Session Number: Parallel Session 8D
Time: Friday, August 27, PM

Paper Prepared for the 31st General Conference of
The International Association for Research in Income and Wealth

St. Gallen, Switzerland, August 22-28, 2010

Spaghetti Unravelled: A Model-Based Description of Differences in Income-Age Trajectories

Stephen P. Jenkins

For additional information please contact:

Name: Stephen P. Jenkins
Affiliation: University of Essex

Email Address: stephenj@essex.ac.uk

Spaghetti Unravelled: A Model-Based Description of Differences in Income-Age Trajectories

Stephen P. Jenkins
Institute for Social and Economic Research
University of Essex

No. 2009-30
November 2009

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

ISER Working Paper Series

www.iser.essex.ac.uk

Spaghetti Unravelled:
A Model-Based Description of Differences in Income-Age Trajectories

Non-Technical Summary

Most descriptions of the income-age relationship are based on comparisons of income across age groups in a particular year and are based on cross-sectional data. In contrast, this paper has taken a longitudinal approach, deriving trajectory estimates using 17 waves of data from the British Household Panel Survey. I propose a framework that provides summary descriptions of not only the way in which incomes among groups of similar individuals change with age on average, but also the way in which trajectories for individuals diverge from the average trajectory of their group. The model is applied to three measures of 'income': the hourly wage, total individual income from all sources, and equivalized net household income.

The paper's main points are as follows:

- Individuals' income-age trajectories collectively look like cooked spaghetti – they are a complex mix of wiggly lines.
- We can, however, use a statistical model to summarize the key features of these trajectories and to highlight differences across groups on average and within groups of individuals with similar characteristics.
- Across twelve social groups defined in terms of combinations of sex, birth cohort, and educational qualifications, there are some clear differences in average income-age trajectories, regardless of which income measure is used.
- Other things equal, the average income-age profile for men lies above that for women; the one for individuals born in or after 1955 is above that for those born before 1955; and the that for individuals with educational qualifications to A-level or higher is above that for individuals with some qualifications which, in turn, is above the profile for individuals with no educational qualifications. There is a distinct dip in income growth for women on average over the age range when many have children.
- Average income-age trajectories derived from longitudinal data look different from those derived from cross-sectional data. For hourly wages for instance, trajectories at the beginning of the working life are steeper – wage growth is greater – according to longitudinal data.
- 'Average' trajectories are potential misleading. Within each social group, there are substantial differences across individuals in the shapes of income-age trajectories, where differences can be usefully summarised in terms of:
    1. Individual-specific differences in incomes at the start of the working life;
    2. Individual-specific differences in income growth rates; and
    3. A close association between initial incomes and income growth rates – those with a lower initial income experience greater income growth on average, so there is a tendency for trajectories to cross;
    4. Transitory variations – income-age trajectories also differ because of substantial individual-specific income changes from one year to the next, representing the effects on income of genuine transitory variation, measurement error, or lifecourse events such as having children, or family formation or dissolution.
- Over the working life, income inequality first declines and then rises, but the nature of the U-shape differs substantially between birth cohorts.

Spaghetti Unravelled:
A Model-Based Description of Differences in Income-Age Trajectories

Stephen P. Jenkins
Institute for Social and Economic Research
University of Essex
Colchester CO4 3SQ
United Kingdom
Email: stephenj@essex.ac.uk

Revised 30 October 2009

Abstract

A modelling framework is developed for describing income-age trajectories that is useful for summarizing not only the average profile for a group of individuals with similar characteristics, but also how individual trajectories differ from the group average. Using data from waves 1-17 of the British Household Panel Survey, the model is estimated separately for twelve groups of individuals differentiated in terms of educational qualifications, birth cohort and sex. The results indicate significant differences in the shapes of average trajectories across groups, and substantial variations in trajectories across individuals within groups.

JEL Codes: D31, C23, J31
Key words: lifecycle, income, wages, earnings, mobility, earnings, income-age trajectory

Acknowledgements

This research was commissioned by the UK National Equality Panel. Financial support from the ESRC and the University of Essex to ISER's Research Centre on Micro-Social Change is also gratefully acknowledged. For comments and suggestions, I am grateful to my NEP colleagues, Lorenzo Cappellari, Xavi Ramos, and Philippe Van Kerm, and participants in the workshop on "Earnings and Income Dynamics" at the National University of Ireland Maynooth, 3 July 2009.

1. Introduction

This paper provides new evidence for Britain about the shape of people's income-age trajectories – how income varies with age – and how these trajectories differ between individuals. I make comparisons within and between twelve groups defined in terms of educational qualifications, birth cohort, and sex, and for three income definitions (employment earnings, total individual income from all sources, and needs-adjusted household net income).

The research was commissioned by the UK's National Equality Panel (NEP).[1] The panel's report, to be published in January 2010, provides a comprehensive picture of the inequalities that exist in contemporary Britain. The NEP's work has three distinctive features. It focuses on inequalities in economic outcomes (household income, individual earnings, employment, and educational achievement), rather than the wider perspective taken by, for instance, the UK Equality and Human Rights Commission. The NEP aims to document inequalities both within and between 'social groups' defined in a number of ways, including by sex, ethnic minority group, social class, religion, region, and combinations of these characteristics. The panel is taking a lifecourse perspective, examining inequalities in childhood, the working life, and old age.

Most of the evidence currently available about the relationship between income and age is derived from cross-sectional data. The pictures of income-age trajectories are derived from survey data for a given year about a large sample of individuals of different ages. By contrast, my research uses longitudinal data that tracks the same people over time and accumulates information about the income-age trajectory for each person in the sample as each person ages. Data about how income varies between the age of 30 and 40 years (say) is derived by following 30 year olds over a decade until they are 40 rather than comparing today's 30 year olds with today's 40 year olds. If one is interested in documenting the nature of individual's income-age trajectories, including how income varies between one year and the next for each person, while also describing the heterogeneity across individuals in income-age trajectories, then a longitudinal approach is essential.

Knowledge of how income varies with age on average, and the extent to which individual trajectories differ from an average profile, is relevant to many aspects of social policymaking. How your income varies over your life is an important determinant of your spending possibilities (and hence consumption and economic well-being) at different ages, and your ability to save for old age, whether privately or through company, occupational, or state pension schemes. It is important to identify the characteristics of not only the groups who, on average, have persistently low incomes and hence low abilities to save, but also whether a 'group average' is potentially deceptive. Even if income increases with age on average, this is consistent with considerable year-on-year fluctuation in the incomes of a minority, or a mixture of subgroups with rising income and subgroups whose income is falling. These features complicate the design of effective policies for fostering saving by all.

It should be stressed, however, that this paper provides evidence relevant to policy discussion rather than an analysis of policy alternatives. The paper develops a framework to summarize

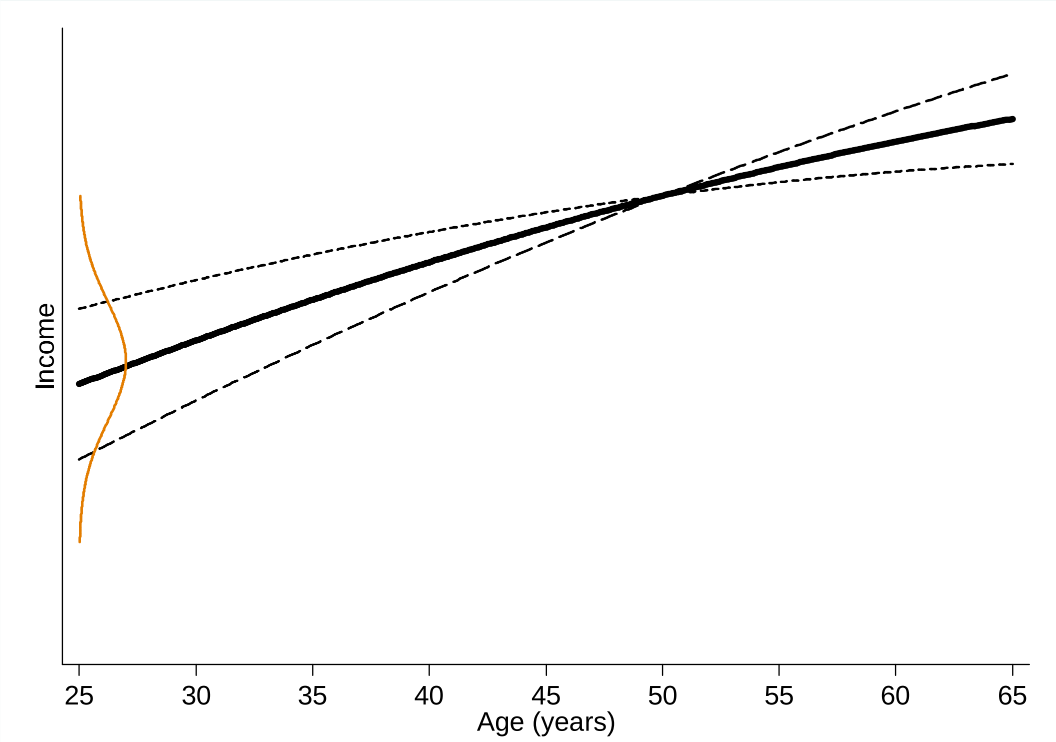
[1] The NEP was established in November 2008 by the British Minister for Women and Equality with a brief to provide an independent report on the nature of inequalities in Britain. The panel is chaired by Professor John Hills (LSE), and has nine other academic members. It is assisted by a secretariat from the Government Equalities Office. See http://www.equalities.gov.uk/national_equality_panel.aspx for further information.

individuals' income-age trajectories in a tractable manner, and applies it to data for Britain. I use the word 'summarize' intentionally for the shapes of income-age trajectories in contemporary Britain are complex, as I show below. Charts plotting raw survey data about income against age look like a plateful of cooked spaghetti. A statistical model is essential for characterizing the key features of income-age trajectories.

I propose an approach that incorporates elements of models from earlier literatures addressing different aspects of income dynamics. My brief from the NEP, in particular the requirement for analysis of different social groups and different income definitions within a common framework, led to an amalgam of these earlier approaches.

The key ideas underpinning my approach are as follows. First, I differentiate twelve 'social groups', with group membership defined in terms of similarity of birth year, educational qualifications and sex. Then, second, within each group, I summarise income-age trajectories in terms of an average group profile combined with individual-specific divergences from the group average. Figure 1 helps explain the idea (the formal statistical model is presented later).

Figure 1. Stylized income-age trajectories for two individuals and the average trajectory



Note. Chart shows stylized income-age trajectories for two individuals (dashed lines) and an average trajectory (solid line): see text for further explanation.

The dashed lines show stylized income-age trajectories for two individuals from the same social group (men born in the same year who both left school with GSCEs but without any A-levels, say). John's profile is summarised by a relatively low income at the beginning of the working-life (taken to be 25 here) combined with a relatively large growth rate in income with age (long dashed line). The other profile (short dashed line) combines a relatively high initial income but a relatively low growth rate in income with age – the slope of the trajectory is less steep than in the first case. Think of the first situation as characterizing someone who qualified as plumber. The starting salary is relatively low but increases over the working life, reflecting the return to the investment in training. The second situation represents Mike who instead trained as a motor mechanic. Initial earnings are higher than for John and remain so until both individuals are nearly 50 at which point, John's earnings are overtaken by Mike's. The solid line represents the average of the two individual profiles.

The key differences between John's and Mike's trajectories are, first, the difference in the initial incomes (one is below the average trajectory initially; the other is above the average initially) and, second, the difference in income growth with age (again one rate is above the average and the other is below it). A third important feature is that initial incomes and income growth rates are negatively correlated: John has a lower income than Mike to start with but experiences greater income growth. Their trajectories cross.

Now suppose that we wish to summarize the trajectories for all of the many individuals in this group, not only those for John and Mike. Given the average trajectory for the group as a whole, we can think of there being a distribution of initial incomes around the average and also a distribution of income growth rates, and some correlation between initial incomes and growth rates. Although most individuals within the group are located relatively close to the average, there are a few outliers either with relatively low (or high) initial incomes or growth rates. The relative frequencies of high and low deviations from the average initial income are illustrated in Figure 1 using the curvy solid line. Most people are located close to the average value at age 25 (the curve is 'higher'), with relatively small numbers with extreme values (where the curve is 'lower'). In the analysis below, the joint distribution of initial incomes and growth rates with age is characterized using a bivariate normal distribution. The advantage of this is that the distribution is completely characterised using only three numbers – the standard deviation of initial incomes around the average, the standard deviation of growth rates around the average, and the correlation between initial income and growth rate – and these parameters can be estimated from longitudinal survey data along with the parameters that describe the group average income trajectory. This characterization is consistent with both the trajectories increasing with age for a majority within the group, and declining with age for a minority.

The model implies that not only is there within-group inequality in income at each age, but also that this inequality varies with age. Intuitively, the less dispersion there is in initial incomes, or in income growth rates, the lower the within-group inequality at any age. Substantial dispersion in the income growth rate will tend to increase age-specific within-group inequality levels as the group members age. The cumulative effect of persistent differential income growth is to magnify initial income differences, providing an impetus for profiles to fan out with age.

The framework can also be used to illustrate differences in income trajectories between groups. It is straightforward to compare average patterns of income-age profiles using the estimated group average trajectories. One can also compare income dispersion at each age across the groups, examining for example whether at a given age, inequality among men is more or less equal than among women with similar educational qualifications and birth year. In addition, one can explore the extent to which income levels at each age overlap across groups, examining for example whether at a given age, even the poorest men earn more than the richest woman with similar educational qualifications and birth year, or whether there is substantial overlap in income levels.

The discussion has highlighted some key features of the statistical model employed in the empirical analysis. There are additional features that complicate the model to make it more realistic. These are discussed in greater mathematical detail later but two features in particular may be noted now. First, the group average trajectory is allowed to have more ‘wiggles’ than the stylised trajectories shown in Figure 1. Second, an additional year-by-year source of idiosyncratic variation in an individual’s income from the group average is introduced to account for the substantial longitudinal variability in incomes that arises in real life. This variation might conceivably arise from several sources, including genuine transitory variation, measurement errors in income, or reflect the impact on income of major life events such as the birth of a child or divorce.

As famous statistician George Box once said, ‘Essentially, all models are wrong, but some are useful’.[2] My models are definitely wrong, but I believe they usefully summarize the main features of income-age trajectories in a manner consistent with the NEP’s brief. The strengths and weaknesses of the modeling framework are discussed further later.

The rest of the paper unfolds as follows. In Section 2, I discuss the longitudinal data drawn from the British Household Panel Survey (BHPS) that are used in the study. In particular, I discuss the three different measures of income that I employ (hourly wages, individual income, and needs-adjusted household income), as well as the definitions of the twelve social groups characterized in terms of birth year, educational qualifications and sex. Section 3 provides a first look at the raw data on income-age trajectories and shows that they look like (cooked) spaghetti. The rest of the paper is concerned with unravelling that spaghetti.

Section 4 reviews related previous literature in order to provide context for my own approach. I refer to research on short-term income mobility, earlier descriptions of average income-age trajectories which are essentially an update of the celebrated portrayal of five alternating periods of want and plenty by Rowntree (1901), dynamic microsimulation modeling, and ‘variance component’ modelling which has been particularly concerned to identify the contributions of transitory and permanent inequality to total inequality overall. My approach is closest to the latter. Section 5 sets out my statistical model, using some mathematical notation to make the discussion above more precise. I also discuss a number of statistical issues that arise with fitting the model and interpreting the estimates.

The model estimates are reported in Section 6. I focus discussion on the results for income defined as the hourly wage for employees. All the analysis was repeated using income defined instead as total individual income and as needs-adjusted household income. The most significant differences between corresponding results for the different variables arise for the shapes of average trajectories rather than aspects relating to within-group differences around the average. So, I do discuss the differences in average trajectories across income variables in the main text. Other results for individual income and needs-adjusted household income are reported in Appendices A–C but not discussed for brevity. Section 7 provides a brief summary and conclusions.

[2] See http://en.wikiquote.org/wiki/George_E._P._Box.

2. Longitudinal data from the British Household Panel Survey (BHPS)

The analysis is based on longitudinal data from interview waves 1–17 of the British Household Panel Survey, corresponding to survey years 1991–2007.[3] The first wave of the BHPS was a nationally representative sample of the population of Great Britain living in private households in 1991. Original sample respondents (including both partners from dissolved marital partnerships) have been followed and they and their co-residents are interviewed subsequently at approximately one year intervals. Children in original sample households are also interviewed when they reach the age of 16 years. The first wave of interviews was in the Autumn of 1991, and subsequent interviews have also been in the Autumn: the modal interview months are September and October. The sample design ensures that the data collected are broadly representative of the population of Britain as it changed through the 1990s and 2000s. BHPS documentation of sampling design, questionnaire and variable definitions is available from http://www.iser.essex.ac.uk/survey/bhps/documentation.

Three measures of income

Three measures of 'income' are used in the analysis:
1. Hourly wage (£ per hour, expressed in January 2008 prices);
2. Individual income (£ per week, expressed in January 2008 prices); and
3. Equivalised net household income (£ per week, expressed in January 2008 prices).

The measures differ in the extent to which one can unabiguously attribute the income to a particular individual without employing assumptions about who benefits from it, in the number of income sources included in the income total, and the proportion of people who derive any income from the source.

The hourly wage refers to current usual employment income from a main job divided by the number of hours worked, assuming hours of overtime work are paid at time-and-a-half, and is expressed in pounds per hour pro rata.[4] The measure exists only for employees; it is not defined for self-employed workers or for those who do not currently have a job at all. It does not differentiate between wages derived from a full-time job or a part-time job. In order to compare income levels across years taking account of inflation, all hourly wages were converted to January 2008 prices using information about the month and year of the interview and the monthly all-items Retail Prices Index (RPI). The income recipient is clearly the wage earner himself or herself.

Individual income is a broader measure than wages because it includes more income sources and for this reason has non-zero values for many more people, not only for individuals in employment. It refers to total income from all sources received personally by an individual, including income from income from the labour market (main and secondary jobs, and self-employment), from savings and investments, and from cash social security benefits.[5] It is expressed in pounds per week (pro rata) in January 2008 prices using information about the month and year of the interview and the monthly all-items RPI.

[3] I do not use data from the extension samples for Scotland, Wales, and Northern Ireland which began in the late 1990s because of difficult issues concerning how to combine the data with those for the original main sample.
[4] The variable is derived using BHPS variables wPAYGU, wJBHRS, and wJBOT in file wINDRESP, where 'w' is the letter identifying the panel wave.
[5] The measure of individual income is BHPS variable wFIMN in file wINDRESP, where 'w' is the letter identifying the panel wave.

As with wages, an individual income variable provides information about who receives the income, and thus potentially who has control over its disbursement and eventual distribution within multi-person families and households. However, with survey data, whether a person actually receives a particular income source is not always clearcut and involves a degree of judgement for the allocation of social security benefits in particular. (For many benefits, assessment depend on family or household means.) For all non-earned income sources such as benefits, BHPS respondents are asked whether they receive each of a large number of income sources (from a list on a showcard) and if they say yes, they are also asked hether receipt of that source is sole or joint. Within a household, if person A reports receipt that is joint and person B does too, the BHPS editing rules split the total income from that source between A and B. But if person A reports receipt of a source, but not person B (even if person A reports joint receipt), the whole amount from that source is attributed to person A.[6] Hence allocations of benefits recorded in the data may depend on respondent reporting behaviour, and lead to undesirable inconsistencies across respondents. However the main argument conventionally advanced in favour of an individual income measure is that it indicates differences in control over resources within families and households, and therefore is also suggestive about the actual distribution of resources. So, arguably, the use of respondent reports about receipt is informative: if no report is made, there is no feeling of control over allocation expressed; conversely, if a benefit is reported as jointly received even if the person is not the official claimant, the response may reflect the respondent's feeling that they have some personal control over the distribution of that source.

The third measure of income, equivalised net household income, is the broadest of the three measures because it covers the most income sources and in principle has non-zero values for all individuals. Equivalised net household income is total household money income from all sources less income taxes and National Insurance contributions and some other deductions, which is then adjusted ('equivalised') to take account of differences in household size and composition using the 'modified OECD' equivalence scale.[7] It is expressed in pounds per week (pro rata) in January 2008 prices using information about the month and year of interview and a modified monthly all-items Retail Prices Index.[8] This income measure is the BHPS counterpart to the 'net income before housing costs' measure that is used in Britain's official income distribution statistics (cf. Department for Work and Pensions 2009). This measure is currently available only for BHPS waves 1–16 rather than for 17 waves as for the other two measures, and is missing for households in which there is at least one non-respondent adult. See Levy and Jenkins (2008) for full details of the variable's derivation.

By constrast with wages, which can be unambiguously attributed to the person earning them, receipt of this income measure is derived by an equal sharing assumption: every individual within the same household is assumed to receive the equivalized income of the household to which they belong, i.e. the same amount. Equivalized net household income also depends on the presence of others in the household in ways that the other two measures do not: household members other than the respondent in question may contribute income to the household's total money income, and they (and any children present) also affect the adjustment for differences in needs that is summarized by the household's equivalence scale factor. These

[6] I thank my ISER colleague Nick Buck, the BHPS PI, for this information.

[7] The scale factor for each household is equal to $1 + 0.6(N_A - 1) + 0.3N_C$ where N_A is the number of adults in the household and N_C. is the number of dependent children. Dependent children are individuals aged less than 16 or aged 16–18 and in school or non-advanced further education, not married and living with a parent.

[8] The index is the series 'all items RPI excluding Council tax (agg4111)' provided to me by the DWP's HBAI team.

interdependencies make it harder to model the underlying determinants of equivalised net household income than to model the the underlying determinants of wages. But it is widely agreed that equivalised net household income is better measure of personal living standards than wages. Individual income falls somewhat in the middle and is valuable for exploring the implications of alternative within-household income sharing assumptions. The main purpose of employing three alternative income measures in this paper is to examine the extent to which conclusions concerning income-age trajectories and their heterogeneity are similar or differ in a descriptive sense. Causal modelling of underlying determinants is not the goal.

The individuals included in the analysis differ according to the income variable considered. When describing trajectories in hourly wages, I consider only individuals of working age, i.e. aged at least 25 and less than 60 (women) or 65 (men). The analysis of individual income and equivalized net household income is based on individuals aged 25 or more, but with no upper age limit imposed.

Age 25 is used to demarcate the start of the working life to ensure that dispersion and variability of initial incomes are not unduly affected by the relatively high turnover among new labour market entrants. In addition, I sought an age by which educational careers had been completed for the vast majority of individuals. As explained shortly, I classify individuals into groups according to highest educational qualification, seeking a definition such that group membership is fixed throughout the lifecourse. Age 25 fits this requirement, as I do not separately distinguish individuals with degrees (see below).

When analyzing income distribution data, some researchers exclude observations with outlying incomes at the top and the bottom of the distribution, for example excluding the poorest one percent and richest one percent of observations in each year. The argument typically made for this selection criterion is that outliers are likely to represent measurement error and unduly and inappropriately influence results. I have not dropped any outlier observations; the data have been used ‘as is’. This is because I believe it is more difficult to identify problematic outliers than is sometimes assumed, especially in a longitudinal context.[9]

The multivariate analysis uses the logarithm of income rather than income as the dependent variable (for reasons explained later). Because a small number of observations had zero or negative values recorded for individual income or equivalized net household income, they were dropped (the logarithm of income is undefined in these cases). A difference in log(income) can be interpreted as a proportionate difference in income levels. A income of £5 per hour corresponds to a log(income) of 1.61; £10 per hour to 2.30; and £20 per hour to 3.00. The absolute difference between a wage of £5 per hour and £5.30 per hour is 30 pence, which is 6 percent of £5. In logarithmic terms, the difference is $\log(5.30) - \log(5) \approx 0.06$, i.e. 6 percent.

Twelve social groups defined by birth cohort, educational qualifications, and sex

As explained earlier, comparisons within and between social groups are a focus of the NEP’s work, with characteristics such as age, sex, household or family type, ethnic minority group, social class, religion, and region of residence used to define group membership. To define

[9] For instance, the definition of an outlier most commonly used in the longitudinal context – based on multivariate versions of the Mahalanobis distance measure – cause problems when there is more than two waves of data and when the panel is unbalanced (as here). Application of these definitions leads to selection of a balanced sample on a sample much reduced in size.

social groups for my analysis, I aimed to use similar characteristics, but there was additional constraint. Since the analysis is intrinsically longitudinal, I wanted individuals to retain the same group membership regardless of their age. This ruled out use of characteristics such as family type or residential location which change over time. Some other characterisics were ruled out because either the BHPS does not collect the information (e.g. about religious affiliation) or because sample sizes were prohibitively small (e.g. for almost all ethnic groups apart from white British).

As a result, the characteristics used to define social groups are restricted to: birth cohort, educational qualifications and sex.

For birth cohort, I distinguish two groups: individuals born before 1955, and those born in 1955 or later. Three levels of education qualification are distinguished:

- 'none': having no qualifications at all;
- 'some': having some educational qualifications but below A-level standard; and
- A-level(s) +: having at least 1 A-level or equivalent (e.g. Highers in Scotland), or some higher qualification such as a degree.

A-level exams are usually taken around age 18, and provide qualifications for university entry. Those who gain an undergraduate degree typically do so by the age 25. By choosing to examine trajectories from age 25 onwards only, I ensure that virtually all individuals remain in the same educational qualifications group throughout their life.[10] By also distinguishing between the sexes, twelve social groups in total were defined.

The number of groups and their definitions represent some compromise between seeking to explore fine detail in between-group difference (leading to more groups) and maintaining reliability (leading to fewer groups each with larger sample numbers). An additional factor arose because there has been a marked increase in average educational qualification levels in Britain over the period covered by the BHPS: the proportion with no qualifications has fallen significantly while the proportion with A-levels or more has risen. This prevented me using a larger number of birth cohorts because, when the number of cohorts was increased, it was difficult to maintain sufficient numbers of individuals in groups defined by sex, birth cohort and educational level. A similar problem arose if more qualification levels were distinguished, in which case the numbers of individuals from earlier cohorts with high qualification levels became too small.

The birth year used to define the two birth cohorts was chosen to maintain the spread of sample number across groups. I experimented with different definitions (e.g. cut-offs of 1950 or 1960), but the general patterns of between-cohort results that are reported later did not change. The cohort of individuals born in 1955 or after includes birth years from 1955 through to 1982, and the respondents range in age between 25 and 52 years over the period of the panel (1991–2007). For the analysis of wages, the cohort of individuals born before 1955 includes birth years from 1927 (men) or 1932 (women) through to 1982, and the respondents range in age between 37 and 64 (men) or 59 (women) over the period of the panel. For the other two income measures, the cohort includes earlier birth years as well and hence some men aged 65 or more and women aged 60 or more.

[10] There was a very small number of persons who upgraded their educational qualifications after age 25 in a way that changed their group membership. In these rare cases, individuals' trajectories were split into sections corresponding to the years spent in each educational qualification group, and then analysed separately. I.e. a long sequence of data for a changer was treated as if it were two shorter sequences for two individuals from different educational qualification groups.

The number of BHPS respondents with valid log(income) values in each of the twelve groups is shown in Table 1. Observe that group numbers differ slightly depending on the income variable considered, because of the different sample selections for each variable that were described earlier. Numbers are smallest for the analysis of wages. The numbers of individuals equivalized net household income analysis is smaller than the number used for the individual income analysis because, in the former case, one fewer wave of data is available and imputations for incomplete within-household response are not as comprehensively employed (see above). The table shows numbers of distinct individuals contributing data. Some of these persons contributed the data for all of the the 17 year period covered by the survey; some contributed many few waves' data. On average, each person contributed data for between 5 to 6 years.

Table 1
Numbers in groups defined by sex, birth cohort and educational qualifications

	Men	Women
Log(hourly wage), individuals aged 25+, and less than 60 (women) or 65 (men)		
Pre-1955 birth, no educational qualifications	456	561
Born 1955+, no educational qualifications	266	225
Pre-1955 birth, some educational qualifications	393	526
Born 1955+, some educational qualifications	883	994
Pre-1955 birth, A-level(s) or higher	808	676
Born 1955+, A-level(s) or higher	2022	2012
Log(individual income), individuals aged 25+		
Pre-1955 birth, no educational qualifications	1221	1892
Born 1955+, no educational qualifications	365	363
Pre-1955 birth, some educational qualifications	739	948
Born 1955+, some educational qualifications	1063	1211
Pre-1955 birth, A-level(s) or higher	1275	1083
Born 1955+, A-level(s) or higher	2298	2250
Log(equivalised net household income), individuals aged 25+		
Pre-1955 birth, no educational qualifications	1138	1779
Born 1955+, no educational qualifications	340	329
Pre-1955 birth, some educational qualifications	708	880
Born 1955+, some educational qualifications	985	1120
Pre-1955 birth, A-level(s) or higher	1211	990
Born 1955+, A-level(s) or higher	2075	2004

3. Individuals' income-age trajectories look like cooked spaghetti

In this section, I present charts summarizing the raw data on individuals' income-age trajectories for each of the three income variables, and argue that the pictures look rather like plates of cooked spaghetti. With BHPS data for thousands of individuals, it is infeasible to show the raw data for everyone and so, instead, I focus on the experiences of selected individuals, and highlight some of the differences by sex, educational qualifications, and birth year.

Figure 2 has two panels, each summarizing how hourly wages vary with age. Panel (a) refers to hourly wages per se, whereas panel (b) refers to the logarithm of hourly wages. In each panel, there are six graphs arranged in two columns and three rows. The three graphs on the left hand side of each panel refer to men, and the three on the right hand side refer to women. The first two rows refer to men and women born in 1966, with a contrast between the first row (those with educational qualifications to A-level or more) and the second row (those with no educational qualifications at all). The third row refers to men and women born in 1946 with educational qualifications to A-level or more. Within each graph, there is a separate line connecting the raw income values, or log(income) values, for each individual. The length of a line shows the number of years for which there was valid wage data for the respondent.[11] Observe that the data do not cover the complete working life for any individual, only a maximum of 17 years.

Among the individuals in the sample born in 1966 (members of the younger birth cohort group defined earlier), there are relatively few people with no qualifications compared to the number with A-levels or more, as Table 1 would lead us to expect. Nonetheless it appears that, for both cohorts, the no-qualifications group has a lower average wages than the more highly qualified group, for both men and women. At each age, the average wage appears greater for men than for women but for both sexes there is also substantial dispersion around the average. Among those born in 1966 with A-levels or higher qualifications, there is wage data covering the beginning of the working life. Over this period, income appears to rise slightly with age on average, for both men and women. For the corresponding group born in 1946, the income data covers the end of the working life. For them, it is hard to discern a rise or fall in average income with age.

Regardless of whether wages are measured on a natural or logarithmic scale, it is apparent from every graph that there is dispersion in wages at the start of the working life with a high prevalence of small year-on-year fluctuations experienced thereafter for most individuals, combined with occasional very large temporal variation for a small minority. In general, trajectories cross and interwine. This is what I call spaghetti.

The patterns seen in Figure 2 for wages can also be seen in the charts for individual income and equivalized net household income. Look at Figures 3 and 4, which summarize raw trajectories for these variables in the same format as Figure 2. The principal difference between corresponding graphs is there appears to be greater dispersion at each age for these other income measures than there is for wages. This is partly but not wholly an artefact of the change in units of measurement from pounds per hour to pounds to week. Another difference is that a small number of outlier trajectories seem more apparent for individual income, especially when looked at using a logarithmic scale. Both patterns are readily explicable in terms of the different definitions of the income variables. By construction, the distribution of individual incomes includes more people who may have a small income, e.g. from non-wage sources. The same is true for equivalized net household income, though the impact is more muted because the equal sharing within households assumption applies in this case.

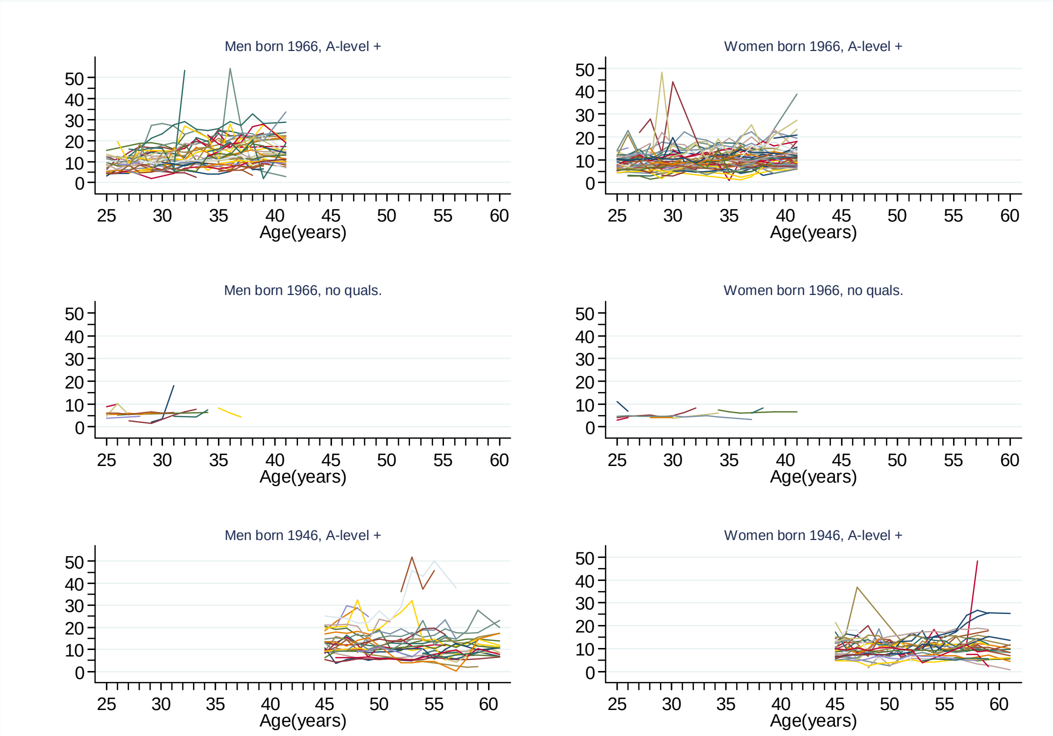
The ubiquity of trajectory spaghetti, and the similarities in patterns for the three income variables and across groups, are the main features to take from Figures 1–3. The first feature is important because it emphasizes the potential role that a statistical model can play in summarizing these apparently complex patterns. The second feature is important because it suggests that the same statistical approach can be applied to each income variable and to each group. Echoing the quotation from George Box in the Introduction, these models are likely to

[11] No line segment is drawn for years in which an individual is self-employed or has no job at all.

be wrong, but there are substantial advantages to having a unified common framework for comparisons across groups and variables. The statistical approach to unravelling spaghetti is set out in the next section more formally and precisely than was done in the Introduction.

Figure 2

(a) Income-age trajectories for hourly wages (£ per hour)



(b) Income-age trajectories for log(hourly wages)

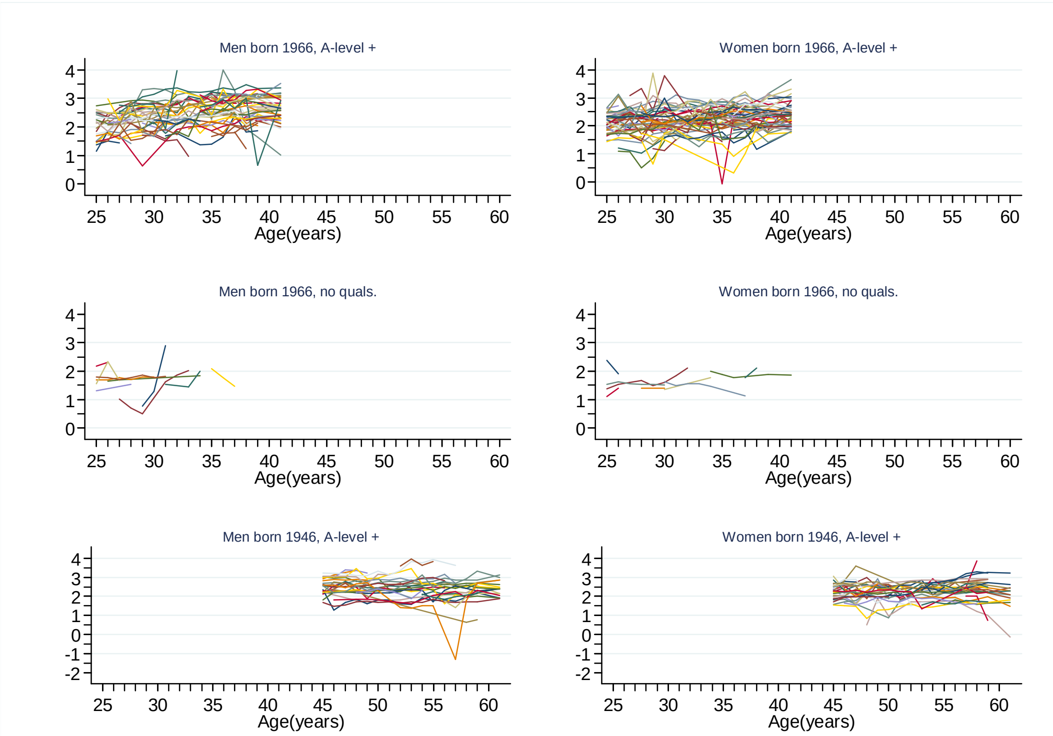
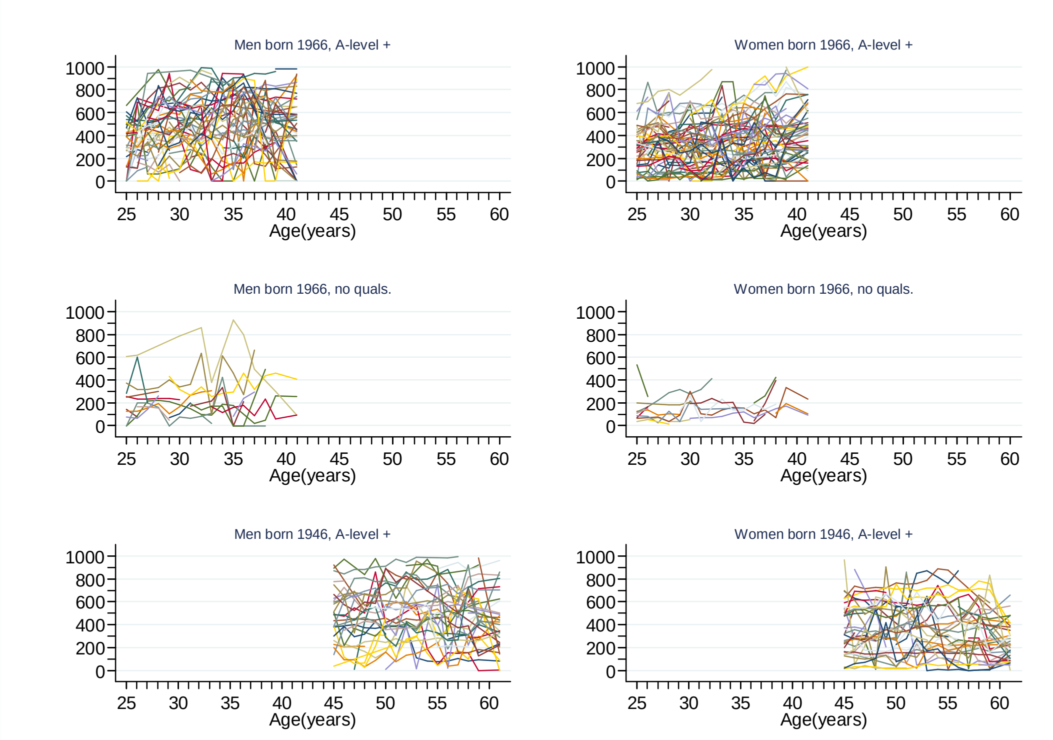


Figure 3

(a) Income-age trajectories for individual income (£ per week)



(b) Income-age trajectories for log(individual income)

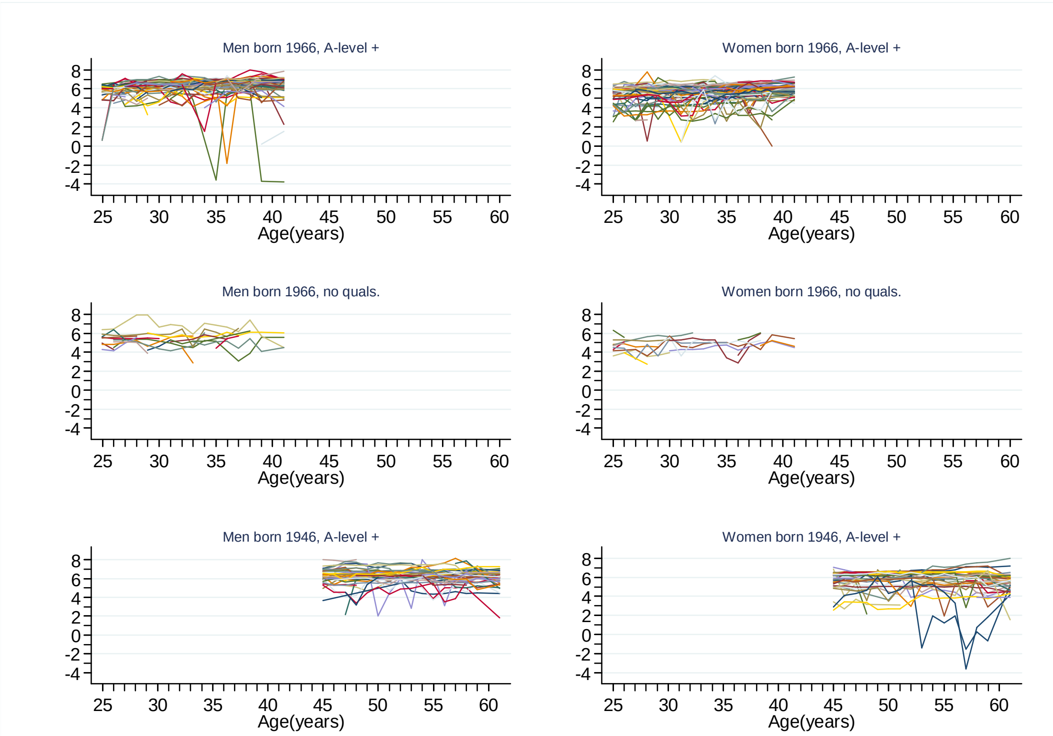
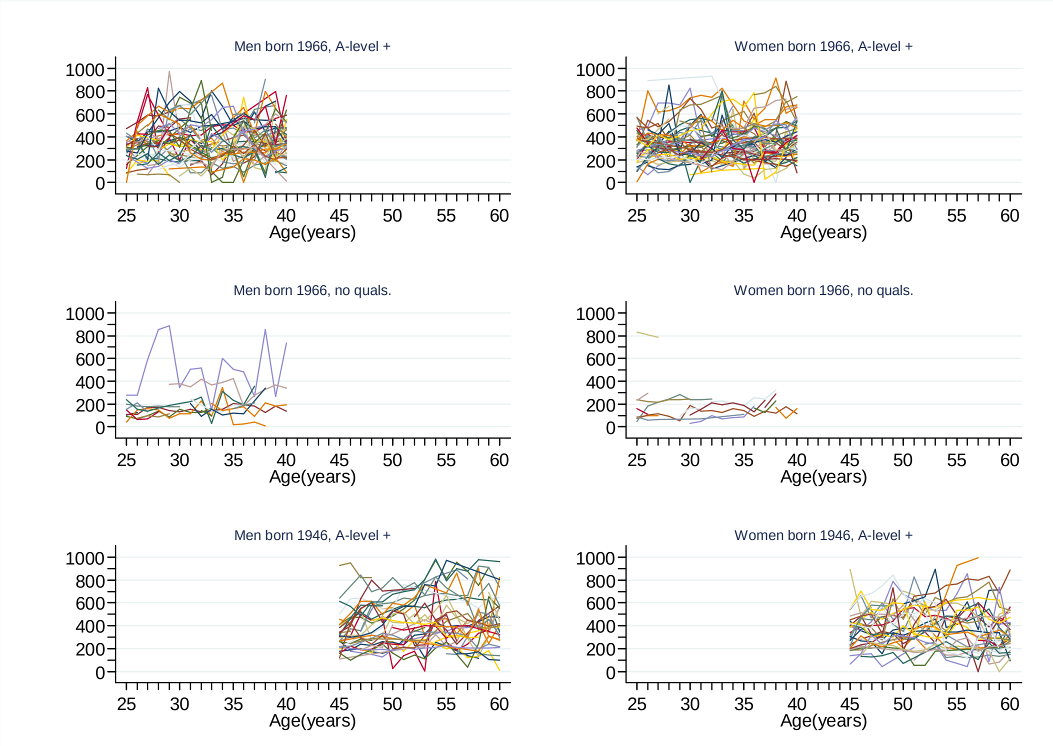
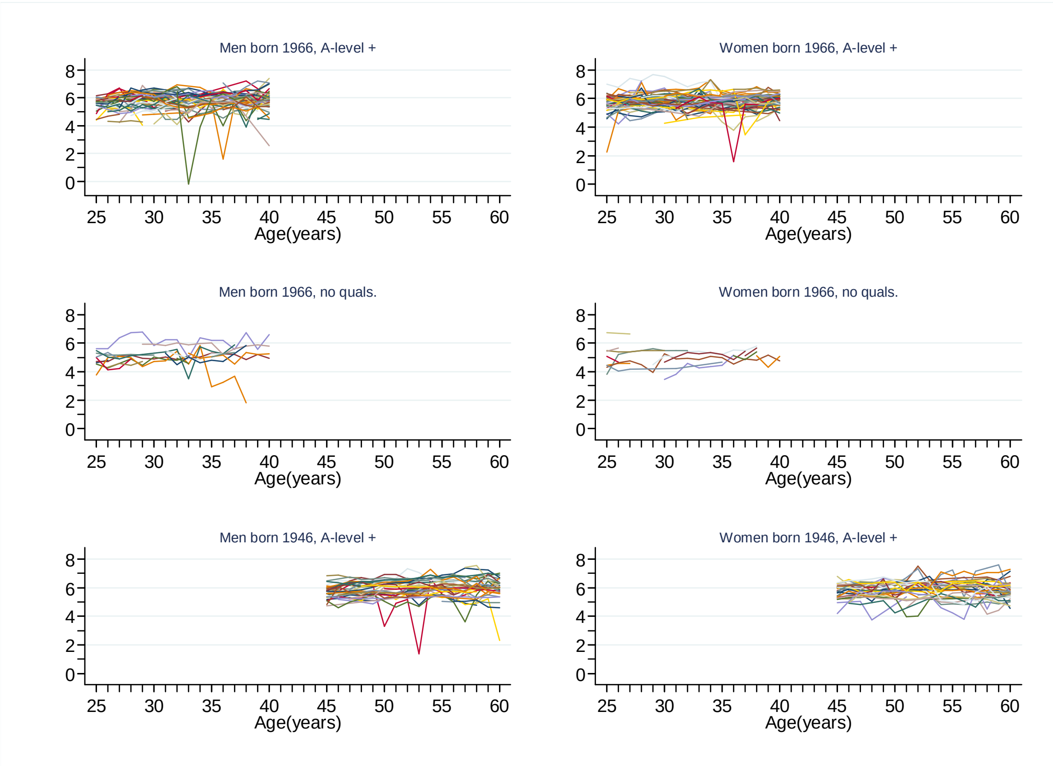


Figure 4

(a) Income-age trajectories for equivalized net household income (£ per week)



(b) Income-age trajectories for log(equivalized net household income)



4. Four areas of research about income-age trajectories

The statistical model of income-age trajectories presented in the next section does not come out of a vacuum, but draws on previous work. So, before presenting the model, I review related research so that it can be placed into context, and to help assess its strengths and weaknesses.

There are four literatures on which I draw. There is research on: (i) short-term income mobility, (ii) descriptions of average income-age trajectories, (iii) dynamic microsimulation modeling, and (iv) 'variance component' modelling. Although there have been empirical applications for many countries, my citations to them are very selective, referring mostly to empirical studies based on British data. Key differences between approaches are the extent to which they consider the evolution of income over the short-run or longer-run (such as the whole lifecourse), and the extent to which they focus on the average experience or variations across individuals.

Short-run mobility

The literature on short-run income mobility has two strands. The first strand is the large literature summarizing the association between individuals' income in one year and their income in another year, or individual income changes over a short sequence of years. Applications have been to both wages and to broader measures of income such as equivalized net household income. Two examples using British data are Jarvis and Jenkins (1998) and Dickens and McKnight (2008). Jarvis and Jenkins, using data about equivalized net household income from BHPS waves 1–4, asked how much income mobility there was in Britain, providing answers refering to a number of summary mobility measures, combined with some regression-based investigations of differences between groups defined by age and sex (finding, surprisingly, more mobility among the elderly than the young). Dickens and McKnight used administrative data on annual employment earnings from the Longitudinal Labout Market DataBase covering 1978/9–2005/6. Using several mobility measures, they document the trends in mobility over the period. They report a decline in short-run mobility, for men and for women, through 1980s and 1990s, with a possible rise in the 2000s.

The second strand in the short-run mobility literature describes the changes in needs-adjusted household income that occur round the time of a particular lifecourse event conditional on experiencing the event. Income changes round partnership dissolution were studied by Jarvis and Jenkins (1999) and Jenkins (2009). For retirement, see Bardasi, Jenkins and Rigg (2002); for widowhood, see Zaidi (2001); and for disability onset, see Jenkins and Rigg (2004). Closely related is the research relating poverty entry and exit transition probabilities to experience of the lifecourse 'trigger' events: see e.g. Jenkins and Rigg (2001) and Jenkins (2008). Rigg and Sefton (2006) extend this literature by relating experience of these events to the incidence of each of six differently-shaped trajectories over 10 years (labelled flat, flat with blips, rising, falling, fluctuating, other).[12]

In the current context, the short-run mobility literature is important because it emphasises the heterogeneity of income change across individuals (first strand), and also points to the importance of lifecourse events as correlates of large income changes (second strand). But the

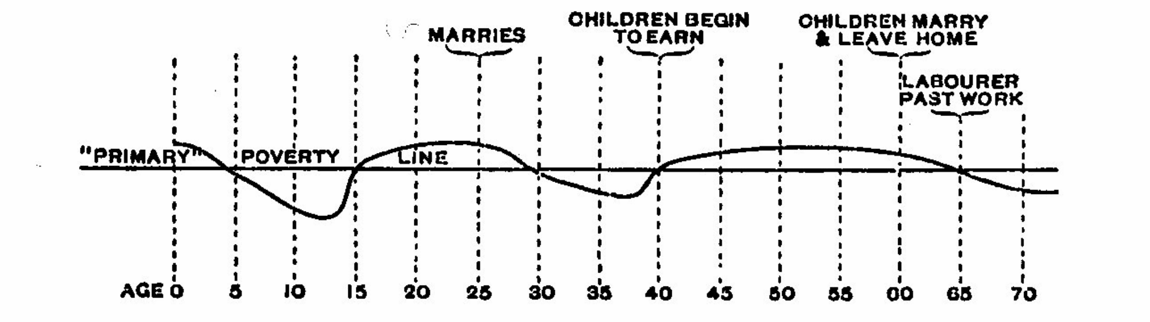
[12] The classification is based on that of Gardiner and Hills (1999) who studied four year income sequences using BHPS data.

focus is predominantly on changes over relative short intervals. Lifecourse income trajectories are not the object of study.

Descriptions of average income-age trajectories

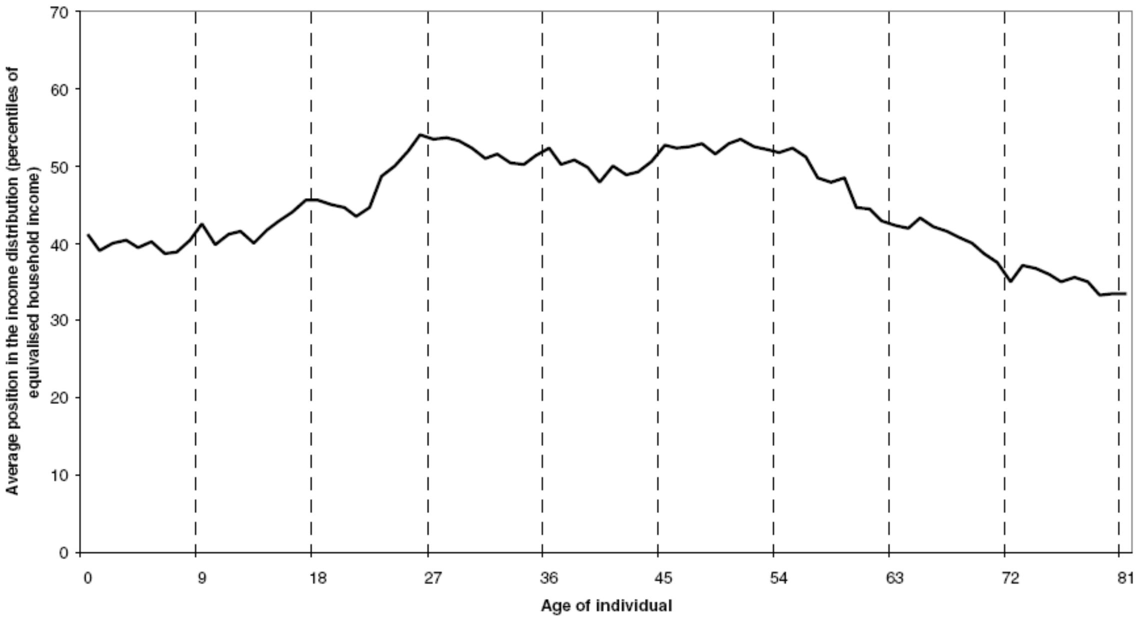
Trajectories are central to the second literature area, however. Research builds on the celebrated portrayal of the lifecourse variation in needs-adjusted income by Seebohm Rowntree, who characterised the life of a labourer as 'characterised by five alternating periods of want and comparative plenty' (2000 [1901], p. 136). Figure 5 reproduces his schematic summary of this: the shape of the income trajectory is closely related to important lifecourse stages such as childhood, marriage, the arrival and departure of children, and retirement. Rigg and Sefton (2006) provide a late-20th century update to this picture, drawing on 10 waves of BHPS data: see Figure 6. The pictures are not fully comparable because one is a stylised summary for a particular type of worker inspired by a pioneering cross-sectional survey (Rowntree) and the other is derived from representative panel survey data (Rigg and Sefton). Nonetheless it is interesting that both charts point to a quasi-M-shape in trajectories of needs-adjusted income in the middle of the lifecourse and a clear decline after retirement.

Figure 5
Five alternating periods of want and plenty for a labourer (Rowntree)



Source: Rowntree (2000 [1901]), p. 137.

Figure 6
Changes in income rank over the lifecourse (Rigg and Sefton)



Source: Rigg and Sefton (2006, Figure 1). Based on BHPS waves 1–10. Vertical axis shows average of each individual's percentile rank by age, with ranks calculated using percentiles defined using the wave 1 distribution of equivalized net household income.

For the purposes of the current study, I note the emphasis in this literature on how trajectories relate to lifecourse stages and events. However, what is generally missing is characterisation of the individual heterogeneity in profiles: the focus is on the average trajectory but not divergences from them. This task was beyond Rowntree's purpose. Sefton and Rigg (2006) recognise the issue of heterogeneity as something that needs to be addressed. After relating their six trajectory types first to lifecourse stages, and then to lifecourse events, they ruefully acknowledge that '[a]lthough many of these events are related to specific income trajectories in the way we might expect, there is a large amount of heterogeneity in people's income trajectories following each of these life-cycle events … Typically, each life-cycle event increases the probability of experiencing a particular trajectory by a factor of approximately two, but most individuals will still follow one of the other trajectory types' (2006, p. 406).

Dynamic microsimulation of lifecourse incomes

A third area of literature is about the dynamic microsimulation of lifecourse incomes, of which the leading British application is Falkingham and Hills (1995). They estimate statistical models of employment, earnings, lifecourse events including household formation and dissolution, and childbearing, and then use the estimates to simulate the economic and demographic lives of a 'synthetic' population from birth to death. They derive a broadly hump-shaped profile for income over the lifecourse on average. The modelling is sophisticated and comprehensive in many respects and lifecourse income trajectories are at the heart of the project. But, for the purposes of the current paper, dynamic microsimulation modelling is less helpful because model-building is very time-intensive and involves many statistical building blocks. The goal at hand also differs: I wish to describe the key features of income-age profiles succinctly, rather than develop a framework for analysis of redistributive features of welfare state policies across the lifecycle.

Variance component modelling

The fourth area of literature is the variance components literature. This research uses models that are the most closely related to one that I use, but the applications to date have had quite a different focus from mine. Virtually all studies under this heading have modelled the dynamics of labour market earnings of men employed full-time and, correspondingly, lifecourse demographic events of the type addressed in the three other literatures are ignored because they are less relevant. Moreover, the implications of the models for income-age trajectories is rarely drawn out. Of principal interest instead is the decomposition of total earnings inequality – typically measured using the variance of log(earnings) – into transitory and permanent components, and how the relative importance of the two components changes with (calendar) time. Hence the 'variance components' label.

Papers developing these variance components models of earnings include Abowd and Card (1989), Baker (1997), Baker and Solon (2003), Haider (2001), Hause (1980), Lillard and Weiss (1979), MaCurdy (1982), and Moffitt and Gottschalk (1995, 2002, 2008). There have been few applications to broader measures of income: notable exceptions are Biewen (2003), Stevens (1999), and earlier work by Duncan (1983).

A simple prototypical model (Lillard and Willis 1978) has the form

$$u_{it} = u_i + v_{it} \tag{1}$$

where u_{it} is interpreted as either the log(wage) for person i in year t or, more commonly, as the residual derived from a regression – or year-specific regressions – of log(wage) on a set of person-specific characteristics such as education level, birth cohort, age or work experience, etc. I use the second interpretation in this review. The idea is that the prior regression summarizes differences in wages on average, and the modelling focus is then on the evolution of deviations of earnings from this average.

The u_i is a fixed random individual-specific component with mean zero and constant variance σ_u^2 (common to all individuals), and v_{it} is a year-specific idiosyncratic random component with mean zero and constant variance σ_v^2 (common to all individuals) which is uncorrelated with u_i. Thus, total inequality equals permanent inequality plus transitory inequality:

$$\sigma^2 = \sigma_u^2 + \sigma_v^2. \tag{2}$$

It is also conventional nowadays to consider models of the form

$$u_{it} = \kappa_t u_i + \lambda_t v_{it}. \tag{3}$$

This is the same model as in (1), except that the permanent and transitory components are weighted by calendar time-specific weights (or 'factor loadings') κ_t and λ_t. The evolution of inequality is then summarized by:

$$\sigma_t^2 = (\kappa_t)^2\sigma_u^2 + (\lambda_t)^2\sigma_v^2. \tag{4}$$

This specification allows the relative importance of transitory and permanent inequalities to vary directly with calendar time and hence to relate trends in this to the business cycle or changes in labour market institutionsl, for example.

The model is completed with assumptions about whether shocks to income have effects that persist over time, and hence how the variance components evolve. The persistent effects of transitory shocks are usually modelled by having the effects decay over time. This is modelled using a so-called autoregressive moving average process for v_{it}, labelled ARMA(p,q), in which parameters p and q characterise the nature of the persistence over time. For example, an ARMA(1,1) process has the form

$$v_{it} = \rho\, v_{it-1} + \theta\, \varepsilon_{it-1} + \varepsilon_{it}. \tag{5}$$

If $\theta = 0$, then the variance of the transitory component this year is equal to a fraction – the square of the autoregression parameter (ρ^2) – of its variance in the previous year, a fraction ρ^4 of the variance two years ago, and so on. So shocks die out quickly if ρ is small ($0.3^4 = 0.0081$ but $0.9^4 = 0.6561$). If $\rho = 0$, then the variance of the transitory component this year is equal to a weighted average of the variance of shocks this year and last year, with the latter receiving less weight (the weight is square of the moving average parameter θ). Whereas we expect ρ to be positive (but no more than one), θ may be positive or negative in principle. If someone is struck by bad luck two years in a row (ε_{it-1} and ε_{it} both negative), a negative value for θ implies that the effect of the past bad luck is dampened. The larger that p or q is in the ARMA(p,q) process, the longer the shadow that past shocks cast over present outcomes.

The models have been developed in two main directions, originally distinct but now often combined. One approach is to relax the assumption that the permanent component (u_i) is fixed and to allow variation over time via a 'random walk': this year's value is equal to last year's value plus or minus some random element. Arguably, some shocks (arising from e.g. major job or health changes) can lead to changes in earnings that are permanent. Instead of u_i, the 'permanent' component in (3) becomes

$$\mu_{it} = \mu_{it-1} + \pi_{it}. \tag{5}$$

The second approach allows for individual-specific rates of growth in wages, and brings us closer to the model sketched informally in the Introduction. The permanent component in (3) is supplemented so that it varies directly with time. Instead of u_i, we have

$$\mu_{it} = u_i + \beta_i t. \tag{6}$$

This is a 'random growth' model: β_i is the growth rate, equal to zero on average but varying across individuals.[13] Both the random walk and random growth approaches lead to a fanning out of the earnings distribution over time, other things equal. Rankings in the earnings distribution are preserved: those at the bottom stay at the bottom but fall further behind those at the top, who stay at the top. It is increases in the transitory variance that increase earnings mobility in the sense of reranking.

The two most well-known applications of these models to British data on earnings are by Dickens (2000) and Ramos (2003). Dickens modelled earnings dynamics for men aged 22–59 using longitudinal data from the New Earnings Survey Panel covering 1975–1995. His model was essentially that described by (3), but with the permanent component modelled as a random walk (with the variance of π_{it} allowed to be age-dependent), and with the transitory component modelled as an ARMA(1,1) process. He found a permanent component of earnings differences that increased with age over the life cycle and significant persistence in the transitory component. Earnings inequality in total grew over the period, and the variances of both permanent and transitory components were found to have risen over the two decades, with each explaining about half the rise in inequality.

Ramos (2003) also earnings dynamics for men aged 22–59, but using BHPS waves 1–9. His statistical model was more complicated than that of Dickens (2000) in that he also allowed for random growth effects. When he fitted the model to raw log earnings rather than log earnings residuals, he found evidence of a negative correlation between initial earnings and growth rates, and hence crossing over of earnings trajectories – as portayed in a stylized manner in Figure 1. He also found that the contribution of the transitory component to total inequality increased over the period. However, when he applied the same model to earnings

[13] Sometimes age or work experience is used instead of calendar time t.

residuals computed from a prior regression of log earnings on education, work experience, region, etc, his preferred model had a much simpler structure with, for example, no signficant individual-specific heterogeneity in growth rates, and no random walk in the permanent component. Put another way, much of the persistence in the data was summarized by the observed systematic earnings differences associated with the characteristics included in the prior regression to calculated the residuals. The simplicity of the earnings covariance structure in this case is somewhat out of line with findings by other researchers and might be due in part to the relatively short panel that Ramos had access to.

Devicienti's (2001) analysis is the only variance components study using British data about equivalized net household income. He fitted several versions of the model to regression-computed income residuals derived from BHPS waves 1–8 including one specification with time-varying weights (as in equation 3) combined with an ARMA(1,0) process for the transitory component, and another without weights (as in equation 1) but with an ARMA(1,1) for the transitory component.[14] He reports statistically significant persistence in transitory shocks, with, for example, the estimate of ρ ranging between 0.42 and 0.76 depending on model. He also comments that '[t]here is some evidence that the permanent component ... has attracted higher returns … over time, particularly so until 1996, while the weight of the transitory component … does not seem to have changed much' (2001, p. 26).

From the point of view of the current study, variance components models are a key reference point as they are a commonly-used method for modelling income dynamics. Their great strength is the modelling of inequality and how it evolves over calendar time, and to allow differences in the process for different birth cohorts.

The problem for me is the models' relative neglect of how incomes evolve with age. I wish to place age centre stage (when characterising both average trajectories and divergences from them), while also making comparisons across groups and for several income measures. As a result, the modelling approach set out in the next section is a compromise. On the one hand, it is inspired by the variance components literature, but it uses a simplified specification relative to most contemporary applications in the literature. On the other hand, it also provides a framework for summarizing average income-age trajectories, though unlike the trajectory studies for Britain cited earlier, it uses a succcinct model-based approach that can illuminate between-group differences as well as within-group ones. The most similar model to mine that I have seen in the literature is that of Gangl (2005) who has similar goals to mine except that he wishes to compare countries rather than social groups within a country.

5. A statistical model to describe individuals' income-age trajectories

This section presents a statistical model to describe income-age trajectories for individuals within a social group, on average as well as the individual-specific divergences from the average. I refer to 'income' in the generic sense; in the empirical work, I fit the model using each of the three measures of income described earlier.

[14] See Devicienti (2001) Table 10, columns (3) and (4) respectively. He also estimated a version of this model with an ARMA(1,0) process for the transitory component supplemented with a random growth specification in age as in (6), but reports that the estimate of the variance of β_i did not differ significantly from zero (Table 10, column 5). It may be that it is difficult to estimate models with complex dynamics from only eight waves of data. He reports some difficulties of achieving convergence of the model with random growth, and he does not report estimates for a model with ARMA(1,1) and year-specific weighting factors.

By contrast with the variance components literature, I do not first control for systematic observed differences in income by running regressions of log(income) on personal characteristics. Instead, I assume that the same model specification applies to each of the twelve social groups separately, but with different values of the model parameters applicable to each group. Both the regression and group approaches are ways of 'conditioning' on characteristics. Using a group approach facilitates the between-group comparisons that I wish to do and is more flexible in principle than the regression approach in the sense that all parameters including error variances and covariances are group-specific rather than homogeous. On the other hand, whereas I identify 12 groups, regression-based approaches typically define many more, at least implicitly, because they use a large number of explanatory variables.

Model specification

The model now set out should be understood as a generic model for each of the twelve social groups. The outcome variable is taken to be log(income) rather than income itself because it led to better fitting models. The estimation method assumes that the residual error terms have a normal distribution and, because the distribution of income is skewed in shape, taking logs make ensure the normality assumption more appropriate. The cost is that any observation with a non-positive income is dropped from the analysis, but this was rare in any case (most likely for the individual income measure).

Individuals in each group are differentiated by their age and their income. For person of *i* in calendar year *t*, let age be represented by A_{it} and the logarithm of income by y_{it}. The model for y_{it} is described by:

$$y_{it} = (\alpha_t + \alpha_i) + (\beta_0 + \beta_i)A_{it} + \gamma(A_{it})^2 + \delta(A_{it})^3 + \phi(A_{it})^4 + v_{it}. \tag{7}$$

Equivalently,

$$y_{it} = [\alpha_t + \beta_0 A_{it} + \gamma(A_{it})^2 + \delta(A_{it})^3 + \phi(A_{it})^4] + \{\alpha_i + \beta_i A_{it}\} + (v_{it}). \tag{8}$$

So, the model has three main building blocks.

The terms in […] characterize the average trajectory for the group. This is a fourth-order polynomial function of age, hence allowing a flexible variety of shapes for the profile. It allows for a period-specific intercept, α_t, so the whole income-age trajectory may shift up or down depending on the calendar year. Without such a term, increases in an individual's income arising from secular growth in the economy are attributed to age: the slope of the income-age profile would be over-estimated. I discuss issues related to the specification of α_t in more detail below.

The terms in {…} characterize individual-level deviations from the group's 'average' profile. These deviations arise from differences in initial income (α_i) and differences in how income grows with age (β_i), as in the random growth model described earlier.[15]

It is assumed that α_i and β_i each have a mean of zero – they represent deviations from an average – and follow a bivariate normal distribution. Thus individual specific variation is captured by three parameters – two standard deviations and one correlation. There is the

[15] Because log(income) is a polynomial function of age, the growth in income with age also depends on age itself (see below). It was infeasible to allow individual-specific differences in the coefficients on higher-order terms in age; there are only a maximum of 17 time points observed for each individual.

variation in 'intercepts' captured by standard deviation σ_α; the variation in 'slopes' captured by σ_β; and the correlation between slopes and intercepts, $\sigma_{\alpha\beta}$. These moments are fixed; they do not vary with age or calendar year. With a negative correlation, configurations of trajectories, with crossings may arise (as shown in Figure 1).

The term in (…) introduces another source of individual-specific deviation from the average profile, that arising from idiosyncratic year-by-year variations from the average. This term is also assumed to be normally distributed with mean zero and dispersion summarized by σ_v, which does not vary with age or calendar year. These idiosyncratic deviations are assumed to be uncorrelated with α_i and β_i. In variance components modelling of wages, v_{it} is the 'transitory component', and discussed as arising from transitory variations per se, or from measurement error. An example of transitory variation is an occasional increase in wages negotiated in a collective bargaining agreement, or occasional overtime working leading to a change in the wage rate. For broader measures of income such as individual income and equivalized net household income, the v_{it} component may also reflect shocks to income arising from major lifecourse events including job loss or gain and changes in household composition.[16] Assuming that income changes arising from these sources have a normal distribution with a smooth symmetric distribution of deviations around the average is potentially questionable, and so some tests of normal fit based on quantile plots are reported in the Appendices. As it happens, the assumption appears remarkably good in the sense that the normal assumption appears consistent with at least 95 percent of the observed data – it is only at the extreme tails of the implied v_{it} distribution that the fit is noticeably poor.

More important is the working assumption that these transitory shocks have no effects that persist beyond the year in which they occur. The assumption clearly conflicts with the assumptions of and lessons from the variance components literature. When I referred earlier to the necessity for compromises in modelling specification, this is the principal example of what I was referring to. In principle, persistence could be incorporated in an extension of my modeling framework, but a lack of suitable software, and time, ruled out explorations of this kind for the current paper.

Underlying explanations

A number of theories suggest that wages increase over the working life but at a decreasing rate. The conventional human capital story (Mincer 1974; Becker 1993) is that investments in education and training are largely financed by earnings foregone at the beginning of the working life and rewarded by faster growing earnings subsequently. Even among groups with similar educational qualifications, one would expect configurations of trajectories as shown in Figure 1 because of differences in human capital investments other than in educational qualifications, for example on-the-job training.[17] Differences in initial earnings may also represent genuine differences in 'ability', work readiness and other factors affecting earnings. If these differences are observed by employers, one would expect trajectories to be higher for employees with greater 'ability' throughout the lifecourse. But if 'ability', etc., is not observed initially by employers, trajectory crossings may arise as a result of employer learning: 'ability' is revealed with the passage of time and pay is adjusted upwards or

[16] One might argue that these types of event might lead instead to changes in the permanent component of earnings. However, incorporation of temporal change in this component as usually done, with a random walk specification, is also implausible.
[17] Among the A-level(s) + group, there are also differences in education qualification levels, between those without degrees or with different types of degrees.

downwards accordingly. Personnel economics provides different arguments for upward sloping wage profiles: an employment contract combining relatively low pay earlier in the working life with higher pay (including pensions) later provides incentives to employees to reduce shirking behaviours that might lead to dismissal and hence loss of the higher pay deferred until later in the working life (see e.g. Lazear 1995).

Various other matters complicate these stories and may lead to different average trajectory shapes for men and women, and for different birth cohorts. For example, women are more likely than men to work part-time, and part-time work is less well-rewarded. This is likely to produce slower growing earnings for women relative to men over childbearing ages. Individuals from different birth cohorts may have the same level of educational qualification in name, but the knowledge and skills encapsulated in them may change over time, and correspondingly the labour market rewards associated with them.

Since earnings are the principal income source for the majority of households, one would expect trajectories for wages and broader measures of income to be similar. Many of the differences in shape are likely to relate to periods when children are more likely to be present, not only because of the effects on labour force participation as discussed, but also because many social security benefits are child-related. For equivalized net household income, there is an additional effect: changes in the number of children (or adults) change this income measure, via changes in the equivalence scale factor, even if household money income remains the same.

Model estimation and some additional issues

The model is fitted separately to data for each group using module `xtmixed` in Stata version 10 (StataCorp 2007). Computations of the standard errors of the parameter estimates do not adjust for the fact that there may be repeated observations from same household within the same group, which conflicts with the assumption of independence across observations.[18] I conjecture that this is a relatively minor issue compared with other issues such as the modelling of persistence of transitory errors, and some other complications that are now elaborated.

Differential non-response in the initial interview wave in 1991 (and subsequently), together with differential attrition (sample drop-out) after the wave one interview may lead to biased estimates of the statistics of interest. Although application of sample weights is the conventional way to mitigate against these potential biases, they have not been used in the current analysis. This is because, first, the single set of longitudinal weights supplied with the BHPS refers to a rather special sample. The wave T longitudinal weights are positive only for panel members responding at each and every wave from when they first joined the panel wave 1 through to wave T. This means that any respondent with intermittent response is dropped from any analysis, which is undesirable. Although it would have been possible to derive special weights, there was also the second issue. Stata's `xtmixed` module does not allow sample weights.[19] So, I use unweighted data, which is what most researchers fitting multivariate models of income dynamics do. This should be kept in mind in what follows. For instance, if respondents who drop out of the panel are those with lower incomes, holding

[18] Biewen (2003) addresses the correlation issue in the context of variance component models.

[19] There is a user-written Stata module, gllamm, which allows sample weights (http://www.gllamm.org/) but it was not employed due to the limited time available for this project. The question of what is the appropriate weights remains, of course.

other characteristics constant, then income levels at older ages (corresponding to longer time elapsed since originally sampled) may be over-estimated and the heights of income-age trajectories over-estimated.[20]

In any case, there are arguably other data issues that are more important. For example, observe that 17 years does not span a complete working life let alone a complete lifetime. For respondents who were 40 in 1991 (born in 1951), the panel covers the 17 years from age 40 until age 57; respondents who were 25 in 1991 (born in 1966), the panel covers the 17 years from age 25 until age 42. So, if one wishes to describe income-age trajectories over the full working life, one has to assume some commonality of experience between people from different birth cohorts – which may not be appropriate. Alternatively, one allows for differences in trajectories between groups with similar birth years, and concedes that inference about complete lifecycle trajectories is constrained. Given the NEP's interest in differences across social groups, the latter approach is the one followed here.

A related matter is that the 17 years covered by the BHPS cover a particular period of calendar time. At the beginning of the 1990s, Britain's economy was at bottom of the economic cycle and the unemployment rate peaked in 1992/1993 at around 10 percent. Over the subsequent decade and a half, the state of the economy improved and by the peak of the cycle in 2007, the unemployment rate had halved. Incomes rose with economic growth, and as the labour market improved more people previously without work took a job and those who might have otherwise lost their job or left the labour force (e.g. by retiring) remained in work. This raises two potential issues.

The first arises from the association between the passage of calendar time and age, meaning that it is tricky to prevent estimates of the relationship between income and age from being contaminated by the effects of period. Identification is secured by exploiting the fact that the panel contains individuals born in different years: for each calendar year, the panel contains individuals of different ages. But this in turn constrains the extent to which differences in income-age trajectories across birth cohorts can be identified, since age equals calendar year minus birth year.[21] The approach taken here is to eschew estimation of fine-grained birth cohort effects, distinguishing only two groups, defined by whether a respondent's birth year was before 1955 or 1955 and afterward. With a small number of broadly-defined birth cohort groups, there is independent within-group variation in income by calendar year and age. The cut-off year of 1955 is to some extent arbitrary, and chosen to ensure there are sufficient sample numbers in each group. I experimented with alternative cut-off years, and also with three groups rather than two, but this analysis did not change the broad tenor of the conclusions reported below. My allowance for calendar time (period) effects is relatively crude, and the empirical analyis simply distinguishes between the 1990s and the 2000s.[22] In terms of (7), α_t is specified as a binary indicator equal to one if the survey year is 1990–2000, and zero otherwise. Generalizing across income measures and groups, incomes were about five or six percent lower in the 1990s than the 2000s.

The second issue concerns changes in the composition of the labour force with the economic cycle or with age and other characteristics. Estimation of income-age profiles is based on data about those currently earning but one would expect labour force attachment propensities to be

[20] Uhrig (2008) discusses of attrition in the BHPS and its correlates.
[21] For more discussion of age-period-cohort identification issues, see e.g. Deaton and Paxson (1994).
[22] This choice was based on inspections of estimates from a series of preliminary OLS regressions of income against a fourth-order polynomial in age and a full set of binary indicators for survey year.

positively associated with earnings potential, other things equal. For example, when the economy is in the doldrums, as it was in the early 1990s, there may be an under-representation of those with low earnings potential relative to boom periods. This 'selection' issue may be of particular relevance when estimating the average income of those at either end of the working life – young people entering the labour force, and older people approaching retirement age – relative to those aged in between, and also comparisons between women and men especially over parenting ages since women with low earnings potential may be less likely to work or return to work. In the later discussion of empirical estimates (Section 6), I attribute some apparent anomalies in the shapes of income-age trajectories to selection issues.[23]

Implications of the model

The model's parameter estimates are used to summarize both average trajectories for each group and the variation in income levels at each age within groups.

The average income-age trajectory for a particular group is given by the expected log income of a person at each age a:

$$\mathrm{E}[y_{it} \mid A_{it} = a\ ,\ t = \tau] = \alpha_\tau + \beta_0 a + \gamma a^2 + \delta a^3 + \phi a^4. \tag{9}$$

Thus the average trajectory is described by a fourth-order polynomial in age. The profile is period-specific because the intercept α_τ is year-specific. The estimated profiles described later refer to those for the 2000s rather than the 1990s (estimates of α_τ are reported in Appendix Tables A1, B1, and C1).

A useful property of the statistical model is that it implies that, within each group, log(income) is normally distributed at each age, and hence the shape of the distribution is characterized by the mean and variance of income at each age. The relevant mean is shown in (9). The variance of log(income) at age a is:

$$\sigma^2(a) \equiv \mathrm{V}[y_{it} \mid A_{it} = a, t = \tau] = \sigma_\alpha^2 + a^2\sigma_\beta^2 + 2a\sigma_{\alpha\beta} + \sigma_v^2 \tag{10}$$

where σ_α^2 (σ_β^2) is the variance of α_i (β_i), and $\sigma_{\alpha\beta}$ is the covariance between α_i and β_i. Thus, at each age, there is greater dispersion of income – a larger prevalence of deviations from the average group trajectory – the greater the dispersion of initial incomes or in income growth rates or in the dispersion of transitory income shocks. A negative correlation between intercepts and slopes is an inequality-reducing influence.

More generally, and assuming $\sigma_{\alpha\beta}$ is negative, the inequality-age relationship is U-shaped, i.e. first declining with age and then increasing with age. More specifically, income inequality increases with age if[24]

$$-\sigma_{\alpha\beta}\,\sigma_\alpha / \sigma_\beta < a \tag{11}$$

and inequality decreases with age if

$$-\sigma_{\alpha\beta}\,\sigma_\alpha / \sigma_\beta > a. \tag{12}$$

Thus age-specific incomes are more likely to fan out as age increases, the larger is the dispersion in income growth rates and the less dispersion there is in incomes at the start of the working life.

[23] See Blundell, Reed and Stoker (2003) for discussion of related selection issues in the context of estimating aggregate wage growth.
[24] Baker (1997, p. 345) derives a similar expression.

For a lognormal distribution, there is a one to one relationship between the variance of logarithms inequality index and other Lorenz-consistent inequality measures. Thus, for instance, the widely-used Gini inequality index is given by the expression

$$G(a) = 2\Phi([\sigma^2(a)/2]^{0.5}) - 1 \qquad (13)$$

where $\Phi(.)$ is the normal probability distribution function. A U-shaped graph for $\sigma^2(a)$ against age a implies a U-shaped graph of $G(a)$ against age, with inequality changing from falling to rising at the same age. I report patterns using both measures below.[25] The normality property also enables me to describe the whole range of incomes at each age. I focus on two specific ages, one at the start of the working life (25) and one in the middle (40), and use estimates of the lower and upper quartiles (the 25th and 75th percentiles) for each group to explore the extent to which the distributions of income for different social groups overlap.

One can also use the model to examine year-to-year mobility in individual incomes – the extent to which individuals may move up or down the distribution relative to others of the same age. If (im)mobility is summarized using the correlation in log incomes between the two ages, its extent is intimately connected to the evolution with age of age-specific inequality. It can be shown that the correlation is definitely positive when inequality increases with age. And immobility is not constant over the lifecourse. For example, there is greater mobility when inequality is rising with age. I eschew discussion of income mobility, however, in this report. As discussed earlier, my model assumes that transitory shocks only have an effect on income in the year in which they occur. So, any income mobility predicted by the model is likely to be over-estimated. I focus on other aspects instead.

6. Estimates of income-age trajectories: group averages and individual divergences

This section discusses the shapes of income-age trajectories on average and how they differ across individuals within and between groups. The discussion focuses on the estimates for hourly wages, and summarizes them almost exclusively using graphs. Corresponding results for the other two income measures, and the parameter regression estimates for all measures are reported in Appendix Table A1, B1, and C1.

Average trajectories, by group

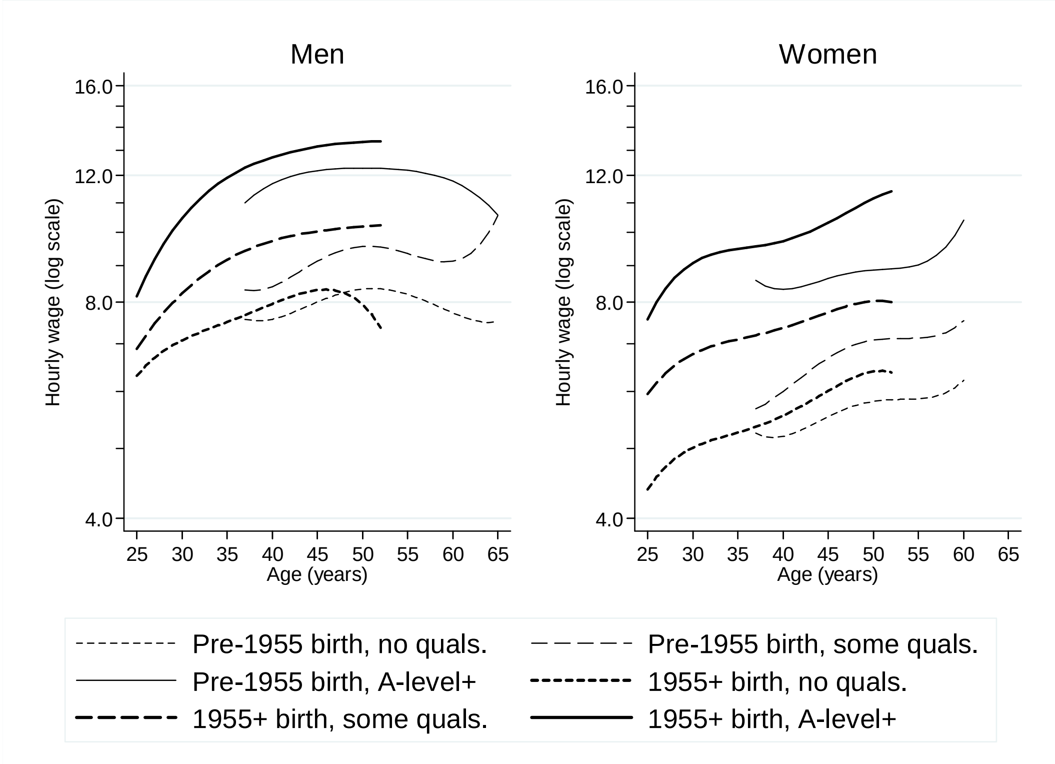
The average trajectories for wages for employees of working age are shown for the twelve groups in Figure 7, and derived using equation (9) assuming the period corresponds to the 2000s.[26] The trajectories are plotted using a logarithmic scales, so that the slope of the trajectory shows how the *proportionate* growth rate of wages changes with age. (If wages increased at the same percentage rate each year, the profile would be a straight line.) The trajectories are shown only for the age ranges covered by the various estimation samples, so the pictures for the 1955+ birth cohort cover the age range 25–52 and those for the pre-1955

[25] I do not examine inequality in the population – for all groups combined – at each age. This depends on three factors: inequality within each group (greater inequality in a group raises overall inequality); the mean income of each group (the greater the spread in the means, the greater is overall inequality); and the relative numbers in each group (the larger a group's population share, the greater the contribution to overall inequality of that group's inequality).

[26] The estimates of the coefficients for the age variables are sometimes not statistically significant (though the variance component estimates are always very significant). These are virtually always cases in which the average trajectory is predicted to be near linear and so reflect multi-collinearity. In the interests of applying a common regression specification to all groups and income measures, I work throughout with estimates from models with a fourth order polynomial in age.

birth cohort cover ages ranging from 37 to 64 (men) or 59 (women). This reminds us that conclusions based on extrapolations outside these age ranges (as in some later graphs) should be done with caution.

Figure 7
Estimated average wage-age trajectories, by group, for employees of working age



Some clear patterns emerge from the estimates, and are in line with expectations. First, hourly wages increase with age from the beginning of the working life, but at a decreasing rate (with some anomalies that I return to shortly). On average, and regardless of group, men's wage grow continuously from the start of the working life but at a decreasing rate, then peak in the late 40s and fall thereafter. In contrast, women's profiles do not have such a distinct peak – wage growth declines up until the late 30s but then appears to rise again. The growth slowdown for women is consistent with their greater prevalence of part-time work, which is less well paid, particularly over the ages when many have children.

Second, for both men and women, and for both birth cohorts, having higher educational qualifications is associated with higher wages, with the return to additional qualifications greater for women than for men up until middle age (women's trajectories appear more parallel than divergent). But, third, among persons with similar educational qualifications and birth cohort, men are paid more on average than women at every age. Fourth, individuals from the later-born birth cohort are on higher trajectories than those from the earlier-born cohort, other things equal.

The returns to different levels of education and between the sexes are substantial. For example, for men aged 40 from the 1955+ birth cohort, the difference on average between those with no qualifications and qualifications of at least A-level standard is a difference of around 50 percent. (just over £12 per hour compared with just under £8). For women, the corresponding difference is around 55 percent. But the difference between the hourly wage of a 40 year old man and a 40 year old woman, both from the younger cohort, is more than one third in his favour on average (around 35 percent). The average trajectory for women with at A-level qualifications lies below that for men with some qualifications. The average trajectories for men with no qualifications lie almost everywhere above the average trajectories for women with some qualifications.

There are some potentially anomalous aspects to some profiles at the beginnings and and ends of the working life, notably for the pre-1955 birth cohort: observe the upward twists in these cases. My explanation for these is that they reflect the impact of the selection effects cited earlier. For instance, arguably the women most likely to remain in the work force as the state retirement age (60) approaches, are those for whom the pay rates are relatively high; those with relatively low pay rates retire. So, the pay rates used to estimate average trajectories over that age range are an over-estimate relative to the average that would be calculated were all women to have remained in work. Similar arguments can be made concerning older men, but it is a puzzle why the increase in the average is so pronounced for men with some qualifications but not for those with no qualifications. There is also a slight decline in average wages for men and women just prior to age among the pre-1955 birth cohort. Arguably this reflects a period effect. For this group, these years correspond to the recession years of 1991–1993 and, again, men with relatively low earnings propensities were less likely to work, thereby raising the average calculated from those who were in employment.

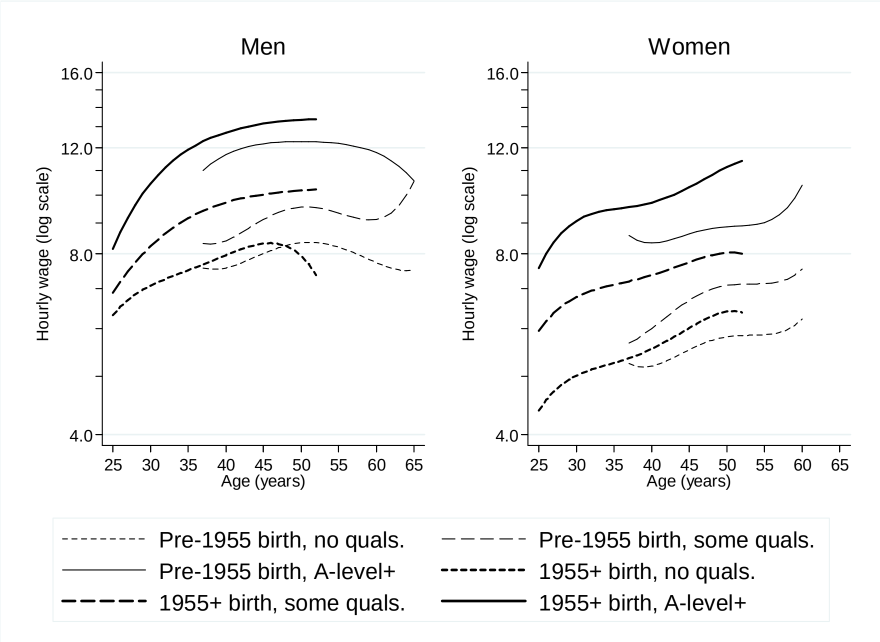
Differences between longitudinal and cross-sectional estimates

Average income-age trajectories derived from longitudinal data look different from those derived from cross-sectional data. Figure 8 illustrates this point for hourly wages. The cross-sectional data used as the reference point are drawn from the UK Labour Force Survey, pooled data for 2006–2008 and, in the right-hand chart, I have plotted median wages (in 2008 prices) by age group, where the age range is the same as was used in Figure 7. (The left hand chart reproduces Figure 7.)

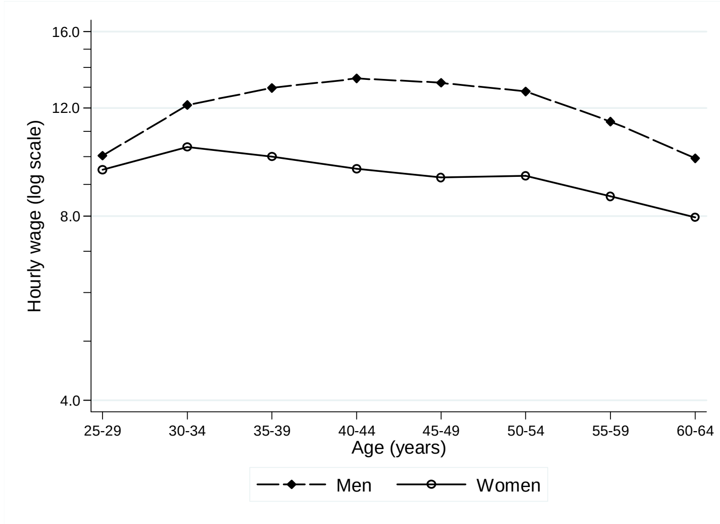
It is clear that both types of data source point to some common features of trajectories, specifically that, on average, they have a ‘hump’ shape with age at least for men, and men are paid more than women at each age. However, there are some important differences. First, trajectories at the beginning of the working life are steeper – wage growth is greater – according to longitudinal data. And, second, wages continue to grow after age 40 whereas, according to the cross-sectional data, hourly wages peak during the forties. Third, and related to the first two points, wage-age trajectories for women differ quite markedly between the two types of data source, and by more than for men. According to the LFS estimates, women’s wage rates are fairly flat or decline from age 30 onwards (on average). By contrast, the BHPS longitudinal data estimates suggest that women earnings continue to increase throughout the working life, albeit with a dip in earnings growth rates associated with child-bearing ages.

Part of income growth associated in the longitudinal estimates may, in fact, reflect the impact of secular growth in income – this is the issue of identification of age effects separately from cohort and vintage effects cited earlier. Illustrating this point, note that when I made *no* allowance for period effects when estimating average trajectories (α_t was constrained to be the same regardless of survey year), the profiles were even steeper (less concave) than shown.

Figure 8
Longitudinal and cross-sectional pictures of 'average' wage-age profiles differ



(a) Longitudinal perspective (as in Figure 7)



(b) Cross-sectional perspective

Source: Author's calculations from National Equality Panel report. Data derived from UK Labour Force Survey, pooled data 2006–08). Estimates refer to age group medians.

Deviations from the average trajectory, by group, and overlapping group distributions

The estimates reveal that there are substantial difference across groups in average income-age trajectories. But how much dispersion is there within groups around the average, and to what extent to the distributions across groups overlap? Table 2, based on the estimates reported in the Appendix, provides a first look at the prevalence of within-group heterogeneity. It shows the estimates of σ_α, σ_β, $\sigma_{\alpha\beta}$ and σ_v for women born in or after 1955 with at least A-level qualifications, and the corresponding estimates for the other eleven groups expressed relative to those of this group.

The statistics shown in Table 2 illustrate the obvious but fundamental point that, for every group, there is substantial heterogeneity around the group average trajectory. This takes the form of significant differences in initial wages (σ_α and σ_v are positive for all groups, but the former is generally larger than the latter), combined with significant differences in slopes ($\sigma_\beta > 0$). Moreover there is a strong tendency for within-group trajectories to cross: the negative estimate for correlation $\sigma_{\alpha\beta}$ means that employees with lower (higher) initial wages tend to have faster (slower) growth rates.

There are also some marked differences across groups, though there are few clear cut patterns. The sharpest difference is between the earlier and later born birth cohorts. For both men and women, and for each educational group, estimates of σ_α, σ_β, and $\sigma_{\alpha\beta}$ are notably smaller for the 1955+ cohort relative to the pre-1955 cohort, implying less within-group deviation from the average profile. But this impetus is offset for some groups by a rise in the transitory variance between earlier and later cohorts, and observe that the estimates of $\sigma_{\alpha\beta}$, vary little (all the ratios are close to 1.00).

Table 2
Between-group differences in variance component parameters

Educational Qualifications	Men		Women		
	Pre-1955 birth	Born 1955+	Pre-1955 birth	Born 1955+	
sd(intercept): σ_α					
None	1.24	0.81	1.47	0.92	
Some	1.75	0.84	1.36	0.92	
A-level(s) +	1.63	0.90	1.66	1.00	[0.993]
sd(age coefficient): σ_β					
None	0.73	0.71	1.02	0.86	
Some	1.07	0.80	0.91	0.80	
A-level(s) +	1.06	0.89	1.07	1.00	[0.030]
corr(int., age coeff.): $\sigma_{\alpha\beta}$					
None	1.05	1.02	1.06	1.04	
Some	1.07	1.02	1.05	1.02	
A-level(s) +	1.05	0.99	1.05	1.00	[–0.919]
sd(error): σ_v					
None	0.78	0.88	0.90	1.02	
Some	0.90	0.79	0.85	0.95	
A-level(s) +	0.86	0.86	1.08	1.00	[0.278]

Note: group parameters expressed as a ratio of the parameters for women born 1955+ with A-level(s)+ (shown in brackets).

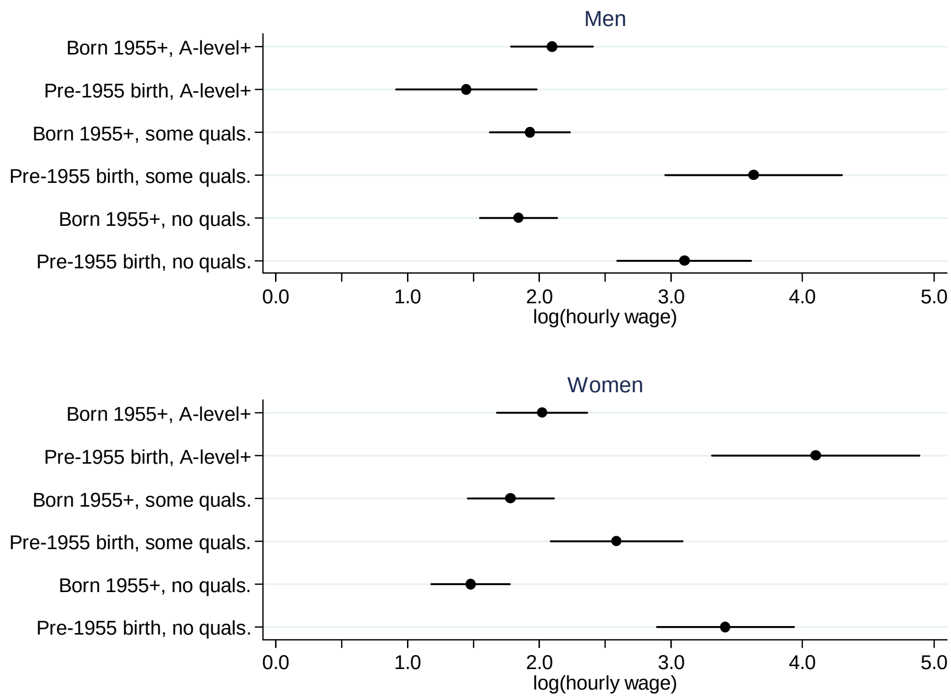
The nature of the within- and between-group differences in income levels at different ages is illustrated by Figures 9 and 10. Figure 9 refers to the start of the working life (age 25); Figure 10 refers to nearer the middle of the working life (age 40). For each group, I show using filled circles, the income of the person in the middle of the distribution (the median which is also the mean given the normality assumptions). Put another way, the filled circles show the income differences at age 25 that are shown in Figure 7. The lines extending from each filled circle show the within-group dispersion in terms of a range of real income levels, specifically the distance between someone one quarter of the way up from the bottom of the group distribution and someone three-quarters of the way up from the bottom (i..e. one quarter of the way down from the top) – the difference between the 25th and 75th percentiles, otherwise known as the inter-quartile range. The Figure shows that these correspond to quite substantial within-group differences in income. For example, for the 1955+ cohort, the inter-quartile range for both men and women regardless of education group is about 0.5, i.e. the 75th percentile is some 50 percent greater than the 25th percentile. The estimates for the pre-1955 birth cohort are less reliable because they are based on out-of-sample predictions, and so I give them less emphasis.[27]

Figure 9 illustrates how the finding reported in Table 2 of smaller heterogeneity parameters for the later birth cohort relative to the earlier one translates into differences in the within-group spread of incomes. The sets of inter-quartile range estimates for the 1955+ cohort are noticeably smaller – the lines are shorter – than those for pre-1955 cohort.

In addition, Figure 9 shows that there is substantial overlapping in income distributions for the different groups. Even though having more educational qualifications, for example, is associated with significantly higher initial wages on average, at age 25 there is a substantial number of employees with no educational qualifications who are paid more than employees with some qualifications or indeed at least A-levels, for both men and women. Among the 1955+ cohort at age 25, a man three-quarters of the way up the distribution of those with some educational qualifications earns more per hour than someone in the middle of the distribution of those with at least A-levels. The same is true for women of this cohort, but note that, in general, the extent of overlapping in distributions across women's educational groups is less for men's. In addition, although Figure 7 highlighted that women have lower average income-age trajectories than men, Figure 9 shows that there is substantial overlapping of men's and women income distributions. Among the 1955+ cohort, the overlapping is smallest among those with no educational qualifications. In this case, the woman with her group's median wage earns less than the man whose wage equals the 25th percentile for his group.

[27] The anomalous upward twists in the average profiles for this group discussed earlier translate into average incomes at age 25 in Figure 9 that are implausibly large.

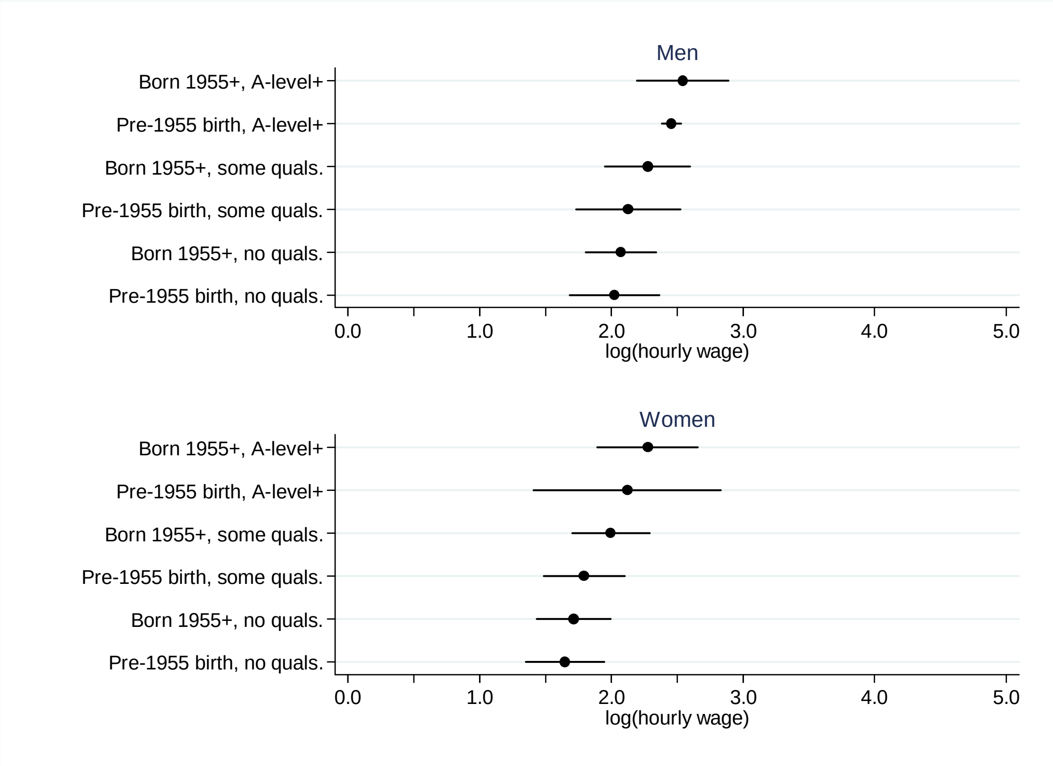
Figure 9
The distribution of log(hourly wage) at age 25, by group



Note. The line for each group shows the group-specific interquartile range (distance between the 25th and 75th percentiles). The filled circles show the group medians (50th percentile), which is the same as the mean. The estimates for the pre-1955 birth cohort are less reliable because they are based on out-of-sample predictions: see text.

Figure 10 is in the same format as Figure 9, except that it refers to the inter-quartile range for those aged 40, by group. (These are within-sample predictions for both birth cohorts.) The graph shows the same rankings of average wages at this age as reported in Figure 7, confirming for example that, on average, men are paid more than women, and having more educational qualifications is associated with higher wages. But, again, as at age 25, there is substantial dispersion of wages within each group, and this implies substantial overlapping in the wage distributions of different groups. Among those born in or after 1955, the man at the 75th percentile of the group with no qualifications earns slightly more than the man at the middle of the group with some qualification. But the man at the middle of the no qualifications group earns more than the woman at the 75th percentile of the group with no qualifications.

Figure 10
The distribution of log(hourly wage) at age 40, by group



Note. The line for each group shows the group-specific interquartile range (distance between the 25^{th} and 75^{th} percentiles). The filled circles show the group medians (50^{th} percentile), which is the same as the mean.

The model estimates can also used to show how within-group inequality varies right across the age range covered by the working life. (Again remember that it is the middle age ranges for which there are within-sample predictions that are more reliable.) Figures 11 and 12 chart the inequality-age relationship, by group, using the variance of logs and Gini coefficient measures of inequality respectively.[28] The figures indicate quite large differences in inequality by age. For reference, observe that the Gini coefficient for wages among all employees increased from around 0.30 to 0.35 between the late 1970s and the mid-1980s – an increase widely regarded as historically large for Britain.[29] The differences between the age-specific Gini coefficients for the beginning and end of the working life are of even larger magnitude according to Figure 8.[30]

Two cross-group differences stand out from the Figures. The first is the constrast between the profiles for the earlier-born and later-born birth cohorts. For those born in or after 1955, wage inequality increases with age throughout the working life after about age 35; before that age, there is little variation with age. By contrast, for those born before 1955, inequality declines with age until the mid-50s (men) or late-40s (women) and only then increases. (The reasons for the different age turning points in the profiles are not obvious.) The cross-cohort differences are related to differences in the ratio of the dispersion in initial wages to the dispersion of income growth rates, i.e. the ratio σ_α / σ_β (see equations 11 and 12). Not only

[28] See equations (10) and (13) for the expressions used to draw the graphs.
[29] See e.g. Figure 2 of Wren-Lewis, Muriel, and Brewer (2009). Their report also provides inequality decompositions by age group based on cross-sectional rather than longitudinal data.
[30] See also footnote 24.

was each parameter smaller for the later-born cohort than its counterpart for the earlier-born one (Table 2), but so also was its ratio.

The second contrast is between those with A-level(s) or higher qualifications and the other two groups. In all but one case (men, pre-1955 cohort), those with A-levels experience distinctly greater inequality at each age throughout the working life than do those with fewer qualifications. There is a straightforward explanation for this. The other two groups are each relatively homogenous in terms of formal educational qualifications, but the A-level(s)+ group includes people for whom A-levels is their highest qualification, as well as those with undergraduate and postgraduate degrees, and one would expect earnings differences associated with these differences in qualifications.

Figure 11
Inequality and age: variance of log(hourly wage), by group

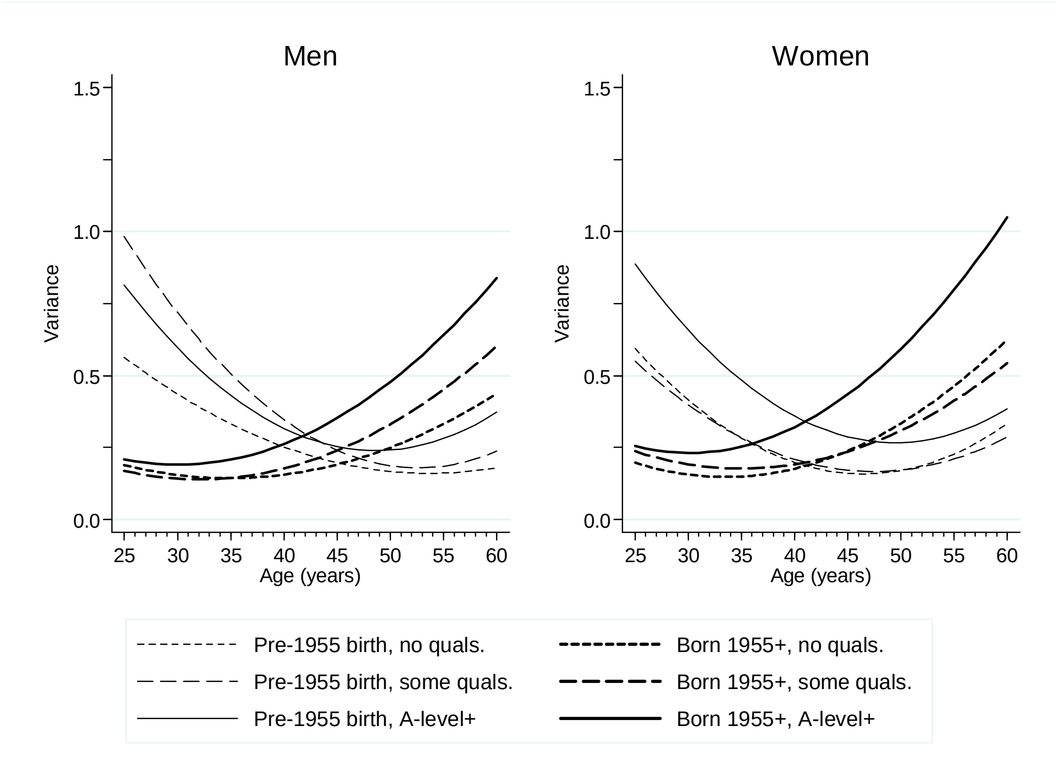
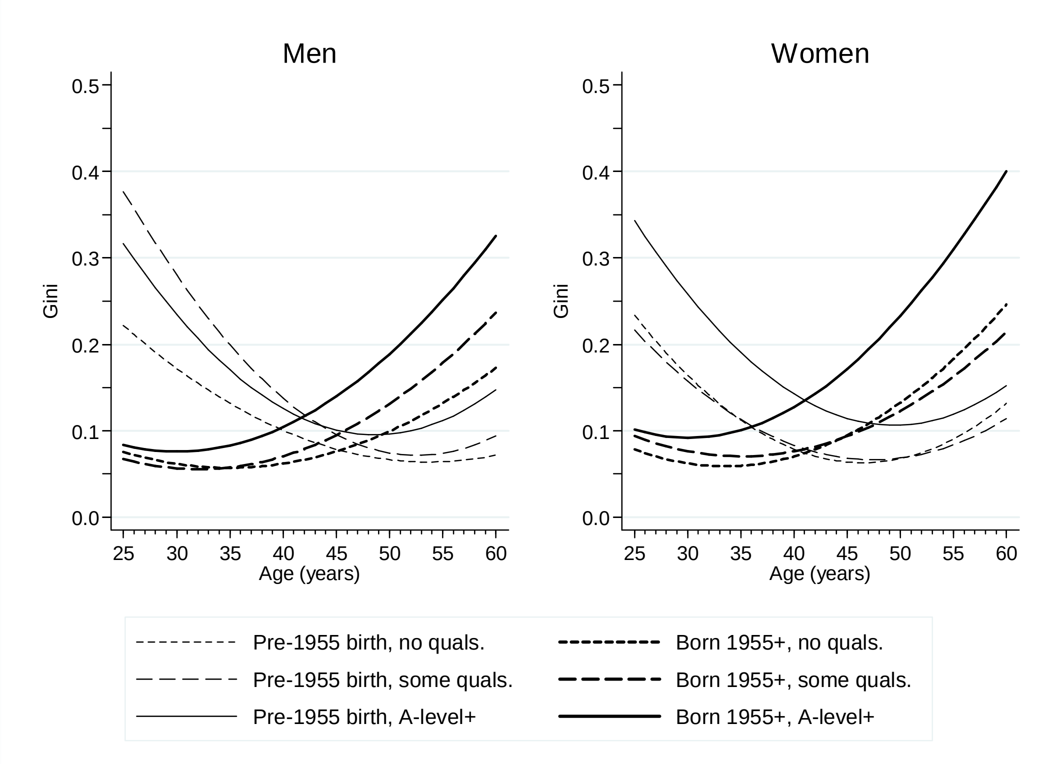


Figure 12
Inequality and age: Gini coefficient of hourly wages, by group



More on within-group differences in trajectories

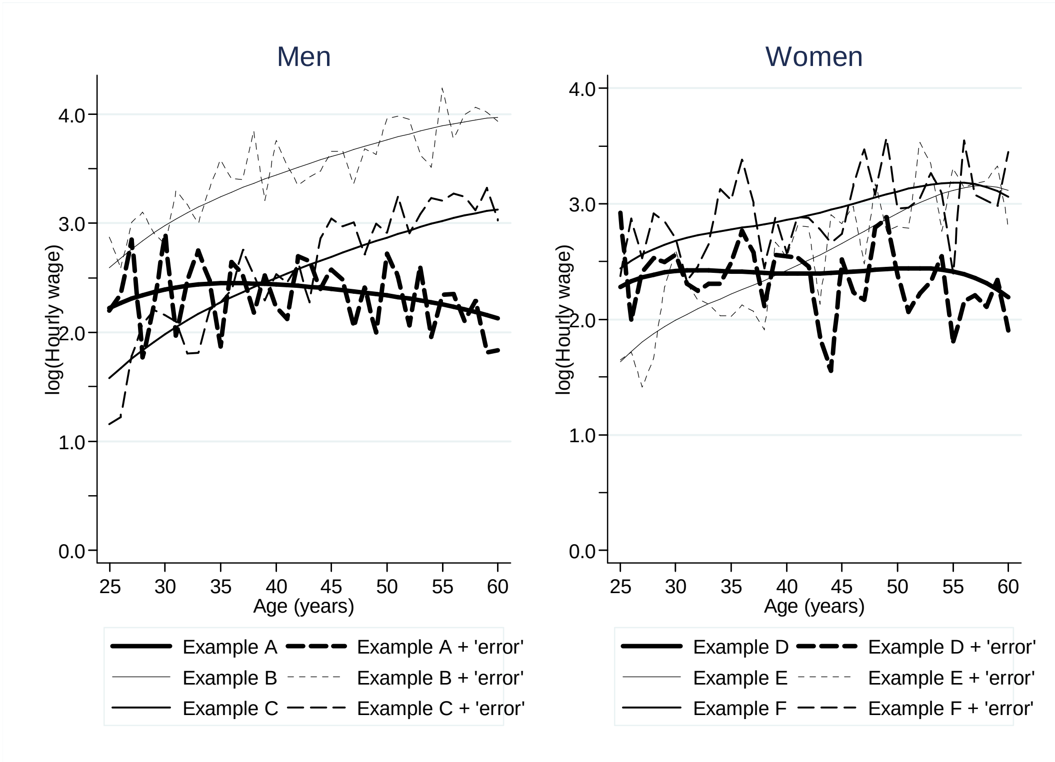
The discussion so far has emphasised the importance of individual deviations from average trajectories in terms of the income differences within groups at each age; what the discussion did not do was describe the shapes of of complete trajectories for different individuals and show how these profiles differed from the average trajectory. So far, there has not been a model-based counterpart to Figure 1. I now provide one, in two ways.

First, I use the model estimates to simulate complete trajectories for a set of individuals with the same observed characteristics. Within-group heterogeneity is summarised by the joint distribution of the individual-specific differences in intercepts, slopes and transitory errors. These are fully characterized by σ_α, σ_β, $\sigma_{\alpha\beta}$, and σ_v for the relevant group, and so I randomly draw several sets of values – one for each hypothetical individual – from the joint distribution using the estimates of these parameters. Then, combined with the estimates of β_0 , γ, δ, and ϕ (common within the group), I plot the wage-age trajectory that is implied for each of the hypothetical individuals.

Figure 13 shows the results of this exercise. It refers to men and women born in or after 1955 with A-level(s)+ qualifications, and the simulated trajectories refer to three men and three women. The average within-group trajectory is shown in Figure 7, and increases with age for both men and women (albeit at different rates). In Figure 13, there is a pair of trajectories shown for each person. The solid line shows the trajectory implied were there no transitory variation in wages (random values for v_{it} were not used in the simulation), and the accompanying dashed line shows the trajectory including simulated transitory variation.

The graphs show substantial differences in complete trajectories even within the same group. Features emphasized earlier such as the dispersion in initial incomes and dispersions in income growth rates are readily apparent again. But the diagram also brings out other features. In particular, the model set out in (7) is consistent with within-group trajectories not crossing as well as not crossing: there are examples of both scenarios for men and for women in Figure 13. In addition, individual trajectories can be negatively sloped over most of the working life even though the group average trajectory increases with age over the full age range.

Figure 13
The heterogeneity of individual trajectories: simulated data example



Note: trajectories simulated using model estimates for men and women born in or after 1955 with A-level(s)+ qualifications. See text for further details.

Figure 13 also shows that transitory variation plays a major role in generating trajectory ‘spaghetti’. Without the simulated transitory error term, profiles are relatively smooth. I emphasize the amplitude of the transitory variations, and not the temporal pattern of the errors for a given individual, because the model specification does not allow transitory shocks to have effects on wages that persistent beyond the year in which they initially occur (see the earlier discussion).

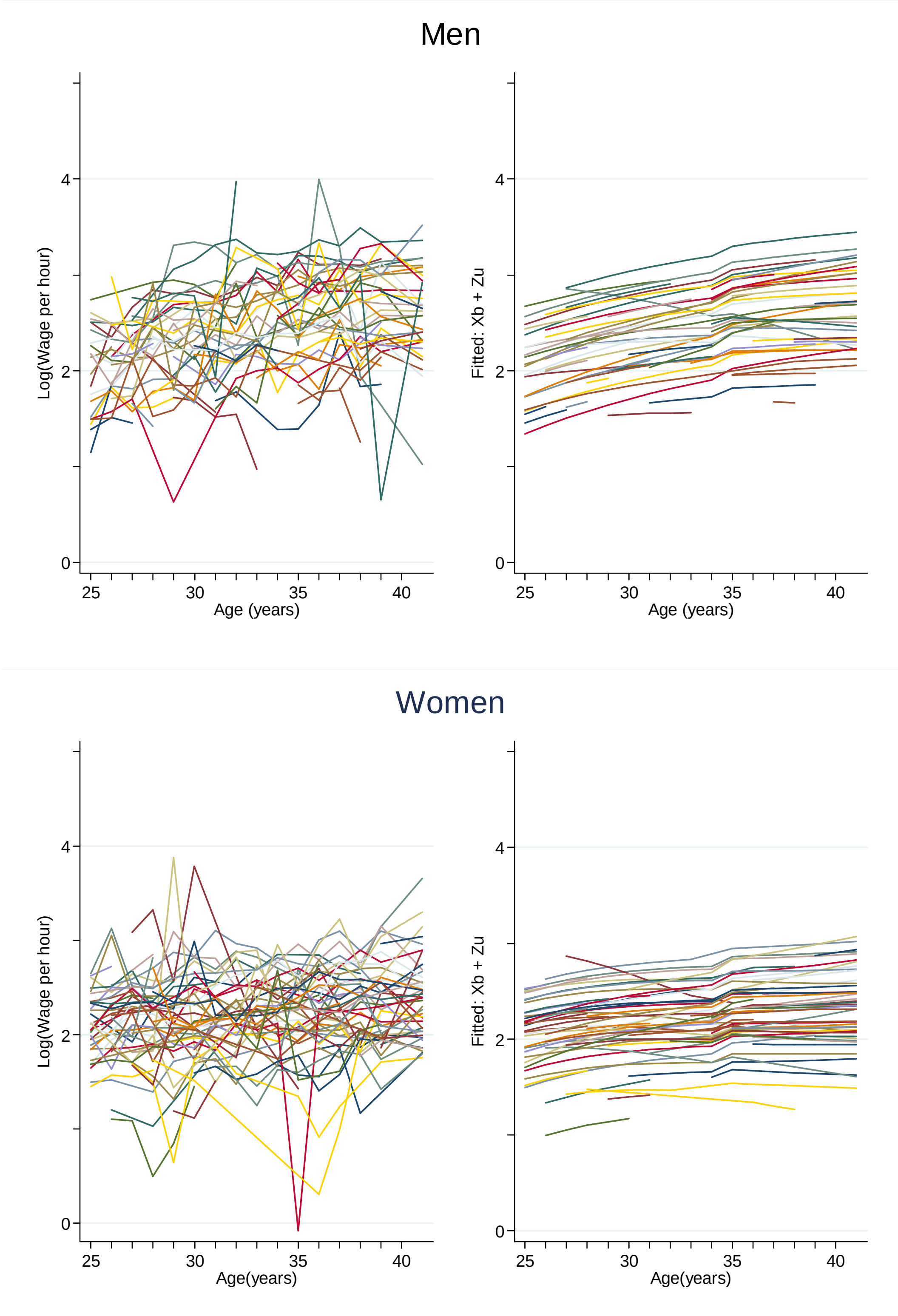
A second approach to examine the model’s predictions of the complete trajectories for individuals observed in the analysis sample. Moreover, by comparing these individual-level predictions with the actual trajectories, we get an additional perspective on the role of transitory errors. To make the example more manageable, I focus on men and women born in 1966 (i.e. belong to the later-born cohort) and with A-level(s)+ qualifications – these are the employees whose trajectory spaghetti was summarized in Figure 2 (bottom panel, row 1) earlier. The ‘fitted’ curves for each individual show predicted log wages taking into account observed characteristics (age and the period-specific intercept in this context) and the best linear unbiased predictors of the individual-specific error components (α_i, β_i).[31] These fitted values do not include the effects of transitory variation.

The graphs shown in Figure 14 underline the point that there is substantial variation in individual income-age trajectories in reality. Even with the smoothing of profiles incorporated into the derivation of the predicted profiles for each individual, it is clear that there is substantial variation in complete profiles. There are large differences in fitted log wages at the start of the working life, and thereafter the fitted profiles for many move broadly in parallel, increasing with age.[32] But there is a relatively high prevalence of fitted profiles that cross-over, and there is a minority for whom the fitted trajectory is distinctly downwards, in contrast to the average pattern for the group (the prevalence of the latter feature appears greater for women in this case).

Figure 14 also underlines that it is the transitory component to income that cooks the spaghetti, introducing many of the year-to-year wiggles in trajectory shapes. Without transitory variation, trajectories are similar to what uncooked spaghetti looks like when it comes out of the packet.

[31] The formulae used to derive the fitted values is documented in the `xtmixed` entry in the Stata manuals (StataCorp 2007).
[32] The shift upwards in the middle of the fitted profiles arises from the allowance for a period effect – a different intercept for the 1990s and the 2000s.

Figure 14
Log(wage) trajectories for men and women born in 1966 with A-level(s)+ qualifications: observed (LHS) versus fitted (RHS)



How do findings change if broader measures of income are used?

All of the analysis reported so far for hourly wages was replicated using each of the other two income measures, namely individual income and equivalized net household income. The corresponding tables of parameter estimates and derived figures are reported in Appendices B and C. I do not discuss all of these results because the main conclusions about the nature of the heterogeneity of income-age trajectories within groups were very much the same as for wages. One exception was that inequality at each age was markedly higher for individual income than for the other two variables (as remarked on earlier in the context of the raw data on trajectories).

The results that were most different across the measures concerned the shapes of the average trajectories. Figure 15 shows the average trajectories estimated for the three measures of income. All the graphs use a logarithmic scale, but observe that the range differs across the charts: compare the shapes of the graphs in panels (a), (b), (c), rather than their heights per se.

There are several similarities with the results for hourly wages. For example, other things equal, higher trajectories are associated with having more educational qualifications, and being a member of the 1955+ birth cohort rather than the pre-1955 one. And as with wages, profiles are higher for men rather than women of the same educational level, though the sex differential is less than for wages, especially for equivalized net household income – which is what is expected. The assumption of equal income sharing within households ensures this among couples, and they form a large proportion of households.

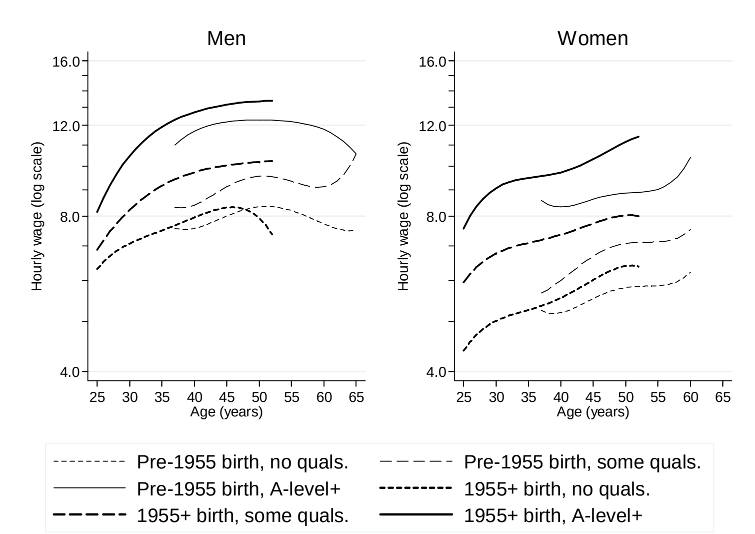
However there are noticeable differences across the measures in the shapes of the profiles for women, especially but not wholly those from the pre-1955 birth cohort. Whereas wages for women do not tend to fall as the state retirement pension age approaches, there is a clear decline in average individual income and especially equivalized net household income. Individual income and equivalized net household income reach a lifecourse peak in the 45–50 age range on average.

In addition, compared to the wage trajectories, there is more marked dip in income growth rates over child-raising ages for the other two measures (necessarily this result refers to the 1955+ birth cohort). This is particularly marked for individual income among women with A-level or high qualifications. For equivalized net household income, the impact on the most educated group of women is even more pronounced; indeed average incomes are predicted to fall in real terms before rising again at around age 40 rather than simply grow at a slower rate. According to this income measure too, there is a dip in the 30s for the two less qualified groups of women. And a dip is perceptible for men too, especially for those with no qualifications. An explanation for this might be that the partners of men without qualifications need to work to maintain family living standards much more than do partners of men with more qualifications. So if the partner stops work or reduces work hours for child-related reasons, this also affects the household income share that men (are assumed to) get.

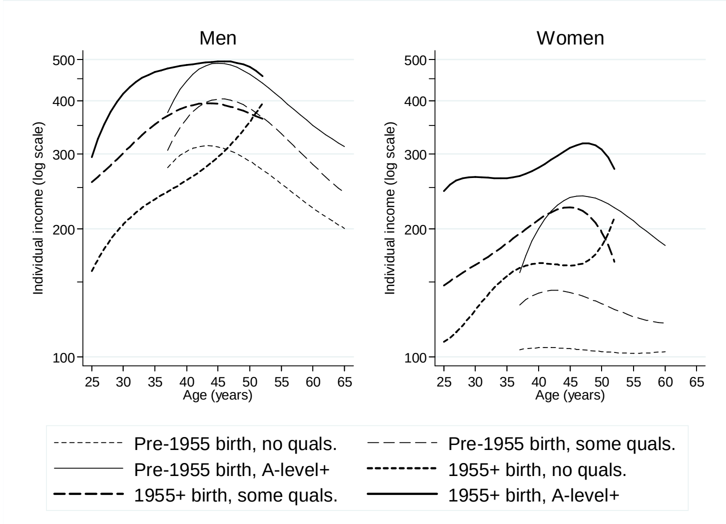
There continue to be some apparently anomalous results that are difficult to explain. For example, for individual income, average individual income does not flatten off in the late 40s for men, and around age 50 for women among those with no qualifications from the 1955+ cohort as it does for all other groups. It is hard to rationalize this this result with the selection arguments made earlier

Figure 15
Average income-age trajectories, by income measure (log scale)

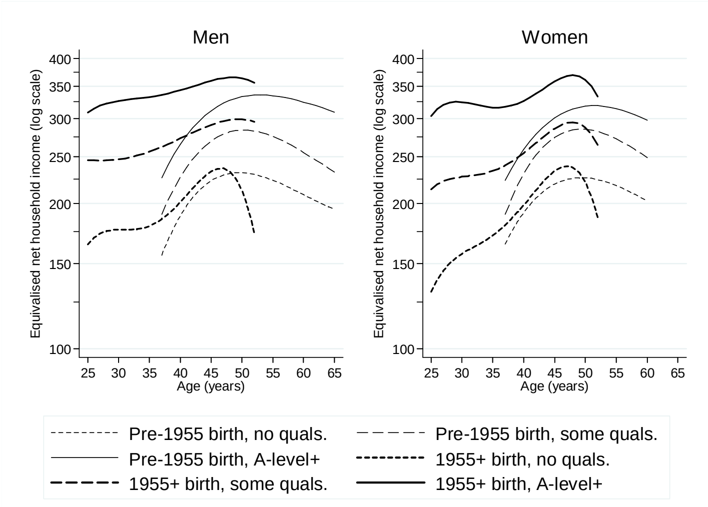
(a) hourly wage



(b) individual income



(c) equivalized net household income



7. Summary and conclusions

Most descriptions of the income-age relationship are based on comparisons of income across age groups in a particular year and are based on cross-sectional data. In contrast, this paper has taken a longitudinal approach, deriving trajectory estimates using 17 waves of data from the British Household Panel Survey. I have proposed a framework that provides summary descriptions of not only the way in which incomes among groups of similar individuals change with age on average, but also the way in which trajectories for individuals diverge from the average trajectory of their group.

The modelling framework is inspired by the literature on average income-age trajectory pioneered by Rowntree (2000 [1901]) but draws especially on the econometric literature on the modelling of the variance components of wages. Although my model's specification involves some compromises and hence disadvantages, I argue that it is fit for the purposes for which it was commissioned – to summarize variation with age of mean income and its variability, for different types of individuals and three measures of income (the hourly wage, total individual income from all sources, and equivalized net household income).

The analysis draws attention to the cooked spaghetti nature of income-age trajectories observed in panel data sets. I have argued that this pattern can usefully be summarized in terms of a number of factors. Looking at groups of individuals with similar observed characteristics, one can distinguish an average income-age trajectory for each group. Within groups, one can summarize differences across individuals in terms of: differences in incomes at the start of the working life; differences in income growth rates; and the nature of the association between initial incomes and income growth rates. (I found that those with lower initial incomes experience greater income growth on average.) In addition, income-age trajectories differ because of substantial individual-specific transitory income changes from one year to the next.

Across twelve groups defined in terms of combinations of sex, birth cohort, and educational qualifications, I find that some clear differences in group average income-age trajectories, regardless of the income measure used Other things equal, the income-age profile for men lies above that for women; the one for individuals born in or after 1955 is above that for those born before 1955; and the that for individuals with educational qualifications to A-level or higher is above that for indidividuals with some qualifications which, in turn, is above the profile for individuals with no educational qualifications. There is a distinct dip in income growth for women on average over the age range when many have children.

Average income-age trajectories derived from longitudinal data look different from those derived from cross-sectional data. For hourly wages for instance, trajectories at the beginning of the working life are steeper – wage growth is greater – according to longitudinal data.

Nonetheless, within each of the twelve groups, there are substantial differences across individuals in the shapes of income-age trajectories, with each of the sources identified above – differences in intercepts, slopes and their correlation, and transitory variation – playing a role. The model implies that over the working life, income inequality first declines and then rises, but the nature of the U-shape differs substantially between birth cohorts.

The paper has also argued that it is the transitory error component of income that cooks the spaghetti. These transitory changes may represent genuinely transitory effects on income, measurement error or, for broader measures of income, the effects on lifecourse events such having children, and family formation or dissolution. A task for future research is to incorporate more sophisticated assumptions about its nature and persistence over time. This is likely to facilitated by access to even longer panels than used in this study (perhaps from administrative record data, which may also have less measurement error than survey data). Long panels are necessary to help study the nature of income persistence in all its complex detail, including the extent to which observed short-run income changes for individuals are genuine.

References

Abowd, J. and Card, D. (1989), 'On the covariance structure of earnings and hours changes', *Econometrica*, 57, 411–445.

Baker, M. (1997), 'Growth-rate heterogeneity and the covariance structure of life-cycle earnings', *Journal of Labor Economics*, 15, 338–375.

Baker, M. and Solon, G. (2003), 'Earnings dynamics and inequality among Canadian men, 1967–1992: evidence from longitudinal income tax records', *Journal of Labor Economics*, 21, 289–322.

Bardasi, E., Rigg, J., and Jenkins, S.P. (2002), 'Retirement and the economic well-being of the elderly: a British perspective', *Ageing and Society*, 22 (2), 131–159.

Becker, G.S. (1993), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, third edition, University of Chicago Press, Chicago.

Biewen, M. (2003), 'The covariance structure of East and West German incomes and its implications for the persistence of poverty and inequality', *German Economic Review* 6 (4), 445–469.

Blundell, R., Reed, H. and Stoker, T.M. (2003), 'Interpreting aggregate wage growth: the role of labor force participation', *American Economic Review*, 93 (4), 1114–1131.

Deaton, A. and Paxson, C. (1994), 'Saving, growth and aging in Taiwan', in David A. Wise (ed.), *Studies in the Economics of Aging*, Wise, University of Chicago Press, Chicago, 331–357.

Department for Work and Pensions (2009), *Households Below Average Income. An Analysis of the Income Distribution 1994/95 – 2007/08*, Department for Work and Pensions, London. http://research.dwp.gov.uk/asd/hbai_arc.asp

Devicienti, F. (2001), 'Estimating poverty persistence in Britain', Working Paper No. 1, LABORatorio Riccardo Revelli, Collegio 'Carlo Alberto', Turin. http://repec.org/res2002/Devicienti.pdf

Dickens, R. (2000), 'The evolution of individual male wages in Britain, 1975–95', *Economic Journal*, 110, 27–49.

Dickens, R. and McKnight, A. (2008), 'Changes in earnings inequality and mobility in Great Britain 1978/9-2005/6', CASEPaper 132, London School of Economics, London. http://sticerd.lse.ac.uk/dps/case/cp/CASEpaper132.pdf

Duncan, G.J. (1983), 'The implications of changing family composition for the dynamic analysis of family economic well-being', in: A.B. Atkinson and F.A. Cowell (eds) *Panel Data on Incomes*, Occasional Paper No. 2, ICERD, London School of Economics, London.

Falkingham, J. and Hills, J. (eds) (1995), *The Dynamic of Welfare: The Welfare State and the Life-cycle*, Prentice Hall/HarvesterWheatsheaf, Hemel Hempstead.

Gangl, M. (2005), ‘Income inequality, permanent incomes, and income dynamics. Comparing Europe to the United States’, *Work and Occupations*, 32 (2), 140–162.
Gardiner, K. and Hills, J. (1999), ‘Policy implications of new data on income mobility’, *Economic Journal*, 109, F91–F111.

Haider, S.J. (2001), ‘Earnings instability and earnings inequality of males in the United States: 1967–1991’, *Journal of Labor Economics* 19, 799–836.

Hause, J.C. (1980), ‘The fine structure of earnings and the on-the-job training hypothesis’, *Econometrica*, 48 (4), 1013–1029.

Jarvis, S. and Jenkins, S.P. (1998), ‘How much income mobility is there in Britain?”, *Economic Journal* 108 (447), 428–443.

Jarvis, S. and Jenkins, S.P. (1999), ‘Marital splits and income changes: evidence from the British Household Panel Survey’, *Population Studies*, 53 (2), 237–254.

Jenkins, S.P. (2008), ‘Poverty transitions and life course events’, Plenary lecture at the “A decade of the lifecourse” Conference, 29–30 September 2008, Australian National University, Canberra.

Jenkins, S.P. (2009), ‘Marital splits and income changes over the longer term’, in M. Brynin and J.F. Ermisch (eds), *Changing Relationships*, Routledge, London, 217–236.

Jenkins, S.P. and Rigg, J.A. (2001), *The Dynamics of Poverty in Britain*, Department for Work and Pensions Research Report No. 157, Corporate Document Services, Leeds.

Jenkins, S.P. and Rigg, J.A. (2004), ‘Disability and disadvantage: selection, onset, and duration effects’, *Journal of Social Policy*, 33 (3), 479–501.

Lazear, E. (1995), *Personnel Economics*, The MIT Press, Cambridge MA.

Levy, H. and Jenkins S.P. (2008), ‘Derived net current and annual income variables to accompany BHPS waves 1–16’, Dataset deposited at the UK Data Archive (Study Number 3909), November 2008.

Lillard, L. and Weiss, Y. (1979), ‘Components of variation in panel earnings data: American scientists, 1960–1970’, *Econometrica*, 47, 437–454.

Lillard, L. and Willis, R. (1978), ‘Dynamic aspects of earnings mobility’, *Econometrica*, 46, 985–1012.

MaCurdy, T. (1982), ‘The use of time series processes to model the error structure of earnings in a longitudinal data analysis’, *Journal of Econometrics*, 18, 83–114.

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

Moffitt, R. and Gottschalk, P. (1995), ‘Trends in the covariance structure of earnings in the U.S.: 1969–1987’, unpublished paper, Economics Department, Johns Hopkins University. http://www.econ.jhu.edu/people/moffitt/

Moffitt, R. and Gottschalk, P. (2002), ‘Trends in the transitory variance of earnings in the United States’, *Economic Journal*, 112, C68–C73.

Moffitt, R. and Gottschalk, P. (2008), ‘Trends in the transitory variance of male earnings in the U.S., 1970-2004’, unpublished paper, Economics Department, Johns Hopkins University.

Ramos, X. (2003), ‘The covariance structure of earnings in Britain, 1991–1999’, *Economica*, 70, 353–374.

Rigg, J. and Sefton, T. (2006), ‘Income dynamics and the life cycle, *Journal of Social Policy*, 35 (3), 411–435.

Rowntree, B.S. (2000, originally 1901) *Poverty: a Study of Town Life*, Centennial Edition, Policy Press, Bristol (originally: Macmillan, London).

StataCorp. (2007), *Stata Statistical Software: Release 10.0*. StataCorp LP, College Station, TX:.

Stevens, A.H. (1999), ‘Climbing out of poverty, falling back in: measuring the persistence of poverty over multiple spells’, *Journal of Human Resources*, 34 (3), 557–588.

Uhrig, S.C.N. (2008), ‘The nature and causes of attrition in the British Household Panel Survey’, ISER Working Paper 2008-05, Institute for Social and Economic Research, University of Essex, Colchester. http://www.iser.essex.ac.uk/publications/working-papers/iser/2008-05

Wren-Lewis, L., Muriel, A., and Brewer, M. (2009), ‘Accounting for changes in inequality since 1968: decomposition analyses for Great Britain’, Report prepared for the National Equality Panel, Institute for Fiscal Studies, London.

Zaidi, M.A. (2001), ‘Snakes and ladders: an analysis of life-course events and income mobility in old age’, SAGE Discussion Paper 8, London School of Economics, London.

Appendix A

Results for log(hourly wage)

Table A1
Model parameter estimates: log(hourly wage), by group

Pre-1955 birth, no educ. quals	**Men**		**Women**	
Age (years)	–1.9829		–3.5350	
	(1.285)		(2.327)	
Age^2	0.0606		0.1116	
	(0.038)		(0.072)	
$Age^3/100$	–0.0805		–0.1548	
	(0.050)		(0.100)	
$Age^4/10000$	0.0393		0.0797	
	(0.024)		(0.051)	
Year: 1991–2000	–0.0982	***	–0.0907	***
	(0.019)		(0.022)	
Intercept	25.8361		43.0843	
	(16.062)		(27.841)	
sd(slope)	0.0221	***	0.0308	***
	(0.002)		(0.002)	
sd(intercept)	1.2354	**	1.4643	***
	(0.109)		(0.110)	
corr(intercept, slope)	–0.9624	***	–0.9775	***
	(0.008)		(0.004)	
sd(error)	0.2176	***	0.2502	***
	(0.004)		(0.004)	
Log-likelihood	–361.133		–811.797	
No. person-years	2379		2948	
No. individuals	456		561	
Wald test p-value	0.0010		0.1549	

Born 1955+, no educ. quals	**Men**		**Women**	
Age (years)	0.8574		1.0345	
	(0.842)		(0.997)	
Age^2	–0.0350		–0.0405	
	(0.035)		(0.040)	
$Age^3/100$	0.0641		0.0701	
	(0.062)		(0.072)	
$Age^4/10000$	–0.0438		–0.0447	
	(0.041)		(0.047)	
Year: 1991–2000	–0.1086	***	–0.0878	***
	(0.027)		(0.032)	
Intercept	–6.0161		–8.2809	
	(7.572)		(9.093)	
sd(slope)	0.0216	***	0.0262	***
	(0.004)		(0.004)	
sd(intercept)	0.8076		0.9142	
	(0.137)		(0.156)	
corr(intercept, slope)	–0.9336	***	–0.9588	***
	(0.028)		(0.017)	
sd(error)	0.2440	***	0.2828	***

	Men		Women	
	(0.006)		(0.007)	
Log-likelihood	–281.462		–386.693	
No. person-years	1226		1148	
No. individuals	266		225	
Wald test p-value	0.0461		0.5746	

Pre-1955 birth, some educ. quals	**Men**		**Women**	
Age (years)	–3.1882	**	–2.6168	
	(1.559)		(2.057)	
Age^2	0.1005	**	0.0861	
	(0.047)		(0.064)	
$Age^3/100$	–0.1382	**	–0.1231	
	(0.061)		(0.088)	
$Age^4/10000$	0.0702	**	0.0650	
	(0.030)		(0.045)	
Year: 1991–2000	–0.0378	*	–0.0492	***
	(0.023)		(0.018)	
Intercept	39.3945	**	30.9134	
	(19.433)		(24.639)	
sd(slope)	0.0325	***	0.0276	***
	(0.003)		(0.002)	
sd(intercept)	1.7415	***	1.3493	***
	(0.131)		(0.103)	
corr(intercept, slope)	–0.9806	***	–0.9691	***
	(0.004)		(0.005)	
sd(error)	0.2507	***	0.2362	***
	(0.004)		(0.003)	
Log-likelihood	–632.351		–699.198	
No. person-years	2230		3469	
No. individuals	393		526	
Wald test p-value	0.0026		0.0003	

Born 1955+, some educ. quals	**Men**		**Women**	
Age (years)	0.2208		0.8170	**
	(0.411)		(0.386)	
Age^2	–0.0055		–0.0311	*
	(0.017)		(0.016)	
$Age^3/100$	0.0061		0.0526	*
	(0.032)		(0.029)	
$Age^4/10000$	–0.0025		–0.0329	*
	(0.022)		(0.020)	
Year: 1991–2000	–0.0637	***	–0.0892	***
	(0.012)		(0.013)	
Intercept	–0.9997		–6.1305	*
	(3.591)		(3.419)	
sd(slope)	0.0244	***	0.0243	***
	(0.002)		(0.002)	
sd(intercept)	0.8379	***	0.9134	***
	(0.049)		(0.053)	
corr(intercept, slope)	–0.9331	***	–0.9331	***

	(0.009)		(0.009)	
sd(error)	0.2198	***	0.2619	***
	(0.003)		(0.003)	
Log-likelihood	–671.021		–1722.923	
No. person-years	4599		5935	
No. individuals	883		994	
Wald test p-value	0.0000		0.0000	

Pre-1955 Birth, A-level(s) +	**Men**		**Women**	
Age (years)	1.1379		–3.9504	*
	(0.875)		(2.271)	
Age^2	–0.0333		0.1262	*
	(0.026)		(0.071)	
$Age^3/100$	0.0437		–0.1778	*
	(0.035)		(0.097)	
$Age^4/10000$	–0.0216		0.0934	*
	(0.017)		(0.050)	
Year: 1991–2000	–0.0603	***	–0.0617	***
	(0.013)		(0.018)	
Intercept	–12.1514		48.1305	*
	(10.826)		(27.212)	
sd(slope)	0.0321	***	0.0323	***
	(0.002)		(0.002)	
sd(intercept)	1.6187	***	1.6505	***
	(0.080)		(0.099)	
corr(intercept, slope)	–0.9646	***	–0.9670	***
	(0.004)		(0.005)	
sd(error)	0.2385	***	0.3000	***
	(0.003)		(0.004)	
Log-likelihood	–1435.598		–1970.318	
No. person-years	5926		4529	
No. individuals	808		676	
Wald test p-value	0.0010		0.1244	

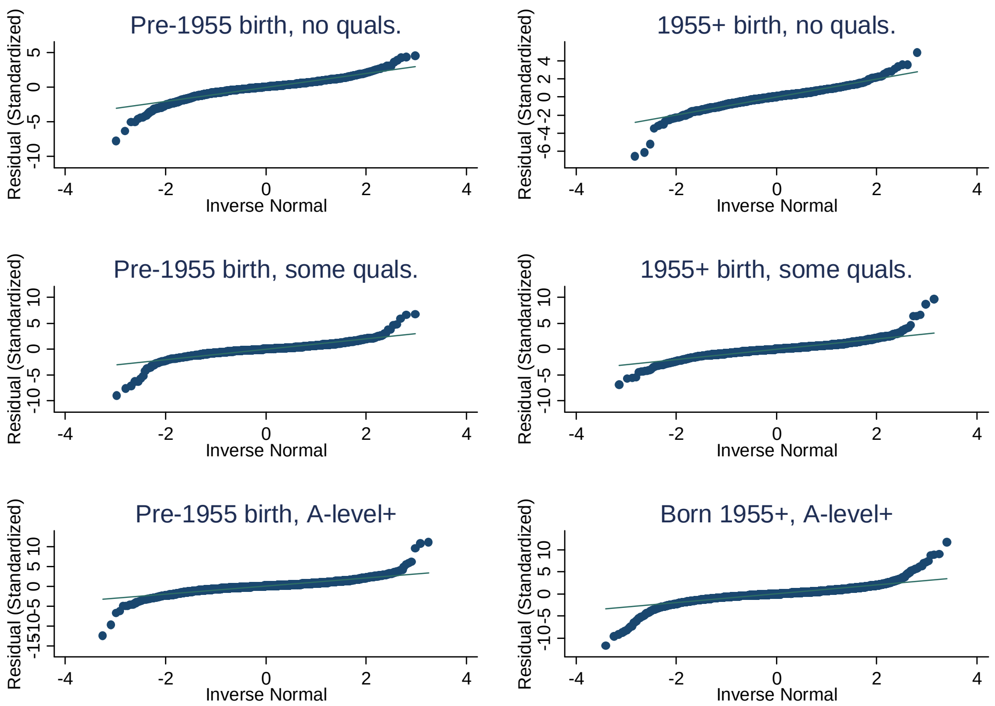
Born 1955+, A-level(s) +	**Men**		**Women**	
Age (years)	0.5426	**	1.1806	***
	(0.232)		(0.271)	
Age^2	–0.0169	*	–0.0439	***
	(0.010)		(0.011)	
$Age^3/100$	0.0240		0.0719	***
	(0.018)		(0.021)	
$Age^4/10000$	–0.0130		–0.0433	***
	(0.012)		(0.014)	
Year: 1991–2000	–0.0757	***	–0.0947	***
	(0.008)		(0.009)	
Intercept	–4.1421	**	–9.5724	***
	(2.064)		(2.400)	
sd(slope)	0.0269	***	0.0303	***
	(0.001)		(0.001)	
sd(intercept)	0.8889	***	0.9931	

	(0.033)		(0.037)	
corr(intercept, slope)	–0.9115	***	–0.9188	***
	(0.007)		(0.007)	
sd(error)	0.2382	***	0.2775	***
	(0.002)		(0.002)	
Log-likelihood	–2824.082		–4473.029	
No. person-years	12476		11961	
No. individuals	2022		2012	
Wald test p-value	0.0000		0.0000	

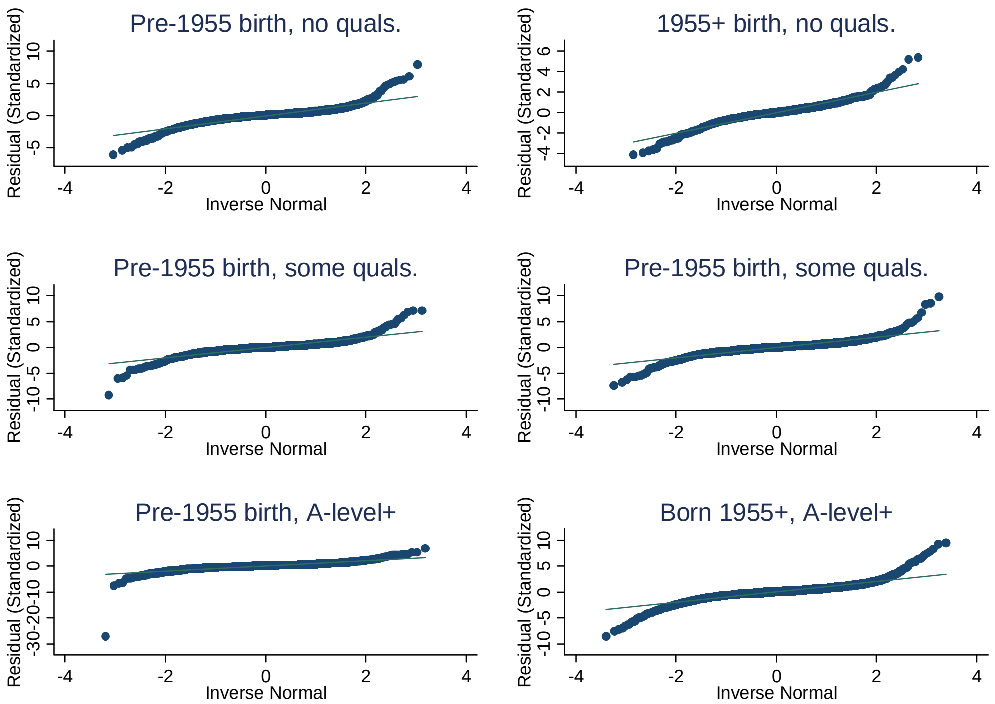
***: $p < 0.01$. **: $p < 0.05$. *: $p < 0.10$.

Wald test p-value: p-value from Wald test that coefficients on all age variables jointly zero.

Figure A1
Quantile plots of standardized residuals against standard normal, by group
Men



Women



Note: Quantile plot graphs quantiles of the distribution of estimated residuals (standardized by their estimated variance) against quantiles of a standard normal distribution. If the plot lies wholly on the 45° ray through the origin, the normal distribution is appropriate. The plots show consistency with normality, except at the extreme tails of the residual distributions

Appendix B

Results for log(individual income)

Table B1
Model parameter estimates: log(hourly wage), by group

Pre-1955 birth, no educ. quals	**Men**		**Women**	
Age (years)	0.9806	***	0.2878	
	(0.209)		(0.185)	
Age^2	–0.0235	***	–0.0076	*
	(0.005)		(0.004)	
$Age^3/100$	0.0239	***	0.0084	*
	(0.005)		(0.005)	
$Age^4/10000$	–0.0088	***	–0.0032	*
	(0.002)		(0.002)	
Year: 1991–2000	–0.1069	***	–0.1483	***
	(0.017)		(0.015)	
Intercept	–8.8875	***	0.7183	
	(3.211)		(2.838)	
sd(slope)	0.0203	***	0.0394	***
	(0.001)		(0.001)	
sd(intercept)	1.5167	***	2.9532	***
	(0.088)		(0.087)	
corr(intercept, slope)	–0.9717	***	–0.9883	***
	(0.004)		(0.001)	
sd(error)	0.5870	***	0.6044	***
	(0.005)		(0.004)	
Log-likelihood	–9959.615		–17084.297	
No. person–years	10089		16129	
No. individuals	1221		1892	
Wald test p-value	0.0000		0.0041	

Born 1955+, no educ. quals	**Men**		**Women**	
Age (years)	0.6520		–1.9112	
	(1.623)		(1.315)	
Age^2	–0.0208		0.0850	
	(0.067)		(0.054)	
$Age^3/100$	0.0290		–0.1605	*
	(0.121)		(0.097)	
$Age^4/10000$	–0.0140		0.1101	*
	(0.081)		(0.065)	
Year: 1991–2000	0.1023	**	–0.2194	***
	(0.052)		(0.044)	
Intercept	–2.2202		20.1408	*
	(14.527)		(11.829)	
sd(slope)	0.0472	***	0.0568	***
	(0.006)		(0.005)	
sd(intercept)	1.7669	***	2.0857	***
	(0.206)		(0.166)	
corr(intercept, slope)	–0.9386	***	–0.9667	***
	(0.018)		(0.007)	
sd(error)	0.6928	***	0.5960	***
	(0.012)		(0.010)	

Log-likelihood	–2584.133		–2663.717	
No. person–years	2145		2505	
No. individuals	365		363	
Wald test p-value	0.1742		0.0019	

Pre-1955 birth, some educ. quals	**Men**		**Women**	
Age (years)	1.2307	***	1.0952	***
	(0.330)		(0.324)	
Age^2	–0.0278	***	–0.0282	***
	(0.008)		(0.008)	
$Age^3/100$	0.0263	***	0.0308	***
	(0.009)		(0.009)	
$Age^4/10000$	–0.0089	***	–0.0121	***
	(0.003)		(0.003)	
Year: 1991–2000	–0.0935	***	–0.0431	*
	(0.024)		(0.024)	
Intercept	–13.4285	***	–10.4345	**
	(4.920)		(4.804)	
sd(slope)	0.0214	***	0.0528	***
	(0.002)		(0.002)	
sd(intercept)	1.4454	***	3.2255	***
	(0.130)		(0.129)	
corr(intercept, slope)	–0.9424	***	–0.9595	***
	(0.011)		(0.004)	
sd(error)	0.6695	***	0.6474	***
	(0.007)		(0.005)	
Log-likelihood	–7450.024		–10609.584	
No. person–years	6644		9074	
No. individuals	739		948	
Wald test p-value	0.0000		0.0000	

Born 1955+, some educ. quals	**Men**		**Women**	
Age (years)	–0.4806		0.7289	
	(1.088)		(0.699)	
Age^2	0.0215		–0.0330	
	(0.046)		(0.029)	
$Age^3/100$	–0.0383		0.0676	
	(0.085)		(0.053)	
$Age^4/10000$	0.0237		–0.0513	
	(0.058)		(0.036)	
Year: 1991–2000	–0.0203		–0.1105	***
	(0.030)		(0.023)	
Intercept	9.1642		–1.1696	
	(9.544)		(6.176)	
sd(slope)	0.0352	***	0.0712	***
	(0.004)		(0.003)	
sd(intercept)	1.3899	***	2.6088	***
	(0.120)		(0.093)	
corr(intercept, slope)	–0.9195	***	–0.9674	***
	(0.017)		(0.003)	

sd(error)	0.7706	***	0.5943	***
	(0.008)		(0.005)	
Log-likelihood	–8143.492		–9837.087	
No. person–years	6382		9008	
No. individuals	1063		1211	
Wald test p-value	0.0000		0.0000	

Pre-1955 Birth, A-level(s) +	**Men**		**Women**	
Age (years)	1.5401	***	1.6583	***
	(0.249)		(0.266)	
Age^2	–0.0369	***	–0.0383	***
	(0.006)		(0.007)	
$Age^3/100$	0.0378	***	0.0375	***
	(0.007)		(0.007)	
$Age^4/10000$	–0.0141	***	–0.0133	***
	(0.003)		(0.003)	
Year: 1991–2000	–0.0416	**	–0.0175	
	(0.019)		(0.021)	
Intercept	–17.0468	***	–20.4471	***
	(3.657)		(3.971)	
sd(slope)	0.0399	***	0.0546	***
	(0.002)		(0.002)	
sd(intercept)	2.2947	***	2.9963	***
	(0.097)		(0.115)	
corr(intercept, slope)	–0.9585	***	–0.9603	***
	(0.004)		(0.003)	
sd(error)	0.6620	***	0.6421	***
	(0.005)		(0.005)	
Log-likelihood	–14908.800		–12394.817	
No. person–years	13030		10758	
No. individuals	1275		1083	
Wald test p-value	0.0000		0.0000	

Born 1955+, A-level(s) +	**Men**		**Women**	
Age (years)	1.6517	***	2.0023	***
	(0.559)		(0.559)	
Age^2	–0.0601	***	–0.0841	***
	(0.023)		(0.023)	
$Age^3/100$	0.0980	**	0.1541	***
	(0.042)		(0.042)	
$Age^4/10000$	–0.0600	**	–0.1038	***
	(0.028)		(0.029)	
Year: 1991–2000	–0.0210		–0.0909	***
	(0.017)		(0.018)	
Intercept	–10.9760	**	–12.0472	**
	(4.969)		(4.953)	
sd(slope)	0.0559	***	0.0708	***
	(0.002)		(0.002)	
sd(intercept)	1.8326	***	2.4678	***
	(0.070)		(0.068)	

corr(intercept, slope)	–0.9572	***	–0.9611	***
	(0.004)		(0.003)	
sd(error)	0.6864	***	0.6812	***
	(0.004)		(0.004)	
Log-likelihood	–18649.852		–19256.509	
No. person–years	15768		15879	
No. individuals	2298		2250	
Wald test p-value	0.0000		0.0001	

***: $p < 0.01$. **: $p < 0.05$. *: $p < 0.10$.

Wald test p-value: p-value from Wald test that coefficients on all age variables jointly zero.

Figure B1
Estimated average income –age trajectories, by group, for individuals aged 25+

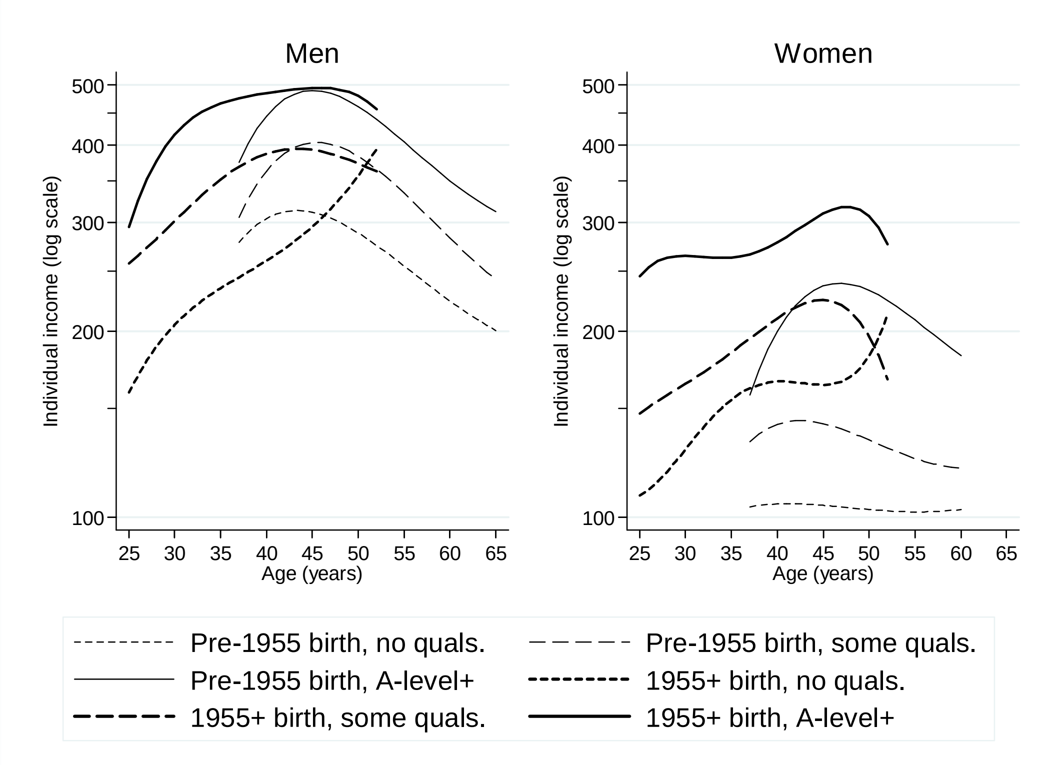
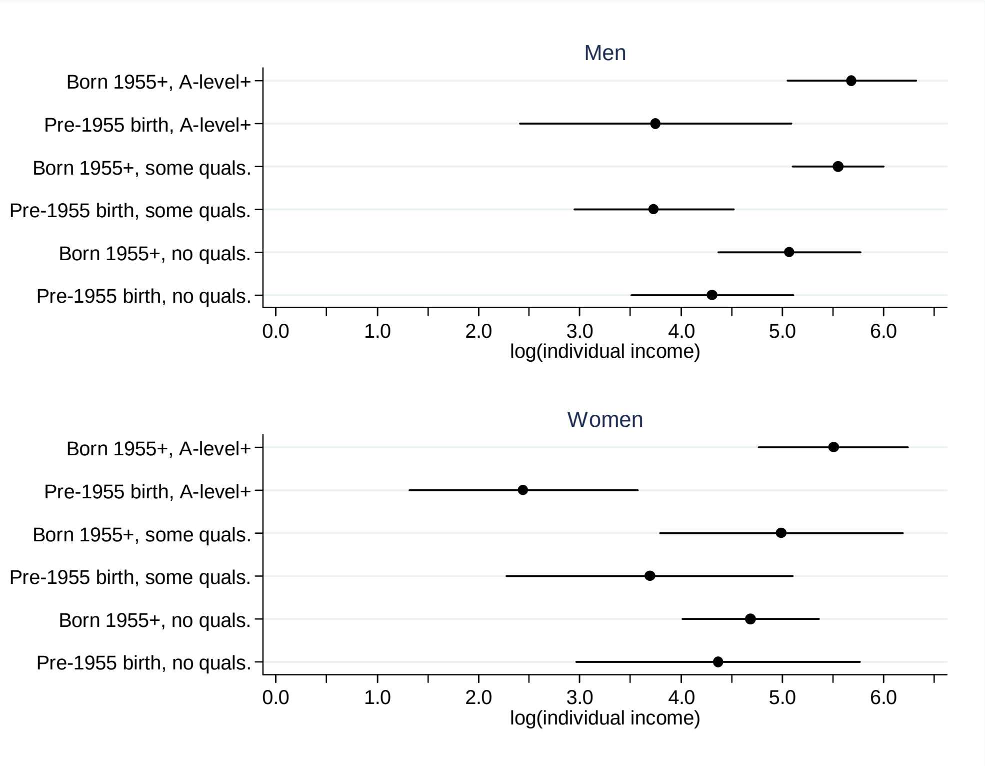


Table B2
Between-group differences in variance component parameters

Educational Qualifications	Men		Women		
	Pre-1955 birth	Born 1955+	Pre-1955 birth	Born 1955+	
sd(intercept): σ_α					
None	0.61	0.72	1.20	0.85	
Some	0.59	0.56	1.31	1.06	
A-level(s) +	0.93	0.74	1.21	1.00	[2.609]
sd(age coefficient): $\sigma_{\alpha\beta}$					
None	0.29	0.67	0.56	0.80	
Some	0.30	0.50	0.75	1.01	
A-level(s) +	0.56	0.79	0.77	1.00	[0.071]
corr(int., age coeff.): $\sigma_{\alpha\beta}$					
None	1.01	0.98	1.03	1.01	
Some	0.98	0.96	1.00	1.01	
A-level(s) +	1.00	1.00	1.00	1.00	[–0.967]
sd(error): $\sigma_{\nu\beta}$					
None	0.86	1.02	0.89	0.88	
Some	0.98	1.13	0.95	0.87	
A-level(s) +	0.97	1.01	0.94	1.00	[0.594]

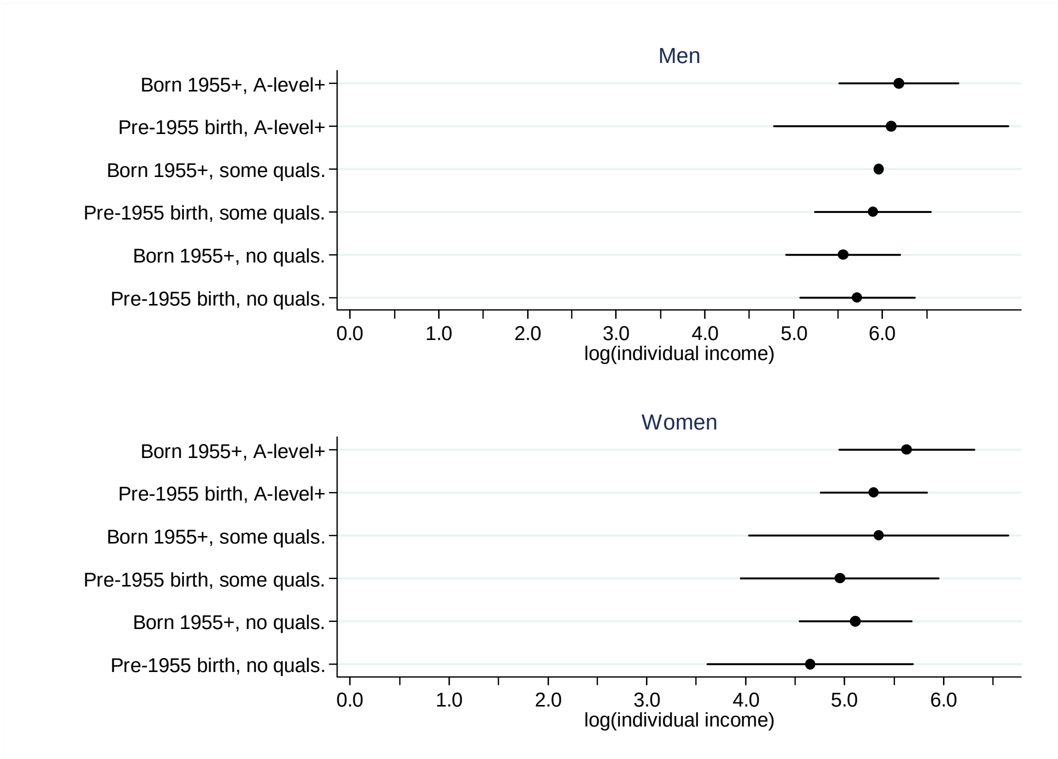
Note: group parameters expressed as a ratio of the parameters for women with A-level(s)+ born 1955+ (shown in brackets).

Figure B2
The distribution of log(individual income) at age 25, by group



The line for each subgroup shows the interquartile range (distance between the 25th and 75th percentiles). The filled circle shows the median.

Figure B3
The distribution of log(individual income) at age 40, by group



The line for each subgroup shows the interquartile range (distance between the 25th and 75th percentiles). The filled circle shows the median.

Figure B4
Inequality and age: variance of log(individual income), by group

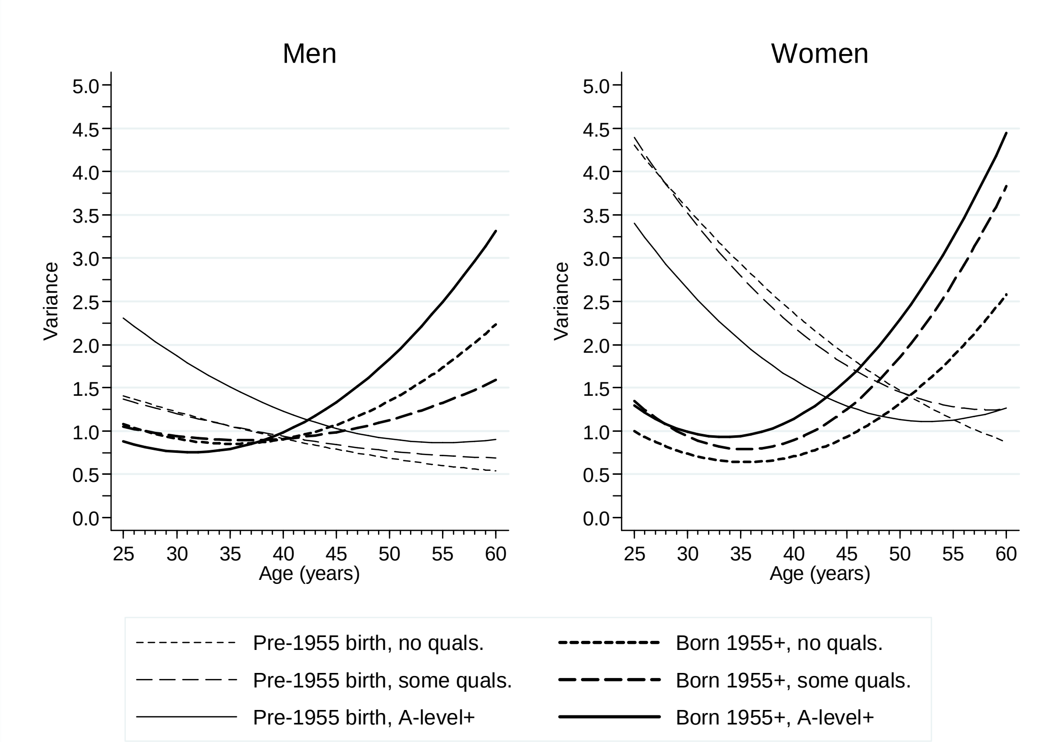


Figure B5
Inequality and age: Gini coefficient of individual income, by group

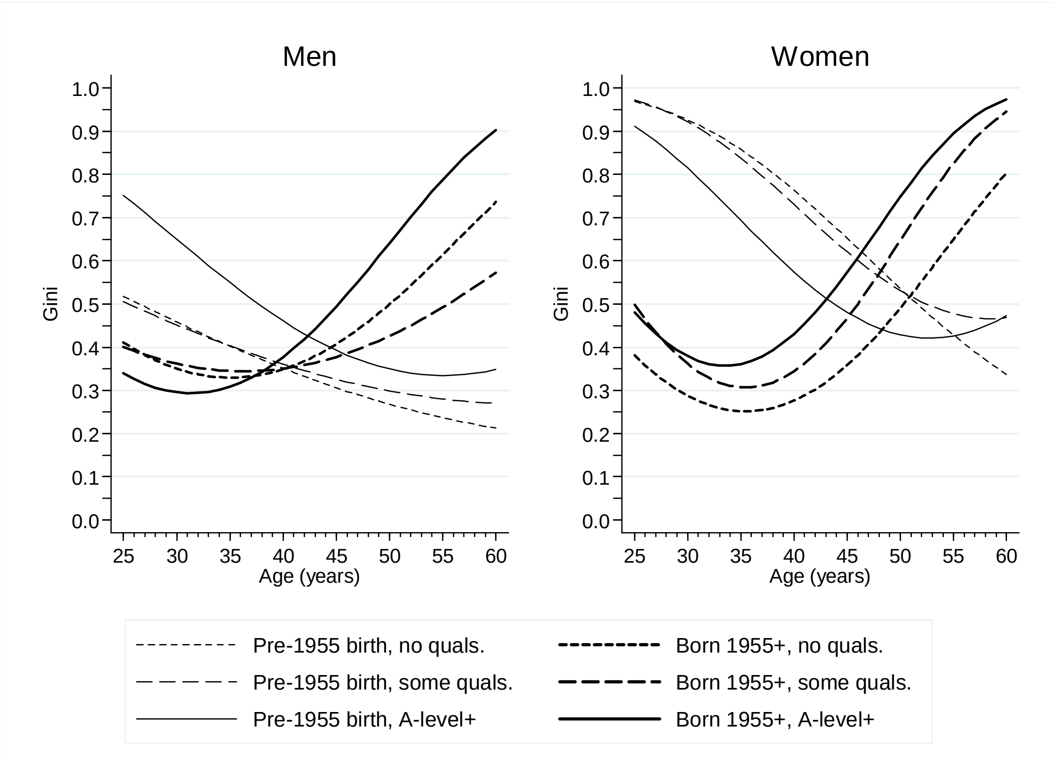


Figure B6
The heterogeneity of individual trajectories: simulated data example

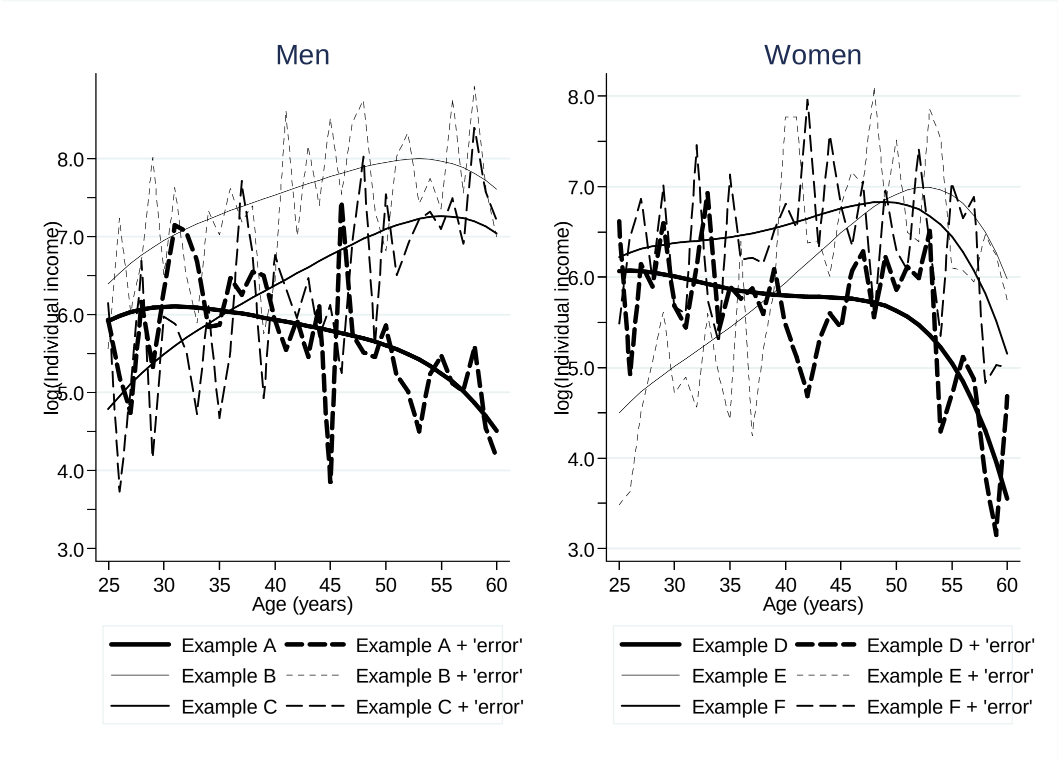
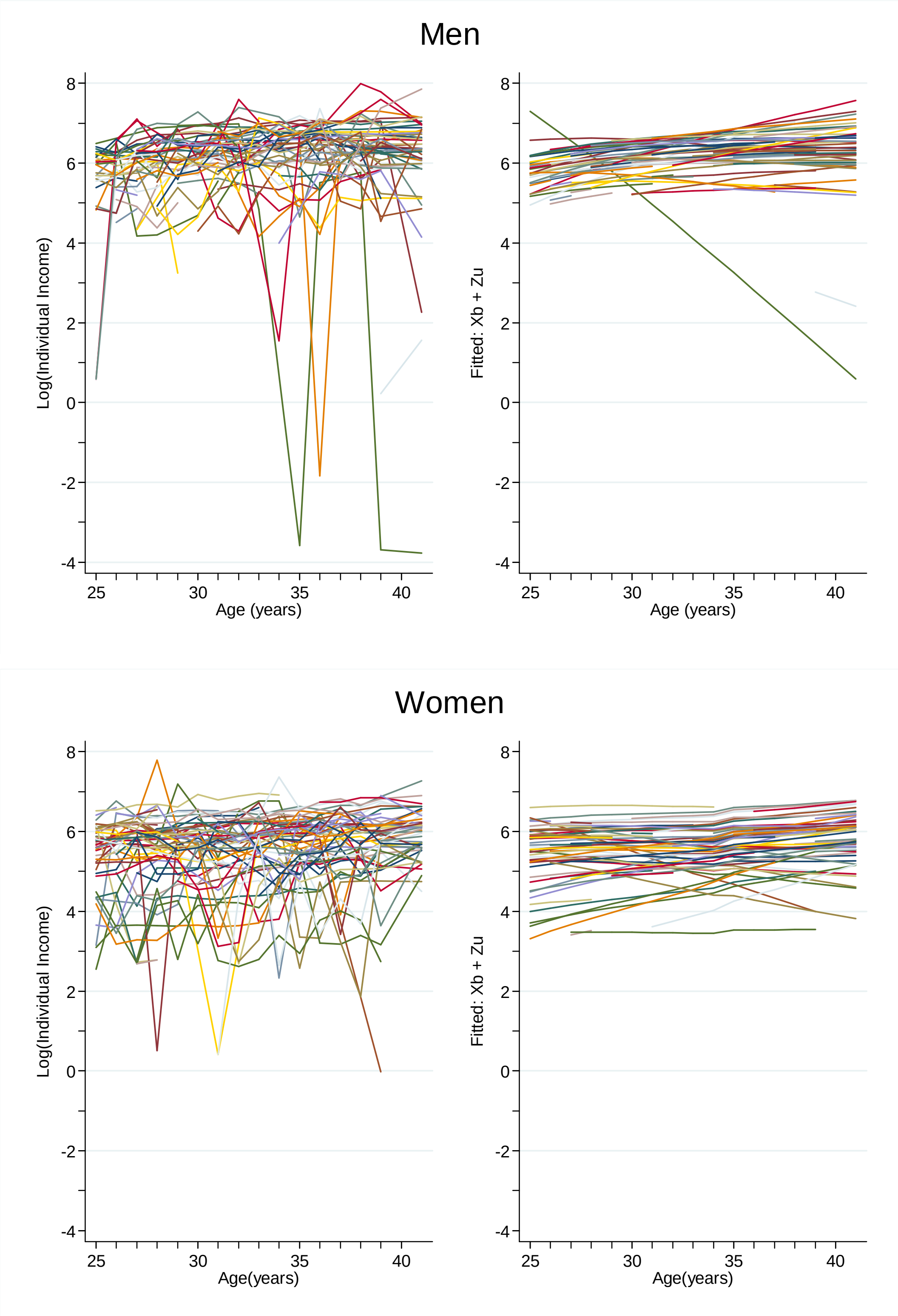
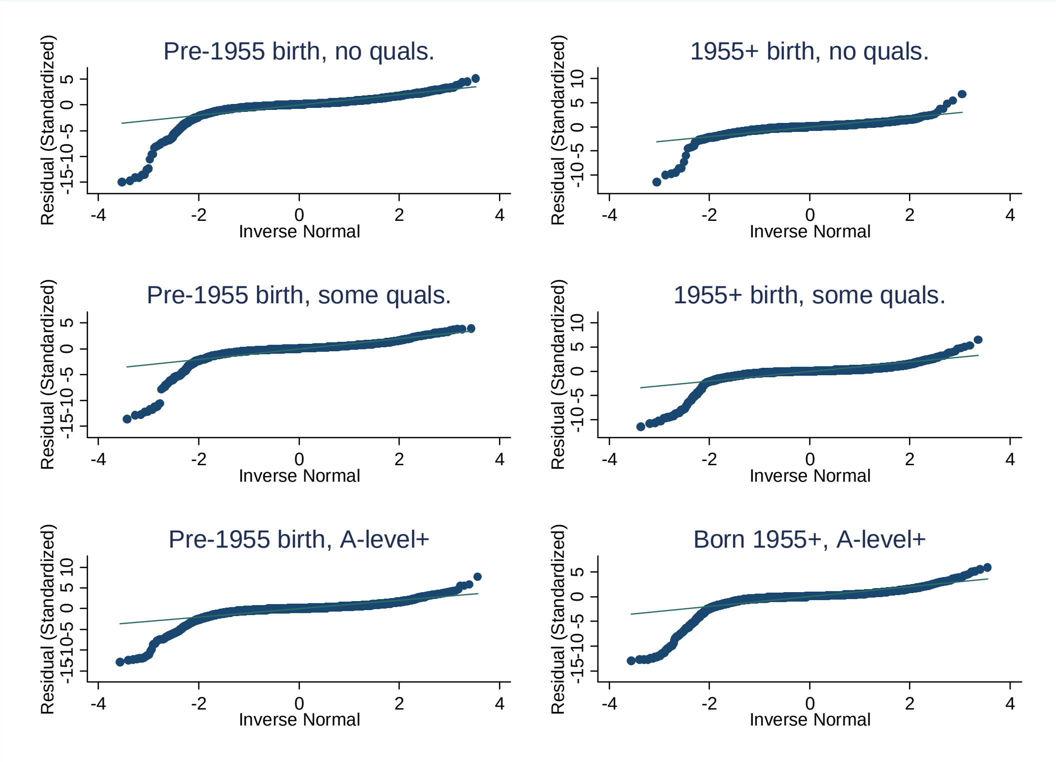


Figure B7
Log(individual income) trajectories for men and women born in 1966 with A-level(s)+ qualifications: observed (LHS) versus fitted (RHS)

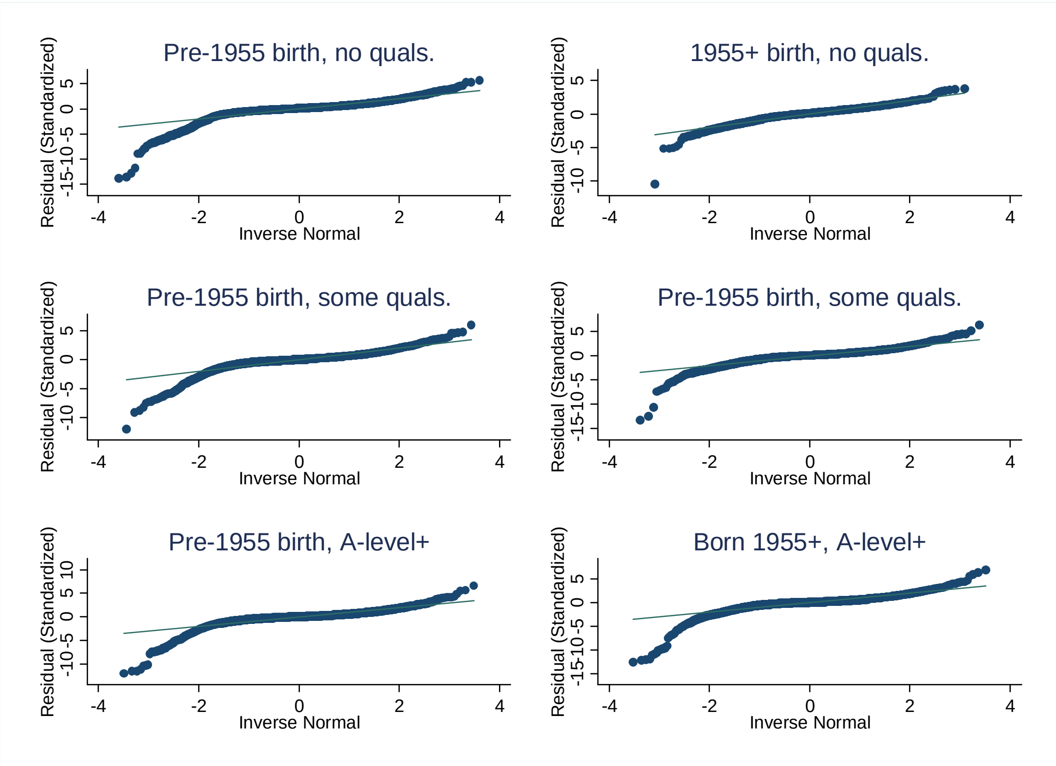


Note: see main text for explanation of the derivation of the graphs.

Figure B8
Quantile plots of standardized residuals against standard normal, by group
Men



Women



Note: Quantile plot graphs quantiles of the distribution of estimated residuals (standardized by their estimated variance) against quantiles of a standard normal distribution. If the plot lies wholly on the 45° ray through the origin, the normal distribution is appropriate. The plots show consistency with normality, except at the extreme tails of the residual distributions.

Appendix C

Estimates for log(equivalised net household income)

Table C1
Model parameter estimates: log(hourly wage), by group

Pre-1955 birth, no educ. quals	**Men**		**Women**	
Age (years)	1.0164	***	0.6717	***
	(0.147)		(0.121)	
Age^2	–0.0228	***	–0.0140	***
	(0.004)		(0.003)	
$Age^3/100$	0.0220	***	0.0122	***
	(0.004)		(0.003)	
$Age^4/10000$	–0.0078	***	–0.0037	***
	(0.001)		(0.001)	
Year: 1991–2000	–0.0945	***	–0.1348	***
	(0.012)		(0.009)	
Intercept	–11.0575	***	–6.0522	***
	(2.251)		(1.874)	
sd(slope)	0.0211	***	0.0212	***
	(0.001)		(0.001)	
sd(intercept)	1.4011	***	1.5220	***
	(0.067)		(0.056)	
corr(intercept, slope)	–0.9686	***	–0.9718	***
	(0.003)		(0.003)	
sd(error)	0.3477	***	0.3426	***
	(0.003)		(0.002)	
Log-likelihood	–4700.005		–7375.205	
No. person–years	9046		14444	
No. individuals	1138		1779	
Wald test p-value	0.0000		0.0000	

Born 1955+, no educ. quals	**Men**		**Women**	
Age (years)	2.4474	**	2.3795	**
	(1.127)		(1.044)	
Age^2	–0.1059	**	–0.1001	**
	(0.047)		(0.043)	
$Age^3/100$	0.2004	**	0.1861	**
	(0.086)		(0.079)	
$Age^4/10000$	–0.1391	**	–0.1276	**
	(0.058)		(0.053)	
Year: 1991–2000	–0.0144		–0.0658	**
	(0.033)		(0.030)	
Intercept	–15.7600		–16.1301	*
	(9.992)		(9.273)	
sd(slope)	0.0365	***	0.0277	***
	(0.004)		(0.003)	
sd(intercept)	1.4312	***	1.1823	*
	(0.127)		(0.113)	
corr(intercept, slope)	–0.9632	***	–0.9516	***
	(0.008)		(0.013)	
sd(error)	0.3957	***	0.3759	***
	(0.007)		(0.007)	

Log-likelihood	–1325.260		–1273.405	
No. person–years	1959		2078	
No. individuals	340		329	
Wald test p-value	0.0520		0.0223	

Pre-1955 birth, some educ. quals	**Men**		**Women**	
Age (years)	0.6386	***	0.9188	***
	(0.229)		(0.188)	
Age^2	–0.0123	**	–0.0196	***
	(0.006)		(0.005)	
$Age^3/100$	0.0094		0.0178	***
	(0.006)		(0.005)	
$Age^4/10000$	–0.0023		–0.0058	***
	(0.002)		(0.002)	
Year: 1991–2000	–0.0900	***	–0.0843	***
	(0.017)		(0.014)	
Intercept	–5.9146	*	–9.7765	***
	(3.414)		(2.790)	
sd(slope)	0.0247	***	0.0250	***
	(0.001)		(0.001)	
sd(intercept)	1.4678	***	1.4738	***
	(0.084)		(0.071)	
corr(intercept, slope)	–0.9640	***	–0.9620	***
	(0.005)		(0.004)	
sd(error)	0.3871	***	0.3722	***
	(0.004)		(0.003)	
Log-likelihood	–3782.594		–4783.944	
No. person–years	6042		8134	
No. individuals	708		880	
Wald test p-value	0.0000		0.0000	

Born 1955+, some educ. quals	**Men**		**Women**	
Age (years)	0.0975		1.4042	***
	(0.661)		(0.498)	
Age^2	–0.0059		–0.0602	***
	(0.028)		(0.021)	
$Age^3/100$	0.0144		0.1130	***
	(0.052)		(0.039)	
$Age^4/10000$	–0.0117		–0.0777	***
	(0.036)		(0.027)	
Year: 1991–2000	–0.0778	***	–0.0689	***
	(0.018)		(0.015)	
Intercept	4.9780		–6.7288	
	(5.757)		(4.358)	
sd(slope)	0.0396	***	0.0325	***
	(0.002)		(0.002)	
sd(intercept)	1.4054	***	1.2537	***
	(0.073)		(0.053)	
corr(intercept, slope)	–0.9643	***	–0.9483	***
	(0.005)		(0.005)	

sd(error)	0.3787	***	0.3585	***
	(0.004)		(0.003)	
Log-likelihood	–3722.604		–4625.424	
No. person–years	5868		8043	
No. individuals	985		1120	
Wald test p-value	0.6073		0.0233	

Pre-1955 Birth, A-level(s) +	**Men**		**Women**	
Age (years)	0.6442	***	0.4489	***
	(0.157)		(0.158)	
Age^2	–0.0137	***	–0.0081	**
	(0.004)		(0.004)	
$Age^3/100$	0.0127	***	0.0057	
	(0.004)		(0.004)	
$Age^4/10000$	–0.0044	**	–0.0011	
	(0.002)		(0.002)	
Year: 1991–2000	–0.0609	***	–0.0524	***
	(0.012)		(0.013)	
Intercept	–5.2451	**	–2.7323	
	(2.311)		(2.365)	
sd(slope)	0.0298	***	0.0296	***
	(0.001)		(0.001)	
sd(intercept)	1.6364	***	1.6414	***
	(0.061)		(0.069)	
corr(intercept, slope)	–0.9653	***	–0.9623	***
	(0.003)		(0.004)	
sd(error)	0.3749	***	0.3653	***
	(0.003)		(0.003)	
Log-likelihood	–7181.063		–5539.896	
No. person–years	11939		9473	
No. individuals	1211		990	
Wald test p-value	0.0000		0.0000	

Born 1955+, A-level(s) +	**Men**		**Women**	
Age (years)	0.6500	*	1.9537	***
	(0.351)		(0.368)	
Age^2	–0.0266	*	–0.0813	***
	(0.015)		(0.015)	
$Age^3/100$	0.0479	*	0.1474	***
	(0.027)		(0.029)	
$Age^4/10000$	–0.0318	*	–0.0980	***
	(0.018)		(0.019)	
Year: 1991–2000	–0.0601	***	–0.0755	***
	(0.010)		(0.011)	
Intercept	–0.1450		–11.5135	***
	(3.099)		(3.232)	
sd(slope)	0.0355	***	0.0360	***
	(0.001)		(0.001)	
sd(intercept)	1.2699	***	1.3136	***
	(0.041)		(0.042)	

corr(intercept, slope)	–0.9442	***	–0.9444	***
	(0.004)		(0.004)	
sd(error)	0.3594	***	0.3716	***
	(0.002)		(0.003)	
Log-likelihood	–8618.360		–8721.338	
No. person–years	14744		14081	
No. individuals	2075		2004	
Wald test p-value	0.0001		0.0000	

***: $p < 0.01$. **: $p < 0.05$. *: $p < 0.10$.
Wald test p-value: p-value from Wald test that coefficients on all age variables jointly zero.

Figure C1
Estimated average income -age trajectories, by group, for individuals aged 25+

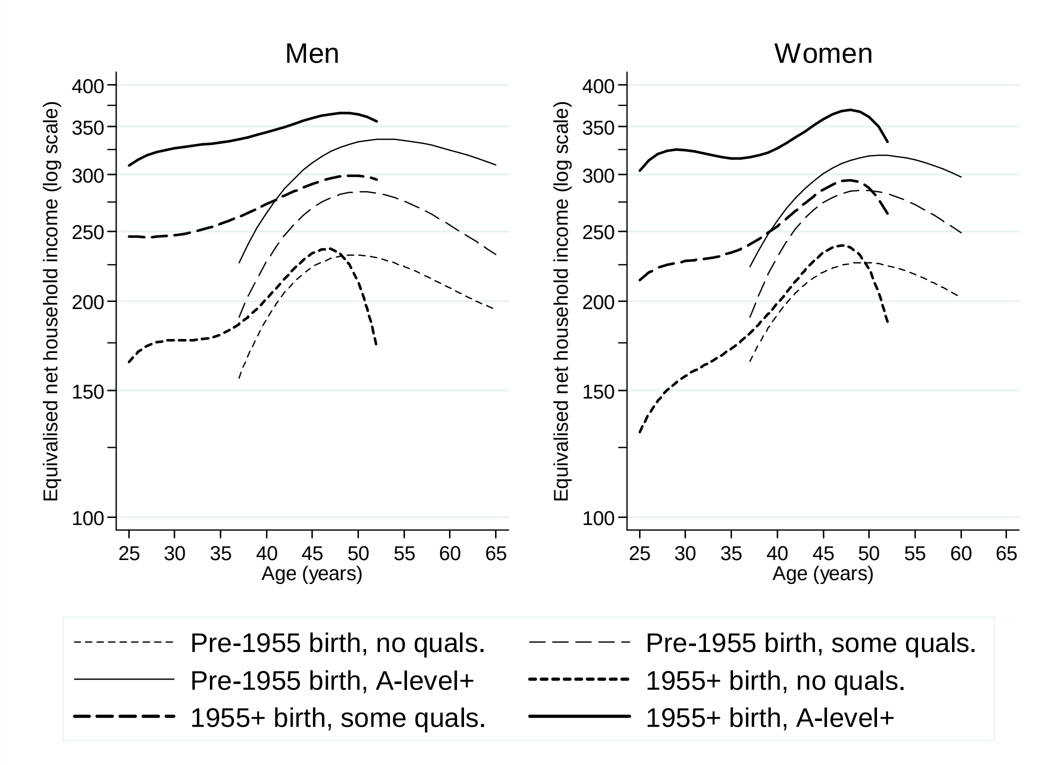
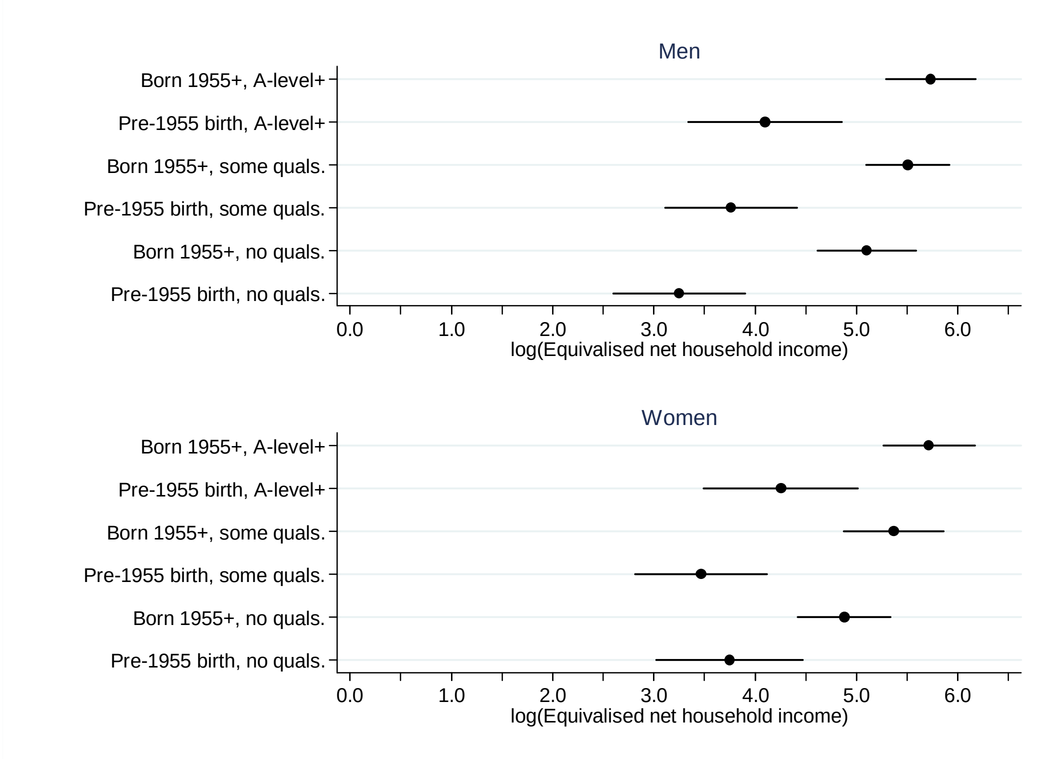


Table C2
Between-group differences in variance component parameters

Educatinal Qualifications.	Men		Women		
	Pre-1955 birth	Born 1955+	Pre-1955 birth	Born 1955+	
sd(intercept): σ_α					
None	1.07	1.09	1.16	0.90	
Some	1.12	1.07	1.12	0.95	
A-level(s) +	1.25	0.97	1.25	1.00	[1.314]
sd(age coefficient): σ_β					
None	0.59	1.01	0.59	0.78	
Some	0.68	1.10	0.69	0.90	
A-level(s) +	0.83	0.98	0.83	1.00	[0.036]
corr(int., age coeff.): $\sigma_{\alpha\beta}$					
None	1.03	1.02	1.03	1.01	
Some	1.02	1.02	1.02	1.00	
A-level(s) +	1.02	1.00	1.02	1.00	[–0.944]
sd(error): σ_v					
None	0.94	1.07	0.92	1.01	
Some	1.04	1.02	1.00	0.97	
A-level(s) +	1.01	0.97	0.98	1.00	[0.372]

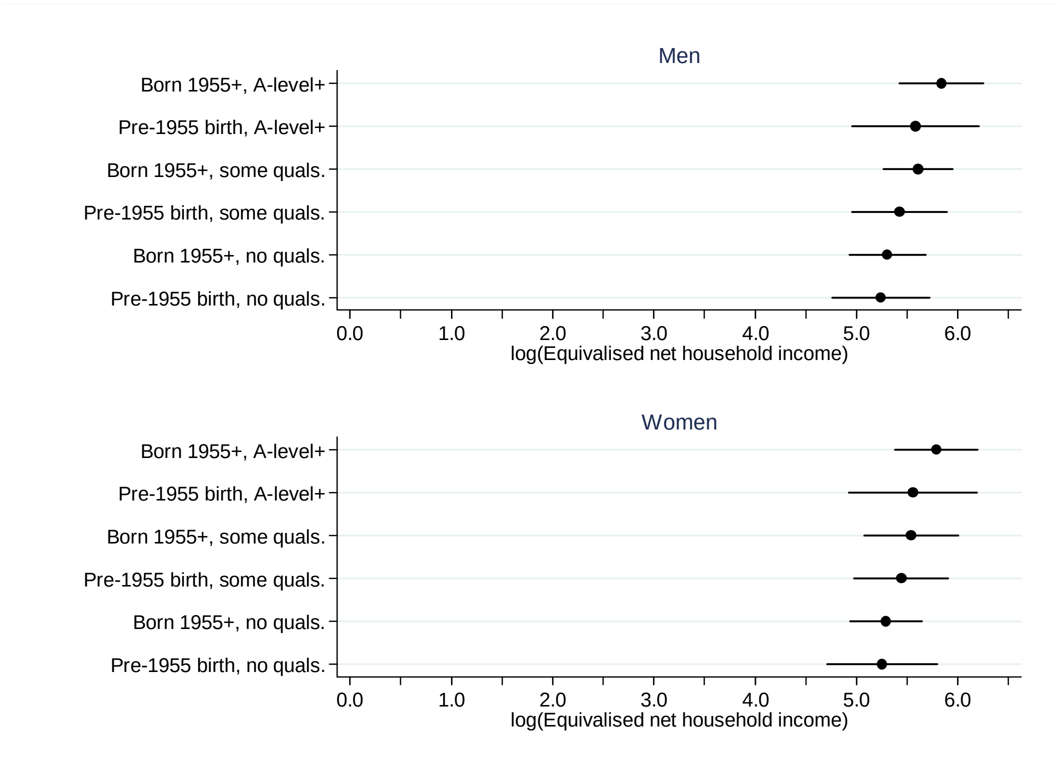
Note: group parameters expressed as a ratio of the parameters for women with A-level(s)+ born 1955+ (shown in brackets).

Figure C2
The distribution of log(equivalized net household income) at age 25, by group



The line for each subgroup shows the interquartile range (distance between the 25th and 75th percentiles). The filled circles show the group medians.

Figure C3
The distribution of log(equivalized net household income) at age 40, by group



The line for each subgroup shows the interquartile range (distance between the 25th and 75th percentiles). The filled circles show the group medians.

Figure C4
Inequality and age: variance of log(equivalized net household income), by group

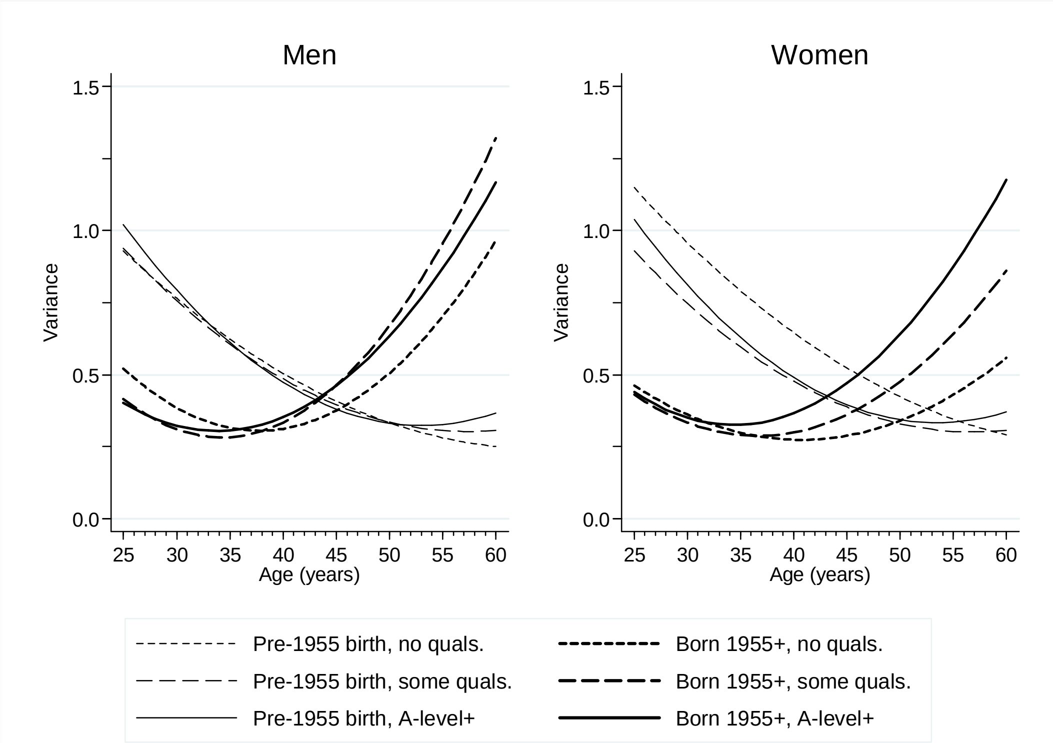


Figure C5
Inequality and age: Gini coefficient of equivalized net household income, by group

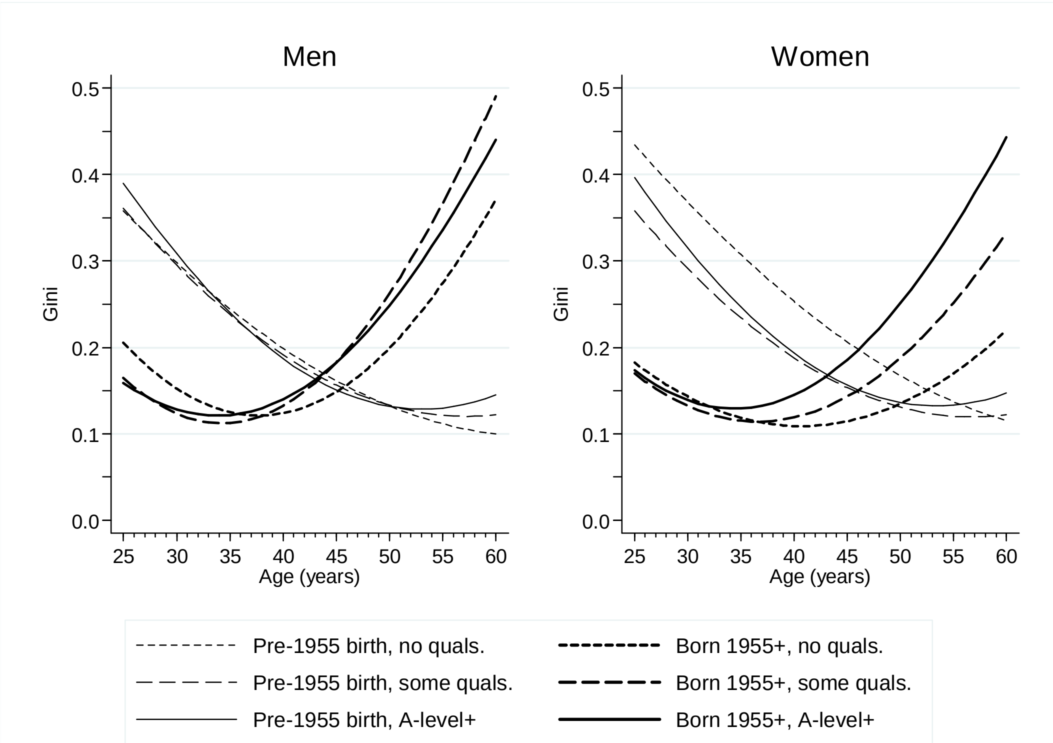


Figure C6
The heterogeneity of individual trajectories: simulated data example

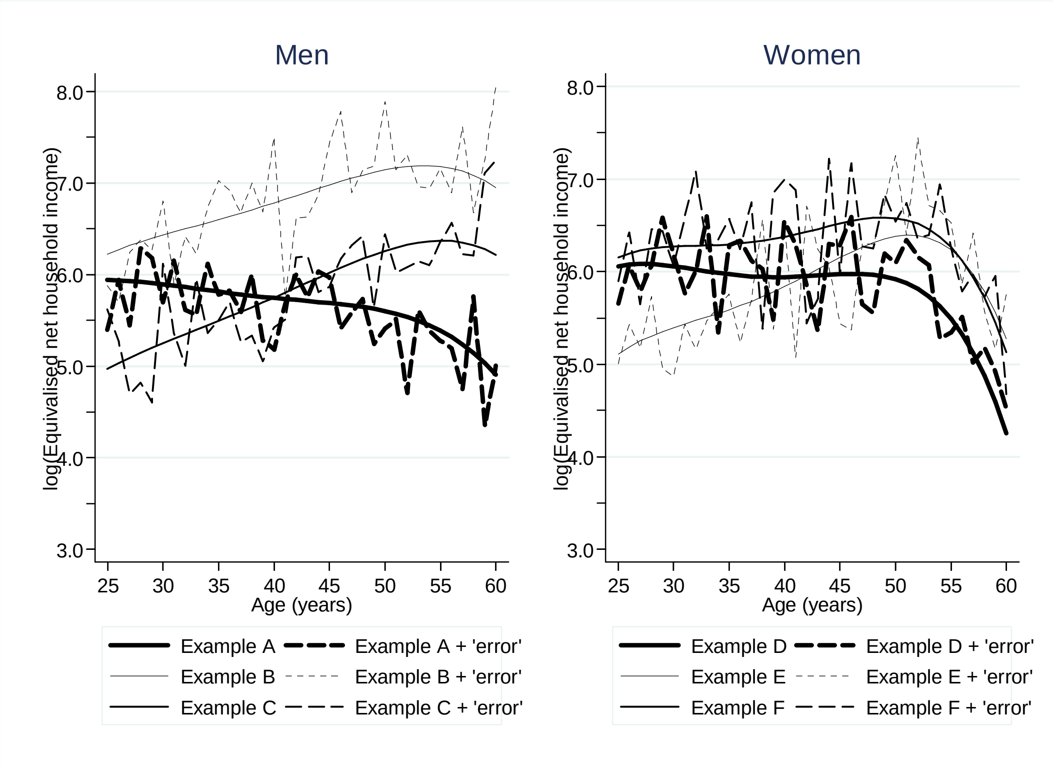
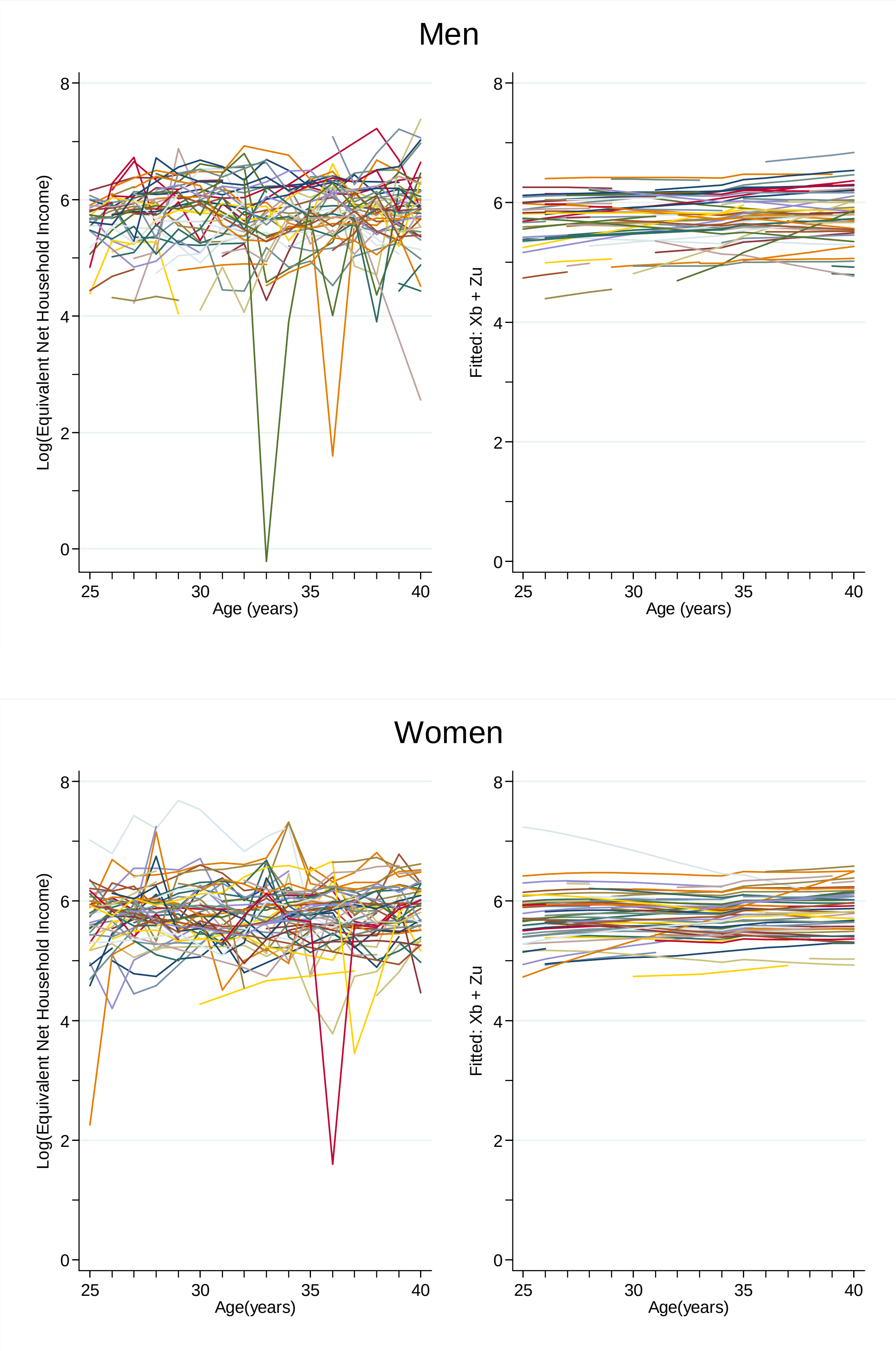


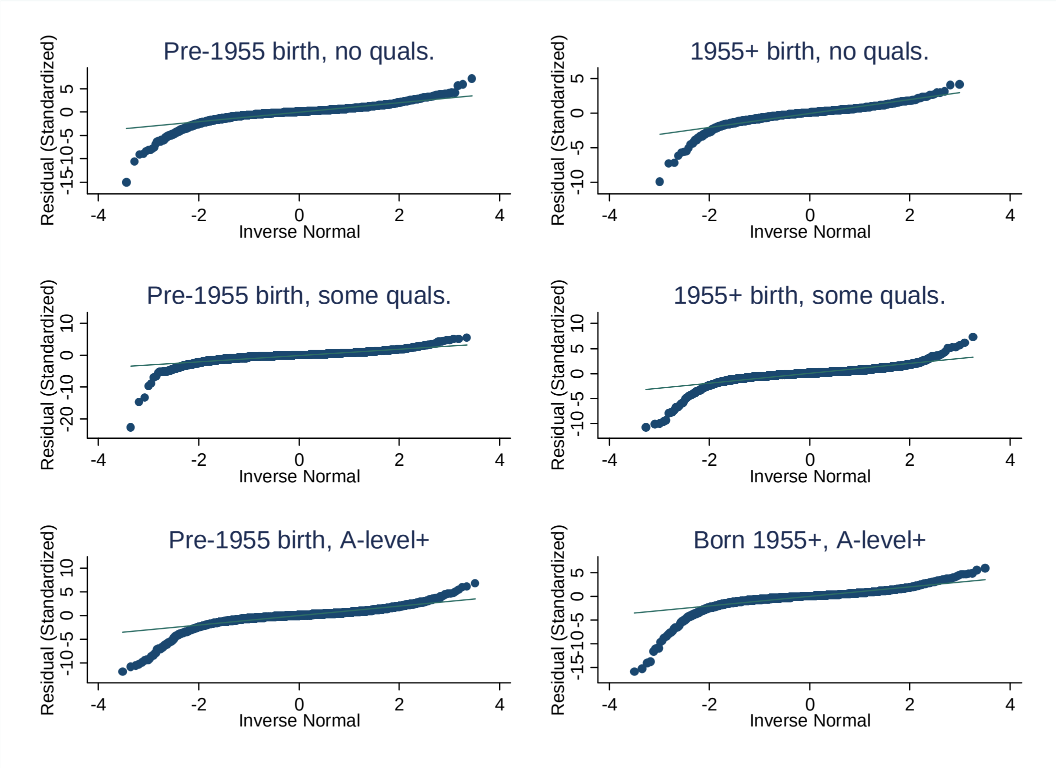
Figure C7

Log(equivalized net household income) trajectories for men and women born in 1966 with A-level(s)+ qualifications: observed (LHS) versus fitted (RHS)

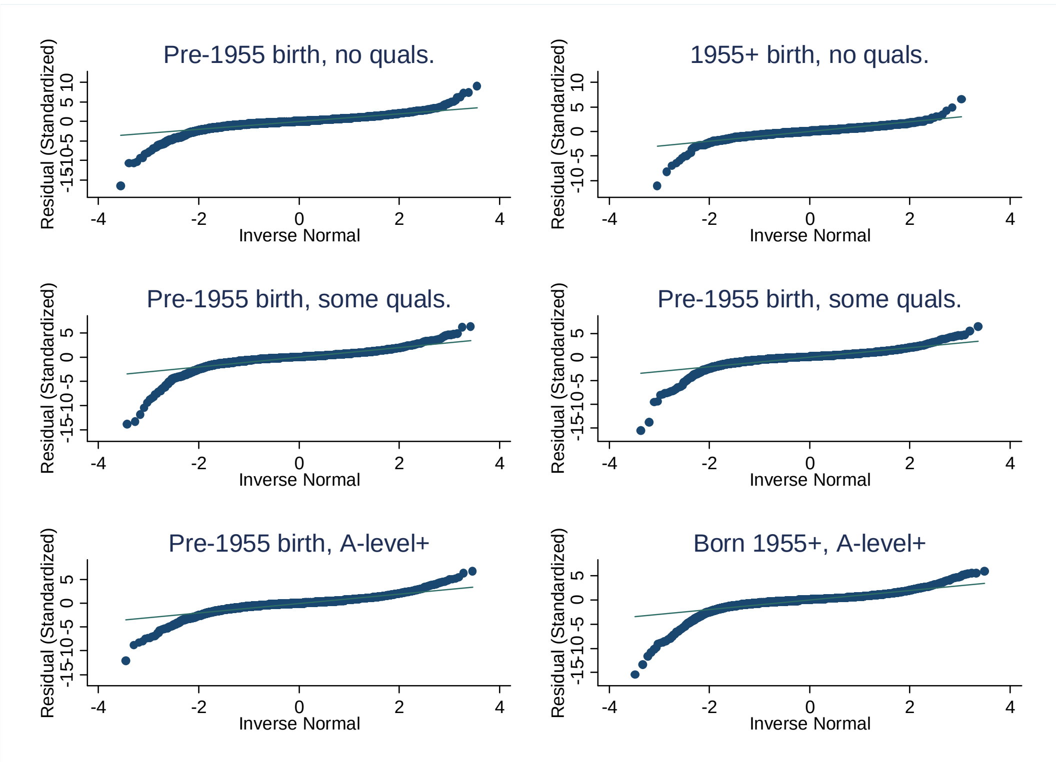


Note: see main text for explanation of the derivation of the graphs.

Figure C8
Quantile plots of standardized residuals against standard normal, by group
Men



Women



Note: Quantile plot graphs quantiles of the distribution of estimated residuals (standardized by their estimated variance) against quantiles of a standard normal distribution. If the plot lies wholly on the 45° ray through the origin, the normal distribution is appropriate. The plots show consistency with normality, except at the extreme tails of the residual distributions.