. (2014). Sosioekonomiset erot: työurat, eläkkeelle siirtyminen ja eläkejärjestelmä. Eläketurvakeskuksen raportteja 01/2014. http://urn.fi/URN:ISBN:978-951-691-194-9

Kautto, M. ed. (2019). Työeläkevarat ja eläkkeiden rahoitus. Eläketurvakeskus. http://urn.fi/URN:978-951-691-299-1

Kautto, M. and Salonen, J. (2013). Työurat ansaintarekisterin valossa. In Järnefelt, N., Kautto, M., Nurminen, M. and Salonen, J. (2013). Työurien pituuden kehitys 2000-luvulla. Eläketurvakeskuksen raportteja 01/2013.

Kiander, J. and Lönnqvist, H. (2002). Hyvinvointivaltio ja talouskasvu. WSOY, Dark Oy, Vantaa.

Klevmarken, A. (1997). Behavioral Modeling in Micro Simulation Models. A Survey. Working Paper, No. 1997:31, Uppsala University, Department of Economics, Uppsala, http://nbn-resolving.de/urn:nbn:se:uu:diva-2400

Klevmarken, A. and Lindgren, B. eds. (2008). Simulating an Ageing Population: A microsimulation approach applied to Sweden. Emerald Group Publishing Ltd. Bingley.

Knuuti, J. and Palomurto, M. (2015). Työtulon ja yritystoiminnan verotettavien tulojen suhde ammatin- jaj liikkeenharjoittajilla. In Salonen, J. ed. (2015). Yrittäjien lakisääteinen eläketurva: työurat, työtulot ja rahoitus. Eläketurvakeskuksen raportteja 10/2015. http://urn.fi/URN:978-951-691-232-8

Knuuti, J. and Ritola, S. (2019). Total pension in Finland: How are earnings-related pensions, national pensions and taxation determined? Finnish Centre for Pensions, Reports 04/2019. http://urn.fi/URN:NBN:fi-fe2019090526951

Kokko, K., Pulkkinen, L., Mesiäinen, P. and Lyyra, A-L. (2008). Trajectories Based on Postcomprehensive and Higher Education: Their Correlates and Antecedents. Journal of Social Issues, 64 (1), p. 59–76.

Korkman, S., Lassila, J., Määttänen, N. and Valkonen, T. (2007). Hyvinvointivaltion rahoitus – Riittävätkö rahat, kuka maksaa? Elinkeinoelämän Tutkimuslaitos ETLA B230, Yliopistopaino, Helsinki.

Korpiluoma, R., Alaluusua, A., Heinonen, K., Kaarre, S., Kekäläinen, J., Kouvonen, K., Lilius, S., Mustonen, P., Perälehto-Virkkala, A., Takanen, M. and Tuomikoski, J. (2006). Työeläke. Finva. Finreklama Oy, Sulkava.

Koskenvuo, K., Ahola, E. and Kalliomaa-Puha, L. (2014). Perhe-eläkkeiden mahdollisten uudistamisvaihtoehtojenhyvinvointivaikutukset. In Hietaniemi, M. & Ritola S. (eds.) (2014). Näkökulmia perhe-eläkkeen kehittämistarpeisiin. Eläketurvakeskus, Eläketurvakeskuksen raportteja 04, 2014: s. 111–126. http://urn.fi/URN:ISBN:978-951-691-198-7

Kotamäki, M. and Kärkkäinen, O. (2017). Mistä lääke eläkeläisköyhyyteen?. Yhteiskuntapoitiikka 82 (2017):5., p. 586–596. http://urn.fi/URN:NBN:fi-fe201703162250

Kotamäki, M., Mattila, J. and Tervola, J. (2018). Distributional Impacts of Behavioral Effects – Ex-Ante Evaluation of the 2017 Unemployment Insurance Reform in Finland. International Journal of Microsimulation (2018) 11(2), p. 146–168.

Kuitto, K., Salonen, J. and Helmdag, J. (2019a). Perhevapaat työurariskinä? Trajektorianalyysi nuorten aikuisten työurapoluista. Yhteiskuntapolitiikka 84 (2019):2. http://urn.fi/URN:NBN:fi-fe2019041011828

Kuitto, K., Salonen, J. and Helmdag, J. (2019b). Gender Inequalities in Early Career Trajectories and Parental Leaves: Evidence from a Nordic Welfare State. Social Sciences 2019, 8, 253. doi:10.3390/socsci8090253

Kuivalainen, S., Rantala, J., Ahonen, K., Kuitto, K. and Liisa-Maria Palomäki eds. (2017). Eläkkeet ja eläkeläisten toimeentulo 1995–2015. Eläketurvakeskuksen tutkimuksia 01/2017. http://urn.fi/URN:978-951-691-261-8

Kurvinen, A., Jolkkonen, A., Koistinen, P., Lipiäinen, L., Nummi, T. and Virtanen, P. (2016). Työpaikan menetys työuran loppuvaiheessa: yli 45-vuotiaiden irtisanottujen työllisyysurat ja eläkkeelle siirtyminen. Eläketurvakeskuksen tutkimuksia 02/2016. http://urn.fi/URN:978-951-691-240-3

Kurvinen, A., Jolkkonen, A., Koistinen, P., Lipiäinen, L., Nummi, T. and Virtanen, P. (2018). Työn menetys työuran loppuvaiheessa Tutkimus 45–60-vuotiaana rakennemuutoksessa työnsä menettäneiden työllisyysurista ja riskistä päätyä työttömäksi tai työvoiman ulkopuolelle. Yhteiskuntapolitiikka 83 (2018): 5–6.

Kuula, S. (1997). Verohallituksen henkilöverotuksen suunnittelumalli. Verotus 47(1997):2, p. 205–207.

Kyyrä, T. and Paukkeri, T. (2015). Työkyvyttömyyseläkkeiden maksuluokkamallin kannustinvaikutukset. Eläketurvakeskuksen raportteja 07/2015. http://urn.fi/URN:978-951-691-226-7

Laaksonen, M., Rantala, J., Järnefelt, N. and Kannisto, J. (2016a). Työkyvyttömyyden vuoksi menetetty työura. http://urn.fi/URN:978-951-691-247-2

Laaksonen, M., Blomgren, J. and Gould, R. (2016b). Sickness allowance trajectories preceding disability retirement: a register-based retrospective study. The European Journal of Public Health, Vol. 26, No. 6, p. 1050–1055. doi:10.1093/eurpub/ckw081

Laesvuori, A. and Tuominen, E. (1996). Työeläkevakuutuksen ulkopuolelle jäävä palkkatyö. Eläketurvakeskuksen raportteja 4/1996, Helsinki.

Lappo, S. (2015). Uncertainty in microsimulation Assessing sampling variability in inequality and poverty indicators through variance estimation. Master's thesis. University of Helsinki.

Larja, L., Törmäkangas, L., Merikukka, M., Ristikari, T. Gissler, M. and Paananen, R. (2016). NEET-indikaattori kuvaa nuorten syrjäytymistä. Tieto& Trendit.

Lassila, J. and Valkonen, T. (2000). Työeläkkeiden indeksointi, elinaikakorjaus ja väestön ikääntyminen. ETLA Sarja B172.

Lassila, J., Määttänen, N. and Valkonen, T. (2014). Linking retirement age to life expectancy – what happens to working lives and income distribution? Finnish Centre for Pensions, Reports 02/2014. http://urn.fi/URN:NBN:fi-fe2015112619983

Lassila, J. and Valkonen, T. (2019). Eläkevarat ja riskien jako työeläkejärjestelmässä. In Kautto, M. ed. (2019). Työeläkevarat ja eläkkeiden rahoitus. Eläketurvakeskus. Hansaprint Oy, Turenki. http://urn.fi/URN:978-951-691-299-1

Lehtelä, K-M., Hannikainen-Ingman, K., Mukkila, S., Kauppinen, T. and Moisio, P. (2016). Vuoden 2016 talousarvion vaikutukset perusturvaan. Terveyden ja hyvinvoinnin laitos. Työpapereita 5/2016, Helsinki. http://urn.fi/URN:ISBN:978-952-302-629-2

Leombruni, R. and Mosca, M. (2014). The Redistributive Features of the Italian Pension System. The Importance of Being Neutral. In Dekkers, G., Keegan, M. and O'Donoghue, C. eds. (2014). New Pathways in Microsimulation. Ashgate, United Kingdom.

Li, J. and O'Donoghue, C. (2013). A survey of dynamic microsimulation models: uses, model structure and methodology. International Journal of Microsimulation, 6(2), p. 3–55.

Li, J., O'Donoghue, C. and Dekkers, G. (2014). Dynamic models. In O'Donoghue, C. Habndbook of Microsimulation Modelling. Contributions to economic analysis. Emerald.

Li, J., O'Donoghue, C., Loughrey, J. and Harding, A. (2014). Static models. In O'Donoghue, C. Habndbook of Microsimulation Modelling. Contributions to economic analysis. Emerald.

Li, X., Wang, W., Gao, Q., Wu, L. and Luo, Y. (2012). The Trajectories and Correlation between Physical Limitation and Depression in Elderly Residents of Beijing, 1992–2009. PLoS ONE 7(8): e42999. doi:10.1371/journal.pone.0042999

Liukko, J. (2005). Hyvinvoinnin eetos ja henkivakuutus. Riskin, vastuun ja solidaarisuuden muodonmuutokset. Kelan tutkimusosasto. Sosiaali- ja terveysturvan tutkimuksia 82, Helsinki. http://hdl.handle.net/10250/3293

Liukko, J. (2013). Solidaarisuuskone. Elämän vakuuttaminen ja vakuutusajattelun muutos. Gaudeamus Helsinki University Press, Tallinna.

Liukko, J., Polvinen, A., Kesälä, M. and Varis, J. (2017). Effectiveness and incentive effects of experience rating in disability insurance: A questionnaire study on employers' views. Finnish Centre for Pensions, Studies 03/2017. http://urn.fi/URN:978-951-691-267-0

Lo, Y., Mendell, N. and Rubin, D. (2001). Testing the number of components in a normal mixture. Biomiietrika (2001), 88, 3, p. 767–778.

Lyytinen, H., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A.-M., & Lyytinen, P. (2006). Trajectories of reading development: A follow-up from birth to school age of children with and without risk for dyslexia. Merrill-Palmer Quarterly, 52 (3), 514-546. doi:10.1353/mpq.2006.0031

MacCallum, R. and Austin, J. (2000). Applications of Structural Equation Modeling in Psychological Research. Annu. Rev. Psychol. 2000. 51, p. 201–226.

Mani, D. and Nandkumar, A. (2016). The differential impacts of markets for technology on the value of technological resources: An application of group-based trajectory models. Strategic Management Journal, 37(1), p. 192–205.

McLachlan, G. and Peel, D. (2000). Finite Mixture Models. John Wiley and Sons Inc. New York.

McLachlan, G. and Rathnayake, S. (2014). On the number of components in a Gaussian mixture model. WIREs Data Mining Knowl Discov 2014, 4, p. 341–355. doi: 10.1002/widm.1135

Meuleman, B., Davidov, E. and Billiet, J. (2017). Modeling Multiple-country Repeated Cross-sections. A Societal Growth Curve Model for Studying the Effect of the Economic Crisis on Perceived Ethnic Threat. methods, data, analyses 2017, pp. 1-25. DOI: 10.12758/mda.2017.10methods

Meyer, K. (2009). New Directions in Life Course Research. Annual Review of Sociology. Volume 35, 2009, p. 413–433. https://doi.org/10.1146/annurev.soc.34.040507.134619

Moisio, P., Mukkila, S., Ilmakunnas, I., Mäkinen, L. and Saikkonen, P. (2016). Perusturvan riittävyys ja köyhyys. Terveyden ja hyvinvoinnin laitos. Tutkimuksesta tiiviisti 23.

Mukkila, S., Tervola, J. and Moisio, P. (2015). Perusturvan riittävyys yksinasuvilla. Palkansaajien tutkimuslaitos. Talous ja yhteiskunta 4/2015.

Mukkila, S., Ilmakunnas, I., Moisio, P. and Saikkonen, P. (2017). Perusturvan riittävyys ja köyhyys 2017. Terveyden ja hyvinvoinnin laitos. Tutkimuksesta tiiviisti 31. http://urn.fi/URN:ISBN:978-952-302-938-5

Mukkila, S., Hannikainen-Ingman, K., Moisio, P. and Saikkonen, P. (2017). Vuoden 2017 talousarvion vaikutukset perusturvaan. Terveyden ja hyvinvoinnin laitos. Työpaperi 6/2017. Juvenes Print – Suomen Yliopistopaino Oy, Tampere. http://urn.fi/URN:ISBN:978-952-302-822-7

Mukkila, S., Hannikainen, K., Ilmakunnas, I., Moisio, P. and Saikkonen, P. (2018). Vuoden 2018 talousarvion vaikutukset perusturvaan. Terveyden ja hyvinvoinnin laitos. Työpaperi 10/2018. Helsinki. http://urn.fi/URN:ISBN:978-952-343-080-8

Murphy, R., Hagaman, A., Reinders, I., Steeves, J., Newman, A., Rubin, S., Satterfield, S., Kritchevsky, S., Yaffe, Ayonayon, H., Nagin, D., Simonsick, E., Penninx, B. and Harris, T. (2016). Depressive Trajectories and Risk of Disability and Mortality in Older Adults: Longitudinal Findings From the Health, Aging, and Body Composition Study. J Gerontol A Biol Sci Med Sci, 2016, Vol. 71, No. 2, p. 228–235.

Myles, G. (1995). Public economics. Cambridge University Press.

Määttänen, N. and Valkonen, T. (2008). Ikääntyneiden varallisuus ja sen muuttaminen kulutukseksi, ETLA Discussion Papers No. 1141.

Määttänen, N. (2014). Evaluation of Alternative Pension Policy Reforms Based on a Stochastic Life Cycle Model. In Lassila, J., Määttänen, N. and Valkonen, T. (2014). Linking retirement age to life expectancy – what happens to working lives and income distribution? Finnish Centre for Pensions, Reports 02/2014. https://www.etla.fi/wp-content/uploads/B_260_eng.pdf

Nagin, D. (1999). Analyzing Developmental Trajectories: Semiparametric, Group-Based Approach. Psychological Methods, 4, p. 39–177.

Nagin, D. (2005). Group-Based Modeling of Development. Harvard University Press, Cambridge, MA.

Nagin, D. (2016). Group-based Trajectory Modeling and Criminal Career Research. Journal of Research in Crime and Delinquency, 53 (3), p. 356–371.

Nagin, D., Jones, B., Lima Passos, V. and Tremblay, R. (2016). Group-based multi-trajectory modelling. Statistical Methods in Medical Research, 0 (0), p. 1–9.

Nagin, D. and Odgers, C. (2010). Group-Based Trajectory Modeling (Nearly) Two Decades Later. J Quant Criminol, 26 (4), p. 445–453.

Nagin, D. and Tremblay, R. (2001). Analyzing Developmental Trajectories of Distinct but Related Behaviors: A Group-Based Method. Psychological Methods 2001, Vol. 6, No. 1, p, 18–34.

Nathalie, M., Fontaine, G. and Petersen, I. (2017). Developmental Trajectories of Psychopathology. An Overview of Approaches and Applications. In Centifanti, L. & Williams, D. eds. (2017). The Wiley Handbook of Developmental Psychopathology. John Wiley & Sons Ltd.

National Pensions Act 11.5.2007/568.

Niemelä, H. (2011). Ansioeläkejärjestelmien muotoutuminen Suomessa ja Ruotsissa – kaksi erilaista vallankäyttötapaa. In Johanson, J-E., Lassila, J. and Niemelä, H. eds. (2011). Eläkevalta Suomessa. ETLA, Helsinki.

Niinivaara, R. and Viitamäki, H. eds. (2005). TUJA-käsikirja. Verotuksen ja etuuksien mikrosimulointimalli. Valtion taloudellinen tutkimuskeskus VATT-muistioita 72.

Nummi, T., Hakanen, T., Lipiäinen, L., Harjunmaa, U., Salo, M., Saha, M-T. and Vuorela, N. (2014). A trajectory analysis of body mass index for Finnish children. Journal of Applied Statistics, 41(7), p. 1422–1435.

Nummi, T., Virtanen, P., Leino-Arjas, P. and Hammarström, A. (2017a). Trajectories of a set of ten functional somatic symptoms from adolescence to middle age. Archives of Public Health, 75, 11.

Nummi, T., Salonen, J., Koskinen, L. and Pan, J. (2017b). A semiparametric mixture regression model for longitudinal data. Journal of Statistical Theory and Practice. http://dx.doi.org/10.1080/15598608.2017.1298062

Nurminen, M. (2011). Working-life expectancy in Finland: development in 2000–2009 and forecast for 2010–2015: a multistate life table approach. Finnish Centre for Pensions, Working Papers, 06/2011. http://urn.fi/URN:NBN:fi-fe2015112619993

Nurminen, M. (2012). Working-life expectancy in Finland: trends and differentials 2000–2015. A multistate regression modeling approach. Finnish Centre for Pensions, Reports 03/2012. http://urn.fi/URN:ISBN:978-951-691-166-6

Nylund, K., Asparouhov, T. and Muthén, B. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Structural Equation Modeling: A Multidisciplinary Journal, Volume 14, 2007 - Issue 4, p. 535–569.

O’Donoghue, C. (2014). Introduction. In O’Donoghue, C. (2014). Handbook of Microsimulation Modelling. Contributions to economic analysis. Emerald.

OECD (2005). Pensions at a Glance. Public Policies across OECD Countries.

OECD (2019a). Education at a Glance. OECD.

OECD (2019b). Pensions at a Glance 2019. OECD and G20 Indicators. OECD Publishing, Paris, https://doi.org/10.1787/b6d3dcfc-en

Official Statistics of Finland (2019a). Total statistics on income distribution [e-publication]. Helsinki: Statistics Finland [referred: 1.2.2019].Access method: http://www.stat.fi/til/tjkt/index_en.html

Official Statistics of Finland (2019b). Labour force survey [e-publication]. ISSN=1798-7857. Helsinki: Statistics Finland [referred: 2.1.2019]. Access method: http://www.stat.fi/til/tyti/tau_en.html

Onogi, A., Ogino, A., Sato, A., Kurogi, K., Yasumori, T. and Togashi, K. (2019). Development of a structural growth curve model that considers the causal efect of initial phenotypes. Genetics Selection Evolution (2019) 51:19. https://doi.org/10.1186/s12711-019-0461-y

Orcutt, G. (1957). A New Type of Socio-Economic System. The Review of Economics and Statistics, Vol. 39, No. 2 (May, 1957), p. 116–123.

Orcutt, G. (1960). Simulation of Economic Systems. The American Economic Review, 50 (December), p. 895–907.

Orcutt, G., Greenberger, M., Korbel, J. and Rivlin, A. (1961). Microanalysis of Socioeconomic Systems: A Simulation Study. Harper & Row, New York.

Paananen, R. and Gissler, M. (2013). Cohort Profile: The 1987 Finnish Birth Cohort. International Journal of Epidemiology. Volume 41, Issue 4.

Palomäki, L. (2018). Pensioners' subjective economic well-being in European countries: comparisons behind the income satisfaction paradox. Finnish Centre for Pensions, Studies 04/2018. http://urn.fi/URN:ISBN:978-951-691-292-2

Parpo, A. ed. (2006). SOMA 2003 Sosiaaliturvan ja verotuksen mikrosimulointimalli. Stakesin työpapereita 20/2006, Helsinki.

Pauly, M. (1974). Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection. Quarterly Journal of Economics, Vol 88, p. 44–62.

Pension indicators 2019. Finnish Centre for Pensions, reports 08/2019. http://urn.fi/URN:NBN:fi-fe2019103136005

Pentikäinen, T. (1997). Muistelua ja kommentteja työeläkejärjestelmästä. In Varoma, P. ed. (1997). Eläkepolitiikka 2000-luvulle. Eläketurvakeskus ja Työeläkelaitosten liitto, WSOY, Juva.

Perhoniemi, R., Blomberg, J. and Laaksonen, M. (2019). Sources of income following a rejected disability pension application: a sequence analysis study. Disability and rehabilitation. https://doi.org/10.1080/09638288.2018.1555619

Peutere, L., Vahtera, J., Kivimäki, M., Pentti, J. and Virtanen, P. (2015). Job Contract at Birth of the First Child as a Predictor of Women's Labor Market Attachment: Trajectory Analyses over 11 Years, Nordic Journal of Working Life Studies, 5(1), p. 9–30.

Polvinen, A. (2016). Socioeconomic status and disability retirement in Finland. Causes, changes over time and mortality. Finnish Centre for Pensions, Studies 05/2016. http://urn.fi/URN:NBN:fi-fe2016080322595

Polvinen, A., Laaksonen, M., Rantala, J., Hietaniemi, M., Kannisto, J. and Kuivalainen, S. (2018). Working while on a disability pension in Finland: Association of diagnosis and financial factors to employment. Scandinavian Journal of Public Health 2018;46(19 suppl), p. 74–81.

Pietiläinen, S. (2005). Kun yrittänyttä vakuutukseen laitettiin. Yrittäjien eläkelain synty ja varttuminen. Eläketurvakeskuksen keskustelualoitteita 2005:1.

Pylkkänen, E. (2015). Kotitalousvähennys arjen tukena. Sitra. Sitran selvityksiä 95. Erweko, Helsinki.

Rantala, J. (2014). Eläkeläisten toimeentulon kehitys ensimmäisten eläkevuosien aikana. Analyysi vuonna 1999 eläkkeelle jääneiden ekvivalenttirahatulosta vuosina 1997–2008. Yhteiskuntapolitiikka 79 (2014):2. http://urn.fi/URN:NBN:fi-fe2014040923365

Rantala, J. and Suoniemi, I. (2011). Toimeentulon muuttuminen eläkkeelle siirryttäessä - ekvivalenttitulosuhteen tarkastelua. Yhteiskuntapolitiikka 76 (2011):1. http://urn.fi/URN:NBN:fi-fe201209117806

Rantala, J. and Gould, R. (2015) Pienituloisuuden dynamiikka työkyvyttömyyseläkkeelle siirryttäessä. Yhteiskuntapolitiikka 80 (2015):3. http://urn.fi/URN:NBN:fi-fe2015061010176

Rantala, J., Hietaniemi, M., Nyman, H., Laaksonen, M. and Kuivalainen, S. eds. (2017). The economic welfare of disability pensioners' in the 2000s. Finnish Centre for Pensions, Studies 04/2017. http://urn.fi/URN:978-951-691-269-4

Rantala, J. and Pentikäinen, T. (2009). Vakuutusoppi. Finva, Vammalan Kirjapaino, Sastamala.

Rao, C. (1958). Some statistical methods for comparison of growth curves. Biometrics 14, p. 1–17.

Riekhoff, A-J. (2018). Retirement trajectories in the Netherlands and Finland: Institutional change, inequalities, de-standardisation and destabilisation. Finnish Centre for Pensions. Studies 05/2018. http://urn.fi/URN:ISBN:978-951-691-296-0

Risku, I. (2015). Yksityisalojen palkansaajien työeläkkeet syntymävuoden ja sukupuolen mukaan. Eläketurvakeskuksen raportteja 09/2015. http://urn.fi/URN:978-951-691-230-4

Rissanen, J., Grönlund, H., Herrlin, O., Kouvonen, K., Levander, M., Lilius, S. Mannonen, H., Mustonen, P., Perälehto-Virkkala, A. and Suotunen, U. (2017). Työeläke. Finva, Hansaprint Oy, Turenki.

Rubinstein, A. (2003). "Economics and Psychology"? The Case of Hyperbolic Discounting. International Economic Review, Vol. 44, No. 4, p. 1207–1216.

Sallila, S. (2009). Hyvinvointivaltion optimaalisen köyhyyspolitiikan määrittelyä. Terveyden ja hyvinvoinnin laitos, tutkimus 1, Jyväskylä. http://urn.fi/URN:NBN:fi-fe201205085081

Sallila, S. (2015). Mikrosimulointi ja politiikka-analyysi Analyysi eriarvoisuudesta ja köyhyydestä Suomessa vuosina 2006 ja 2012 ja skenaario näiden vähentämiseksi. Terveyden ja hyvinvoinnin laitos, työpaperi 36/2015, Juvenes Print – Suomen yliopistopaino, Tampere. http://urn.fi/URN:ISBN:978-952-302-591-2

Salmela-Aro, K. and Tynkkynen, L. (2009). Trajectories of Life Satisfaction Across the Transition to Post-Compulsory Education: Do Adolescents Follow Different Pathways? Journal of Youth and Adolescence. August 2010, Volume 39, Issue 8, p. 870–881.

Salmela-Aro, K. and Upadysya, K. (2014). Developmental trajectories of school burnout: Evidence from two longitudinal studies. Learning and Individual Differences. Volume 36, December 2014, p. 60–68.

Salminen, K. (1993). Pension schemes in the making. A Comparative Study of the Scandinavian Countries. The Central Pension Security Institute, Studies 1993:2, Hakapaino Oy, Helsinki.

Salonen, J. and Takala, M. (2011). Working career and income of part-time pensioners in Finland. Finnish Centre for Pensions. Working Papers 2011:2. http://urn.fi/URN:NBN:fi-fe2015112620052

Salonen, J., Virtanen, P., Lipiäinen, L., Nummi, T. and Saloniemi, A. (2014). Tutkintoja suorittaneiden nuorten työurapolut. Eläketurvakeskuksen tutkimuksia 01/2014. Juvenes Print – Suomen Yliopistopaino Oy: Tampere. http://urn.fi/URN:ISBN:978-951-691-212-0

Salonen, J. ed. (2015). Yrittäjien lakisääteinen eläketurva: työurat, työtulot ja rahoitus. Eläketurvakeskuksen raportteja 10/2015. http://urn.fi/URN:978-951-691-232-8

Salonen, J. and Möttönen, J. (2019). Osa-aikaeläkeläisten työuran pituus vertailussa kokoaikatyössä jatkaneisiin. Kansantaloudellinen aikakauskirja – 115. vsk. – 2/2019.

Saloniemi, A., Salonen, J., Lipiäinen, L., Nummi, T. and Virtanen, P. (2013). Opinnot ja työt. Kehityspolkuanalyysi ylemmän korkeakoulututkinnon suorittaneiden nuorten työurista. Yhteiskuntapolitiikka 78 (2013):4. http://urn.fi/URN:NBN:fi-fe201309055633

Saloniemi, A., Salonen, J., Nummi, T. and Virtanen, P. (2020). The diversity of transitions during early adulthood in the Finnish labour market. Journal of Youth Studies. https://doi.org/10.1080/13676261.2020.1776229

Samuelson, P. (1958). An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money. Journal of Political Economy Vol. 66, No. 6 (Dec., 1958), p. 467–482.

Sankala, M., Reipas, K. and Kaliva, K. (2018). Sijoitusriskien vaikutus TyEL:n rahoitukseen. Eläketurvakeskuksen raportteja 05/2018. http://urn.fi/URN:ISBN:978-951-691-297-7

Sen, A. (2017). Collective Choice and Social Welfare. An Expanded Edition. Harvard University Press, Cambridge Massachusetts.

Serracant, P. (2014). A Brute Indicator for a NEET Case: Genesis and Evolution of a Problematic Concept and Results from an Alternative Indicator. Social Indicators Research 117 (2), p. 401–419.

Sihvonen, H. (2015). Eläketurvakeskuksen ELSI-mikrosimulointimallin laajennus Kelan eläkkeisiin ja verotukseen. Eläketurvakeskuksen keskustelualoitteita 03/2015. http://urn.fi/URN:NBN:fi-fe201601213396

Simmons, R. and Thompson, R. (2011). NEET Young People and Training for Work: Learning on the Margins. Trentham Books Limited. London.

Singer, J. and Willett, J. (2013). Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, New York.

Statistics Canada. https://www.statcan.gc.ca/eng/microsimulation/modgen/modgen

STM (2019). Yrittäjän työeläketurvan kehittäminen. Työryhmän raportti. Sosiaali- ja terveysministeriön raportteja ja muistioita 2019:23. http://urn.fi/URN:ISBN:978-952-00-4054-3

Stull, D., Wiklund, I., Gale, R.., Capkun-Niggli, G., Houghton, K. and Jones, P. (2011). Application of latent growth and growth mixture modeling to identify and characterize differential responders to treatment for COPD. Contemporary Clinical Trials, 32 (2011), p. 818–828.

Sullivan, D. (1971). A single index of mortality and morbidity. Health Services and Mental Health Administration Health Reports 86, p. 347–354.

Takala, M. (2015). Luottamus eläketurvaan vuosina 2011 ja 2014. Eläketurvakeskuksen raportteja 04/2015. http://urn.fi/URN:ISBN:978-951-691-220-5

Takala, M. ed. (2013). Katsaus perhe-eläkkeeseen. Eläketurvakeskuksen raportteja 03/2013. http://urn.fi/URN:ISBN:978-951-691-185-7

Takala, M., Salonen, J. and Lampi, J. (2015). Survivors' pensions in Finland. Finnish Centre for Pensions, Working Papers 02/2015. http://urn.fi/URN:NBN:fi-fe201601203195

Tarvainen, T. (2017). Ansiotulojen ennustaminen stokastisessa mikrosimulointimallissa. Master's thesis, University of Jyväskylä.

Taylor-Gooby, P. ed. (2005). New Risks, New Welfare: The Transformation of the European Welfare State. Oxford University Press.

Tein, J-Y, Coxe, S. and Cham, H. (2013). Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. Structural Equation Modeling: A Multidisciplinary Journal, Volume 20, 2013 - Issue 4, p. 640–657.

Tenhunen, S. and Salonen, J. (2016). Maatalousyrittäjien työurat ja eläketurva. Eläketurvakeskuksen tutkimuksia 06/2016. http://urn.fi/URN:978-951-691-259-5

Tenhunen, S. (2017). Eläkeikä nousee mutta joustot säilyvät. Kyselytutkimus vuoden 2017 eläkeuudistuksesta ja työssäjatkamisaikeista. Eläketurvakeskuksen tutkimuksia 05/2017. http://urn.fi/URN:978-951-691-274-8

Tenhunen, S., Airio, I., Kangas, O., Koskenvuo, K. and Kuivalainen, S. eds. (2017). Eläketiedon merkitys Suomalaisten mielipiteet vuoden 2017 eläkeuudistuksesta. Eläketurvakeskuksen tutkimuksia 02/2017. http://urn.fi/URN:978-951-691-263-2

Tenhunen, S., Nivalainen, S., Järnefelt, N. and Salonen, J. (2018). Ketkä valitsevat OVEn? Tutkimus osittaiselle varhennetulle vanhuuseläkkeelle siirtymisen taustatekijöistä. Eläketurvakeskuksen tutkimuksia 06/2018. http://urn.fi/URN:978-951-691-300-4

The Journal of Risk and Insurance (1999). American Risk and Insurance Association, June 1999, Vol 66, No. 2.

THL (2015). Perusturvan riittävyyden II arviointiryhmä. Perusturvan riittävyyden arviointiraportti 2011–2015. Terveyden ja hyvinvoinnin laitos (THL). Työpaperi 1/2015. Helsinki.

Tikanmäki, H., Sihvonen, H. and Salonen, J. (2014). Microsimulating Finnish earnings-related pensions. Finnish Centre for Pensions, Working Papers 02/2014. http://urn.fi/URN:NBN:fi-fe2015112620030

Tikanmäki, H., Sihvonen, H. and Salonen, J. (2015). Työurien ja eläkeajan kehitys sekä eläkkeiden ja niiden jakautumisen muutos. In Kautto, M. and Risku, I. eds. (2015) Laskelmia vuoden 2017 työeläkeuudistuksen vaikutuksista. Eläketurvakeskuksen raportteja 02/2015. http://urn.fi/URN:ISBN:978-951-691-216-8

Tikanmäki, H., Ahola, I., Koponen, J., Reipas, K. and Suikkanen, H. (2017a). Skenaariolaskelmia työn muutoksen vaikutuksesta eläkejärjestelmään ja julkiseen talouteen. Eläketurvakeskuksen raportteja 03/2017. http://urn.fi/URN:978-951-691-272-4

Tikanmäki, H., Appelqvist, J., Reipas, K., Sankala, M. and Sihvonen H. (2017b). Statutory pensions in Finland – Long-term projections 2016. Finnish Centre for Pensions, Reports 02/2017. http://urn.fi/URN:978-951-691-265-6

Tikanmäki, H., Lappo, S., Merilä, V., Nopola, T., Reipas, K. and Sankala, M. (2019). Statutory pensions in Finland – Long-term projections 2019. Finnish Centre for Pensions, Reports 07/2019. http://urn.fi/URN:ISBN:978-951-691-009-6

Titterington, D., Smith, A. and Makov, U. (1985). Statistical Analysis of Finite Mixture Distributions. Wiley, New York.

Tolvanen, A. (2007). Latent Growth Mixture Modeling: A Simulation Study. University of Jyväskylä, PhD dissertation. University of Jyväskylä, Department of Mathematics and Statistics, Report 111. Jyväskylä.

Tucker, L. (1958). Determination of parameters of a functional relationship by factor analysis. Psychometrika 23, p. 19–23.

Tuomala, M. (2004). Miksi sosiaalivakuutusta tarvitaan? In Hellsten, K. and Helne, T. eds. (2004). Vakuuttava sosiaalivakuutus? Kela.

Tuomala, M. (2009). Julkistalous. Uudistettu painos. Gaudeamus, Helsinki.

Tuomala, M. (2016). Optimal Redistributive Taxation. Oxford University Press.

Turunen, T. (2015). Työeläkettä kartuttavat palkattomat jaksot sekä niiden jakautuminen Suomessa vuosina 2005–2013. Eläketurvakeskuksen raportteja 11/2015. http://urn.fi/URN:978-951-691-234-2

Työeläkelainsäädäntö 2017. Eläketurvakeskuksen käsikirjoja 01/2017. Edita Publishing Oy, Keuruu.

Uusitalo, H. ed. (2011). Työeläkejärjestelmän uudistamisen tavoitteiden mittaaminen: taustaselvitys työuraryhmälle. Eläketurvakeskuksen selvityksiä 2011:2. http://urn.fi/URN:NBN:fi-fe2015112620086

Uusitalo, R. and Nivalainen, S. (2013). Vuoden 2005 eläkeuudistuksen vaikutus eläkkeellesiirtymisikään, The Office of the Finnish Government, report series 5/2013.

Vaalavuo, M. ed. (2018). Sosiaali- ja terveydenhuollon asiakasmaksujen kohdentuminen, vaikutukset ja oikeudenmukaisuus. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 30/2018. http://urn.fi/URN:ISBN:978-952-287-534-1

Vaittinen, R. and Vanne, R. (2011). National Transfers Accounts for Finland. In Ronald, L. and Mason, A. eds. (2011). Population Ageing and the Generational Economy – A Global Perspective, Edgar Elgar, Cheltenham.

Vaittinen, R. and Vanne, R. (2017). Finland's slow recovery from the financial crisis: A demographic explanation, The Journal of the Economics of Ageing. https://doi.org/10.1016/j.jeoa.2017.06.003

Valtiovarainministeriö (2018). Eri sijoitusmuotojen verokohtelu. Työryhmäraportti. Valtiovarainministeriön julkaisu 14/2018, Helsinki.

van der Geest, V., Bijleveld, C., Blokland, A. and Nagin, D. (2016). The Effects of Incarceration on Longitudinal Trajectories of Employment: A Follow-Up in High-Risk Youth From Ages 23 to 32. Crime & Delinquency, 62(1), p. 107–140.

van Sonsbeek, J-M. (2011a). Microsimulation as a Decision Making Tool In Social Security Policy. PhD dissertation. Vrije Universiteit, The Netherlands.

van Sonsbeek, J-M. (2011b). Micro simulations on the effects of ageing-related policy measures: The Social Affairs Department of the Netherlands Ageing and Pensions Model. International Journal of Microsimulation (2011) 4(1), p. 72–99.

Varis, J. (2018). Eläketurvakeskuksen koneoppimiskokeilu – näin se tehtiin! ETK blogi. https://www.etk.fi/blogit/elaketurvakeskuksen-koneoppimiskokeilu-nain-se-tehtiin/

Varoma, P. ed. (1997). Eläkepolitiikka 2000-luvulle. Eläketurvakeskus ja Työeläkelaitosten liitto, WSOY, Juva.

Warren, J., Luo, L., Halpern-Manners, A., Raymo, J. and Palloni, A. (2015). Do Different Methods for Modeling Age-Graded Trajectories Yield Consistent and Valid Results? American Journal of Sociology 2015 May; 120(6), p. 1809–1856. doi:10.1086/681962.

Wiesner, M. and Windle, M. (2004). Assessing Covariates of Adolescent Delinquency Trajectories: A Latent Growth Mixture Modeling Approach. Journal of Youth and Adolescence, Vol. 33, No. 5, October 2004, p. 431–442.

Williamson P, Zaidi, A. and Harding, A. (2009). New Frontiers in Microsimulation Modelling: Introduction. In Williamson, P., Zaidi, A. and Harding, A. (2009). New Frontiers in Microsimulation Modelling. Ashgate.

Wolfson, M, (2009). Preface – Orcutt's Vision 50 Years On. In Williamson, P., Zaidi, A. and Harding, A. (2009). New Frontiers in Microsimulation Modelling. Ashgate.

Zaidi, A. and Rake, K. (2001). Dynamic Microsimulation Models: A Review and Some Lessons for SAGE. SAGE Discussion Paper no. 2. The London School of Economics.

APPENDIX

Appendix 1. ELSI model yearly information contents.

No.	Contents
1	Anonymous personal identifier
2	Gender
3	Population and labor market state (21 states)
4	Year of birth
5	Year of death
6	Earnings-related pension paid as of this date
7	Duration of working life since age 18 (years)
8	Accrued earnings-related pension (€/month)
9	Earnings-related pension in payment (€/month)
10	Earnings from paid work (€/year)
11	Earnings from self-employment (YEL, MYEL) (€/year)
12	Share of earnings-related child benefits for which pension accrues (€/year)
13	Share of earnings-related unemployment allowance for which pension accrues (€/year)
14	Share of earnings for other unpaid periods for which pension accrues (€/year)
15	Share calculated for the projected pension component of earnings for unpaid periods (€/year)
16	VEKL (Care for own child/ren under the age of 3 and study leading to a degree, €/year)
17	Level of education (5 levels)
18	Country of residence (1=Finland, 0=country with social security agreement, -1=other)
19	Earnings-related pension paid over one year, i.e. pension expenditure (€/year)

No.	Contents
20	Pension accrued in retirement excl. accrual for period of partial old-age pension (€/month)
21	Paid earnings-related pension contributions (€/year)
22	Projected pension component (€/year)
23	Earnings that the projected pension component is based on (€/year)
24	Deduction for early retirement on an earnings-related pension (€/year)
25	Increment for late retirement on an earnings-related pension (€/year)
26	Earnings-related pension that has accrued after retirement age (€/year)
27	Pension accrual under VEKL (€/year)
28	Earnings-related surviving spouse's pension + pensions from abroad (€/month)
29	Type of family or type of housing (1=marriage, 2=cohabiting, 3=single)
30	National pension (€/month)
31	Time of residence in Finland as of age 16 (years)
32	Guarantee pension (€/month)
33	National pension expenditure (€/year)
34	Guarantee pension expenditure (€/year)
35	Net income (Wages + Earnings from self-employment + Earnings-related pensions + Kela pensions – Taxes, €/year)
36	Net total pension (net income *(Gross pension / Gross earnings), €/year)

Appendix 2. ELSI model population and labor market states.

No.	Abb.	Contents
Full-time employed		
1	A1	Full-time employed for the first consecutive year
2	A2	Full-time employed for the second consecutive year
3	AK	Full-time employed for at least the third consecutive year
On social security benefits		
4	AT	Unemployed (excluding those on an unemployment pathway to retirement)
5	AL	On an unemployment pathway to retirement
6	AS	Sickness benefits preceeding full disability pension

No.	Abb.	Contents
Retired and employed		
7	OE	Part-time pension
8	OV	Partial old-age pension and employed (as of 2017)
9	OT	Partial disability pension and employed
Full-time retired		
10	TU	Years-of-service pension (as of 2018)
11	Ov	Partial old-age pension and not employed (as of 2017)
12	VE	Old-age pension
13	T1	Full disability pension for the first consecutive year
14	TK	Full disability pension for at least the second consecutive year
15	Ot	Partial disability pension and not employed
16	KV	National old-age pension only
17	KT	National disability pension only
Outside the labor force for other reasons		
18	Ak	Outside the labor force but has accrued earnings-related pension
19	SY	Permanently outside the labor force but has accrued earnings-related pension
20	Mu	Outside the labor force and has not accrued earnings-related pension
Deceased		
21	KU	Deceaced

Appendix 3. Social security benefits included in ELSI model.

No.	Social benefit
	Child benefits
1	Maternity allowance
2	Special maternity allowance
3	Paternity allowance
4	Parental allowance
5	Partial parental allowance
6	Parental allowance during period of sickness allowance

No.	Social benefit
	State-payed benefits, VEKL*
7	Child home care allowance
8	Basic vocational qualification
9	Lower university degree
10	Degree from university of applied sciences
11	Higher university degree
	Earnings-related unemployment benefits
12	Earnings-related daily allowance
13	Earnings-related daily allowance for YEL/MYEL insured
	Other social security benefits
14	Sickness allowance
15	Partial sickness allowance
16	Earnings-related job alternation compensation

PUBLICATIONS

PUBLICATION I

Statistical analysis of labor market integration: A mixture regression approach

Nummi Tapio, Salonen Janne & O'Brien Timothy

New Advances in Statistics and Data Science, ICSA Book Series in Statistics.

(https://doi.org/10.1007/978-3-319-69416-0_18S)

Statistical Analysis of Labor Market Integration: A Mixture Regression Approach

Tapio Nummi, Janne Salonen and Timothy O'Brien

Abstract In this paper we investigate the labor market integration of young males in Finland. The study is based on individual-level registry data which contains information about working, studying, sickness and maternity benefits. Using data from 2005 to 2013, we have studied the labor market attachment of the birth cohort of 1987, which totals 29,383 individuals. The statistical methodology is a multivariate version of mixture modeling applied to longitudinal data. We have modeled the information about labor market outcomes as a binary dependent variable, and thus we apply a multivariate logistic mixture regression model. We found that there are ten main clusters or groups that lead to different labor market outcomes. Our results suggest that our mixture regression approach applied to register-based population can reveal new information that may remain hidden in more formal, census-based labor market statistics. To our knowledge, application of finite mixture model methods have not yet been applied to extensive and comprehensive pension data of this sort. Our analysis also provides valuable information for policy-makers and good statistical tools for corresponding analyzes of register data.

1 Introduction

Integration of young people into the labor market is a complex and socially important issue which must be understood as a process in time. The majority of young

Tapio Nummi
University of Tampere, Faculty of Natural Sciences, Finland, e-mail: tan@uta.fi

Janne Salonen
The Finnish Centre for Pensions, Research Department, Finland e-mail: Janne.Salonen@etk.fi

Timothy O'Brien
Loyola University of Chicago, Department of Mathematics and Statistics and Institute of Environmental Sustainability, USA
e-mail: teobrien@gmail.com

people attach to the labor market quite quickly, some after their studies although some do remain unemployed (e.g. Pareliussen 2016). Moreover, different stages may not necessarily follow a straightforward progression. For example, it is quite common in the Finnish system for students to work during their studies. Meaning, one person can have several different statuses at the same time (throughout one year). Obtaining an overall picture of such complex and heterogeneous longitudinal data is a challenging task.

In this paper we present one possible approach to this complex data analysis problem. The approach is based on the mixture regression applied to multivariate longitudinal binary data. In applied statistics, these methods are often referred to as latent class regression models, or trajectory analysis (Nagin, 2005). The idea is that data consists of unknown sub-populations with some common properties that can be revealed through longitudinal data. Recently these methods have been very popular in many fields of science, including psychology, education, sociology, marketing and health sciences (Korpela et al. 2017, Kokko et al. 2008, Jolkkonen et al. 2017, Mani et al. 2016, Nummi et al. 2014 and Nummi et al. 2017).

In this paper we present a 4-dimensional binary mixture regression model that is used to identify the subgroups in the data gathered. We find that there are ten main groups that lead to different development paths of young men. Most people integrated into the labor market quite quickly after various intermediate stages, but in a few groups the integration is weaker or slower. In terms of society, the groups of weak attachment are of central interest, because they may later require special support or action from society. For instance, in some countries a special youth guarantee policy has been promoted (Keränen 2012; Escudero and Mourelo 2015).

2 Methods

2.1 Data

The data comes from the administrative registers of The Finnish Centre for Pensions and Statistics Finland. In the administrative registers there is a range of information pertaining to all of the pension insured (total population) people in Finland. For this study we choose the male cohort born in 1987. For other studies of the same cohort we can refer to Paananen and Gissler (2013). We follow all individuals between 2005 to 2013, when the cohort is 18 to 26 years of age. We take a sub-population of those who are Finnish citizens and who have lived in Finland during the specified period. The research population is 29,383 males.

Labor market attachment is measured using days when working, in education, in unemployment and on various social benefits per year. This yields the 4-dimensional response vector as follows:

- Variable 1 (Employed): Individual employed for days/year in private or public sector or self-employed.

Table 1 Clusters k, their estimated absolute N_k and relative sizes $\hat{\pi}_k$.

Group(k)	N_k	$\hat{\pi}_k$ (%)
1:	545	1.9
2:	3009	10.2
3:	1971	6.7
4:	1975	6.7
5:	5942	20.2
6:	5933	20.2
7:	1510	5.1
8:	3141	10.7
9:	4093	13.9
10:	938	3.2

- Variable 2 (Education): Individual in education leading to a degree and/or is receiving student financial aid.
- Variable 3 (Unemployed): Individual receives unemployment benefits, either earnings related or paid by the state.
- Variable 4 (Leave): Individual receives sickness benefits or is on vocational rehabilitation. Parental leave and the permanently disabled are included here as well.

The original data is measured as days/year. For our analysis, the data was dichotomized (Yes/No) because in this type of longitudinal data, the most important factor in the formation of an individuals career trajectory is the several different statuses of the individuals. This makes the analysis of data much simpler and more uniform.

2.2 Multivariate binary mixture

Our aim is to identify clusters of individuals with the same kind of mean developmental profiles (trajectories). Let $\mathbf{y}_i = (y_{ij1}, y_{ij2}, \ldots, y_{ijT})'$ represent the sequence of measurements on individual i for the variable j over T periods and let $f_i(\mathbf{y}_i|\mathbf{X}_i)$ denote the marginal probability distribution of $\mathbf{y}_i$ with possible time dependent covariates $\mathbf{X}_i$ that are same to all $j = 1, \ldots, s$ variables. It is assumed that $f_i(\mathbf{y}_i|\mathbf{X}_i)$ follows a mixture of K densities

$$f_i(\mathbf{y}_i|\mathbf{X}_i) = \sum_{k=1}^{K} \pi_k f_{ik}(\mathbf{y}_i|\mathbf{X}_i), \quad \sum_{k=1}^{K} \pi_k = 1 \text{ with } \pi_k > 0, \tag{1}$$

where π_k is the probability of belonging to the cluster k and $f_{ik}(\mathbf{y}_i|\mathbf{X}_i)$ is the density for the kth cluster (see e.g. McLachlan and Peel, 2000). The natural choice is to use the Bernoulli distribution for the mixture components. It is assumed that s variables in $\mathbf{y}_i$, $i = 1, \ldots, N$, are independent. Also measurements given the kth cluster are

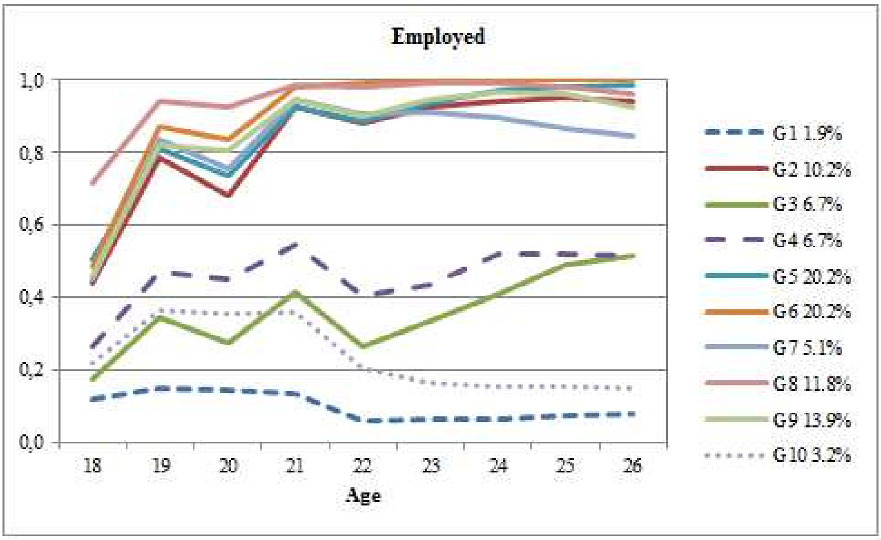


Fig. 1 Time-point means (proportions) over trajectories for the variable Employed.

assumed to be independent. This yields the density

$$f_{ik}(\mathbf{y}_i|\mathbf{X}_i) = \prod_{j=1}^{s}\prod_{t=1}^{T} p_{ijtk}^{y_{ijt}}(1-p_{ijtk}^{1-y_{ijt}}), \tag{2}$$

where p_{ijtk} is a function of covariates $\mathbf{X}_i$. For modeling the conditional distribution of p_{ijtk} we use the logistic regression model. For the ith individual we can then write

$$p_{ijtk} = \frac{\exp(\mathbf{x}_i'\beta_{jk})}{1+\exp(\mathbf{x}_i'\beta_{jk})}, \tag{3}$$

where $\mathbf{x}_i'$ is the tth row of $\mathbf{X}_i$, β_{jk} is the parameter vector of the jth variable in the kth cluster. For our analysis we took the second degree model

$$\mathbf{x}_i'\beta_{jk} = \beta_{0jk} + \beta_{1jk}t + \beta_{2jk}t^2 \tag{4}$$

for modeling the probabilities within the variable j and cluster k in time t. Maximum likelihood estimates can then be obtained by maximizing the log likelihood $\log\sum_{i=1}^{N} f_i$ over unknown parameters $\beta_{jk},\ j = 1,\ldots s; k = 1,\ldots K$ (Nagin, 1999; Jones et al., 2001; Nagin and Tremblay, 2001; Jones and Nagin, 2007, Nagin and Odgers 2010a & 2010b). When the model parameters are estimated the posterior probability estimate provides a tool for assigning individuals to specific clusters. Individuals can then be assigned to specific clusters to which their posterior probability is the largest.

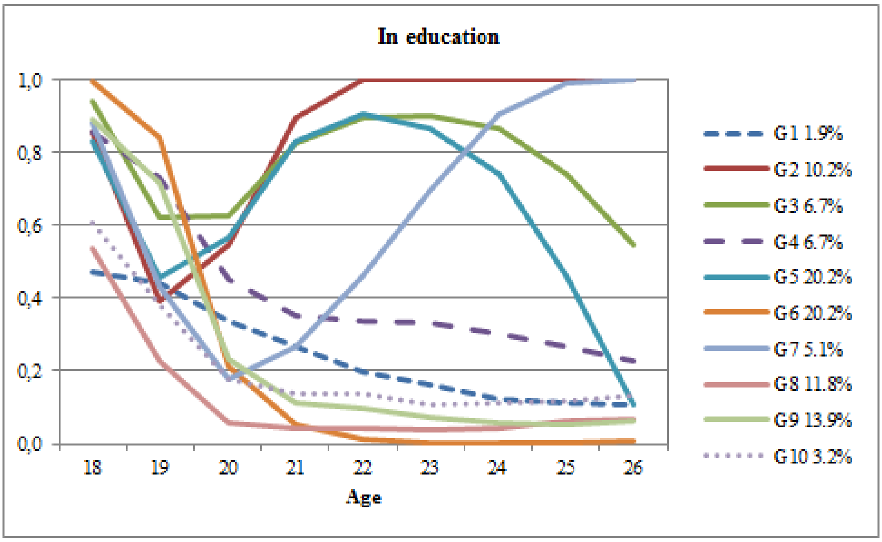


Fig. 2 Time-point means (proportions) over trajectories for the variable Education.

3 Analysis

Choosing the number of trajectories (clusters) K is an important issue when applying mixture modeling. The selection of K can be based on technical criteria, substantive examination, or both. We used the information criteria BIC, which is perhaps the most widely used in this context. Here we present the values of BIC for $k = 5, \ldots, 10$: -397509.4, -394554.5, -392966.5, -389504.3, -388897.6 and -383596.6. The maximum (note: SAS implementation) is obtained for $k = 10$ and is therefore our choice for the number of clusters K. Increasing K could yield to more difficult to interpret and insignificantly small clusters. The final estimated model is summarized in Table 1 and Table 4 (in the Appendix). The trajectory plots (Figures 1- 4) present conditional point means calculated for each of the four variables. These plots are used as the main tool for the interpretation of the results obtained from the mixture regression fit. The computations were carried out by SAS proc traj procedure.

3.1 Normal life-course

In the trajectory plots the solid lines indicate groups (total of 88.1%) where labor market integration is good (Figure 1). It is quite common for young people ages 18 to 21 to be in post-secondary or vocational education (Figure 2). From Figure 1 we note that about 74.6% of young males are in the trajectories where the percentage of people who have employment status after 21 years of age is surprisingly high

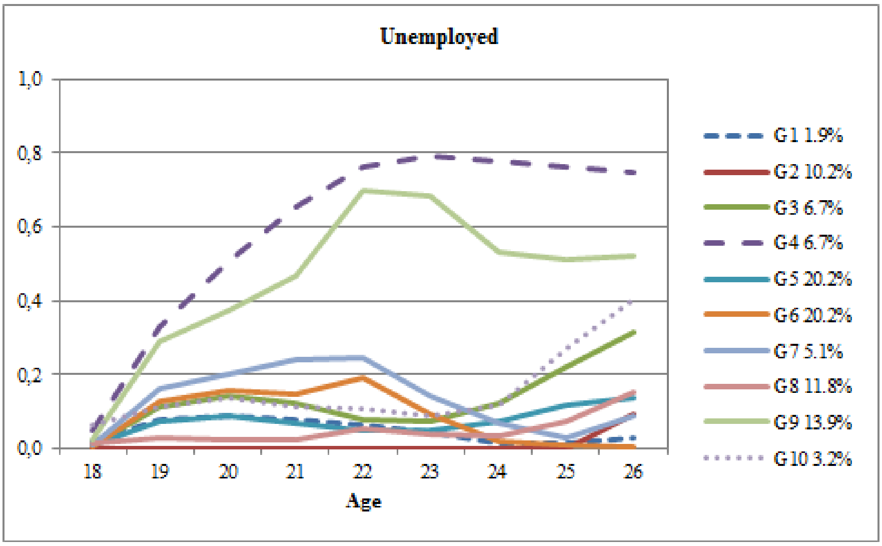


Fig. 3 Time-point means (proportions) over trajectories for the variable Unemployed.

(> 80%). We refer to this group of trajectories (groups 2, 5, 6, 7, 8 and 9) as the HES group (high employment status group). Trajectories 5 and 2 (30.4%) from the HES group both have high percentage of people (> 50%) in education after 21 years of age (Figure 2). This reflects the well-known fact that the majority of students in Finland work during their studies. Group 2 (10.2%) contains typical university students. Note that this group also has a high percentage of people in employment.

Group 9 (13.9%) is interesting as we can see from Figure 3 that this group has a low percentage of people in education after the age of 21 (Figure 2). However, the percentage of people in employment is still high and increases with age (Figure 1). This group contains low-educated young men who work in various part-time or temporary jobs and experience unemployment periods.

In general most of the trajectory groups (Figure 3) have low percentage of people who are unemployed, with many young men experiencing short unemployment periods. These unemployment spells are usually short throughout their normal life-course. This is also true for sickness and disability periods (Figure 4). As the trajectories indicate, young Finnish men participate in family life (take parental leave), and this participation increases with age. However, due to short follow-up time, a more thorough analysis of parental leave is not possible as the first child is usually born after this time.

From employment and education trajectories we can see a change at age 20 (Figure 1 and Figure 2). This is due to the fact that the Finnish male cohort (80% of males) enters military service at this age. Service lasts less than a year, and therefore these men in the military have no status in the labor market or on education during this time.

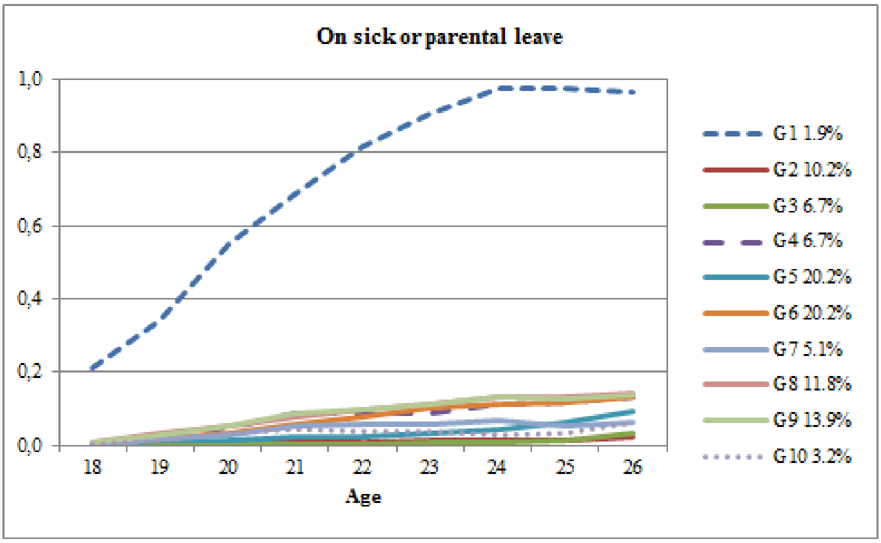


Fig. 4 Time-point means (proportions) over trajectories for the variable Leave.

3.2 Weak labor market integration

In the figures the dotted lines indicate groups (total of 11.8%) where labor market integration is poor for longer time periods. We refer to this group of trajectories (groups 1, 3 and 10) as the LES group (low employment status group) which is our main interest and focus. These young men are experiencing difficulties in the labor market and have therefore been the target government support programs.

From Figure 2 we can identify four groups (1, 3, 4 and 10) with low percentages in terms of employment. There is one positive indication, however, as group 3 contains higher percentage in education (Figure 2). The LES group is not particularly active in education, as they perhaps receive some vocational education in their late teens.

Unemployment and leave trajectories (Figures 3 and 4) explain the LES group in more detail. Group 4 (6.7%) is clearly unemployed after secondary and vocational education. The percentage of unemployment in this group is nearly 80% at the end of the follow up period.

The sickness and parental leave trajectory plot indicates that Group 1 (1.9%) has the highest percentage (Figure 4). In fact this group is not on parental leave, but receives occupational rehabilitation or a disability pension instead. This is clearly the group with the most difficulties in the labor market. In Finland occupational rehabilitation is rather effective, so these young men may have a chance to attend school or work later in life.

As an ex-post validation of these trajectory groups we can measure or sum up the working days over the follow-up period. The length of working life is on average 5.1

years for the HES group and only 1.3 years for the LES group. The results confirm that this analysis has found clusters that also have practical importance.

4 Concluding remarks

It is clear that our mixture regression analysis is an effective tool for the identification of different clusters of register-based data. Naturally, the central interest is on those who have difficulties with labor market integration. We think that our analysis provides new insight into this important social issue. We mainly concentrated on a descriptive analysis of results. However, it would also be interesting to analyze the identified clusters further using covariates like social class, living area, parents education, or parents income. The best way to proceed may be the joint modeling of clusters and mixing percentages using multinomial regression. This more subject-oriented analysis remains a topic of further research.

Acknowledgements The authors wish to thank The Finnish Centre for Pensions and Statistics Finland for providing the research data for this study. We also like to thank the referees for the comments that led to improvements of the paper.

References

1. Escudero Veronica and Mourelo Elva Lopez (2015). The Youth Guarantee programme in Europe: Features, implementation and challenges. ILO.
2. Keränen Henna (2012). Young people within services best practices for the promotion of the youth quarantee. The Ministry of Employment and the Economy in Finland.
3. Jolkkonen A., Kurvinen A., Virtanen P., Lipiäinen L., Nummi T., Koistinen P. (2017). Labour market attachment following major workforce downsizings: A comparison of displaced and retained workers, Work, Employment and Society (2017). DOI: 10.1177/0950017017706305.
4. Jones Bobby L., Nagin Daniel S. and Roeder Kathryn (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. Sociological Methods % Research, Vol. 29, No 3, 374–393.
5. Jones Bobby L. and Nagin Daniel S. (2007). Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them. Sociological Methods % Research, Vol. 35, No 4, 542–571.
6. Kokko Katja, Pulkkinen Lea, Mesiäinen Päivi, Lyyra Anna-Liisa (2008).Trajectories Based on Postcomprehensive and Higher Education: Their Correlates and Antecedents. Journal of Social Issues. Volume 64, Issue 1, March 2008, pages 59-76.
7. Korpela, K., Nummi, T., Lipiäinen, L., De Bloom, J., Sianoja, M., Pasanen,T., Kinnunen, U., Nature exposure predicts well-being trajectory groups among employees across two years, Journal of Environmental Psychology (2017), doi: 10.1016/j.jenvp.2017.06.002.
8. Mani Deepa, Nandkumar Anand (2016). The differential impacts of markets for technology on the value of technological resources: An application of group-based trajectory models. Strategic Management Journal. Volume 37, Issue 1. January 2016 Pages 192-205.
9. McLachlan, G. and Peel, D. (2000). Finite Mixture Models, Wiley: Hoboken, NJ.

10. Nagin Daniel S. and Tremblay Richard E. (2001). Analyzing Developmental Trajectories of Distinct but Related Behaviors: A Group-Based Method. Psychological Methods, Vol. 6, No. 1, 18–34.
11. Nagin Daniel S. (2005). Group-Based Modeling of Development. Harvard University Press.
12. Nagin Daniel S. and Odgers Candice L. (2010). Group-Based Trajectory Modeling (Nearly) Two Decades Later. Journal of Quantitative Criminology 2010 Dec; 26(4): 445-453.
13. Nagin, D.S. and Odgers, C.L. (2010). Group-Based Trajectory Modeling in Clinical Research, Annual Review of Clinical Psychology, 6, 109–138.
14. Nummi Tapio, Hakanen Tiina, Lipiäinen Liudmila, Harjunmaa Ulla, Salo Matti, Saha Marja-Terttu,Vuorela Nina. (2014). A trajectory analysis of body mass index for Finnish children. Journal of Applied Statistics, 41 (7), 1422–1435.
15. Nummi T., Virtanen P., Leino-Arjas P., Hammarström A. (2017). Trajectories of a set of ten functional somatic symptoms from adolescence to middle age. Archives of Public Health 75, 11.
16. Paananen Reija and Gissler Mika (2013). Cohort Profile: The 1987 Finnish Birth Cohort. International Journal of Epidemiology. Vol 41.
17. Pareliussen Jon Kristian (2016). Age, skills and labour market outcomes in Finland. OECD Economics Department Working Papers, No. 1321, OECD Publishing, Paris.

Appendix

Table 2 Summary of the estimated model: Variables, clusters, parameter estimates and their standard errors.

Variable	Group	$\hat{\beta}_0$	SE($\hat{\beta}_0$)	$\hat{\beta}_1$	SE($\hat{\beta}_1$)	$\hat{\beta}_2$	SE($\hat{\beta}_2$)
1	1	2.398	4.0937	-0.3086	0.3809	0.0043	0.0088
1	2	-33.1751	1.6662	2.8697	0.1567	-0.0572	0.0036
1	3	-1.3307	1.4851	-0.082	0.1356	0.0052	0.0031
1	4	-10.4101	1.4206	0.8602	0.1308	-0.0177	0.003
1	5	-18.5665	1.4307	1.4459	0.1371	-0.0223	0.0033
1	6	-30.6046	3.7969	2.2395	0.3836	-0.0297	0.0097
1	7	-49.6895	2.1645	4.5499	0.201	-0.0993	0.0046
1	8	-64.2231	2.9188	5.9415	0.2719	-0.1288	0.0062
1	9	-48.3674	1.6773	4.3293	0.1584	-0.0912	0.0037
1	10	-10.4439	2.556	1.0078	0.2392	-0.0264	0.0055
2	1	11.1319	2.9527	-0.8535	0.2754	0.013	0.0064
2	2	399.1368	10.4145	-41.2003	1.0772	1.0621	0.0278
2	3	-20.6646	1.6869	2.0785	0.156	-0.0485	0.0036
2	4	36.9462	1.8208	-3.1096	0.1655	0.0635	0.0037
2	5	-54.0624	0.9532	5.2318	0.0885	-0.1231	0.002
2	6	165.358	4.3881	-14.2717	0.3962	0.2974	0.0089
2	7	151.7117	4.8462	-14.7589	0.4754	0.3563	0.0116
2	8	72.251	2.3522	-6.5804	0.2146	0.1428	0.0048
2	9	83.5415	2.0619	-7.2462	0.1871	0.1515	0.0042
2	10	45.0733	2.7618	-4.0274	0.2565	0.0857	0.0059
3	1	-37.1495	7.5126	3.338	0.7054	-0.0805	0.0164
3	2	-3.5307	99.786	-15.2704	54.119	0.5892	2.0791
3	3	7.1229	2.1854	-1.0465	0.1995	0.0284	0.0045
3	4	-61.0014	2.0817	5.2682	0.1901	-0.1111	0.0043
3	5	-1.4631	1.5541	-0.2562	0.1412	0.0092	0.0032
3	6	-110.2661	3.8561	10.3492	0.3723	-0.246	0.009
3	7	-54.1435	3.2046	4.9123	0.2998	-0.1141	0.007
3	8	8.3929	3.0624	-1.3156	0.2777	0.0354	0.0062
3	9	-58.2828	1.2639	5.0717	0.1153	-0.1092	0.0026
3	10	19.9764	2.9505	-2.2136	0.2699	0.0548	0.0061
4	1	-22.1081	4.2901	1.4843	0.4084	-0.0185	0.0097
4	2	-16.5855	5.8831	0.9372	0.5318	-0.0173	0.0119
4	3	6.3987	8.9441	-1.345	0.7973	0.0372	0.0176
4	4	-33.001	3.2404	2.5058	0.2873	-0.0505	0.0063
4	5	-11.5583	2.6232	0.3925	0.2318	-0.0015	0.0051
4	6	-36.7212	2.2235	2.7618	0.1953	-0.0548	0.0043
4	7	-38.7528	4.6765	3.0266	0.4167	-0.0633	0.0092
4	8	-30.3185	2.4405	2.2794	0.2174	-0.0456	0.0048
4	9	-33.8236	2.1217	2.6022	0.1885	-0.0529	0.0042
4	10	-20.8289	6.5442	1.4166	0.591	-0.0281	0.0132

PUBLICATION II

Underinsurance in the statutory pension scheme: A trajectory analysis of self-employed earnings in Finland

Salonen Janne, Koskinen Lasse & Nummi Tapio

Manuscript accepted for publication into the International Social Security Review

PUBLICATION III

Distributional effects of the forthcoming Finnish pension reform – A dynamic microsimulation approach

Tikanmäki Heikki, Sihvonen Hannu & Salonen Janne

International journal of microsimulation 8(3) 75–98.

INTERNATIONAL JOURNAL OF MICROSIMULATION (2015) 8(3) 75-98

INTERNATIONAL MICROSIMULATION ASSOCIATION

Distributional Effects of the Forthcoming Finnish Pension Reform – a Dynamic Microsimulation Approach

Heikki Tikanmäki

Finnish Centre for Pensions
FI-00065 Eläketurvakeskus
e-mail: heikki.tikanmaki@etk.fi

Hannu Sihvonen

Finnish Centre for Pensions
FI-00065 Eläketurvakeskus
e-mail: hannu.sihvonen@etk.fi

Janne Salonen

Finnish Centre for Pensions
FI-00065 Eläketurvakeskus
e-mail: janne.salonen@etk.fi

ABSTRACT: The Finnish pension system consists of earnings-related pensions (almost 90% of the total pension expenditure) and the national pension. The earnings-related pension system will undergo a major reform in 2017. The main objective of the reform is to ease the burden on public finances by extending working lives.

The earliest eligibility age for old-age pension will gradually rise from the current 63 years to 65 years by the 1962 birth cohort. After that, the earliest eligibility age will be linked to the growth in life expectancy. The age-specific accrual rates will be unified after a transition period. The early retirement schemes will also be reformed. The level of the disability benefits will improve.

The forthcoming reform has been analysed using the ELSI microsimulation model of the Finnish

Centre for Pensions. The model is based on administrative register data and it covers both the earnings-related and the national pension system of Finland. ELSI is a model with dynamic ageing without behavioural adjustments. The behavioural changes resulting from the reform have been taken into account explicitly: the age-gender-specific behaviour is obtained from our macro model using a micro-macro link. The differences in transition probabilities between educational groups are extrapolated from the register data.

The reform indicates changes in income distribution. We analyse the impacts of the reform on working lives, the pension level as well as the pension distribution. The results are presented by gender and educational level.

KEYWORDS: dynamic microsimulation, Finnish pension system, pension reform, working lives.

JEL classification: **H55, C63**

1. INTRODUCTION

The most important part of the statutory pension system in Finland is the earnings-related pension system, which corporates approximately 90 per cent of the total pension expenditure. In practice, the earnings-related pension system is a pillar I system with some pillar II features. In addition, there is a pillar I national pension system and a guarantee pension system that supplement earnings-related pensions for those who have accrued no or only a relatively small earnings-related pension. It has been suggested that the limited role of private pensions (III pillar) is due to the fact that there is no ceiling in the earnings-related pension scheme (Andersen *et al.* 2014, p. 124).

The Finnish pension system was last reformed in 2005. As a result of that reform, some early retirement benefits were discontinued, while the flexible retirement age and the life expectancy coefficient were introduced (see appendix A.2). The effects of these reforms are currently starting to show (e.g. Hietaniemi *et al.* 2006; Lassila and Valkonen 2007). The Finnish pension system is facing its next major reform in 2017. The main changes include increasing the earliest eligibility age for old-age pension and harmonizing the age-specific accrual rates. There will also be some changes to the early retirement options. For example, the new disability pensions will improve, the part time pension will be discontinued, new partial old-age pension[1] and years-of-service pensions will be introduced. Most of the changes will have a direct impact only on earnings-related pensions. However, the age limits of the national pension system will also change when the earliest eligibility age in the earnings-related pension system rises.

In this paper, we analyse the reform using the ELSI microsimulation model developed by the Finnish Centre for Pensions. Our LTP (Long Term Projection) macro model (Risku *et al.* 2014) has also been used for modelling the reform (Finnish Centre for Pensions 2014, Kautto and Risku 2015)[2]. In aggregate level, the two models give essentially analogous results. The added value of the microsimulation model is that we can obtain distributional information on the pension levels and replacement rates and receive information on the different impacts on various population sub-groups.

ELSI is based on administrative register data. It is a dynamic microsimulation model in the sense of dynamic aging (discrete-time/cross sectional aging). The model contains functional modules for population dynamics, earnings, earnings-related pensions, national pensions and income taxation. It is also possible to run pension calculation for example cases. The sample size contains 250,000 individuals in the starting year 2008. New 18-years-olds and immigrants are added to the sample annually. A typical simulation run covers years 2009–2080.

The model does not contain any behavioural equations. However, some of the behavioural adjustments resulting from the rising age limits are modelled explicitly in the macro model. The age-gender specific behaviour is obtained from the macro model using a micro-macro link. The differences among different educational groups are estimated from the register data. The ELSI model is described in detail in the report Tikanmäki *et al.* 2014. Since then, we have added new modules for national and guarantee pensions as well as for income taxation. The updated structure of the model is described in appendix A.1.

This paper is mainly based on chapter 6 of the report Kautto and Risku 2015 (in Finnish), with some extensions and updates. Some details of the 2017 pension reform were specified after the report was finished. However, the main lines of the reform and of the results are the same as in that report. In this paper, we have extended our model to national pensions in order to get more accurate results on the impacts on the income distribution of the retired population.

2. BACKGROUND OF THE REFORM

The main objectives for the reform were to increase the employment rate by postponing the retirement and to ease the sustainability gap[3] of the public finances. The reform was also addressed to improve the sustainability of the pension system and adequacy of the pensions as well as to guarantee intergenerational fairness. No other primary goals for income redistributions have been set, but impacts on these will come on the side. Naturally, the labour market organizations that took part in the negotiation had their own goals.

In addition to the Finnish Centre for Pensions (Finnish Centre for Pensions 2014, Kautto and Risku 2015), the reform has been analysed in detail also by the Research Institute of the Finnish Economy (Lassila *et al.* 2015). Their pre-reform study in English on income distribution, public finances and working lives can be found in Lassila and Valkonen 2014.

The sustainability gap (S2 indicator) is estimated to be at around 5 per cent of GDP at an annual level in 2019. Taking into account the pension reform that is scheduled for implementation as of the beginning of 2017 the sustainability gap in public finances will be reduced by approximately one percentage point (Ministry of Finance 2015a).

Working lives have been extended in recent years. Working life is usually analysed in population sub-groups. Recent figures for people retiring on an old-age pension in 2012 show that their median working life spans 38.6 (males 39.4, females 37.7) years (Pension indicators 2014). Regardless of the pension scheme, the median working life of people with a higher educational level is longer than that of people with a lower educational level. It has been estimated that the median working life of the less educated (ISCED 2) retiring on an old age pension is 35.4 years, whereas the figure for those with a higher education level (ISCED 5a) is 36.5 years (Järnefelt *et al.* 2014, p. 47).

When comparing with other countries, we have to rely on Eurostat statistics (Table 1). Working lives are somewhat shorter in Finland than in other Nordic countries. In comparison to Denmark, Sweden and Norway, Finnish working lives are roughly 2–4 years shorter.

Table 1 Average length of expected working life in some European countries, years[4]

Country	Male	Female	Total
EU 28	37.7	32.5	35.2
Denmark	40.3	37.6	39.0
Finland	37.8	36.6	37.2
Sweden	42.1	39.6	40.9
Iceland	47.2	43.8	45.5
Norway	40.6	38.3	39.5

Source: Eurostat

The reform, which will take effect in January 2017, is based on the Agreement on 2017 Earnings-Related Pension Reform (2014) between the central labour market organizations. The government has proposed a bill on the reform which Parliament is scheduled to pass in the fall of 2015 (Prime Minister Sipilä's Government Program 27 May 2015).

3. THE 2017 REFORM IN DETAIL

In the following, we will describe the main changes following the reform. The currently valid rules are described in appendix A.2. The earliest eligibility age for old-age pension will rise by 3 months per birth cohort, from 63 years (1954 birth cohort) to 65 years (1962 birth cohort). From the 1965 birth cohort onward, the earliest eligibility age will be linked to the growth in life expectancy.

The upper age limit for pension accrual is now 68 years. It will be 69 years for birth cohorts 1958–1961 and 70 years as of the 1962 birth cohort. A new target retirement age will be introduced for each birth cohort.[5]

Although the pensionable age will be linked to the growth in life expectancy, the life expectancy coefficient will remain. However, the increase in the pensionable age above 65 years[6] will be taken into account when calculating the coefficient. Thus, the cutting effect of the coefficient will be smaller than it would be without the reform.

Currently, the pension accrual rates are age-specific, so that older employees accrue more pension rights than do younger employees. Currently the income that accrues a pension equals the salary from which the employee's pension contribution has been deducted. As a result of the reform, the accrual rates will be unified to 1.5 per cent, following a transition period. Until the year 2025, the accrual rate for 53–62-year-olds will be 1.7 per cent. In addition, the employee's pension contribution will not be deducted from the pensionable earnings after the reform. Effectively, these two changes combined mean that the accrual rates for the under 53-year-olds will improve while those for the above 53-year-olds will deteriorate.

Currently there is a so-called accelerated accrual rate of 4.5 per cent for those who have deferred their retirement past the earliest eligibility age for old-age pension. This accelerated accrual rate will be replaced by the regular accrual rate of 1.5 per cent, combined with an actuarial deferral rate of 0.4 per cent per month.

From 2017 onward, pension will accrue already as of age 17, (in contrast to the current lower age limit of 18 years).

The disability pensions will improve, because the exit age of the projected pensionable service will be the earliest eligibility age for retirement in the year of occurrence of the disability instead of current 63 years. However, after 2027, the life expectancy coefficient will be applied to the entire disability pension instead of to the part accrued by the time of retirement. This will mitigate the increase of the disability pensions starting from the year 2027.

In addition to the disability pension, a years-of-service pension will be introduced. It may be granted to persons aged 63 who have worked for at least 38 years mostly in strenuous and wearing work. The required reduction in work capacity is rather minor.

The current part-time pension based on the final wage will be replaced by an actuarial partial old-age pension. The age limit for the partial old-age pension will be 61 years for people born in 1963 or earlier and 62 years for those born in 1964 or later. The partial old-age pension can amount to either 25 or 50 per cent of the accrued pension, with an actuarial reduction. The conditions for working less and working at all will be abolished.

According to current legislation, unemployed people born in 1957 or later, whose earnings-related unemployment period (of approximately 2 years) ends at the age of 61 or above, will get additional earnings-related unemployment benefits until they reach the retirement age. This age limit will increase to 62 for people born in 1961 or later.

Some of the amendments included in the reform were suggested by Barr (2013) in his evaluation of the Finnish pension system. These include linking the earliest eligibility age to the growth in life expectancy, allowing for more flexible partial retirement, changing the compensation mechanism for deferring retirement to a more actuarial one and harmonizing the age-specific accrual rates. Barr further suggested that the terminology should be revised so that the earliest eligibility age would not become the social norm of retirement. Introducing the target retirement age is an attempt to fulfil this objective.

In addition to the pension benefit rules, the reform contains several updates to the financing rules of earnings-related pensions. However, as these changes do not have direct effects on the results of this paper, they have been omitted.

4. MODEL ASSUMPTIONS

The impact of the reform is analysed using two different projections. The *baseline* projection is based on the current pension rules. The *reform* projection is a modification of the baseline projection, where we have taken the impacts of the reform into account. The difference between the two projections is interpreted as the impact of the reform. Most of the assumptions are common for both projections.

The base population in ELSI is the socially insured adult population of Finland. The development of the base population follows the official population projection of Statistics Finland until 2060. The most important assumptions of the population projection are the following:

1. the total fertility rate is 1.82,
2. net immigration is 17,000 persons each year,
3. the decreasing mortality trend of 1987–2011 is extended for the projection period.

Age and gender specific retirement rates are expected to develop as in LTP macro model (Finnish Centre for Pensions 2014; Kautto and Risku 2015). In the baseline scenario under current legislation, we have preserved the observed differences in transition probabilities among the educational groups.

Even though we get the age and gender specific retirement rates from the macro model, we have to define explicitly what the differences in transition probabilities among the educational groups are, when the age limits increase. We have decided to use the best available statistical information to do this. We have assumed that, for over 63-year-olds, the differences between educational groups in the retirement rate for disability pension will be the same as for 60–62-year-olds observed in the past. These figures are also applied for the new years-of-service pensions. The same methodology is applied for old-age retirement for over 68-year-olds, where the differences among the groups are estimated from the historical figures for 63–67-year-olds. In practice, these assumptions mean that the higher the education level, the lower the age-specific retirement rate. This phenomenon is highlighted in the case of the disability pension.

The age and gender specific retirement rates for partial old-age pension are also obtained from the LTP macro model. In general, the new partial old-age pension is expected to be much more attractive because of abolished conditions on labour market participation. We have used the historical data of the part-time pension for the differences in the retirement rates among the various educational groups for partial old-age pension.

The long-term inflation assumption is 1.7 per cent, and the wages are expected to grow by 1.6 per cent in real terms. The short-term behaviour follows the short-term economic forecast of the Finnish Centre for Pensions per April 2015.

Our simulation utilizes information on the educational background of individuals. Annual data on individuals' highest education level has been added to the primary data. In the presentation of the results, we have divided the population into two groups: people with a secondary or lower degree (ISCED 0–3) and people with a Bachelor's or higher degree (ISCED 4–6).

5. MICROSIMULATION RESULTS

Most of the results are represented for 75-year-old people who live in Finland and have an earnings-related pension in payment[7]. Unless otherwise explicitly mentioned, we have considered only earnings-related pensions, excluding survivors' pensions. In some cases, we have also included national and guarantee pensions.

When considering the results by educational groups, we have excluded immigrants entering Finland during the simulation period. The information on the immigrants' educational level is not completely comparable to that of the native population. However, immigrants are included in the figures covering the whole population.

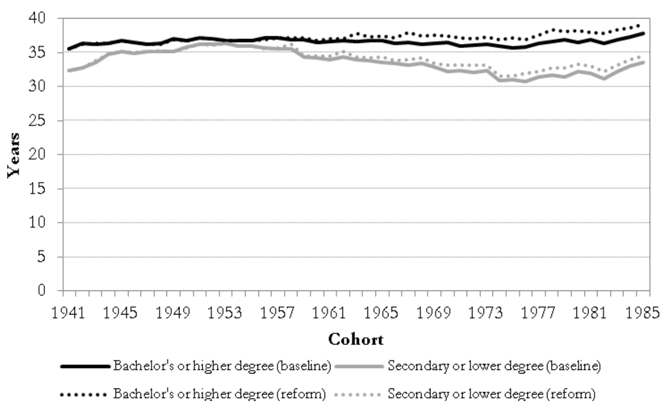
The ELSI model is based on the Monte Carlo simulation. This means that there is some random noise in the results. In addition, in the projection, the age limits are always in full years although, in reality, the age limits increase on a monthly basis. Hence, one should not make too strong statements based on single figures.

5.1. Impact on working lives and time in retirement

In the following, we consider the impact of the reform on the length of working lives in the future. The working life begins at the age of 18 and is based on working days for which pension accrues. The lengths of working lives in the baseline projection are compared to those in the reform projection. The most important factor driving the results is the gradually increasing retirement age. The assumptions regarding the behavioural responses on the rising age limits also have a considerable effect on the results.[8]

We have used register-based information on the insured working history of each individual included in the sample until 2007 and simulated working years as of the year 2008. This means that there are some differences in how working lives are measured before and after the year 2008. Primarily, one should only make conclusions of the effects of the reform. In general, the younger cohorts are better educated than the older ones. This phenomenon has some selection impact between the two educational groups. However, the decreasing trend in the working lives of the lower educational group may be a consequence of the difference in data between the observed past and the simulated future.

Figure 1 Average working life by education, median length in years



As a result of the reform, the working lives will be extended by 1–2 years as of the 1970 birth cohort. The working lives of the highly educated will be extended more than the working lives of those in the lower educational group (Figure 1). The difference in the impact between genders is negligible.

In Table 2 we present the impact of the reform for the lower, median and upper quartiles (25%, 50% and 75% percentage points respectively) and the mean of the length of working life distribution. It turns out that the long working lives are expected to grow slightly more than the shorter ones.

Table 2 Difference in working life of 75-year-olds by gender
Difference of reform to baseline in years

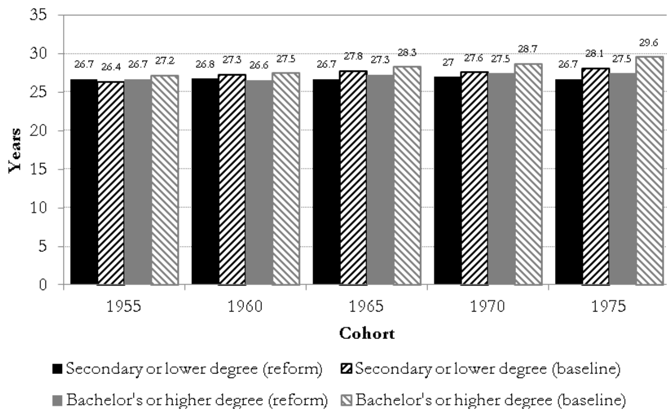
Gender/Cohort	P25%	P50%	P75%	Mean
Male 1960	-0.1	0.2	0.3	0.1
Male 1970	0.5	0.8	1.7	0.9
Male 1980	1.0	0.9	1.2	1.2
Male 1990	1.0	1.0	1.0	1.2
Male 2000	1.0	2.0	2.0	1.4
Female 1960	0.3	0.4	0.7	0.5
Female 1970	0.9	0.9	1.1	1.0
Female 1980	1.1	1.4	1.8	1.5
Female 1990	1.0	1.0	2.0	1.6
Female 2000	1.0	1.0	2.0	1.5

Figure 2 lists the average time spent in retirement by educational groups, in the baseline and the reform projections. One might expect that the group with a higher education would spend more time in retirement. However, the difference is small, especially in the reform projection, because the less educated people tend to retire earlier and to draw disability pensions more frequently. The differences between the educational groups in the baseline projection are in line with the results of Järnefelt *et al.* 2014, p. 114.

The time spent in retirement will be lower for the future pensioner cohorts than it would be without the reform (Figure 2). However, the increasing longevity means that, despite the reform, the time spent in retirement will increase slightly from the present level. The reform will have a greater impact on the groups with a higher education. Those with a lower education are much more prone to draw a disability pension or the forthcoming years-of-service pension.

The gender differences will remain fairly stable after the reform. However, in relative terms, the time males will spend in retirement will be altered more because of their shorter life expectancy.

Figure 2 Average time spent in retirement, by education and in years



In Figure 3, we demonstrate the impacts of the reform for the 1988 birth cohort. For both genders, the average time spent in retirement will be reduced by two years because of the reform. The working lives will be extended by one year. The time spent outside of the labour force will also increase by approximately one year. This outlines the fact that postponing retirement does not have one-to-one correspondence with extending working lives. The earliest eligibility age for old-age pension for this cohort is increased by about four and half years. This means that the working lives are extended by around one fourth of the rise in the earliest pensionable age.

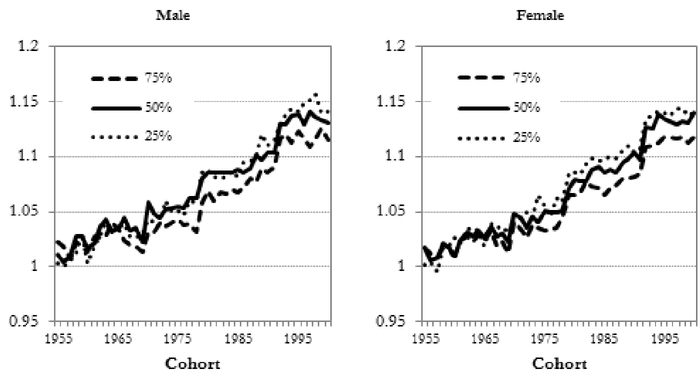
Figure 3 Partition of the life course of the 1988 birth cohort, years on average

It may seem strange that the working lives of females are slightly longer than those of males. This is mainly due to the fact that, for technical reasons, parental leaves are typically included in the length of working lives, at least partly.

5.2. Impact on earnings-related pension levels

In the following we interpret how different percentage points of the earnings-related monthly pensions change as a result of the reform (Figure 4). The different groups are compared by considering the relative change of the median within each group (Figure 5).

Figure 4 Relative changes of the percentage points of the pension levels of 75-year-olds, by birth cohort and gender

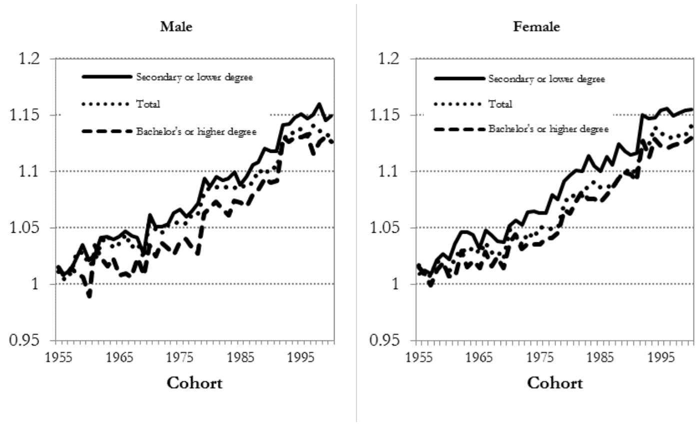


For the pension levels (€/month), see appendix A.3.

It turns out that high pensions (upper quartile) will increase less than will median-sized and small pensions (lower quartile). As of the birth cohort 1975, this difference will be 2–3 percentage points. For earlier cohorts, the difference will be smaller (Figure 4). These differences are mainly a consequence of the fact that the higher accruals at the end of the working lives used to favour high-earners. In this sense, the new accrual rules are more neutral. Improving the disability pensions will also have some impact on improving the lower end of the pension distribution.

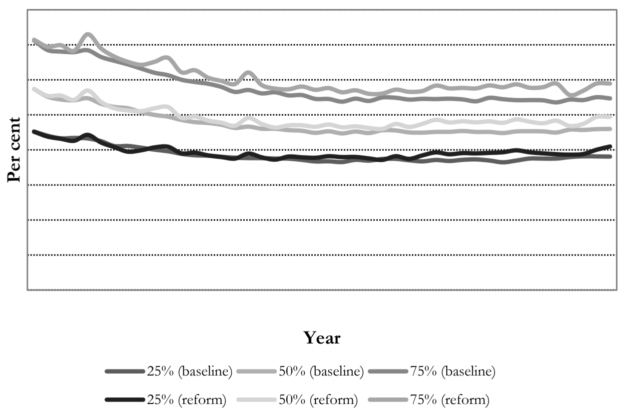
Figure 5 presents the change of the median pension levels in different educational groups. The pension levels will improve in all groups studied. However, the pensions in the lower educational group will increase approximately 2 percentage points more than in the higher educational group.

Figure 5 **Relative changes of the medians of the earnings-related pensions of 75-year-olds, by gender and educational level**



One of the main indicators of the adequacy of pensions is the replacement rate i.e. the fraction between the starting pension and the final wage. We present the development of the replacement rates by calendar years both in the baseline and the reform projections (Figure 6). We show the replacement rates in the entire statutory pension system, including also the national and the guarantee pensions. This extension improves the lowest pensions. However, low pensions may occur in any part of the replacement rate distribution as they also depend on the final wage. We decided to show the replacement rates in the whole statutory pension system in order to facilitate a comparison with other countries.[9]

Figure 6 **Replacement rates of the statutory pensions 2017–2060 of people moving directly from work to old-age-pension, in per cent**



In general, all the curves shown here would be around 1 per cent lower if we only considered earnings-related pensions. The difference is so low because the people retiring from work tend to be the ones with at least a moderate earnings-related pension accrual. If we also considered starting disability pensions, the role of the national pension would be more important.

There is a decreasing trend in replacement rates in the future. There are at least two main reasons. Firstly, the life expectancy coefficient will cut the replacement rates directly if working lives are not extended. The working lives are expected to extend but not enough to compensate the impact of the cutting effect of the life expectancy coefficient. This impact is highly dependent on the assumption on the future demographic development. Secondly, the impact of the transition rule of the 2005 pension reform relating to the old final wage principle will vanish in the long run. The accrual rules applied to pre 2005 earnings are more generous in many typical cases than the present ones. As a result of the forthcoming reform, the replacement rate distribution will stabilize on a higher level than it would have done without the reform.

5.3. Impact on income distribution

We have analysed the reform's effect on income distribution using the Gini coefficient. The Gini coefficient is a widely used measure for income inequality. In this paper the coefficient has been computed for the pensioner population for the baseline and reform scenarios separately. We have analysed the change between these.

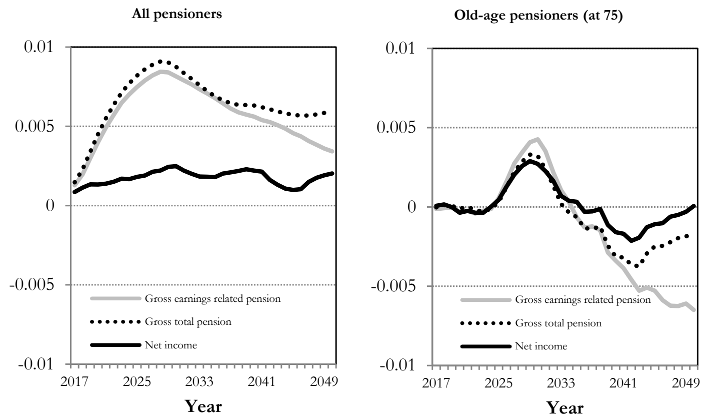
We have computed three different variations on the change of the Gini coefficient (Figure 7). The first one involves earnings-related pensions only, while the second includes all statutory pensions (earnings-related, national and guarantee pensions). The third variation is on the net income of the pensioner population, including also their wage income and income taxation. However, capital income, survivors' pensions and social security benefits (other than pensions) have been excluded.

The figures are computed separately for the whole pensioner population and for 75-year-old pensioners. The latter is done because the age structure of the pensioner population changes over time and the reform affects pensions for different birth cohorts in different ways (Finnish Centre for Pensions 2014).

At this scope the reform will increase the gross pension income differences in the pensioner population to some extent. Because of the growth in wages and the indexing rules, the starting pensions are higher, in general, than the ones that have been paid out for decades. The reform will even improve the starting pensions but will not alter the pensions in payment.

However, taxation and other income will overturn most of this effect since, as a result of the reform, the share of partially retired people with other income sources will grow.

Figure 7 Change in Gini coefficient in 2017–2050



For the 75-year-old pensioners, the picture is slightly different. Short term, before the year 2035 (i.e. those born before 1960), the reform will increase the income differences. After that, the result will be the opposite. This is mainly due to the fact that the disability pensions starting before the reform are in general lower than the ones starting after the reform. For the cohorts born in the 1950's, the disability pensions have typically started already before the reform, but the pensions for these cohorts with extended working lives will improve. When comparing these results to those in Figure 4, please note the different time scale.

Overall, the impacts on the Gini coefficient are rather minor. Taxation mitigates the impact of the changes on the Gini coefficient in all populations studied.

6. CONCLUSIONS

With the help of the microsimulation model, the decision makers had better and more in-depth analyses of the impacts of the proposed reform options. During the negotiations for the pension reform (in 2014), the ELSI model was used in addition to the LTP macro model to provide additional information on the reform proposals. The forthcoming pension reform in 2017 is probably the best studied reform of the Finnish pension scheme in a long time.

As a result of the reform, working lives will be extended. However, the increase in the length of working lives will be less than the increase in the effective retirement age. The earnings-related pension levels will be higher in the future because of the reform. The younger the cohort, the greater is the impact. Within each age and gender group, the earnings-related pensions will increase slightly more in the lower than in the higher education group. In the long run, the lower earnings-related pensions will improve slightly more than the higher ones.

However, when considering the whole pensioner population, the reform will give rise to increased differences in pensions. One reason for this is that the reform will not affect pensions that have already started. Typically, the level of such pensions that have been paid out for long time is lower than average. The reform will even improve new starting pension but, in the long run, the income differences will decrease within each age group.

The results on taxation presented here are preliminary and subject to changes. The Finnish economy has faced hardships since 2008 and public finances are imbalanced. The pension reform plays a part in the balancing of public finances in the long run. However, there is a pressure to increase taxes and widen the tax bases. The pressure to increase municipality tax rates is at least 3 percentage points in 50 per cent of the municipalities (Ministry of Finance 2015b, p. 58). However, the Programme of Prime Minister Sipilä's Government (page 8) states that the total tax rate will not increase during the current term 2015–2019. Some changes in the structure of the taxation are to be expected, however. The Government Program indicates some cuts to the indexing of the pensions. On the other hand, the level of the guarantee pension is subject to increase slightly. We have not taken the changes stated in the Government Program into account in the simulation results presented in this article.

REFERENCES

Agreement on 2017 Earnings-related Pension Reform (2014). Available at http://www.etk.fi/wp-content/uploads/agreement_on_2017_earnings_related_pension_reform_final.pdf

Andersen T B, Määttänen N and Valkonen T (2014). Pension reforms: Longevity and retirement. In Valkonen T and Vihriälä V (eds.) (2014). The Nordic model – challenged but capable of reform. Nordic Council of Ministers 2014, TemaNord 2014:531.

Barr N (2013). The pension system in Finland: Adequacy, sustainability and system design. Finnish Centre for Pensions, Evaluation of the Finnish Pension System / Part 1.

European Commission (2015). The 2015 Ageing Report: Economic and budgetary projections for the 28 EU Member States (2013−2060).

Eurostat (2015). Duration of working life. Available at http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdde420&plugin=1

Finland, a land of Solutions. Strategic Programme of Prime Minister Juha Sipilä's Government, 29 May 2015. http://valtioneuvosto.fi/documents/10184/1427398/Ratkaisujen+Suomi_EN_YHDISTETTY_netti.pdf/8d2e1a66-e24a-4073-8303-ee3127fbfcac

Finnish Centre for Pensions (2014). Preliminary consequence analysis of the results of the 2017 pension reform negotiations. Available at http://www.etk.fi/wp-content/uploads/etk_vaikutusarvio_26092104_hn2_en-1.pdf

Hietaniemi M, Risku I, Ritola S, Salonen J and Uusitalo H (2006). How do earnings-related pensions develop? Calculations until 2050. Finnish Centre for Pensions, Reports 2006:1.

Järnefelt N, Nivalainen S, Salokangas S and Uusitalo H (2014). Sosioekonomiset erot – työurat, eläkkeelle siirtyminen ja eläkejärjestelmä (In Finnish). Finnish Centre for Pensions, reports 01/2014.

Kautto M and Risku I (2015). Laskelmia vuoden 2017 työeläkeuudistuksen vaikutuksista (In Finnish). Finnish Centre for Pensions, reports 02/2015.

Lassila J, Määttänen N and Valkonen T (2015). Työeläkeuudistus 2017: vaikutukset työuriin, tulonjakoon ja julkisen talouden kestävyyteen (In Finnish). Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 1/2015.

Lassila J and Valkonen T (2007). The Finnish Pension Reform of 2005. The Geneva Papers on Risk and Insurance. Issues and Practice. Vol. 32, No. 1, January 2007.

Lassila J and Valkonen T (2014). Longevity, Working Lives and Public Finances. ETLA Working Papers no 24, 9 April 2014. The Research Institute of the Finnish Economy.

Ministry of Finance (2015a). Economic Survey, Spring 2015. Ministry of Finance publications – 17b/2015.

Ministry of Finance (2015b). Kuntatalousohjelma 2016–2019 (In Finnish). Valtiovarainministeriön julkaisu – 16a/2015.

Pension indicators (2014). Finnish Centre for Pensions, Reviews 02/2014.

Risku I, Appelqvist J, Sankala M, Sihvonen H, Tikanmäki H and Vaittinen R (2014). Statutory pensions in Finland - long-term projections 2013. Finnish Centre for Pensions, Reports 03/2014.

Statistics Finland (2012). Population projection 2012−2060. http://www.stat.fi

Statistics Finland (2014). Population structure [e-publication]. ISSN=1797-5395. Helsinki: Statistics Finland [referred: 14.6.2015]. Access method: http://tilastokeskus.fi/til/vaerak/tau_en.html

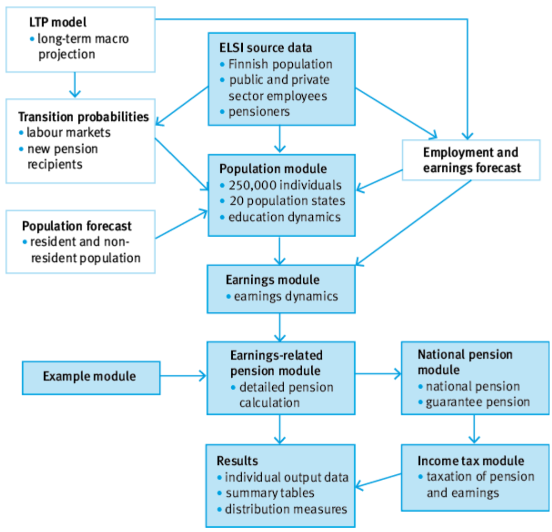
Tikanmäki H, Sihvonen H and Salonen J (2014). Microsimulating Finnish earnings-related pensions. Finnish Centre for Pensions, Working Papers 02/2014.

APPENDIX

A.1 The model structure

The model structure is described in detail in Tikanmäki *et al.* (2014). Since then, we have added modules for the national and the guarantee pension, as well as for income taxation. The updated structure of the model is described in what follows.

Figure A.1 **Structure of the ELSI model. Each dark box represents a module of the model. Each white box represents an external source of information**



The national pension module is used for calculating the national pension and the guarantee pension for people whose earnings-related pensions are small or who receive no earnings-related pension at all. The national pension consists of the old-age and the disability pension. The amount of the national pension is reduced by the amount of the earnings-related pension. The national pension is smaller for those living with a spouse or partner than for those who are living alone. In the ELSI model, the marital status is stochastic so that, on average, the share of people living alone stays at the same level for each gender and age group as in the years 2008–2013. The length of time lived abroad also affects the level of the national pension. The minimum requirement is that the individual has resided in Finland for at least three years after the age of 16. The national pension is paid abroad if Finland has a social security agreement with the country in question. In addition, simulation assumptions on insurance periods in Finland and social security agreement countries greatly affected the results. In 2014, the full monthly national pension for a person living alone in Finland was approximately EUR 630 (the average monthly total pension was ca EUR 1,590 and the average monthly wage EUR 3,300).

The Act on Guarantee Pension came into force in 2011. Its purpose is to provide a minimum pension level for people living in Finland. The guarantee pension is not paid abroad. It is granted to people aged 65 or above or as a disability allowance as of the age of 16 years. The minimum residence requirement in Finland is three years. The amount of the earnings-related pension and the national pension reduce the amount of the guarantee pension. In 2014, the full monthly guarantee pension was approximately EUR 750, providing the person does not receive any other pension income.

In the microsimulation model we have not simulated survivors' pensions and pensions paid from abroad. These factors would reduce the national and the guarantee pension. As a result of the pension reform in 2017, the earliest eligibility age for the national and the guarantee pension will rise gradually.

From a technical point of view, there were some challenges in the national pension module. In the lower income groups, the assumptions on times resided in Finland had great influence on the level of the national pension and thus the total pension level. In the long run, it is hard to estimate the proper level for the national pension.

The income tax module is used for calculating the direct taxation of wages and pension income. As a result, we have obtained the net income for each person each year. In the taxation module we use the official parameters for income taxation for the past years and the present year. These include, for example, income tax rates for different income levels and reductions. For the simulation years, we have indexed parameters using the income level index so that the future taxation level will be the same as the current with respect to the wage level. The module is purely deterministic.

The tax module includes the state tax and the municipality tax. The church tax is not included in the income tax module. Members of the Evangelical Lutheran Church or Orthodox Church pay an average monthly church tax of 1.43 per cent of the wages). In practice, 73 per cent of the Finnish population belongs to the Evangelical Lutheran Church. Among the older cohorts (65+ yrs.), 82 per cent are enrolled with the Church (Statistics Finland 2014).

A.2 Current earnings-related pension rules in a nutshell

Accrual rules. Earnings-related pension accrues from income earned between the ages of 18 and 67 in accordance with the following accrual rates:

1. 1.5 per cent of earnings for people between 18 and 52 years old.
2. 1.9 per cent of earnings for people between 53 and 62 years old.
3. 4.5 per cent of earnings for people between 63 and 67 years old.
4. 1.5 per cent of earnings during retirement.
5. 1.5 per cent of earnings for projected pensionable service[10]
6. 1.5 per cent of certain social security benefits[11].

Persons under the age of 18 and persons aged 68 or above do not accrue a pension, nor do they fall under the insurance obligation. The income that accrues a pension equals the salary from which the employee's pension contribution has been deducted[12]. Pension accrues during the following social benefit periods: earnings-related unemployment allowance and labour market training periods, parental allowance, sickness allowance and job alternation leave periods. Earnings-related pension also accrues from a few other benefit periods that are less significant from the point of view of pension expenditure.

Regardless of age, the accrual rate for social benefit periods is 1.5 per cent per year. The accrual is based on the same earnings that the actual benefit is based on. For the parental allowance, the basis for the pension is 117 per cent of the earnings; for earnings-related unemployment benefits, the percentage is 75, and for other types of daily allowance, except job alternation leave, the percentage is 65. For job alternation leave, the basis for the pension is 55 per cent of the earnings.

Pension accrues also from studies leading to a vocational or university-level degree, as well as from child care at home for children under the age of three. This accrued pension is calculated using an earnings base of EUR 676 per month, at 2012 prices, with an annual accrual rate of 1.5 per cent. The earnings base is indexed to the wage coefficient.

Indexing. When calculating the initial pension amount, the income from different years is adjusted using the wage coefficient, which is a weighted average equal to 80 per cent of wage changes and 20 per cent of price changes. Pensions in payment are adjusted using an earnings-related pension index, which is a weighted average equal to 20 per cent of wage changes and 80 per cent of price changes. A one-time raise in pension is carried out for young and middle-aged disability pensioners after the pension has been paid for five years. The increase is 25 per cent for pensioners under the age of 32. For those over 32, the increase will be lowered by one percentage point for each year of age, until it ceases altogether.

Pension benefits. Earnings-related pension benefits include disability, part-time, old-age and survivors' pensions.[13]

Disability pension. Disability pension can be granted either as a full pension or a partial pension, depending on the degree to which the work ability of the insured has decreased. The partial disability pension is equal to half of the full disability pension. The disability pension is equal to the pension accrued up to the date on which the disability occurred, plus an additional amount based on projected pensionable service. Projected pensionable service is calculated from the time of the pension contingency until the time of turning 63. The annual accrual rate applied to the projected pensionable service is 1.5 per cent, and the salary applied is the average salary computed over the five years preceding the occurrence of the disability. The life expectancy coefficient affects the starting amount of the disability pension as explained later in this appendix.

Part-time pension. A part-time pension may be granted to insured persons who reduce their working hours in such a manner that the earnings decrease to 35−70 per cent of their stabilised earnings level. The age limit for a part-time pension is 58 years for those born in 1952 or earlier; 60 years for those born in 1953 and 61 years for those born in 1954 or later. The size of the part-time pension is half of the earnings reduction caused by the decrease in working hours. Pension is accrued in a regular manner from work carried out during part-time pension. This pension benefit will be discontinued as a result of the reform.

Old-age pension. The insured are entitled to an old-age pension at the age of 63. In some special cases, the age limit of old-age pension may be lower than 63 years. If the insured continues to work after turning 63 and do not take out old-age pension, the pension accrual rate is 4.5 per cent per year. However, for those who engage in gainful employment while receiving a pension, the accrual rate is 1.5 per cent per year. After reaching the age of 68, the pension accrual and the insurance obligation end. If the insured do not take out their old-age pension after turning 68, an increment for deferred retirement of 0.4 per cent per month is added to the pension.

Life expectancy coefficient. Using the life expectancy coefficient, the initial amount of old-age pension and disability pension is adjusted to reflect changes in life expectancy at age 62. The starting amount of old-age pension is determined by multiplying the accrued pension by the life expectancy coefficient. For starting disability pensions, the accrued pension is also multiplied by the life expectancy coefficient. However, the coefficient is not applied to the projected pensionable service. Hence, the closer to the age of old-age pension the individual is when becoming disabled, the larger is the effect of the life expectancy coefficient on the size of the disability pension. The life expectancy coefficient is defined in such a way that the capital value (actuarial present value) of the old-age pension remains unchanged, even if the mortality of those of a pensionable age were to differ from the mortality observed in the years 2003−2007. The life expectancy coefficient affects the pensions of those born in 1948 and later. The value of the coefficient is determined separately for each birth cohort.

A.3 Earnings related pension by education

Table A.3 Earnings related pension by education, median EUR/month at 2014 prices*

Cohort	Male			Female		
	25% (baseline)	50% (baseline)	75% (baseline)	25% (baseline)	50% (baseline)	75% (baseline)
1960	1,240	1,790	2,470	1,050	1,500	1,960
1970	1,440	1,960	2,650	1,220	1,640	2,140
1980	1,640	2,180	2,900	1,470	1,950	2,610
1990	1,930	2,510	3,350	1,700	2,210	2,880
2000	2,350	3,010	3,960	2,050	2,610	3,400
	25% (reform)	50% (reform)	75% (reform)	25% (reform)	50% (reform)	75% (reform)
1960	1,250	1,820	2,500	1,080	1,520	1,970
1970	1,490	2,070	2,740	1,270	1,720	2,230
1980	1,780	2,370	3,110	1,590	2,110	2,780
1990	2,150	2,770	3,640	1,880	2,430	3,110
2000	2,680	3,410	4,420	2,330	2,980	3,800
	Low (baseline)	High (baseline)	Total (baseline)	Low (baseline)	High (baseline)	Total (baseline)
1960	1,550	2,610	1,790	1,230	1,890	1,500
1970	1,670	2,780	1,960	1,270	2,000	1,640
1980	1,960	3,170	2,180	1,550	2,500	1,950
1990	2,250	3,670	2,510	1,820	2,720	2,210
2000	2,740	4,340	3,010	2,190	3,250	2,610
	Low (reform)	High (reform)	Total (reform)	Low (reform)	High (reform)	Total (reform)
1960	1,590	2,590	1,820	1,250	1,900	1,520
1970	1,780	2,870	2,070	1,340	2,080	1,720
1980	2,130	3,390	2,370	1,700	2,660	2,110
1990	2,520	4,002	2,770	2,030	2,980	2,430
2000	3,150	4,880	3,410	2,530	3,680	2,980

* Low=Secondary or lower degree, High=Bachelor's or higher degree.

1 In some sense, the part time pension is replaced by the partial old-age pension. The partial old-age pension is considered to be more actuarial and less bureaucratic. The main differences of the benefits are the following. The old partial pension can be granted for employees who have agreed with the employers to reduce their working hours. The new partial old-age pension can be granted to any insured person working full time, part-time or not working at all. The amount of partial pension is based on the final wage. Taking the partial pension does not reduce the old-age pension later. The amount of the partial old-age pension is based on the accrued pension with actuarial reduction.

2 The macro model is the same that is used for producing the results on the Finnish earnings-related pension system of the AWG report (European Commission 2015). The impact of the reform is not present in the AWG results.

3 The sustainability gap refers to the additional financing (in relation to GDP) needed to balance the public sector in the long run.

4 The duration of the working life indicator measures the number of years a person aged 15 is expected to be active in the labor market throughout his/her life.

5 The target retirement age is higher than the earliest eligibility age. It is a psychological target rather than a parameter of the pension system or an economical incentive.

6 The increase of the earliest eligibility age from 63 to 65 is not taken into account when redefining the life expectancy coefficient.

7 75-year-olds are all retired in both the baseline and the reform projections.

8 In the projection, the target retirement age is not assumed to have any impact on the retirement decisions.

9 According to the legislation, national and guarantee pensions are indexed using the consumer price index. However, in the past, there have been decision-based improvements (and freezings) to this indexation. In the long run, we have assumed that the indexing follows price changes to 50 per cent and wage changes to 50 per cent.

10 The accrual for the projected pensionable service is calculated from the time of the pension contingency to the time of turning 63 years.

11 The basis for the accrual is a per-benefit-specified share of the earnings prior to the benefit period.

12 Exceptions for the self-employed, farmers and seafarers.

13 Survivor's pensions are not subject to the pension reform and thus not covered in this paper.

PUBLICATION IV

Using trajectory analysis to test and illustrate microsimulation outcomes

Salonen Janne, Tikanmäki Heikki & Nummi Tapio

International journal of microsimulation 12(2).

(https://doi.org/10.34196/ijm.00198)

International Microsimulation Association

RESEARCH ARTICLE

Using trajectory analysis to test and illustrate microsimulation outcomes

Janne Salonen[1]*, Heikki Tikanmäki[1], Tapio Nummi[2]

[1]Finnish Centre for Pensions, Helsinki, Finland; [2]University of Tampere, Tampere, Finland

Abstract We propose a new data-driven way of testing and visualizing dynamic microsimulation outcome data. The proposed statistical methodology is based on trajectory analysis (***Nagin, 1999***), which can be used to identify several sub-populations from a population measured longitudinally. We briefly introduce the statistical basis of trajectory analysis and discuss its use in the context of microsimulation. Finally, we report our results from the Finnish microsimulation model ELSI (***Tikanmäki et al., 2014***; ***Tikanmäki et al., 2015***) to illustrate the possibilities and benefits of this technique. Trajectory analysis is available in many statistical software packages (e.g., SAS, R, Stata and Mplus). We conclude that trajectory analysis is a useful tool for investigating microsimulation outcomes.
JEL classification: C63, C10, H55
DOI: https://doi.org/10.34196/ijm.00198

1. Introduction

This paper has two goals. First, we present a statistical technique called group-based trajectory analysis and demonstrate its usefulness in a microsimulation context. We use trajectory analysis to identify unknown groups or sub-populations that can yield valuable information that is not necessarily easily accessible by other means. Second, we use three examples to show how this technique, among others, can be used to validate models and to reveal possible misspecification of the microsimulation model.

Trajectory analysis has recently gained much popularity in a number of fields including psychology, criminology (***Nagin and Odgers, 2010***; ***Nagin, 2016***; ***van der Geest et al., 2016***), sociology (***Hynes and Clarkberg, 2005***; ***Don and Mickelson, 2014***), education (***Kokko et al., 2008***), marketing (***Mani and Nandkumar, 2016***) and health sciences (***Nummi et al., 2014***; ***Nummi et al., 2017a***). It has increasingly been used in studies on labor market attachment (***Peutere et al., 2015***; ***Nummi et al., 2017b***). To our knowledge, trajectory analysis has very seldom been used in a microsimulation context.

Dynamic microsimulation results are often reported using some ex ante classification (e.g., education, age, or labor market state). The trajectory approach offers key advantages over ex ante classification, which stem from the a priori use of an economic taxonomy. The basic advantage of trajectory analysis is that it can reveal latent patterns in longitudinal data that might otherwise remain hidden, hence complementing ex ante classifications. Furthermore, the use of a formal statistical methodology has the capacity to distinguish chance variation across individuals from real differences caused by latent sub-groups (***Nagin and Odgers, 2010***).

In this paper, we propose two kind of trajectory analysis implementations. First, the data can be stratified by various background factors (e.g., by labor market state) and thus the latent sub-groups are to be found within the strata. This would yield information on, for example, how common the earnings trajectories (by population state) are in the stratas investigated. Second, another way would be to classify the trajectory groups by some background factor after trajectory analysis. This would yield information on how the ex ante classifier (e.g., labor market state) is divided in trajectory groups.

***For correspondence:** janne.salonen@etk.fi

Author Keywords: finite mixtures, trajectory analysis, group-based modeling, dynamic microsimulation

DOI: https://doi.org/10.34196/ijm.00198

Population heterogeneity is an important topic in dynamic microsimulation. Credible heterogeneity in a simulated population is a desired property of any microsimulation model. Trajectory analysis is a powerful method for demonstrating this heterogeneity. It allows for the analysis of any microsimulation outcome with interesting population heterogeneity. For instance, in the Finnish ELSI model we could analyze the individual population state, pension contributions, working lives or incomes. The technique is especially suited to cohort-based analysis, but it can also handle several cohorts simultaneously.

The approach we propose has a number of advantages. First, trajectory analysis has the potential to reveal interesting sub-groups of individuals. Second, the technique of trajectory analysis with normal distribution (like other members of the exponential family) is a well-established method, with software packages readily available for microsimulation practitioners (e.g. ***Jones et al., 2001***; ***Leisch, 2004***; ***Grun and Leisch, 2007***; ***Haughton et al., 2009***; ***Muthén and Muthén, 2010***). Third, trajectory analysis is a flexible method that can be applied to both cross-sectional data (single period measurements), and more importantly, to longitudinal data (multiple periods of measurements). The most common metric for indexing time is age or year. Another possible metric would be time before or since a life-event.

Descriptive analysis is often inadequate for the purposes of empirical research. Trajectory analysis can provide a more rigorous examination with time-dependent covariates or risk factors affecting trajectory group membership (***Nagin, 2005***; ***Jones and Nagin, 2007***). In addition to single outcome analysis, trajectory analysis also allows for the simultaneous analysis of multiple outcomes ***Nagin et al., 2016***). Indeed, there is now a growing body of research that uses multiple trajectory analysis (e.g. ***Hsu, 2015***; ***Nummi et al., 2017b***). ***Nummi et al. (2017b)*** provide an example of the simultaneous modeling of employment, education, unemployment and parental leaves using a multivariate trajectory model. These abovementioned extensions also provide interesting possibilities for microsimulation data analysis.

The method has also its disadvantages. First, trajectory analysis can only be performed with discrete-time data. Most dynamic microsimulation models are defined in discrete-time, with time-point intervals of one year (***Zaidi and Rake, 2001***; ***Li and O'Donoghue, 2013***; ***Li et al., 2014***). This leads to a second disadvantage, which can be described as state vs. event microsimulation. Trajectory analysis is easiest to implement with the calendar year as the time interval, that is, using state microsimulation models.

Trajectory analysis is often applied to normally distributed data, but it is also applicable to discrete distributions such as binomial, Poisson, multinomial, etc., making this method a useful tool for the exploratory analysis of data sets. In this paper, we show three examples of trajectory analysis in microsimulation contexts. These examples cover labor market simulation topics such as wage earnings, education and pensions. Trajectory analysis of these outcomes illustrates the technique with different data distributions.

2. Trajectory analysis in microsimulation

2.1. Trajectory analysis

Trajectory analysis is, in essence, the application of finite mixture modeling to longitudinal data. It can be used for modeling the unobserved heterogeneity of individuals measured longitudinally (e.g. ***Nagin, 1999***; ***Nagin, 2005***). In what follows we describe the statistical background for our application of trajectory analysis.

The statistical foundation of trajectory analysis is on finite mixture modeling. Well-established in the field of statistics, the mixture modeling method concerns modeling a statistical distribution by a mixture (or weighted sum) of distributions (see ***Titterington et al., 1985***; ***McLachlan and Peel, 2000***). ***Böhning et al. (2007)*** give examples of the use of finite mixture modeling in various statistical applications.

Two basic modeling approaches are called growth mixture modeling and group-based trajectory modeling. They share the same analytical objective of measuring and explaining differences across population members in their developmental course. The difference between approaches lies in the way they model individual-level heterogenity in developmental trajectories.

Growth mixture models depict the average trend of outcome and individual-specific variation around the average trend with random effects using the same parameters of change (***Nagin and Odgers, 2010***).

Our specific aim is to identify individuals or objects in microsimulation data with the same kind of unknown developmental profiles (trajectories or sub-groups). The microsimulation population is then splitted into several sub-populations. Let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$ represent the sequence of measurements on an individual *i* over *T* periods and let $f_i(\boldsymbol{y}_i|\boldsymbol{X}_i)$ denote the marginal probability distribution of $\boldsymbol{y}_i$ with possible time-dependent covariates $\boldsymbol{X}_i$. It is assumed that $f_i(\boldsymbol{y}_i|\boldsymbol{X}_i)$ follows a mixture of *K* densities

$$f_i(\boldsymbol{y}_i|\boldsymbol{X}_i) = \sum_{k=1}^{K} \pi_k f_{ik}(\boldsymbol{y}_i|\boldsymbol{X}_i), \quad \sum_{k=1}^{K} \pi_k = 1 \; with \; \pi_k > 0, \tag{1}$$

where π_k is the probability of belonging to the sub-group *k* and $f_{ik}(\boldsymbol{y}_i|\boldsymbol{X}_i)$ is the density for the *k*th sub-group. Trajectory analysis can handle discrete or continuous data. The simplest choice is to use the Bernoulli distribution {0, 1} for the mixture components $f_{ik}(\boldsymbol{y}_i|\boldsymbol{X}_i)$: but other members of the exponential family are often applied as well. It is assumed that given *k*th sub-group measurements are independent. For Bernoulli mixtures we can write

$$f_{ik}(\boldsymbol{y}_i|\boldsymbol{X}_i) = \prod_{t=1}^{T} p_{itk}^{y_{it}} \left(1 - p_{itk}^{1-y_{it}}\right), \tag{2}$$

where the probability p_{itk} is a function of covariates $\boldsymbol{X}_i$. For modeling the conditional distribution of p_{itk} we use the logistic regression model. For the *i*th individual, we can then use the equation

$$p_{itk} = \frac{\exp\left(\boldsymbol{x}_i' \beta_k\right)}{1+exp\left(\boldsymbol{x}_i' \beta_k\right)}, \tag{3}$$

where $\boldsymbol{x}_i'(t)$ is the *t*th row of $\boldsymbol{X}_i$ and $\boldsymbol{\beta}_k$ is the parameter vector for the *k*th sub-group. For the analysis of continuous data, one alternative is the multivariate normal distribution

$$f_{ik}(\boldsymbol{y}_i|\boldsymbol{X}_i) = (2\pi)^{-\frac{p_i}{2}} |\Sigma_{ik}|^{-\frac{p_i}{2}} exp\left\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{ik})' \Sigma_{ik}^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_{ik})\right\}, \tag{4}$$

where $\boldsymbol{\mu}_{ik}$ is a function of covariates $\boldsymbol{X}_i$ with parameters $\boldsymbol{\beta}_k$ and $\Sigma_{ik} = \sigma_k^2 \boldsymbol{I}$ a covariance matrix within *k*th sub-group. Thus, the measurements are assumed to be independent within sub-group *k*, with the variance σ_k^2 . One advantage of this assumption is that it considerably simplifies the likelihood function and thus yields a computationally lighter and more stable analysis.

For modeling the trajectory mean in time *t*, simple linear models are usually applied, e.g. low-degree polynomials. For our three examples we used cubic polynomial model

$$\boldsymbol{x}'\boldsymbol{\beta}_k = \beta_{0k} + \beta_{1k}t + \beta_{2k}t^2 + \beta_{3k}t^3 \tag{5}$$

to model the development within the sub-group *k* in time (age). The mean model can also include other time-dependent covariates or time-stable covariates (risk-factors).

For parameter estimation, Maximum Likelihood (ML) estimates can be calculated by maximizing the log-likelihood $\sum_{i=1}^{N} f_i$ over unknown parameters $\boldsymbol{\beta}_{ik}$ and σ_k, $k = 1, \ldots, K$ (see e.g. ***Nagin, 1999***; ***Jones et al., 2001***; ***Jones and Nagin, 2007***). In most software packages, the method used for ML estimation is the EM (Expectation and Maximization) algorithm (see ***Dempster et al., 1977***; ***McLachlan and Peel, 2000***). The algorithm in an iterative technique involving two steps. E step finds the expected log likelihood under current parameter estimates, the subsequent M step maximizes the expected log likelihood function. These steps are iterated until the estimates converge. When applied to trajectory analysis, the E step calculates the posterior probability for sub-group membership

$$w_{ik} = \frac{\pi_k f_k(y_i|X_i, \varnothing_k)}{f(y_i|X_i, \varnothing_k)} \tag{6}$$

under all parameter estimates $\widehat{\varnothing}_k$. The estimated sub-group probabilities are then

$$\hat{\pi}_k = \frac{1}{N}\sum_{i=1}^{N}\hat{w}_{ik} \tag{7}$$

Once the model parameters have been estimated, the posterior probability estimates provide a way to assign each individual to a specific sub-group or trajectory. Individuals can be assigned to the specific sub-group in which their posterior probability is highest. From the equations, we see that the assignment of individuals into groups takes into account regression parameter estimates, the number of groups, probability distribution, membership probability estimates and time span of longitudinal dataset. These are major topics in trajectory analysis and need careful consideration from microsimulation practitioner. ***Nagin (2005)*** discusses the selection of the number of groups (p. 78–87) and related statistical information criteria (p. 63–76) and the question of conditional independence (p. 26–27).

The question of groups as real entities is discussed in ***Nagin and Odgers (2010)*** and ***Nagin (2016)***. ***Nagin (1999)*** discusses the use of group membership probabilities in the calculations, as well as the links between group membership probabilities and other time-dependent covariates besides age (or time). The nature of trajectory groups in growth mixture modeling and group-based trajectory analysis is discussed in ***Nagin, 2005***, pp. 54–56), ***Nagin (2016)*** and ***Nathalie et al. (2017)***. ***Don and Mickelson (2014)*** discuss the good practices of trajectory analysis, especially in terms of model selection.

Note that the sub-groups revealed by trajectory analysis are not fixed constructs. They are just approximations of a more complex reality. Each individual belong to a specific group with a certain probability. It would be good if we could investigate the fit of the assumed model using the identified trajectory groups. However, since the groups are based on maximum posterior probability, this would probably introduce some correlation to within-group residuals, as the actual group could also contain individuals from other groups with a small probability or weight. The correct approach from a statistical point of view, however, would be based on the investigation, using alternative covariance structures (for example, likelihood-ratio type test statistics or some other information criterion). However, this kind of testing is impossible in Nagin's basic model. This topic has been discussed in ***Nagin and Odgers (2010)*** and ***Nagin (2016)***.

As with any statistical analysis of dynamic microsimulation outcome data, a change in the period of observation would naturally change the results of the model. The trajectory groups could also be affected to a certain extent. However, we believe that in our examples the basic main grouping structure would remain quite similar.

2.2. Finnish ELSI microsimulation model

ELSI is a longitudinal microsimulation model (***Tikanmäki et al., 2014***; ***Dekkers and Van den Bosch, 2016***) that is used to assess the development of the statutory pensions in Finland. The dynamic ageing model has been developed at the Finnish Centre for Pensions, a statutory co-operation body providing research and expertise services related to the Finnish pension system.

ELSI model has been designed to assess the future earnings-related pensions and the national pensions (***Tikanmäki et al., 2017***). It can also be used to analyze changes in the pension system and in the underlying demographic or macroeconomic conditions. One of its uses has been to assess the distributional effects of the pension reform of 2017 (***Tikanmäki et al., 2015***).

The Finnish pension system is mainly based on pension rights accrued on the basis of the individual's life-time earnings. The only exception is the survivor's pension, which accounts for no more than 6% of total pension expenditure. The ELSI model is therefore based on individual-level information and calculations of pensions received in one's own right. The model comprises both pension recipients and those still working. The model simulates each individual's working life prior to retirement.

The base population consists of all adults aged 18 or over covered by the social insurance system in Finland. Most of the material is drawn from administrative records maintained by the Finnish Centre for Pensions and the Social Insurance Institution of Finland. Information on educational level from Statistics Finland is also added to the model. The ELSI model provides the opportunity to run large populations. For the present analysis we used a random sample of 32% of the base population, which in numerical terms translates into 1.5 million individuals in the baseline year of 2012. The population is then simulated until 2085. Deceased people remain in the model and a new cohort as well as new immigrants enter the model each year. Consequently the population increases over the course of the simulation.

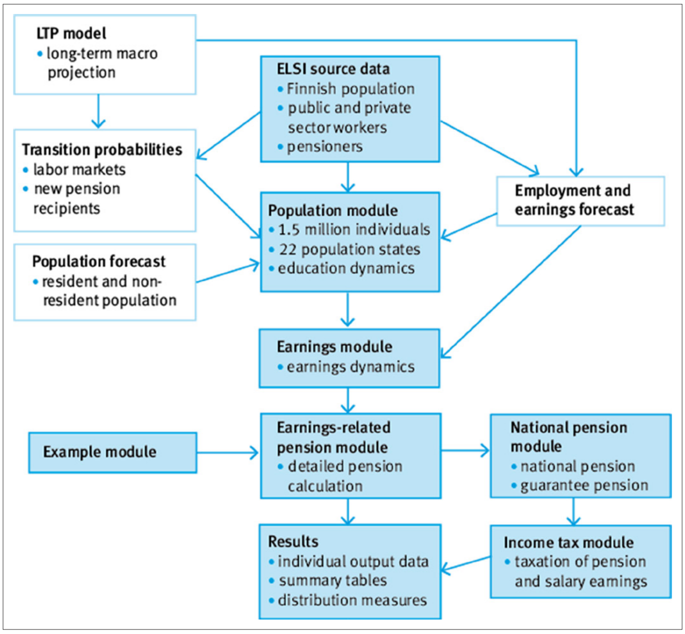


Figure 1. Structure of the ELSI microsimulation model

The ELSI model has a modular structure (***Figure 1***). Each colored box represents a module of the model, while white boxes represent external sources of information. There are no feedback loops from later to earlier modules. The simulation starts from the population module, followed by the earnings module, pension modules, and the taxation module, and finally brings together the results.

The population module has several functions. It simulates population and labor market transitions as well as educational changes. The population module is based on transition probabilities that are estimated from historical data for 2010–2014. The module also uses Statistics Finland's official population forecast to replicate general trends in the sample population. Transition probabilities, with one-year time steps, are by and large deterministic, that is, based on exogenous information.

In the population module, we simulate a new population or labor market state for each individual based on transition probabilities. There are 22 states in the model, the most important of which are: active (employed), inactive, unemployed, various pension states for different pension benefits, and deceased.

The labor market transitions are of Markovian type, which means that the transition probabilities are based on current state rather than former history. However, it is possible to add memory to a Markov process by extending the state space. For instance, in model ELSI we have three different active states. One for those employed first consecutive year, another for those employed second year and third one for the rest. Hence, unemployment risks may be higher for those who do not yet have an established labor market position.

Education is also simulated in the population module. Post-basic education dynamics is based on age and gender-specific transition probabilities. Changes in education level are possible at any age, although they are not very common after age 35.

The transition probabilities are updated each simulation year using the population level information produced by the semi-aggregated LTP model (***Tikanmäki et al., 2017***). There is thus a simple alignment of microsimulation outcomes to macro level aggregates.

Wage earnings are simulated in the earnings module. Wages are simulated for each individual annually based on the labor market state and an underlying earnings-equation, which is a time-series model with a stochastic component. The earnings equation takes also into account gender, age and level of education. Wages are simulated for active workers and those in partial retirement. The earnings module is described in ***Tikanmäki et al. (2014)***.

The earnings-related pension module calculates pension amounts based on the simulated time of retirement (retirement age) and life-course earnings. The pension calculation takes no shortcuts but is as detailed as possible given the data in use. The national pension and guarantee pension amounts are calculated as a residual of the earnings-related pension, since they are income tested. The calculation of national pensions is described in ***Sihvonen (2015)***.

The taxation module finalizes the substantive simulation. The previous modules have produced gross wages and pensions. In the taxation module, the current (2016) rules for income taxation are applied to both simulated wages and pension earnings. The calculation of income taxes is described in ***Sihvonen (2015)***.

After the simulation run, the results are analyzed in the results module, which calculates aggregate results over the course of the simulation, based on individual-level outcomes. Measures of the distribution (mean, percentage points, Gini coefficient) of pensions can be produced by ex ante classifiers such as gender, level of education and year of birth. Aggregate measures on the duration of working life and partition of life-course into active and passive stages are also calculated in the result module. The module collects individual-level output data containing information on labor market state, wage earnings, residence, education level, pension earnings, pension benefit, working life, pension accrual, etc. Therefore the material is also available for the statistical analysis illustrated in this paper. The proposed statistical technique could be used with many other outcomes as well.

In the following analysis we use a longitudinal individual-level data set that covers the period from 2008–2085 with a 25% sub-sample of the original output data. The analysis is based on three outcomes: wage earnings, education level and pension earnings. The wage and education trajectories are presented for the male cohort born in 1995 and pension trajectories for the male cohort born in 1960. With trajectory analysis, it would be possible to analyze several cohorts at the same time so long as the data is longitudinal.

Similar longitudinal data is produced in many other dynamic microsimulation models, too. For example, the Swedish SESIM model and the Norwegian MOSART model are in this respect similar to ELSI, providing rich individual or household level labor market output data (***Fredriksen and Stølen, 2007***; ***Klevmarken and Lindgren, 2008***; ***Flood et al., 2012***).

2.3. Trajectory model selection

Choosing the number of mixture components *K* is an important stage of finite mixture modeling. The trajectory method requires the researcher to specify the assumed number of sub-groups in the data. The optimal number of sub-groups *K* can be assessed using information criteria, which are widely used in this context.[[1]] One commonly used criterion to assess the model fit is the Bayesian information criterion (BIC). The model with largest BIC is preferred (SAS implementation). One may also calculate the group-specific average posterior probability over individuals to measure the fit. If this values exceeds the minimum threshold (at least 0.7) the fit is considered satisfactory. ***Nagin (2005***, p. 63–77) provides an overview of model selection in trajectory analysis context. Ultimately, meaningful real-world interpretations of sub-groups requires not just information criteria, but also theoretical evaluation and good judgment.

1. SAS package PROC TRAJ counts BIC and AIC routinely.There are other information criteria, like cross-validation available, if the number of sub-groups is critical for the microsimulation practitioner (see ***Nielsen et al., 2014***).

Table 1. BIC scores and number of sub-groups k

Outcome	3	4	5	6
Earnings (N=3,325)	−50288.7	−48178.8	−45974.7	−46375.9
Education (N=13,173)	−549399.6	−544342.9	−537412.2	−535026.8
Pension (N=10,761)	−834881.4	−811092.8	−786210.9	786234.1

Notes: BIC = log(L) −.5 * log(n) k, where L = log likelihood, n = sample size and k = number of parameters.

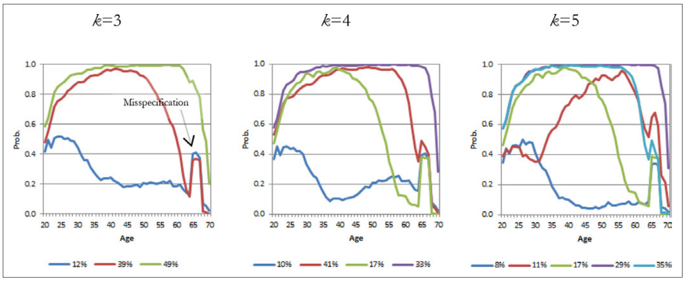


Figure 2. Earnings trajectories

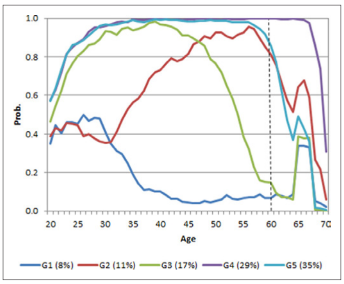


Figure 3. Earnings trajectories, *k*=5

In trajectory analysis missing data is considered missing completely at random according to the taxonomy layed by ***Little and Rubin (1987)***. Software packages handle missing data in such a way that all available data is used in the estimation. In other words, all individuals or objects with some missing longitudinal data values are included in the analysis.

From ***Table 1*** we can see the BIC values of our example analyses on three outcome-variables (earnings, education and pension). The BIC score indicates that, a five-group solution fits the wage earnings and pension variables, while for education BIC increases with more groups. However, increasing the number of education groups further would yield some infinitesimal groups. Therefore our choice is the five-group solution for the wage earnings and pension outcomes, and the six-group solution for the education outcome variable.

In this study the computations were carried out using SAS software package with accompanying PROC TRAJ application, an easy to use PC SAS procedure for analyzing Nagin's model (e.g. ***Jones et al., 2001***; ***Jones and Nagin, 2007***; ***Andruff et al., 2009***).[2] Appendix A.1 gives an example of the programming code for wage earnings trajectories (Example 1, *k*=5).

The trajectory plots (***Figures 2–4***) present conditional means of time points calculated over the simulation period. Relative sizes $\hat{\pi}_k$ of the sub-groups are also presented in the figures. These plots are the main tool for interpreting the results obtained. The estimated model is summarized in Appendix (A.2), which includes group-specific parameter estimates. The SAS procedure also produces a confidence interval (95%) and model-predicted values by sub-group if necessary.

2. SAS and STATA (see ***Jones and Nagin, 2013***) trajectory analysis package can be downloaded from the website: https://www.andrew.cmu.edu/user/bjones/. R package for trajectory analysis is Flexmix (see ***Leisch, 2004***).

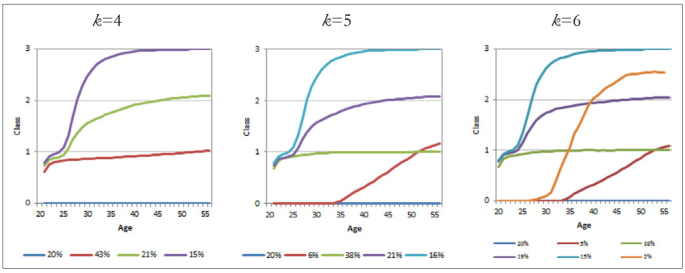


Figure 4. Education trajectories

It must be emphasized that the trajectories in the following examples are group means, which indicates that there is a range of individual tracks around the trajectory curves. Software packages routinely output the group estimates and confidence intervals to visualize within-group population heterogeneity. For the sake of readability, the model fit curves and confidence intervals are not presented in this paper.

In the examples we use cubic age model (see Equation 5). The choice depends on the assumed development of the data over time. One advantage in PROC TRAJ is that the age or time model can be group-specific, which means that for example one group can be fitted with a second order polynomial model, whereas another group can be fitted with a cubic model etc.

3. Examples

3.1. Earnings trajectories: binary outcome

For this example, we have chosen a young cohort to illustrate a simulated life-course. The cohort, born in 1995, enters labor markets at age 17 (year 2012), and subsequent life-course including individual wage earnings is simulated in the ELSI model's population and earnings modules.

The example shows a trajectory analysis with binary outcome data. Individual yearly wage is dichotomized (yes = 1/no = 0), so wage earnings could also be interpreted as employment. In terms of content, wage earnings could be analyzed without transformations using continuous data (euros).

The BIC score (***Table 1***) indicates that a five-group solution yields the best fit for the earnings outcome, but we also present the solutions for three and four-groups solutions. ***Figure 2*** shows that the relative sizes $\hat{\pi}_k$ vary somewhat with an increasing number of groups. Nevertheless all solutions have essentially the same major sub-groups.

In terms of content, we can see that there is a large group (k = 3: 49%, k = 4: 33% and k = 5: 29%) with strong labor market integration, or a strong earnings profile, until retirement age. Other groups show an early declining trend in employment. Labor market integration is weak in one group (k = 3: 12%, k = 4: 10% and k = 5: 8%), especially at older age. Unemployment and permanent disability draw individuals out of employment prior to old-age retirement. Retirement age for old-age pension in this cohort is 67 years and 9 months. Partial old-age pension can be drawn three years before retirement age.

3.1.1. ELSI model recalibration

This trajectory analysis experiment revealed a slight misspecification in the ELSI model's earnings module (see spikes at age 65 in all trajectory solutions). The spikes are visible in mean-based trajectories, but they would not be seen in fitted means models. This goes to show how useful group-based mean curves can be in revealing model misspecification.

In the earnings module the mechanism for working after retirement did not work as intended. During the course of simulation, many of those who had stopped working earlier suddenly decided

Table 2. Composition of earnings trajectories at age 60, per cent

	G1 (8%)	G2 (11%)	G3 (17%)	G4 (29%)	G5 (35%)	Total
Active first year	0.8	2.3	2.3	0.5	3.4	2.1
Active second year	0.4	0.6	0.7	0.3	0.9	0.6
Active (working)	2.3	51.1	1.2	89.9	54.2	50.6
Unemployed or in education	1.5	9.7	6.2	1.6	10.4	6.5
Sickness benefits	.	1.1	0.4	.	2.5	1.1
Partial disability pension and active	.	1.4	0.2	2.3	0.4	1
Partial disability pension and non-active	0.8	1.1	0.9	.	0.7	0.6
Full disability pension	23.4	2.8	26.9	0.5	2.9	7.9
Full disability pension second year	1.1	1.4	3.4	.	2	1.6
Old-age pension	.	.	0.5	0.1	0.7	0.4
Only national disability pension	15.5	.	.	.	.	1.2
Out of labor markets I	18.9	1.4	17.2	0.3	1.8	5.3
Out of labor markets II	14.3	.	2.5	.	0.8	1.9
Inactive	5.7	18.8	30.8	1.5	15.5	13.8
Deceased	15.5	8.2	6.9	2.8	3.8	5.5
Total	100	100	100	100	100	100

to return to work after drawing their pension (***Figure 2***). Of course, such behavior is not plausible on a larger scale.

The Finnish pension system allows full-time employment even after retirement on an old-age pension. However, wage earnings for old-age pensioners are typically quite low. The misspecification therefore did not have a major impact on the main results and was not observed in comparison with the LTP macro model.

In the ELSI model earnings are calibrated with the LTP model. In practice, this is done by comparing projected average earnings each year by gender, population state and age. Any deviations imply changes in the ELSI model's parameters.

In the LTP model, working after retirement is not modeled in detail, and the relevant point of reference is provided by the corresponding statistical figures. In the ELSI model, the share of new retirees working after retirement is the same as in the observed statistics. Trajectory analysis showed that the pool of retirees working after retirement also included individuals who had not worked in the year preceding retirement, which was not intended.

Following the trajectory analysis of simulated earnings, the ELSI model was recalibrated appropriately. There was no change in the aggregate figures, namely wage sum and number of people working after retirement.

Trajectory analysis is not yet routinely part of ELSI model testing, but it could be incorporated as part of its validation procedures.

3.1.1.1. Composition of trajectory groups

Checking the composition of the groups is a good way to validate also the trajectory analysis results. It is good practice to cross-tabulate the trajectory groups by the available background factors of the microsimulation. We have added an example which shows the labor market state or population state of earnings groups ($k = 5$) at age 60 in the simulated life-course (***Figure 3***). Depending on the available information of the microsimulation model, also other background factor could be cross-tabulated.

The labor market states (***Table 2***) confirm the earnings trajectory results. The high and stable wage earnings (employment) group G4 (29%) consists of people who are mainly (89.9%) in active labor market states. The weak attachment group G1 (8%) consists of individuals who had a disability or inactive status at the time. Also the share of deceased (15.5%) is higher in this group than in the other groups, which is quite natural. Another weak attachment group is G3 (17%), where the share of disabled and inactive individuals is relatively high compared to the other groups.

3.2. Education trajectories: class outcome

In this second example we continue to study the young cohort born in 1995 and illustrate a simulated life-course in terms of education level. This cohort finishes compulsory education at age 16 (in 2011). The subsequent education dynamics is simulated in the education module of the ELSI model.

The example shows a trajectory analysis with class-level outcome data. The distribution of the levels of education skewed to left (low education), therefore the outcome is modeled using the zero-inflated Poisson model. Yearly individual education information records the highest level of education. In practice, each individual level of education either remains unchanged or rises during the course of the simulation. Individuals enter the model at age 18 when 99% are at the basic education level. The ELSI model includes four levels of education: 0 = basic education (ISCED 0–3), 1 = secondary education (ISCED 4), 2 = lower academic degree (ISCED 5, 5B) and 3 = higher academic degree (ISCED 5A, 6).

The BIC score (***Table 1***) indicates that a six-group solution yields the best fit for the education outcome, but we also present the four-group and five-group solutions. ***Figure 4*** shows that with the six-group solution, the smallest group comprises no more than 2% of the population.

The trajectory means clearly show how education dynamics play out in practice. For example, education increases sharply until around age 30, and then is virtually stagnant at age 40. In the four-group solution, the results are dominated by the final education level. The four-group and five-group solutions show adult and further education in action, with 6–7% of the cohort still on an upward education trajectory at later ages (k = 5: 6%, k = 6: 7%).

Overall, we have observed that 20% have a low education, 40% complete secondary education, 21% complete a lower university degree and 15% a higher university degree.

3.3. Pension trajectories: continuous outcome

The cohort in this third example was born in 1960. For this cohort, earnings, labor market state, accrued pensions and disability pension data are covered up to age 52 (in 2012). Thereafter, pension stock and new pensions are simulated in the ELSI model's earnings-related pension module.

The example shows a trajectory analysis with continuous normal data. The outcome variable is earnings-related pension (log euros/month), excluding survivor's pensions, expressed in nominal value. If an individual is not retired (not yet retired or deceased) the outcome value is zero, otherwise the individual is retired and receives a pension (***Figure 5***). The level of pension is determined by current pension rules for different types of pension benefit, such as disability pensions, partial old-age pensions and old-age pensions.

Pension levels are affected by two factors. First, the most important factor that affects all retirees is indexation: earnings-related pensions are index-linked to prices (80%) and wages (20%). Second,

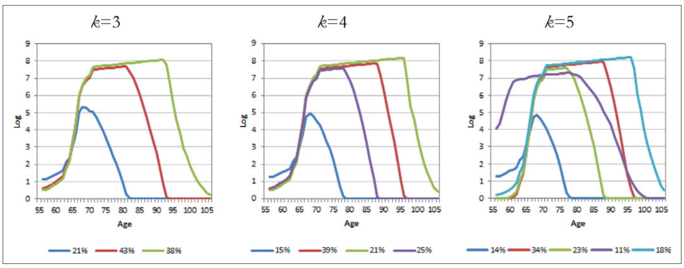


Figure 5. Pension trajectories

some individuals continue to work while drawing their pension, which accrues additional pension benefits.

The BIC score (***Table 1***) indicates that the five-group solution yields the best fit for the pension outcome, but we also present the solutions for three-group and four-group solutions.

Trajectory analysis reveals several substantive issues. First, pensions increase as the cohort ages and approaches minimum retirement age at 64 years and 5 months (persons born in 1960 can retire flexibly between ages 64 years 5 months and 69 years). Pensions taken earlier are for the most part either disability pensions or partial old-age pensions. Second, the trajectories are affected by mortality as all trajectory means go to zero with the death of individuals in the group. For instance, one group (k = 5: 14%) dies early and its pension level remains low. This group, along with (k = 5: 11%), enters disability pensions at a higher rate than other groups. In addition, some groups have a higher life expectancy. The group with the highest life expectancy (k = 5: 18%) also has the highest pensions. Third, there is some indication of segregation among pensioners here.

It is also worth noting that pension indexation rules increase pensions in those sub-groups where life expectancy is high. Finally, an increase in the number of groups leads to the emergence of some interesting sub-groups. The early transition to disability pension is evident in the five-group solution (k = 5: 11%).

4. Conclusion

We have applied finite mixture modeling to longitudinal data in the context of dynamic microsimulation. It is useful in visualizing dynamic microsimulation results because it demonstrates, within a formal statistical structure, population heterogeneity in simulated datasets. Furthermore, the sub-groups identified in the analysis may reveal model misspecifications if such issues are present.

Our first goal was to illustrate the results of dynamic microsimulation, which are often presented as statistical moments or other distribution measures for a given population. Statistical moments presented on the basis of ex ante classification rules (e.g. gender, level of education, occupation or region) are a common way of reporting results for specific population sub-groups. We have shown how trajectory analysis can be applied as a data-driven method for illustrating longitudinal results in a novel way. These two ways of presenting results are mutually complementary.

Our second goal was to test the individual-level or object-level outcomes of dynamic microsimulation. Testing should be understood here by locating groups with unwanted or otherwise peculiar outcomes. Microsimulation easily yields biases and model misspecification. Microsimulation practitioners apply various techniques to calibrate models and to locate misspecification. Trajectory analysis could also be a useful tool in testing simulation model assumptions, as model misspecification may be revealed in developmental trajectories. Trajectory analysis together with other testing methods can improve microsimulation model reliability, especially in connection with large-scale revisions of model parameters.

There are some limitations with the proposed technique. First, it takes some effort to analyze all the cohorts included in the microsimulation model. It might be advisable to focus on some key cohorts if the simulation model assumes similar parameters across cohorts. Second, given the computational requirements of the EM algorithm, it is apparent that extremely large data sets (hundreds of thousands of individuals) should be avoided. A fraction or sample of the simulated data set usually reveals the underlying sub-populations with sufficient reliability. Third, it is necessary to mention conditional independence, even though this is not a serious drawback in a microsimulation context. Finally, basically trajectory analysis cannot handle nonlinear models. The group-specific regression model needs to be linear in parameters. However, reasonable approximations can often be provided by low-degree polynomial models.

Trajectory analysis is also a highly flexible method. In our case we performed the analysis for one outcome at the time, but the technique and software is in place to model several outcomes at the same time. This is called multi-trajectory modeling, which identifies groups of individuals or objects that follow similar trajectories across multiple outcomes. For example, it would be possible to analyze income, education and labor market status simultaneously. This would yield a slightly different probability for belonging to a specific group. Another possibility would be to add time-dependent (e.g. earnings) or time-stable (e.g. gender, year of birth or nationality) covariates alongside age or time polynomial. This would be a case of multinomial logistic regression within a population sub-group, and the result would consist of risk factors influencing group membership.

ORCID iD
Janne Salonen http://orcid.org/0000-0002-0595-6226

Acknowledgments

The authors wish to thank Statistics Finland for providing material for this study and two anonymous referees as well as chief editor Matteo Richiardi for their many thoughtful comments that greatly improved this paper.

Data availability

The data used in the study is from ELSI-microsimulation model. The policy regarding ELSI is; Proprietary, with executable not available.

Code availability

The statistical model used in the article (in SAS-environment) is given in the article Appendix.

References

Andruff H, Carraro N, Thompson A, Gaudreau P, Louvet B. 2009. Latent class growth modelling: a tutorial. *Tutorials in Quantitative Methods for Psychology* **5**:11–24. DOI: https://doi.org/10.20982/tqmp.05.1.p011
Böhning D, Seidel W, Alfó M, Garel B, Patilea V, Walther G. 2007. Advances in mixture models. *Computational Statistics & Data Analysis* **51**:5205–5210. DOI: https://doi.org/10.1016/j.csda.2006.10.025
Dekkers G, Van den Bosch K. 2016. Prospective microsimulation of pensions in European Member States. In: Dekkers G, Mészáros J (editors). *Applications of microsimulation modelling*.
Dempster A, Laird N, Rubin D. 1977. Maximum likelihood estimation for incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* **B 39**:1–38.
Don BP, Mickelson KD. 2014. Relationship satisfaction trajectories across the transition to parenthood among low-risk parents. *Journal of Marriage and Family* **76**:677–692. DOI: https://doi.org/10.1111/jomf.12111
Flood L, Jansson F, Pettersson T, Pettersson T, Sundberg O, Westerberg A. 2012. *SESIM III - A Swedish dynamic micro simulation model*.
Fredriksen D, Stølen N. 2007. Model 1: MOSART (Dynamic Cross-Sectional Microsimulation Model). In: Gupta Anil, Harding Ann (editors). *Modelling Our Future: Population Ageing, Health and Aged Care (International Symposia in Economic Theory and Econometrics, Volume 16)*. Emerald Group Publishing Limited. p. 433–437
Grun B, Leisch F. 2007. FlexMix: an R package for finite mixture modelling. Faculty of Commerce – Papers, University of Wollongong, Australia.
Haughton D, Legrand P, Woolford S. 2009. Review of three latent class cluster analysis packages: latent GOLD, poLCA, and MCLUST. *The American Statistician* **63**:81–91. DOI: https://doi.org/10.1198/tast.2009.0016
Hsu H-C. 2015. Trajectories of multimorbidity and impacts on successful aging. *Experimental Gerontology* **66**:32–38. DOI: https://doi.org/10.1016/j.exger.2015.04.005
Hynes K, Clarkberg M. 2005. Women's employment patterns during early parenthood: a group-based trajectory analysis. *Journal of Marriage and Family* **67**:222–239. DOI: https://doi.org/10.1111/j.0022-2445.2005.00017.x
Jones BL, Nagin DS. 2007. Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociological Methods & Research* **35**:542–571. DOI: https://doi.org/10.1177/0049124106292364
Jones BL, Nagin DS. 2013. A note on a STATA plugin for estimating group-based trajectory models. *Sociological Methods & Research* **42**:608–613. DOI: https://doi.org/10.1177/0049124113503141
Jones BL, Nagin DS, Roeder K. 2001. A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research* **29**:374–393. DOI: https://doi.org/10.1177/0049124101029003005
Klevmarken A, Lindgren B. 2008. eds. *Simulating an Ageing Population: A microsimulation approach applied to Sweden*. Bingley: Emerald Group Publishing Ltd.
Kokko K, Pulkkinen L, Mesiäinen P, Lyyra A-L. 2008. Trajectories based on Postcomprehensive and higher education: their correlates and antecedents. *Journal of Social Issues* **64**:59–76. DOI: https://doi.org/10.1111/j.1540-4560.2008.00548.x
Leisch F. 2004. FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R. *Journal of Statistical Software* **11**:1–18. DOI: https://doi.org/10.18637/jss.v011.i08
Li J, O'Donoghue C. 2013. A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of Microsimulation* **6**:3–55.
Li J, O'Donoghue C, Dekkers G. 2014. Dynamic models. In:O'Donoghue C (editor). *Handbook of Microsimulation Modelling. Contributions to economic analysis*. Emerald.
Little R, Rubin D. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.
Mani D, Nandkumar A. 2016. The differential impacts of markets for technology on the value of technological resources: an application of group-based trajectory models. *Strategic Management Journal* **37**:192–205. DOI: https://doi.org/10.1002/smj.2457
McLachlan G, Peel D. 2000. *Finite Mixture Models*. New York: John Wiley and Sons Inc.
Muthén L, Muthén B. 2010. *MPlus Statistical Analysis With Latent Variables, User's Guide*. Los Angeles, CA.

Nagin DS. 1999. Analyzing developmental trajectories: a semiparametric, group-based approach. *Psychological Methods* **4**:139–157. DOI: https://doi.org/10.1037/1082-989X.4.2.139
Nagin D. 2005. *Group-Based Modeling of Development*. Cambridge, MA: Harvard University Press.
Nagin DS. 2016. Group-Based trajectory modeling and criminal career research. *Journal of Research in Crime and Delinquency* **53**:356–371. DOI: https://doi.org/10.1177/0022427815611710
Nathalie M, Fontaine G, Petersen I. 2017. Developmental Trajectories of Psychopathology. An Overview of Approaches and Applications. In: Centifanti L, Williams D (editors). *The Wiley Handbook of Developmental Psychopathology*. John Wiley & Sons Ltd.
Nielsen JD, Rosenthal JS, Sun Y, Day DM, Bevc I, Duchesne T. 2014. Group-Based criminal trajectory analysis using cross-validation criteria. *Communications in Statistics - Theory and Methods* **43**:4337–4356. DOI: https://doi.org/10.1080/03610926.2012.719986
Nagin D, Jones B, Passos L, V, Tremblay R. 2016. Group-Based multi-trajectory modelling. *Statistical Methods in Medical Research* **0**:1–9.
Nagin DS, Odgers CL. 2010. Group-Based trajectory modeling (nearly) two decades later. *Journal of Quantitative Criminology* **26**:445–453. DOI: https://doi.org/10.1007/s10940-010-9113-7
Nummi T, Hakanen T, Lipiäinen L, Harjunmaa U, Salo MK, Saha M-T, Vuorela N. 2014. A trajectory analysis of body mass index for Finnish children. *Journal of Applied Statistics* **41**:1422–1435. DOI: https://doi.org/10.1080/02664763.2013.872232
Nummi T, Virtanen P, Leino-Arjas P, Hammarström A. 2017a. Trajectories of a set of ten functional somatic symptoms from adolescence to middle age. *Archives of Public Health* **75**:11. DOI: https://doi.org/10.1186/s13690-017-0178-8
Nummi T, Salonen J, O'Brien T. 2017b. Statistical Analysis of Labor Market Integration: A Mixture Regression Approach. In: Chen D-G, Jin Z, Li G, Li Y, Liu A, Zhao Y (editors). *New Advances in Statistics and Data Science*. ICSA Book Series in Statistics, Springer.
Peutere L, Vahtera J, Kivimäki M, Pentti J, Virtanen P. 2015. Job Contract at Birth of the First Child as a Predictor of Women's Labor Market Attachment: Trajectory Analyses over 11 Years. *Nordic Journal of Working Life Studies* **5**:9–30. DOI: https://doi.org/10.19154/njwls.v5i1.4763
Sihvonen H. 2015. Eläketurvakeskuksen ELSI-mikrosimulointimallin laajennus Kelan eläkkeisiin ja verotukseen. Finnish Centre for Pensions Working papers 03/2015.
Tikanmäki H, Sihvonen H, Salonen J. 2014. Microsimulating Finnish earnings-related pensions. *Finnish Centre for Pensions Working Papers 02/2014*.
Tikanmäki H, Sihvonen H, Salonen J. 2015. Distributional effects of forthcoming Finnish pension reform – a dynamic Microsimulation approach. *International Journal of Microsimulation* **8**:75–98.
Tikanmäki H, Appelqvist J, Reipas K, Sankala M, Sihvonen H. 2017. Statutory pensions in Finland – long-term projections 2016. Finnish Centre for Pensions Reports 02/2017.
Titterington D, Smith A, Makov U. 1985. *Statistical analysis of finite mixture distributions*. Great Britain: John Wiley & Sons.
van der Geest V, Bijleveld, C, Blokland A, Nagin D. 2016. The effects of incarceration on longitudinal trajectories of employment: a follow-up in high-risk youth from ages 23 to 32. *Crime & Delinquency* **62**:107–140.
Zaidi A, Rake K. 2001. Dynamic Microsimulation Models: A Review and Some Lessons for SAGE. SAGE Discussion Paper no. 2. The London School of Economics.

International Microsimulation Association Research article

Methodology, Dynamic microsimulation

Appendix

A.1 An example SAS PROC TRAJ-code.

Earnings trajectories: Five-group solution

- Males born 1995
- Simulated yearly wage earnings 2015–2080
- Binary outcome 1 = some wage, 0 = no wage
- WMALE_1995 is the individual level data set with yearly observations on binary wage (bwage) and age (age).

The PROC TRAJ produces four output datasets. OP includes the group mean estimates and confidence intervals (95%). OS includes the mixture components (sizes of the sub-groups). OF includes individual-level information on group posterior probabilities and group assignment. This is an important data set where we can merge any other relevant background information. Finally, OE includes the regression model coefficients and information criteria (AIC and BIC). ITDETAIL statement shows the progression of maximum likelihood estimation in the SAS log window.

Algorithm 1: PROC TRAJ.

```
PROC TRAJ DATA=WMALE_1995 OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
ID idno;
VAR bwage2015-bwage2080;
INDEP age2015-age2080;
MODEL logit;
NGROUPS 5;
ORDER 3 3 3 3 3;
RUN;
```

The actual model is specified with VAR, INDEP, MODEL and ORDER statements. The MODEL logit states that we analyze binary outcomes. Other possibilities are censored normal outcome (cnorm) and zero-inflated Poisson outcome (zip). NGROUP states the number of sub-groups. ID states the number of individuals. ODRER defines the number of mixture components (=NGROUPS) and respective degree of age-polynomial. Each sub-group can have different degrees of age-polynomial.

Algorithm 2: TRAJPLOT-macro.

```
%TRAJPLOT(OP,OS,,);
RUN;
DATA WMALE_1995; MERGE WMALE_1995 OF (KEEP=idno group); BY idno;
RUN;
```

The TRAJPLOT macro statement invokes the result plot, which by default plots the trajectory means and their estimates. Finally, individual level results are collected and merged with microsimulation output dataset.

A.2 Summary of Estimated Model: Outcomes, Groups, Parameter Estimates and their Standard Errors.

Outcome	Groups	$\hat{\beta}_0$	$SE(\hat{\beta}_0)$	$\hat{\beta}_1$	$SE(\hat{\beta}_1)$	$\hat{\beta}_2$	$SE(\hat{\beta}_2)$	$\hat{\beta}_3$	$SE(\hat{\beta}_3)$
Earnings	1	–0.60816	0.93616	0.17156	0.07229	–0.00845	0.00173	0.00008	0.00001
Earnings	2	10.77241	0.7512	–1.05243	0.05715	0.03048	0.00135	–0.00026	0.00001
Earnings	4	–24.0041	0.69707	1.88334	0.05421	–0.04084	0.00128	0.00026	0.00001
Earnings	5	2.89272	0.93197	–0.55508	0.08212	0.02763	0.00225	–0.00029	0.00002
Education	1	150.7634	13.69866	–23.324	1.03504	0.60943	0.01352	–0.00442	0.00018
Education	2	–38.2479	1.84202	1.95442	0.10906	–0.03343	0.00214	0.00019	0.00001
Education	3	–1.40435	0.08235	0.09431	0.00638	–0.00206	0.00016	0.00001	0.00000
Education	4	–3.82746	0.09413	0.27969	0.00712	–0.00568	0.00017	0.00004	0.00000
Education	5	–6.14041	0.09384	0.46159	0.00701	–0.00958	0.00017	0.00006	0.00000
Education	6	–34.1529	1.02762	2.02103	0.06446	–0.03841	0.00133	0.00024	0.00001
Pension	1	1352.226	1.29912	–66.1983	0.03537	1.07053	0.0006	–0.00572	0.00001
Pension	2	–220.279	3.93005	5.68868	0.15254	–0.03308	0.00196	–0.00003	0.00001
Pension	3	–464.879	3.77328	14.13735	0.14801	–0.12224	0.00197	0.00024	0.00001
Pension	4	27.60321	4.88433	–1.86771	0.19721	0.03987	0.00262	–0.00025	0.00001
Pension	5	–127.459	2.91581	3.18956	0.10992	–0.01691	0.00136	–0.00002	0.00001

02
2020
FINNISH CENTRE FOR PENSIONS, STUDIES

New methods in pension evaluation

Applications of trajectory analysis and dynamic microsimulation

This study gives an overview of two new methods in pension evaluation finite mixture modeling technique called trajectory analysis and dynamic microsimulation. Trajectory analysis and microsimulation can be used to study longitudinal datasets and various topics related to life courses. In this study, the applications of these methods focus on early career labor market attachment, insurance behavior of the self-employed and evaluation of the 2017 pension reform. Furthermore, this study discusses measurement and classification questions in pension research. The analyzes are based on the register data of the Finnish Centre for Pensions.

FINNISH CENTRE FOR PENSIONS, STUDIES

The Finnish Centre for Pensions, an expert on earnings-related pensions, is a statutory body that develops pension provision and produces joint services for all parties to the scheme. Our Studies series includes research reports that meet high-quality scientific criteria. The subjects of our studies relate to pension provision and pension systems.

Finnish Centre for Pensions
ELÄKETURVAKESKUS