05/2012
FINNISH CENTRE FOR PENSIONS, WORKING PAPERS

Earnings profiles of Finnish wage earners in 2000–2010

Tapio Nummi, Janne Salonen and Lasse Koskinen

Finnish Centre for Pensions
ELÄKETURVAKESKUS

05/2012

FINNISH CENTRE FOR PENSIONS, WORKING PAPERS

Earnings profiles of Finnish wage earners in 2000–2010

Tapio Nummi, Janne Salonen and Lasse Koskinen

Finnish Centre for Pensions
ELÄKETURVAKESKUS

Finnish Centre for Pensions
FI-00065 ELÄKETURVAKESKUS, FINLAND
Telephone +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

Helsinki 2012

ISSN-L 1795-3103
ISSN 1797-3635 (online)

FOREWORD

The outcomes of an earnings-related pension scheme depend on earnings to a great extent. Earlier, only the last years before retirement were important for determining the pensionable wage. Since 2005, all earnings from 18 years onwards matter.

How will future pensions develop? How does the change in calculating the pensionable wage affect the distribution of pensions? For addressing these kinds of questions, the standard solution is to calculate average earnings and average pensions, or to make model-family calculations and simulations. However, for more empirically grounded answers, in-depth knowledge of the distribution and development of earnings is needed.

This working paper presents the first results of research co-operation, where the main aim was to address variation in earnings by identifying differing earnings profiles with the help of trajectory analysis and longitudinal data.

Mikko Kautto
Head of Research Department
Finnish Centre for Pensions

ABSTRACT

One primary aim of this report is to identify and describe typical earnings profiles in Finland. The statistical method applied is that of modern trajectory analysis, which is commonly used for the clustering of longitudinal data. The range of earnings profiles identified is quite heterogeneous. The trajectory analysis distinguishes six groups with different earnings profiles that can be separately modelled. Rather than pre-established groups, the clustering was based on the categories suggested by the statistical method. The models show a good statistical fit to the data. According to our results group modelling is a more effective technique than using a single model applied to the whole sample using the same explanatory factors. In the case of most wage earners in Finland, earnings profiles are in line with the moderate pay rises agreed through collective bargaining. However the analysis also identified a group with a high profile career and on the other hand groups showing weaker labour force attachment or earnings profiles.

Age-earnings profiles are described in this report not only by means of clusters, but also by using statistical mixed models. A mixed model developed to explain the earnings profile of an individual trajectory group includes explanatory factors related to the individual's age, employment career and the general economic environment. Again, group modelling provides a more effective tool than a single model applied to the whole sample using the same explanatory factors. According to the estimation results real wages begin to fall in all groups after age 50. Length of career or work experience has a positive effect on earnings. A favourable general economic climate is reflected not only in employment rates, but also in the development of wages. With the exception of one group, GDP growth drives up earnings as well.

The results are consistent and have interesting implications for the pension system. It is apparent that age-earnings profiles cannot be exhaustively described by just one single model that is based on the total employed population. Statistical clustering of the data reveals many different kinds of profiles. Pension accrual is based on employment career and earnings. Looking at the age group 55–60 who are approaching retirement age, we find that their accrued earnings-based pension rights reflect their employment career to date. A strong earnings profile translates into a high expected pension. The group with a strong earnings can look forward to an earnings-related pension that ranks among the top decile of pension recipients. In low earnings groups the earnings-related pension will probably remain so low that it will need to be topped up by the national pension.

Keywords: Mixed model, Curve clustering, Trajectory analysis, Pension, Earnings

ABSTRAKTI

Tämän raportin keskeinen tavoite on tutkia millaiset ansioprofiilit ovat tyypillisiä. Tilastollisena menetelmänä sovelletaan modernia trajektorianalyysiä, joka soveltuu pitkittäisaineiston luokitteluun. Ansioprofiilien joukko on varsin heterogeeninen. Trajektorianalyysin perusteella otoksesta tunnistetaan kuusi ansioprofiileiltaan erilaista ryhmää, joiden palkkakehitystä voidaan erikseen mallintaa. Ryhmittely perustuu ennalta sovittujen kriteerien sijaan tilastolliseen menetelmään. Tilastollisessa mielessä mallit sopivat hyvin aineistoon. Voidaan osoittaa, että ryhmäkohtainen mallintaminen toimii paremmin kuin koko otokselle sovellettava malli samoilla selittäjillä. Valtaosalla ansiokehitys noudattelee maltillista linjaa, joka noudattaa palkkaneuvotteluissa sovittuja korotuksia. Joukosta erottuu korkean profiilin työuraa tekevien ryhmä sekä toisaalta ryhmiä, joiden työurakiinnittyminen tai ansiokehitys on heikko.

Ryhmittelyn lisäksi iän mukaista ansiokehitystä kuvataan tilastollisilla sekamalleilla. Yksittäiselle trajektori-ryhmälle laaditussa sekamallissa on mukana ansiokehitystä selittämässä henkilön ikään, työuraan ja yleiseen talouskehitykseen liittyviä selittäjiä. Ryhmäkohtainen mallintaminen toimii paremmin kuin koko otokselle sovellettava malli amoilla selittäjillä. Estimointitulosten mukaan reaali-palkka kääntyy laskusuuntaan kaikissa ryhmissä 50-ikävuoden jälkeen. Työuran kestolla/työkokemuksella on positiivinen palkkavaikutus. Yleinen positiivinen talouskehitys heijastuu paitsi työllisyyteen, myös palkkakehitykseen. Yhtä ryhmää lukuun ottamatta bruttokansantuotteen kasvu nostaa myös ansioita.

Eläkejärjestelmän kannalta tulokset ovat mielenkiintoisia ja johdonmukaisia. Iän mukaista ansiokehitystä ei ilmeisesti voida täydellisesti kuvata vain yhdellä profiililla tai koko työlliseen väestöön perustuvalla mallilla. Aineistoon perustuva luokittelu tuo esiin monenlaisia profiileja. Henkilön eläke perustuu työhön ja ansioihin. Tarkasteltaessa ryhmittäin vanhuuseläkeikää lähestyviä, 55–60-vuotiaita, karttunut työeläke heijastelee toteutunutta työuraa. Hyvä ansiokehitys heijastuu myös hyvänä odotettavissa olevana eläkkeenä myös euromäärin mitattuna. Korkean profiilin ryhmä voi odottaa aikanaan saavansa työeläkkeen, joka on eläkkeensaajien tulonjakauman ylimmässä kymmenyksessä. Matalan ansiokehityksen ryhmissä työeläke jää niin pieneksi, että työeläkkeen rinnalle todennäköisesti lisätään kansaneläke.

Avainsanat: Sekamalli, Kehityspolkuanalyysi, Trajektorianalyysi, Eläke, Ansiot

ACKNOWLEDGMENTS

Tapio Nummi, Professor of Biometry at University of Tampere, wishes to thank The School of Health Sciences for the time devoted for this study.

Lasse Koskinen, Senior Risk Expert, wishes to thank Financial Supervisory Authority, for the time devoted for this study. The views expressed in this study are those of Lasse Koskinen and do not necessarily reflect the views of the Financial Supervisory Authority.

Mikko Kautto, Head of the Research Department at the Finnish Centre for Pensions, contributed useful practical points on how to interpret the results.

Pia Hansson, Head of translation at the Finnish Centre for Pensions and especially David Kivinen at Translation agency Transmasters, translated the manuscript into English.

Anja Kallio, Department Secretary at the Finnish Centre for Pensions, has taken care of transforming the manuscript into a publication.

CONTENTS

1 **Introduction** 9
2 **Research data** 12
3 **Wage trajectory model** 15
Wage trajectory clusters 16
Explanatory factors 19
4 **Estimation results** 21
Age-wage profiles in practice 22
5 **Discussion** 25
Annexes
References

1 Introduction

Earnings-related pension is accrued on the basis of future wages. Calculations to determine the future sustainability of the pension system and pension levels must therefore give close consideration to the wage or earnings profile they use (for an earlier discussion of this theme, see e.g. Elo and Salonen 2004). It has been shown earlier that the gender pay gap and the shape of the wage profile are key factors in predicting future wages and pensions.

Individual wage profiles vary widely and can hardly be adequately described using a single profile for men and for women; this is too crude an approach. Our aim in this research is to identify groups of wage earners who share similar earnings profiles by using existing register datasets. The statistical technique we use is called trajectory analysis, a method initially applied in behavioural sciences and marketing research, for instance. These groups are then modelled in closer detail using a linear mixed model that takes account of individual differences.

It has long been known that earnings profiles differ at different stages of the life cycle. Factors impacting the age-wage profile have been an area of long-standing interest in microeconomics research. Wages tend to rise quite sharply in the early stages of the individual's employment career, but then begin to plateau in middle age, around age 50. When the individual begins to approach retirement age, earnings often begin to trend down (Fig. 1). This is most apparent in a cross-sectional analysis, but even cohort level analyses often reveal a concave wage profile (Annex Fig. 1).

In microeconomics, age-earnings profiles are explained by reference to the theory of human capital: observations based on that theory show that these profiles tend to have a concave shape with age (Mincer 1974).

It is also well-known that wage profiles depart from the concave shape most particularly in the public sector, where earnings often remain on an upward trajectory through to retirement age. Furthermore, it is known that men and women have distinct wage profiles: the age-wage profile for men follows the traditional concave shape, whereas women's profile tends to continue to rise through to the end of their career.

The Finnish labour market is highly gender-segregated (Korkeamäki and Kyyrä 2006 and Kyyrä 2007). In principle men and women receive more or less equal pay for equal work, but since they tend to work in very different occupations, their earnings do in fact differ quite substantially. Table 1 illustrates the gender pay gap by age group based on Statistics Finland sources. In the age group 35–54 when employment rates are highest, women's earnings are about 70–80% of men's earnings. The wage gap is a persistent phenomenon, even though it has narrowed to some extent. In 2005 the gender pay gap ratio was 72%, in 2010 it had been reduced to 77%.

Figure 1

Age-wage profile for the sample, mean.

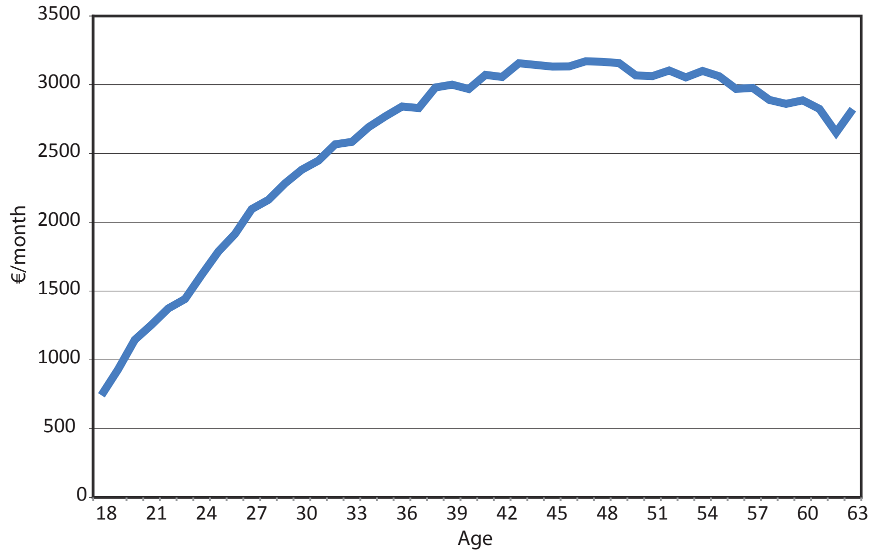


Table 1.

*Women's earnings as per cent of men's earnings in 2010**

Age	%
-16	85
17-24	89
25-34	67
35-44	70
45-54	83
55-64	88
65-	57
Total	77

*Source: Statistics Finland

Recently there has been a growing trend in the Finnish labour market from full-time to part-time employment; the same is true in the other Nordic countries (see e.g. Kjeldstad and Nymoen 2012). Part-time employment is often taken up by women as they seek to reconcile family and work obligations. This is reflected in the earnings of younger women who are in the process of setting up a family, although it is not uncommon at later stages of the employment career either. One form of part-time employment is to retire on part-time pension, an option taken by both women and men. Eligibility for part-time pension in Finland currently starts at age 60.

The development of earnings over the life cycle is explained by age, work experience and education. The theory of human capital suggests that earnings peak at a certain stage in middle age, at a point when the individual is at his or her most productive. Most statistical models therefore include an age (or work experience) factor and the square of age. Agespecific productivity has recently been described in a UK dataset by Dickerson and McIntosh (2011). In Finland the phenomenon has been studied by Ilmakunnas and Maliranta (2005).

This report uses register data to identify typical age-earnings profiles in Finland and to try to establish how to explain those profiles. In contrast to many earlier studies our aim here is not to extract a single wage profile for all persons in the sample, but rather to undertake a data-driven search for typical profiles. The statistical technique used in this search is trajectory analysis, which is used to cluster typical earnings categories throughout the period under review. The six groups that are identified in this search are described in closer detail using a linear mixed model.

In Finland trajectory analysis has recently been applied in the field of career research (Virtanen et al. 2011 and Liukkonen 2012). However to the best of our knowledge this is the first time that clustering trajectory analysis is used in earnings research. The most important new feature of the method is that it allows us to identify, even in large longitudinal datasets, trajectories that cause heterogeneity in the material and that could not be detected in a study of background factors alone.

Once the sample has been classified based on trajectory analysis, the next stage uses a mixed model to provide a more in-depth understanding of the factors that have a bearing on the development of wages in the different earnings profiles categories. Mixed models are particularly well suited for modelling at the individual level, and indeed they have been widely applied in the field of social insurance.

Data classified by trajectory analysis will cast new light on the prevalence of the traditional concave age-earnings profile as well as on the background factors of persons with a typical earnings profile.

2 Research data

The earnings-related pension system in Finland has mechanisms in place to collect information about work performed and wages paid. Pensions are based on register information about earnings. Maintained by the Finnish Centre for Pensions, the administrative register contains information about wages and earnings, job careers, employment contracts, employers, sectors of employment and pensions. Our research material covers all wages, salaries and compensation paid to the employee in both the private and public sector in 2000–2010, giving a total of 11 annual observations. There are a total of 58 age groups (persons born in 1935–1992). Self-employed persons are not included in the material.

Sample restrictions

Given the sheer size of the administrative register dataset, we have been forced for practical reasons to work with a sample. In the first stage we took a random sample of about 10% from the population aged 18–68. Representing the population of working age, this sample comprises 184,310 persons. This first-stage sample is used in the early part of the report to describe the research data and the phenomenon under investigation. For the purposes of the statistical analysis in this report, the dataset was further reduced to a random sample of 5,000 persons (Table 2). This second-stage sample is necessary because the statistical technique employed (trajectory analysis) is an iterative method that requires substantial computing capacity. However the sample is large enough so that it can be considered reasonably representative of the total wage-earning population. It is worth noting that for clarity of interpretation, those 439 persons were removed from the sample who were on part-time pension.

Table 2.
Variables describing the sample (n=4,561): summary statistics.

	Birthyear	Wage	Age	Workday	Employers	Career
Min	1935	1	14.0	1.0	1.0	0.0
1st Qu.	1955	1399	29.0	360.0	1.0	5.8
Median	1965	2184	40.0	360.0	1.0	15.8
Mean	1966	2361	39.5	323.9	1.7	16.8
3rd Qu.	1976	2953	50.0	360.0	2.0	26.5
Max	1992	124500	68.0	360.0	52.0	49.9

Wage

Pension rights accrue on the basis of wages paid. Wages (our response variable) are compensation paid to employees for work performed. They do not include any other income items, such as investment income or current transfers. Wages are indicated as monthly earnings (eur/month) and are calculated by dividing annual wages by annual working time. As is common practice, wages are expressed in the prices of the last year under review, i.e. in this case wages are expressed in 2010 prices.

Workday

Annual working time is measured in terms in workdays per year, or 360 days.

Career

The individual's employment career begins on the day they turn 18 and is calculated through to the end of the year under review. Employment career is expressed in years carried to one decimal place. It is assumed that this explanatory variable increases the amount of wages received with increasing work experience.

Employers

Number of employment contracts describes the number of employers for whom the person has worked during the course of the year. Working with the same employer adds to career stability, but on the other hand changing jobs brings the opportunity to move into better paid positions. Our assumption is that this explanatory variable correlates positively with wages, which would suggest the dominance of the latter option.

Part-time pension

One of the shortcomings of our dataset is that it provides no information about employment intensity, i.e. we do not know exactly when a person has been in full-time employment and when they have been working part-time. In part-time employment wages are obviously lower than in full-time jobs. Young adults and women in particular often work part-time for family reasons. This, unfortunately, remains hidden to us. Persons over 57, for their part, may be in part-time retirement. This information is recorded in pension registers, and therefore years spent in part-time retirement can be excluded from the analysis. A total of 439 persons were in part-time retirement; 4,561 persons were thus included in the analysis proper.

GDP growth

Wage research often uses some explanatory variable in its models to describe the development of the real economy. The purpose of this is to control the impact of cyclical movements and the general economic climate on wage profiles. For the present analysis we have included GDP volume measured in billion euros to one decimal place. The Finnish economy has been on a fairly steady path during the period under review, with very few major shocks impacting either employment or wages. It is assumed that this explanatory variable correlates positively with wages.

In 2000–2001, during the first years of the period under review, the IT stock market overheated and share prices in Finland tumbled. This had some disruptive effect on the real economy and was reflected in output. However the impacts on employment were quite limited. Following the IT crisis the Finnish economy recovered to a strong path for several years. The situation in the labour market was settled, and exports from our open economy showed healthy growth. In the aftermath of the IT crisis unemployment edged down steadily.

The financial crisis that started from the United States made landfall in Finland in autumn 2008. The effects escalated in the course of 2009, when the number of layoffs started to accelerate and unemployment to increase. At the time Finnish businesses were still in reasonably good financial health. Central government invested heavily in stimulus. Unemployment did not rise as sharply as had been feared. In 2009 GDP shrank by 8.4%, clearly demonstrating the sensitivity of a small open economy to international shocks.

Distribution of wages

In this analysis we used a version of trajectory analysis that assumes a normal distribution. Wages do not follow a normal distribution, but the distribution is highly skewed to the right. It follows that methods based on the assumption of normality are not directly applicable to this analysis. One option is to transform the distribution so as to bring it close to a normal distribution again. It turned out that the square root of wages yielded the best result.

Neither gender nor sector of employment are separately included in the model as explanatory factors. Our aim here is not to measure the size of the gender wage gap, but rather to see how trajectory analysis distinguishes men and women into different groups. Gender is reflected in the distribution of wages such that women are overrepresented in the below mean income groups.

3 Wage trajectory model

Trajectory analysis is an application of finite mixtures theory that is used for purposes of modelling heterogeneity appearing in longitudinal datasets (see e.g. Nagin 1999 and 2005, Jones et al. 2001). The aim is to identify groups with similar behaviour over time, who can then be subjected to various further analyses. If the probability distribution of measurements y is $f(y|X)$, then the aim is to estimate the mixed distribution

$$f(y|X) = p_1 f_1(y|X) + \cdots + p_k f_1(y|X)$$

so that the sum of parameters $p_1, \cdots, p_k (0 < p_i < 1, i = 1, \cdots, k)$ is 1, where matrix X includes the model's explanatory factors. In other words parameter pi indicates the probability of membership of group *i*. If distribution *f* belongs to the exponential family, we get a mixture of generalized linear models, including the special case of a mixture of ordinary regression models (normal distribution). Estimation of the model parameters is usually performed using the maximum likelihood method.

Basic trajectory analysis (Nagin, 1999) works with the assumption that observations are independent, which slightly simplifies the likelihood function to be maximized. An application of latent growth curve models (Muthén, 2001) takes the curve parameters as latent factors (i.e. random variables), implying that dependence within trajectories can in principle be modelled. Unfortunately this sharply increases the number of parameters to be estimated, which may easily cause problems (e.g. convergence) with the estimation of model parameters.

In practice, the parameter estimation is performed iteratively via the so-called EM (Expectation and Maximization) algorithm (Dempster et al., 1977), which is often used in difficult estimation settings, for instance in situations of missing information. The algorithm is based on two-step iteration. The E-step involves computing the expected value of a logarithm of likelihood function. In trajectory analysis the likelihood function is constructed such that the membership of an observation to a certain cluster is considered a random variable that follows multinomial distribution, for which an expected value is then calculated. The M-step then uses the expected values obtained at the E-step, which are then maximized with respect to unknown parameters. The procedure is repeated until the algorithm has converged (i.e. the likelihood function no longer increases). The basic theory of the algorithm guarantees that the probability function increases at each iteration. However convergence to local maximum may cause some problems. Indeed in order to corroborate the solution it is commonplace to repeat the iterations several times using different initial values. Examples of software suited for trajectory analysis include *SAS Proc Traj*, MPLUS and the R software package *Flexmix*, which was use to produce the solution reported here.

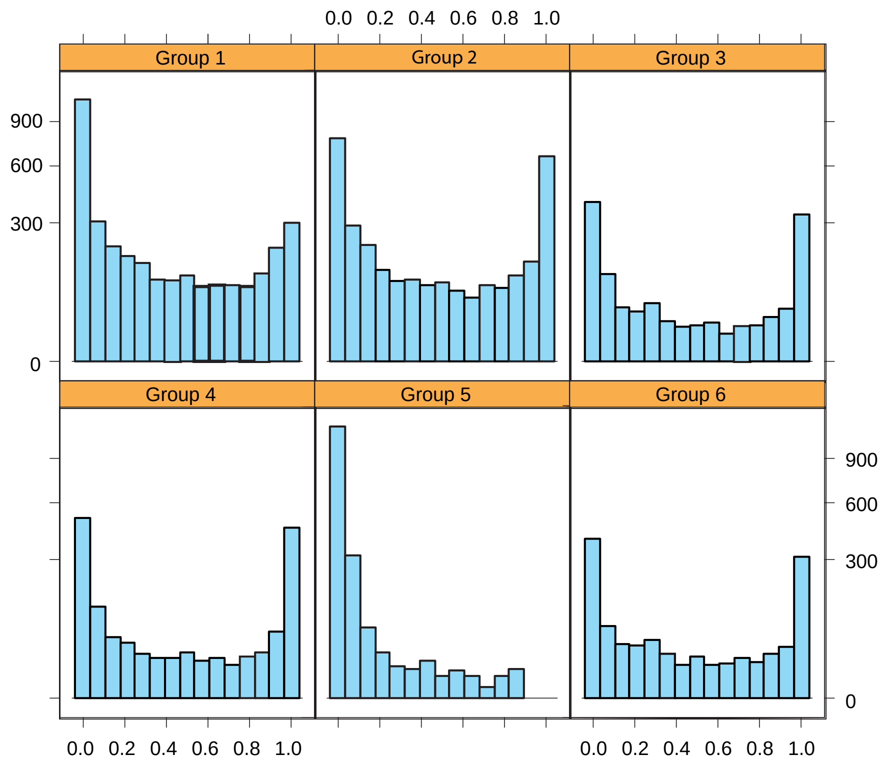
Wage trajectory clusters

The first stage of trajectory analysis involves the choice of a suitable probability model. In this study we use normal distribution, in which the wage profile is modelled using a linear model:

$$Wage_{it} = \beta_0 + \beta_1 Age_{it} + \beta_2 Age_{it}^2 + \beta_3 Workday_{it} + \beta_4 Career_{it} + \beta_5 GDP_t + \beta_6 Employers_{it} + \varepsilon_{it}$$

The model is computed for each individual *i* over the period under review *t*, *t* = 2000, … , 2010. However based on the dataset monthly wages are not normally distributed. Box-Cox transformation was applied to find the appropriate transformation. The value of the parameter needed for the transformation was estimated using the basic model: this yielded an estimate value of approx. λ = 0.5. The transformation obtained corresponds roughly to the square root transformation, which is less commonly used for wage modelling purposes. The more common method of log transformation was also used, but the results with this method were not as good. Annex Figures 2 and 3 illustrate how well the log and square root transformations perform in our sample. They show that the log transformation does not perform satisfactorily at the higher and lower end of the income distribution. The square root transformation worked over a wider income range; only in the very highest income percentages is the assumption of normality not fully satisfied.

Figure 2.
Rootgram of posterior probabilities › 0.001.



The next stage of trajectory analysis is to search for the optimum number of trajectories (or groups). Software packages provide various tools for this purpose, such as criterion functions AIC and BIC and various diagnostic graphs (e.g. rootgram in *Flexmix*, see Fig. 2). Another important criterion is that the solution obtained is interpretable. In our sample we arrived at six trajectory solutions. As is shown in Figure 2, the material is well distributed by the six components across the different trajectories because the computed posterior probabilities are close to zero or one (U-shape). The solution obtained is still interpretable, and no component was estimated as being too small.

The solution is illustrated in Figure 3 below, which shows the development of wages as a function of age separately for each trajectory. Table 3 shows the basic results for the trajectory solution.

Table 3.
Trajectory solution results.

Group	Prior Prob	Size	Males %	Females %
1	0.243	1116	44.6	55.4
2	0.296	1331	34.1	65.9
3	0.114	542	55.7	44.3
4	0.150	703	73.3	26.7
5	0.084	345	73.6	26.4
6	0.110	524	37.8	62.2

Women have a slightly majority in our initial sample: the number of men is 2,221 (48.7%) and the number of women 2,340 (51.3%). Trajectory 1 comes closest to the gender breakdown in the sample, while the other trajectories differ quite clearly from one another. Men are in the majority in trajectories 3, 4, and 5, women in turn have a clear majority in trajectories 2 and 6.

Table 4 describes the career backgrounds of the individuals in different trajectory groups. Individuals in groups 1, 2 and 5 come predominantly from a wage-earning background in the private and public sector. People with a public sector background are overrepresented in groups 3 and 6; these are mainly civil servants working in the public sector. In terms of sector of employment, group 4 most closely resembles group 3.

We have seen earlier that persons with a private and public sector wage-earning background in groups 1 and 2 are predominantly women. Women are also overrepresented in group 6, which has the largest proportion of persons coming from a civil service background. Groups 3, 4 and 5 are male-dominated, but with the exception of group 3 the proportion of public sector wage earners is relatively small.

Annex Table 1 provides a more detailed picture of the statistical model's explanatory factors in each group. It is clear from the Table that the group compositions are rather heterogeneous in terms of age. The mean age is lowest in group 1 (34.7 years) and highest in group 6 (45.3 years). Based on the quartile range (P25–P75), persons in group 1 are aged 24–45 years, in group 2 aged 27–48 years, in group 3 aged 37–52 years, in group 4 aged 31–50 years, in group 5 aged 29–51 years and in group 6 aged 39–53 years.

The dependent variable, i.e. monthly wages, varies markedly between groups. There are two underlying reasons for this. Firstly, the gender wage gap is reflected in our groups. Secondly, because of their different age compositions, the groups are formed of people who are at different stages of their careers. Groups 1, 2 and 6 have the lowest median wages (less than 2,050 €/month). Group 5 has by far the highest median wage at 4,905 €/month.

Table 4.
Employment in private and public sector by group.*

Group	Private	Public	Private & public	Private (%)	Public (%)	Private & public (%)
1	442	61	613	39.6	5.5	54.9
2	448	109	774	33.7	8.2	58.2
3	297	58	187	54.8	10.7	34.5
4	392	49	262	55.8	7.0	37.3
5	155	15	175	44.9	4.4	50.7
6	255	81	188	48.7	15.5	35.9

*Public = Central and local government sector.

Accrued pension by group

Information on the amount of accrued pension as at 31 Dec 2004 is available for all persons in the sample. Accrued pension rights describe the individual's current level of pension based on his or her employment career to date. In the case of older age groups, current accrued pension already gives a rough indication of anticipated final pension.

Age limits for the accrual of earnings-related pension have changed during the period under review (2000–2010). Prior to 2005 pension rights began to accrue from age 23, whereas since 1 Jan 2005 the age limit for pension accrual was lowered to 18 years. Both these rules are applied to the persons in our sample. For older age groups (up to 1982) pension rights have accrued from age 23, whereas for younger age groups progressively from age 18.

An examination of accrued pension makes most sense in the case of persons who are approaching general retirement age (which in Finland is 63 years). Table 5 shows the current level of accrued pension for employed persons aged 55–60.

Table 5.
Accrued pension as at 31 Dec 2004 for persons aged 55–60, median.

Group	Accrued pension 2004, €/month
1	744
2	750
3	1153
4	1366
5	1855
6	977

The six groups differ quite markedly from one another. In groups 1 and 2 the level of earnings-related pension looks set to remain quite low, whereas in groups 3 and 6 the accrual rate suggests a future pension that will be around the average for all pension recipients. Groups 4 and 5 clearly include individuals whose careers and earnings have developed more favourably than average. Assuming these individuals do not fall victim to labour market or population risks[1], the level of old age pension in group 5 will be very high.

Explanatory factors

Career length mainly reflects the age distribution within the group. The older the individual in question, the longer their career on average. For instance, persons in group 6 have an average employment career of 32.,1 years. In group 1, where the persons are the youngest, the average length of career is just 12.,4 years. Of course career is not directly explained by age, but it will also be affected by unemployment and other possible interruptions.

The annual number of employers can be interpreted by reference to the wage effect of changing jobs. There are no marked differences in this regard between the groups, which is quite natural. Most people continue to work with the same employer, although some job changes and possibly some spells of unemployment are bound to occur during the period under review. However an examination of means and medians brings to light some differences especially with regard to wages in the highest earning group (group 5) and in the lowest earning groups (groups 1 and 2). These two sharply differing groups have a much higher number of job changes than other groups when assessed on the basis of means and the upper quartile P75. The reasons for this may differ, however. In low income groups job changes are probably due more to fragmented employment careers rather than career advancement. One might expect to see people in the high income group change jobs more often as a result of career advancement.

The final variable included in the model to explain wage profiles is the amount of work done per year. Based on the median and quartiles (P25 and P75) it is very common for all groups to work full years, i.e. 360 days a year. However there are also some individuals who work a fewer number of days a year, which might be due to their taking up or terminating an employment contract in the middle of the year. Working for less than a full year does not necessarily imply a negative or adverse career turn. Based on an examination of means we find that groups 1 and 2 have the shortest work years (312 and 302 days, respectively). The longest year is recorded for group 3 at 347 days. The differences are not very great. The explanatory variable in our model, i.e. monthly wage, is effective wage, and therefore working less than a full year does not necessarily mean a poor year in terms of earnings.

The model equation estimated from the sample is thus of the form:

$$\sqrt{Wage_{it}} = \beta_0 + \beta_1 Age_{it} + \beta_2 Age_{it}^2 + \beta_3 Workday_{it} + \beta_4 Career_{it} + \beta_5 GDP_t + \beta_6 Employers_{it} + \varepsilon_{it}$$

1 These risks include death, long-term unemployment and inability to work.

In this equation *i* refers to individuals and *t* to the years under review 2000–2010. When calculated for the whole sample as in traditional regression analysis, the coefficient of determination is about 23% (R = 0.232). Annex Tables 3 and 4 describe the parameter estimates more precisely.

Table 6.
Goodness of the fit.

Group	Multiple R-squared	Residual standard error
1	0.567	8.161
2	0.217	12.110
3	0.578	3.140
4	0.530	5.808
5	0.397	18.580
6	0.557	3.042

Table 6 shows the coefficients of determination for each group. They are at a consistently high level. With the exception of group 2 they are in practice clearly higher than in the model estimated for the whole dataset.

The results of regression analysis discussed above are intended to provide a rough description of the trajectory solution. Traditional regression analysis does not take account of the fact that the dataset investigated is longitudinal, which means that observations cannot be considered independent.

A linear mixed model was used in order to obtain a more accurate description of the trajectory solution, with the dependence of observations modelled by means of random effects. The random effects included in this model (see Annex Table 5) were intercept and age, which are assumed to be normally distributed, but independent of random errors that were assumed to be independent and normally distributed.

4 Estimation results

Overall the explanatory factors included in the model seem to work quite well. With the exception of the number of employment contracts, all variables are highly significant in all groups. Annex Table 2 shows the parameter estimates for the mixed models related to each trajectory. Interpretation of the parameters is complicated by the square root transformation applied to wages. If the transformation is not reversed, the parameters are interpreted in relation to the square root of wages. We were also keen to interpret some of the coefficients without square root transformation. Annex 1 describes the method that has been used to assess the contribution of the variable to the original untransformed response variable.

The parameter values obtained by the age terms are in line with expectations. Initially wages tend to rise, but at around age 50 the effect of age turns negative. The age parameter (β_1) showed the highest values in groups 1 and 5 that had strong career profiles. The second term of age (β_2) is negative in all groups (Annex Figure 4).

Amount of work input (β_3) is rather harder to interpret. In groups 3, 4 and 6 the parameter has a positive value, suggesting that an increased labour input increases wages as well. On the other hand in groups 1, 2 and 5 the parameter has a negative value, which suggests the opposite conclusion. The distribution of the variable is highly skewed towards full years (360 days a year), which means that the explanation for the parameter probably follows from individuals whose employment is interrupted during the course of the year. There may be many reasons for this. One explanation has to do with job changes and atypical employment more generally. A negative parameter is observed for the female-dominated groups 1 and 2 and to the 'career-making' group 5.

Career length (β_4) has a positive impact on the development of earnings in all groups. The effect is strongest in group 5. Even in group 1, career length has a positive effect on wages. If the parameter is used to assess[2] the one-year wage effect at the mean point for each trajectory (when all other variables have been scaled to match their means), in group 5 the effect is around 157 euros. In other groups the effect is smaller. In practice career length has hardly any wage effect in groups 3 and 6 (Table 7) – although the estimate is computed at the mean of employment career, which means that the age effect often turns earnings onto a downward trend.

Economic growth (β_5) has a predominantly positive effect on the development of earnings. Employment effect is a separate concern altogether, but that cannot be directly seen in this dataset. Over the period from 2000 to 2010, real GDP increased on average by 1.8% a year. The mean GDP value stood at around 172 billion euros. When the parameters are used to assess the wage effect of a one million euro increase in mean GDP, it increases in group 5 to 47 euros. In other groups the effect is quite small, and in group 2 economic growth actually has a negative wage effect (Annex Figure 5).

2 The parameter value is equivalent to the effect of 0.1 years on wages.

The effect of changing jobs (β_6) is visible in some groups. In groups 1, 2 and 6 the parameter is highly significant, and in groups 3 and 4 significant. In all groups the wage effect is negative. This is a rather confusing result, for one would expect to see job changes have a clearer positive wage effect. However based on our material this is not the case. In group 5 the parameter is not significant. Apparently the wage effects for the individuals in this group largely cancel each other out, making it impossible to detect any clear direction.

Table 7.
Parameter effects on monthly wages, €.

Group	GDP effect	Career effect
1	27	114
2	-4	55
3	10	28
4	21	60
5	47	157
6	6	27

Age-wage profiles in practice

The estimated parameters provide first-hand information about the shape of the wage profile. The results are based on a fit of a second order model, which in these groups take a rather predictable shape. The parameters of the age terms explaining wages have the 'right' signs and are significant. The second order term of age is indisputably negative in all cases. The parameter estimates do not provide a very clear picture of the variation appearing in the data.

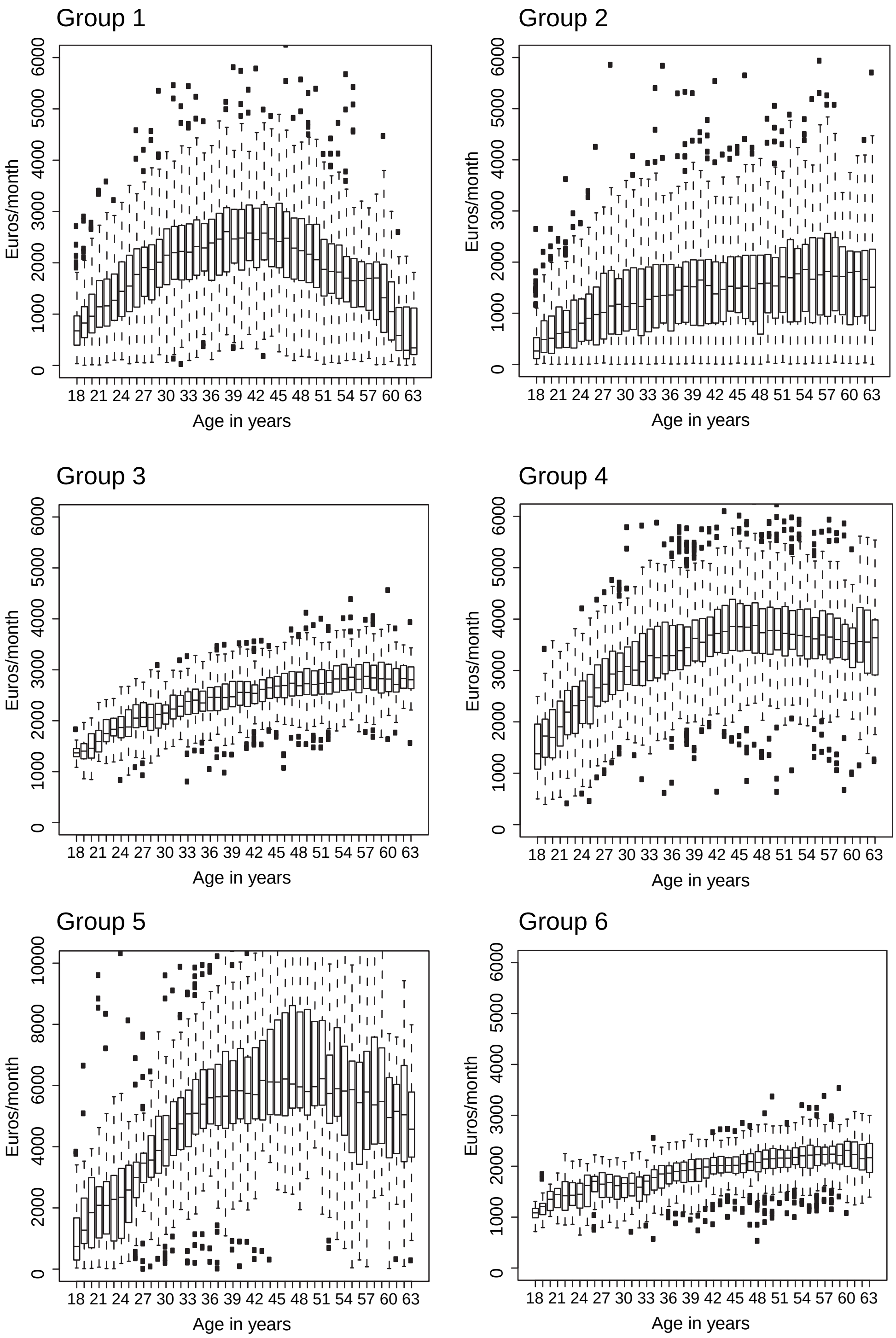
We proceed now to look at how age-wage profiles vary in practice. In order to illustrate the point we have chosen to display the wages in each group in the form of boxplots. The boxplots show the minimum value of wages, the median limit for the lowest income quartile (P25), the limit for the highest income quartile (P75) and the highest observations. The boxplot makes no assumptions about the statistical distribution of the material, so in this sense it is entirely non-parametric.

Figure 3 reveals some interesting points. Firstly, a rising wage profile is typical in all groups. In practice median wages increase with advancing age and increasing experience. Based on visual overview wages increase in all age groups from 18 to 30. In most groups wages continue to rise through to the very last stages of the employment career, which is consistent with observations made earlier. However it is important to note that a selection process takes effect among individuals in wage employment as they approach retirement age. After age 55 inability to work and unemployment begin to reduce the numbers in employment. People holding the strongest positions in the labour market are more successful in retaining their jobs. In practice, wage earners aged 60 or over are represented by white-collar employees, which obviously accelerates the growth of wages. In groups 2, 3 and 6, wages continue to rise virtually to the very end.

Secondly, the groups differ from one another in their wage levels. In the biggest group (2) wages remain quite low; gross median monthly earnings at the end of the employment career rise to 1,700 euros. The situation is very similar in group 6. Wages are indicated in 2010 prices, which means that the figures show the development of real wages. It is quite surprising to find that there are so many people in these low income brackets (more than 40% of the sample belong to groups 2 and 6). It is worth pointing out that during the period under review, Finnish wage earners' average monthly pay was 2,300 euros. The closest fit to this would be the wage profile of group 3. The trajectory analysis also throws up some surprises. Wages in group 5 represent the very high end of the Finnish wage distribution. Thirdly, patterns of age-wage variation differ considerably between the groups. Groups 3 and 6 show the least variation in relation to quartile points (P25–P50) and extreme values. In groups 1, 2 and 5 there is marked variation, which is even reflected in the annual variation of the median.

We saw from Table 5 that towards the end of the employment career, at age 55–60, there were marked differences between the groups in accrued pension rights. Groups 1 and 2 had the lowest expected pension (median about 750 euros). Groups 4 (median 1,366 euros) and 5 (median 1,855 euros), on the other hand, could look forward to a higher earnings-related pension. Replacement ratio serves as a measure of pensioners' income in relation to their earlier income. The same logic can be applied while people are still in wage employment by calculating the ratio between accrued pension and earnings. Calculations of mid-career expected replacement ratio for persons still working in 2004, i.e. the ratio between accrued pension rights and earnings, is 52–60% in groups 1, 2 and 6; 49% in group 3; and about 45% in groups 4 and 5.

Figure 3.
Age-wage profile for groups 1–6. Note the different scale for the y-axis in trajectory 5.



5 Discussion

Factors impacting the age-earnings profile have been an area of long-standing interest in applied microeconomics research. It has been shown that education and work experience has a significant impact on the individual's earnings over time. The shape of the age-earnings profile is an interesting subject of study in the sense that it sheds useful light on how earnings develop towards the end of the labour market career, when people approach retirement age.

One of the primary aims of this report was to identify and describe typical earnings profiles in Finland. The statistical method applied was trajectory analysis, a modern technique that is particularly suited for the clustering of longitudinal data. Statistically, the model used in this study showed good fit with the data. The dataset comprised a sample of 5,000 wage earners in Finland. Drawn from administrative registers and containing details about each individual's earnings and employment career, the dataset covers the period from 2000 to 2010.

The range of earnings profiles identified was quite heterogeneous. The trajectory analysis distinguished six groups that allowed us to describe the most distinctive age-earnings profiles. The earnings profiles of most people included in the sample are in line with the moderate pay rises agreed through collective bargaining. However the analysis also identified a group with a high profile career and on the other hand groups showing weaker labour force attachment or earnings profiles.

Age-earnings profiles are described not only by means of clustering, but also using statistical models. A mixed model developed to explain the earnings profile of an individual trajectory group includes explanatory factors related to the individual's age, employment career and the general economic environment. The evidence is clear that group modelling is a more effective technique than using a single model applied to the whole sample using the same explanatory factors. According to the estimation results real wages begin to fall in all groups after age 50. Length of career or work experience has a positive effect on earnings. A favourable general economic climate is reflected not only in employment rates, but also in the development of wages. With the exception of one group, GDP growth drives up earnings as well.

The results are consistent and have interesting implications for the pension system. It is apparent that age-earnings profiles cannot be exhaustively described by just one single model that is based on the total employed population. Statistical classification of the data reveals many different kinds of profiles. Furthermore, there is marked intra-group variation in earnings. Under the earnings-related pension system the accrual of pension rights is based on employment and earnings. In the age group 55–60 years the expected ratio of earnings-related pension to earnings (the expected replacement ratio) is 44–60%. This figure is fairly constant across all groups. A strong earnings profile is also reflected in a high expected pension. The high profile group can look forward to an earnings-related pension that ranks among the top decile of highest pensions. In low earnings groups the earnings-related pension will probably remain so low that it will need to be topped up by the national pension.

Earnings profiles are most commonly estimated using a second order age term. In the next stage of our research we plan to review this tradition by using this same dataset. Although a second order model often provides quite a good approximation for wage development, our own experience is that it does not always adequately describe wages and their development, particularly in the later stages of the individual's employment career. One option for modelling of the development of earnings is to use methods of non-parametric regression, such as splines, where the shape of the function is not tied to any particular parametric curve shape.

Annexes

Annex 1.

The following procedure was used to clarify the interpretation of estimated parameters. Let us first consider the model

$$y(x) = (a + bx)^2 + \varepsilon^2,$$

where *x* is the variable under analysis, *a* is constant and ε is a normally distributed error term. An approximation to the fitted model at *x** is obtained by using the first two terms of Taylor's series expansion

$$\hat{y}(x) \approx \hat{y}(x^*) + \hat{y}(x^*)(x - x^*).$$

Our focus of interest now turns to variable x, whereby at *x** it is possible to approximate

$$\widehat{y'}(x^*)x \approx \left(2\hat{a}\hat{b} + 2\hat{b}\hat{b}x^*\right)x.$$

When the value of explanatory factor x at *x** changes by one unit, the contribution to the response is on average approximately

$$2\hat{b}\left(\hat{a} + \hat{b}x^*\right),$$

where the standard term $\hat{a}$ includes the values of the intercept and the values of the other variables from analysis fitted to their mean, and $\hat{b}$ is the estimate of the regression coefficient of variable *x*.

Annex Table 1.
Descriptive statistics of the sample by group.

Group 1	Cohort	Age	Wages	Career	Employers	Workdays
Min	1938	14.0	4	0.0	1.0	2
P25	1960	24.0	1037	3.0	1.0	340
Median	1973	32.0	1768	8.4	1.0	360
Mean	1971	34.7	1789	12.4	1.8	312
P75	1982	45.0	2432	20.2	2.0	360
Max	1992	68.0	7887	49.9	41.0	360
Group 2	**Cohort**	**Age**	**Wages**	**Career**	**Employers**	**Workdays**
Min	1936	14.0	1	0.0	1.0	2
P25	1957	27.0	553	3.7	1.0	284
Median	1968	37.0	1213	12.0	1.0	360
Mean	1967	37.9	1314	14.2	2.0	302
P75	1978	48.0	1904	23.2	2.0	360
Max	1992	68.0	13960	49.9	46.0	360
Group 3	**Cohort**	**Age**	**Wages**	**Career**	**Employers**	**Workdays**
Min	1935	14.0	789	0.0	1.0	18
P25	1953	37.0	2288	14.0	1.0	360
Median	1960	45.0	2583	22.5	1.0	360
Mean	1961	43.6	2561	21.8	1.2	347
P75	1967	52.0	2862	29.9	1.0	360
Max	1992	67.0	4564	47.9	35.0	360
Group 4	**Cohort**	**Age**	**Wages**	**Career**	**Employers**	**Workdays**
Min	1936	15.0	392	0.0	1.0	2
P25	1955	31.0	2801	8.4	1.0	360
Median	1964	40.0	3384	16.9	1.0	360
Mean	1965	40.3	3344	17.8	1.3	336
P75	1975	50.0	3931	26.7	1.0	360
Max	1992	68.0	7956	49.1	25.0	360
Group 5	**Cohort**	**Age**	**Wages**	**Career**	**Employers**	**Workdays**
Min	1965	15.0	2	0.0	1.0	1
P25	1954	29.0	3244	6.3	1.0	360
Median	1967	39.0	4905	14.8	1.0	360
Mean	1965	39.7	5271	16.3	1.7	328
P75	1976	51.0	6321	25.6	2.0	360
Max	1992	68.0	124500	47.5	52.0	360

Group 6	Cohort	Age	Wages	Career	Employers	Workdays
Min	1936	16.0	533	0.0	1.0	2
P25	1952	39.0	1778	15.8	1.0	360
Median	1957	47.0	2026	24.7	1.0	360
Mean	1960	45.3	1991	32.1	1.2	341
P75	1966	53.0	2245	31.3	1.0	360
Max	1992	68.0	3532	47.7	21.0	360

Annex Table 2.

Parameter estimates by group.

	Estimate	Std. Error	t value	Pr(>|z|)
Group 1				
Intercept	-63.519	1.618	-39.262	0.000***
Age	3.724	0.060	62.310	0.000***
Age^2	-0.058	0.001	-70.235	0.000***
Workday	-0.012	0.001	-11.200	0.000***
Career	1.108	0.029	37.713	0.000***
GDP	0.260	0.008	33.881	0.000***
Employers	-0.409	0.041	-9.885	0.000***
Group 2				
Intercept	9.940	2.463	4.035	0.0001
Age	1.617	0.087	18.576	0.000***
Age^2	-0.023	0.001	-20.313	0.000***
Workday	-0.009	0.001	-6.799	0.000***
Career	0.723	0.039	18.472	0.000***
GDP	-0.048	0.011	-4.383	0.000***
Employers	-0.070	0.049	-1.431	0.152
Group 3				
Intercept	9.793	1.324	7.394	0.000***
Age	0.589	0.058	10.172	0.000***
Age^2	-0.007	0.001	-10.576	0.000***
Workday	0.016	0.001	15.459	0.000***
Career	0.289	0.020	14.648	0.000***
GDP	0.100	0.004	26.428	0.000***
Employers	-0.163	0.039	-4.152	0.000**

Group 4				
Intercept	-13.340	1.716	-7.773	0.000***
Age	1.707	0.075	22.905	0.000***
Age^2	-0.022	0.001	-24.192	0.000***
Workday	0.007	0.001	6.138	0.000***
Career	0.505	0.034	14.907	0.000***
GDP	0.172	0.006	28.930	0.000***
Employers	-0.227	0.069	-3.271	0.001
Group 5				
Intercept	-80.757	7.184	-11.241	0.000***
Age	5.637	0.277	20.313	0.000***
Age^2	-0.068	0.004	-18.753	0.000***
Workday	-0.064	0.004	-15.035	0.000***
Career	0.968	0.154	6.301	0.000***
GDP	0.290	0.027	10.855	0.000***
Employers	0.147	0.119	1.230	0.219
Group 6				
Intercept	18.224	1.286	14.175	0.000***
Age	0.302	0.052	5.840	0.000***
Age^2	-0.004	0.001	-6.732	0.000***
Workday	0.008	0.001	9.844	0.000***
Career	0.311	0.017	18.650	0.000***
GDP	0.068	0.004	17.742	0.000***
Employers	-0.496	0.038	-13.205	0.000***

Annex Table 3.
Parameter estimates for whole sample (fixed effects).

	Estimate	Std. Error	t value	Pr(>|z|)
Intercept	-22.410	2.093	-10.710	0.000***
Age	2.205	0.042	52.360	0.000***
Age^2	-0.030	0.001	-57.360	0.000***
Workday	0.007	0.001	7.670	0.000***
Career	0.621	0.016	39.510	0.000***
GDP	0.120	0.007	18.450	0.000***
Employers	-0.658	0.036	-18.400	0.000***

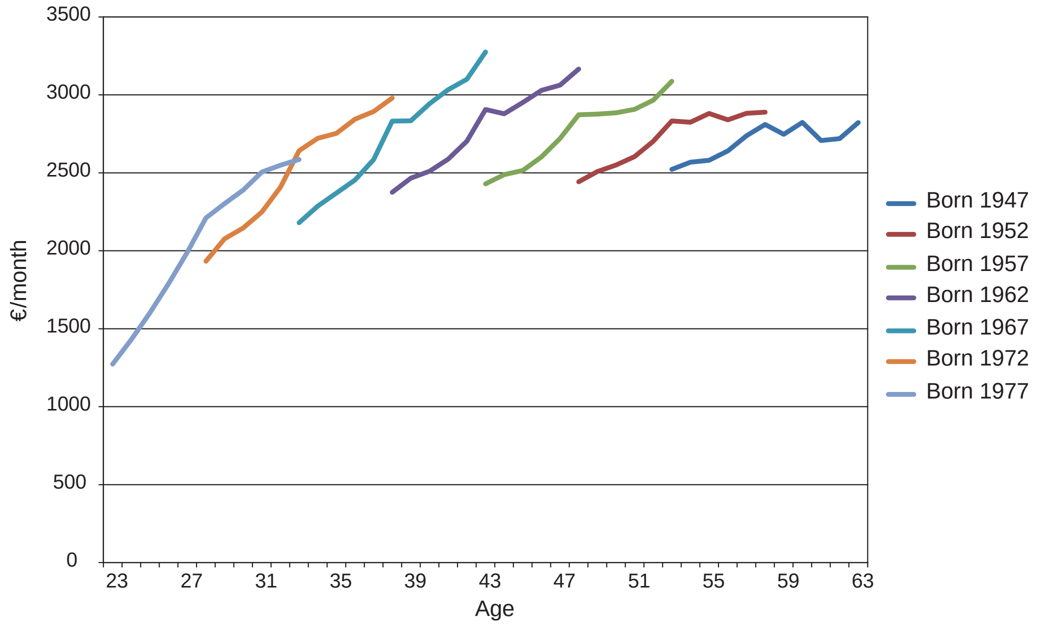
Annex Table 4.
Correlation matrix for fixed effects.

Effect	Intercept	Age	Age^2	Workday	Career	GDP	Employers
Intercept	1.0000	-0.3538	0.3261	-0.02211	0.0776	-0.5424	-0.0268
Age	-0.3538	1.0000	-0.9365	-0.2740	-0.1491	0.0389	0.0751
Age^2	0.3261	-0.9365	1.0000	0.3152	-0.1665	-0.0589	-0.0996
Workday	-0.0221	-0.2740	0.3152	1.0000	-0.2331	-0.0195	-0.1256
Career	0.0776	-0.1491	-0.1665	-0.2331	1.0000	0.0315	0.1247
GDP	-0.5424	0.0389	-0.0589	-0.0195	0.0315	1.0000	-0.0294
Employers	-0.0268	0.0751	-0.0996	-0.1256	0.1247	-0.0294	1.0000

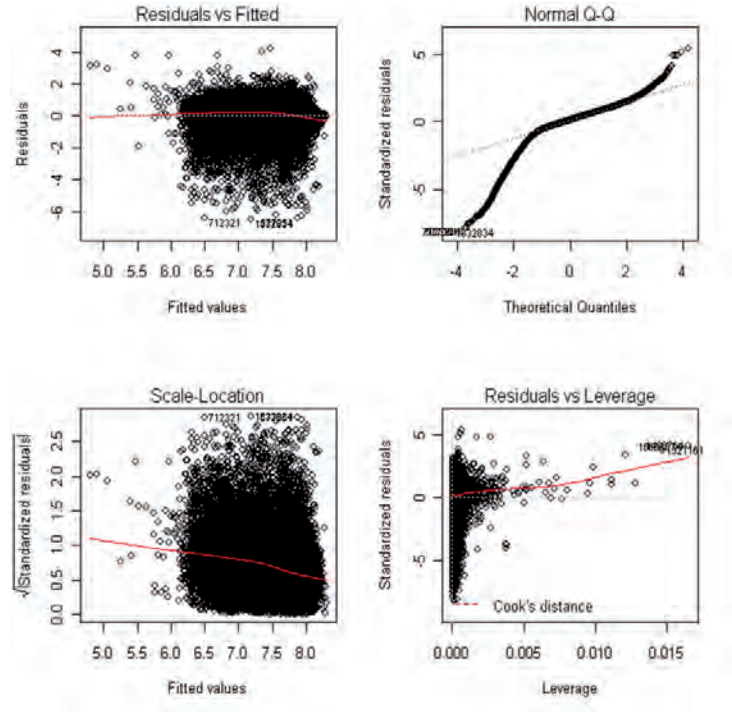
Annex Table 5.
Solution for random effects.

Effect	Estimate	Std. Error Pred.	DF	t value	Pr > |t|
Intercept	1.09E-12	1.5824	36E3	0.00	1.0000
Age	5.64E-18	0.003222	36E3	0.00	1.0000

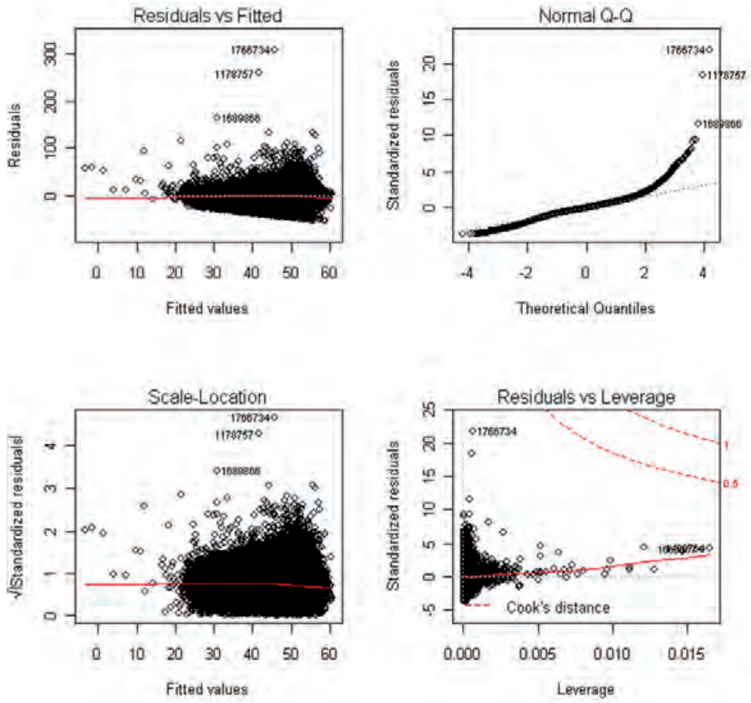
Annes Figure 1.
Cohort wage profiles.



Annex Figure 2.
Residual diagnostics for log wages.

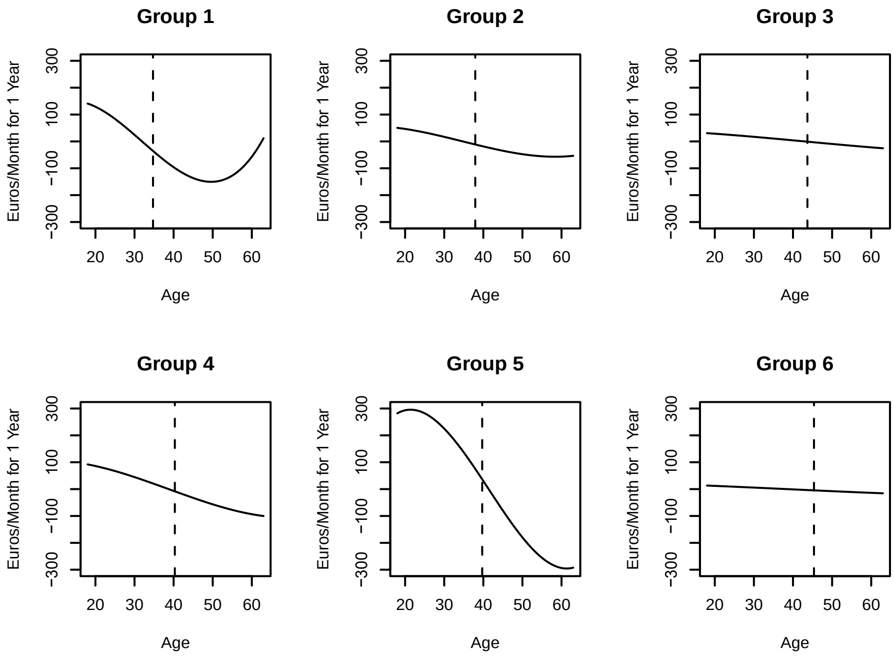


Annex Figure 3.
Residual diagnostics for sqrt wages.



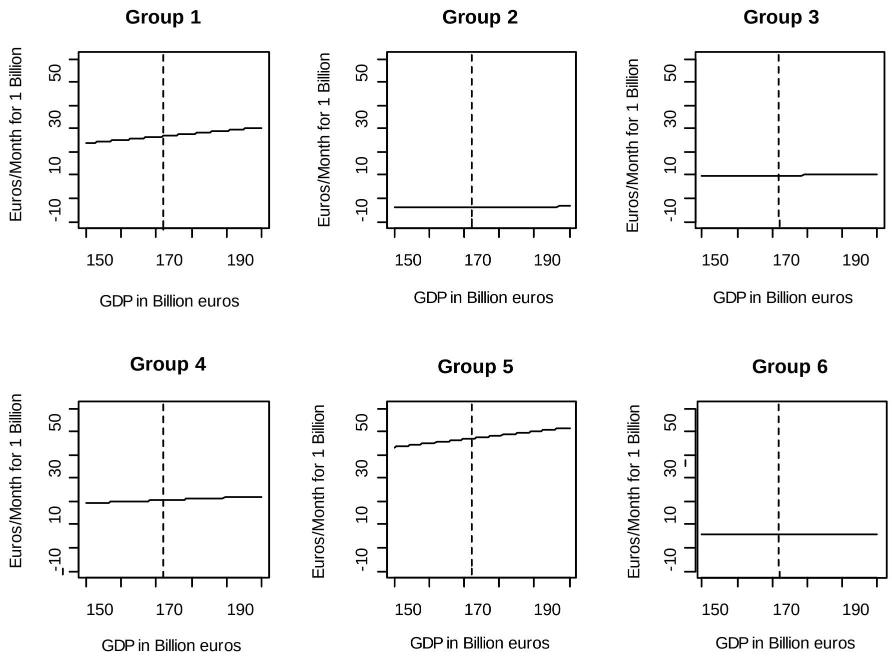
Annex Figure 4.

Effect of age on response variable when other variables in the model are fitted to their mean. The broken line marks the mean age for the group.



Annex Figure 5.

Effect of GDP growth on response variable when other variables in the model are fitted to their mean. The broken line marks the mean GDP.



References

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.

Dickerson A. and McIntosh S. (2011) An investigation into the relationship between productivity, earnings and age in the early years of a working life. Report Prepared for the Low Pay Commission,
Department of Economics, University of Sheffield.

Elo K. and Salonen J. (2004) Age-Wage Profiles for Finnish Workers. Nordisk Försäkringstidskrift 4/2004.

Ilmakunnas, P. and Maliranta M. (2005) Technology, worker characteristics, and wage-productivity gaps. Oxford Bulletin of Economics and Statistics, vol. 67, 623–645.

Jones, B., Nagin D. and Roeder K. (2001) A SAS procedure based on mixture models for estimating developmental trajectories. Sociological Methods & Research, 29, 374–393.

Kjeldstad R. and Nymoen E. (2012) Part-time work and gender: Worker versus job explanations. International Labour Review, vol. 151, No 1–2.

Korkeamäki O. and Kyyrä T. (2006) A gender wage gap decomposition for matched employer-employee data. Labour Economics, vol. 13(5), pp. 611–638.

Kyyrä T. (2007) Gender Wage Differentials: Evidence from Matched Employer-Employee Data. VATT Research reports 133.

Liukkonen V. (2012) Non-standard Employment and Health with Respect to Sense of Coherence and Social Capital. University of Tampere.

Mincer J. (1974) Schooling, Experience, and Earnings. Columbia University Press.

Muthén B. (2001) Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), New developments and techniques in structural equation modeling (pp. 1–33). Lawrence Erlbaum Associates.

Nagin D. (1999) Analyzing developmental trajectories: Semi-parametric, group-based approach. Psychological Methods, 4, 39–177.

Nagin D. (2005) Group-based modeling of development. Cambridge, MA: Harvard University Press.

Virtanen P., Lipiäinen L., Hammarström A., Janlert U., Saloniemi A. and Nummi T. (2011) Tracks of labour market attachment in early middle age: A trajectory analysis over 12 years. Advances in Life Course Research 16, 55–64.

The Finnish Centre for Pensions is a statutory co-operation body, expert and producer of joint services for the development and implementation of earnings-related pension provision.
The aim of our research is to produce high-quality, widely applicable information for the evaluation and development of pension provision.

Eläketurvakeskus on työeläketurvan kehittämisen ja toimeenpanon lakisääteinen yhteistyöelin, asiantuntija ja yhteisten palveluiden tuottaja. Tutkimustoiminnan tavoitteena on tuottaa korkeatasoista ja laajasti hyödynnettävää tietoa eläketurvan arvioimiseen ja kehittämiseen.

Pensionsskyddscentralen är ett lagstadgat samorgan och sakkunnig inom verkställigheten och utvecklingen av arbetspensionsskyddet. Vi producerar gemensamma tjänster för arbetspensionssystemet. Vår forskning har som mål att ta fram högklassig information som nyttiggörs på bred front vid bedömningen och utvecklingen av pensionsskyddet.

Finnish Centre for Pensions
ELÄKETURVAKESKUS