Modelling Poverty

Technical development of the LINDA model for policy analysis

A report submitted to the Joseph Rowntree Foundation by the National Institute of Economic and Social Research

June 2016

Paolo Lucchino, Jonathan Portes and Justin Van de Ven

Table of Contents

Summary

This report describes the tasks that have been undertaken to adapt an existing model developed at the National Institute for use as part of the JRF programme to develop anti-poverty strategies for the UK. Discussion focusses upon LINDA, a structural dynamic microsimulation model of the UK population cross-section that the National Institute delivered to HM Revenue and Customs and Treasury in 2014.

The version of LINDA that existed prior to this project was well adapted for exploring the influence of a wide range of policy counterfactuals on the employment savings incentives that individuals face. It is also fundamentally designed to explore the medium and long-term effects of policy, which is consistent with the JRF's interest in considering the sustainability of alternative policy strategies.

Nevertheless, LINDA had a number of limitations that were likely to be crucially important for use in the JRF programme, and which have been relaxed by the current project. Chief among these limitations is that the model was structured to report the lifetime circumstances of a single population cross-section, in contrast to the circumstances of the evolving population cross-section. This means that the model could only consider the budgetary implications of policy alternatives that were relevant for a limited sub-group of the evolving population; it could not be used to project the government budget through time. This limited the scope for using the model to gauge the sustainability of policy alternatives from a budgetary perspective. Furthermore, as the model did not capture the entire population, it could provide only a blinkered perspective of the evolution of poverty though time.

One of the principal adaptations implemented under the current project has consequently been to extend the existing model so that it can now be used to project the evolving population cross-section forward through time. This substantial task helps to meet many of the most important limitations of the pre-existing model for use in improving the evidence base of the JRF anti-poverty programme. We have also implemented a range of alternative changes to the model structure, each of which is designed to tailor the model for use in exploring the implications of policy alternatives over a 30 year time horizon, for poverty, the evolving government budget, and the circumstances of individuals. This report describes the model adaptations that have been implemented, including discussion of the crucially important task of parameterisation.

Introduction

In September 2012, the Joseph Rowntree Foundation (JRF) launched a four-year programme to develop anti-poverty strategies for the UK. The aim of this programme is to create a set of costed, evidence-based, strategies for all age groups in all parts of the UK. In late 2014, the JRF commissioned the National Institute of Economic and Social Research (NIESR) to deliver a model capable of projecting the implications of alternative policy counterfactuals for poverty and the Government budget in the UK, over a 25 year horizon to 2035. This report describes the work undertaken by the NIESR to meet this objective.

The starting point for analysis was a pre-existing model – referred to as LINDA (Lifetime INcome Distributional Analysis model) – that the NIESR had developed over a period of more than a decade. The most recent variant of this model had been delivered for use by analysts at the UK Treasury and HM Revenue and Customs, in October 2014.

LINDA was designed to start from data reported for a representative cross-sectional sample of the population of Great Britain reported in 2006, and project the circumstances of each adult in the cross-sectional sample at annual intervals forward and backward through time, building-up a lifetime description for each. The model projected a range of characteristics, including each individual's evolving relationship status, the birth and aging of dependent children, employment, consumption, private income, transfer payments, and evolving benefit unit assets.

An important feature of LINDA is that employment, consumption, and investment decisions can all be projected as solutions to a formal utility maximisation problem. This is in contrast to the standard approach taken in the associated literature, where behaviour is often projected on the basis of reduced form regression equations. The distinction is important, because the utility framework permits LINDA to project behaviour that responds to the incentives that are embedded in policy counterfactuals; incentives that can otherwise be difficult to discern or predict. This feature of the model becomes increasingly important as the time-horizon of the analysis lengthens, due to the feedback effects that decisions taken in the near-term can have on circumstances in the future.

Nevertheless, the variant of LINDA that was delivered to Treasury and Revenue and Customs required a number of important amendments to meet the needs of the JRF as part of its programme to develop costed anti-poverty strategies for the UK. First, and most importantly, the model has been altered to project the circumstances of the evolving population cross-section forward through time, rather than the circumstances of a single reference cross-section. This adaptation was necessary to permit the model to project government budgets and the incidence of relative poverty, which are central issues of concern for the JRF programme.

Secondly, the preference relation has been altered to better capture consumption incentives toward the bottom of the income distribution. This involved implementing a more sophisticated description of "committed expenditure" than had previously existed in the model.

Thirdly, the model has been fully re-parameterised to reflect data for 2011, the most recently available data at the time of writing. In addition, the policy environment represented in the model has been updated, and now includes reforms announced as part of the 2015 Budget statement (programming in progress). These adaptations were necessary to ensure the contemporary relevance of the policy structure, from which counterfactuals of interest to the JRF programme are to be projected.

Fourth, the model has been extended to permit decisions concerning employment, consumption, pension scheme participation, and the timing of pension access to be simulated based on reduced form statistical regressions, rather than the utility maximising structure that is a central feature of LINDA. It is also possible to adopt a mixed approach, whereby some decisions are simulated using the statistical descriptions and others are simulated using solutions to the utility maximisation problem. This model adaptation facilitates sensitivity analysis of results to simulated behaviour. It can also aid in the interpretation of results, by suppressing the incentive effects that complicate interpretation of projections based on utility maximisation. Furthermore, it is possible that some readers will strictly prefer projections that reflect current behaviour, if only because such behaviour is immediately recognisable. We consequently hope that this new feature of the model will enhance its use as a tool for practical policy analysis.

Three additional variations of the model remain to be completed. First, we are currently in the process of introducing additional heterogeneity in relation to educational qualifications, much along the lines of the distinction between graduates and non-graduates that currently exists in the model (see the technical report for a full description of the current model). Secondly, we will be augmenting the model to distinguish between individuals with respect to their prevailing health status (which will vary through time). And finally, we will introduce a reduced form regression into the model, which is designed to disaggregate benefit unit consumption into consumption attributable to individual benefit unit members. These outstanding issues will be written up as they are implemented.

The remainder of this paper describes in turn each of the model adaptations that have been implemented as part of our project for the JRF.

Committed Expenditure

A basic concept in the analysis of poverty and inequality is the view that there exists a basket of goods and services that represent the basic necessities of life, and which individuals therefore exercise little discretion in regards to their purchase. This minimum consumption is referred to as "committed expenditure". Previous incarnations of the model include two components of committed expenditure: housing and childcare costs. One task under the project is to extend this feature of the model to improve reflection of the (financial) constraints that individuals face, particularly those in financial difficulty. This section describes the approach implemented to model committed expenditure, which distinguishes between three classes of consumption: housing, childcare, and other necessities. The first two of these are distinguished separately, as they have a bearing on the benefits that individuals are eligible for. Each consumption class is discussed in a separate section below.

1. Housing

Accommodating housing wealth in the model explicitly is beyond the scope of the current project. Rather, LINDA aggregates net housing equity with all other assets and liabilities held outside of public and private pensions, and Individual Savings Accounts (where included for analysis). Nevertheless, the importance of housing, both as a component of committed expenditure and a determinant of benefits eligibility, makes it useful to distinguish within the model structure.

A stylised approach has been adopted to identify the following housing related features within the model: gross housing equity, gross mortgage debt, realised return on gross housing equity, unrealised return on gross housing equity, mortgage interest costs, and rent. These factors are identified using the following procedure.

1. Identify home-owners
2. If home owner, then go to step 3. Otherwise go to step 10.
3. Identify net housing equity
4. Identify mortgage holders
5. If mortgage holder, then go to step 6. Otherwise go to step 8.
6. Identify mortgage debt
7. Identify annual interest charge on mortgage debt
8. Identify returns to gross housing equity
9. Go to step 11
10. Identify rent paid
11. Evaluate council tax payment

1.1 Incidence of home-ownership and mortgage holders

Logit regression models are used to describe the incidence of home ownership and mortgage holding in the model, with reference to the simulated year, and each benefit unit's age, relationship status, and aggregate non-pension wealth. The logit regressions were estimated on data reported by the WAS for 2011, using the same variable definitions as considered for the reference population cross-section (see Section 4). Associated regression statistics are reported in Appendix A.1.

Figure 1.1 displays a useful summary of the incidence rates upon which imputation of home-ownership in the model is based. This figure indicates that home-ownership was close to universal amongst individuals with some reported net assets and aged 40-80 in 2011. Nevertheless, 40 per cent of all adults are reported to be living in benefit units with net non-pension wealth worth less than £5,000 in 2011, 95 per cent of whom do not own their home. As Figure 1.1 suggests, the majority of these individuals – 78 per cent – are single adults.

Figure 1.2 reports similar statistic as in figure 1.1, but for the incidence of mortgage holding amongst home owners. This figure reveals that mortgage holding is almost universal amongst home owners under age 35, and declines gradually as age rises. Importantly, however, and unlike the incidence of home ownership, the likelihood of holding a mortgage does not vary very substantially with net non-pension wealth. Nevertheless, non-pension wealth is included as an explanatory variable in the logit regressions for mortgage holding to accommodate future flexibility in this relationship.

1.2 Net housing equity

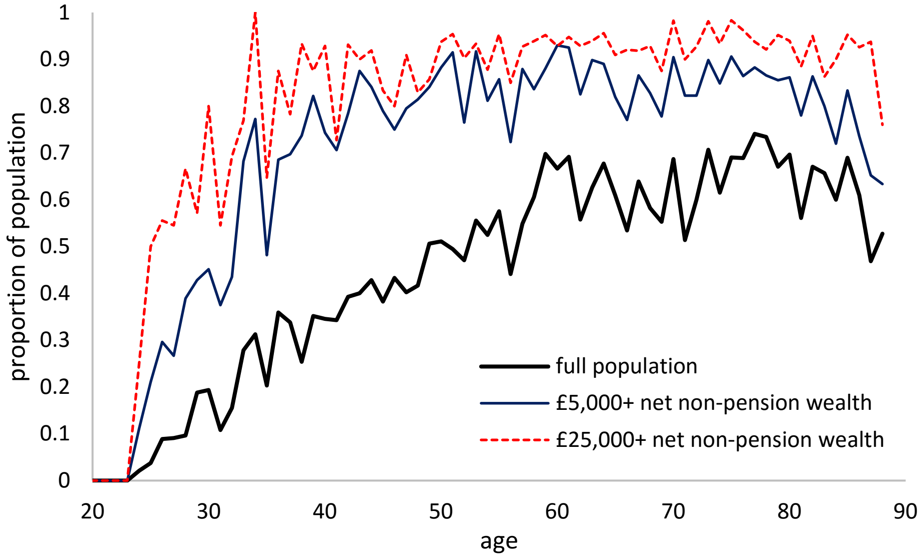
Net housing equity is defined as a ratio of net non-pension wealth. Data describing the weight of housing in the net asset portfolio of home owners, by age and relationship status are displayed in Figure 1.3. This figure indicates that housing comprises around 80 per cent of all net non-pension wealth for both singles and couples throughout the life course. Given the stylised approach taken to model housing, and the compelling nature of this relationship, net housing equity in the model is imputed as a simple age specific ratio, reported in Appendix A.1 for completeness.

1.3 Mortgage debt

Mortgage debt, for home owners who are also identified as mortgage holders in the model, is imputed on the basis of a Tobit regression equation for the ratio of gross mortgage debt to net housing equity. The regression model describes the ratio of mortgage debt to housing equity as a function of age, time, relationship status, and net non-pension wealth, and was estimated on the same WAS data as considered for other aspects of housing. Regression statistics are reported in Appendix A.1.

Figure 1.4 provides age specific averages of the ratio of mortgage debt to net housing equity. This figure indicates that the value of mortgage debt relative to net housing equity among mortgage holders peaks in the early 20's, and falls away fairly sharply into the 40's, before training away into retirement. Although similar profiles are reported for singles and couples, the regression model adopted to simulate mortgage debt includes a full set of relationship-specific interaction effects.

Figure 1.1: Incidence of home ownership in Great Britain in 2011 by age, net non-pension wealth and relationship status



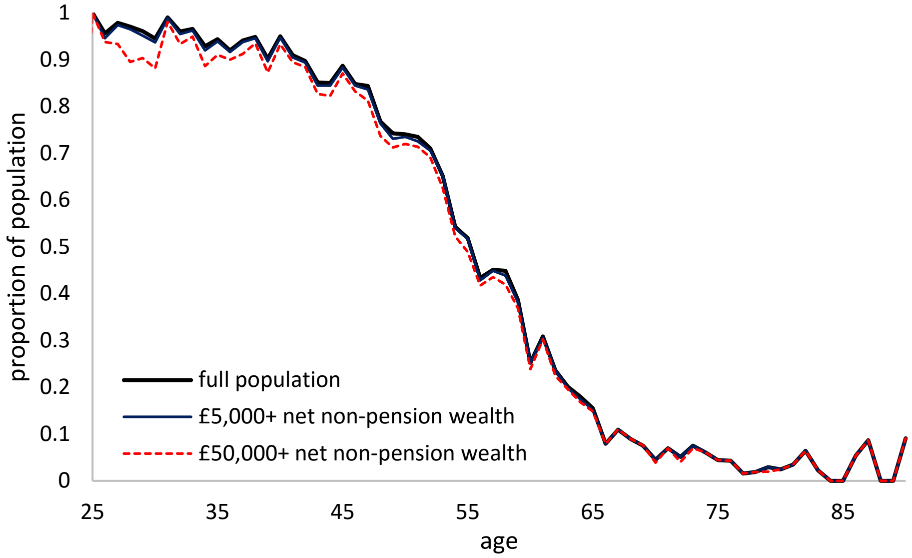
Panel A: singles

Panel B: couples

Source: Authors' calculations using data reported for 2011 by wave 3 of the Wealth and Assets Survey
Notes: Non-pension wealth includes all wealth held outside of private and occupational pensions.

Figure 1.2: Incidence of mortgage holding amongst home owners in Great Britain in 2011 by age, net non-pension wealth and relationship status

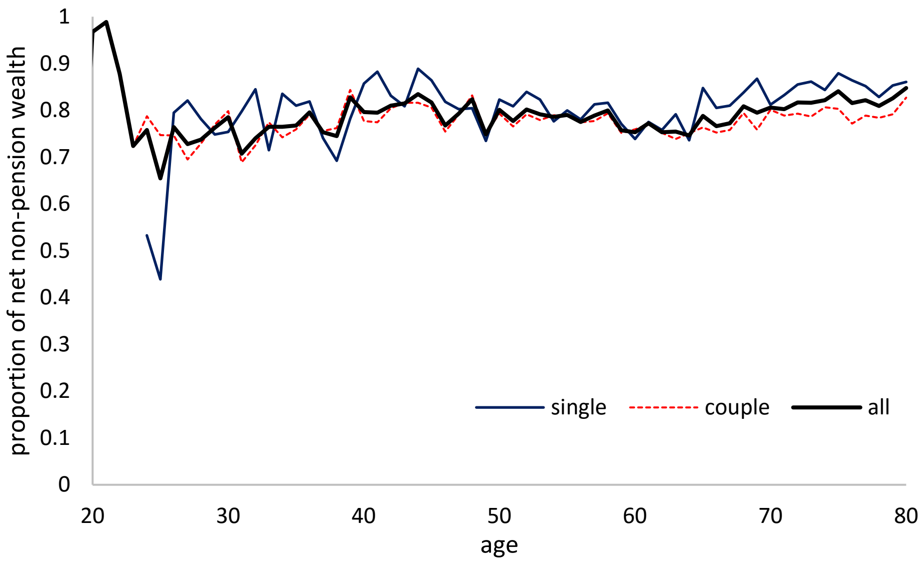
Panel A: singles



Panel B: couples

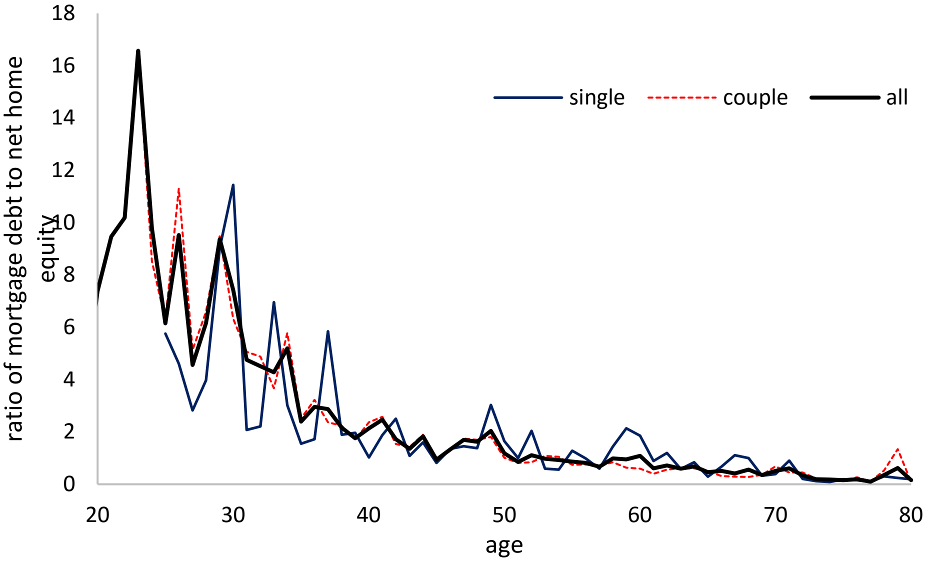
Source: Authors' calculations using data reported for 2011 by wave 3 of the Wealth and Assets Survey
Notes: Non-pension wealth includes all wealth held outside of private and occupational pensions.

Figure 1.3: Proportion of net non-pension wealth invested in owner occupied housing



Source: Authors' calculations using data reported for 2011 by wave 3 of the Wealth and Assets Survey
Notes: Non-pension wealth includes all wealth held outside of private and occupational pensions.

Figure 1.4: Ratio of mortgage debt to net equity in owner occupied housing for the population subsample of mortgage holders by age and relationship status



Source: Authors' calculations using data reported for 2011 by wave 3 of the Wealth and Assets Survey
Notes: Non-pension wealth includes all wealth held outside of private and occupational pensions.

1.4 Housing costs and returns

Housing has an impact on family budgets in three ways in the model: returns to gross housing wealth, interest charges on outstanding mortgages, and rent payable. Each of these is described in turn.

1.4.1 Returns to gross housing wealth

Returns to gross housing wealth were calculated from the ONS mix adjusted house price index, and discounted to real terms by the National Accounts final consumption deflator. Rates of return assumed by the model are reported in Appendix A.1. The principal calibration assumes that the return to housing wealth in forward projections is equal to the mean return observed between 1970 and 2010. The negative calibration reduces the return in forward projections by half of one standard deviation of the observed time-series.

1.4.2 Interest on outstanding mortgage debt

Interest on outstanding mortgages has a direct impact on our measure of committed expenditure, and can affect the benefits to which a benefit unit is eligible (discussed in Section 5). Rather than identifying a market rate of interest on mortgage debt (available, for example from the Bank of England's website), the objectives of our analysis have motivated the decision to base mortgage interest on the maximum interest rate eligible for subsidy through Income Support; 3.85% per annum nominal – which we take to be 1.85% real (assuming that target inflation is achieved). The same interest charge is assumed for both the principal and negative calibrations.

1.4.3 Rent

Rents in 2011 are set equal to the Local Housing Allowance rates averaged over all Local Authorities, reported for June 2011 by the Valuation Office Agency. Rents are assumed to grow in line with wages. The rental rates are defined in terms of numbers of bedrooms, which are simulated in the model as defined for the Housing Benefit; see Section 5.2. The only exception is in relation to single adults aged 30 and under without children, who are assumed to share their accommodation, and who consequently incur the lower rental charge associated with shared accommodation by the LHA.

Two sets of rents are supplied to the model: Local Housing Allowance (LHA) rates for low income people, and "market rents" for high income people. Both sets of rents are defined in terms of numbers of bedrooms, which are simulated in the model as defined for the Housing Benefit; see Section 5.2. The only exception is in relation to single adults aged 30 and under without children, who are assumed to share their accommodation, and who consequently incur the lower rental charges.

Any benefit unit with equivalised income under a lower threshold is assumed to be subject to the LHA rates, and any benefit unit with equivalised income above an upper threshold is assumed to be subject to market rents. Between these two thresholds, rents are assumed to increase from the LHA up to the market rents in proportion to income. The lower threshold was set equal to 60% of median gross full-time earnings in 2011, and the upper threshold to 120% of median full-time earnings.

The starting point for LHA rents were the rental averages reported over all Local Authorities for June 2011 by the Valuation Office Agency. Market rates were assumed to be twice the assumed LHA rates. These rental charges were then adjusted as part of the calibration process to match poverty rates reported generated the model to poverty rates reported in the *Households Below Average Income* publication issued by the Department for Work and Pensions.

Both the principal and negative calibrations assume that rental charges grow in line with wages.

2. Childcare

Childcare costs are modelled on the aggregate of two variables reported by the LCFS. Variable cc4121 reports expenditure on crèche, nurseries, playschool, and playgroups. Variable cc4122 reports expenditure on babysitters, child minders, childcare, and nannies. A linear regression model was estimated for the average aggregate childcare costs implied by these two variables, distinguished by child age, parental labour supply, and survey year. Figure 2.1 displays selected results from this analysis, and the full set of statistics is reported in Appendix A.2. Figure 2.1 indicates a number of key features of the childcare costs reported by the LCFS. First, childcare costs are substantively higher for children under 5 years of age, than for older children. Furthermore, childcare costs tend to increase with parental employment status, with the highest average costs reported by benefit units in which all adults are employed full-time. Finally, the real value of reported childcare costs have tended to rise in the two decades from 1990, by an annualised rate of 2.2 percentage points for benefit units with children under 5 years of age and all adults full-time employed.

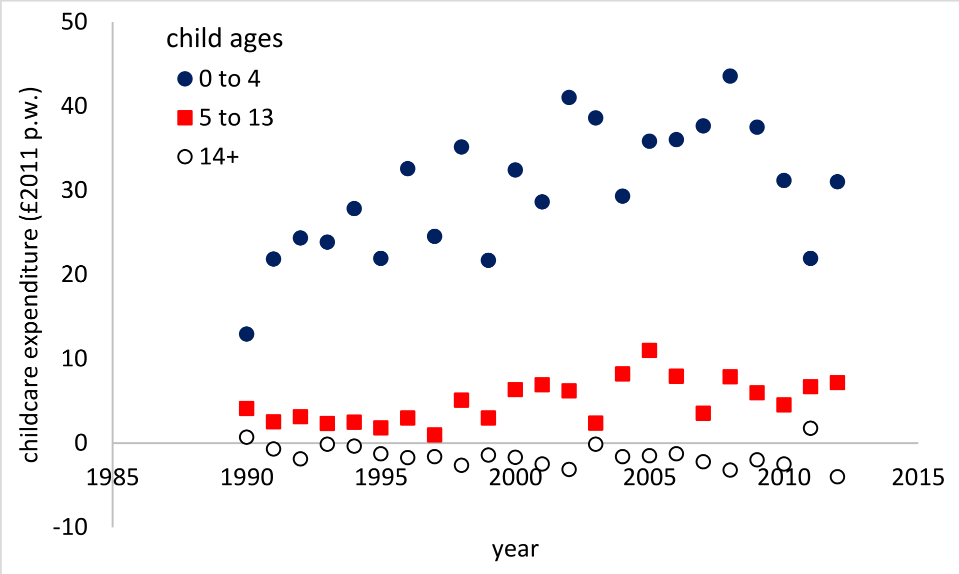
The model has been structured to reflect the statistics reported here. Childcare costs are assumed to be incurred only by benefit units with children under age 14, and in which all adults supply some labour. Childcare costs are assumed to vary between children aged 0-4 and those aged 5-13. Childcare costs are also assumed to vary between benefit units in which all adults are full-time employed and those working less than full-time. Given the volatility of costs displayed in Figure 2.1, the childcare costs incurred in the reference cross-section have been set equal to the averages reported over the three years 2010 – 2012.

Hence, the model assumes that childcare costs equal to £28.07 per week for each child aged 0-4, and £6.17 per week for each child aged 5-13, are incurred in 2011 by full-time employed benefit units. These costs fall to £16.08 per week for each child aged 0-4, and £0.50 per week for each child aged 5-13, for benefit units with at least one adult less than fully employed. Childcare costs are assumed to grow at a real rate of 2.2 per cent per annum in both the principal and negative calibration.

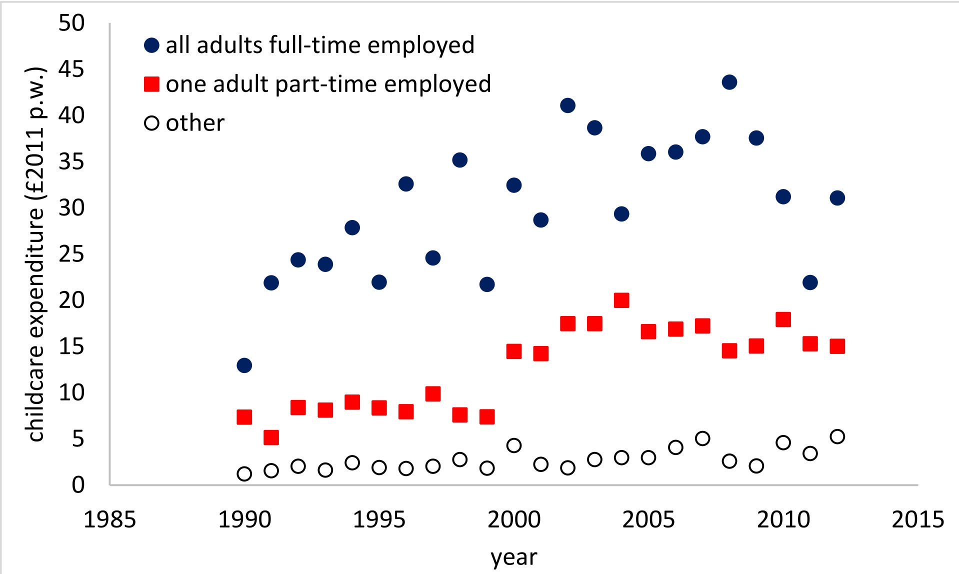
The measure of childcare expenditure derived by aggregating these two variables is not ideal for at least two reasons. First, the variables referred to here do not distinguish between expenditures that are eligible for relief through the transfer system (see the Working Tax Credit described in section 4.6). Second, the LCFS reports “out of pocket” expenditure, which is likely to be net of childcare payments from employers, salary sacrifice schemes, childcare payments or grants from a government scheme (e.g. to assist entry into work), or childcare costs paid for by an educational or local authority (e.g. free early learning or nursery education).

Relative to practice, the model is consequently likely to (i) understate full expenditure on childcare costs (to the extent that these costs are covered by schemes like salary sacrificing), (ii) understate full relief on childcare costs (to the extent that salary sacrificing provides a tax advantage), and (iii) understate gross earnings. In aggregate, these distortions are likely to off-set one another – albeit imperfectly – with respect to the budget that is available to finance discretionary expenditure.

Figure 2.1: Childcare costs per dependent child by age band, parental employment status and survey year



Panel A: by child age, all adults full-time employed



Panel B: by parental employment, all children aged under 5

Source: Authors' calculations using data reported by 1971 to 2012 waves of the Living Costs and Food Survey
Notes: Childcare expenditure set equal to LCFS variables cc4121 + cc4122
Per child costs calculated using weighted linear regression over benefit units
Expenditure discounted to 2011 prices using the National Accounts final consumption deflator

3. Other basic necessities

Consumption on "other basic necessities" in the model is parameterised based on data reported for the 5% of benefit units with the lowest aggregate consumption. Data for calibrating this aspect of the model were evaluated from the LCFS (and its predecessors) between 1971 and 2012.

The LCFSs were used to identify age, year, equivalised total consumption (in 2011 prices), and equivalised consumption net of childcare and rental charges for the full reported sample of benefit units. The equivalence scale used for analysis is the revised OECD scale, which is also assumed for the model's preference relation (see Section 4.1). The 5% of benefit units with the lowest equivalised total consumption, distinguished by year and across 7 age bands (18-24; 25-34; 35-44;…; 65-74; 75+) were selected from the full sample population. Averages for equivalised consumption net of childcare and rental charges, by age band and year, were then evaluated. These averages of equivalised consumption are loaded into the model to define consumption on "other basic necessities".

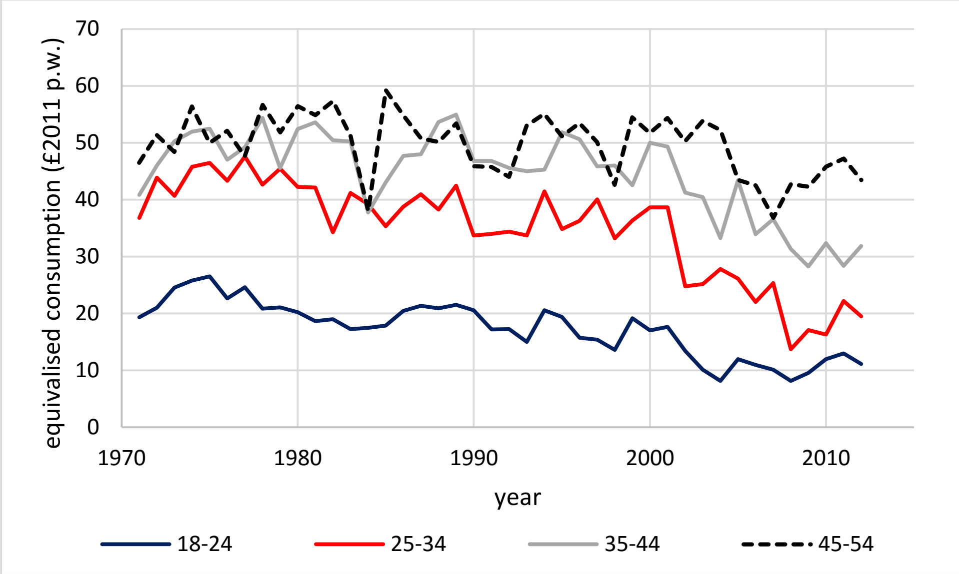
The approach taken to parameterise "other basic necessities" in the model is loosely designed to reflect the minimum expenditure required to participate in society, consistent with the target definition of poverty used for the JRF anti-poverty programme. These statistics are reported in Figure 3.1.

Figure 3.1 reveals that substantive differences exist between the time profiles of equivalised consumption after childcare and rental charges of the age groups considered for analysis. Whereas equivalised consumption has tended to decline amongst younger benefit units over the sample period, it has risen fairly consistently amongst older benefit units. The most conspicuous examples of these trends are at the extremes of the age distribution, where (real) equivalised consumption is reported to have halved in the four decades between 1970 and 2010 for 18-24 year olds, and to have tripled amongst those aged 75 and over.

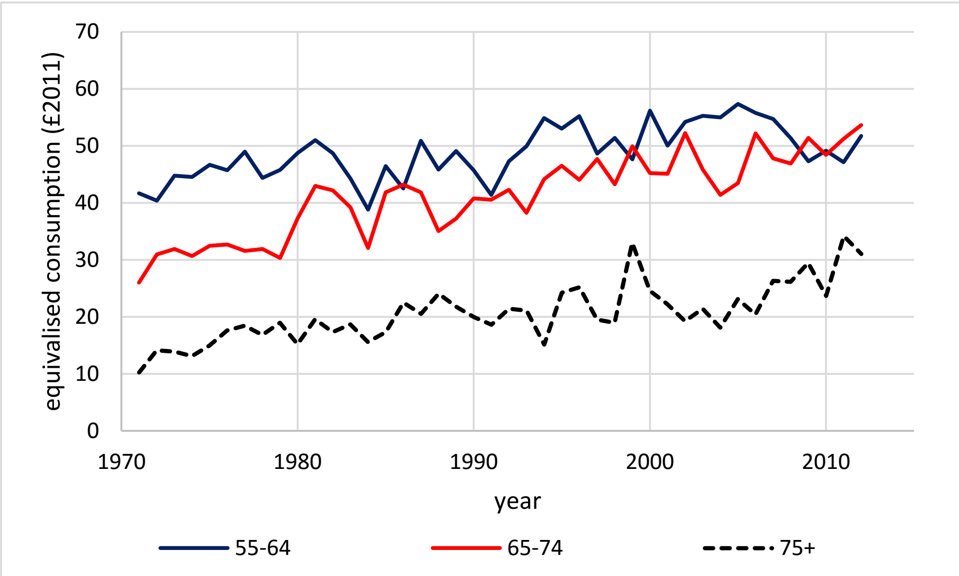
These trends are predominantly driven by shifts in disposable income and rental charges. Associated data are reported in Figure 3.2. Panel A of Figure 3.2 indicates that disposable income amongst the 5% of benefit units with the lowest equivalised total consumption increased strongly for individuals aged 75 and over, from £20 per week in the early 1970's to £140 per week in the early 2010's. In contrast, the incomes of 18-24 year olds at the bottom of the consumption distribution has remained broadly unchanged in real terms, equal to around £30 per week throughout the 40 year sample period.

Rental charges reported for the limited sample population considered here describe very similar profiles for both 18-24 year olds and benefit units with reference adults 75 and over, rising from a very low figure in the early 1970's to around £10 per week in the early 2010's. This rise in rental charges is responsible for the falling equivalised consumption net of rental charges reported for 18-24 year olds (in context of flat disposable income). In the case of individuals aged 75 and over, however, the rise in rental charges represents only a fraction of the coincident increase in disposable income, leaving room for non-rental consumption to rise appreciably over the period as observed.

Figure 3.1: Average equivalised consumption net of childcare and rental costs of 5 per cent of benefit units with lowest equivalised total consumption, by age and year



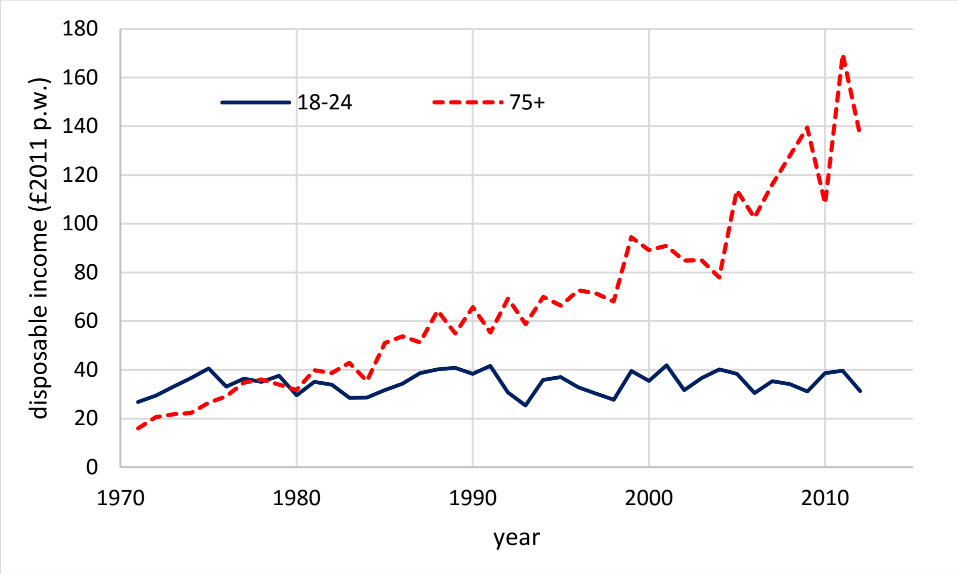
Panel A: ages 18 to 54



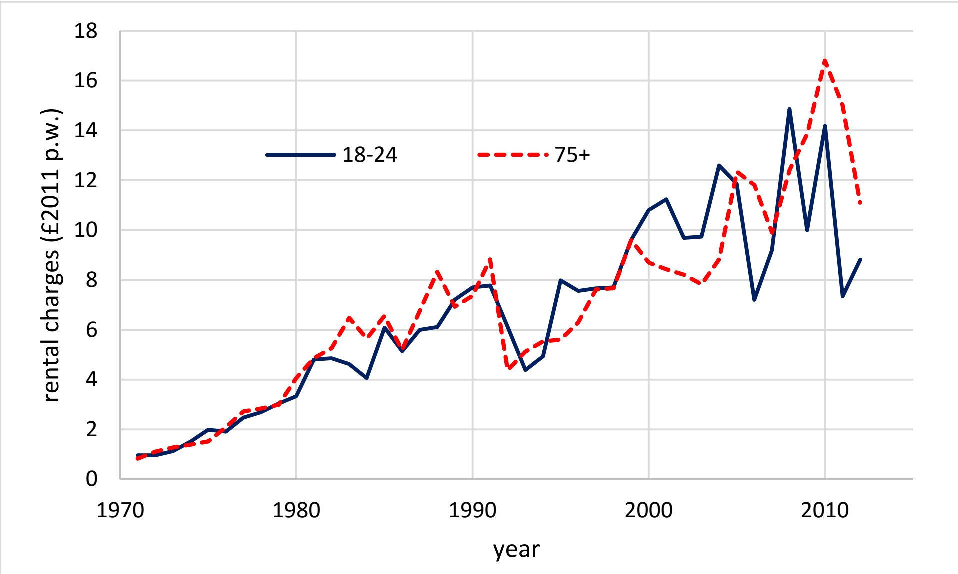
Panel B: ages 55 and over

Source: Authors' calculations using data reported by 1971 to 2012 waves of the Living Costs and Food Survey
Notes: Consumption reported net of childcare and rental costs
Consumption averaged over benefit units
Consumption equivalised using the revised OECD equivalence scale
Consumption discounted to 2011 prices using the National Accounts final consumption deflator

Figure 3.2: Average disposable income and rental charges of 5 per cent of benefit units with lowest equivalised total consumption, by age band and year



Panel A: disposable income



Panel B: rental charges

Source: Authors' calculations using data reported by 1971 to 2012 waves of the Living Costs and Food Survey
Notes: All figures discounted to 2011 prices using the National Accounts final consumption deflator

Updating Model Parameters

The parameters of LINDA have been updated to reflect circumstances from 2011, based on the most recently available data at the time of writing. Calibration of the model parameters was divided into a number of discrete tasks. First, a model specification was selected which describes the characteristics that are accommodated by default in the model, and the ways that those characteristics are assumed to vary through time. This specification was selected to ensure sufficiently fast run-times, off-set by the objective to capture as much realism as possible. Data for a base population cross-section were then compiled, and model parameters were adjusted to match the model against survey data. Finally, the set of pre-programmed policy structures was extended to account for reforms announced in the 2015 Budget Statement (the most recently available information at the time of writing).

Two sets of model parameters were identified: a principal set reflecting a balanced appraisal of the unknown future, and a negative counterfactual. Relative to the principal model calibration, the negative calibration is designed to generate a negative bias for the government budget in forward projections, and a positive bias for poverty rates. Key differences between the two model calibrations are summarised in Box 1. This section describes each step undertaken during the model re-parameterisation, including reference to the data sources used. A detailed analysis of projected trends implied by the principal calibration is provided in Appendix G.

Box 1: differences between model calibrations

Wage growth:	lower in negative calibration, especially amongst lower educated
Wage offers:	less prevalent in negative calibration, especially amongst lower educated
Rates of return:	reduced in negative calibration
Longevity:	increased in negative calibration
Costs inflation:	higher in negative calibration
Benefits growth:	lower in negative calibration

The model specification considered for calibrating the model is described in Section 4. Section 5 describes derivation of the data for the reference population cross-section, from which model projections are made. Section 6 describes the method used to calibrate unobserved model parameters, and Section 7 describes specification of the policy environment to reflect reforms announced as part of the 2015 Budget Statement.

4. Model specification

The model is designed to project through time the evolving circumstances of a sample of reference adults, with the sample specified to reflect the evolving UK population. The model allows for a great deal of flexibility concerning the characteristics that are included for analysis, and in the ways that selected characteristics are projected through time. Full details of the modelling alternatives that are available are described in the companion technical report. In this section, we describe the model features that were enabled to update unobserved parameters, and which characterise the default options assumed by the model.

Endogenous decisions for parameterising the model were limited to labour supply, consumption, participation in private pensions, and the timing of pension access. These decisions were assumed to depend on the age, birth year, education status, relationship status, number and age of dependent children, and survival of the reference adult of each benefit unit, the wage potential of both the reference adult and their spouse (if one exists), the non-pension wealth and pension wealth held by the benefit unit, and whether pension wealth had been previously accessed. These terms are summarised in Box 2. Each of the principal components of this model specification, including the respective parameterisation, is described in turn below.

Box 2: Default model options adopted for calibration

Endogenous Decisions

- consumption
- labour supply
- pension participation
- pension take-up

Benefit unit characteristics

- age
- birth year
- education status
- relationship status
- dependent children
- wage potential
- non-pension wealth
- pension wealth
- pension draw-down
- survival

4.1 Preferences

Preferences are assumed to take a nested Constant Elasticity of Substitution (CES) specification described by:

$$U_{i,b,a} = \frac{1}{1-\gamma}\left\{u\left(\frac{c_{i,a}}{\theta_{i,a}}, l_{i,a}\right)^{1-\gamma} + E_a\left\{\sum_{j=a+1}^{A}\delta^{j-a}\left[\phi^b_{j-a,a}u\left(\frac{c_{i,j}}{\theta_{i,j}}, l_{i,j}\right)^{1-\gamma} + \left(1-\phi^b_{j-a,a}\right)\zeta_1\left(\zeta_0 + w^+_{i,j}\right)^{1-\gamma}\right]\right\}\right\}$$

$$u\left(\frac{c_{i,a}}{\theta_{i,a}}, l_{i,a}\right) = \left[\left(\frac{c_{i,a}}{\theta_{i,a}}\right)^{1-1/\varepsilon} + \alpha^{1/\varepsilon} l_{i,a}^{1-1/\varepsilon}\right]^{\frac{1}{1-1/\varepsilon}}$$

where $U_{i,b,a}$ is the expected lifetime utility of benefit unit *i*, of birth year *b*, at age *a*, *c* denotes period specific discretionary non-durable consumption, θ is the revised OECD adult equivalence scale (which varies by the number and age of benefit unit members), *l* is period specific labour supply, E_a is the expectations operator at age *a*, *A* is the upper bound assumed for life expectancy, $\phi^b_{j,a}$ is the probability of surviving an additional *j* years, given survival to age *a* for birth cohort *b*, and w^+ denotes non-pension wealth when this is positive and zero otherwise.

The modified OECD scale assigns a value of 1.0 to the benefit unit reference person, 0.5 to each additional benefit unit member over age 13, and 0.3 to each child aged 13 and under. This scale is currently the standard for adjusting incomes in European Union countries. *A* and ϕ are defined with respect to ONS lifetables as described in Section 4.8. The remaining terms described in the preference relation are unobserved preference parameters that have been calibrated to match the model to survey data as described in Section 6.

4.2 Employment income

A benefit unit's employment income *g* depends upon its latent wage, *h*, whether it is considered to receive a job offer, λ^o, whether it had previously elected to access its pension wealth (see Section 4.4), λ^{ret}, and its labour supply decision, λ^{emp}:

$$g_{i,a} = \max\left(h_{i,a}, h_{a,t}^{\min}\right)\lambda_{i,a}^{o}\lambda_{i,a}^{ret}\lambda_{i,a}^{emp}$$

where h^{min} denotes the minimum wage applicable to time *t*. Each of these components of employment income are described in turn.

4.2.1 Latent wages

A benefit unit's latent wage is the employment income that it would earn if it received a job offer, chose to be fully employed, and had not elected to access its pension wealth at a preceding age; $\lambda^o = \lambda^{emp} = \lambda^{ret} = 1$. Latent wages are modelled at the benefit unit level, and in most periods are assumed to evolve following a regression toward the mean specification that includes an experience effect:

$$\log\left(\frac{h_{i,a}}{m_{i,a}}\right) = \log\left(\frac{h_{i,a}}{m_{i,a}}\right) + \kappa_{i,a-1}\frac{\left(1-l_{i,a-1}\right)}{\left(1-l_{ft}\right)} + \omega_{i,a-1}$$

where κ and *m* are model parameters, and ω denotes random shocks drawn from a normal distribution. The parameters of the above equation were all adjusted internally to the model structure, as described in Section 6.

The only exceptions to the above equation are when the reference adult is identified as changing their education status or labour class (self-employed / employee), in which case new random draws were taken from log-normal distributions.

4.2.2 National minimum wage

The national minimum wage rate was set to £4.98 per hour in 2011 for individuals aged 18 to 20, and £6.08 for those aged 21 or over. These rates grow at rates that match official increases to 2017, and beyond that period at the assumed rates of wage growth.

4.2.3 Employment offers

The model includes an allowance for (involuntary) unemployment by allowing for job offers. Only employees are affected by job offers (self-employed are unaffected). Job offers are assumed to be randomly allocated, governed by age, year, and education specific transition probabilities. If a job offer in any given year is not received, then the "principal earner" of the benefit unit (defined below) is assumed to be unemployed. In the case of couples, this approach permits some labour supply from the spouse. The transition probabilities assumed for the model are based on statistics calculated from the annual Labour Force Survey from 1975 to 1991, and on the spring quarter Labour

Force Survey from 1992 to 2014.[1] These probabilities are further distinguished by graduate status from 1977, when education status was first reported by the Labour Force Survey.

Beyond the observed time-series, the principal calibration assumes that age-specific unemployment rates through time will take the average of the rates described by the observed time-series. In contrast, the negative calibration sets age and education specific rates of unemployment in forward projections to the respective means, plus half a standard deviation of the observed time-series for graduates, and 1 standard deviation for non-graduates. As a guide, the average unemployment rate for graduates across all ages between 18 and 64, is 5.2%, and the standard deviation is 2.6%. These compare with an average unemployment rate of 7.6% for non-graduates, and a standard deviation of 2.7%. Furthermore, assuming a normal independent and identical distribution, there is approximately a 30% probability that unemployment will be more than half a standard deviation above the mean, and a 15% probability that it will be more than one standard deviation above the mean.

4.2.4 Retirees

Retirees are defined as benefit units that have previously chosen to access their private pension wealth (see Section 4.5.2). Pension take-up is assumed to cause a disruption to employment that incurs a wage penalty in all future ages. This feature of the model is designed to limit the ability of individuals to fluctuate in and out of employment late in the working lifetime, resulting in a retirement decision that exhibits greater persistence than would otherwise be the case. The parameters governing this aspect of the model were also adjusted as part of the model calibration discussed in Section 6.

4.2.5 Labour supply

Employment is assumed to increase labour income and reduce the leisure time of the benefit unit, resulting in a trade-off between consumption and leisure that characterises the endogenous labour supply decision.

Each adult in a benefit unit is assumed to choose from a discrete number of employment alternatives, n. Selecting the lowest alternative returns zero employment income and incurs no leisure penalty. Selecting the highest alternative returns the wage potential associated with the respective adult, and incurs the highest leisure penalty. Labour supply alternatives between these two extremes are assumed to alter employment income proportionally, so that selecting option x from n alternatives returns an employment factor $\lambda^{emp} = (x - 1) / (n - 1)$. Similarly, intermediate labour supply alternatives are assumed to have a proportional impact on the leisure penalty, after a fixed time cost of any employment is accounted for. The fixed time cost of any employment reduces the effective hourly wage rate of work for lower labour supply alternatives, and can be taken to represent the time spent preparing for work, and travelling to and from the workplace.

Each adult is assumed to have 98 hours per week of time to allocate between leisure and work. Any employment is assumed to incur a fixed leisure cost of 10 hours per week (2x5), and full time employment is assumed to incur a further cost of 40 hours per week (8x5). The base calibration for the model assumes that each adult between ages 18 and 74 chooses between three labour supply

[1] The annual LFS was carried out in the spring.

alternatives, corresponding to not employed, part-time, and full-time employment; all individuals are assumed to be out of the labour market from age 75. For singles, the wage potential of the adult is the same as that of the benefit unit. For couples, the model distinguishes between “principal” and “secondary” earners (as discussed in Section 4.2.3), and assumes that 60% of the benefit unit’s wage potential is associated with the former and the remaining 40% with the latter.

4.3 Self-employment

Self-employed are included in the model primarily in recognition of the higher year-on-year volatility that is commonly associated with their earnings stream, relative to employees. Self-employed people are distinguished from otherwise similar employees in three respects. First, although the same analytical equations govern the evolution of earnings of both self-employed and employees, the two groups are subject to different earnings parameters. Second, self-employed benefit units have no risk of incurring a low wage offer (discussed in Section 4.2.3). Thirdly, although the private pensions of the self-employed and employees are the same (discussed in Section 4.5), the self-employed are assumed to contribute a different share of their (pre-tax) earnings to the pension, and do not benefit from an employer pension contribution. These terms have been defined to limit the increase in computational burden of including the self-employed in the model.

Transitions into and out of self-employment are generated with respect to age specific probabilities, distinguished by self-employment status in the preceding year. These probabilities were set equal to the average year-on-year transitions reported by the British Household Panel Survey, pooled over the first 18 waves of the survey between 1991 and 2008.

Beyond the transition probabilities referred to above, the only other model parameters that differ between the self-employed and the remainder of the population are the parameters governing intertemporal evolution of earnings. These parameters were calibrated endogenous to the model structure, as described in Section 6.

4.4 Non-pension wealth

Non-pension wealth is assumed to follow the simple accounting identity:

$$w_{i,a} = w_{i,a-1} + \tau_{i,a-1} + B_{i,a-1} - t.c_{i,a-1}$$

where $w_{i,a}$ represents wealth of benefit unit *i* at age *a*, $\tau_{i,a}$ represents disposable income net of non-discretionary expenditure, *B* describes bequests received, and *t.c* is a term that represents discretionary consumption including associated consumption taxes and duties.

Disposable income net of committed expenditure is comprised of private income, plus (public) transfer payments, less taxes, less committed expenditure. The model evaluates committed expenditure, taxes and transfer payments, based on measures of private income and benefit unit demographics. The evaluation of committed expenditure is discussed in Sections 1 to 3. Private income in the model is comprised of earnings (described in Section 4.2), dispersals from private pensions (described in Section 4.5.2), and investment returns on non-pension wealth. Benefit unit demographics include age, birth cohort, relationship status (described in Section 4.7), and number and age of dependent children (described in Section 4.8). Simulated taxes and transfer payments are described in Section 4.6.

Investment returns to non-pension wealth are comprised of returns to housing and non-housing wealth. Simulation of returns to housing wealth are described in Section 1. All non-housing, non-pension wealth is assumed to attract an interest rate described by:

$$r^s = \begin{cases} r^I & \text{if } w \geq 0 \\ r_l^D + \left(r_u^D - r_l^D\right)\min\left\{-w/g^{ft}, 1\right\} & \text{otherwise} \end{cases}$$

where $r_l^D < r_u^D$. This specification allows the interest charges on debt to rise with the ratio of (net) debt to earnings potential (g^{ft} equal to the earnings potential of the principal earner).

In the principal calibration scenario, r^I is set equal to the average real return on long-term treasury bills reported between 1970 and 2010. r_l^D is set equal to annual averages of the monthly interest rates on sterling personal loans up to £10,000 to households reported by the Bank of England (code IUMHPTL) between 1995 and 2010. r_u^D is set equal to annual averages of the monthly interest rates for sterling credit card lending to households reported by the Bank of England (code IUMCCTL) between 1995 and 2010.

The negative calibration, sets r^I, r_l^D, and r_u^D using the same data, but subtracts half a standard deviation from all of the sample means described in the preceding paragraph.

4.5 Pensions

4.5.1 State pensions

The model is designed to allow state pensions to be simulated in a number of alternative ways that have been designed to capture variation in relation to the basic state pension and the state second pension. The choice between these alternatives, which are described in the technical report, depends upon the focus of the respective analysis. This section focuses exclusively on the base assumptions adopted for the model.

State pensions are assumed to be paid to qualifying benefit units from state pension age. State pension age is specified to increase in line with official guidelines for men, current as of the 2014 Pensions Act – see Section 2.2 on the Pension credit for further details.

The 2014 Pensions Act effectively brought forward the planned transition to a single-tier pension, which will now be available from April 2016. As the single tier pension can reasonably be assumed to influence behaviour in forward projections, the model adopts this policy amendment as though it was effective from 2011. The maximum value of the single tier pension is set equal to the guarantee credit, worth £137.35 per week to a single adult in 2011, and £209.70 to a couple (see Section 5.2). This benefit is indexed to assumed wage growth, consistent with prevailing legislation (although the government has stated its intention to index the benefit using the "triple lock").

Rights to the single tier pension are accrued through accreditation for national insurance contributions. An adult is accredited with national insurance contributions in any given year, if they pay class 1 contributions (see Section 5.1), if they are identified as involuntarily unemployed (receive a low wage offer in the model), or if they have a dependent child under age 12 in the household. At least 10 years' of national insurance contributions are required to receive any income under the single tier pension, and 35 years' contributions are required for the full single tier pension.

The above description is applied when simulating the population through time. An alternative approach is, however adopted when solving the lifetime decision problem. To limit computational burden when solving the lifetime decision problem, all benefit units are assumed to be eligible to a flat rate pension worth 90% of the maximum defined above – that is the condition on national insurance contributions is suppressed.

4.5.2 Private pensions

The model is structured to permit private pensions to be represented in a variety of ways. We have adopted a highly stylised specification for the base model structure, to maximise flexibility over other margins of interest. We describe this structure here.

Private pensions are modelled at the benefit unit level as defined contribution schemes. In each year, a benefit unit with earnings exceeding a minimum threshold, g_l, can choose whether to make fresh contributions to its pension scheme. If a benefit unit chooses to contribute to its pension, then a fixed share of its total pre-tax labour income, π, is added to its accumulated pension fund. Contributing benefit units also receive an employer contribution to their pension fund, which is specified as a fixed share of pre-tax labour income, π_{ec}. Eligible employer contributions to a benefit unit's pension fund in any given year are lost if the benefit unit chooses not to contribute to its scheme in the respective year. Wealth held in the private pension fund of benefit unit *i* at age *a*, $w^p_{i,a}$, is assumed to be illiquid, and attracts a fixed rate of return *r*. In most periods prior to pension receipt, pension wealth follows the accounting identity:

$$w^p_{i,a} = rw^p_{i,a-1} + (\pi + \pi_{ec})\, g_{i,a-1} \lambda^p_{i,a-1}$$

where *g* is labour income and $\lambda_{i,a}$ is an indicator variable, equal to one if the benefit unit of reference adult *i* at age *a* contributes to its pension, and zero otherwise. The only departures from this equation are following relationship transitions, where relationship formation doubles pension wealth and relationship dissolution halves it. Furthermore, a cap on the maximum value of the pension is imposed, set equal to £1.25 million in 2011.

A benefit unit can choose to access its accumulated pension fund at any time between ages 55 and 75, at which time 25% of the fund is taken as a tax-free lump-sum transfer into non-pension wealth, and the remainder is used to purchase an inflation adjusted life annuity (which is assessable for income tax discussed in Section 5.1). If a benefit unit chooses to access its pension in a given year, then all adult members must not work in that year, and a wage penalty is applied to any future labour income. The annuity rates assumed for analysis are calculated with reference to the survival rates assumed for individual birth cohorts, an assumed return to capital, and an assumed transaction cost levied at the time of purchase.

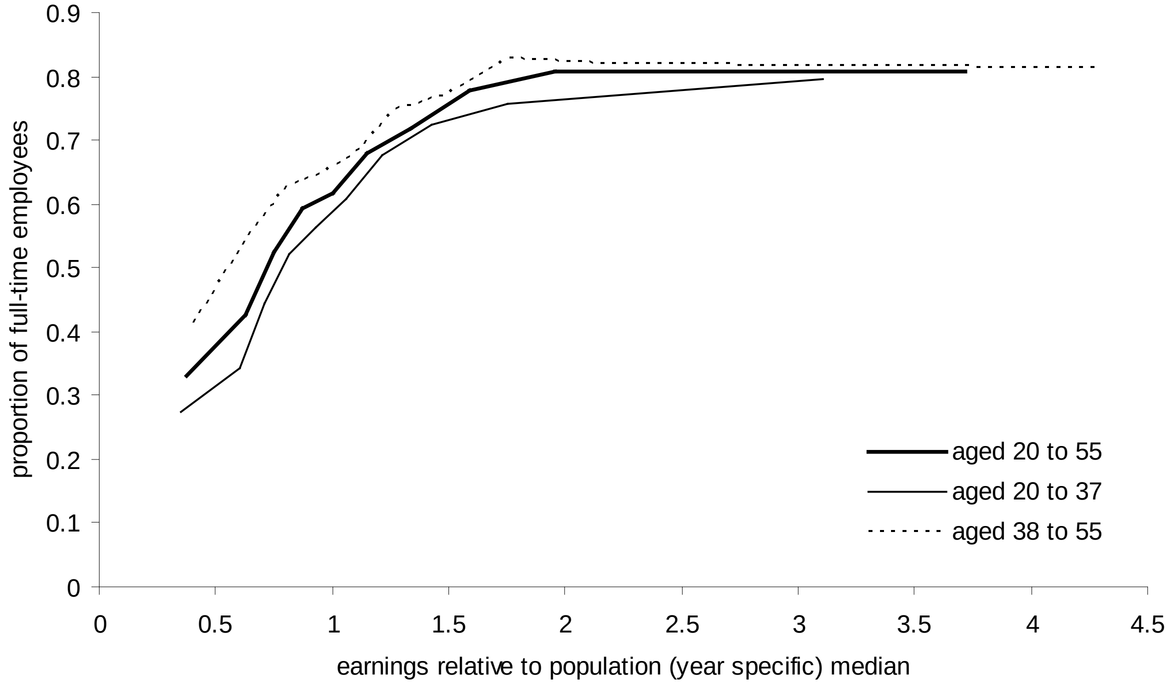
This specification of private pensions depends upon six parameters: the rate of return to pension wealth *r*, the minimum earnings threshold for pension contributions g_l, the rate of private contributions to pensions out of employment income π, the rate of employer pension contributions π_{ec}, the return assumed for calculating the price of pension annuities, and the fixed capital charge associated with purchasing a pension annuity.

There is a great deal of diversity in private pension arrangements in the UK, and in the details of occupational pensions in particular. Panel A of Figure 4.1 reveals that – although not universal – a sizeable majority of employees were offered some form of contribution in respect of participation in an employer sponsored pension. Eligibility to an employer sponsored pension is reported to increase from a low of between 30 and 40 per cent among individuals on less than half of median earnings (increasing by age group), to between 75 and 85 per cent among individuals on more than one and a half times median earnings. The figure also indicates that eligibility for an employer pension contribution exhibited a stronger relationship with employee earnings than it did with age. Following these observations, we set g_l equal to 75% of median earnings.

Panel B of Figure 4.1 indicates that, for employees who received an employer pension contribution, the distribution of employer pension contributions was dominated by a single mode between 12.5 and 15 per cent of employee wages. Bearing in mind that the decision by an employee not to participate in their employer's sponsored pension plan would usually result in the forfeiture of any matching employer pension contributions on offer, the scale of the employer contributions reported in Panel B provides an indication of how important these contributions are in supporting the UK system of private sector pension provisions. Panel B of Figure 4.1 also reveals that there was very little difference between the distributions of employer pension contributions offered in low-pay industries and the wider labour market, with the principal disparity being that employer contributions in excess of the mode were less frequent among employees in low pay industries. We consequently set the rate of employer contributions to 14%; the rate of private contributions to pension wealth was set to the `normal' contribution rate stated in the guidance to interviewers for the FRS, equal to 8%.

In the principal calibration, we set the (real) return assumed to pension wealth during the accrual phase, r, to 3.5% per annum, based on typical assumptions of a 60:40 split between equities and bonds, a 5% real return to equities, and a 1.2% return to Gilts. The capital return assumed for calculating the price of pension annuities was set equal to 1.5%, reflecting the assumed (real) rate of wage growth, and the associated capital charge was set to 4.7% based on "typical" pricing margins reported in the pension buy-outs market (see Lane (2008), p. 22). All rates of return to private pension wealth were reduced by 1.85% per annum under the negative calibration, equal to half of one standard deviation of the (real) returns reported for long-term treasury bills between 1970 and 2010.

Figure 4.1: Eligibility rates of full-time employees to employer sponsored pensions by age and earnings



Panel A: eligibility to any employer pension contribution

Panel B: employer contribution where some contribution was made

Source: Panel A: author's calculations on individual level data from the 2006/07 and 2007/08 waves of the FRS
Panel B: author's calculations using data from waves 2005 to 2009 of the Annual Survey of Hours and Earnings
Notes: Earnings deciles defined within survey waves, and averaged across waves
Excludes employees in the public sector and the self-employed
low pay industries as defined by the Low Pay Commission (2010), Appendix 4

4.6 Taxes and benefits

The policy structure implemented for re-parameterising the model broadly reflects the UK policy context as it applied in April 2011. Each scheme considered for analysis is described under a dedicated subheading below.

4.6.1 Taxes[2]

- *Income taxes*

Income taxes are calculated separately for each simulated adult. The 'taxable income' of an adult is evaluated by aggregating their income from employment (both employee and self-employed), retirement pensions, and investment returns on assets held outside of pensions and Individual Savings Accounts, and deducting their respective personal allowance. In contrast to practice, this definition of taxable income omits jobseeker's allowance, a stylisation that is likely to have a negligible bearing on simulations as individual employment circumstances are assumed to be stable within each year (so that individuals in receipt of Jobseekers Allowance do not typically pay income tax).

The personal allowance for individuals under state pension age is set equal to £7475 per year in 2011, and is withdrawn by 50p for each £1 of income in excess of £100,000 per year. The personal allowance for individuals in excess of state pension age is set equal to £10,015 per year[3], and is withdrawn in stages, at the rate of 50p for each £1 of private income between £24,000 and £29,080 per year, and then again at the rate of 50p for each £1 of private income in excess of £100,000 per year.

Income taxes are paid at the rate of 20% on taxable income up to the basic rate limit of £35,000 per year, at the rate of 40% on income between £35,001 and the higher rate limit of £150,000, and at the additional rate of 50% on all taxable income in excess of £150,000 per year. All thresholds and personal allowances are assumed to grow in line with wages under both the principal and negative calibrations.

- *National insurance contributions*

Class 1 (employee) National Insurance contributions are calculated on the earnings of each adult under state pension age. Contributions are paid at the rate of 12% on earnings in excess of the Primary Threshold (worth £139 per week in 2011) and under the Upper Earnings Limit (£817 per week), and at the rate of 2% on earnings in excess of the Upper Earnings Limit. Furthermore, individuals who are identified as having contracted out of the State Second Pension, receive a contracted out rebate worth 1.6% on all earnings in excess of the Primary Threshold. All thresholds are assumed to grow in line with wages under both the principal and negative calibrations.

- *Taxes on consumption*

---

[2] See Pope and Roantree (2014) for a concise review of the UK tax system.

[3] This personal allowance is mid-way between the personal allowances provided to individuals aged 65-74 and those aged 75+ in 2011-12.

Simulating taxes on consumption for the UK is complicated by the different rates of Value Added Tax and excise duties that are payable on different goods and services, which contrasts with the aggregate measure of non-durable consumption projected by LINDA. To accommodate this variation, we have estimated a series of reduced form regressions that describe the fraction of total non-durable expenditure on each of six consumption categories: goods liable to the full rate of VAT, goods liable to the reduced rate of VAT, alcohol, tobacco, fuel, and insurance premia distinguishing those liable to the standard and higher rates of tax. These reduced form models are used to approximate benefit unit consumption of each of the six expenditure categories during each year, and the associated tax burden is evaluated by multiplying by the relevant tax or duty rate.

In the case of VAT and taxes on insurance premia, the relevant rates are unambiguously defined: the full-rate of VAT in 2011 was 20% and the reduced rate was 5%; the standard rate of tax applied to insurance premia was 6% and the higher rate was 20%. In contrast, the duties levied on alcohol, tobacco and fuel vary substantially across individual products. Averages were consequently calculated for the rates applied to individual products within these three broad consumption categories, weighting by expenditure share (data supplied by HM Treasury). These calculations yielded indirect tax rates (inclusive of VAT) of 50% for alcohol, 66% for fuel, and 80% for tobacco.

The reduced form models for consumption shares were calculated using data drawn from the Living Cost and Food Survey, pooled over the years 2001 to 2010 (115,827 observations). Each of the six categories of consumption were evaluated following coding used for the IGOTM effective as at 31 August 2012. After testing a number of alternatives, we selected a linear regression specification for all six consumption categories, defining the respective consumption share as a function of age, relationship status, numbers of children, education status, disposable income, and consumption. Regression statistics are reported in Appendix D.

The regression statistics reported in Appendix D indicate that many of the specifications fail to capture much of the variation in expenditure shares that is described by the data. This is not particularly surprising: the expenditure of smokers on tobacco products, for example, is very different to that of non-smokers, and the model is ill-equipped to accommodate this type of variation. Such considerations imply that the way that we have implemented taxes on consumption is better adapted to reflecting population averages of associated incidence, than it is to reflect variation between benefit units.

- *Council Tax*

Council tax is assumed to depend on relationship status, and the number and age of dependent children. It is modelled based on imputed bedrooms, consistent with the definition of the "bedroom entitlement" applied for housing benefit (see Section 5.2). One bedroom is imputed for each single adult / cohabitating couple, another for each child aged 13 or over, and another for every 2 children aged under 13 years, subject to a maximum of four bedrooms. Furthermore, higher rates of council tax are assumed to apply to couples than to singles. The rates assumed for 2011 are: £14.17 per week for single adults and £19.23 for couples in a one bedroom house; £16.19 for singles and £22.26 for couples in a two bedroom house; £18.22 for singles and £27.32 for couples in a three bedroom house, and £20.24 for singles and £32.38 for couples in a four bedroom house. These rates are all indexed to wage growth.

4.6.2 Benefits[4]

All benefits values and thresholds, unless otherwise stated, are assumed to be frozen (in nominal terms) from 2016 to 2020, and to be indexed to prices thereafter.

- *Income support prior to state pension age*

Support to benefit units with low incomes who are not affected by a health condition (see employment and support allowance below), and do not satisfy the age requirements of the pension credit (see below) is modelled on income support. The maximum benefit payable under income support is comprised of a number of components. Single adults are eligible for income support worth £71 per week in 2011, and couples for £111.45 per week. Benefit units with dependent children under age 6 are assumed to be eligible for healthy start vouchers worth £6.20 per week for new-borns, decreasing to £3.10 per week for children aged 1 to 5. Free school meals are also included, worth £5.52 for each child aged 5 to 10.

Furthermore, benefit units that are identified as home owners with a mortgage are eligible for assistance with mortgage interest payments, worth up to 3.85% per annum (nominal, 1.85% real) on a mortgage of up to £200,000.

The benefits referred to here are means-tested, withdrawn at the rate of £1 for each £1 of private income, and at the rate of £1 per week for every £250 of non-housing capital in excess of a £6,000 disregard for those on income support.

- *Employment and Support Allowance*

Benefit units are assumed to be eligible for the employment and support allowance (exogenously assumed to be) during the ten years prior to state pension age (see below).

Employment and Support Allowance (ESA) is designed to support working-aged individuals with a 'limited capability for work'. Claimants must be aged between 16 and state pension age, and must satisfy a related health test. The model abstracts from the contributory benefit and (13 week) assessment phase, focussing on income related ESA 'main phase' payments. The benefit tops-up income to the relevant applicable amount, equal to the sum of personal allowances and premia.

The personal allowance is worth £67.50 per week in 2011 for single and sole parents, and £105.95 per week for a couple, with premia worth £32.35 per week (for support receipients). These benefits are subject to similar means-tests as described for income support, with the exception that the assets test deducts £1 per week of benefit for every £500 over a £10,000 threshold.

- *Pension credit*

The pension credit is comprised of two elements, both of which are modelled explicitly. The guarantee credit is payable from age 60 in all years prior to 2010, rising to age 61 in 2010, 62 in 2012, 63 in 2014, 64 in 2016, and from simulated state pension age thereafter. Simulated state pension age

[4] See Hood and Oakley (2014) for a concise review of the benefits system in Great Britain.

is projected to remain at 65 years until 2018, increasing to 66 in 2019, to 67 in 2026, and to 68 in 2034.

The age of eligibility for the guarantee credit is defined with respect to planned variation of state pension age for women, taking into consideration the division of time into discrete annual intervals that is a feature of the model. Specifically, state pension age for women increased to 60 years and 1 month in May 2010, 61 years and 1 month in May 2012, 62 years and 1 month in May 2014, and is projected to rise to 63 years and 3 months by July 2016, 64 years and 1 month by November 2017, and 65 years by November 2018 at which time the state pension age of men and women will be equalised. Thereafter, existing legislation that will see state pension age of both men and women increase to 65 years and 1 month by January 2019, 66 years and 1 month by May 2026, and the government has announced its intention to increase state pension in line with longevity thereafter.[5]

The guarantee credit provides a maximum benefit worth £137.35 per week to a single adult in 2011, and £209.70 to a couple. Furthermore, benefit units that are identified as home owners with a mortgage are eligible for assistance with mortgage interest payments, worth up to 3.85% per annum on a mortgage of up to £200,000. These payments are subject to a means-test on total benefit unit private income and non-housing capital, with benefits reduced £1 for every £1 of private income, and by £1 for every £500 of non-housing wealth in excess of a £10,000 disregard.

The savings credit provides additional income to benefit units that have private income in excess of a minimum threshold and are above state pension age. Savings credit increases at the rate of 60p for every £1 of private income in excess of £103.15 per week for singles in 2011, and £164.55 per week for couples, up to a maximum benefit worth £20.52 per week for singles in 2011 and £27.09 per week for couples. Any private income earned beyond the sum required to obtain the maximum savings credit benefit, reduces the benefit at the rate of 40p in the £1.

The maximum benefits payable under the guarantee credit are assumed to increase with real wages. In contrast, the maximum value of the savings credit is assumed to be indexed to prices. This implies that the savings credit threshold grows faster than earnings.

- *Working tax credit*

Working tax credit is payable to any benefit unit over age 24, with at least one adult in full-time employment, or with at least one adult in some employment and with a dependent child.[6] The maximum benefit payable is comprised of a basic element worth £1920 per year in 2011, plus an additional element worth £1950 per year for single-parents and couples, plus an element worth £790

---

[5] The 2012-based projected mortality data reported by the ONS indicate that the period life expectancy at birth in the UK was 79 for men, rising to 84.1 years in 2037. This implies an increase in life expectancy of 0.2 years for each year that birth is delayed. The Chancellor as part of the 2013 Autumn statement that people should expect to spend one third of their lives in retirement. Combining these two details suggests that each year that birth is delayed should increase state pension age by 0.133 years, suggesting that state pension age should increase by 1 year in every 7.5 years.

[6] These conditions are designed to approximate the rules concerning employment applied by the WTC, which require most claimants to work at least 30 hours per week. Disabled individuals, lone-parents, and individuals aged 60 and over can work 16 hours per week, and couples with dependent children must work a combined total of at least 24 hours per week.

per year for benefit units with at least one adult working at least 30 hours per week. Furthermore, benefit units in which all adult members work at least 16 hours per week are eligible to subsidised childcare for children aged 14 and under, equal to 70% of qualifying childcare costs up to a maximum of £175 per week for a single child, and £300 per week for two or more children. See Section 2 for details concerning the simulation of childcare costs.

The benefits payable under the working tax credit are means-tested in conjunction with benefits payable under the child tax credit described below.

- *Child tax credit*

The child tax credit is payable to benefit units with dependent children under 20 years of age. The maximum benefit payable is comprised of a family element worth £545 per year in 2011, and a child element worth £2,555 per year for each qualifying child.

Eligible benefits under the child tax credit are aggregated with those of the working tax credit and subject to a common means test. Any benefit unit income in excess of a minimum threshold reduces the total tax credits payable, at the rate of 41p per £1. The minimum threshold is £6,420 per year in 2011 for benefit units eligible to both the working tax credit and the child tax credit, and £15,860 per year for those eligible to only the child tax credit.[7]

- *Child benefit*

Benefit units with dependent children receive Child Benefit. The benefit is worth £20.30 per week in 2011 for the first child and £13.40 per week for each subsequent child. Child benefits are not means-tested.

- *Disability related benefits*

Key disability related benefits, including Employment and Support Allowance, Disability Living Allowance, Personal Independence Payment, Attendance Allowance, and Carer’s Allowance, have all been included in the model as part of the model extensions to accommodate varying health status. Section 15 describes these benefits.

- *Housing benefit*

Housing benefit is paid in respect of simulated rent payments to benefit units that have sufficiently low incomes. The maximum benefit payable under housing benefit is equal to rent expenditure capped at the respective local housing allowance. The local housing allowance is determined by rates graded by the bedroom entitlement of the benefit unit. The bedroom entitlement allows one bedroom for each single adult / cohabitating couple, another for each child aged 13 or over, and another for every 2 children aged under 13 years, subject to a maximum of four bedrooms. The local housing allowance is £162.50 per week in 2011 for one bedroom, £188.50 per week for two

[7] This means test ignores the second threshold used to means test child tax credits in 2011/12 as this feature was withdrawn from 2012/13.

bedrooms, £221 per week for three bedrooms, and £260 per week for four bedrooms.[8] Rental charges actually incurred in the model are described in section 1.4.3.

Housing benefit is means-tested with respect to a broad measure of income that is net of child care costs eligible to subsidies via the working tax credit (see above), and most taxes and benefits (excluding the disability living allowance). The measure of income also includes imputed returns on non-housing wealth, calculated as £1 per week for every £250 of wealth in excess of a £6,000 disregard for individuals under the age of eligibility for Employment and Support Allowance (see above), and £1 per week for every £500 of non-housing wealth in excess of a £10,000 disregard otherwise. This measure of income is then reduced by an earnings disregard, worth £5 per week for single people, £10 per week for couples, and £25 per week for lone-parents, with an additional £17.10 disregard applied for benefit units with at least one adult working full-time, or at least one adult working and with dependent children.

Any excess of this measure of income above the benefit unit's relevant applicable amount reduces housing benefit at the rate of 65p in every £1. The applicable amount is equal to £67.50 per week for single adults, £72.30 for lone-parents, and £105.95 for couples, increasing by £62.33 for each dependent child (aged 18 and under). Couples and lone-parents are also eligible to an additional family premium worth 17.40 per week.

- *Council tax benefit*

Council tax benefit is provided to subsidise the costs to households of council taxes (see Section 4.6.1). Although council tax benefit was available in 2011, it has been amended to a localised benefit from 2013/14. Nevertheless, council tax benefit is modelled by default in the form that it took prior to localisation of the benefit. This structure assumes that council tax benefit is subject to the same means-testing rates and thresholds as housing benefit described above, with the exception that the benefit is withdrawn the rate of 20p in every £1. As council tax benefit and housing benefit are withdrawn simultaneously, this implies 85p of aggregate (council tax and housing) benefit is withdrawn for every £1 of income above the benefit unit's relevant applicable amount so long as both benefits are received.

- *Universal credit*

Universal credit is designed to replace income support (including income based jobseeker's and employment and support allowances), housing benefit, the working tax credit, and child tax credit. Although universal credit was not available in 2011, it has been in planning since the conservative party annual conference in 2010. The scheme has been rolled-out progressively since April 2013, and is scheduled to have replaced the existing benefit schemes by 2017/18. This policy shift is consequently built into the model.

The maximum payment under universal credit is simulated by aggregating the maximum sums payable under income support, child tax credit, and housing benefit, and then adding the maximum

[8] In practice, the bedroom entitlement distinguishes dependent children by sex of dependent children, which is omitted in the model assumptions defined here. The assumed local housing allowances are set equal to 65% of the relative caps imposed in April 2011, an assumption that is broadly based on the (now discontinued) DWP Tax Benefit Model Tables.

childcare component payable under the working tax credit. This maximum benefit is then withdrawn at the rate of 65p for each £1 of earned income net of taxes and national insurance contributions earned above the benefit unit's relevant "work allowance". Furthermore, any pension income received is deducted from the universal credit payable, and each £250 of non-housing wealth in excess of a £6,000 disregard reduces the universal credit by £1 per week (in 2011 prices).

The work allowance depends upon whether housing related support would have been received through either income support (off-setting mortgage interest) or housing benefit (off-setting rent). If so, then so-called "lower work allowances" apply, worth £25.21 per week to single adults in 2011, £59.72 to lone parents, £25.21 to couples, and £50.41 to couples with children. Otherwise, "higher work allowances apply, which are identical to the lower allowances for benefit units without children, but are worth £166.67 per week to lone parents, and £121.71 to couples with children.

- *Allowance for imperfect benefits take-up*

Take-up of benefits by a benefit unit in any year is modelled as a random event. Each benefit unit in each year is assigned a number drawn from a uniform [0,1] distribution. This number is compared against the "take-up rate" associated with the benefit unit's circumstances in the respective year. If the number is less than the take-up rate, then the benefit unit is assumed to take-up their benefits; otherwise they do not.

Take-up rates distinguish between benefit units, based on labour supply, family demographics, and age, as described by the following table:

no adult employed			
working lifetime		from state pension age	
without children	with children	singles	couples
57%	62%	64%	56%
at least one adult employed			
without children		with children	
income to £10,000 p.a.	income £10,000+ p.a.	income to £10,000 p.a.	income £10,000+ p.a.
47%	28%	97%	69%

The take-up rates for benefit units without any employment were set equal to figures reported for 2013/14 in DWP (2015). Benefit units under state pension age are subject to take-up rates reported for (income based) Job Seekers Allowance (table 4.3.1), and those in excess of state pension age are subject to rates reported for the Pension Credit (table 2.3.4). The take-up rates and income thresholds assumed for analysis are drawn from Tables 4 (with children) and 11 (without children) of HMRC (2015). These tables report take-up rates for Child Benefit, Child Tax Credit and Working Tax Credit estimated for 2013-14 (updated in February 2016).[9]

4.6.3 Tax residuals

---

[9] A common rule is assumed for disability related benefits (DLA, AA, and CA), as for other non-housing benefits, due to the unique difficulties associated with identification of disability benefit take-up rates (on which, see Kasparova *et al*., 2007, for DLA and AA, and Berthoud, 2010, for CA).

Tax residuals are model parameters that can vary by birth year and age. The influence that these residuals have on disposable income is determined within the tax function, and so is otherwise unconstrained. These residuals have been suppressed in the current calibration for the model by setting them to zero.

4.7 Relationship status

The model takes explicit account of the evolving relationship status of reference adults during the course of their respective lives. Relationships are assumed to form exclusively between adults within the simulated population. Relationship transitions are simulated based on exogenous age, year, and education specific probabilities of relationship formation and dissolution.

4.7.1 Parameterisation

The model requires rates of marriage formation, divorce, and widowhood. The last of these three states is projected on the basis of projections of individual specific mortality (see Section 4.9). This section describes derivation of the model parameters for marriage formation and divorce.

At the time of writing, the ONS reports historical data for the number of marriages in England and Wales by age, sex and calendar year at annual intervals between 1851 and 2011. The ONS also makes available for modelling purposes the component factors that underlie its population projections, which describe official estimates for the number of marriages by age and sex at annual intervals between 2008 and 2033. Furthermore, ONS population estimates by age, sex and marital status are available for England and Wales at annual intervals between 1971 and 2033. These statistics permit age and gender specific marital rates to be calculated for England and Wales at annual intervals between 1971 and 2033 inclusive.

Furthermore, the ONS reports age and sex specific divorce rates for England and Wales at annual intervals between 1950 and 2010, which can be extended to 2032 by the component factors of the ONS population projections that are noted above.

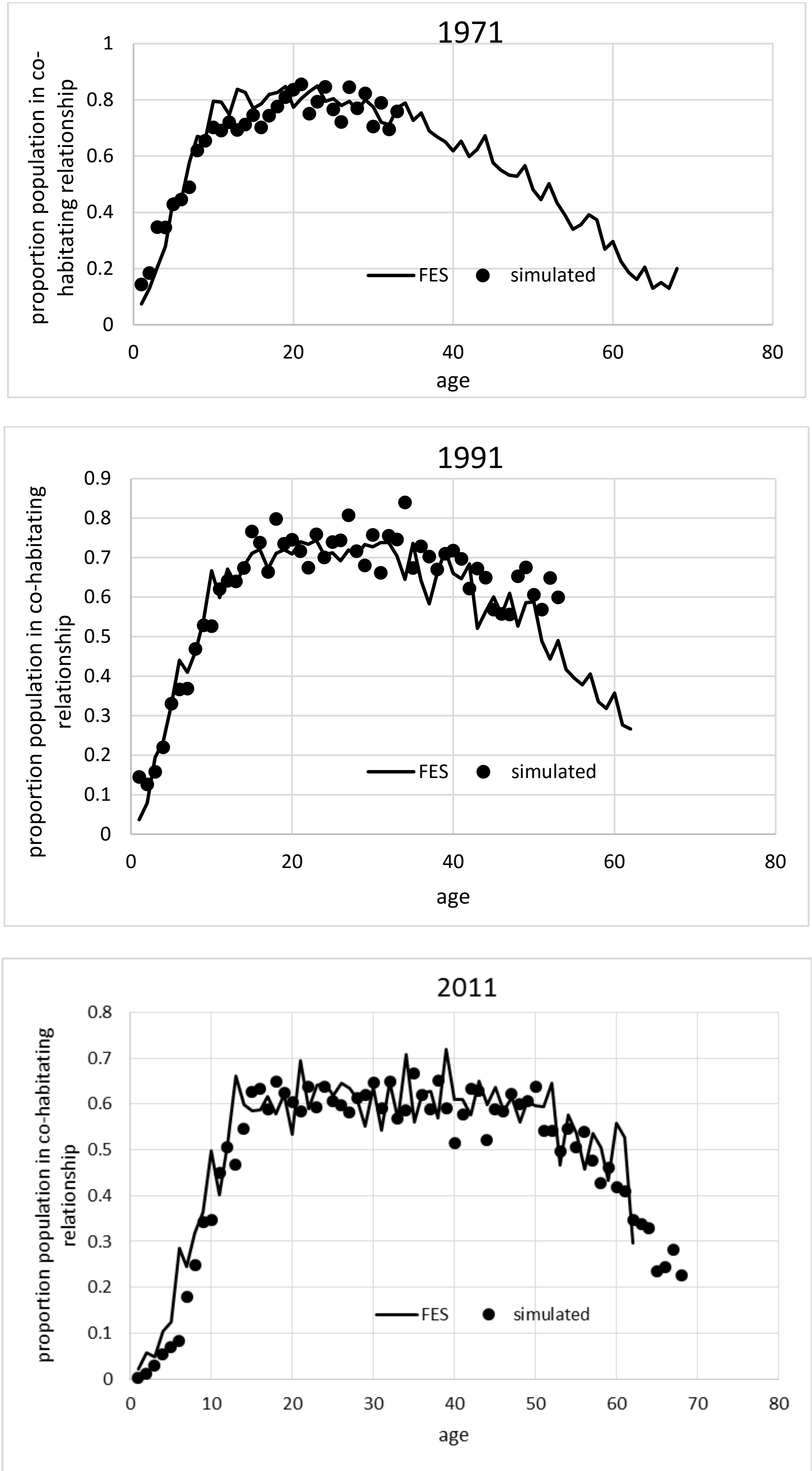
The rates of marriage formation and divorce that are described above are imperfect for modelling purposes in (at least) three important respects. First, the transition rates for marriage and marital dissolution that are described above are not distinguished by education status. Second, marriage rates calculated on historical data do not account for marriages that are performed abroad. Thirdly, the majority of the statistics that are reported by the ONS focus on legal marital status, and do not extend to include civil partnerships or cohabitation.

The focus of ONS statistics on legal marriage is problematic for modelling purposes due to the rise of civil partnerships and cohabitation, and the fact that couples who share the same address often engage in some pooling of consumption and income. This pooling of financial resources is recognised by the system of social security in the UK, which treats cohabitating couples in the same way as registered married couples when determining eligibility for most benefits (excluding state pensions and bereavement allowances).

The rates of marriage and divorce described above were consequently adjusted to match the implied proportion of the population married, by age and year, to marriage rates described by LCFS data reported between 1978 and 2012. These adjustments were implemented as follows, focussing discussion here on women (with the same approach adopted for men). First, all women were assumed to be single at age 16 in all survey years. Given the year specific probabilities of marriage for 16 year olds derived as described above, the proportion of women married at age 17 was calculated for each year. Marriage rates were then projected forward one year to age 18, based on the probabilities of marriage and divorce of women aged 17 – as described above – and the probabilities of death of men aged 17. This last set of probabilities was used to evaluate widowhood, on the assumption that all marriages are between identically aged individuals, and that the probability of mortality is the same for married and non-married individuals.

The proportions of women married at age 18 derived as described above were then compared against associated proportions calculated from LCFS data. If the "simulated" series indicated a smaller proportion of women married than the survey data, then the associated marriage rates were increased. If the simulated series indicated a larger proportion of women married than the survey data, then the associated divorce rates were increased. The procedure then advanced to age 19. This approach was used to derive sex, age, year, and education specific probabilities of marriage and divorce for all years between 1978 and 2012.

Figure 4.2: simulated and sample rates of cohabitation by year



Source: Author's calculations using data reported for selected cross-sections by the Living Costs and Food Survey and its predecessor surveys (FES), and simulated by the LINDA model (simulated)

4.8 Fertility and dependent children

The model explicitly simulates the birth of dependent children to reference adults. Children who are born into the simulation are assumed to remain dependents of their respective reference adult until they reach 17 years of age. The only exception to this is in the case of relationship dissolution, when half of the dependent children in the household are assumed to leave with the departing spouse. [10]

The way that child birth is allowed for in the model has been carefully selected to provide a high degree of control over the computational burden imposed on the lifetime decision problem. This is important because the objective to take account of both the number and age of dependent children in a household can have a very pronounced impact on computation times. If, for example, reference people could receive a new dependent child at any age between 20 and 45, with no more than one birth in any year, and no more than six dependent children at any one time, then this would add an additional 406,624 state variables to the computation problem (with a proportional increase in the associated computation time).[11]

The model is consequently structured so that child birth can be limited to an exogenously defined discrete set of “birth ages” (not to be confused with the “birth years” distinguishing individual cohorts). Multiple births are accommodated at each birth age to ensure that the model is capable of matching the distribution of family sizes observed in the data, subject to the limitation that births only occur at one of a few birth ages. This structure brings with it a set of complexities that may not be evident at first glance. Some simulated characteristics will, for example, reflect discrete jumps as child birth and maturation are synchronised between large numbers of households in a way that we do not observe in the survey data.[12] The focus on a limited number of birth ages also complicates identification of the transition probabilities upon which the model is based.

The transition probabilities required to simulate child birth at the assumed child birth ages are calculated internally by the model, based upon a set of disaggregated fertility rates that constitute basic parameters of the model. The base model parameters include fertility rates that vary by individual age, year, relationship status, and the number of existing dependent children in the household. Defining the model in terms of disaggregated model parameters has the dual advantages of permitting variation of the terms assumed for fertility, and aligning the model parameters with statistics reported for the UK population.

4.8.1 Parameterisation

The model requires fertility rates by age, year, relationship status, and number of previous births to simulate dependent children. Unfortunately these rates are not readily available for the UK, and so

---

[10] In the case of odd-numbered children, *n*, the number of children assumed to depart the household is set equal to (*n*-1)/2. Children are withdrawn from the reference adult’s benefit unit in order from eldest to youngest.

[11] assuming that children remain dependents up to their 17th year.

[12] This issue is usually most evident when considering simulations for individual birth cohorts than for groups of reference adults that vary by birth year.

were constructed based on a set of identifying assumptions and a selected set of publically available data sources.

The ONS reports the number of births by age of mother and registered marital status at annual intervals between 1938 and 2010 for England and Wales. It also reports the number of births within marriage/civil partnerships by age of mother and number of previous live-born children, at annual intervals between 1938 and 2010 for England and Wales. As noted previously, ONS population estimates/projections by age, sex and marital status are available for England and Wales at annual intervals between 1971 and 2033. Furthermore, the proportions of women by age, year, and marital status, recorded as having 0, 1, 2, 3, and 4+ dependent children can be estimated on the sample reported by the LCFS (and its predecessor surveys) for the period between 1971 and 2012.

We assume that the data reported by the ONS for England and Wales, and by the LCFS, are representative for the United Kingdom. Assume also, that the number of children reported by the LCFS for women to age 45 is equivalent to the total number of live-born children. Then, combining the ONS population estimates with the proportions estimated on LCFS data, we can calculate the number of married women with 0, 1, 2, 3, and 4+ dependent children by age and year. Ignoring the distinction between marriage and civil partnerships, we then combined the 'number of births' data reported by the ONS with our population estimates to obtain birth rates for married women that distinguish between the mother's age, year, and the number of previous births. These birth rates were extended to single women by applying a proportional adjustment equal to the age and year specific ratio of the average birth rate of single women to married women, described by the ONS population estimates that are cited above.

The fertility rates that were obtained as outlined above can be treated as approximate only. To improve the ability of the model to capture the evolving population structure in the UK, we adjusted the statistics that were obtained as described above to align the model to ONS estimates for England and Wales of the percentage distribution of women of childbearing age by number of live-born children, age and year of birth of woman between 1920 and 1991. This was achieved as follows.

Each birth cohort of women was considered in turn, starting with those born in 1920 and then increasing at annual intervals to those born in 1990. It was assumed that all 16 year old women were unmarried without children, and that no more than a single child could be born between adjacent ages. We began by assuming the age specific probabilities for relationship transitions (for women) that are described in the preceding subsection, and the first approximations to fertility rates that are described above. Given these assumptions, we used the same Monte Carlo routine that is used to generate demographic transitions in the model (assuming a full set of 30 birth ages between ages 16 and 45) to project relationship status and number of dependent children for a simulated population of 100,000 women forward one year to age 17. The total proportion of women with 0 children at age 17 can then be calculated, and compared against the associated population estimate reported by the ONS.[13]

[13] The ONS data report estimates at five year intervals between birth cohorts, and five year intervals between women's ages. Linear interpolation was used to identify intermediate values, using the "interp2" command in MATLAB.

If the proportion of 17 year old women with 0 children was too low, relative to the respective ONS estimate, then fertility rates were adjusted down (and vice versa), until a match to within half a percentage point was obtained.[14] Given the population composition identified for 17 year-old women that is described above, analysis proceeded by stepping forward one year to age 18 based on the assumed rates of marriage and marital dissolution, and assumed starting values for associated fertility rates. This involved identifying the proportion of women falling into each of six mutually exclusive categories (married/unmarried, 0/1/2 children), comparing the implied population distribution against ONS estimates, and adjusting the assumed fertility rates to minimise the disparities obtained.[15] Repeated application of this procedure to successively higher ages allowed fertility rates to be identified for the entire life course.

The ONS estimates report the proportions of women with 0, 1, 2, 3, and 4 or more children. Fertility rates for women with 4 or more children were calculated by applying a cohort specific multiplicative factor to the fertility rates identified for 3 or more children to ensure that cohort specific completed fertility rates matched to ONS population estimates.

The above analysis produced estimates of fertility rates distinguished by relationship status, age, year, and number of previous children for women born between 1920 and 1990. Obtaining a similar set of statistics for men is frustrated by the relative scarcity of data that are available. The ONS produces period fertility rates for both men and women, by age and year of observation and these indicate that, relative to men, the fertility rates of women are both compressed and skewed toward lower ages. Alternative methods could be devised to capture this variation. Nevertheless, the complexity of the temporal variation that is implied by alternative fertility assumptions, combined with the paucity of suitable data for validation, argues in favour of a simplified approach in the current context.

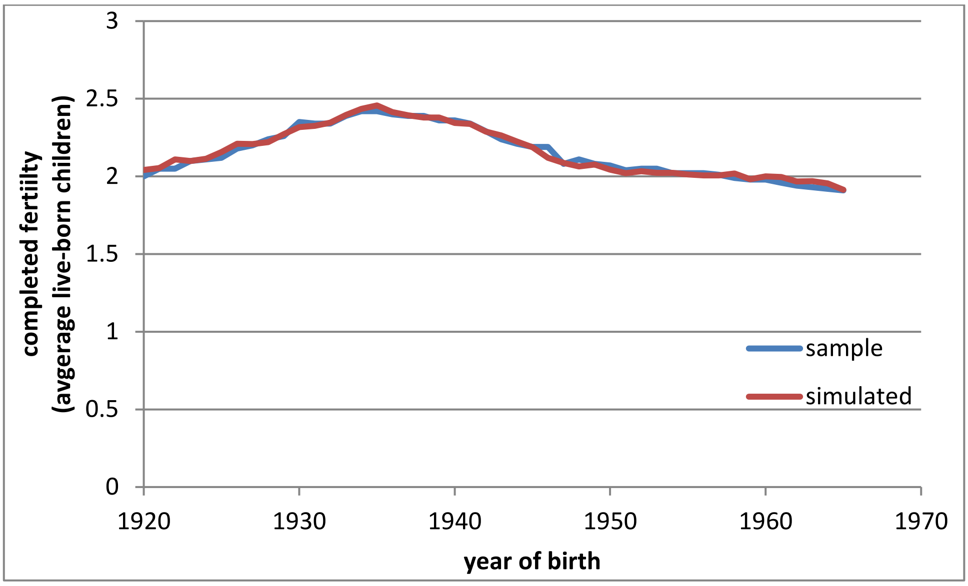
The gender neutrality that is a feature of the current model was accommodated in relation to the fertility statistics considered for analysis by adjusting the estimates that are discussed above by a fixed proportional factor to ensure that the model matched cohort specific completed fertility rates when the simulated population was generated based upon the gender neutral rates of marriage and marriage dissolution that are described in the preceding subsection.

---

[14] The same proportional adjustment was applied to fertility rates of married and single women.

[15] Fertility rates were adjusted in an ordered fashion, starting with the fertility rates of childless women, and working up in the number of dependent children.

Figure 4.3: Fertility rates by year of birth; simulated and sample data



4.9 Mortality

The mortality of each benefit unit's reference adult is modelled as uncertain, subject to a maximum potential length of life of 131 years. Death is simulated by generating a uniform random number between zero and one in each potential year for each benefit unit. If the random number is less than the respective age and year specific mortality rate, then the reference person of the respective benefit unit is considered to die and the benefit unit exits the simulated population.

The age and year specific mortality rates are set to observed rates between 1951 and 1980. The principal calibration sets age and year specific mortality rates between 1981 and 2062 to rates adopted by the ONS to generate the 2012 principal projection for the UK from 1981 to 2062. In contrast, the negative calibration assumes mortality rates between 1981 and 2062 used to generate the 2012 "high life expectancy variant" of the ONS population projections for the UK These two series differ for mortality projections between 2013 and 2062.

All dates prior to and following the time series reported by the ONS (1951 to 2062) are imputed by employing a geometric convergence rate of 2.5 per cent per annum from the last age-specific statistic loaded into the model (1951 if back in time, 2062 if forward in time) and a series assumed to represent the "very long-run". This "very long-run" projection is obtained by amplifying the year-to-year variation in the two nearest years loaded into the model by a factor of 15.

The sex specific mortality rates published by the ONS are aggregated into a gender neutral series by weighting each series by the relative likelihood that the respective sex survives to each age, given the assumption that sexes are equally weighted at birth.

5. Data for the reference population cross-section

LINDA makes its projections, starting from data reported for a reference population cross-section. These data are drawn from the Wealth and Assets Survey (WAS), which is currently the micro-dataset that provides the most complete description of household demographics, income, and wealth that is available for the UK.

The sample frame for the WAS is the small users Postcode Address File covering residential addresses in Great Britain, comprised of England, Wales and Scotland excluding North of the Caledonian Canal and the Isles of Scilly. At the time of writing, the WAS is comprised of three waves. Wave 1 data were collected between July 2006 and June 2008. The achieved sample from Wave 1 was re-interviewed two years later in Wave 2 (conducted between July 2008 and June 2010). In Wave 3, responding and non-contact households in Waves 1 and 2 were re-surveyed two years later, between July 2010 and June 2012. The Wave 3 sample was also augmented by a new random sample. All surveys were administered by Computer Assisted Personal Interviewing

The sample of households for Wave 1 was spread evenly across the sample period, and was selected to be geographically representative of the population of Great Britain, subject to over-sampling of high wealth households.[16] Of the 55,835 households invited to participate in wave 1 of the survey, responses are reported for 30,511 households, implying an achieved response rate of 55 per cent. Wave 2 reports data for 20,009 households (response rate 68 per cent), and Wave 3 reports data for 21,251 households (response rate of 61 per cent).

The model parameters have been updated using a reference population comprised of 10,771 households reported by wave 3 of the WAS between January and December 2011. This sample period aligns with the period of data reported by the 2011 Living Costs and Food Survey, which is the second most important data source used to parameterise the model.[17] The representativeness of the WAS sample for the British population was considered, by comparing the distribution of total gross household income reported by the WAS against associated data reported by the Family Resources Survey. Results of this analysis suggest a close degree of agreement exists between these two data sources; see Appendix B for details.

Formatting the data for use in the model is performed by a single STATA "do" file, which is provided with the model and is available from the authors upon request. Data from the household level file are merged with those reported in the personal level file for wave 3 of the WAS. Each individual is then allocated to a benefit unit, with each benefit unit comprised of a single adult or cohabitating couple and their dependent children. All individuals under age 16, or under age 19 and full-time students, are identified as dependent children. Age, relationship status (single/couple), and the number and age of all dependent children in the benefit unit are evaluated.

Indicator variables are evaluated for each adult to distinguish those with graduate education, and those currently enrolled in a tertiary level education.

All earnings from employment are evaluated for each adult, and identifiers calculated to distinguish self-employed, full-time employees, part-time employees, and unemployed. Furthermore, identifiers are calculated to distinguish those with non-contributory pension schemes (predominantly public sector employees), those eligible for a (contributory) occupational pension scheme, members of occupational pensions, and whether defined benefit and defined contribution pensions are held. For those who are identified as members of an occupational pension, their private contribution rate to

[16] Households in the top 10% of the wealth distribution were designed to have three times the likelihood of sampling than those in the remainder of the population.

[17] The distribution of WAS

the pension is also evaluated. Furthermore, the value of all state pensions currently in payment is recorded.

Regression models for log earnings are evaluated that adjust for sample selection via a Heckman correction, separately for men and women, and for individuals aged under 50 and those aged 50 and over. These regression models are calculated on the full set of data reported by wave 3 of the WAS and include, in addition to the range of data saved for loading into the LINDA model, identifiers for health status, stated preferences for saving, and housing tenure. The regression results are used to adjust the earnings of part-time employed adults to their full-time equivalents, and to impute predicted wages for those identified as not employed.

To impute a full-time wage for each adult not identified as employed in the WAS data, predicted values at the coefficient estimates for the wage equation, x_b, and the selection equation, z_g, are evaluated by STATA. These data are transferred to LINDA, along with regression estimates for sigma (the standard deviation of the residual of the wage regression) and lambda (the standard deviation of the wage regression times the correlation between the residuals of the target and selection equations). These data permitted a full-time wage to be imputed for each affected individual, after generating a random draw from a standard normal distribution.[18]

The value of savings held in Individual Savings Accounts, own businesses, property other than the main home, financial and non-financial assets and pensions are identified at the benefit unit level by aggregating up the value of each asset class held by all benefit unit members. The principal exception to this approach is the value of equity held in the main home, which is allocated entirely to the benefit unit of the household reference person.

The model is designed to track the evolving household circumstances of a sample of "reference people". Each adult aged 18 or over in the WAS is represented as a reference person of a benefit unit in the reference cross-section, so that the benefit units of couples are represented twice in the base data – once for each spouse. An indicator variable is included in the model, which identifies which reference adults are married to one-another. This is consistent with the approach taken to simulate the evolution of relationship status in the model, where marriages are considered to be between individuals represented in the simulated population; see Section 4.7 for further detail.

Two adjustments were applied to the data reported by the WAS to obtain the base data-set from which model projections are made. First, the cross-sectional weights reported by the WAS are designed to aggregate up to the 2010 principal projections for the population of Great Britain. These estimates have subsequently been revised upward in light of data reported by the 2011 Census. The cross-sectional weights reported for wave 3 data are consequently adjusted to align the aggregate weighted population reported by the WAS to ONS mid-year estimates for Great Britain in 2011.[19]

---

[18] $y_{i,t} = x_b - \lambda \,.\phi(z_g) / (1 - \Phi(z_g)) + (\sigma^2\text{-}\lambda^2)^{0.5}\varepsilon_{i,t}$ where $\phi(.)$ denotes the standard normal pdf, and $\Phi(.)$ denotes the standard normal cdf, $\varepsilon_{i,t}$ is the random draw from a standard normal distribution, and $y_{i,t}$ is the projected full-time wage.

[19] The aggregate population size for Great Britain reported by the WAS prior to adjusting sample weights is 59.3 million people, which understates the 61.4 million people in Great Britain reported in the 2011 census.

Secondly, a pseudo population for Northern Ireland was imputed, by randomly selecting observations reported by the WAS for Great Britain, structured to reflect the age, relationship status, and income distributions of Northern Ireland. This was achieved, by first analysing data reported by the Family Resources Survey (FRS) for the UK in 2011/12.

A measure of total gross income, comprised of earnings, self-employment, social benefits, pensions, investments, and other income, was extracted from the FRS for each reported benefit unit, along with the age group of the reference adult distinguished by 10 year intervals[20] Age and relationship specific quintile thresholds for total gross income were evaluated for the sample of benefit units reported to be living in Great Britain, weighting each benefit unit by their respective sample weight (GROSS3) and the number of members of the benefit unit (ADULTB+DEPCHLDB). This produced 80 mutually exclusive and collectively exhaustive population subgroupings, distinguishing 5 income-ranges for each of 8 age-bands, separately for singles and couples. The proportion of the Northern Irish population reported by the FRS as corresponding to each of the 80 population subgroups was the evaluated (reported in Appendix C).

Comparable age bands and income measures to those considered for the FRS were evaluated for each benefit unit reported by the WAS.[21] Each benefit unit reported by the WAS was then sorted into a unique age-band and relationship-specific quintile group, based on the considered measure of income. Random selections (based on the WAS household identifier) from each of the age-band/income quintile groups were then taken to match the distribution of the Northern Irish population, as calculated using FRS data.

The base data set derived as described here is comprised of 20,247 adults and 5,177 children in 13,592 benefit units. This includes the pseudo population for Northern Ireland comprised of 657 adults and 212 children in 428 benefit units. Associated weighted populations are 61,371,315 for Great Britain (the ONS mid-year population estimate), and 1,812,671 for Northern Ireland (compared with the ONS mid-year population estimate of 1,810,863).

---

[20] FRS variables BUINC and BUAGEGR3.

[21] WAS variables DVGIEMPW3, DVGISEW3, DVTOTALLBENANNUALW3_I, DVGIPPENW3, and DVGIINVW3.

6. Calibrating unobserved model parameters

Preference parameters and wage parameters were adjusted to match profiles simulated by the model to survey data. Preference parameters are unobservable, and must consequently be identified within the context of the model. Although wages are observable, associated parameters were included for adjustment within context of the model to account for selection effects. Parameters adjusted internally to the model's structure can be identified either by manual calibration or optimisation of a loss function using an econometric criterion.[22] We have elected to adjust these parameters via a manual calibration process, which we describe in this section.

The assumed preference relation (see Section 4.1) includes six parameters: relative risk aversion, γ; an exponential discount factor, δ; two parameters for the warm-glow model of bequests, $\zeta_{0/1}$; the intra-temporal elasticity, ε; and the utility price of leisure, α. In contrast, the specification adopted for wages (see Section 4.2) includes a very large number of parameters: the parameters governing wage growth, *m*; experience effects, κ; earnings volatility σ_ω^2; and the factor effects of pension take-up, λ^{ret}. Following extensive experimentation, we settled upon the following step-wise procedure to identify these various parameters.

We divided the calibrated model parameters into two sets; set *A* comprising the parameters governing wage growth and earnings volatility, and set *B* comprising all other calibrated parameters. We began by setting all wage growth parameters *m*=1, and made initial guesses for earnings volatility, σ_ω^2. Given these assumptions for set *A* parameters, and the model parameters identified exogenously from the model structure in the first stage of the analysis, we adjusted the parameters in set *B* to reflect behaviour observed at a single point in time for a reference population cross-section. Having obtained first approximations for set *B* parameters, we then adjusted the parameters *m* and σ_ω^2 to reflect historical earnings data. This procedure was then repeated until convergence in the two sets of parameters *A* and *B* was obtained. We found that it was not necessary to iterate between these two sets of model parameters more than two times to obtain convergence, a property that is attributable to the invariance of the cross-sectional population characteristics considered for adjusting parameters in set *A*, which we discuss further below.

6.1 Parameters identified on data for a reference population cross-section

All six preference parameters of the model and the factor effects of pension take-up λ^{ret}, were identified by matching the model to moments evaluated on survey data reported for a single (reference) population cross-section. This is notable, given that preference parameters are often a central focus of interest in the related literature. It is also extremely useful because it simplifies specification of the policy context underlying the behaviour considered for identification, and omits the feed-back effects that can otherwise complicate parameter adjustments.

The feed-back effects that are mentioned here complicate any empirical analysis that refers to dynamic behaviour described for an appreciable period of time. Suppose, for example, that we were interested in matching a structural model of savings and retirement to data observed during the life-

---

[22] Econometric methods include Simulated Minimum Distance (Lee and Ingram, 1991), Method of Simulated Moments (Stern, 1997), Indirect Estimation (Gourieroux et al., 1993) and Efficient Method of Moments (Gallant and Tauchen, 1996).

course of a single birth cohort. If a given set of model parameters implied savings early in the life course that over-stated observed data, then this might suggest that preferences should reflect greater impatience. Adjusting preferences in this way might then imply lower wealth later in life, and thereby influence model implications for the timing of retirement. Such feed-back effects can be ignored in an empirical analysis of household sector savings that focuses exclusively on behaviour described for a single point in time (as population characteristics such as wealth holdings are exogenously defined), which considerably simplifies the identification problem.

Our calibration of parameters identified on cross-section survey data started with the assumption of a high value for γ (=5), a high value for δ (=1), a low value for ζ_1 (=0.2) and ζ_1 (=0), a moderate value for ε (=0.5), and no retirement penalty λ^{ret} (=0). Parameterisation then proceeded in four concentric 'loops'.

(1) The inner-most loop, which was repeated most frequently, focussed on adjusting α and λ^{ret}. Increasing the utility price of leisure α tends to decrease labour supply throughout the working lifetime. Exaggerating the wage discount for benefit units that have previously accessed their private pensions tends to decrease labour supply late in the working lifetime. These two model parameters consequently provide a degree of control over the employment profile throughout the life-course, and were jointly adjusted to match the model to age and relationship-specific means for employment participation.
(2) The second loop of our calibration jointly adjusted δ and ζ_1 to reflect age and relationship specific geometric means for consumption. Increasing the discount factor δ makes benefit units more patient, and consequently tends to decrease consumption throughout the working lifetime. In contrast, exaggerating the bequest motive by increasing ζ_1 tends to lower consumption late in the life course when the probability of imminent mortality is appreciable. Taken together δ and ζ provide control over the age profile of consumption implied by the structural model.
(3) The third loop of the calibration strategy adjusted the parameter of relative risk aversion γ. γ has an important influence on savings incentives throughout the life course, in common with the discount factor δ. Raising δ ceteris paribus tends to imply lower consumption and higher pension scheme participation as benefit units are made more patient. Raising γ, in contrast, exaggerates precautionary savings motives, implying lower consumption and lower pension scheme participation (due to the illiquidity of pension wealth). Hence, if the rates of pension scheme participation implied by the model following the second loop of the calibration were too low (high), we reduced (increased) γ and returned to the inner-most loop. Otherwise we proceeded to the outer-most loop.
(4) ε was adjusted to match the model to distributional variation described by data for the ratio between equivalised consumption and leisure. If the utility maximisation problem was separable, and labour supply was a decision on a continuous domain, then the assumed preference relation would imply the following relationship between the decision variables *c* and *l* in the region of the optimum:

$$\alpha \frac{\hat{c}_{i,a}}{l_{i,a}} = \hat{h}^{\varepsilon}_{i,a}$$

where $\hat{c}$ denotes equivalised consumption and $\hat{h}$ is the equivalised post-tax and benefit wage rate. This relationship will approximately hold late in the simulated working lifetime, when benefit units exhibit substantial variation over labour supply decisions and continue to possess multiple periods over which they can choose between (discrete) labour supply alternatives. This relationship can be used to compare the decisions taken by any two benefit units, 0 and 1, as described by the ratio:

$$\left(\frac{\hat{c}}{l}\right)_1 \Big/ \left(\frac{\hat{c}}{l}\right)_0 = \left(\frac{\hat{h}_1}{\hat{h}_0}\right)^{\varepsilon}$$

This relationship indicates that increasing ε will tend to shift period specific expenditure in favour of (equivalised) consumption, relative to leisure, for benefit units with relatively high (equivalised) wage rates.

Model implications were consequently evaluated for the ratio between equivalised consumption and leisure for every benefit unit with a reference adult aged 55 to 60 in the reference population cross-section. Two separate averages were calculated over these ratios, distinguishing benefit units with and without reference adults educated to graduate level. If the value of the ratio of the graduate average divided by the non-graduate average was too low (high), then ε was increased (decreased). The calibration then proceeded back to the inner-most loop, and the entire process repeated until a convergence was obtained.

Conceptually, the strategy described here (implicity) assumes that the state variables of the dynamic programming problem are sufficient statistics to account for the full historical background of the control (flow) variables considered for empirical identification. Any modelling structure of the life-course necessarily involves a compromise between computational burden and descriptive detail, and the model described here is no exception. A form of omitted variable bias arises where the set of state variables included in the dynamic programming problem are insufficient to capture the circumstances that influence considered control variables, and the severity of this bias will depend upon the importance of the omitted detail. The scale of this issue will consequently diminish with time, as improvements in computational power allow us to include increasingly detailed descriptions of individual specific circumstances in dynamic programming frameworks.

6.2 Parameters identified on historical earnings data

Earnings data were derived from the LCFS and its predecessor surveys. These data were used to evaluate means and variances of log earnings for population subgroups distinguished by employment, education, relationship status, age, and year. Any benefit unit with at least one adult identified as self-employed, was distinguished from households with employees only. Amongst employee benefit units, those with an adult who had continued to study full time until at least age 19 (graduates), were distinguished from others (non-graduates). Benefit units comprised of a single adult were distinguished from couples. Age and year were divided into annual increments.

Model wage parameters were adjusted to match age and year specific earnings moments reported between ages 23 and 55 for single graduate employees, between ages 23 and 60 for couple graduate employees, and between ages 20 and 60 for single and couple non-graduate employees, for all years between 1978 and 2012. The first year of this sample period corresponds to the introduction into the LCFS (FES) of the age at which individuals (aged between 16 and 69) first left full-time education. The last year of the sample period corresponds to the most recently available data at the time of writing.

Similarly, earnings parameters were adjusted to match age and year specific earnings moments reported between ages 20 and 60, and years 1972 to 2012 for the self-employed. The proportion of benefit units reported to include a self-employed individual increase discontinuously from under 2 per cent of the survey population in 1971, to just under 10 per cent of the survey population from 1972, which motivates the choice of the sample period considered for the self-employed.

The drift parameters, *m*, and the dispersion parameters, σ_{ω}^2, were calibrated against historical data by projecting the reference population cross-section backward through time. The drift parameters were adjusted to reflect geometric means of employment income. For this purpose, age specific geometric means of employment income were imputed beyond the observed time series to cover the 100 year period between 1951 and 2050, based on assumed rates of wage growth. The principal calibration assumes out of sample (real) wage growth of 1.5% p.a.. The negative calibration assumes out of sample wage growth of 1.0% p.a. for graduates, and 0.5% p.a. for non-graduates (implying average wage growth of approximately 0.6% p.a.).

As the model includes a separate drift parameter for each age, year, education and relationship combination, a close match could be obtained to the associated sample moments. Given the large number of model parameters involved, this stage of the parameterisation was undertaken using an automated procedure. First, age, year, education and relationship specific means of log employment income implied by the model under any given parameter combination were calculated from simulated panel data projected back in time for the reference population cross-section. These simulated moments were subtracted from associated sample moments estimated from survey data. The differences so obtained were then multiplied by a 'dampening factor', which is a common approach used to improve the convergence properties of any simple gradient search routine. The exponent of the result was taken, and multiplied by the prevailing drift parameter to obtain an updated value for the parameter. The model was then re-run using the updated wage parameter and the process repeated until the average absolute variation of parameters over ages for any year, education, and relationship combination fell below 10 percentage points.

Similarly, the variance parameters were adjusted to reflect age, year, and relationship specific variances of log employment income calculated from survey data. Unlike the drift parameters, however, only six parameters -- distinguish singles from couples, and self-employed from graduate employees and non-graduate employees -- were adjusted to reflect the dispersion of employment income. These model parameters were adjusted manually.

7. Reflecting reforms introduced in the 2015 Budget Statements

We envisage a sensible starting point for using LINDA to explore the implications of policy alternatives will involve the following:

1) Project the reference population cross-section forward and backward through time, using the base parameters supplied with the model. As discussed in Section 4, these parameters are structured around transfer policy that was effective in 2011.
2) Create a base population using the simulated data generated by (1).
3) Starting from (2), project the population forward through time from 2016, using a model specification adapted to reflect the most recently announced policy environment.
4) Create a base population using the simulated data generated by (3).

Policy counterfactuals could then be projected, taking the simulated specification defined under (4) above as the base-case.

LINDA has been structured to facilitate the steps defined above, by introducing a range of policy initiatives announced as part of the 2015 Budget Statements. Section 7.1 provides details of each aspect of the revised policy environment that have been included in the model. A number of announcements were made that affect thresholds and uprating rates. LINDA accommodates a full parameterisation for the selected tax year (2016 in this case), and includes 6 uprating rates to govern evolution of the system through time. Section 7.2 provides some guidance concerning how alternative parameter options available in the model can be used to approximate the commitments made by Government. Finally, some announcements from the 2015 Budget Statement fall outside the scope of the existing model, and have consequently not been included; these reforms are listed in Section 7.3.

Please note that the full set of tax and benefit rates, allowances, premia etc. for 2016/17 were not publically available at the time of writing. Where a policy parameter could not be found, rates applicable for 2015/16 have been included by default for the model. It is recommended that these parameters be updated as relevant information becomes available.

7.1 Accommodated Budget announcements

Pay to Stay – The government will require high income social tenants to be charged a market or near market rent, with the additional rental income raised by Local Authorities to be returned to the Exchequer. The government will consult on how this is implemented. (47)

- Details concerning how this reform will be implemented will only emerge after the consultation. The model has consequently been extended in a way that provides a high degree of flexibility.

From September 2017, the government will extend the free childcare entitlement to 30 hours a week for working parents of 3 and 4 year olds. (6)

- The model now allows for childcare costs to be differentiated by age (0-2;3-4;5-13) and by labour supply of the second earner (part time; full time). (Column AS of JobFile). The user can therefore apply discounts for ages 3-4 to reflect the incidence of the Free Childcare Entitlement.

National Living Wage – The government will introduce a new premium for those aged 25 and over starting at 50 pence leading to a new National Living Wage (NLW) of £7.20 in April 2016. The government's ambition is for the NLW to increase to 60% of median earnings by 2020, and it will ask the Low Pay Commission to recommend the premium rate in light of this ambition going forward. On OBR forecasts, this means the NLW is expected to reach the government's target of over £9 by 2020.

- The model now allows for the NMW to be differentiated across up to 4 age bands. £7.20 is 62% of median hourly earnings in 2014. It will be somewhat lower than this in 2016, but it is already not far off from the target of 60%. We suggest setting the NLW to £7.20 in 2016 and setting further uprating in line earnings.

Personal savings allowance – As announced at March Budget 2015, the government will introduce an allowance from 6 April 2016 to remove tax on up to £1,000 of savings income for basic rate taxpayers and up to £500 for higher rate taxpayers. Additional rate taxpayers will not receive an allowance. Automatic deduction of 20% income tax by banks and building societies on non-ISA savings will cease from the same date and the government will shortly publish a public consultation on whether changes are required to the deduction arrangements in place for other savings income. (Finance Bill 2016)

- We have implemented a script within the module that computes taxes and benefits, which assigns a tax free Personal Savings Allowance (PSA) as either fully exempt (Y<16k), Basic rate tax payer with PSA 1000, or higher rate tax payer with PSA 500.[23] The PSA is subtracted from investment income (up to the full investment income) when calculating taxable income. Any excess investment income is taxed at each individual's marginal rate. Associated parameters can be found in cells (35,170) to (35,175) of the Job File.xls, which can be altered through the Excel front-end (when the 2016 policy structure is activated).

Insurance premium tax standard rate – From 1 November 2015, the standard rate of insurance premium tax (IPT) will be increased by 3.5 percentage points to 9.5%. From this date all premia received by insurers using the IPT cash accounting scheme will be charged at 9.5%. For insurers using the special accounting scheme, there will be a 4 month concessionary period that will begin on 1 November 2015 and end on 29 February 2016, during which premia received that relate to policies entered into before 1 November 2015 will continue to be liable to IPT at 6%. From 1 March 2016 all premia received by insurers will be taxed at the new rate of 9.5%, regardless of when the policy was entered into. (Summer Finance Bill 2015) (19)

- Cell (73,11) increased from 0.06 to 0.095. The change is legislated for 2015 but we will simulate it from 2016.

Lifetime Allowance for pension contributions – The government will reduce the Lifetime Allowance for pension contributions from £1.25 million to £1 million from 6 April 2016. Transitional protection for pension rights already over £1 million will be introduced alongside this reduction to ensure the change is not retrospective. The Lifetime Allowance will be indexed annually in line with CPI from 6 April 2018. (Finance Bill 2016)

- The model does not impose a Lifetime allowance on pension contributions, but does impose an upper limit on the value of aggregate pension wealth. This is set to £1,250,000 for each adult in the household by default. The lack of sophistication here is balanced against the fact that this measure only affects a small fraction of the simulated population.

[23] See SUBROUTINE CalcTax() in file UK_4.f90 in the TAXES.sln that is supplied with LINDA.

Changes to taper rates in tax credits – From April 2016 the taper rate in tax credits will be increased from 41% to 48% of gross income. Once claimants earn above the income threshold in tax credits, their award will be withdrawn at a rate of 48 pence for every extra pound earned. (41)

- This reform involved altering pre-existing model parameters (see cells (35,31) (35,33) (35,41) and (35,43) of the Job File.xls).

Removing the Family Element in tax credits, the first child premium in Universal Credit and the Family Premium in Housing Benefit – From April 2017, the Family Element in tax credits and the equivalent in Universal Credit will no longer be awarded when a first child is born. This will also apply for families with children making their first claim to Universal Credit. Households who have been in receipt of tax credits or Universal Credit with an interruption of less than 6 months will be protected. Furthermore, children with disabilities will continue to receive the Disabled Child Element or Severely Disabled Child Element in tax credits and the equivalent in Universal Credit. In Housing Benefit, the family premium will be removed for new claims and new births from April 2016. (40)

- This reform involved setting the family premia to zero in the model (job file.xls, cells (35,37) and (35,50)). The model ignores protected rights in this regard. Furthermore, although the measure is set to apply from 2017, it is modelled from 2016.

Limit Child Element in tax credits and Universal Credit – The Child Element of tax credits and Universal Credit will no longer be awarded for third and subsequent children born after 6 April 2017. This will also apply to families claiming Universal Credit for the first time after April 2017. Households who have been in receipt of tax credits or Universal Credit, with an interruption of less than 6 months, will be protected. Furthermore, children with disabilities will continue to receive the Disabled Child Element or Severely Disabled Child Element in tax credits and the equivalent in Universal Credit. Multiple births will be protected in both systems. The Department for Work and Pensions and HMRC will develop protections for women who have a third child as the result of rape, or other exceptional circumstances. Consequential changes will be made in Housing Benefit from April 2017. (39)

- An additional model parameter (cell (35, 110)) now determines the maximum number of children that receive the "child element" of benefits. This parameter influences the revised tax code that can be accessed as part of the TAXES.sln that is supplied with the model. Although the measure is effective from 2017, it is assumed to apply from 2016 in the model.

Lowering the household benefit cap – The government will lower the household benefit cap, which caps the amount of benefits out-of-work working-age families can receive, to £20,000, except in Greater London where the cap will be £23,000. The current exemptions to the cap will continue to apply. (38)

- A benefits cap has been implemented in the revised tax code. At present, this cap is not differentiated by Government Office Region (as this would require the GOR to be included in the state space of the model).

Restricting Housing Benefit entitlement for young people – From April 2017, those out of work aged 18 to 21 making new claims to Universal Credit will no longer be automatically entitled to the housing element. Parents whose children live with them, vulnerable groups, and those who were living independently and working continuously for the preceding 6 months will be exempt from this measure.

- An age restriction for housing benefits has been included in the revised tax code.

Changes to tax credits income thresholds and Universal Credit work allowances – From April 2016 the income threshold in tax credits will be reduced from £6,420 to £3,850 per year. Work allowances in Universal Credit will be abolished for non-disabled childless claimants, and reduced to £192 per month for those with housing costs and £397 per month for those without housing costs. Claimants earning below these amounts will retain their maximum award. (42)

- This reform involved altering associated pre-existing model parameters governing the receipt of Universal Credit

Personal allowance increase – The government will increase the income tax personal allowance from £10,600 in 2015-16 to £11,000 in 2016-17.

- This reform involved altering associated pre-existing model parameters

Higher rate threshold increase – The government will increase the higher rate threshold from £42,385 in 2015-16 to £43,000 in 2016-17. The NICs Upper Earnings Limit will also increase to remain aligned with the higher rate threshold. (Summer Finance Bill 2015) (2)

- This reform involved altering associated pre-existing model parameters

7.2 Uprating and the Budget announcements

The Government has made a number of announcements broadly related to 'uprating', with reference to the prospective 4 years. The model includes a set parameters that permit a range of factors – including tax thresholds, housing costs, child care costs, benefits – to growth through time at different (time invariant) rates. Care must be taken to select these uprating parameters in a way that is consistent with the period of interest. If the period of interest is very short (<10 years), uprating rates can be set fairly freely. If the period of interest is long, different uprating rates may generate odd incentive structures, which may adversely affect behavioural projections. A benefits freeze over 4 decades with positive income growth, for example, could marginalise the welfare state.

This section discusses the implications of the 2015 Budget announcements for model parameters assumed for forward projections.

The Government made a number of announcement limiting or freezing benefits in the near term:

- The welfare cap is set for the Parliament at the level of the OBR's forecast for welfare spending. The cap covers welfare spending excluding the state pension and automatic stabilisers. The forecast margin will be 2% of the cap in each year that the cap applies.
- Benefits uprating – Most working-age benefits will be frozen for 4 years from April 2016.
    - This will apply to: Jobseekers' Allowance; Employment and Support Allowance; Income Support; Child Benefit; applicable amounts for Housing Benefit; and Local Housing Allowance rates, with provision for high rent areas.
    - This excludes: Maternity Allowance; Statutory Sick Pay; Statutory Maternity Pay; Statutory Paternity Pay; Statutory Shared Parental Pay; and Statutory Adoption Pay; disability, carers and pensioners' premia in the frozen benefits; the Employment and Support Allowance Support Group component; and other disability, carer and pensioner benefits, which will continue to be uprated in relation to prices or earnings as applicable. (37)
- Tax credits uprating – The uprating freeze will extend to the Child Tax Credit and Working Tax Credit (excluding disability elements). All disability elements will continue to be uprated by prices each year. (37)

The Government also made a number of announcements in relation to tax thresholds:

- Personal allowance – will increase to £11,200 from 2017-18. (Summer Finance Bill 2015) (1)
- Personal allowance indexation change – The government will legislate to ensure that once the personal allowance reaches £12,500 it will be uprated in line with the National Minimum Wage (NMW), ensuring that anyone on the NMW working 30 hours per week or less, does not pay income tax. Until then, the government will also have a legal duty to consider the impact of the level of the personal allowance on an individual working 30 hours a week on the NMW and to report on this at each fiscal event. To underpin this, the government has also announced a review of the NMW timetable to align it with the tax year. (Summer Finance Bill 2015)
- Higher rate threshold increase – The government will increase the higher rate threshold to £43,600 in 2017-18. The NICs Upper Earnings Limit will also increase to remain aligned with the higher rate threshold. (Summer Finance Bill 2015) (2)

The rates of increase of tax thresholds implied above are roughly in line with inflation. This means a 0 real rate of growth. The NMW has tended to grow with average earnings. Pegging the Personal Allowance to the NMW would imply a positive real growth. However, as this kicks-in when the PA reaches £12,500, it may occur beyond the forecast horizon at the growth rates suggested above. An assumption of 0 real wage growth may be appropriate for very near term analysis.

Given the above, uprating parameters in the near-term (to 2020) have been defined as follows:

- NMW: growth between 2011 and 2020 adjusted to meet targets for the living wage.
- Tax thresholds: wage growth
- Benefits: frozen (-1% real, assuming price growth of 1% p.a. between 2011 and 2020)
- Pension Contribution Thresholds: wage growth

Over the long-term, there is likely to be closer alignment between wage growth and the transfer system. If this were not the case, there would be a secular shift in the relative weight of taxation or welfare compared to earnings. This is unlikely to be what the user intends to model.

For long-term analysis, the relevant uprating parameters have therefore be set as follows:

- NMW: wage growth
- Tax thresholds: wage growth
- Benefits: price growth
- Pension Contribution Thresholds: wage growth

7.3 Budget announcements not reflected in LINDA

As noted above, a range of policy initiatives included in the 2015 Budget Statements were not included in the model. This section lists omitted initiatives, along with a brief explanation for the omission.

- The government will fund public sector workforces for a pay award of 1% for 4 years from 2016-17 onwards.
    - The current parameterisation does not distinguish public and private sector workers

- The government will reduce rents paid by tenants in social housing in England by 1% a year for 4 years from 2016
    - This is a transitional change over the short-term, so unlikely to be relevant in a model designed for long term projections.
- Lifetime tenancies – The government will review lifetime tenancies in social housing to limit their use. The review will seek to ensure that households are offered tenancies that match their needs, and make best use of the social housing stock.
    - Lifetime tenancies not included in the model
- Increasing the employer National Insurance contributions Employment Allowance from £2,000 to £3,000. (13)
    - Employer NICs, like company taxes, are not included in the model.
- National Insurance contributions Employment Allowance; from April 2016, companies where the director is the sole employee will no longer be able to claim the employment allowance. (25)
    - The current parameterisation does not distinguish the employees and the self-employed
- Making ISAs more flexible – March Budget 2015 announced that the government will change the ISA rules in the autumn to allow individuals to withdraw and replace money from their cash ISA in-year without this replacement counting towards their annual ISA subscription limit. This policy will also cover cash held in stocks and shares ISAs. These changes will commence from 6 April 2016.
    - Intra-year allocations are beyond the scope of the current parameterisation
- Taxation of pensions at death – As announced at Autumn Statement 2014, the government will reduce the 45% tax rate that applies on lump sums paid from the pension of someone who dies aged 75 and over to the marginal rate of the recipient from 2016-17. (Summer Finance Bill 2015)
    - Inheritances are not simulated in forward projections
- Inheritance tax and the main residence nil-rate band – The government will introduce an additional nil-rate band when a residence is passed on death to direct descendants. This will be £100,000 in 2017-18, £125,000 in 2018-19, £150,000 in 2019-20, and £175,000 in 2020-21. It will then increase in line with CPI from 2021-22 onwards. Any unused nil-rate band will be transferred to a surviving spouse or civil partner. It will also be available when a person downsizes or ceases to own a home on or after 8 July 2015 and assets of an equivalent value, up to the value of the additional nil-rate band, are passed on death to direct descendants. This element will be the subject of a technical consultation. There will also be a tapered withdrawal of the additional nil-rate band for estates with a net value of more than £2 million. This will be at a withdrawal rate of £1 for every £2 over this threshold. (Summer Finance Bill 2015, Finance Bill 2016) (3)
    - Inheritances are not simulated in forward projections
- Inheritance tax and the nil-rate band – The inheritance tax nil-rate band is currently frozen at £325,000 until April 2018. The government will continue to freeze the nil-rate band at £325,000 until April 2021. (Summer Finance Bill 2015) (3)
    - Inheritances are not simulated in forward projections
- Employment and Support Allowance – From April 2017 new claimants of Employment and Support Allowance who are placed in the Work-Related Activity Group will receive the same

rate of benefit as those claiming Jobseeker's Allowance, alongside additional support to help them take steps back to work. (50)

  - The model focuses exclusively on Jobseeker's Allowance in this regard, so the reform will actually provide a closer match with the model structure

- Tax-free childcare – As announced on Wednesday 1 July, the government reaffirms its commitment to introducing Tax-Free Childcare to support working parents with the costs of childcare. Due to a legal challenge, Tax-Free Childcare will now be launched from early 2017. The government will hold the existing scheme, Employer Supported Childcare, open to new entrants until the new scheme is introduced. (7)
  - The model currently assumes net childcare costs vary by age of child and labour supply status of the second earner. Conditional on these, it does not model heterogeneity. Implementing this reform, without overhauling the current modelling approach for childcare would fail to capture the heterogeneous distributional and incentive effects of the policy. At the same time, introducing a more sophisticated modelling approach would be beyond the scope of this project. Finally, the legal uncertainty around the policy adds to the arguments against modelling this explicitly at this stage. It is recommended that this be captured in the existing model framework – if desired – by suitable adjustment of the (exogenously imposed) costs of childcare.
- Dividend taxation – The government will abolish the Dividend Tax Credit from April 2016 and introduce a new Dividend Tax Allowance of £5,000 a year. The new rates of tax on dividend income above the allowance will be 7.5% for basic rate taxpayers, 32.5% for higher rate taxpayers and 38.1% for additional rate taxpayers. (Finance Bill 2016) (16) (26)
  - The model currently distinguish interest from dividend income.
- Pensions: reduced Annual Allowance for top earners – The government will restrict the benefits of pensions tax relief for those with incomes, including pension contributions, above £150,000 by tapering away their Annual Allowance to a minimum of £10,000. This policy will come into effect from April 2016. (Summer Finance Bill 2015) (4) The Annual Allowance (maximum annual amount of pension contributions that can attract tax relief) is set to £40,000 for everyone in the base specification of the model. The new measure will taper this amount by 50p for every pound of income above £150,000, down to a floor of £10,000.
  - This reform could be implemented, but has been omitted in the current specification

Projecting the Evolving Population Cross-section Forward Through Time

LINDA has been adapted to project the evolving population cross-section forward through time, which is necessary to generate the implications of policy alternatives for the aggregate government budget. This involved altering the way that the model simulates marriage, including the effects of migratory flows, and allowing for the "birth" of young cohorts not represented in the reference population cross-section. Given these innovations, the time horizon over which the evolving population cross-section is projected is now defined by the total number of years that the model projects forward through time and the maximum number of simulated adults (whichever condition is the more restrictive). We begin by discussing how the simulation of marriage has been altered, before discussing how migration has been included in the model, and finally describe the methods used to introduce younger birth cohorts.

8. Simulating marital status

There are two principal approaches to simulating marital status in dynamic microsimulation models. The first assumes that the partners of newly married reference adults are drawn from outside the simulated population, and the second identifies married partners from within the simulated population.

The advantage of the first approach is that it does not require a matching routine for identifying which adults within the simulated population marry one another – the characteristics of a spouse are generated at the time of marriage, and then leave the model sample frame at the time of marital dissolution. This is the approach that has been used in LINDA prior to the current project. However, drawing the characteristics for each simulated spouse from outside the simulated population introduces important complications when applied to a framework that projects the evolving population cross-section through time.

Suppose, for simplicity, that a simulated population provided a perfect reflection of a population cross-section at a given point in time, and consider the implications of simulating marital status forward one year. Each individual who was projected to marry would increase the number of adults in the population cross-section by one, and each individual who was projected to experience a marital dissolution would reduce the population by one. Variation of the simulated population size attributable to these fluctuations in marital status could be suppressed by adjusting the population weight of each simulated benefit unit in inverse proportion to the number of adult members of the benefit unit. But these changes to population weights would affect other population aggregates, including the numbers of children, wealth holdings, and so on. It is possible to assume that individuals marry identical copies of themselves, which would suppress variation of "other" population aggregates. Unfortunately, this last assumption is not very plausible.

The model structure has consequently been altered to identify married partners from within the simulated population. This approach has the appeal that it bears close similarities with the way that benefit units are organised in practice. It does, however, introduce two key challenges. First, how to combine the circumstances of benefit units that are associated with different survey weights.

Secondly, how to match benefit units that marry. The first of these issues is the simpler of the two and we discuss it first, before moving on to the second.

8.1 Population weighting

The survey data that are supplied with the WAS include individual weights designed to aggregate the sample population up to the population cross-section. The WAS survey weight for wave 3 data (w3xswgt) is a continuous variable, that takes values between 68.898 and 9999.0, with a mean of 1206.563: the associated histogram is reported in the figure below.

Two benefit units can only be combined simply into a single married unit that leaves population aggregates unaffected if they are associated with the same population weights. This is in contrast to the population heterogeneity in sample weights supplied by the WAS, as reported above. The model has consequently been structured to construct an x% sample of the UK population, where x can be freely selected, subject to available computing resources (both RAM and hard-disk space).

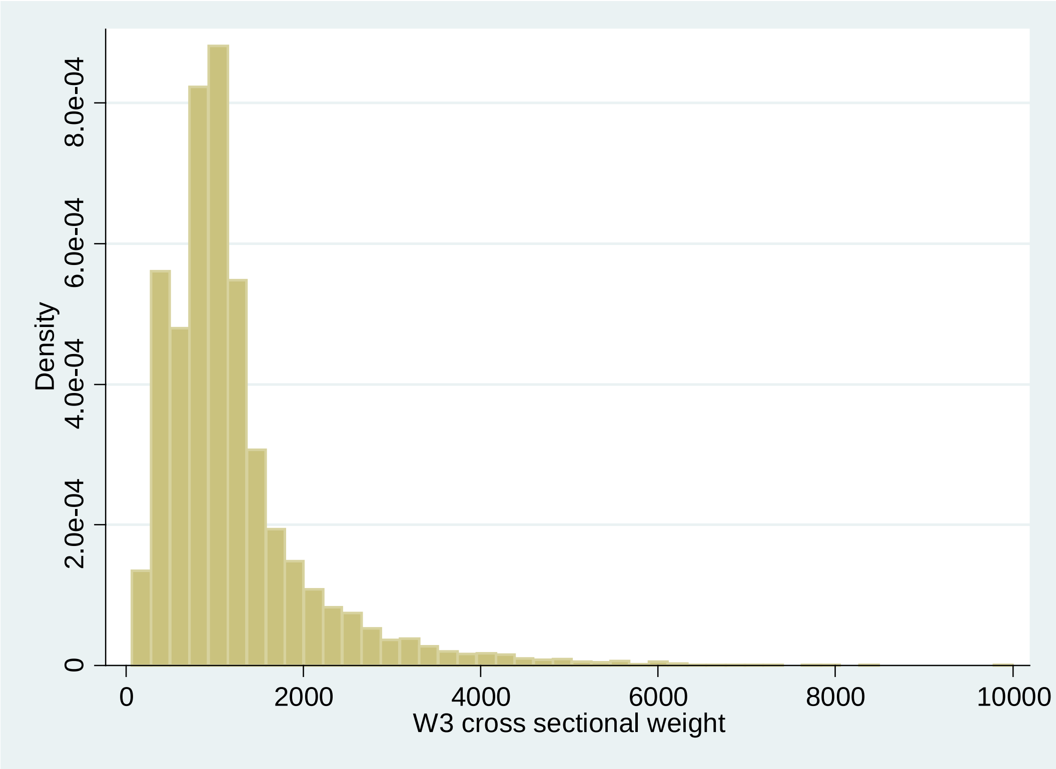
The base population cross-sectional data that are loaded into the model describe all of the characteristics of each reference adult, and their spouse if one exists (see Section 5 for details). These data include a population weighting variable, as defined by w3xswgt above. The model now also includes a parameter that defines the number of individuals in the population cross-section represented by each simulated person; denote this variable *y*. The first task that the model performs after it loads in the base population cross-sectional data is to construct a simulated population in which each reference unit has an equal weight (equal to *y*).

The data for each reference adult in the base population cross-section are considered in turn. First, the sample weight (w3xswgt) is divided by the simulation weight (*y*), and the result is rounded to the nearest integer. This term is referred to as the "replication factor", *n*. The data for the reference adult are then copied into the simulated population, and replicated *n* times. Any reference adult with a replication factor equal to 0 is dropped from the simulation.

A smaller value for the simulation weight, *y*, implies fewer surveyed adults are omitted from the simulated population, at the expense of increasing the simulated population size. This trade-off is displayed in the figure below. The figure indicates that no survey adults are dropped from the considered sample when *y* = 500, in which case the number of simulated reference adults is 99,560. Increasing *y* to 1500 reduces the simulated number of reference adults to 32,900 at the expense of dropping 8 per cent of the survey sample. Given the statistics reported in the figure, we have set *y* = 1000 by default, resulting in a simulated population of 50,241 reference adults, and 1.4 per cent of sampled adults dropped from the analysis (equivalent to 0.2% of the weighted sample).

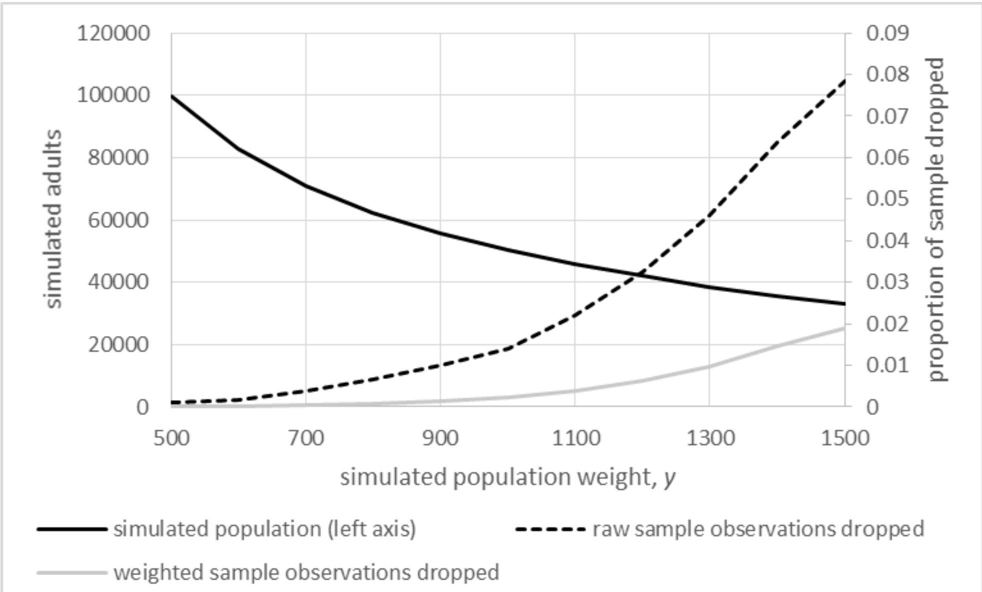
The bearing that the above adjustments to population weights have on the projected size of the aggregate population depends upon the skew of rounding of the weighting factors that are loaded into the model. Analysis reveals that rounding will generally inflate the implied size of the UK population, given the distribution of weights supplied with wave 3 of the WAS. Specifically, the weighted size of the UK population cross-section loaded into the model, without rounding, is 63,183,986, as discussed at the end of Section 5. The rounding of weights implied by a simulated population in which each adult represents 1000 individuals as described above, inflates the projected UK population size to 63,728,000 (comprised of 50,616,000 adults and 13,121,000 children).

Figure 8.1: Histogram of sample weights reported for Wave 3 of the Wealth and Assets Survey



Source: Authors' calculations on data for variable w3xswgt from wave 3 of the Wealth and Assets Survey.

Figure 8.2: Impact of simulation population weight on numbers of simulated adults and sample omitted from analysis



Source: Authors' calculations using data reported for variable w3xswgt from wave 3 of the Wealth and Assets Survey

8.2 Matching married reference adults

Marriage is defined in the model to reflect relationships recognised by the benefits system, consistent with our focus on benefit units. Two new variables have been added to the model that are designed to keep track of married couples. “Psnno” defines a unique person identifier for each simulated adult. “ben_unit” defines a unique benefit unit identifier for each simulated family unit. Ben_unit is set equal to the psnno of the benefit unit reference person.

The population cross-section data that are loaded into the model have been extended to include values for psnno and ben_unit, as defined above. These links between couples are maintained when the population is replicated to generate common simulation weights for all simulated adults, as described in Section 8.1. Consider, for example, two individuals, where psnno = (1001,1002) and ben_unit = 1001. Weights in the WAS are supplied at the household level, and so two adults in the same household are subject to the same sample weight. This is useful, as it implies that adults in a cohabitating relationship are subject to the same replication factor *n*. Suppose that the replication factor for our two adults is 5. Then the model starts by cloning individuals 1001 and 1002 five times over. Each of these clones is assigned a unique psnno, and the ben_unit identifier is re-set for each cloned couple to the new person number assigned to the clone of individual 1001.

An important feature of the approach taken to simulate marital status in the model is that it is exogenous of the decision process. This allows marital status to be simulated for the entire life-course, prior to simulation of other characteristics, including labour supply, consumption, and investment decisions. This form of exogeneity is also true of a number of other simulated characteristics in the model, including survival (section 4.9), internal migration (section 9.1), graduate status, and whether attending tertiary education. We refer to these characteristics as being “simulated exogenously”.

The exogenously simulated characteristics are ordered hierarchically. The first characteristic in this hierarchy is survival, which depends only on age and year. This characteristic is simulated first for all adults. Next, internal migration is simulated, which depends only on age, survival, and current geographic region. Education and student status are then simulated, which depend upon age, survival, and current education/student status. Marriage and fertility are then simulated, which depend upon age, year, survival, geographic region, education, student status, and current marriage status and dependent children. Finally, immigration and emigration are simulated, which depend upon age, year, survival, region, education, marital status, dependent children, and past migrant status. In this section we discuss the methods used to simulate marriage.

The model is designed to permit the entire life course to be simulated for each adult present in a reference population cross-section, which requires projecting individual circumstances both forward and backwards through time. The model has been extended under the current project to project the evolving population cross-section *forward* through time. We have consequently adjusted the way that we simulate relationship status in forward projections, retaining the methods used to simulate relationships backward through time (on which, see the technical report).

Model parameters include probabilities of marriage formation and divorce, distinguished by age, year, and education status. Prior to the current project, omission of explicit consideration of partner characteristics beyond periods of marriage permitted divorce and widowhood to be treated identically – and not distinguished separately – when simulating marital dissolutions. The shift to draw spouses from within the simulated population requires that divorce be distinguished from widowhood, and the probabilities of marriage have consequently been re-specified to focus on divorce (as described in Section 4.7). Widowhood, in contrast, is projected with regard to simulated mortality, which is also an exogenously projected characteristic as noted above.

Matching of spouses is achieved within the model by projecting marital status, dependent children, and migration jointly, one year at a time, for the evolving population cross-section. First, each adult in the evolving cross-section is considered in turn, including new migrant entrants and dependent children who mature to adulthood in the prevailing year by reaching 18 years of age (see Section 10), omitting adults who die or emigrate. Marital status and fertility are then simulated for the existing resident population. Finally, the pool of individuals who are identified as marrying in the prevailing year is sorted into cohabitating benefit units.

The sorting routine matches spouses using a "points system", with reference to each individual's age, geographic region, and education status. Each newly married individual is considered in turn. First, the model lists the population of newly married individuals in no particular order, from 1 to *n*. Starting with the first listed individual, the model identifies their age, education status, and geographic region. The model then considers the suitability of individual 2 in the list as the spouse for individual 1. It does this by assigning one "point" for each year difference in age, three points if the two individuals are identified as living in different geographic regions, and five points if the two individuals are identified as holding different education levels. The model aggregates the points together, stores the result, and then passes to individual 3 in the list. The model evaluates the points associated with a union between individuals 1 and 3. If the aggregate is less than the points associated with a union between individuals 1 and 2, individual 3 is adopted as the preferred spouse for individual 1. The model then passes to individual 4 in the list, and continues its search until it either identifies a perfect match with individual 1 (zero points in aggregate), or reaches the end of the list. If it reaches the end of the list, the best match to individual 1 (minimum points in aggregate) is adopted as individual 1's spouse, and the two adults are assigned the same benefit unit number, equal to the person identifier for individual 1.

The model proceeds by checking whether individual 2 has been matched to individual 1. If not, then it searches for a match to individual 2 in the same fashion as described for individual 1, with the addition that it first checks whether each individual considered as a potential spouse for individual 2 has already been assigned as a spouse to individual 1. The model works through the entire pool of newly married individuals in this way. If an odd number of newly married individuals is identified by the model, then the last individual in the list is assumed to remain single.

Having matched spouses, the model must simulate endogenous benefit unit characteristics, including consumption and labour supply, jointly for cohabitating adults. It does this by first identifying a "reference adult". The reference adult is set to the individual who has the highest human capital at the time of the marital union. All endogenous characteristics of a married couple are then simulated with reference to the circumstances of the benefit unit reference person, including their age, birth cohort, education, and jointly held assets.

At the time of the union, jointly held assets are the sum of the assets held individually by each spouse. At the time of divorce, jointly held assets are assumed to be divided evenly between the separating spouses. In the event that one spouse dies, then all assets are assumed to be inherited by the surviving spouse.

9. Migration

The model distinguishes between internal migration between UK regions and international migration. The methods used to parameterise and simulate each type of migratory flow are discussed below.

9.1 Internal migration

LINDA currently does not distinguish the geographical location of any simulated benefit unit, beyond the basic design of the simulated sample to reflect the UK population cross-section observed in 2006. This is likely to be important for an analysis of poverty, and distributional analyses more generally, due to regional variation over earnings opportunities and living costs. In some cases – e.g. London relative to Wales – these differences can be large.

Although this issue was noted in our feasibility report to JRF in 2014, it was subsequently agreed that distinguishing benefit units by geographic location would not be included as a formal extension of the model under the current project. This decision was taken due to the non-trivial increase in computational burden associated with allowing for geographic variation, and the increase in the number of parameters required for model calibration. Nevertheless, the approach that we have taken to accommodate external migration in the model – which is a necessary model adaptation to permit projections for the evolving population cross-section – has been designed as far as is practical to facilitate regional analyses.

The WAS includes a range of geographic identifiers, including Government Office Regions (GOR), Local Authorities, and urban/rural identifiers. We have chosen to accommodate Government Office Regions in the model, of which there are 12 (including one region for Northern Ireland, imputed into the WAS as described in Section 5).[24] Furthermore, we assume that each benefit unit's GOR affects only their likely GORs in subsequent time periods, and does not influence the temporal evolution of other characteristics simulated by the model. This approach is designed to off-set the two principal difficulties associated with accommodating regional variation that are cited above.

Limiting the distinction to geographic tracking allows this characteristic to be introduced at a negligible computational cost, as this suppresses any bearing that geographic location might have on behaviour. Furthermore, the approach limits the number of additional parameters that are required for the model to a single set of transition probabilities describing the likelihood of moving from one GOR to another. The result is a model that can be used to consider differences in poverty rates between UK regions, but only to the extent that such differences exhibit persistence from the reference population cross-section.

The transition probabilities used to simulate internal migration were calculated using data reported by the ONS, and which are publically available through the ONS website at the time of writing. Two principal data sources were referenced. The first is an ONS "research series", which reports the number of individuals migrating from one UK Local Authority to another, by sex and single year of

[24] The Government Office Regions are: North East, North West, Yorkshire and The Humber, East Midlands, West Midlands, East, London, South East, South West, Wales, Scotland, and Northern Ireland. The definition of Government Office Regions follows the ONS categorisation as at 1 January 2011.

age during the year ending June 2011.[25] These data were used to calculate the total number of individuals migrating from any one GOR to each of the 11 alternative GORs, for five year age bands.[26] The second data source reports the number of usual residents in five year age bands for each GOR evaluated from the 2011 census data. The required transition probabilities were obtained by dividing the population flow data calculated from the first data source, by the population stock data reported by the second data source. These statistics are supplied to the model via the worksheet entitled "internal migration", in the Excel file MIGRATION PARAMETERS.xls, which is provided in the model subdirectory entitled "\base_files\UK2011_all".

Internal migration forward through time is simulated by organising GORs, and numbering them from 1 to 12. A random draw from a uniform distribution is then generated for each reference adult. If the random draw is less than or equal to the age and region specific probability of migrating to GOR 1, as described by the transition probabilities referred to above, then the individual's benefit unit is assumed to migrate to GOR1 at the very beginning of the succeeding year. Otherwise, if the random draw is less than or equal to the aggregate of the probabilities of migrating to GORs 1 and 2, then the individual's benefit unit is assumed to migrate to GOR2 at the very beginning of the succeeding year. This procedure is repeated for all 12 GORs, where migration to GOR x is generated if the random number is greater than the sum of moving to GORs 1 to (x-1), but less than or equal to the sum of moving to GORs 1 to x.

A similar approach is taken to simulate internal migration backward through time. Simulating migration backward through time, however, is complicated by the fact that the transition probabilities discussed above uniquely describe the distribution of transitions for any benefit unit forward through time, but not backward through time. The indeterminacy is resolved by assuming that the population distribution during the course of the simulated period remains the same as observed in the reference year (2011 for our case).

9.2 International Migration

Prior to this project, the model omitted international migratory flows from its projections. Such flows, however, have an important bearing upon the evolving population cross-section, particularly since the UK entered the European Union. Migration has consequently been incorporated into the model as part of the current project.

As the review by O'donoghue *et al.* (2010) makes clear, there are a wide range of alternative approaches used to simulate the effects of migration in the microsimulation literature. Key modelling decisions include whether to model net migration or immigration and emigration flows separately, the variables that describe the likelihood of emigration, the approach taken to generate the characteristics of immigrants, and whether to accommodate re-entry of emigrants. These decisions depend upon the reasons for the respective model's development, and the data that are available for parameterisation.

---

[25] 352 Local Authorities are reported in the series, which comprises 1,245,836 age/sex/local authority combinations.

[26] Age bands considered for analysis are: 18-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+.

Migration has been accommodated in LINDA to meet two key objectives. First, the model should be capable of reflecting official projections for the age distribution of the population through time. Secondly, the model should reflect the bearing that contemporary trends concerning migration would have on the distribution of income if they were to continue into the projected future. Although the first of these two objectives could be achieved by modelling net migration only, allowing for both immigration and emigration helps to meet the second objective. LINDA is consequently designed to accommodate explicitly both immigration and emigration in each simulated year.

There are two key approaches to generating the characteristics of recent immigrants in a microsimulation context (see, e.g. Duleep and Dowhan, 2008). The ‘regression' based approach involves estimating a system of equations that describe all of the characteristics of interest, and to use these equations to generate the circumstances of new immigrants when they enter the model. Valid implementation of this approach is, however, exceptionally challenging in any context where more than a few characteristics need to be imputed. The alternative approach has been to initialise the circumstances of migrant entrants to the modelling frame by ‘cloning’ individuals selected from a ‘donor’ population pool. LINDA employs a cloning method for generating new migrants.

In contexts that use a cloning approach, and where the evolving circumstances of immigrants are a special subject of interest, then it is usually important to define the donor pool to focus exclusively on data reported for recent migrants. In the current model formulation, however, immigrants are included only to avoid distortions to the population cross-section that their omission would imply. As such, the model is simplified by not distinguishing immigrants from the domestic population when projecting employment and savings behaviour, and by generating the characteristics of new immigrants by cloning individuals from targeted subcategories of the domestic population. This approach allows the same basic methodology to be employed for simulating immigrants and emigrants, as described below.

The model parameters include the total numbers of immigrants and emigrants, distinguished by age and government office region, to be assumed for each prospective year. Importantly, the age distribution is exhaustive, including both the very old and very young. The model parameters also describe the proportion of emigrants who were previously immigrants, and the division of immigrants into age specific (domestic) disposable income quintiles. This set of parameters permits a series of ‘target numbers’ to be calculated, separately for immigrants and emigrants.

The model uses a procedure based on random draws to reflect the target numbers for migrants that are described above. Starting with immigration, the model selects at random and without replacement a candidate donor from the simulated domestic population. The characteristics that are used as targets for the immigrant population – namely, the numbers and age of benefit unit members, region, and disposable income quintile – are then evaluated for the candidate. The influence of the candidate on the target numbers for immigrants is determined. If recognition of the candidate as a donor benefit unit for immigration does not result in any one of the targets numbers being exceeded, then the candidate is accepted; otherwise they are rejected. New random draws continue to be taken until all targets for the immigrant population are met. Simulation of emigration proceeds in the same manner, subject to a slightly different set of targets.

The principal data source for the model parameters used to simulate migration is the statistical release for the 2012-based Principal population projections produced by the ONS. These projections are based on the UN recommendation for defining a long-term international migrant as someone who changes his or her country of usual residence for a period of at least a year, so that the country of destination effectively becomes the country of usual residence.

Data underlying the 2012-based Sub-National Population Projections (SNPP), report projected numbers of international immigrants and emigrants for local authorities in England, at annual intervals between 2012/13 to 2018/19.[27] The projections beyond 2018/19 assume that the numbers of immigrants assumed for 2018/19 remain unchanged into the indefinite future. These data are aggregated to obtain total numbers of immigrants and emigrants for the nine Government Office Regions (GOR) in England, for age bands 17 and under, 18 to 24, 25 to 34, 35 to 44, 45 to 64 and 65 and over. Finally, these data were adjusted to account for differences in the methodology used in the national population projections and the subnational population projections.[28] The age bands were selected to ensure sufficient numbers of migrants per GOR, noting that the minimum unit of measurement in the model represents 1000 individuals (as discussed in Section 5).

Obtaining similar numbers of immigrants and emigrants for the remaining countries of the UK is complicated by the fact that the ONS does not report data at the same level of disaggregation as these England. The parameters for Wales, Scotland, and Northern Ireland were calculated by adjusting age specific measures of net migration to reflect the split between immigration and emigration reported for England.[29]

The income distribution of immigrants to the UK, relative to the wider UK population, is designed to reflect survey data reported by the Family Resources Survey (FRS). Associated statistics are reported in Table 9.1.

The measure of income considered for analysis is total gross benefit unit income, equivalised using the OECD revised scale. As the sample sizes for immigrants tend to be small, data were pooled over three survey years, observed between 2010/11 and 2012/13 (year specific statistics are reported in Appendix E). A recent migrant for the purpose of analysis was defined as any adult in a benefit unit in which all adults reported having arrived in the UK from abroad, and in which the most recent arrival was within five years of the month of survey. Quintile income thresholds were evaluated for age and year specific population subgroups, omitting the benefit units of recent migrants. The relevant quintile of recent migrants was then evaluated. Table 9.1 reports the distribution of recent migrants derived by this approach, averaging across survey years.

Table 9.1 indicates that recent immigrants tend to be skewed toward the bottom of the gross income distribution. This is particularly true of immigrants aged 65 and over, who comprise a small fraction of the total immigrant population. However, there is also a clear skew toward the bottom of the gross income distribution in the other age groups considered for analysis. Amongst immigrants aged 25 to 34, for example, the proportion of migrants in the bottom income quintile is approximately

---

[27] 2012-based Subnational Population Projections. International Migration, mid-2013 to mid-2037.

[28] The year aggregates for net migration for England implied by the SNPP are systematically larger than for the national population projections, and were consequently subject to a year-specific proportional adjustment factor.

[29] Calculations based on country specific tables of assumed net migration by age for the 2012-based principal population projections, and are available from the authors upon request.

twice the proportion in the highest income quintile. These statistics are included in the set of model parameters to ensure that migrants are located appropriately within the wider UK population.

The emigration of previous immigrants is simulated based on citizenship data reported by the ONS as part of their Long-term International Migration (LTIM) series (table 2.01a). These data are displayed in Figure 9.1. This figure indicates that the proportions of all UK migrants holding British citizenship has declined over the sample period. With regard to emigrants from the UK, the decline in the incidence of British citizenship is particularly pronounced following the enlargement of the European Union in 2004. Nevertheless, data for the five years from 2008 to 2013 provide some evidence of a potential stabilisation of the share of British citizens migrating, amongst both immigrants and emigrants. These data motivate the model assumption that 60% of emigrants from the UK in each year had previously immigrated to the UK.

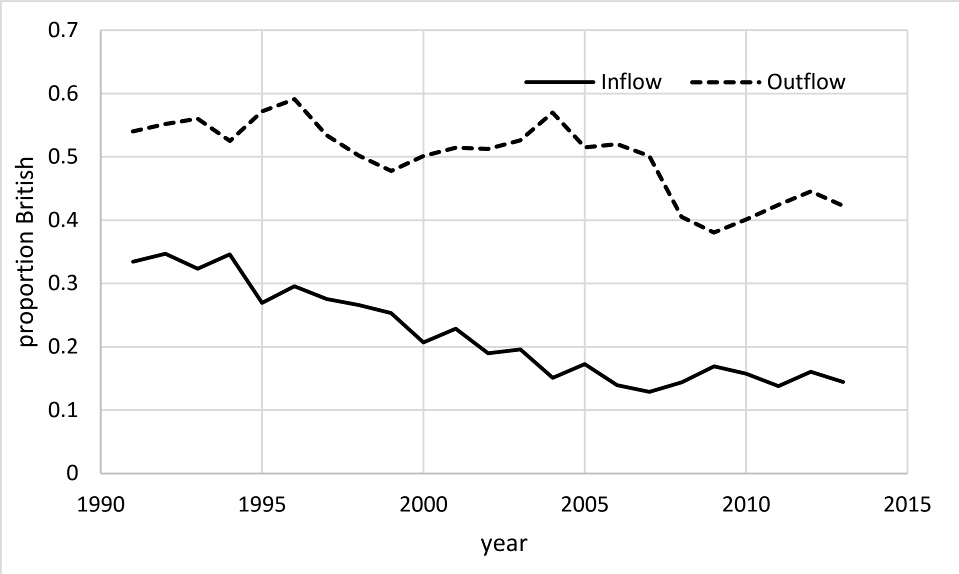
Including past immigration as a descriptive variable for identifying emigrants requires the immigration history of each simulated individual to be stored by the model. This task is complicated by the fact that the Wealth and Assets survey reports only the country of birth of an individual, and omits both their citizenship and additional detail concerning past migratory behaviour. We have elected to define all individuals present in the reference cross-section as British citizens for the purpose of the simulation. This assumption will tend to understate the average duration of stay of immigrants into the simulated population, a distortion that will decline with the projected time horizon.

Table 9.1: Distribution of recent immigrants to the UK relative to the gross equivalised benefit unit incomes of the wider UK population

age	sample	gross equivalised income quintile				
band	size	lowest	2	3	4	highest
18-24	483	0.232	0.222	0.162	0.172	0.205
25-34	1127	0.270	0.248	0.186	0.158	0.137
35-44	476	0.291	0.278	0.143	0.118	0.162
45-64	241	0.277	0.246	0.200	0.111	0.165
65+	31	0.627	0.130	0.071	0.056	0.115

Source: Authors' calculations on Family Resources Survey data pooled over 2010/11 to 2012/13 cross-sections
Notes: Total benefit unit gross income reported by variable buinc
Income equivalised using the OECD revised scale
Income quintiles evaluated by age and year on population omitting recent immigrants
Recent immigrants defined as benefit units in which all adult members reported as moving to the UK and where the most recent immigrant arrived within the 5 years preceding the month of their survey

Figure 9.1: Proportion of UK migrants holding British citizenship by year and destination



Source: Authors' calculations using data reported in Table 2.01a of the ONS Long-term International Migration series

10. The birth of young cohorts and inter-generational dynasties

Prior to our work for the JRF, LINDA could be used to project births, and the aging of dependent children. In the pre-existing model framework, however, a child was assumed to exit the simulation frame when they reached maturity (18 years of age). This aspect of the model required amendment as part of the objective to permit projections for the evolving population cross-section. Maturing children are now assumed to form benefit units of their own when they mature out of the benefit unit of their parents. This involved introducing new routines to impute characteristics for children as they pass into maturity. Following maturation, children are projected through time in an identical fashion to any individual present in the reference population cross-section.

Relative to immigrants, imputing characteristics for maturing children is facilitated by the additional detail that is afforded by the characteristics of their parents, and by the fact that maturing children have limited life experience and consequently present something of a "blank slate". This section describes the methods introduced into LINDA to impute the characteristics of maturing children.

When a child first matures, a set of characteristics are identified in a deterministic fashion:

- An individual specific *Person Identifier* (psnno) is defined.
- *Age* is set to the age of maturity, and follows directly from each individual's age as a dependent child.
- The *birth cohort* of a maturing child is defined with reference to their age and the projected year of the simulation.
- All maturing children are identified as *non-graduates* at the age of maturity
- All maturing children are identified as *private sector employees* if they choose to work at the age of maturity
- All maturing children are assumed to hold *no assets or debts* of any kind at the time of maturity. This assumption is made because the model does not account for child income, and unrequited transfers other than inheritances are not included in the model.
- *Internal geographic location* is equated to the parental region at age 18, after which it evolves as defined for the remainder of the population (independent of prospective parental transitions).
- *Survival* is assumed for all dependent children, to age of maturity.

All other characteristics are allocated, based in part on random draws:

- *Relationship status* is initialised for maturing children in the same fashion as for adults in the reference population cross-section (on which, see the technical report), with the exception that no condition is enforced for their year of maturity.[30] Associated model parameters were evaluated using ONS and LCFS data, as described in Section 4.7. Maturing children who are identified as married at age 18, are matched to spouses from the pool of newly married individual in the respective simulated year (see Section 8.2).
    - The prevailing benefit unit identifier of the maturing child (ben_unit) is defined deterministically, given their marital status and the person identifier of the individual and their spouse (if one exists).

[30] In contrast, a search routine is used to identify a set of random draws that is consistent with the circumstances reported for each adult that is defined in the reference population cross-section.

- *Tertiary student status* is allocated via a random draw from a uniform distribution, where the probability of being a student can vary by year, and parental education qualifications. The probabilities assumed for analysis were estimated from data reported for individuals aged 25 to 30 by the *Understanding Society* survey, and are discussed further in Section 14.1.
- *Wage potential* at age of maturity is based on a random draw from a log-normal distribution, the means and variances of which are age, year, and education specific. These model parameters were evaluated using data from the LCFS, as reported in Section 6.2.
- *Wage-offers* to maturing children are allocated randomly, based on age, year, and education specific probabilities in common with the simulation of wage offers for the remainder of the life course. These model parameters were evaluated using data from repeated waves of the LFS, as reported in Section 4.2.3.

Statistical Descriptions for Behaviour

There are two broad alternatives to simulating behaviour through time, commonly referred to as reduced-form and structural modelling. The utility maximising behaviour that is a central feature of LINDA is a structural approach. Reduced form projections of behaviour, in contrast, are based on statistical descriptions of correlations between the behaviour of interest and selected observable characteristics, which are not driven by an assumed theoretical framework. Each of these approaches has its own strengths and weaknesses when analysing the implications of policy alternatives. Recognising this, we have extended LINDA under the current project to provide greater flexibility concerning the way that behaviour is projected through time.

This section begins by outlining the relative strengths and weaknesses of the structural and reduced form approaches to simulating behaviour. We then provide an overview of how behaviour can now be simulated using LINDA, before describing the reduced form specifications that have been introduced into the model.

11. Reduced-form vs Structural Projections of Behaviour

Assumptions are unavoidable when considering how the future may unfold. The quality of a projection essentially depends upon the plausibility of the assumptions made. The task of selecting a model for projecting behaviour through time therefore resolves to a selection between competing assumptions, and is complicated by the fact that there does not currently exist one set of assumptions that is best for all contexts. The most important decision in this regard, is between simulating behaviour on the basis of either a reduced-form or structural framework.

Reduced-form and structural frameworks of decision making share a great deal in common. Both approaches are based upon the assumption that past behaviour can be used as a basis for anticipating behaviour in the future. Both approaches typically describe past behaviour as a function of a set of observable characteristics. Furthermore, the precise nature of the relationships between decisions of interest and observable characteristics is commonly evaluated statistically drawing on survey data sources, whether a model is based on a reduced-form or structural framework. The key distinguishing feature is that a reduced-form approach to model behaviour is based only on observable correlations, whereas a structural approach is designed to interpret observable correlations through the lens of a theory concerning how decisions are made.

The addition to a model of a theoretical framework for decision making has both advantage and disadvantages. Consider, for example, a model of savings decisions. A reduced form approach could define the proportion of disposable income consumed each year as a function of a range of observable characteristics, including age, relationship status, number and age of dependent children, location in the income distribution, and retirement status. This type of equation could be estimated on survey data observed in 2011, and used to project net savings (and consumption) in future years. Similar equations could be estimated for other decisions, including employment, pension scheme participation, the timing of access of pension wealth, and so on.

This methodology can be a powerful tool in the analysis of policy counterfactuals. It can indicate the impact of policy on household and government budgets, and can provide a measure of responses to past policy reforms. Furthermore, the approach is well adapted to ensuring that projected behaviour is immediately recognisable by contemporary observers. For example, a middle-income couple, one of whom is 59 years old, and who have no dependent children, is likely to save 25% of their disposable income if observed today – a statistical regression equation can force the same types of relationships to be described by the simulated projections. The omission of a structural basis for the reduced-form approach also permits a great deal of flexibility when selecting the functional specification considered for analysis. This allows for the incorporation of a large set of variables, which might be difficult to distinguish within a structural model of agent behaviour, and can allow the analyst to tailor their study to the data that are available.

Unfortunately, the reduced-form approach has an important draw-back. It is almost impossible to ensure that the effects of changing incentives implied by a shift in the decision environment (e.g. a policy counterfactual) feed through to sensible changes in behaviour. Suppose, extending upon the above example, a model was to be used to consider a reform in which income support payments were increased, funded by a delay in the state pension age. This would have the effect of increasing incomes during the working lifetime, particularly toward the bottom of the income distribution, at the expense of less generous state benefits in retirement. This policy counterfactual alters incentives for employment and savings in complex ways: weakened employment incentives at the bottom of the income distribution during the working lifetime; increased incentives for employment late in the working lifetime; increased savings incentives to off-set the lower state pension benefits on offer; weakened savings incentives due to the improved income safety-net during the working lifetime. A reduced form model is ill-equipped for predicting the behavioural implications of these changes to incentives.

In macroeconomics, the behavioural limitations of reduced-form models are often referred to as the Lucas critique. A structural model framework is expressly designed so that changes in incentives feed through to behaviour in a logical fashion, guided by so-called "deep parameters", which are assumed to be invariant with changes in time and the policy environment.

As is suggested by the above text, there are (at least) two important drawbacks of the structural approach. First, the structural approach is often more computationally demanding to implement, implying longer simulation run-times, and complicating associated parameterisation. Secondly, a structural model can increase the difficulty involved in interpreting simulated effects of policy change, as behaviour may adapt to altered incentives in complex ways. These two drawbacks suggest that it may be preferable to adopt a reduced form analytical approach when undertaking preliminary analysis of potential policy reforms, due to differences in simulation run-times. A reduced form approach will also be preferred when interest is focussed on impact effects (omitting behavioural responses) of policy change. In contrast, a structural approach becomes essential when projections of interest need to adapt to the altered incentives associated with a policy counterfactual.

As a personal aside, my view is that behavioural rigidities make reduced form projections preferred for near-term time horizons (up to 5 years), and the importance of capturing the influence of changing incentives make structural projections preferable for medium to long-term time horizons (5 years and over).

12. Reduced-form behaviour in LINDA

Reduced-form descriptions for four decisions were introduced into LINDA under the current project: labour / leisure; consumption / saving; pension scheme participation; and the time of pension access. The function descriptions adopted for all four of these decisions are broadly the same, represented by:

$$y_{i,t} = \theta_1 nk_{i,t}^{0-5} + \theta_2 nk_{i,t}^{6-12} + \theta_3 nk_{i,t}^{13-17} + \theta_4 d_{i,t}^{student} + \theta_5 d_{i,t}^{graduate} + \theta_6 d_{i,t}^{pension} + \sum_{j=1}^{3} d_{i,t}^{ageband(j)} \sum_{k=2}^{10} \beta_{j,k} d_{i,t}^{decile(k)} + \sum_{l=-9}^{46} \alpha_l d_{i,t}^{age(l)} + \varepsilon_{i,t}$$

where $y_{i,t}$ denotes the decision of benefit unit *i* at time *t*, nk^{x-z} is the number of dependent children in the benefit unit within the age band *x* to *z*, d^D denote indicator variables equal to 1 when condition *D* is true and zero otherwise. *student* denotes (tertiary) students, *graduates* denotes (tertiary) graduates, *pension* denotes benefit units in receipt of retirement pensions, *ageband* denotes age bands distinguishing between three periods of life, *decile* denotes the respective decile group into which the benefit unit falls, *age* denote year specific ages, and ε denotes a residual term. Separate functions are considered for single adults and for couples.

Age and relationship specific deciles groups are evaluated using non-pension wealth for evaluating employment and the timing of pension access, using earnings for evaluating pension scheme participation, and using disposable income for evaluating total non-durable consumption (including non-discretionary expenditure). Given the structure of the model, this approach requires employment and the timing of pension access to be evaluated first, which permits earnings to be calculated. Having calculated earnings, pension scheme participation can then be identified, which permits disposable income to be calculated. Finally, given disposable income, consumption can be identified.

State pension ages are currently scheduled to increase, from 65 in 2011, to 66 from 2019, 67 from 2026, and 68 from 2034. To ensure that the reduced form adapts to this important aspect of policy change, age is measured relative to the state pension age of each individual; specifically, state pension age less prevailing age of the individual.

The functional description represented above was adopted because it has a number desirable properties. First, it is a parsimonious specification that depends only on a set of variables that are simulated by the model, which is a necessary pre-condition for use as a basis for projecting simulated decisions. Secondly, it takes into account important distributional relationships concerning the decisions of interest, wealth, earnings, and disposable income in a way that adapts in a plausible manner to temporal shifts in levels and dispersion.

The regression equations included with the model were estimated twice, once on data for 2006 and the second time on data for 2011. These two sets of parameters are designed to facilitate associated sensitivity analysis, and were selected because they coincide to periods just prior to, and following the 2007/8 financial crisis. Figure 12.1, for example, reveals that the unemployment rate among the working aged population in the UK increased steeply from just over 5 percentage points in early 2008 (5.5 percentage points in 2006), to 8 percentage points in late 2009 (8.1 percentage points in 2011). Reduced-form estimates for 2006 and 2011 may consequently provide a useful indication of the extent to which simulated profiles are likely to be sensitive to behavioural fluctuations to the prevailing economic cycle.

An ordered logit was estimated for employment, distinguishing between full-time, part-time and non-employment for each adult member, using the same Wealth and Assets Survey data that are imported into the model for the reference population cross-section (see Section 5 for details). The sample considered for estimation was limited to adults aged between 18 and 74 years, and omitted any benefit unit with at least one adult (involuntarily) unemployed.

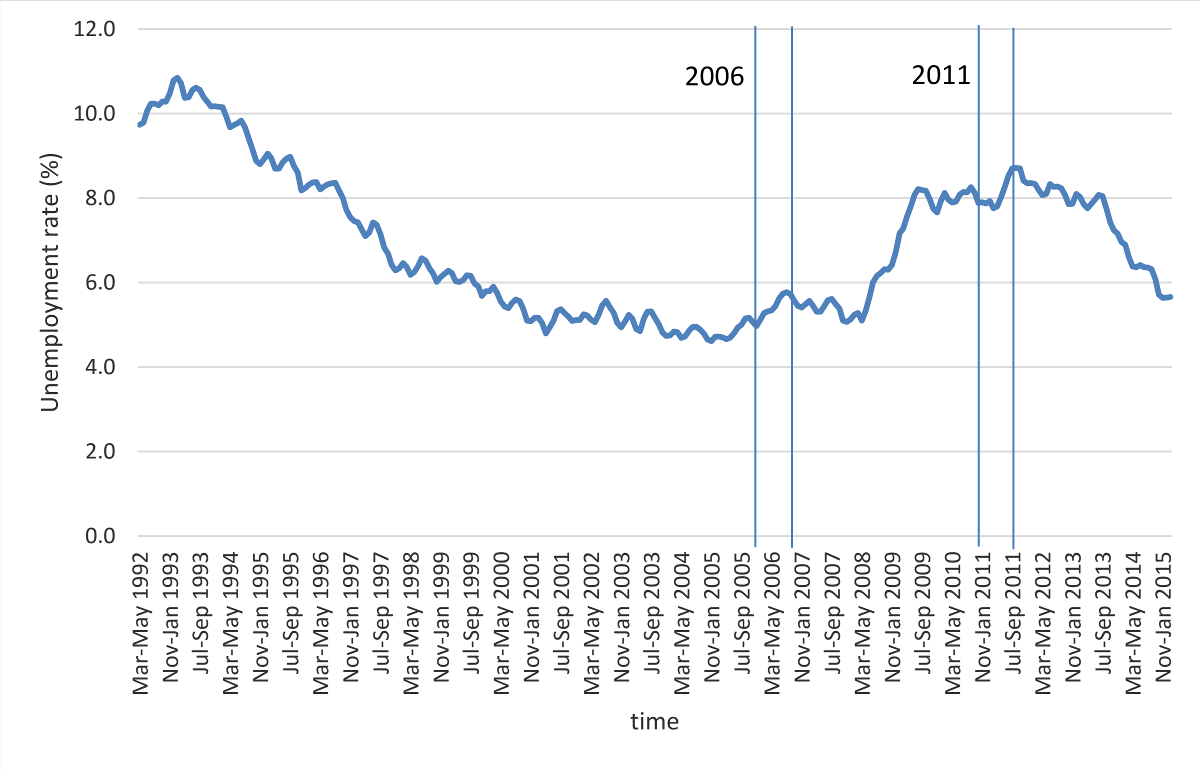
Similarly, a logit model was used to evaluate the timing of pension access, estimated using data from the Wealth and Assets Survey. For pension access, the WAS sample was limited to individuals with some pension wealth, and the same limitation is imposed when simulating benefit units through time. The probability that a given benefit unit accesses their pension wealth at any given age is evaluated by calculating the proportion of the population who are described by the logit model as having accessed their pension by the given age, p^1, and by the immediately preceding age, p^0. The probability of accessing a pension during the given age is then calculated as $(p^1 - p^0)/(1 - p^0)$. To ensure that this term is non-negative, the age specific coefficient estimates calculated for the logit were smoothed using the “lowess” STATA command and a bandwidth of 0.5, which produced parameters that increase monotonically with age.

A logit model was also estimated for the incidence of pension scheme participation, using data derived from the 2006 and 2011 waves of the Family Resources Survey. The samples considered for analysis was limited to benefit units with some labour income and under state pension age.

A linear regression model was estimated for log consumption less log disposable income, using data reported for the benefit units of individuals aged between 18 and 74 years by the 2011 Living Costs and Food Survey and the 2006 Expenditure and Food Survey. The exogenous variable considered here is only identified where consumption and disposable income are both positive, and the sample was consequently limited to benefit units with disposable income of at least £10 per week. The sample was further restricted to omit the 1% of remaining benefit units with, respectively, the lowest and highest disposable incomes, to control of outliers (Windsorisation).

Results from the regression analysis are reported in Tables 12.1 to 12.4.

Figure 12.1: Unemployment rate among all individuals aged 16 to 64 in UK



Source: ONS calculations on Labour Force Survey data, no seasonal adjustment; series UNEM01.

Table 12.1: Logit regression coefficients for pension access, by year and relationship status

	2006				2011			
	singles		couples		singles		couples	
	coef.	std. error	coef.	std. error	coef.	std. error	coef.	std. error
tertiary graduate	-0.1926	0.2079	0.0805	0.0948	-0.6293	0.2174	-0.4168	0.1149
deciles								
decile 2	0.1999	1.4771	0.3573	0.1795	0.8054	1.6976	0.1710	0.2838
3	0.2575	1.3159	0.2768	0.1788	0.4599	0.7999	0.4841	0.2532
4	0.8352	1.3082	0.5124	0.1771	1.0027	0.8164	0.6981	0.2550
5	0.6123	1.3016	0.5522	0.1740	0.8395	0.7761	0.9249	0.2595
6	0.4403	1.3090	0.7353	0.1740	0.8758	0.7672	1.4018	0.2661
7	1.0384	1.2992	0.9901	0.1800	1.4704	0.7731	1.4652	0.2648
8	1.6250	1.2966	1.2976	0.1808	1.7694	0.7853	1.8331	0.2741
9	1.7410	1.3010	1.4095	0.1808	2.3604	0.7911	1.9315	0.2772
highest decile	1.7688	1.3033	1.5178	0.1799	2.6885	0.7854	2.0216	0.2757
constant	-3.2720	1.3416	-4.0218	0.3369	-3.8107	0.8468	-3.3887	0.3489
sample size	1647		6673		1493		5888	
pseudo R2	0.4259		0.4032		0.4345		0.4108	

Source: Authors' calculations on weighted data for benefit units reported by the Wealth and Assets Survey for 2006 and 2011. Sample limited to benefit units that satisfy minimum age requirements for accessing pensions; the minimum age of pension access was increased from 50 to 55 for all individuals from 6 April 2010. Regressions include age specific dummy variables that are omitted from the table for brevity.

Table 12.2: Logit regression coefficients for participation in employer sponsored pension scheme, by year and relationship status

	2006				2011			
	singles		couples		singles		couples	
	coef.	std. error	coef.	std. error	coef.	std. error	coef.	std. error
children by age								
under 6	0.1841	0.1716	0.0228	0.0359	0.4736	0.1670	0.0973	0.0431
6 to 12	0.3922	0.0871	0.0324	0.0338	0.4141	0.1249	0.0795	0.0418
13 to 17	0.1118	0.0912	-0.0069	0.0421	0.0626	0.1211	0.0675	0.0542
tertiary graduate	0.0366	0.0777	0.1861	0.0492	0.3200	0.0941	0.0259	0.0571
pensioner	-0.1710	0.1853	0.2553	0.0722	0.1608	0.2513	0.1692	0.0883
deciles - 19 or more years to state pension age								
decile 2	0.5631	0.1870	1.1186	0.0987	0.2043	0.2789	1.0980	0.1495
3	0.9582	0.1943	1.4817	0.0981	0.7485	0.2696	1.7340	0.1455
4	1.2969	0.1885	1.8453	0.1002	1.0793	0.2544	2.0145	0.1414
5	1.7067	0.1982	2.3231	0.1046	1.2592	0.2640	2.3999	0.1423
6	1.8318	0.1957	2.4402	0.1081	1.7177	0.2524	2.7854	0.1445
7	2.2526	0.1963	2.4740	0.1069	2.0424	0.2509	2.9084	0.1478
8	2.5802	0.1958	2.8050	0.1133	2.7764	0.2524	3.4236	0.1537
9	2.4735	0.1982	3.0335	0.1240	2.8644	0.2568	3.7271	0.1596
highest decile	2.8488	0.2089	2.9194	0.1226	2.8981	0.2695	3.9226	0.1682
deciles - between 6 and 18 years to state pension age								
decile 2	0.3777	0.3450	0.8679	0.1613	1.3233	0.5410	0.6885	0.2030
3	0.9156	0.3251	1.3084	0.1617	1.7940	0.5227	1.4377	0.2042
4	1.4860	0.3202	1.3764	0.1637	2.6649	0.5127	1.9112	0.2079
5	1.3056	0.3280	1.5469	0.1631	2.3664	0.5106	1.9394	0.2332
6	1.7044	0.3303	1.6766	0.1693	2.6022	0.5166	2.4211	0.2162
7	2.1277	0.3294	2.3425	0.1845	2.8202	0.5103	2.3819	0.2193
8	2.4917	0.3529	2.8098	0.2036	3.1282	0.5355	2.6344	0.2281
9	2.6637	0.3406	3.3001	0.2355	3.5678	0.5259	2.9596	0.2440
highest decile	3.1668	0.3939	2.6817	0.2083	3.5515	0.5343	3.0660	0.2500
deciles - within 5 year of state pension age								
decile 2	0.3777	0.3450	1.6779	0.3860	-0.4498	0.8676	0.6255	0.4054
3	0.9156	0.3251	2.1357	0.3826	0.0116	0.8035	0.4733	0.4490
4	1.4860	0.3202	1.8374	0.3837	0.3201	0.8140	1.0478	0.4106
5	1.3056	0.3280	2.5379	0.3771	1.4440	0.7578	1.3071	0.3926
6	1.7044	0.3303	3.1420	0.3818	1.2288	0.7188	1.3632	0.3878
7	2.1277	0.3294	2.9665	0.3796	-0.2146	0.8381	2.3832	0.3858
8	2.4917	0.3529	3.1055	0.3831	1.5744	0.7506	2.8552	0.3994
9	2.6637	0.3406	3.5003	0.3906	1.5761	0.7450	2.5282	0.3949
highest decile	3.1668	0.3939	3.5962	0.4014	3.1572	0.8630	3.1911	0.4194
constant	-3.89855	0.272622	-3.35969	0.263051	-6.29029	1.021144	-3.17162	0.4086
sample size	5585		16860		4629		11378	
pseudo R2	0.1726		0.1588		0.2455		0.2053	

Source: Authors' calculations on weighted data for benefit units reported by 2006/07 and 2011/12 waves of the Family Resources Survey. Sample limited to benefit units with some employment income, and include age specific dummy variables that are omitted from the table for brevity.

Table 12.3: Ordered-logit regression coefficients for employment status, by year and relationship status

	2006				2011			
	singles		couples		singles		couples	
	coef.	std. error	coef.	std. error	coef.	std. error	coef.	std. error
children by age								
under 6	-1.0535	0.0948	-0.7954	0.0306	-0.7443	0.1188	-0.8262	0.0400
6 to 12	-0.6597	0.0662	-0.4107	0.0271	-0.5602	0.0884	-0.5567	0.0382
13 to 17	-0.3167	0.0728	-0.1624	0.0357	-0.0791	0.0952	-0.1753	0.0487
tertiary student	-0.7808	0.1187	-0.0218	0.1182	-1.0848	0.1427	-0.1929	0.2542
tertiary graduate	0.4488	0.0887	0.3488	0.0431	0.3028	0.1000	0.3526	0.0505
pensioner	-1.3054	0.1422	-0.6269	0.0663	-1.4632	0.1513	-1.4204	0.0830
deciles - 19 or more years to state pension age								
decile 2	-1.0919	0.1376	0.3144	0.1152	-1.0427	0.1842	0.0480	0.1581
3	-1.9493	0.1792	0.6924	0.1140	-0.9462	0.2110	0.3331	0.1499
4	-0.4354	0.1765	0.9130	0.1099	-0.3659	0.2018	0.5667	0.1510
5	0.1953	0.1406	0.9502	0.1106	0.0550	0.1885	0.8439	0.1441
6	0.5177	0.1396	1.1368	0.1123	0.6336	0.1874	0.7325	0.1394
7	0.6902	0.1347	1.1084	0.1072	0.6214	0.1869	0.8578	0.1442
8	0.6949	0.1326	1.2234	0.1111	0.8385	0.1727	0.8234	0.1402
9	0.7192	0.1387	1.2226	0.1069	1.0461	0.1819	0.8243	0.1440
highest decile	0.8529	0.1411	1.0611	0.1090	1.0321	0.1882	0.7471	0.1478
deciles - between 6 and 18 years to state pension age								
decile 2	-0.1390	0.3403	0.9749	0.1830	0.0284	0.4029	1.3628	0.3381
3	1.1710	0.2884	1.1754	0.1765	1.2593	0.3511	1.4625	0.3486
4	1.8497	0.3091	1.0929	0.1737	1.9311	0.3697	2.2193	0.3371
5	2.2614	0.3145	1.3760	0.1884	2.2272	0.4458	1.9446	0.3303
6	2.5696	0.3209	1.5503	0.1896	2.5599	0.3369	2.3422	0.3305
7	2.2989	0.2916	1.3665	0.1763	2.8110	0.3755	2.0679	0.3250
8	3.0132	0.3175	1.4348	0.1828	3.0053	0.3693	1.9425	0.3362
9	3.7104	0.3201	1.3701	0.1840	2.7167	0.3558	1.8689	0.3320
highest decile	2.3611	0.3058	1.0520	0.1739	3.3136	0.3451	1.9221	0.3229
deciles - within 5 year of state pension age								
decile 2	0.1346	0.7330	0.7133	0.2007	0.1338	0.5443	1.4998	0.2786
3	1.9900	0.5799	1.0880	0.1899	1.4888	0.4842	1.4828	0.2853
4	2.6453	0.5602	1.0384	0.1821	1.8655	0.4714	1.5976	0.2672
5	2.4322	0.5569	1.0565	0.1878	2.4294	0.4844	1.6839	0.2630
6	2.8315	0.5580	1.2465	0.1953	2.5360	0.4484	1.6764	0.2661
7	2.7747	0.5571	1.1207	0.1964	2.2584	0.4632	1.7395	0.2779
8	3.2771	0.5558	0.9509	0.1919	2.8221	0.4450	1.6939	0.2702
9	3.2702	0.5462	0.8417	0.1837	2.7075	0.4759	1.6666	0.2770
highest decile	3.0811	0.5446	1.0027	0.1860	2.4650	0.4657	1.5341	0.2650
cut1	-1.0618	0.2006	-1.7756	0.9617	0.1871	0.2148	-1.9083	1.2621
cut2	0.1404	0.2014	-0.9472	0.9611	1.4105	0.2162	-1.0620	1.2615
cut3			-0.5641	0.9608			-0.6660	1.2617
cut4			0.6121	0.9616			0.4479	1.2617
cut5			2.0348	0.9623			1.9966	1.2620
sample size	7568		15062		5622		11563	
pseudo R2	0.2329		0.1746		0.2231		0.2021	

Source: Authors' calculations on weighted data for benefit units reported by the Wealth and Assets Survey for 2006 and 2011, omitting benefit units with adult members under age 18 or over 74 years, or with at least one adult unemployed. Regressions distinguish between non-employed, part-time, and full-time employment of each adult member, where two adults part-time employed is ranked under one adult full-time employed and one not employed. Regressions include age specific dummy variables that are omitted from the table for brevity.

Table 12.4: Linear regression coefficients for the ln ratio of consumption expenditure to disposable income, by year and relationship status

	2006				2011			
	singles		couples		singles		couples	
	coef.	std. error	coef.	std. error	coef.	std. error	coef.	std. error
children by age								
under 6	0.0748	0.0406	0.0202	0.0115	0.1353	0.0362	0.0236	0.0112
6 to 12	0.1783	0.0331	0.0339	0.0095	0.0255	0.0258	0.0428	0.0107
13 to 17	0.1077	0.0306	0.0780	0.0126	0.0930	0.0432	0.0504	0.0176
tertiary student					0.0629	0.0378	-0.0072	0.0444
tertiary graduate	0.1393	0.0252	0.0917	0.0142	0.0540	0.0259	0.0892	0.0139
pensioner	0.0569	0.0409	0.1379	0.0210	0.2024	0.0390	0.1191	0.0213
deciles - 19 or more years to state pension age								
decile 2	-0.2359	0.0662	-0.2080	0.0374	-0.0761	0.0590	-0.1034	0.0385
3	-0.2876	0.0631	-0.2827	0.0370	-0.1484	0.0567	-0.1287	0.0383
4	-0.2940	0.0640	-0.3108	0.0361	-0.1900	0.0555	-0.1594	0.0375
5	-0.3179	0.0658	-0.3867	0.0362	-0.1874	0.0600	-0.2416	0.0372
6	-0.4388	0.0642	-0.4012	0.0371	-0.2287	0.0538	-0.2773	0.0379
7	-0.3990	0.0629	-0.3602	0.0389	-0.1789	0.0511	-0.3422	0.0372
8	-0.4581	0.0613	-0.4434	0.0375	-0.2405	0.0547	-0.3105	0.0382
9	-0.5355	0.0625	-0.4510	0.0399	-0.2830	0.0532	-0.4192	0.0386
highest decile	-0.5589	0.0670	-0.4930	0.0420	-0.3345	0.0528	-0.3764	0.0400
deciles - between 6 and 18 years to state pension age								
decile 2	-0.2716	0.1890	-0.1099	0.0687	-0.2223	0.1418	-0.1805	0.0809
3	-0.3995	0.1885	-0.3369	0.0569	-0.3564	0.1275	-0.2536	0.0735
4	-0.2969	0.1768	-0.3138	0.0637	-0.3751	0.1159	-0.3559	0.0724
5	-0.5188	0.1765	-0.4572	0.0589	-0.5862	0.1119	-0.3997	0.0729
6	-0.5227	0.1754	-0.4412	0.0628	-0.3654	0.1185	-0.4774	0.0738
7	-0.5479	0.1761	-0.4161	0.0631	-0.5552	0.1133	-0.4493	0.0733
8	-0.5794	0.1737	-0.5081	0.0732	-0.7182	0.1199	-0.4699	0.0758
9	-0.7950	0.1742	-0.5238	0.0787	-0.7632	0.1180	-0.5220	0.0740
highest decile	-0.7396	0.1827	-0.6580	0.0695	-0.7029	0.1494	-0.5974	0.0828
deciles - within 5 year of state pension age								
decile 2	-0.2055	0.1072	-0.2145	0.0586	-0.2852	0.1075	-0.0990	0.0632
3	-0.2496	0.1019	-0.3171	0.0533	-0.3388	0.1161	-0.1451	0.0643
4	-0.3929	0.0988	-0.4131	0.0545	-0.3577	0.1087	-0.2080	0.0656
5	-0.3809	0.1050	-0.4462	0.0548	-0.4004	0.1062	-0.2889	0.0612
6	-0.4118	0.1255	-0.4619	0.0572	-0.4136	0.1089	-0.3315	0.0593
7	-0.4374	0.1021	-0.6463	0.0514	-0.4231	0.1016	-0.4366	0.0607
8	-0.5348	0.1048	-0.6384	0.0528	-0.4094	0.1098	-0.4297	0.0647
9	-0.6075	0.1088	-0.7131	0.0569	-0.5390	0.0976	-0.4963	0.0627
highest decile	-0.7581	0.1228	-0.7399	0.0631	-0.5702	0.1054	-0.5686	0.0679
constant	0.1920	0.1036	-0.1006	0.0362	-0.0377	0.0634	-0.2460	0.1605
sample size	3267		6174		2982		5273	
pseudo R2	0.1343		0.1599		0.0941		0.1351	

Source: Authors' calculations on weighted data reported by the Expenditure and Food survey, 2006, and the Living Costs and Food Survey 2011. Regression equations calculated on data organised with respect to benefit units, and include age specific dummy variables that are omitted from the table for brevity.

Consumption at the Individual Level

LINDA simulates each family as a single decision maker. Adapting the model to focus on the individual (rather than the benefit unit) as the unit of analysis would be a very significant undertaking that we beyond the scope of the current project. Nevertheless, power relationships and intra-family allocations are a focus of the JRF anti-poverty programme. The model was consequently augmented to permit aggregate benefit unit consumption to be disaggregated down to the individual level. This was achieved by introducing a functional description for the share of family consumption enjoyed by individual family members, as is discussed below.

13. Reduced form disaggregation of consumption

LINDA has been adapted to permit individual specific measures of consumption to be generated for each family member. Benefit unit consumption is disaggregated using a reduced form equation that describes the share of consumption accruing to each family member as a function of a set of individual specific and benefit unit characteristics that are simulated by the model. The coefficients for this reduced form equation were estimated using data reported by the 2012 Living Costs and Food Survey (LCFS), assuming a logit regression specification.

The LCFS reports detailed information concerning household expenditure, derived from responses to both a household questionnaire, and diary entries that respondents fill in daily during a fortnight. The information derived from these sample methods is organised by the survey into individual specific measures of expenditure, and communal expenditure at the household level. Individual expenditure is comprised of 14 different expenditure categories in the survey, and is reported by the survey in aggregate as variable p153(c).[31] Total household expenditure is comprised of aggregate private expenditure, plus various measures of communal expenditure, and is reported by the survey as variable p550tp. Aggregate measures of private consumption account for 55% of total household consumption on average, and 52% at the median.[32]

Each individual's total consumption was evaluated by adding individual specific consumption to their allocated share of communal consumption, subject to the assumption that communal consumption is shared equally between all household members. Having identified each individual's total consumption, it was straightforward to evaluate the consumption share of each individual in their respective benefit unit. These consumption shares are the dependent variables described by the reduced form equation.

The reduced form equation was estimated using the "glm" regression command in Stata, a logit link function, and a binomial distribution; see the Stata user manual for details. Estimated statistics for the weighted regression are reported in Table 13.1.

---

[31] The "c" distinguishes expenditure by dependent children from expenditure by adults.

[32] Mean (median) total household consumption = £487 (£408) per week, compared with mean (median) aggregate household private consumption of £268 (£213) per week; statistics weighted.

The statistics reported in Table 13.1 are included in the set of model parameters for LINDA, and are used to evaluate the consumption undertaken by each individual family member. The approach taken to disaggregate benefit unit consumption can be explained by considering a practical example.

Consider a benefit unit comprised of a single adult age 35 and a dependent child aged 7. Suppose that the adult is full-time employed, is a non-graduate, and that the benefit unit's ratio of income to consumption is 1.15 (spending approximately 15 per cent less than income). The regression parameters reported in Table 13.1 imply a total coefficient value of -1.19 for the child and +0.98 for the adult. Taking the inverse of the logit implies unadjusted consumption shares equal to 0.23 for the child and 0.73 for the adult. The model adjusts these shares so that they aggregate up to 1.0, and uses the result to allocate total benefit unit consumption to individual unit members. The consumption that is allocated to each individual who appears as an adult in the simulated population in at least one year is reported as part of the simulated micro-data, under the variable name cons_ind.

One of the key motivations for decomposing consumption to the individual level is to take into consideration the influence of power-relations within the household. Careful attention has been paid to this issue in defining the specification considered for analysis. We assume that any two children who have identical ages within a benefit unit enjoy the same consumption. This same assumption could, however, mask important variation if applied to adults in married couples. The dummy variable "reference adult" has been added to the specification to reflect systematic variation between the consumption of adult family members.

The initial regression specification that we considered for analysis identified the reference adult as the adult with the highest private income. The estimated coefficient on this identifier was small, negative, and not statistically significant, indicating that adults within benefit units tend to consume very similar shares. We subsequently tested identification of reference adults as the adults with the highest individual specific consumption within the benefit unit, as described by survey data (see above). This second approach results in the highly significant coefficient estimate reported in Table 13.1, and has two important implications. First, it serves to capture the absolute size of systematic differences between the consumption of adult benefit unit members. Secondly, when applied to the simulation model, it tends to provide an upper bound to the influence of power-structures on consumption through time, when used to generate consumption for the same adult in every simulated time period.

Table 13.1: Generalised linear regression statistics for share of total benefit unit expenditure accruing to individual benefit unit members; LCFS 2012

		singles		couples	
		coef	s.e.	coef	s.e.
aged 0-1		-0.9260*	0.0796	-1.3440*	0.0485
aged 2-4		-0.8901*	0.0830	-1.3359*	0.0498
aged 5-9		-0.8862*	0.0678	-1.3326*	0.0497
aged 10-12		-0.8070*	0.0670	-1.2698*	0.0516
aged 13-15		-0.7792*	0.0747	-1.2600*	0.0517
aged 16-17		-0.3851*	0.1017	-1.0244*	0.0533
aged 18-24		2.1435*	0.1302	0.0066	0.0225
aged 25-34		2.0332*	0.0947	0.0476*	0.0176
aged 35-44		1.9439*	0.1039	0.0490*	0.0178
aged 45-54		2.0294*	0.1202	0.0535*	0.0164
aged 55-64		1.7620*	0.1600	0.0373*	0.0143
aged 65-74		1.6045*	0.4156	0.0151	0.0154
aged 75+		0.9758*	0.1119		
constant		-0.4937*	0.0126	-0.4937*	0.0126
covariates interacted with adult indicator					
at least one graduate		-0.0239	0.0677	-0.0148*	0.0043
income to expenditure ratio		-0.1513*	0.0628	-0.0080*	0.0023
individual full-time emp		-0.0784	0.1104	-0.0153	0.0145
individual part-time emp		-0.1569*	0.0783	0.0163	0.0159
reference adult				0.9693*	0.0133
spouse full-time emp				-0.0024	0.0143
spouse part-time emp				-0.0319*	0.0159
children:	couple			-0.2377*	0.0126
	aged 0-1	-0.2416*	0.0679	-0.1191*	0.0118
	aged 2-4	-0.2533*	0.0740	-0.1374*	0.0148
	aged 5-9	-0.2212*	0.0483	-0.1320*	0.0094
	aged 10-12	-0.2480*	0.0626	-0.1467*	0.0114
	aged 13-15	-0.2476*	0.0649	-0.1312*	0.0114
	aged 16-17	-0.4705*	0.1091	-0.2002*	0.0153
covariates interacted with child indicator					
at least one graduate		0.0250	0.0587	0.0716*	0.0216
income to expenditure ratio		0.1963*	0.0601	0.0572*	0.0163
adults full-time employed		-0.0361	0.1009	-0.0049	0.0261
adults part-time employed		0.0593	0.0675	0.0021	0.0253
siblings:	aged 0-1	-0.3094*	0.0736	-0.2708*	0.0295
	aged 2-4	-0.4049*	0.0722	-0.2455*	0.0320
	aged 5-9	-0.3651*	0.0453	-0.2489*	0.0199
	aged 10-12	-0.3763*	0.0574	-0.2424*	0.0238
	aged 13-15	-0.4595*	0.0616	-0.3035*	0.0257
	aged 16-17	-0.4758*	0.1419	-0.2761*	0.0389

Source: Authors' calculations using data reported by the 2012 LCFS for individuals in benefit units with at least two members; sample size = 10,055

Notes: * indicates significant at 95% confidence interval. Regression statistics evaluated using the glm Stata command, a logit link function, and a binomial distribution on un-weighted data. A joint regression specification evaluated for singles and couples; constant reported for each subgroup is the same by construction. Income to expenditure ratio evaluated with reference to aggregate benefit unit gross income from employment and pensions. Reference adults identified as adults with highest consumption share. Robust standard errors, assuming clustered observations around benefit unit identifiers assumed.

Modelling Education

Prior to the current project, LINDA could be used to distinguish between individuals on the basis of whether they held a graduate level qualification (see the technical report, and Sections 4, 5, and 6 above). This feature of the model assumes that graduates are distinguished from non-graduates in relation to their wage parameters, rates of (involuntary) unemployment, marital rates, and divorce rates. The model can also take into consideration periods of graduate level education at the beginning of the simulated adult life.

Education in the model has been extended under the current project to distinguish between four non-graduate education categories, referring to the highest qualification held: 2 or more A levels, 5 or more A*-C GCSEs (AC5), apprenticeships, and none of the above. Individuals within each of the three new education categories are assumed to differ from one another in relation to their wage parameters and rates of unemployment.

14. Distinguishing between sub-degree education qualifications

Expanding simulated heterogeneity in education status as described above required new detail to be included for each individual at their time of entry into the simulated sample, and evaluation of new wage and unemployment parameters for the expanded set of educational categories. These details are described in the subsections that follow.

14.1 Identification of education qualifications held at model entry

Three types of individual enter into the simulated sample: adults represented in the data for the reference population cross-section (the starting point for model projections); children who mature into the simulated population; and international immigrants. In relation to the last of these population subgroups, imputation of characteristics for international immigrants was unaffected by the extension of educational heterogeneity described here; see Section 9.2 of this report for associated details. The remaining two population subgroups are discussed in turn below.

14.1.1 Education qualifications for the reference population cross-section

The central data source for the reference population cross-section is the Wealth and Assets Survey (see Section 5 of this report). This survey reports three variables that describe the education qualifications held by each adult. “edlevelW3” reports the “level of highest education qualification”, and permits identification of individuals holding “degree level qualifications or above”. “EdAttn1W3” permits individuals with some form of education certificate to be distinguished from those without. Finally, the WAS reports the age at which each individual left full-time education, “TEAW3”.[33]

The Labour Force Survey (LFS) is the principal data source in the UK for exploring the variation of labour market circumstances by educational qualifications. In addition to the education specific characteristics reported by the WAS, the LFS reports a suite of questions that permit identification of

[33] Note that TEAW3 reports age left full-time education for only a subset of the population. This information can be supplemented with TEAW* data from waves one and two to obtain decent coverage of the reported population.

the five educational categories considered for analysis. For the population of non-graduates aged 25 to 69, an ordered logit equation was consequently estimated to describe each individual's educational category as a function of their age, sex, relationship status, and the age that they left full-time education. Whether an education certificate was held at the time of interview was dropped from the regression analysis as this variable was found to complicate convergence of the ordered logit. Estimated coefficients are reported in Table 14.1.

The logit regression statistics reported in Table 14.1 are used to impute education status for each adult described by the WAS, with individuals under age 25 treated identically to 25 year olds, and individuals older than age 69 treated identically to 69 year olds. This procedure, which is performed internally by LINDA, involves evaluating for each individual the probabilities implied by the ordered logit for each educational category, and then comparing these probabilities against a random draw from a uniform [0,1] distribution. The educational category of each individual is included in the standard set of micro-data outputs that are generated by the model.

14.1.2 Education qualifications for maturing children

The model is designed to project the education qualifications of maturing children with reference to the education qualifications of their parents. The relationship between parental and child education qualifications is defined by two transition matrices, one specified for the reference cross-section and the other specified for a given future year. These transition matrices define the probability that a child holds education category *B*, given that their parents' highest education category held is *A*. The model assumes that there is a linear shift from the probabilities defined for the reference cross-section to the probabilities defined for the future year, after which the probabilities for the future year are assumed to apply into the indefinite future.

The modelling approach defined above is designed to facilitate alternative assumptions concerning the relationship between parental and child education qualifications in the future. The probabilities defined by the two transition matrices, and the 'given future year' can all be altered via dedicated Excel spreadsheets.

As a starting point, the transition probabilities are assumed to be constant through time. The probabilities that are supplied with the model were evaluated using data reported for individuals aged 25 to 30 by wave 4 of the *Understanding Society* survey, observed in 2013 (the most recently available data at the time of writing). This data set has the advantage that it reports comparable education qualifications to the Labour Force Survey for individuals and their parents. The estimated transition probabilities are reported in Table 14.2. One point of note in relation to Table 14.2 is that, although there is a high degree of correspondence in the rates of alternative educational categories reported by the Labour Force Survey and Understanding Society in general, this does not hold for reported rates of apprenticeships.[34] Rates of apprenticeships are consequently set with reference to the "other qualification" education category listed in Table 14.2. This recognises that level 1

[34] Comparative data for the population aged 18 to 64 in 2013, for the seven education categories defined by variable hiqual_dv are: "no answer" (LFS=1.2%, US=2.2%); "degree" (LFS=26.2%, US=26.4%); "other higher degree" (LFS=9.7%, US=11.6%); "A-level etc" (LFS=22.8%, US=22.6%); "GCSE etc" (LFS=21.1%, US=21.1%); "other qualifications" (LFS=9.6%, US=7.8%); "no qualification" (LFS=9.4%, US=8.2%). In contrast, the LFS reports 8.5% of the population holding an apprenticeship, relative to less than 1% holding a modern apprenticeship in the US (general apprenticeships are not reported).

Certificates and level 2 Diplomas are currently cited as useful preparatory work prior to commencement of an apprenticeship.

14.2 The effects of education on employment opportunities and wages

14.2.1 Education and age specific unemployment rates

Age, year, and sex specific unemployment rates were evaluated from data reported by the April-June quarter of the Labour Force Survey between 1993 and 2014, for each of the five education categories permitted in the model.[35] These probabilities are included in the set of model parameters, and are used to project differences in employment opportunities between education categories through time. Plots of age, sex, and education specific averages of unemployment rates taken over the two decade period for which data are available are provided in Figure 14.1. These plots reveal that unemployment rates are broadly similar for all but the lowest educational level, which are appreciably higher than the others throughout the working lifetime. Hence, employment returns to education, at least in terms of the default model parameters are largely confined to disparities between those with and without some recognised education qualifications.

Age specific employment returns to education that are used for forward projections can be adjusted via dedicated model parameters found in worksheet “forward projection assumptions” in file “NONGRADUATE PARAMETERS.xls”.

14.2.2 The wage returns to education

All non-graduates are subject to the same process governing the intertemporal evolution of latent wages, as described in Section 4.2 of this report. Wage returns to education are distinguished between alternative non-graduate qualifications through the addition of a term, λ^{ngrad}, to the equation that translates wage potential into labour income:

$$g_{i,a} = \max\left(h_{i,a}, h_{a,t}^{\min}\right)\lambda_{i,a}^{o}\,\lambda_{i,a}^{ret}\,\lambda_{i,a}^{emp}\,\lambda_{i,a}^{ngrad}$$

where all terms are described in Section 4.1. The form assumed for λ^{ngrad} accommodates variation by age and year. Associated model parameters, accessible through the “forward projection assumptions” in file “NONGRADUATE PARAMETERS.xls”, are designed to facilitate associated sensitivity analysis.

The default parameters provided with the model are broadly based on the empirical estimates for the marginal wage returns to education that are reported in Hayward *et al*. (2014).[36] The task of estimating wage returns, like any reduced form wage regression, is complicated by selection effects associated with the decision to work. Attendant conceptual difficulties argue in favour of a stylised

---

[35] In any case where fewer than 20 observations were reported for an age / year / sex / education combination, then the associated unemployment rate was usually approximated by the nearest year for which the age / sex / education combination did report at least 20 observations. The only exception to this was in relation to female apprentices, for whom very few observations were reported. In this case, statistics for men were assumed.

[36] Marginal wage returns describe the disparities between wages earned by individuals holding a given level of education as their highest qualification, and the wages earned by a reference population holding some other level of education as their highest qualification.

specification with underlying assumptions that can be readily explained, and subject to sensitivity analysis. The specification assumed by Hayward *et al.* seems fit for this purpose.

The regression estimates reported by Hayward *et al*. were evaluated using Ordinary Least Squares on data pooled from the quarterly Labour Force Survey, from the first quarter 2006 to the first quarter 2013 (thereby spanning the 2008 financial crisis). The sample focuses on wave 1 individuals employed in England, aged 18 to 64, and earning between £0.01 and £100.00 per hour. The regression specifications considered here take a standard Mincer form, describing the log of gross weekly wages from the respondent's main job as a function of a detailed set of explanatory variables. The explanatory variables include controls for government office region, ethnicity, relationship status, dependent children, and the year of observation, in addition to sex, age, and education that are our principal focus of concern.

The estimated wage premia reported by Hayward *et al*. (2014) are effectively averages taken over sex, age, and education specific subgroups. Furthermore, the estimates do not distinguish individuals by hours of work. Hence, differences between population subgroups in terms of hours of work have a direct bearing on the premia that are reported, and any distributional variation is obscured. These limitations are balanced against the relative simplicity and flexibility of the relationship between education and wages that is described by the estimates reported by Hayward *et al.*.

One issue concerning the empirical specification reported by Hayward *et al*. (2014) that we have attempted to off-set in the default parameters supplied by the model is the influence of the assumed quadratic relationship between wage premia and age. A quadratic age profile is a common feature of Mincer type wage regressions. Such a profile tends to be well adapted for capturing the rise in wages typically observed at the beginning of the working lifetime, and the levelling out of wages as a worker matures. There is, however, some contention concerning the extent to which wages decline later in life. In cross-sectional data (similar to the data considered by Hayward *et al*.), a decline in wages amongst older people can reflect the influence of wage growth amongst younger birth cohorts. In longitudinal data, it is difficult to distinguish a wage decline from the influence of (endogenous) employment decisions – it may be, for example, that high wage people also have high savings, and choose to retire earlier as a result.

Consider, for example, estimates reported for marginal wage returns for individuals with five or more A*-C GCSEs including maths and English, relative to those qualified to any lower education level. The wage regression coefficients reported in Table 3 of Hayward *et al*. (2014) imply returns that are displayed graphically in Figure 14.2. Although it seems plausible to us that the returns to GCSE education may be negative at the very beginning of the working lifetime, the negative estimates late in the working lifetime are unlikely to reflect underlying differences in earning potential. If individuals late in working life who hold GCSE qualifications do earn substantively less than those who are lower qualified[37], then it seems reasonable to suppose that these differences are driven by employment decisions rather than differences in earnings potential.

[37] Note that such differences may be a pure artefact of the quadratic functional form that was assumed for the regression, and may not be present in the data at all.

Given the above observations, and the well-recognised end-point problems associated with polynomial specifications[38], the default model parameters assume that the marginal wage returns to education are non-negative beyond age 45. Similarly, the wage return for apprentices is assumed to level out from age 45. These adjustments censor the profiles implied by the regression estimates reported by Hayward *et al*. (2014), resulting in the wage returns that are reported in Figure 14.3.

Model implementation

The form that is assumed to describe the influence of non-graduate education on wages introduces a complication when non-graduates first enter the simulated population, as it drives a wedge between observed wages and simulated wage potential. Consider, for example, an individual entering the simulated population as part of the reference population cross-section. Each individual represented in the reference cross-section is associated with a wage potential and an educational level. Prior to the current project, all non-graduates were treated equivalently, and the distribution of wage potentials for this population subgroup reflected both differences within non-graduate educational categories and differences between them.

With the inclusion of variation between non-graduate education levels as described above, it is necessary to adjust wage potentials of non-graduates at model entry. If this were not done, then an individual with A-level qualifications, for example, would tend to be associated with a higher wage potential, $h_{i,a}$, and would also receive a wage return represented by λ^{ngrad}, resulting in a form of "double counting".

The model is designed to accommodate the complication described above by re-scaling the wage potential of every non-graduate (excluding immigrants) to off-set the implied influence of their education level at the time that they enter the simulated sample. Each individual represented in the reference cross-section, or entering the simulated sample as a maturing child, is associated with both a wage potential and an education level. Where no distinction between non-graduate education qualifications is accommodated, then no adjustment is required. Where non-graduate education qualifications are taken into account, then the wage potential when an individual first enters the population is adjusted by the inverse of their respective educational wage return.

[38] A polynomial defines a specific profile, and a least squares estimated profile may fit observed data well where observations are intensive (during peak working years), but very poorly where observations are sparse (into retirement).

Table 14.1: Ordered logit regression statistics for highest education qualification held, population aged 25 to 69 in 2011

age band	18 to 30		31 to 40		41 to 50		51 to 60		60 and over	
sex	men	women	men	women	men	women	men	women	men	women
married	-0.0188	0.1823	0.2103*	0.2475*	0.2644*	0.3138*	0.2152*	0.2351*	0.3546*	0.2639*
unemployed	-0.5554*	-0.5940*	-0.7306*	-0.5793*	-0.3962*	-0.7195*	0.0548	-0.0912	0.3711*	0.5801
age left full-time education										
under 16	1.5629*	0.7113*	1.2477*	0.7062*	1.4689*	0.9521*	1.6622*	0.2283	1.6502*	-0.0862
16	2.3804*	2.2217*	2.1571*	1.9303*	2.1666*	1.9121*	2.3422*	1.7277*	2.1308*	1.2246*
17	3.2814*	2.9094*	2.7628*	2.6398*	3.0425*	2.9470*	3.2533*	2.8299*	2.5010*	1.7045*
18	4.0382*	3.5557*	3.4245*	3.2463*	3.9672*	3.6474*	4.1850*	3.9163*	3.0388*	2.4310*
19	3.2345*	3.3721*	3.1821*	3.1309*	3.7581*	3.6866*	3.7248*	4.1392*	3.2037*	3.0462*
20	3.2683*	2.8473*	3.4822*	2.6347*	3.5374*	3.4558*	4.0108*	4.3814*	2.2901*	1.8991*
21	3.8874*	3.0890*	3.1282*	3.2222*	3.7196*	3.7556*	3.8596*	4.9603*	2.6953*	2.8777*
over 21	2.2760*	2.3749*	2.4924*	2.2828*	2.9107*	2.6627*	3.8250*	3.7120*	1.9206*	1.5311*

Source: Authors' calculations using weighted data for non-graduates reported by the April-June quarter of the Labour Force Survey, 2011.

Notes: Dependent variable = 3 for two or more A levels, 2 for five or more A*-C GCSEs, 1 for apprentices, 0 otherwise

"married" includes cohabitating couples.

* indicates coefficient statistically significant at 95% confidence interval.

Number of observations = 59650.

Cuts = 2.8936, 3.3711, 5.2547.

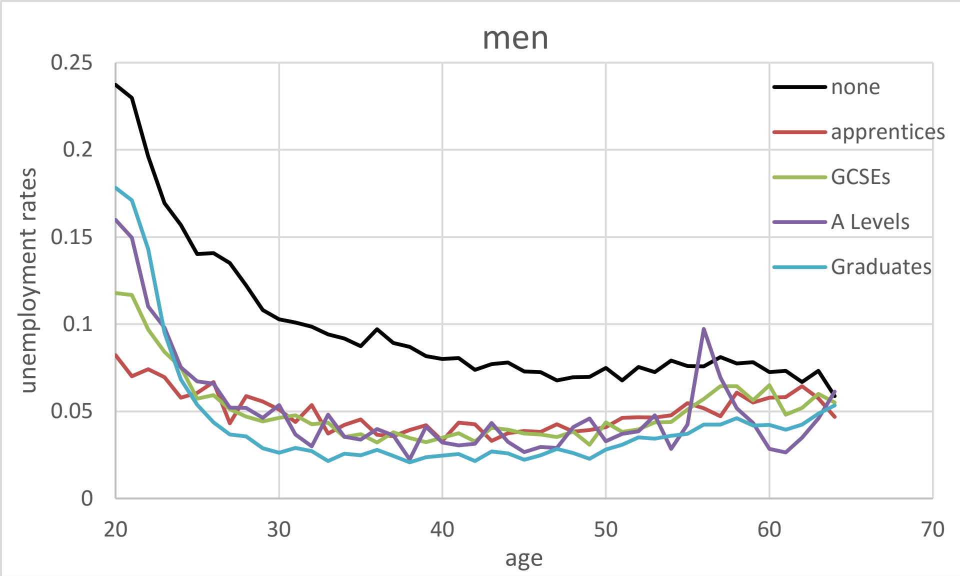
Pseudo R squared = 0.0958.

Table 14.2: Distribution of child educational attainment by parental educational attainment; individuals aged 25 to 30 in 2013

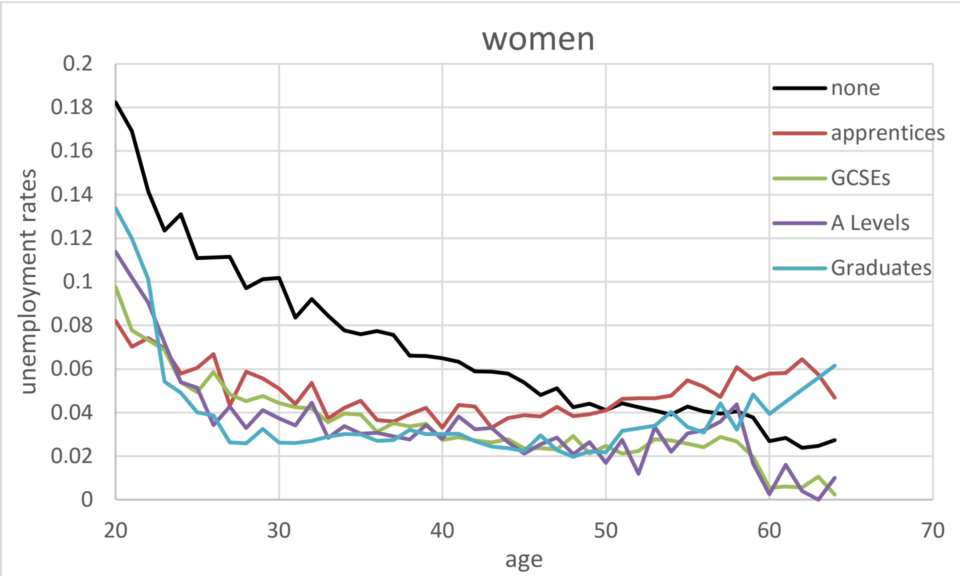
	parental qualifications				
child qualifications	none	other higher	GCSEs	A-levels	degree
no qualification	0.1280	0.0617	0.0307	0.0252	0.0200
other qualification	0.0961	0.0912	0.0429	0.0303	0.0351
GCSEs	0.3337	0.3997	0.2606	0.2310	0.1043
A-levels	0.2778	0.3209	0.4048	0.3712	0.3509
degree	0.1376	0.1165	0.2376	0.3348	0.4821
all	0.9732	0.9900	0.9765	0.9926	0.9923
observations	136	70	205	279	193

Source: Authors' calculations using data reported for 25 to 30 year olds in wave 4 of the Understanding Society survey
Notes: All statistics report weighted averages and omitted observations with missing data
Statistics refer to highest reported qualifications held, variable d_hiqual_dv
A-levels category includes both A-level qualifications and "other higher degrees"
Columns do not sum to 1 due to rounding

Figure 14.1: Unemployment rates by age, gender, and highest education qualification; annual averages between 1993 and 2014



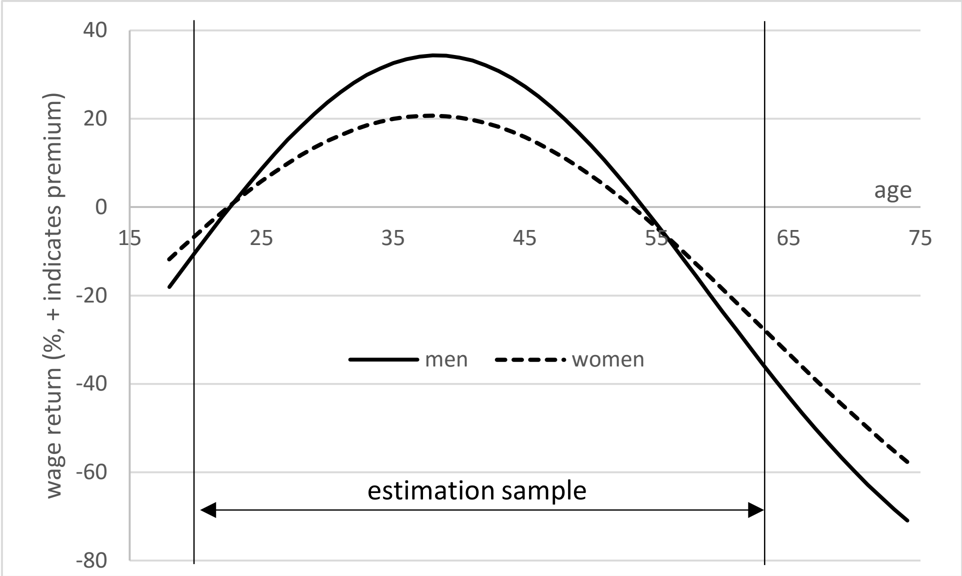
Panel A: men aged 20 to 64



Panel A: women aged 20 to 64

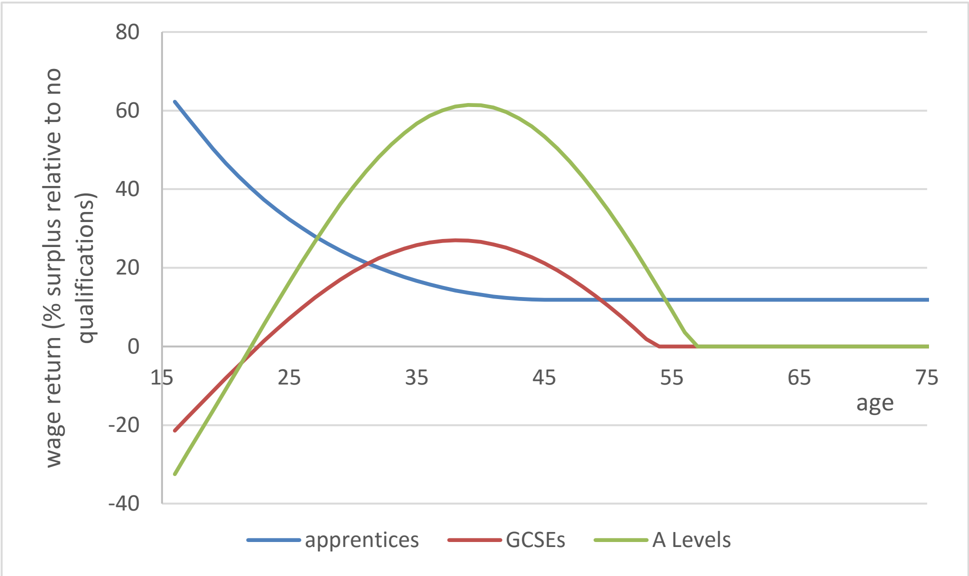
Source: Authors' calculations using data reported by the April-June waves of the Labour Force Survey, 1993 to 2014

Figure 14.2: Wage returns to five or more A*-C GCSEs including English and maths relative to individuals with any lower qualifications (including none)



Source: Authors' calculations using regression estimates reported for "marginal" wage specifications in Hayward *et al*. (2014), Table 3.
Notes: estimation sample refers to age bands included in data used for estimation
Full life period reported in figure spans ages 18 to 74, which is the period of life that individuals are assumed to be able to work in LINDA

Figure 14.3: Wage returns assumed by default model parameters by age and non-graduate education status, relative to population without recognised qualifications



Source: Authors' calculations based on regression estimates reported for "marginal" wage specifications in Hayward *et al*. (2014), Table 3.

Modelling Disability

Prior to the current project, LINDA accounted for health differences between benefit units only in relation to the timing of death. This aspect of the model has been extended under the current project, to distinguish benefit units with respect to:

1) disability status of adults (measured on a discrete scale)
2) the incidence of disabled children
3) the incidence of carer responsibilities amongst adults

The model has been adapted to reflect key support payments relating to the characteristics defined above. Discussion consequently begins by describing these support payments in Section 15. Given the policy context, Section 16 describes the approach taken to simulate each of the three characteristics listed above, including associated issues of parameterisation.

15. Disability and public policy

Four key transfer payments accounted for approximately 90% of state expenditure on benefits for sick and disabled people in 2013/14.[39] Each of these scheme is described in turn, where all monetary values relate to rates applicable in April 2011.

15.1 Employment and Support Allowance (ESA)

The ESA is designed to support working-aged individuals with a ‘limited capability for work’. As noted in Section 4.6.2, a variant of the ESA is included in the model that omits explicit consideration of health status. In this case, ESA is assumed to be payable to any benefit unit within ten years of their state pension age. This feature is suppressed when the model is directed to simulate disability, in which case the model simulates eligibility for ESA benefits with reference to each adult’s prevailing disability status.

ESA payments are simulated based on a set of personal allowances and premia. The simulated personal allowance is worth £67.50 per week (in 2011) for single and sole parents, and £105.95 per week for a couple. Furthermore, the model simulates premia per week equal to:

- £26.75 for work related activity recipients
- £32.35 for support recipients
- £31.00 for carers
- If eligible for the enhanced rate living component of the PIP (see below):
    - £14.05 for single adults
    - £20.25 for couples
- If eligible for the living component of PIP, and:[40]

[39] Based on figures reported in Table 3.1 of Hood and Oakley (2014).

[40] The model assumes that single adults do not receive this benefit, but couples who both receive the living component do receive a single share of the allowance. These assumptions are based on the premise that the carer allowance is always claimed, although only one share of the carer’s allowance is claimed for couples.

- single, without anyone receiving Carer's Allowance in respect of household, or
- couple, both receiving living component of PIP and only one receiving subsidised care under Carer's Allowance:
  - £55.30 per week (doubled if couple and no CA received)

Only the premium for 'support recipients' is reflected in the version of the model that omits health. Furthermore, benefit units that are identified as home owners with a mortgage are eligible for assistance with mortgage interest payments, worth up to 3.85% per annum (nominal, 1.85% real) on a mortgage of up to £200,000.

The applicable amount, equal to the sum of personal allowances and premia, is withdrawn at the rate of £1 for each £1 of private income, and £1 per week of benefit for every £500 of non-pension non-housing wealth over a £10,000 threshold.

15.2 Disability Living Allowance (DLA) and Personal Independence Payment (PIP)

DLA and PIP are designed to help benefit units with disabled members who are under state pension age meet the costs associated with their conditions. Attendance Allowance, which is described in the following subsection, plays the same role as DLA/PIP for claimants in excess of state pension age. The PIP replaced DLA for new claimants from April 2013, and is scheduled to be open to all existing claimants by late 2017. The two schemes share many similarities and are consequently simulated together. The most important differences between the two schemes are that the PIP includes two rates of care payments, relative to three payable under the DLA. The reduction in the number of care (daily living) payments under PIP aligns the benefit with the Attendance Allowance. Furthermore, the health assessment for the PIP focusses on the ability of an individual to 'participate in society', in contrast to the DLA's reference to specific health conditions. The following discussion focusses upon the PIP, as this is the most relevant benefit for forward projections.

PIP has two components, and a claimant can receive either one or both of these depending upon their circumstances. The daily living component is designed to subsidise the costs of care required by a disabled individual. The rate of payment depends on an official assessment of care needs, with a standard rate equal to £49.30 per week, and an enhanced rate of £73.60 per week. The mobility component of PIP pays a benefit to subsidise outdoor journeys by individuals aged 5 years and above, and is payable at a standard rate of £19.55 per week, and an enhanced rate of £51.40 per week. The PIP is non-taxable, non-means tested, and non-contributory.

Receipt of PIP passports individuals to premia under a range of alternative programmes:

- The Child Tax Credit increases by £2,800 for a disabled child (disabled child addition).
- The Working Tax Credit increases by £2,650 in 2011.
- The Working Tax Credit increases by a further £1,130 if the enhanced rate living component is payable.
- The labour condition required to receive the WTC is relaxed to part-time for one adult.
- Disability premia under ESA (see Section 15.1)
- Exempts the benefit unit from the benefits cap.

15.3 Attendance allowance (AA)

AA is payable to individuals from state pension age, and provides similar support to the daily living component of the PIP for working aged people. Like the PIP's daily living component, AA is payable at two rates; the lower rate is equivalent to the standard rate of PIP (£49.30 per week in 2011), and is payable if an individual requires care during the day or night (but not both). The higher rate corresponds to the enhanced rate of PIP (£73.60 per week), and is payable if an individual requires care during both the day and night. Like PIP, AA is non-taxable, non-means tested, and non-contributory.

Although initial claims to the PIP are limited to individuals who are under state pension age, on-going claims to PIP are not discontinued from state pension age. This has the advantage that mobility payments, which are not payable under the AA, can continue to be received into older age.[41]

Eligibility for AA increases the pension credit guarantee credit by £55.30 per week for each qualifying adult, in respect of foregone Carer's Allowance (see conditions defined for ESA).

15.4 Carer's allowance (CA)

CA provides support to individuals who are 16 or over, and who provide regular and substantial care to a person receiving AA or the living component of PIP. The benefit pays a standard rate of £55.55 per week, and cannot be claimed by an individual in full-time education or earning more than £102 per week (in 2014). The benefit is taxable and non-contributory. CA increases the eligible benefit payable under IS, PC, or ESA by £31.00 per week. CA also exempts the spouse from the 16 hour working condition to receive the WTC.

15.4.1 Distinguishing between alternative allowances

Some individuals may be eligible to Carer's Allowance at the same time as they are eligible to either Income Support or Employment and Support Allowance. The model is designed to select the benefit option that maximises each individual's disposable income. It does this by first considering eligibility for a CA payment, and any associated IS/ESA premia. The eligible IS payment is then evaluated, taking into account any relevant CA premium. ESA eligible payments are then evaluated in the same way, and the maximum payment selected between IS and ESA. This maximum is compared against the relevant CA payment to determine which benefit the individual should take-up.

[41] The same issue applies to DLA, with the added advantage in relation to that scheme that the third (lower) rate of care payment – not available under either PIP or AA – also remains payable into older age.

16. Introducing disability into LINDA

In keeping with the remainder of LINDA's structure, each of the three disability-related characteristics that have been added to the model are assumed to evolve through time in a dynamic fashion. Accommodating this type of functionality in a model like LINDA is computationally demanding, which has had an important impact on the design of the module.

The disability module in LINDA has been developed with the objective of limiting the increase in computational burden to a factor of approximately 60. This target implies a total run-time for the base model specification of 2.5 days on prevailing computing technology, which is the upper bound of what we consider useful for the purposes of policy experimentation. We have divided this computational 'budget' between the three new characteristics in the following way.

There is little flexibility in the definition of carer responsibilities introduced into the model, which increases the computation burden by a factor of approximately 3. We have structured the approach taken to introduce disabled children into the model, so that it increases the computational burden of the model by a factor of 5/3. This leaves a factor of 12 (= 60 x 3 / 3 / 5) for the specification of disability states.

Please note that disability, as it has been included in the model, pushes computing technology to the limits of what is currently feasible. One implication is that the model cannot be used to project backward in time when adult disability, child disability, carers and DLA status are all enabled. The Excel front-end has been set up to manage the computational limitations in the background.

We describe the approach taken to simulate each characteristic in turn.

16.1 Adult disability

16.1.1 Summary

A "disability state" is generated for each benefit unit in every simulated year. In the case of singles, this disability state defines the disability of the relevant adult. In the case of couples, the disability state defines the 'disability combination' of the two spouses. The disability states recognised by the model are drawn from a discrete set of alternatives, the number of which is a model parameter that can be adjusted as desired. The simulated disability states evolve through time, based on exogenously defined transition probabilities that vary by each adult's prevailing disability state, education, age, and year.

The simulated disability state can influence benefit units in a variety of ways. As noted above, the disability state of each adult can affect their likely disability condition in the future. This feature is required to capture the persistence that is associated with many forms of disability, which may have an important bearing on the evolution of poverty. The disability state of one adult in a couple can be defined to affect the carer responsibilities of their spouse or a mature child, a subject that is discussed in Section 16.3. The disability state can also be defined to affect the labour opportunities (unemployment rate, see Section 4.2.3), wage potential (see section 4.2.1), non-discretionary costs (see Section 3), and transfer payments of a benefit unit. Furthermore, an adult's disability state can be defined to influence their probable relationship status in prospective years.

Care must be exercised when defining this aspect of the model to ensure feasible computational times in context of prevailing computing technology. Consider, for example, specifying the model to distinguish between 4 discrete disability states in each year throughout each adult's simulated lifetime. In this case, any single adult would be associated with four potential disability states, and any couple would be associated with 16 alternative disability combinations (4 for each spouse). This would expand the size of the state space by a factor of 10 ((4+16)/2). If transitions between simulated years are permitted between any given disability state and any of the alternative disability states, then this would expand the computations involved when evaluating expectations by a factor of 136 (= (4*4 + 16*16)/2). As the standard model runs in approximately 1 hour on contemporary computing technology, increases in computational burden of this magnitude are highly undesirable.

The model is designed to limit the computational problem outlined in the preceding paragraph in two ways. First, some disability transitions occur with low probabilities. The model is designed to ignore evaluation of expectations that are associated with (near) zero probabilities. Secondly, it may make practical sense to ignore some disability combinations for adult couples. This would be the case, for example, if the incidence of two severely disabled adults in a couple was very low over a specific age band. Ignoring very unusual combinations of this sort can help to improve model performance. The model is designed to adapt to such observations, as described in the next subsection.

16.1.2 Base specification

- *Defining disability states*

As noted at the start of this section, disability has been specified in the base specification of the model to increase the model's computational burden by a factor of 12. This is a challenging target to meet. We selected four disability states for inclusion in the base model specification, to balance the target computational burden against the eligibility criteria imposed by key disability-related transfer policies (described in Section 15). Adults are assumed to be able to transition to any one of the four disability states from one year to the next, subject to transition probabilities that vary by age and prevailing disability status

1) *Non-disabled*: This is the default condition, in which individuals are assumed to be physically capable of work, not to require any care, and subject to 'normal' living costs.
2) *Moderate work-limiting disability*: This disability state is assumed to impede work effort, and to receive recognition through the transfer system via eligibility for ESA work related activity benefits. Affected individuals are assumed to be physically capable of work, not to require any care, and subject to 'normal' living costs.
3) *Significant work-limiting disability*: This disability state is assumed to impede work effort, and to receive recognition through the transfer system via eligibility to both ESA support and PIP benefits. Affected individuals are assumed to be capable of limited work effort, may require care up to but not exceeding the day or night but not both. Individuals with significant work-limiting disabilities are also assumed to be subject to higher living costs, reflecting the value of DLA/PIP.
4) *Severe impairment*: This disability state is assumed to prevent any work effort, and to receive recognition through the transfer system via ESA support and PIP benefits. Affected

individuals are assumed to require care during both the day and the evening, and are subject to increased living costs.

- *Disability and the benefits system*

Disability state 1 (non-disabled) is treated as described in Section 4.6. From this base specification, disability state 2 exchanges income support benefits during the working lifetime for ESA work-related activity benefits (described in Section 15.1), if the ESA provides a higher rate of payment (see Section 15.4.1). From state pension age, disability state 2 is treated equivalently to disability state 1.

Individuals with disability states 3 and 4 are assumed to be eligible for the ESA. Furthermore, if these individuals are under state pension age, then they are considered eligible for the DLA/PIP. If they are over state pension age, then they are only considered eligible for the DLA/PIP if the respective benefit was received in the immediately preceding period; otherwise they are considered eligible for the AA.

Disability states 3 and 4 are defined to distinguish individuals eligible for the enhanced/higher rates of daily living allowances payable under the PIP/AA, but do not permit any finer disaggregation. This limitation is sufficient to capture AA in its entirety, but cannot identify eligibility for all feasible award rates for DLA/PIP. The rates of payment simulated for DLA/PIP are calculated by the model to average over constituent case-load numbers observed for the DLA in August 2011. These case-load numbers are reported in Table 16.1.

Table 16.1: Disability Living Allowance caseload numbers in August 2011 ('000)

care	mobility award			
award	higher	middle	lower	nil
		working aged		
higher	390	262	121	7
middle	617	237	339	41
lower	587	233	185	170
nil	235	159	76	-
		pension aged		
higher	247	218	18	10
middle	282	218	46	17
lower	296	191	31	74
nil	231	209	23	-

Source: DWP Information, Governance and Security, Work and Pensions Longitudinal Study.
Notes: state pension age: The age at which women reach State Pension age is gradually increasing from 60 to 65 between April 2010 and April 2016 to November 2018. Under current legislation, State Pension age for men and women is planned to increase to: 66 between November 2018 and October 2020; 67 between 2034 and 2036; 68 between 2044 and 2046. This will introduce a small increase to the number of working age benefit recipients and a small reduction to the number of pension age recipients. Figures from May 2010 onwards reflect this change. For more information see http://statistics.dwp.gov.uk/asd/espa.pdf. Caseload (Thousands) Totals show the number of people in receipt of an allowance, and exclude people with entitlement where the payment has been suspended, for example if they are in hospital.

- *Disability and the reference cross-section*

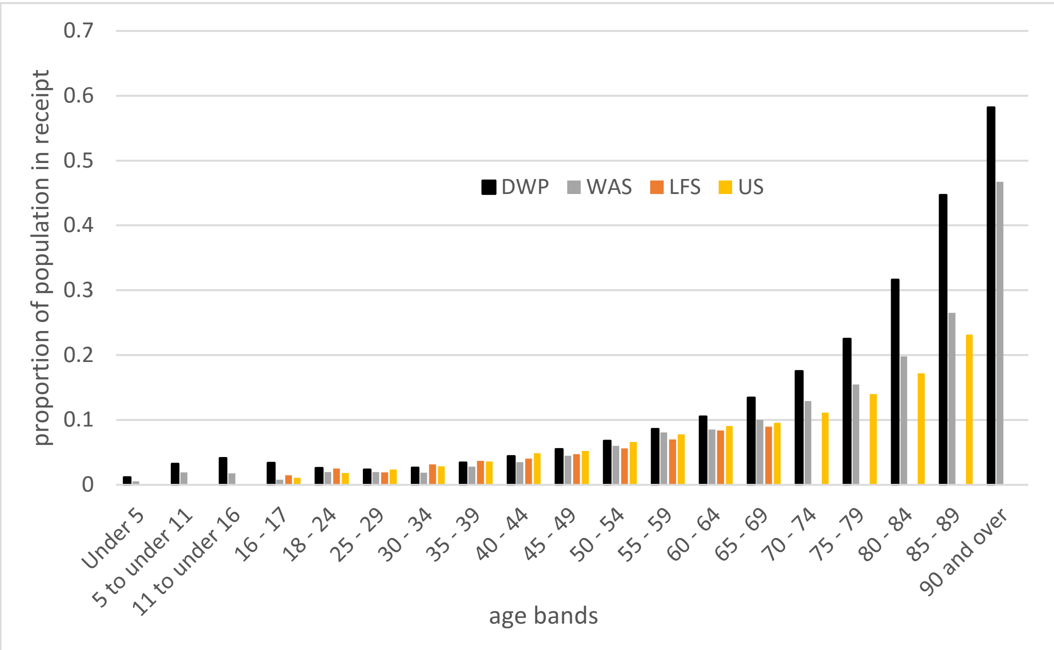
As discussed in Section 5, model projections begin from micro-data reported for 2011 by the Wealth and Assets Survey (WAS). Individuals with disability categories 3 or 4 were identified as those reported to be in receipt of DLA or AA payments in the WAS (either personally or via another household member). Figure 16.1 compares the incidence of receipt of these benefits by age described in wave 3 of the WAS, relative to population data reported by the DWP and ONS.

Figure 16.1 indicates that the incidence of DLA and AA reported by the WAS and the national aggregate statistics describe similar profiles with age, in a range of between 2 and 3 percentage points to age 45-49, before rising steeply into older age. It is notable, however, that the incidence of benefits receipt rises less steeply into old age in the WAS than it does in the national aggregate statistics.

The scale of the disparity identified between the incidence of DLA/PIP/AA receipt implied by DWP client data and the WAS motivated a comparison with associated statistics reported by the April-June quarterly Labour Force Survey (2011), and waves 1 to 4 of the Understanding Society survey. These alternative statistics are also reported in Figure 16.1, and indicate a fairly close correspondence between all three random-sample surveys, which contrast with the (DWP) population data. This suggests that common distortionary effects are likely to affect all three random surveys, including imperfect population coverage (particularly omission of individuals in institutional care), survey response rates, and so on. It is important to bear in mind that these limitations will feed through to the parameters defined on the basis of these survey data sources.

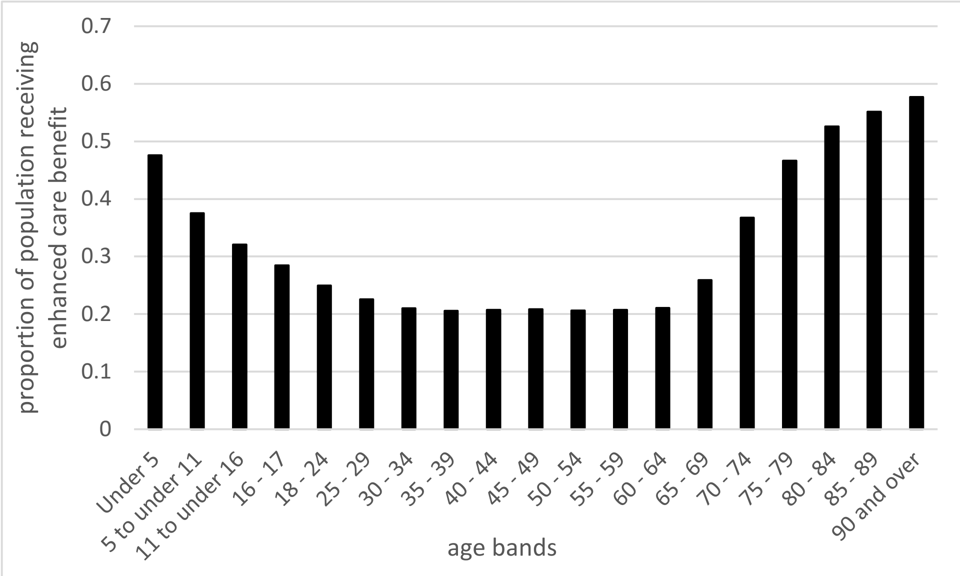
We consider the case load numbers reported by the DWP to be the most accurate description of benefits receipt for the UK. We have consequently imputed eligibility for disability benefits to match the WAS to the DWP rates of DLA/AA receipt. This imputation was implemented with reference to data reported by the WAS concerning the incidence of disabilities that resulted in ‘substantial difficulties’ with mobility.

Figure 16.1: Reported incidence of Disability Living Allowance / Attendance Allowance in 2011



Source: Author's calculations.
DWP series calculated using data from ONS mid-2011 UK population estimates, Table 1, and caseload numbers reported for August 2011 by the DWP Statistical Tabulations Tool
WAS series calculated using data from Wave 3 of the Wealth and Assets Survey
LFS series calculated using data from the April to June quarter of the Labour Force Survey, providing data to age 69
US series calculated using data from waves 1 to 4 of the Understanding Society survey

Figure 16.2: Proportion of Disability Living Allowance or Attendance Allowance recipients paid the higher care benefit in 2011

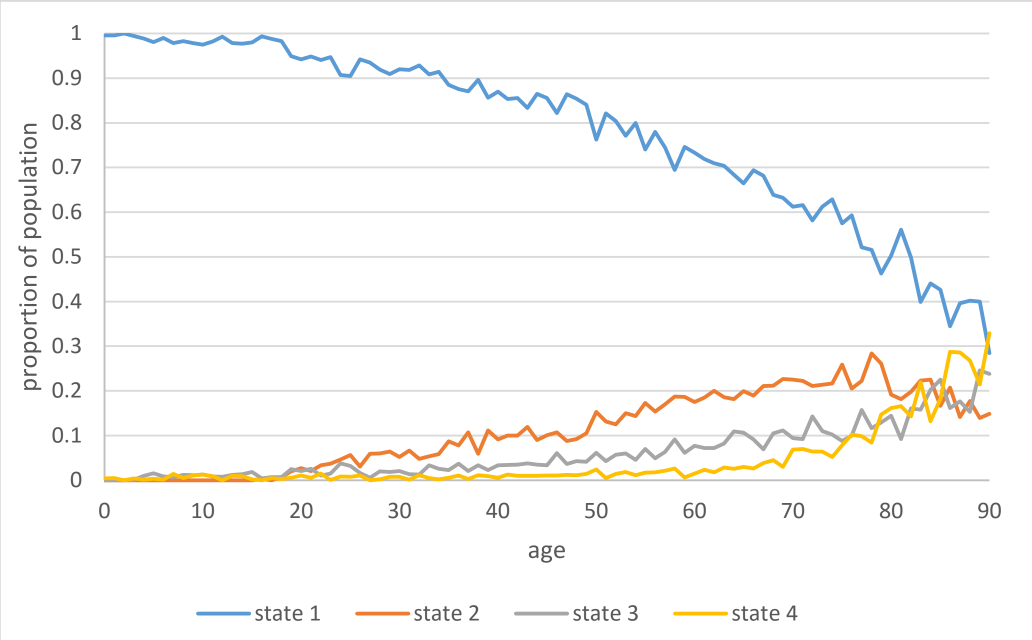


Source: Author's calculations using data from caseload numbers reported for August 2011 by the DWP Statistical Tabulations Tool

As noted previously, the distinction between disability states 3 and 4 defined above is designed to capture the rules governing receipt of standard and enhanced care benefits awarded by PIP and AA. The incidence of these states among the adult population described by the WAS was identified via random selection, to reflect the proportion of all PIP and AA recipients receiving the enhanced rate of care benefits described by national aggregate data. These population shares are reported for information in Figure 16.2.

Adults reported by the WAS were identified as having disability state 2 if they were not identified as having disability state 3 or 4 (discussed above), and indicated in the survey that they had a "long-standing illness, disability or infirmity" that "limit(s) your activities in any way". Given identification of disability state 2 individuals, those with no disability (state 1) were identified as the residual. The age-specific shares of the reference population cross-section allocated to each of the four disability states are reported in Figure 16.3.

Figure 16.3: The incidence of disability conditions among the 2011 population cross-section



Source: Author's calculations using data from Wave 3 of the Wealth and Assets Survey
Notes: 'state 1' = no disability, 'state 2' = moderate work-limiting disability, 'state 3' = significant work-limiting disability, and 'state 4' = severe impairment.

- *Disability and the simulated lifetime*

As noted above, starting from the reference cross-section, disability is projected through time to reflect the incidence of transitions reported in contemporary survey data.

Starting from the reference cross-section, the model evaluates the probability that an individual of a given age, education, and disability status will have any one of the four alternative disability states in the immediately succeeding year. Define the probability associated with disability category i, p(i). The model 'stacks' the evaluated probabilities, to obtain the set {p(1), p(1)+p(2), p(1)+p(2)+p(3), 1}. A random draw, *u*, is taken and compared against each element of the set in turn. If *u* < p(1), then the disability state in the immediately succeeding year is set equal to category 1; otherwise if *u* < p(1)+p(2), then disability state 2 is identified for the succeeding year, and so on. This procedure is repeated for every simulated year to build up an entire life history of disability for each adult in the model.[42]

The transition probabilities required for the procedure described above were calculated from the first four waves of the Understanding Society survey, covering the period 2009 to 2013 (centred upon 2011). This survey identifies, for each survey respondent aged 16 and over, in each year, whether DLA/PIP/AA was received and of what value, and whether the individual reported a persistent (12 months or more) health condition limiting common activities. This information permits identification of disability status in a similar fashion to that described for the WAS above. The additional detail concerning the value of the DLA/PIP/AA benefit received was used to

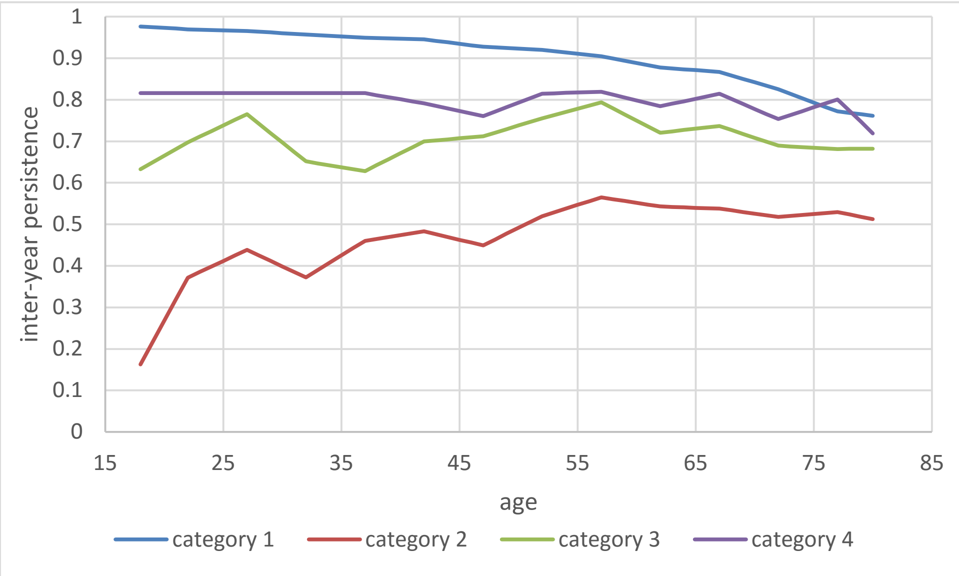
---

[42] A similar approach is used to project disability status backward through time, details of which can be obtained from the authors upon request.

distinguish between disability categories 3 and 4, using the higher rate of care payment for DLA in the prevailing survey year as a cut-off for category 4 individuals.

The full set of transition probabilities evaluated for the model can be found in the HEALTH PARAMETERS.xls file in the model base directory. Figure 16.4 reports the inter-year persistence probabilities for the four categories of disability as an example. These statistics indicate that the highest persistence is reported for the 'non-disabled', category 1. Here, persistence is close to complete early in life, and displays a gentle downward trend into old age. The next most persistent disability category is severe impairment, category 4, which is broadly stable through the life course at approximately 80%. Notably, this indicates that transitions about category 4 are not uni-directional, implying that some individuals may recover from a severely impaired state. Category 2, denoting moderate work-limiting disabilities, is the least persistent state. Most transitions out of category 2 are back into category 1, and vice-versa, so that some individuals may cycle into and out of a moderate work-limiting disability multiple times during their life courses.

Figure 16.4: Probabilities that disability category remains unchanged from one year to the next, by age and disability category



Source: Author's calculations using data from the first four waves of the Understanding Society survey.

The model can be directed to take disability state into consideration when solving the lifetime decision problem, only from age 60, or throughout the life course.

During ages from 61, the lifetime decision problem is solved treating disability category 2 equivalently to category 1. Furthermore, solutions to the decision problem assume that the reference adult is at least as impaired as their spouse. These assumptions reduce the number of considered disability combinations from 16 to 6.

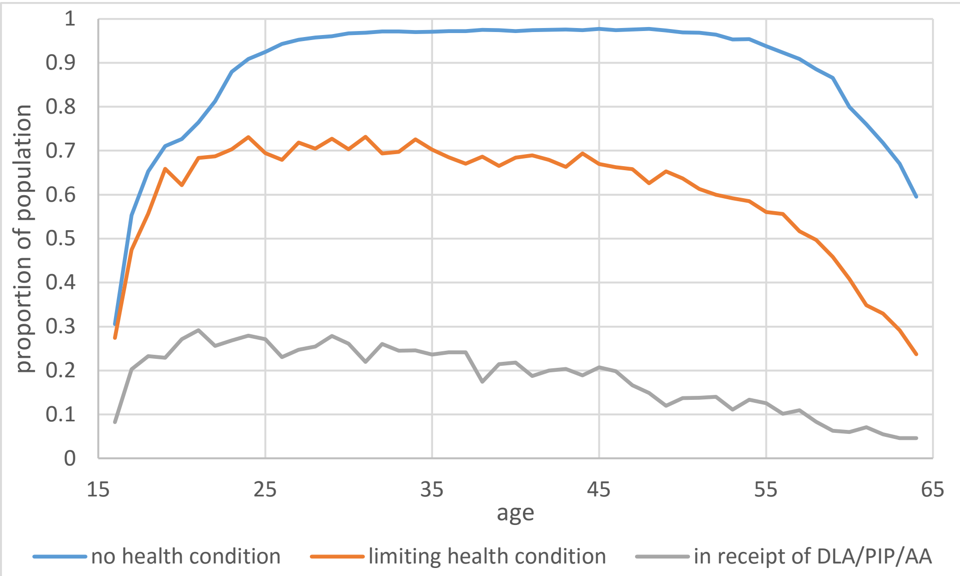
Ages between 18 and 60 treat disability category 3 equivalently to category 4. As in later life, the reference adult is assumed to be at least as impaired as their spouse. Furthermore the option where both spouses have category 2 disabilities is ignored. These restrictions reduced the number of considered disability combinations from 16 to 5. Note, however, that these restrictions are applied only when solving the lifetime decision problem; they are relaxed when the model actually projects individual circumstances through time.

- *Disability and employment opportunities*

The influence of each of the four disability states on simulated employment opportunities is parameterised to reflect data reported by the Labour Force Survey.

Quarterly waves of the LFS since 1998 permit identification of the same population subgroups distinguished by disability status as reported above for the WAS.[43] Discussion here focuses upon population averages calculated using data from the April to June waves of the quarterly LFS, reported between 1998 and 2014 (the most recently available data at the time of writing). These averages are useful because they help to mitigate the temporal volatility and small sample issues that affect year specific statistics. Rates of labour force participation and unemployment by age and disability status are reported for men in Figure 16.5; similar data are displayed in Appendix E for women.

Figure 16.5: Average rates of labour force participation and unemployment by age and disability status reported for men between 1998 and 2014



Panel A: Rates of labour force participation

[43] These waves distinguish individuals in receipt of some DLA/PIP/AA benefit, and individuals with a health problem that limits activity and has lasted at least 12 months.

Panel B: Rates of unemployment

Source: Author's calculations using data from the April to June wave of the quarterly Labour Force Survey between 1998 and 2014.
Notes: 'limiting health condition' defined as a condition that is reported to limit activity and to have persisted for at least 12 months.

The two panels of Figure 16.5 indicate that labour force participation and rates of unemployment display similar trends with age amongst the three disability categories distinguished here. Labour force participation tends to rise steeply from the beginning of adult life, to peak somewhere in the middle of the working lifetime, before falling away toward retirement. Unemployment rates, in contrast, are highest at the beginning of adult life, and display a general downward trend that is more rapid in early years, and trails away later in life. Within these broad similarities, very distinct differences can be observed between the three population subgroups. Rates of labour participation are reported to decrease, and rates of unemployment to rise appreciably with the reported severity of the disability condition.

The model is parameterised to capture the stylised observations discussed above. First, the model assumes that individuals identified with disability category 3 (significant work-limiting disability) can work at most part-time, and individuals with disability category 4 (severe impairment) are unfit for work of any kind. These assumptions are designed to capture the low rates of employment participation reported for individuals in receipt of Disability Living Allowance / Personal Independence Payment / Attendance Allowance.

Furthermore, the differences in unemployment rates by health condition that are described in Panel B of Figure 16.5 are included as parameters in the model. These parameters alter the likelihood of job offers with respect to each individual's simulated disability status.

- *Disability and wage potential*

The effects of alternative disability states on wage potential are parameterised to reflect regression estimates for reduced form descriptions of the (log) hourly wage rate.

Data were obtained from each of the quarterly labour force surveys reported between the first quarter 2004 to the last quarter 2014. The sample was limited to individuals reported in their first wave, aged between 16 and 59 years, and with reported gross hourly pay greater than 0 and less than £100. This delivered a total sample size of 136,366 men and 151,742 women for analysis.

Weighted regressions of log hourly pay were calculated, separately for men and women. Hourly pay was described as a linear function of a set of explanatory variables, including age indicators, age interacted with disability status, and a set of additional controls including region of residence, ethnicity, marital status, number and age of dependent children, and time. Three disability states were considered for analysis, reflecting the "1; non-disabled", "2; moderate work-limiting", and "3 and 4; severely disabled" categories (as discussed in the subsection on 'work opportunities' above). Ordinary Least Squares regression results that describe the estimated relationship between hourly wages and disability status are reported in Table 16.1; full results are reported in Appendix F.

The full regression specification includes, for each observed individual, both an age indicator variable and the age indicator interacted with each of the two disability indicators reported in Table 16.1. The statistics reported in the table can consequently be interpreted as the (marginal) age-specific impact of the respective disability condition on hourly wages, relative to the wider population. The proportional penalty on hourly wages can be obtained by taking the exponent of the respective coefficient and subtracting 1.

The estimates reported in the table indicate that statistically significant wage penalties are associated with the identified disability conditions, relative to the wider population. For men reporting 'moderate' conditions, the wage penalty roughly describes a hump-shape with age, rising from approximately 10%, to peak at just under 20% during the 40-44 age band, before falling gradually into higher age bands. A roughly similar age profile is estimated for men reporting a 'severe disability', although associated penalties are approximately 10 percentage points higher than those with 'moderate' disabilities on average.

Although the estimated wage penalties for women are statistically significant, they tend to be substantially smaller than those estimated for men. There is also a less pronounced difference between the estimates obtained for women with 'moderate' and 'severe' disabilities, relative to men. Averaging over all age bands, the wage penalty for women with moderate disabilities is 10% (c.f. 16% for men), for women with severe disabilities is 13% (c.f. 28% for men).

Table 16.1: Estimated impact of disability status on log hourly wage rates

	men		women	
age band	coef	std error	coef	std error
moderate work-limiting disability				
16-19	-0.0974	0.0351	-0.0430	0.0273
20-24	-0.0739	0.0188	-0.0808	0.0206
25-29	-0.1219	0.0187	-0.1151	0.0166
30-34	-0.2165	0.0176	-0.0810	0.0158
35-39	-0.2064	0.0166	-0.1044	0.0148
40-44	-0.2203	0.0156	-0.1313	0.0141
45-49	-0.2111	0.0149	-0.1256	0.0126
50-54	-0.2071	0.0146	-0.1199	0.0121
55-59	-0.1875	0.0148	-0.1186	0.0118

significant disability				
16-19	-0.0836	0.0724	-0.0214	0.0895
20-24	-0.3204	0.0663	-0.1386	0.0641
25-29	-0.3661	0.0530	-0.1998	0.0459
30-34	-0.3817	0.0771	-0.1716	0.0373
35-39	-0.3600	0.0634	-0.1420	0.0360
40-44	-0.4232	0.0539	-0.1649	0.0299
45-49	-0.3260	0.0608	-0.1078	0.0352
50-54	-0.3560	0.0654	-0.1123	0.0394
55-59	-0.3326	0.0707	-0.2226	0.0577
observations	136366		151742	
R-squared	0.2537		0.1885	

Source: authors' calculations on quarterly Labour Force Survey data, Jan-Mar quarter 2004 to Oct-Dec quarter 2014
Notes: Statistics calculated using the "regress" Stata command, and the "robust" option. Full regression specification includes both age indicator variables and age interacted with disability status. "significant disability" identified as receipt of Disability Living Allowance, Personal Independence Payment, or Attendance Allowance. "moderate work-limiting disability" identifies individuals reporting a long-lasting health condition that limits normal activities, and who are not identified as "significantly disabled".

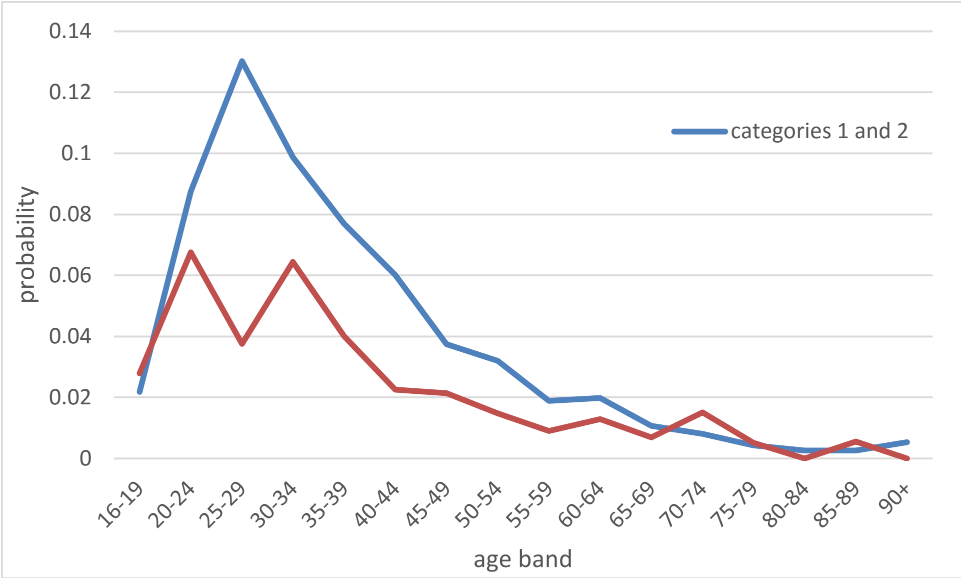
- *Disability and non-discretionary costs*

The model currently sets non-discretionary costs equal to DLA/PIP/AA payments. We may need to refine this aspect of the model in light of discussions with the JRF.

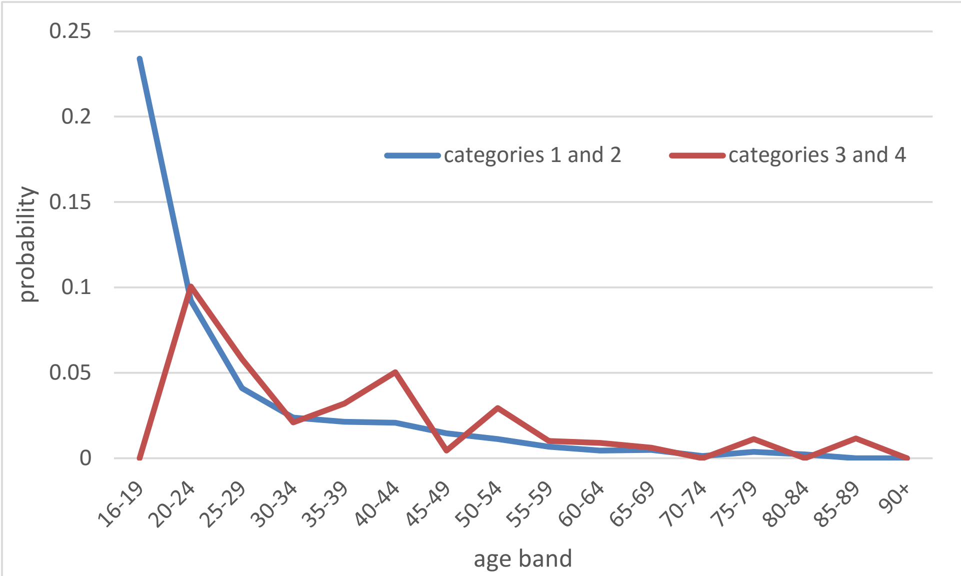
- *Disability and marriage rates*

Disability status is assumed to influence age specific rates of marriage and divorce. The parameters governing this aspect of the model were evaluated from the same Understanding Society data as considered for the transition probabilities between disability states. The model distinguishes between only two disability classes when evaluating probabilities of marriage and divorce, with disability categories 3 and above treated separately from disability categories 2 and below. The distinction between these two population subgroups is reported in Figure 16.6.

Figure 16.5: Age specific probabilities of marriage and divorce reported between 2009 and 2013



Panel A: marriage rates



Panel B: divorce rates

Source: Author's calculations using data from waves 1 to 4 of *Understanding Society*.
Notes: categories 3 and 4 are distinguished from categories 1 and 2 by the fact that they are identified as in receipt of Disability Living Allowance, Personal Independence Payment, or Attendance Allowance. Marriage rates are calculated as the number of marriages divided by the number of non-married. Divorce rates are calculated as the number of divorces divided by the number married.

16.2 Child disability

16.2.1 Summary

Prior to this project, LINDA could be used to simulate the birth and aging of dependent children (see Section 4.8). This aspect of the model has been extended to distinguish between disabled and non-disabled children. As discussed in Section 4.8, the model is designed to limit computational burden by restricting child birth to a discrete set of "birth years". The model can now be directed to identify whether a child is disabled in each considered birth year.

In contrast to the approach taken to simulate disability states for adults, child disability is assigned at birth, and is assumed to remain invariant to the assumed age of maturity. Child disability influences simulated benefit unit costs, benefits (discussed in Section 15), and the carer responsibilities of responsible adults (discussed in Section 16.3). Furthermore, disability status is assumed to follow the child into maturity, and to influence their education status upon maturity. It is consequently possible to use the model to track the influence of disability throughout the life course, from birth through to death.

16.2.2 Base specification

- *Defining disabled children*

Disabled children are defined as those eligible for DLA payments, as described in Section 15.2. This corresponds to adult disability states 3 and 4, as discussed in Section 16.1. Disabled children are assumed to be eligible to DLA payments, in an equivalent fashion to adults with disability state 3.

- *Disabled children and the reference cross-section*

Disabled children were identified as those in receipt of DLA payments in the WAS (either personally or via another household member). This is consistent with the approach adopted to identify disabled adults, as described in Section 16.1.2. In contrast to the approach taken for adults, however, no attempt was made to align the incidence of benefits receipt among children to client numbers reported by DWP.

- *Disabled children and the simulated lifetime*

Up to one child can be disabled in each birth year. The incidence of disability among children is randomly allocated by the model, with the probability of disability set to 1.36%. This probability is equal to the incidence of children in receipt of DLA benefits reported in wave 3 of the WAS. After a disabled child enters a benefit unit, they are assumed to remain a member of the benefit until either they leave the benefit unit due to divorce, or reach the assumed age of maturity (17 years). In the case of divorce, disabled children are withdrawn from the benefit unit of the reference adult after non-disabled children of the same age.

At the time of maturity, a disability state is drawn for each newly maturing adult. This state is drawn from a set comprised of four alternatives for any maturing child who is not identified as disabled, as discussed in Section 16.1. The disability state for maturing children who are identified as disabled is drawn from a set of alternatives limited to states that preclude the possibility that the individual will act as a carer, as described in Section 16.3. In the base model specification, this means that disabled children are assumed to remain eligible for DLA after they reach adulthood, corresponding to disability states 3 and 4 as described in Section 16.1. The division of disabled children between disability states 3 and 4 is randomly selected to reflect the relative incidence of states 3 and 4 reported for 18 year olds in the WAS (as discussed in Section 16.1).

Maturing children who are identified as disabled are subject to a different set of transition probabilities governing the projection of education status than the remainder of the population (reported in Table 14.2). The influence of child disability on education transition probabilities is specified to reflect the relative incidence of alternative education categories described by data reported for individuals aged 25 to 30 by the April-June quarters of the Labour Force Survey between the years 2004 and 2014. Associated statistics are reported in Table 16.2. These data reveal that individuals at the beginning of their working lifetimes who are in receipt of disability related benefits (DLA) are substantively less likely to attain a graduate degree, and more likely to have no recognised qualification. These differences feed into the model, to better capture the bearing of disability on lifetime outcomes.

Table 16.2: Effects of child disability on education attainment

		child health	
child qualifications	all	no DLA	DLA
no qualification	0.0728	0.0679	0.3068
other qualification	0.1187	0.1173	0.1867
GCSEs	0.1960	0.1948	0.2504
A-levels	0.2864	0.2884	0.1886
degree	0.3261	0.3316	0.0675
all	1.0000	1.0000	1.0000
observations	81709	79946	1763

Source: Authors' calculations using data reported for 25 to 30 year olds in April to June quarters of the LFS between 2004 and 2014
All statistics report weighted averages and omitted observations with missing data. Statistics refer to highest reported qualifications held, as reported by variables hiqual4d, hiqual5d, hiqual8d, and hiqul11d. A-levels category includes both A-level qualifications and "other higher qualifications".

16.3 Carer responsibilities

16.3.1 Summary

A "carer state" is generated for one adult in each benefit unit in each simulated year. The carer state evolves through time, based on exogenously defined transition probabilities that vary by the individual's prevailing carer state, the disability state of their spouse (see section 16.1) and/or dependent children (see section 16.2), age, and year. Carers can be distinguished from other adults in regards to the benefits that they are eligible for, their employment opportunities, and the time that they have available for leisure.

16.3.2 Base specification

- *Defining Carers*

A carer is defined in the model as an adult who spends at least 20 hours per week on average caring for a (cohabitating) spouse, their disabled children, or an individual outside of their own benefit unit. Beyond their dependent children, any one individual is assumed to be able to care for only one individual at a time.

- *Carers and the benefits system*

Carers are assumed to be eligible for the Carer's Allowance (CA), as described in section 15.4. This assumes that a compelling case can be made for the gap between the 20 hour minimum assumed for defining carers, and the 35 hour minimum required for the CA. Beyond this, carers are subject to the same taxes and benefits as considered for other adults, as described in section 4.6.

- *Identifying carers*

Carer responsibilities are assigned at the level of the benefit unit. A "carer" benefit unit is one that includes one adult with carer responsibilities. Only benefit units that include at least one adult with disability state 1 or 2 (as defined in Section 16.1) can provide care services. Carer services are provided by the adult with disability state 1 or 2 who has the lowest wage potential within a carer benefit unit. Carer benefit units are identified in the model in a way that is adapted to a number of important limitations of existing publically available data sources.

Wave 3 of the Wealth and Assets Survey, from which data for the reference population cross-section are drawn, does not provide any information concerning carer responsibilities of respondents. Carer responsibilities are consequently imputed for the reference cross-section in a similar fashion to the approach that is adopted for the remainder of the simulated time period.

As discussed in Section 16.2, the definition of child disability in the model is aligned with receipt of DLA benefits. Although this is useful when exploring the implications of associated policy alternatives, it complicates parameterisation of the model as data concerning receipt of DLA by children are scarce. Following an extensive exploration, data reported by waves 1 and 2 of the WAS were found to provide the most complete description of the detail required for identifying transition probabilities into and out of carer status.

Transitions into and out of care are simulated using a logit regression specification. The logit equation describes carer transition probabilities for benefit units as a function of the reference person's age, graduate status, number of dependent children aged under 5, aged 5 to 10, and aged 11 to 17, whether any children in the benefit unit were disabled and under 5, between ages 5 and 10, or between ages 11 and 17 children, whether one spouse has disability state 3 or 4, and whether the benefit unit included a carer in the preceding year.[44]

Wave 2 of the WAS reports data that permit identification of disability status among all individuals in the simulated population, in addition to the other demographic characteristics for benefit units referred to in the preceding paragraph (in common with wave 3 of the survey, as discussed throughout this report). Unlike wave 3 of the survey, however, wave 2 also reports three variables useful for identifying carer responsibilities. "carer1qw2" identifies survey respondents who indicated that they provided some care services. "carehrw2" describes the hours spent providing care, permitting identification of individuals providing at least 20 hours of care. "caredur" reports how long the respondent had been providing care services. This last variable permits identification of individuals who had been providing care for in excess of a year from those who had been carers for 12 months or less.

The set of data referred to in the preceding paragraph describe all of the variables included in the logit regression, except for carer status in the preceding year where no carer exists in the prevailing

---

[44] Marital status was omitted from the specification as associated coefficient estimates were found to be statistically insignificant.

year. This detail is important, as it helps to identify transition rates out of carer status. The missing detail was imputed by drawing upon data reported by wave 1 of the WAS.[45]

Waves 1 and 2 of the WAS are separated by a two year interval. Wave 1 of the WAS reports the variable "carer1qw1", which is analogous to variable "carer1qw2" reported in wave 2 of the survey, but omits any further detail. The incidence of carers in the year preceding wave 2 of the WAS was imputed for benefit units that did not include a carer in wave 2, via the following procedure.

If carer1qw1 = carer1qw2, then it was assumed that the status of the benefit unit remained unaltered between waves 1 and 2 of the survey, implying no carer existed in the year preceding wave 2. Similarly, if carer1qw1 did not indicate a carer in wave 1 of the survey, then it was assumed that no carer existed in the year preceding wave 2 of the survey. If carer1qw1 did indicate a carer in wave 1, but carer1qw2 did not indicate a carer in wave 2 of the survey, then a random draw from a uniform distribution was taken and compared against a proportional incidence statistic to impute carer status in the year preceding wave 2. The incidence statistic was evaluated on wave 2 data, equal to the proportion of benefit units reporting a carer in variable care1qw2, that satisfied both the health condition (disability state 1 or 2), and the hours condition (at least 20 hours of care) imposed on carers in the model. Furthermore, the incidence statistic was divided by two to accommodate the two year gap between waves 1 and 2 of the WAS.

Two logit regressions were estimated using the WAS data referred to above, one that omits the indicator for carers in the preceding year and another that includes the indicator. The former of these regression specifications is used to impute carers for the reference cross-section and the latter for transition probabilities during the simulated lifetime. The impact of random sampling on regression estimates was mitigated for the specification that includes the indicator for carers in the preceding year by re-sampling and estimating the logit equation 20 times. Regression statistics are reported in Table 16.2.

The quadratic specification in age was selected following exploratory regressions using age band dummy variables. This age profile implies increasing probabilities of caring to age 60, followed by a gradual decline into older age. Age is top-coded to 80 for the regression to avoid end-point issues with the polynomial specification at very advanced ages.

The parameter estimates indicate that graduates, and benefit units with young children, none of whom are disabled, are less likely to be carers on average. The reverse is weakly true for benefit units with older dependent children. The logit coefficients on the indicator variables for disabled children and disabled spouses are similar in magnitude, and are both highly significant, indicating a higher likelihood of caring. The coefficient on the indicator variable for carers in the preceding year is also positive and highly significant, underscoring the persistence associated with carer responsibilities. The importance of this variable is also indicated by the substantial increase reported for the psuedo R-squared, when it is included in the regression specification.

Table 16.2: Logit regression statistics for the incidence of benefit units with carers

[45] A detailed search was conducted to identify alternative panel data sources that could be used to estimate the logit regression considered here. A key limitation of both the Understanding Society and the Life Opportunities surveys is that they omit detail concerning receipt of DLA benefits by children.

	coeff	std. error	coeff	std. error
age	0.18180	0.01694	0.10744	0.02659
age^2	-0.00151	0.00015	-0.00092	0.00025
graduate	-0.36334	0.08813	-0.47397	0.15891
number of children by age, none disabled				
aged 0-4	-1.11440	0.31131	-1.27851	0.40884
aged 5-10	-1.57055	0.27224	-1.75018	0.35178
aged 11-17	0.19244	0.07317	0.34444	0.13954
disabled child	2.97482	0.19654	1.87534	0.51177
disabled spouse	2.33706	0.08503	1.86833	0.20638
carer in preceding year			6.78013	0.14370
constant	-8.3720	0.4677	-8.4768	0.6995
psuedo R^2	0.1633		0.7573	

Source: authors' calculations on data from waves 1 and 2 of the Wealth and Assets Survey
Notes: sample limited to benefit units reported in wave 2 of the WAS. Sample size = 23061. Disability defined as in receipt of DLA benefits. Carer in preceding year imputed based on recall questions for carers in wave 2, and via random sampling for non-carers in wave 2 who were carers in wave 1. Reported regression statistics for specification that includes carer in preceding year as an explanatory variable are averages taken over 20 random samples.

- *Carers and employment opportunities*

Carers are assumed to have the same employment opportunities as other simulated adults, with exceptions. First, carers are assumed to be subject to a fixed time penalty representing the hours that they are responsible for providing care. Secondly, carers are assumed to be able to work, at most, half-time; full-time work is not permitted.

References

Andersson, I., Brose, P., Flood, L., Klevmarken, N. A., Olovsson, P. & Tasizan A. (1992), *MICROHUS - A microsimulation model of the Swedish household sector*. International Symposium on Economic Modelling: Gothenburg.

Attanasio, O. P., Bottazzi, R., Low, H. W., Nesheim, L. & Wakefield, M. (2012), “Modelling the demand for housing of the life cycle”, *Review of Economic Dynamics*, 15, pp. 1-18.

Basu, N., Pryor, R. & Quint, T. (1998), “ASPEN: A microsimulation model of the economy”, *Computational Economics*, 12, pp. 223-241.

Blundell, R., Brewer, M. and Shephard, A. (2005), “Evaluating the labour market impact of Working Families’ Tax Credit using difference-in-differences”, *HM Revenue and Customs*, ISBN 1-904983-12-X.

Blundell, R, Meghir, C, Symons, E, and Walker, I. (1984), “A Labour Supply Model for the Simulation of Tax and Benefit Reforms”. In Blundell, R. and Walker, I., (eds.) *Unemployment, Search and Labour Supply*. Cambridge University Press: Cambridge.

Browning, M. Lusardi, A. (1996), “Household saving: micro theories and macro facts”, *Journal of Economic Literature*, 34, pp. 1797-1855.

Berthoud, R. (2010), “The take-up of Carer’s Allowance: A feasibility study”, *Department for Work and Pensions Working Paper* 84.

Caldwell, S. (1997), *Corsim 3.0 User and Technical Documentation.* Ithaca: New York.

Carrol, C. (1992), “The buffer-stock theory of saving: Some macroeconomic evidence”, *Brookings Papers on Economic Activity*, 23, pp. 61-156.

Creedy, J., Duncan, A.S., Harris, M. & Scutella, R. (2002). *Microsimulation Modelling of Taxation and The Labour Market: The Melbourne Institute Tax and Transfer Simulator*. Cheltenham: Edward Elgar.

Curry, C. (1996), “PENSIM: A dynamic simulation model of pensioners’ income”, *Government Economic Service Working Paper* 129.

Deaton, A. (1991), “Saving and liquidity constraints”, *Econometrica*, 59, pp. 1221-1248.

Dueé, M. & Rebillard, C. (2004), “Old age disability in France: A long-run projection”, *INSEE Working Paper* G2004/02.Duncan, A. and Harris, M.N. (2002), “Simulating the effect of welfare reforms among sole parents in Australia”, *Economic Record*, 78, pp. 249-63.

Duleep, H.O. & Dowhan, D.J. (2008), “Adding immigrants to microsimulation models”, *Social Security Bulletin*, 68.

Duncan, A. (2001), Microsimulation and Policy Setting in the United Kingdom, http://www.cerc.gouv.fr/meetings/seminaireoctobre2001/intervention1.doc

DWP (2015), *Income-related benefits: Estimates of take-up* - financial year 2013/14 (experimental).

Emmerson, C., Reed, H. & Shephard, A. (2004), “An assessment of PenSim2”, IFS WP04/12.

Falkingham, J. & Lessof, C. (1992), “Playing god: the construction of LIFEMOD”. In R. Hancock & H. Sutherland (eds), *Microsimulation Models for Public Policy Analysis: New Frontiers*. STICERD Occasional Paper, London School of Economics: London.

Giles, C. and McCrae, J. (1995), “TAXBEN: the IFS microsimulation tax and benefit model”, *IFS Working Paper* W95/19.

Harding, A. (1993), *Lifetime Income Distribution and Redistribution: Applications of a Microsimulation Model*. North-Holland: London.

Hayward, H., Hunt, E. & Lord, A. (2014), “The economic value of key intermediate qualifications: estimating the returns and lifetime productivity gains to GCSEs, A levels and apprenticeships”, *Department for Education Research Report*, December 2014.

HMRC (2015), *Child Benefit, Child Tax Credit and Working Tax Credit Take-up rates 2013-14*. Revised February 2016.

Hood, A. & Oakley, L. (2014), "A survey of the GB benefit system", IFS Briefing Note BN13.

Kasparova, D., Marsh, A. & Wilkinson, D. (2007), "The take-up rate of Disability Living Allowance and Attendance Allowance: Feasibility study", *Department for Work and Pensions Research Report* 442.

Kimball, M.S. (1990), "Precautionary saving in the small and in the large", *Econometrica*, 58, pp. 53-73.

King, A., Bækgaard, H. & Robinson, M. (1999), "The base data for DYNAMOD-2", *NATSEM Technical Paper* 20.

Lane, Clark and Peacock (2008), *Pension Buyouts 2008*, www.lcp.uk.com

Lucas, R. (1976), Econometric Policy Evaluation: A Critique. In Brunner, K. and Meltzer, A. *The Phillips Curve and Labor Markets*. Carnegie-Rochester Conference Series on Public Policy 1. New York: American Elsevier

Lucchino, P. and van de Ven, J. (2013), "Modelling the dynamic effects of transfer policy through the life-course: the LINDA policy analysis tool", *National Institute Discussion Paper* DP405.

Nelissen, J. H. M. (1994), *Income Redistribution and Social Security: An Application of Microsimulation*. Chapman and Hall: London.

Orcutt, G. (1957), "A new type of socio-economic system", *Review of Economics and Statistics*, 58, pp. 773-797.

O'Donoghue, C., Redway, H. & Lennon, J. (2010), "Simulating migration in the PENSIM2 dynamic microsimulation model", *International Journal of Microsimulation*, 3, pp. 65-79.

ONS (2014), *Quality of Long-Term International Migration estimates from 2001 to 2011*. Office for National Statistics.

Orcutt, G. (1957), "A new type of socio-economic system", *Review of Economics and Statistics,* 58, pp. 773-797.

Orcutt, G., Caldwell, S. B. & Wertheimer, R. (1976), *Policy Exploration Through Microanalytic Simulation*. Urban Institute: Washington D.C.

Sutherland, H. and Figari, F. (2013), "EUROMOD: the European Union tax-benefit microsimulation model", *International Journal of Microsimulation*, pp. 4-26.

Steventon, A., Sanchez, C. & Curry, C. (2007), "Increasing the value of saving in Personal Accounts: taking small pension pots as lump sums", *Equal Opportunities Commission Working Paper*, 58.

Van de Ven, J. (2013), "The influence of decision costs on investments in Individual Savings Accounts", *National Institute Discussion Paper* DP407.

Zaidi, A. (2007), "The SAGE model: A dynamic microsimulation population model for Britain". In W. Barnett (eds), *Modelling Our Future: Population Ageing, Health and Aged Care*. Emerald Group Publishing Limited: London.

Appendix A: Committed Expenditure

Appendix A.1: Housing

Appendix A.1.1: Incidence of home ownership and mortgage holding

The logit models used to impute home ownership and mortgage holding in the model are based upon estimated regression coefficients reported in Tables A.1 and A.2 respectively. In addition to the variables that are explicitly listed in the tables, the model includes a time variable for each relationship/age group combination that is specified to discount the respective wealth measure for projected wage growth. Specifically, for any population subgroup, the regression models can be represented by:

$$\begin{aligned} l_{i,t} &= \beta_0 + \beta_1 \log w_{i,t} \left(1+r\right)^{-t} \\ &= \beta_0 + \beta_1 \log w_{i,t} - t\beta_1 \log\left(1+r\right) \end{aligned}$$

The assumed rate of discounting, *r*, is a parameter in the model, which we align to the assumed rate of real wage growth by default.

Table A.1: Logit regression coefficients for incidence of home ownership, 2011

	singles		couples	
age group	constant	log(w)	constant	log(w)
under 25	-12.734	0.778	-11.062	1.231
25-29	-11.763	1.066	-7.677	0.840
30-34	-11.334	1.091	-7.400	0.835
35-39	-18.002	1.815	-5.609	0.697
40-44	-13.404	1.363	-7.186	0.859
45-49	-12.668	1.255	-9.618	1.112
50-54	-13.848	1.413	-12.465	1.362
55-59	-22.764	2.187	-11.661	1.264
60-64	-21.223	2.053	-20.603	2.007
65-69	-21.470	2.026	-32.791	3.052
70-74	-29.865	2.814	-40.564	3.764
75-79	-40.206	3.762	-33.383	3.121
80-84	-37.712	3.451	-143.043	13.525
85 and over	-28.949	2.619	-52.470	4.756
observations	6610		12852	

Notes: authors' calculations on weighted WAS data reported for 2011

log(w) denotes log of non-pension wealth

all parameters significant at 99% confidence interval

Table A.2: Logit regression coefficients for incidence of mortgage holding among home owners, 2011

	singles		couples	
age group	constant	log(w)	constant	log(w)
under 30	48.753	-4.019	13.435	-0.967
30-39	18.729	-1.423	19.819	-1.432
40-49	19.353	-1.549	15.113	-1.094
50-59	15.868	-1.325	11.796	-0.926
60-69	8.420	-0.807	9.237	-0.861
70 and over	14.108	-1.413	10.750	-1.100
observations	2,519		10,552	

Notes: authors' calculations on WAS data for 2011
log(w) denotes log of non-pension wealth
all parameters significant at 99% confidence interval

Appendix A.1.2: Net housing equity

Table A.3: Portfolio share invested in owner occupied housing among home owners

age	share	age	share	age	share
18	0.452	39	0.828	60	0.753
19	0.452	40	0.796	61	0.772
20	0.968	41	0.795	62	0.753
21	0.989	42	0.810	63	0.755
22	0.878	43	0.815	64	0.747
23	0.724	44	0.834	65	0.788
24	0.758	45	0.817	66	0.766
25	0.654	46	0.768	67	0.773
26	0.764	47	0.793	68	0.808
27	0.728	48	0.823	69	0.795
28	0.737	49	0.748	70	0.806
29	0.763	50	0.801	71	0.802
30	0.786	51	0.777	72	0.817
31	0.708	52	0.802	73	0.816
32	0.740	53	0.791	74	0.822
33	0.765	54	0.786	75	0.841
34	0.765	55	0.790	76	0.815
35	0.768	56	0.775	77	0.822
36	0.796	57	0.788	78	0.809
37	0.753	58	0.800	79	0.826
38	0.745	59	0.758	80	0.847

Notes: authors' calculations on WAS data for 2011
weighted portfolio shares calculated for all home owners
figure for age 18 set equal to figure for age 19 due to sample size
weight for 81 and over set equal to average weight reported for ages 76-80.

Appendix A.1.3: Mortgage debt

The Tobit regression used to impute mortgage debt in the model takes a similar form to the logit models used to impute home ownership and mortgage holding, as described in Appendix A.1.1. Time is accommodated in the specification so that non-pension wealth is discounted to 2011 by the assumed rate of real wage growth.

Table A.4: Tobit regression coefficients for the ratio of mortgage debt to net housing equity among mortgage holders

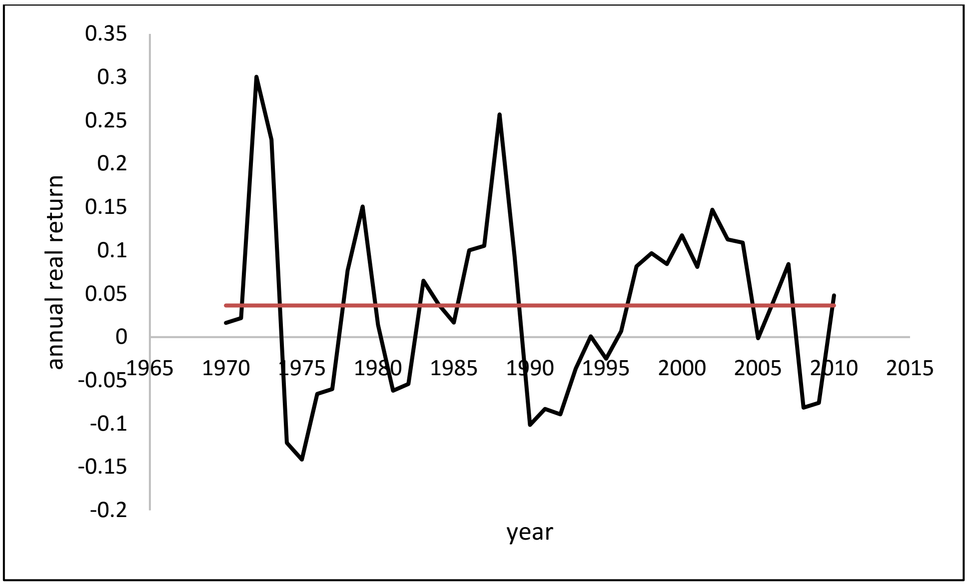
	singles		couples	
age group	constant	log(w)	constant	log(w)
under 25	36.710	-1.959	51.386	-4.406
25-29	59.467	-5.207	25.812	-1.981
30-34	50.890	-4.322	23.327	-1.811
35-39	31.831	-2.641	19.578	-1.532
40-44	20.826	-1.710	16.291	-1.265
45-49	8.691	-0.622	15.935	-1.241
50-54	18.131	-1.483	11.050	-0.846
55-59	10.663	-0.820	9.774	-0.750
60 and over	6.600	-0.485	4.532	-0.334
observations	811		5088	

Notes: authors' calculations on WAS data for 2011
log(w) denotes log of non-pension wealth
all parameters significant at 99% confidence interval

Appendix A.1.4: Real returns to housing wealth

The real rates of return to housing wealth that are assumed by the model are based on the historical mix-adjusted house price index reported by the ONS, discounted by the National Accounts consumption deflator (code YBGA). The historical series is projected through time by taking the average over the entire historical series (equal to 3.65% per annum).

Figure A.1: Real annual rate of growth of house prices by year



Source: ONS mix adjusted house price index deflated by the National Accounts consumption deflator

Appendix A.2: Childcare

Table A.5: Linear regression coefficients for childcare costs, by child age band, year, and parental employment status

	all adults full-time employed			one part-time employed			other		
child ages	0 to 4	5 to 13	14+	0 to 4	5 to 13	14+	0 to 4	5 to 13	14+
	estimated coefficient (£2011 p.w.)								
1990	12.96	4.12	0.71	7.34	0.91	0.81	1.22	0.86	1.04
1991	21.88	2.53	-0.68	5.14	0.99	1.65	1.56	0.50	2.57
1992	24.37	3.16	-1.86	8.38	-0.01	-0.06	2.04	-0.15	-0.13
1993	23.90	2.34	-0.10	8.12	0.27	-0.14	1.64	-0.06	0.00
1994	27.86	2.52	-0.36	8.97	0.26	0.25	2.44	-0.09	0.25
1995	21.96	1.80	-1.26	8.37	0.76	0.14	1.89	-0.04	-0.26
1996	32.59	2.99	-1.72	7.95	0.40	-0.41	1.78	0.06	-0.28
1997	24.57	0.98	-1.58	9.88	0.46	-0.53	2.05	0.09	-0.23
1998	35.18	5.11	-2.63	7.61	0.45	-0.43	2.77	0.03	0.16
1999	21.72	2.98	-1.40	7.40	0.15	-0.40	1.84	0.05	-0.01
2000	32.45	6.36	-1.67	14.47	0.64	0.67	4.29	0.50	-0.69
2001	28.67	6.93	-2.47	14.21	1.10	-0.31	2.23	0.62	-0.40
2002	41.07	6.21	-3.11	17.48	0.36	-0.46	1.87	0.63	-0.71
2003	38.64	2.37	-0.11	17.46	1.77	-1.14	2.77	0.60	0.12
2004	29.34	8.21	-1.60	19.98	0.55	-0.35	2.97	0.11	-0.39
2005	35.88	11.02	-1.46	16.62	1.82	-1.31	2.97	-0.12	0.25
2006	36.04	7.97	-1.25	16.87	0.41	-0.07	4.08	-0.01	0.13
2007	37.68	3.55	-2.19	17.23	1.47	0.16	5.04	0.36	-0.41
2008	43.61	7.90	-3.18	14.53	0.34	4.74	2.61	0.02	-0.30
2009	37.55	5.99	-1.96	15.04	1.21	-2.02	2.06	0.90	-0.70
2010	31.21	4.57	-2.46	17.93	0.45	0.81	4.59	-0.20	-0.47
2011	21.94	6.72	1.77	15.30	0.74	-0.32	3.43	0.52	1.02
2012	31.05	7.22	-3.99	15.02	0.31	2.38	5.24	-0.37	2.66
	range of year specific standard errors								
minimum	2.01	0.45	0.35	0.76	0.16	0.13	0.20	0.08	0.07
maximum	7.09	2.36	1.99	2.33	0.68	2.40	1.31	0.50	2.39

Source: Authors' calculations on LCFS data for specified years
dependent variable is weekly childcare expenditure in 2011 prices
expenditure discounted by the National Accounts final consumption deflator

Appendix B: Testing the Representativeness of the Wealth and Assets Survey Data

As the base population data for the model are drawn from the Wealth and Assets Survey (WAS), it is of fundamental importance that this data source provides information that can be reliably related to the wider UK population. We have consequently compared the base data reported by the WAS for Great Britain in 2011 against alternative data sources, and report the results obtained from this exercise here.

One of the principal uses of the model is to explore distributional implications of policy alternatives. It is therefore important that the base data for the model provide a close reflection of the income distribution prevailing in the base population cross-section. This section reports comparisons between the distributions of total gross household income that are reported by the WAS and the Family Resources Survey (FRS) for Great Britain in 2011.

The FRS collects information on the incomes and circumstances of households in the UK. It is a standard data source for conducting analyses of the income distribution because of the relatively large sample size and financial detail that it reports. The survey is sponsored by the Department for Work and Pensions (DWP), and is the basis for the *Households Below Average Income* publication and the DWP's *Policy Simulation Model*, amongst other uses.

Similar methods are used to draw the samples for Great Britain reported by the FRS and WAS, based upon the small user's Postcode Address File. Important differences do, however, exist between the two data sources that are likely to affect any associated comparison.

Whereas the FRS is designed to aid DWP in projecting the implications of policy alternatives, the WAS is designed to measure household wealth. As noted in Section 4.1, the WAS sample consequently over-samples high net-worth households, which is not the case for the FRS. This suggests that comparisons between WAS and the FRS will help to identify whether the WAS – due to its sampling frame – is a poor basis for considering measures of poverty, for example. Furthermore, whereas the WAS data drawn for the model are for the calendar year ending December 2011, the FRS reports data for the financial year ending 31 March 2012.

In this section we report comparisons between two key variables, measuring total gross household income, HHINC reported by the FRS and DVTOTGIR reported by the WAS. Although both measures of income are designed to be inclusive, in the sense that they aggregate returns from earnings, investments, pensions, and government benefits, the two surveys issue different questions when gathering the component data, and so some definitional differences are unavoidable. Nevertheless, they do represent a useful basis for comparison.

Table B.1 reports a series of age-group specific distributional measures calculated for the two measures of income noted above. This table indicates a reasonably close correspondence between the distributional statistics reported for both surveys. The statistics indicate that the WAS tends to understate slightly inequality reported by the FRS during peak working years (35-59), and overstate inequality in retirement (60+). Of these two, disparities are more pronounced in retirement. Furthermore, the WAS appears to indicate lower incomes toward the bottom of the distribution – and a higher proportion of the population with incomes below 60% of the median – for all but the 60-64 year age group, relative to the FRS. These observations suggest that the WAS is unlikely to generate artificially low projections for the effects of income support policies on the population in general.

Table B.1: Distribution of Gross Household Income by Age Group and Data Source

	age specific deciles relative to population median											
	lowest	2	3	4	5	6	7	8	9	highest	poor*	Gini
						Family Resources Survey						
16-24	0.1684	0.3186	0.4485	0.5430	0.6485	0.7811	0.9558	1.1425	1.4050	2.2055	0.4109	0.3651
25-34	0.3051	0.5669	0.7420	0.9272	1.1022	1.2885	1.5114	1.8034	2.2487	3.9156	0.1672	0.3662
35-44	0.3086	0.6144	0.8076	1.0041	1.2219	1.4626	1.7457	2.1237	2.7447	5.8990	0.1451	0.4264
45-54	0.2927	0.5670	0.7816	1.0145	1.2604	1.5059	1.8019	2.1971	2.8396	5.3701	0.1633	0.4147
55-59	0.2347	0.4638	0.6659	0.8437	1.0752	1.3291	1.6236	2.0428	2.6783	4.5833	0.2080	0.4213
60-64	0.2452	0.4291	0.5540	0.6794	0.8307	1.0020	1.2033	1.4848	1.9041	3.6504	0.2952	0.4054
65-74	0.3024	0.4247	0.4983	0.5797	0.6724	0.7835	0.9208	1.1393	1.4594	2.6300	0.3730	0.3546
75+	0.2613	0.3535	0.4197	0.4837	0.5465	0.6215	0.7108	0.8360	1.0743	1.9763	0.5258	0.3292
						Wealth and Assets Survey						
16-24	0.1186	0.2819	0.3983	0.4862	0.5739	0.6937	0.7930	0.9386	1.1980	2.1393	0.4781	0.3791
25-34	0.2999	0.5731	0.7542	0.9271	1.0880	1.2784	1.4981	1.7884	2.1844	3.7803	0.1691	0.3587
35-44	0.2858	0.5557	0.7508	0.9364	1.1226	1.3271	1.5678	1.8737	2.3599	5.0100	0.1748	0.4108
45-54	0.2794	0.5422	0.7444	0.9771	1.1994	1.4444	1.7225	2.0821	2.5952	5.1364	0.1789	0.4134
55-59	0.2267	0.4267	0.6050	0.8218	1.0306	1.2778	1.5893	1.9750	2.4606	4.0849	0.2468	0.4108
60-64	0.2528	0.4031	0.5473	0.7215	0.9050	1.1038	1.3182	1.6121	2.1290	4.2925	0.2797	0.4318
65-74	0.2620	0.3757	0.4623	0.5492	0.6575	0.7914	0.9721	1.2183	1.5838	3.0841	0.4003	0.4031
75+	0.2248	0.3210	0.3872	0.4519	0.5197	0.6016	0.7106	0.8734	1.1476	2.2836	0.5513	0.3776

source: Family Resources Survey 2011/12, omitting sample for Northern Ireland, variable HHINC
Wealth and Assets Survey, wave 3 reported during the calendar year 2011, variable DVTOTGIR

Notes: * poor defined with respect to poverty line set to 60% of population median income
population median equals £528 per week for FRS, and £32076.40 per year for the WAS

Appendix C: Distribution of Northern Irish Population

Table C.1: Distribution of Northern Irish population by reference adult age, relationship status, and gross income quintile reported for benefit units in Great Britain

age	income quintile thresholds for Great Britain					
band	lowest	2	3	4	highest	total
			singles			
16 to 24	0.0185	0.0125	0.0171	0.0187	0.0140	0.0807
25 to 34	0.0214	0.0191	0.0130	0.0136	0.0120	0.0790
35 to 44	0.0124	0.0162	0.0087	0.0110	0.0046	0.0529
45 to 54	0.0109	0.0104	0.0058	0.0053	0.0064	0.0388
55 to 59	0.0035	0.0035	0.0029	0.0025	0.0013	0.0137
60 to 64	0.0023	0.0031	0.0015	0.0014	0.0008	0.0091
65 to 74	0.0103	0.0039	0.0030	0.0040	0.0026	0.0238
75 +	0.0069	0.0085	0.0052	0.0065	0.0040	0.0311
			couples			
16 to 24	0.0081	0.0047	0.0041	0.0024	0.0015	0.0208
25 to 34	0.0269	0.0294	0.0201	0.0273	0.0127	0.1163
35 to 44	0.0532	0.0552	0.0424	0.0266	0.0210	0.1985
45 to 54	0.0503	0.0319	0.0254	0.0223	0.0183	0.1483
55 to 59	0.0219	0.0150	0.0083	0.0038	0.0039	0.0529
60 to 64	0.0134	0.0122	0.0065	0.0065	0.0043	0.0429
65 to 74	0.0208	0.0124	0.0097	0.0091	0.0062	0.0582
75 +	0.0108	0.0061	0.0063	0.0047	0.0052	0.0330
total	0.2916	0.2440	0.1799	0.1657	0.1188	1.0000

Source: Family Resources Survey
Quintile groups defined within age and relationship specific subgroups

Appendix D: Regression Statistics to Evaluate Indirect Taxes and Duties

Table D.1: Tobit regression results for the proportion of expenditure on selected consumption baskets – pooled data 2001-2010

	VAT full rate	VAT red. rate	Non VAT	Alcohol	Tobacco	Fuels	SR Insurance	HR Insurance
Age	-4.433E-04	1.841E-04	4.874E-04	9.545E-04	1.245E-03	7.984E-04	6.710E-05	-2.254E-04
	(2.95E-06)	(9.00E-07)	(2.91E-06)	(8.96E-07)	(7.61E-07)	(7.89E-07)	(1.59E-07)	(5.79E-07)
Age squared	-6.930E-06	6.650E-06	-2.390E-06	-1.250E-05	-1.410E-05	-1.140E-05	-6.290E-07	3.590E-06
	(3.00E-08)	(9.17E-09)	(2.97E-08)	(9.14E-09)	(7.76E-09)	(8.04E-09)	(1.62E-09)	(5.90E-09)
Disposable income (£)	-8.730E-05	1.900E-06	7.480E-05	6.180E-06	-1.300E-05	1.330E-05	2.390E-06	1.800E-05
	(4.91E-08)	(1.50E-08)	(4.85E-08)	(1.49E-08)	(1.27E-08)	(1.31E-08)	(2.65E-09)	(9.64E-09)
Consumption (£)	1.613E-04	-5.000E-05	-1.118E-04	-9.910E-06	-1.520E-05	-1.870E-05	-5.840E-07	-2.060E-05
	(4.16E-08)	(1.27E-08)	(4.10E-08)	(1.26E-08)	(1.07E-08)	(1.11E-08)	(2.25E-09)	(8.16E-09)
Single no children (ref)	-	-	-	-	-	-	-	-
Couples no children	2.391E-02	-2.757E-03	-1.976E-02	1.866E-03	2.515E-03	1.263E-02	-3.058E-04	4.028E-03
	(2.50E-05)	(7.65E-06)	(2.47E-05)	(7.62E-06)	(6.48E-06)	(6.71E-06)	(1.35E-06)	(4.92E-06)
Single & 1 Child	-2.151E-02	1.216E-02	9.664E-03	-1.773E-02	3.241E-03	-1.114E-02	-4.878E-04	-5.549E-03
	(5.60E-05)	(1.71E-05)	(5.53E-05)	(1.70E-05)	(1.45E-05)	(1.50E-05)	(3.03E-06)	(1.10E-05)
Single & 2 Children	-3.016E-02	1.519E-02	1.649E-02	-2.269E-02	1.420E-03	-1.161E-02	-9.376E-04	-6.532E-03
	(7.28E-05)	(2.22E-05)	(7.19E-05)	(2.21E-05)	(1.88E-05)	(1.95E-05)	(3.94E-06)	(1.43E-05)
Single & 3 Children	-2.939E-02	2.025E-02	7.182E-03	-2.613E-02	6.175E-03	-1.672E-02	-1.509E-03	-1.015E-02
	(1.26E-04)	(3.84E-05)	(1.24E-04)	(3.83E-05)	(3.25E-05)	(3.37E-05)	(6.81E-06)	(2.47E-05)
Single & 4+ Children	-3.118E-02	2.150E-02	7.717E-03	-3.026E-02	5.826E-04	-2.434E-02	-2.217E-03	-1.766E-02
	(2.14E-04)	(6.55E-05)	(2.12E-04)	(6.52E-05)	(5.54E-05)	(5.74E-05)	(1.16E-05)	(4.21E-05)
Couple & 1 Child	-2.821E-02	2.550E-06*	2.868E-02	4.575E-03	-3.458E-03	5.842E-03	3.849E-04	3.818E-03
	(6.71E-05)	(2.05E-05)	(6.62E-05)	(2.04E-05)	(1.73E-05)	(1.80E-05)	(3.63E-06)	(1.32E-05)
Couple & 2 Children	-3.127E-02	-1.905E-03	3.159E-02	5.876E-03	-3.151E-03	4.617E-03	9.687E-04	4.645E-03
	(8.17E-05)	(2.49E-05)	(8.06E-05)	(2.48E-05)	(2.11E-05)	(2.19E-05)	(4.42E-06)	(1.60E-05)
Couple & 3 Children	-3.564E-02	-5.193E-03	4.230E-02	4.890E-03	-5.724E-03	1.113E-02	5.660E-04	8.660E-03
	(1.42E-04)	(4.35E-05)	(1.41E-04)	(4.33E-05)	(3.68E-05)	(3.81E-05)	(7.70E-06)	(2.80E-05)
Couple & 4+ Children	-2.894E-02	-4.099E-03	3.384E-02	5.395E-03	2.022E-03	1.568E-02	8.466E-04	1.335E-02
	(2.45E-04)	(7.49E-05)	(2.42E-04)	(7.47E-05)	(6.34E-05)	(6.57E-05)	(1.33E-05)	(4.82E-05)
Non-graduate (ref)	-	-	-	-	-	-	-	-
Graduate	-3.291E-02	2.573E-03	3.227E-02	-4.414E-03	-8.461E-03	-1.877E-03	-1.093E-04	7.715E-04
	(2.57E-05)	(7.86E-06)	(2.54E-05)	(7.83E-06)	(6.65E-06)	(6.89E-06)	(1.39E-06)	(5.05E-06)
Constant	4.688E-01	4.296E-02	4.887E-01	2.664E-02	5.310E-03	2.854E-02	3.733E-04	3.477E-02
	(5.86E-05)	(1.79E-05)	(5.79E-05)	(1.78E-05)	(1.51E-05)	(1.57E-05)	(3.17E-06)	(1.15E-05)
R-squared	0.07239	0.17096	0.03450	0.02828	0.04536	0.04660	0.00594	0.03773

Source: authors' calculations using 115,827 observations drawn from the Living Cost and Food Survey, pooled over the years 2001 to 2010
Notes: standard errors reported in parentheses; all coefficients significant at 99% confidence interval, except parameter indicated by *
when calculating indirect taxes, model imposes top-codes for age (80), income (£50,000 p.a.), and annual consumption (£50,000).

Appendix E: Distribution of Recent Immigrants to the UK

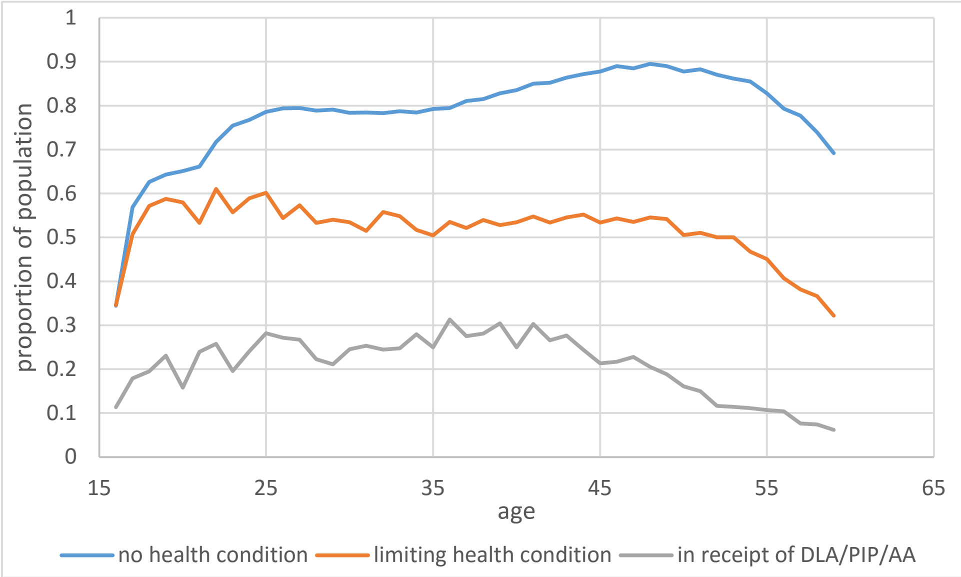
Table E.1: Distribution of immigrant benefit units to the UK, relative to the gross equivalised incomes of the wider UK population, by age group and survey year

age band	sample size	gross equivalised income quintile: lowest	2	3	4	highest
		Family Resources Survey 2012/13				
18-24	111	0.177	0.156	0.199	0.220	0.248
25-34	350	0.226	0.253	0.183	0.162	0.173
35-44	151	0.318	0.214	0.149	0.140	0.173
45-64	79	0.156	0.402	0.180	0.080	0.183
65+	14	0.533	0.000	0.047	0.168	0.252
		Family Resources Survey 2011/12				
18-24	158	0.241	0.312	0.122	0.135	0.176
25-34	355	0.305	0.223	0.189	0.147	0.135
35-44	167	0.267	0.328	0.149	0.127	0.129
45-64	74	0.451	0.188	0.144	0.082	0.136
65+	14	0.540	0.201	0.166	0.000	0.093
		Family Resources Survey 2010/11				
18-24	214	0.277	0.199	0.164	0.160	0.189
25-34	422	0.279	0.268	0.185	0.165	0.103
35-44	158	0.287	0.293	0.131	0.088	0.185
45-64	88	0.225	0.147	0.277	0.173	0.178
65+	3	0.809	0.191	0.000	0.000	0.000
		Family Resources Survey 2009/10				
18-24	139	0.180	0.201	0.168	0.194	0.252
25-34	252	0.304	0.199	0.219	0.160	0.113
35-44	96	0.326	0.165	0.163	0.140	0.205
45-64	57	0.244	0.387	0.069	0.158	0.143
65+	4	0.628	0.186	0.186	0.000	0.000

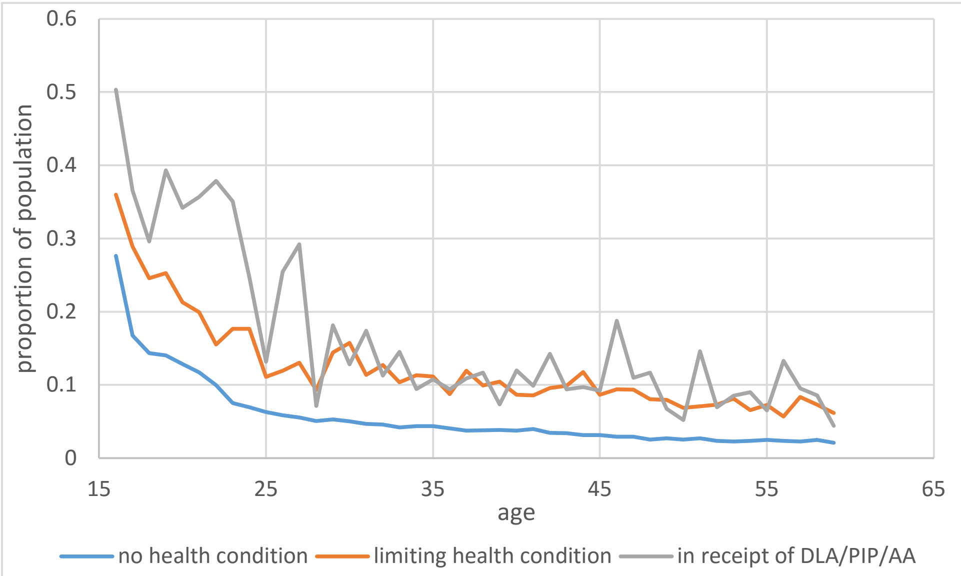
Source: Authors' calculations on Family Resources Survey data
Notes: Total benefit unit gross income reported by variable buinc
Income equivalised using the OECD revised scale
Income quintiles evaluated for population omitting recent immigrants
Recent immigrants defined as benefit units in which all adult members reported as moving to the UK and where the most recent immigrant arrived within the 5 years preceding the month of survey

Appendix F: Simulating Disability

Figure F.1: Average rates of labour force participation and unemployment by age and disability status reported for women between 1998 and 2014



Panel A: Rates of labour force participation



Panel B: Rates of unemployment

Source: Author's calculations using data from the April to June wave of the quarterly Labour Force Survey between 1998 and 2014.

Notes: 'limiting health condition' defined as a condition that is reported to limit activity and to have persisted for at least 12 months.

Table F.1: Full regression statistics for log gross hourly pay

	men		women	
	coef	std error	coef	std error
age band				
20-24	0.3440	0.0080	0.2496	0.0076
25-29	0.5842	0.0084	0.4933	0.0075
30-34	0.7552	0.0085	0.6254	0.0078
35-39	0.8418	0.0085	0.6873	0.0077
40-44	0.8689	0.0084	0.6696	0.0075
45-49	0.8707	0.0084	0.6344	0.0075
50-54	0.8390	0.0088	0.5959	0.0077
55-59	0.7786	0.0092	0.5350	0.0081
moderate disability				
16-19	-0.0974	0.0351	-0.0430	0.0273
20-24	-0.0739	0.0188	-0.0808	0.0206
25-29	-0.1219	0.0187	-0.1151	0.0166
30-34	-0.2165	0.0176	-0.0810	0.0158
35-39	-0.2064	0.0166	-0.1044	0.0148
40-44	-0.2203	0.0156	-0.1313	0.0141
45-49	-0.2111	0.0149	-0.1256	0.0126
50-54	-0.2071	0.0146	-0.1199	0.0121
55-59	-0.1875	0.0148	-0.1186	0.0118
significant disability				
16-19	-0.0836	0.0724	-0.0214	0.0895
20-24	-0.3204	0.0663	-0.1386	0.0641
25-29	-0.3661	0.0530	-0.1998	0.0459
30-34	-0.3817	0.0771	-0.1716	0.0373
35-39	-0.3600	0.0634	-0.1420	0.0360
40-44	-0.4232	0.0539	-0.1649	0.0299
45-49	-0.3260	0.0608	-0.1078	0.0352
50-54	-0.3560	0.0654	-0.1123	0.0394
55-59	-0.3326	0.0707	-0.2226	0.0577
married	0.1295	0.0038	0.0612	0.0030
dependent children by age band				
2 and under	0.0080	0.0047	0.0108	0.0055
3 to 4	0.0090	0.0042	0.0286	0.0044
5 to 9	0.0034	0.0034	-0.0504	0.0032
10 to 15	-0.0070	0.0028	-0.0741	0.0025
16 to 19	0.0170	0.0048	-0.0432	0.0040

Notes: Table continued on next page

Table F.1: Full regression statistics for log gross hourly pay – time dummies (cont.)

		men		women	
year	quarter	coef	std error	coef	std error
2004	2	0.0234	0.0125	0.0242	0.0112
2004	3	0.0117	0.0118	0.0200	0.0113
2004	4	0.0145	0.0121	0.0370	0.0110
2005	1	0.0405	0.0124	0.0714	0.0113
2005	2	0.0699	0.0119	0.0624	0.0113
2005	3	0.0744	0.0116	0.0658	0.0114
2005	4	0.0690	0.0122	0.0947	0.0115
2006	1	0.1062	0.0121	0.1036	0.0113
2006	2	0.0803	0.0124	0.1107	0.0112
2006	3	0.0852	0.0123	0.1144	0.0114
2006	4	0.0989	0.0121	0.1120	0.0115
2007	1	0.1131	0.0127	0.1265	0.0114
2007	2	0.1172	0.0123	0.1389	0.0115
2007	3	0.1199	0.0123	0.1486	0.0112
2007	4	0.1264	0.0122	0.1794	0.0117
2008	1	0.1456	0.0123	0.1655	0.0114
2008	2	0.1461	0.0124	0.1708	0.0112
2008	3	0.1608	0.0122	0.1833	0.0116
2008	4	0.1700	0.0126	0.1841	0.0115
2009	1	0.1632	0.0136	0.1874	0.0118
2009	2	0.1846	0.0130	0.2053	0.0119
2009	3	0.1598	0.0123	0.2061	0.0114
2009	4	0.1932	0.0128	0.2155	0.0118
2010	1	0.1784	0.0134	0.2151	0.0122
2010	2	0.1827	0.0128	0.2319	0.0117
2010	3	0.1768	0.0133	0.2246	0.0118
2010	4	0.1880	0.0132	0.2300	0.0119
2011	1	0.1895	0.0130	0.2342	0.0119
2011	2	0.1878	0.0136	0.2220	0.0123
2011	3	0.1906	0.0133	0.2326	0.0125
2011	4	0.1940	0.0127	0.2507	0.0120
2012	1	0.1986	0.0131	0.2439	0.0120
2012	2	0.2059	0.0135	0.2617	0.0123
2012	3	0.2141	0.0136	0.2669	0.0123
2012	4	0.2387	0.0135	0.2564	0.0124
2013	1	0.2181	0.0131	0.2755	0.0122
2013	2	0.2387	0.0135	0.2714	0.0128
2013	3	0.2146	0.0141	0.2619	0.0124
2013	4	0.1987	0.0134	0.2568	0.0122
2014	1	0.2317	0.0131	0.2820	0.0121
2014	2	0.2141	0.0138	0.2772	0.0120
2014	3	0.2328	0.0134	0.2835	0.0121
2014	4	0.2256	0.0133	0.2849	0.0123

Notes: Table continued on next page

Table F.1: Full regression statistics for log gross hourly pay – time dummies (cont.)

	men		women	
	coef	std error	coef	std error
ethnicity				
mixed	-0.0141	0.0186	0.0182	0.0158
Asian	-0.2079	0.0084	-0.0787	0.0079
Black	-0.2812	0.0116	-0.1204	0.0099
Chinese	-0.0972	0.0297	-0.0320	0.0267
other	-0.2908	0.0157	-0.1502	0.0153
UK region				
Rest of North East	0.0083	0.0121	-0.0346	0.0110
Greater Manchester	0.0631	0.0108	0.0440	0.0100
Merseyside	0.0226	0.0130	0.0052	0.0119
Rest of North West	0.0607	0.0108	0.0094	0.0098
South Yorkshire	0.0008	0.0123	-0.0267	0.0112
West Yorkshire	0.0425	0.0109	0.0271	0.0100
Yorkshire & Humber.	0.0335	0.0118	-0.0310	0.0107
East Midlands	0.0620	0.0101	0.0159	0.0093
West Midlands Metro	0.0502	0.0116	0.0367	0.0105
Rest of West Mid.	0.0674	0.0110	0.0172	0.0100
East of England	0.1842	0.0101	0.1014	0.0093
Inner London	0.4453	0.0136	0.4253	0.0123
Outer London	0.3218	0.0111	0.2859	0.0102
South East	0.2317	0.0097	0.1292	0.0090
South West	0.0893	0.0102	0.0214	0.0093
Wales	0.0040	0.0109	0.0089	0.0100
Strathclyde	0.0599	0.0115	0.0591	0.0102
Rest of Scotland	0.0947	0.0106	0.0653	0.0098
Northern Ireland	-0.0355	0.0118	0.0168	0.0107
constant	1.3591	0.0137	1.4165	0.0128
observations	136,366		151,742	
R-squared	0.2537		0.1885	

Source: authors' calculations on quarterly Labour Force Survey data, Jan-Mar quarter 2004 to Oct-Dec quarter 2014

Notes: dependent variable is log hourly wage rate. "significant disability" identified as receipt of Disability Living Allowance, Personal Independence Payment, or Attendance Allowance. "moderate work-limiting disability" identies individuals reporting a long-lasting health condition that limits normal activities, and who are not identified as "significantly disabled". Statistics calculated using the "regress" command in Stata, with the "robust" option. 'region': Government Office Regions, omitting "Tyne and Wear"; 'ethnicity': omits "white"

Appendix G: Validating Projected Trends

Executive Summary

This appendix checks that the final LINDA model base scenario projects trends that are plausible and free from unexpected dynamics.

This is found to be the case across all key variables. In most cases, projected distributions are not very different from the distributions observed in the reference year survey data. This is an important result given the long horizon over which the model projects forward.

The baseline scenario exhibits systematic trends in some variables, although these are either what is expected or within what we consider plausible. These trends are:

- increased life expectancy;
- a delay in couple formation later in life and consequently a decline in marriage/cohabitation rates at a point in time;
- a modest reduction in employment, in the form of some shift into part-time employment and non-employment. This may be plausible, if efforts to increase flexible working are sustained in the future;
- positive real earnings and income growth.

The detailed shape of the income distribution in later years is very similar to that in the reference year. This is an important observation, given its importance for poverty analysis.

In this appendix, we present the apparent drivers of the trends in poverty rates for selected family types. Specifically, we first display the trends in poverty rates by family type, and we then analyse the driving factors for groups where we find a marked trend.

The document also presents the trends in relative BHC poverty rates by family type, with view of identifying any trends that may appear as too strong. We find a notable increase in the poverty rate of couples without children and, to a lesser extent couples with children. These trends appear to be driven by a decline in labour supply. This is, in turn, driven more by declining earnings potential among the lower tail of the potential earnings distribution, rather than by abundant accumulation of savings.

Aims and objectives

This appendix presents and discusses a descriptive analysis of the current LINDA model base scenario provided by NIESR to the JRF. The aims are twofold:

1. to validate that projected trends are plausible and free from unexpected dynamics; and,
2. to illustrate the main apparent drivers of any significant trends in poverty rates affecting a key population subgroups.

Methodology

The validation of projected trends is inherently subjective, in that an objective benchmark free from uncertainty does not exist. As such, our default validation approach consists of checking that the distribution of key variables does not exhibit any systematic evolution, or indeed sudden change, over time.

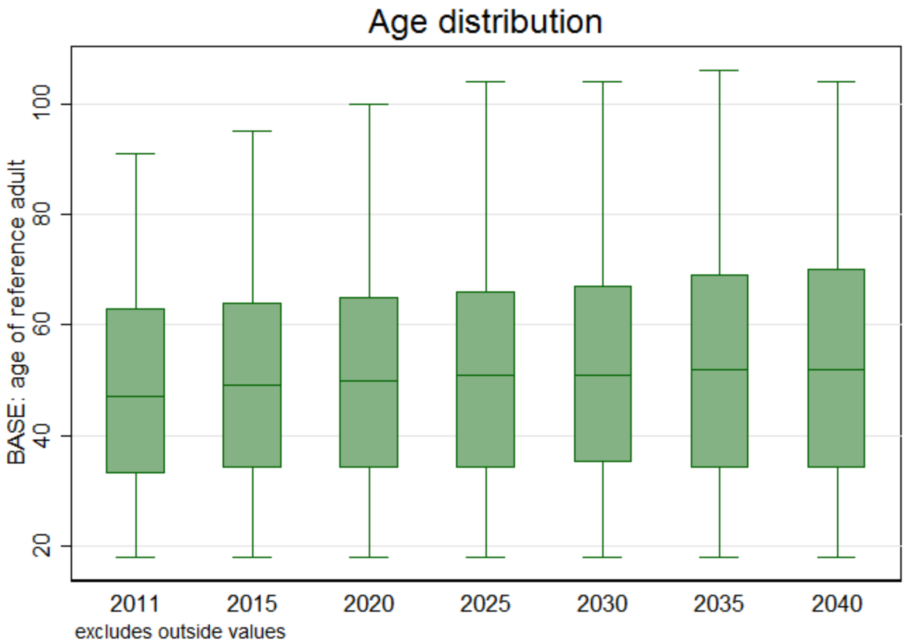
Where we find strong trends in poverty rates for key family types, we carry out an initial exploration of the possible underlying drivers. We do this by presenting trends in key income variables over the same time period, to establish temporal coincidences and correlations which can suggest the origin of the observed trends in poverty rates.

Appendix G.1 – Validating projected trends

This section presents a brief headline commentary on the trends in the distribution of the main variables of interest.

Appendix G.1.1. Age

The box plots below show the median, upper and lower quartile and upper and lower adjacent values[46] of the age distribution over selected years between 2011 and 2040.



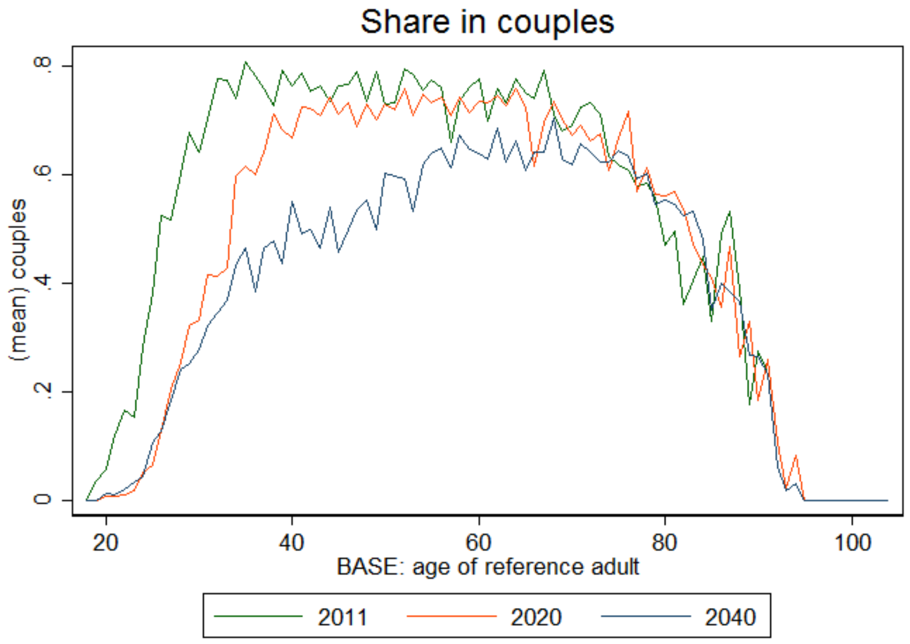
[46] The upper adjacent value (UAV) is the largest observation that is less than or equal to the upper inner fence (UIF), which is the third quartile plus 1.5*IQR. The lower adjacent value (LAV) is the smallest observation that is greater than or equal to the lower inner fence (LIF), which is the first quartile minus 1.5*IQR.

- Distribution is broadly stable over time.
- Higher tail is increasing over time, reflecting increased life expectancy.

This is what we would expect, and is in line with ONS principal projection for the UK population, upon which the simulated mortality rates are based.

Appendix G.1.2 Marriage/cohabitation rates

The graph below displays the share of adults in couples by age, in 2011, 2020 and 2040.



- Marriage/cohabitation rates beyond the age of 70 are broadly stable over time.
- Couple formation however appears to be postponed to later in life as we progress into the future.
- As a consequence, the share of the cross-section which is married/cohabiting at any point in time declines over time.

This trend toward later coupling reflects contemporary trends in the associated data considered for calibration, including ONS projections to 2032 (see Section 4.7.1).

Appendix G.1.3 Employment rates

- The above charts shows there are no major or sudden changes in employment rates over time.
- The tables below allow us to take a closer look at the trends.
- There appears to be a shift from full-time to part-time work for singles. This is plausible, if efforts to increase flexible working are sustained.
- Full-time employment rates among couples declines less than for singles, but is not offset by an increase in part-time work.

Full-time employment rates, singles

	b_age_grp						
year	Age 18-21	Age 22-30	Age 31-40	Age 41-50	Age 51-60	Age 61-70	Age 70+
2011	.3703433	.5361485	.5712165	.6628745	.5216064	.1747229	.0279208
2016	.3640951	.4801555	.5061977	.5921268	.4968184	.1514011	.0276849
2020	.3715991	.4668486	.4693419	.5582602	.4656318	.1458572	.0281644
2030	.278334	.4770682	.408439	.5686736	.457023	.0962128	.0133224
2040	.3753494	.4400732	.4563485	.5286522	.4095642	.1122772	.0093499

Part-time employment rates, singles

	b_age_grp						
year	Age 18-21	Age 22-30	Age 31-40	Age 41-50	Age 51-60	Age 61-70	Age 70+
2011	.0847108	.1157433	.1762226	.1007566	.1375858	.0726054	.012813
2016	.1239227	.1119255	.2210407	.1456281	.1239124	.075405	.0075109
2020	.139819	.1197191	.2088873	.1535166	.1264919	.0791119	.0123287
2030	.1676907	.1222425	.2097338	.1368301	.1529133	.073583	.0073129
2040	.0876337	.179323	.208201	.1528643	.1528122	.11991	.0133034

Full-time employment rates, couples

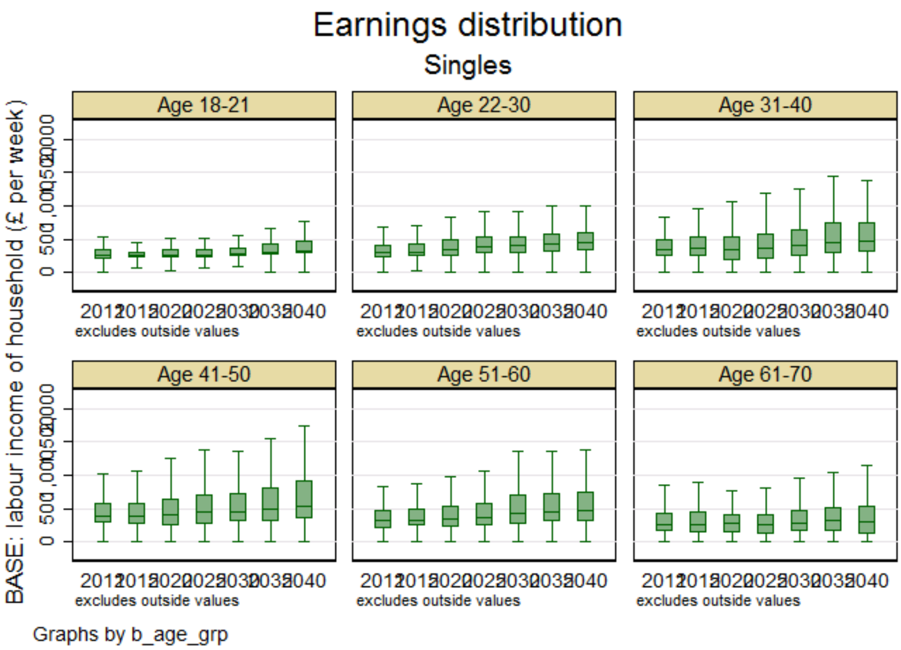
year	Age 18-21	Age 22-30	Age 31-40	Age 41-50	Age 51-60	Age 61-70	Age 70+
				b_age_grp			
2011	.4704759	.7861897	.7507473	.7694786	.5368957	.1352602	.0171051
2016	.5064	.6285196	.7136052	.7377921	.5070797	.1703686	.0393579
2020	.75	.6259653	.68902	.7321233	.4684207	.1883548	.0467953
2030	1	.6959505	.6763742	.676488	.461531	.140954	.0363086
2040	.6536428	.6695764	.7546012	.6915027	.4515852	.1765935	.0345486

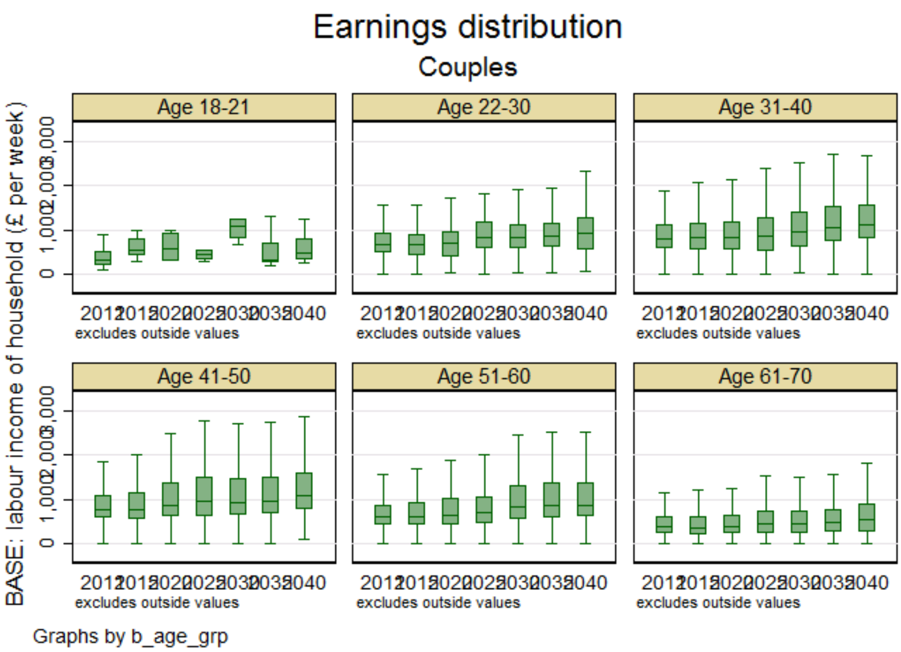
Part-time employment rates, couples

year	Age 18-21	Age 22-30	Age 31-40	Age 41-50	Age 51-60	Age 61-70	Age 70+
				b_age_grp			
2011	.0987364	.0981966	.1486388	.0925715	.1205413	.0581453	.0074529
2016	.1602667	.1440898	.135827	.1241484	.1268761	.0883927	.0147812
2020	0	.1289616	.1341625	.1159772	.1320837	.0987517	.0223198
2030	0	.0932504	.1276875	.1022811	.159303	.0769462	.0191227
2040	.0606429	.1919918	.1072319	.1154682	.1443807	.1045858	.0195555

Appendix G.1.4 Earnings distribution

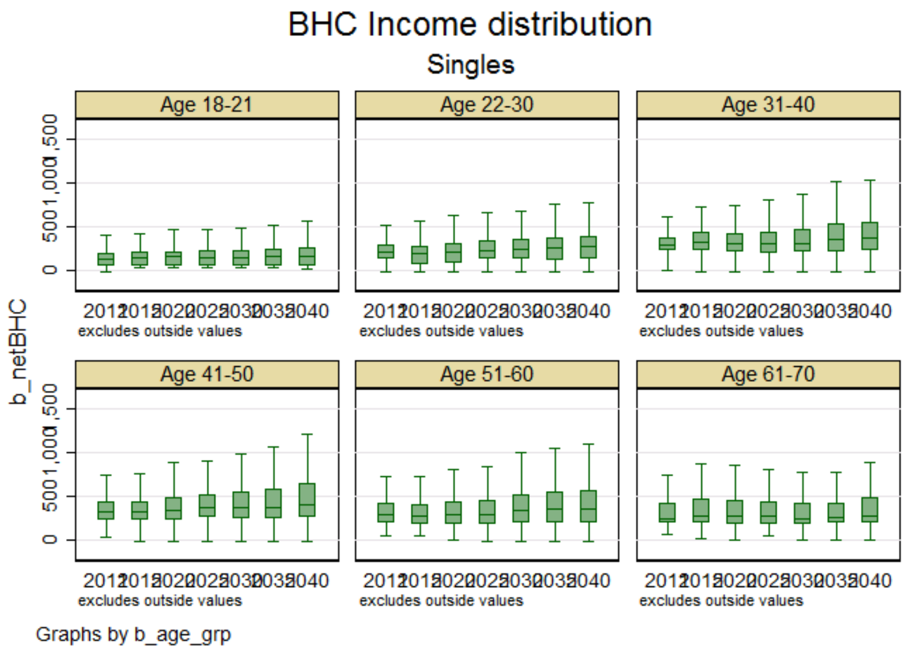
- The shape of the distribution of earnings among those in work, analysed by age group, is maintained over time.
- There is an increasing trend in the distribution, reflecting the assumption of positive real earnings growth.

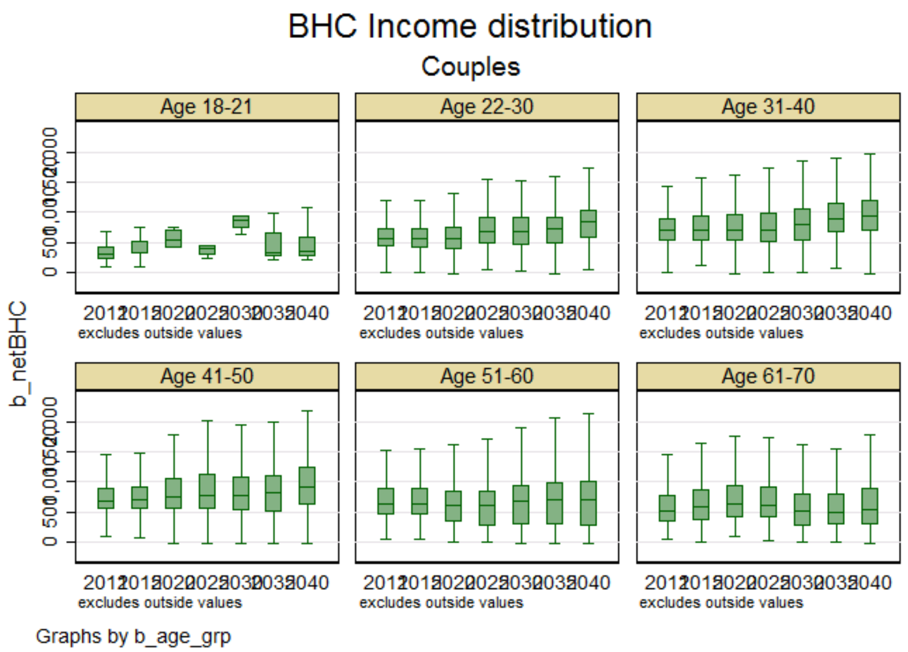


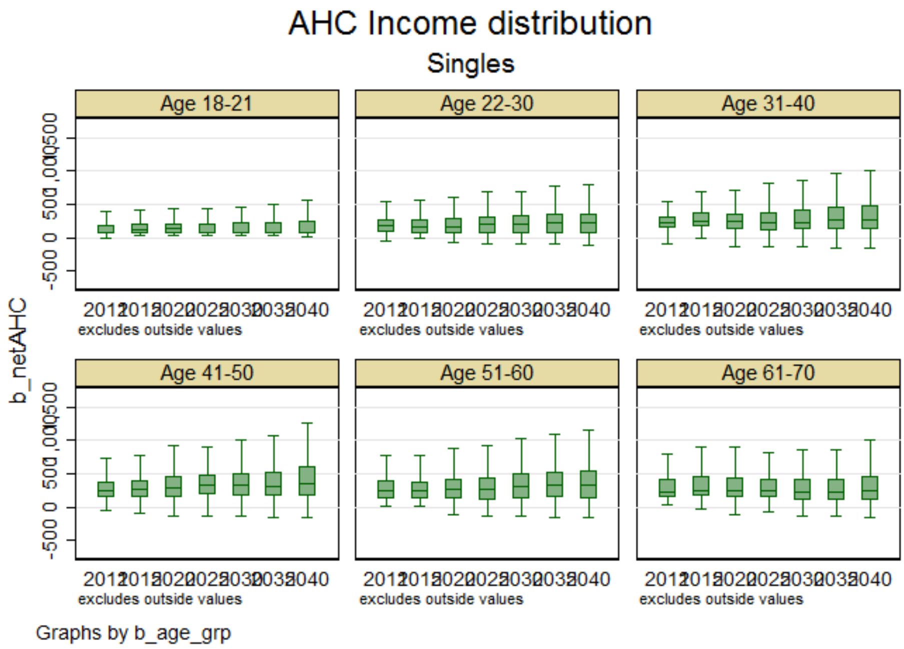


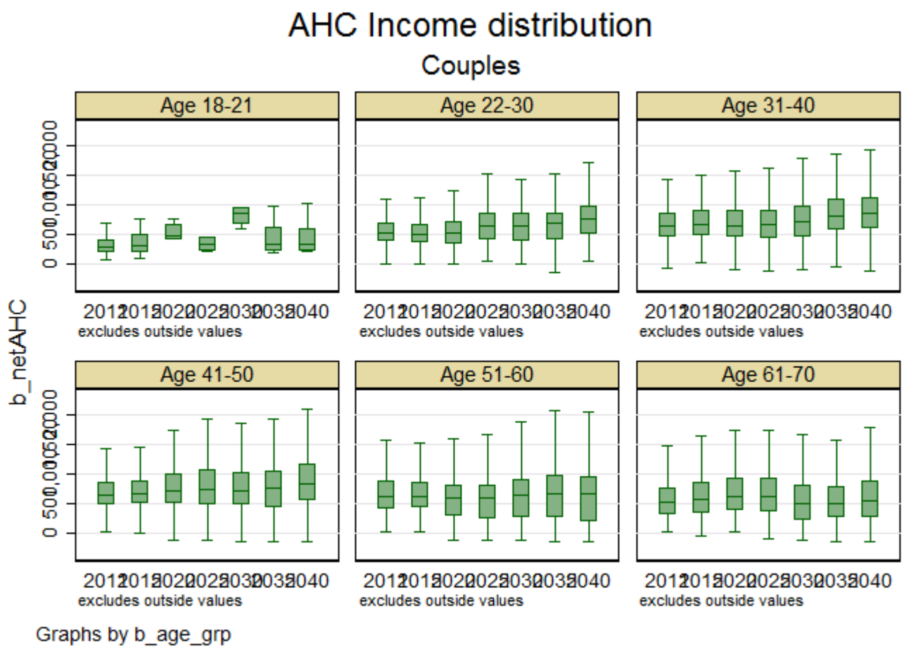
Appendix G.1.5 Net income distribution

- The equivalised net income distribution, by age group, is maintained over time.
- There is an increasing trend in the distribution, reflecting the assumption of positive real earnings growth.
- The above holds for incomes defined both before and after housing costs.



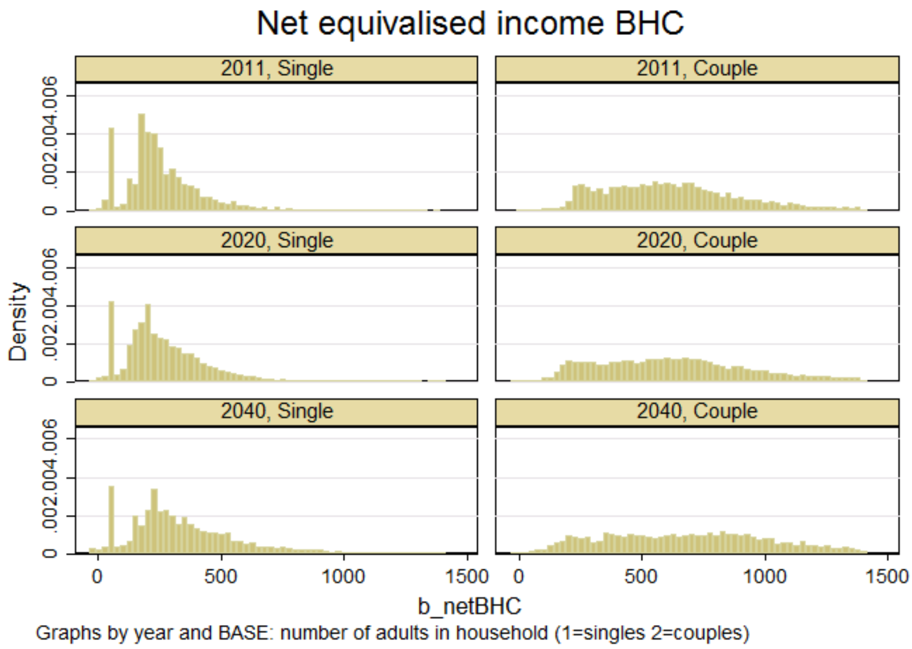


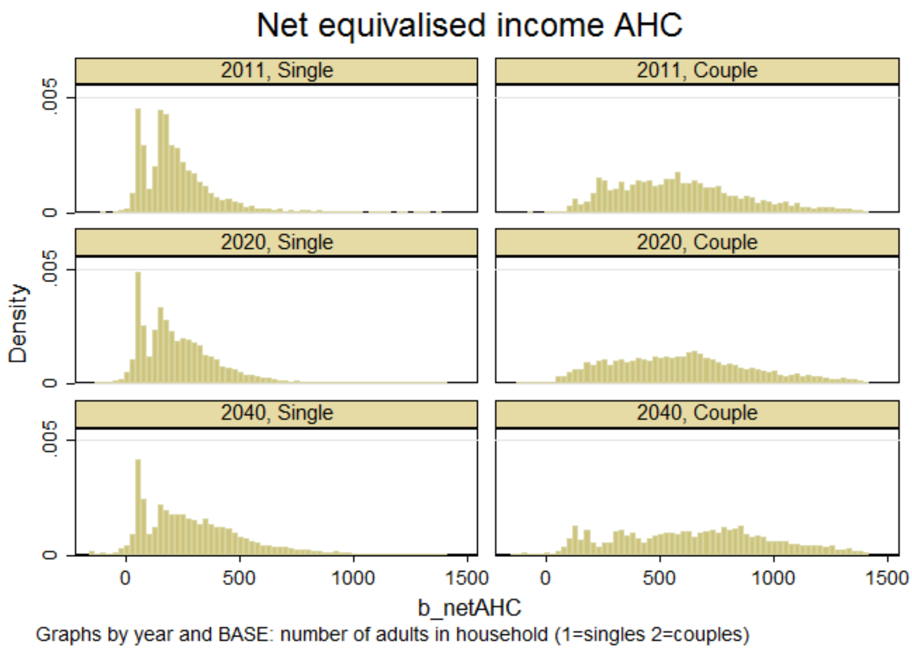




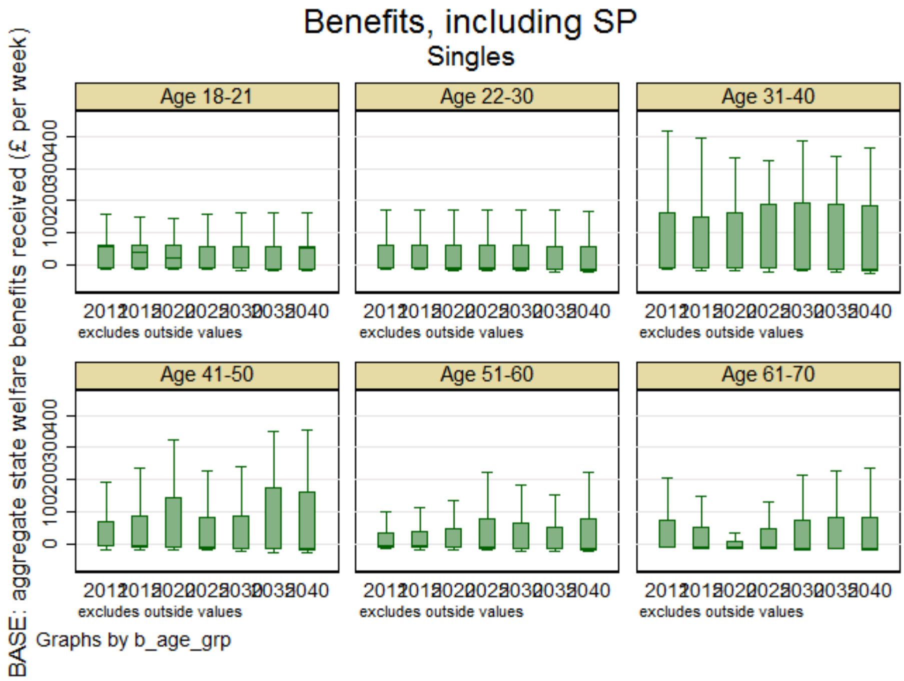
Given the importance of the shape of the income distribution for poverty analysis, we present a detailed histogram of the income distribution in the reference year (2011) and in selected later years.

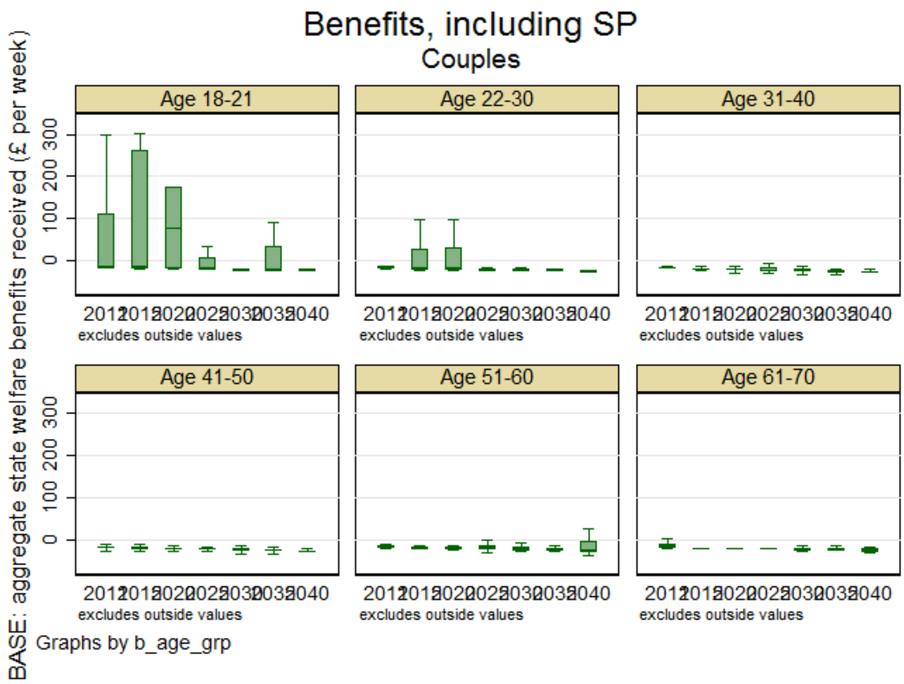
- The shape in later years is very similar to that in the reference year.
- The above holds for incomes defined both before and after housing costs.


Net equivalised income BHC
2011, Single
2011, Couple
2020, Single
2020, Couple
2040, Single
2040, Couple
Density
0
.002
.004
.006
0
500
1000
1500
b_netBHC
Graphs by year and BASE: number of adults in household (1=singles 2=couples)


Net equivalised income AHC
2011, Single
2011, Couple
2020, Single
2020, Couple
2040, Single
2040, Couple
Density
0
.005
0
500
1000
1500
b_netAHC
Graphs by year and BASE: number of adults in household (1=singles 2=couples)

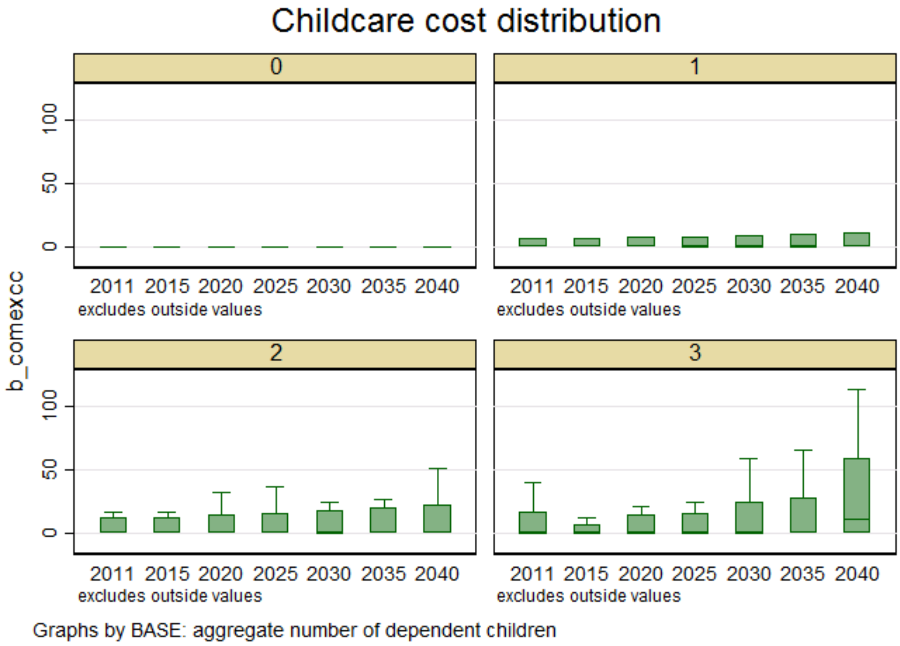
Appendix G.1.6 Benefit incomes (including State Pension)



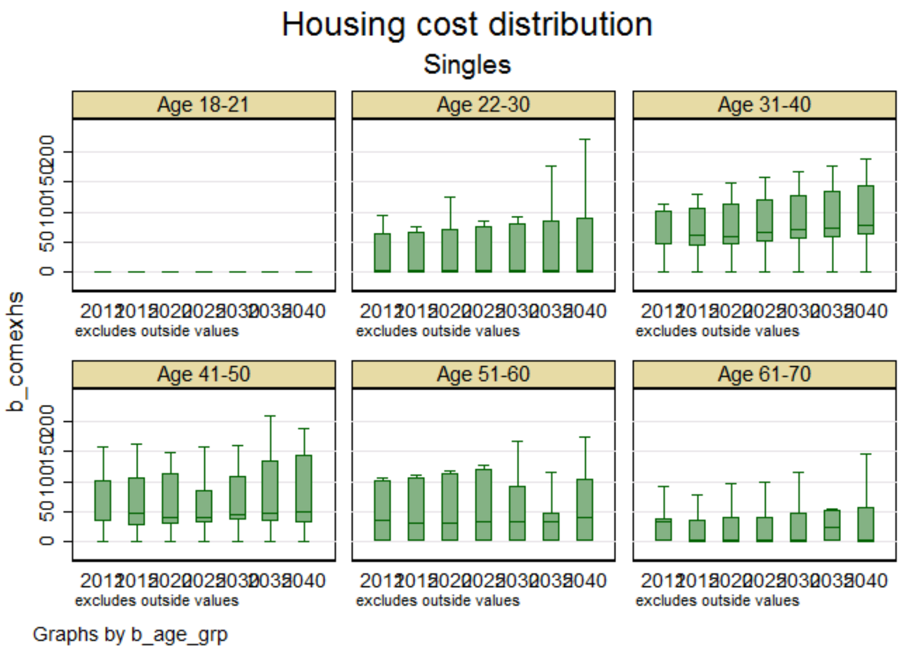


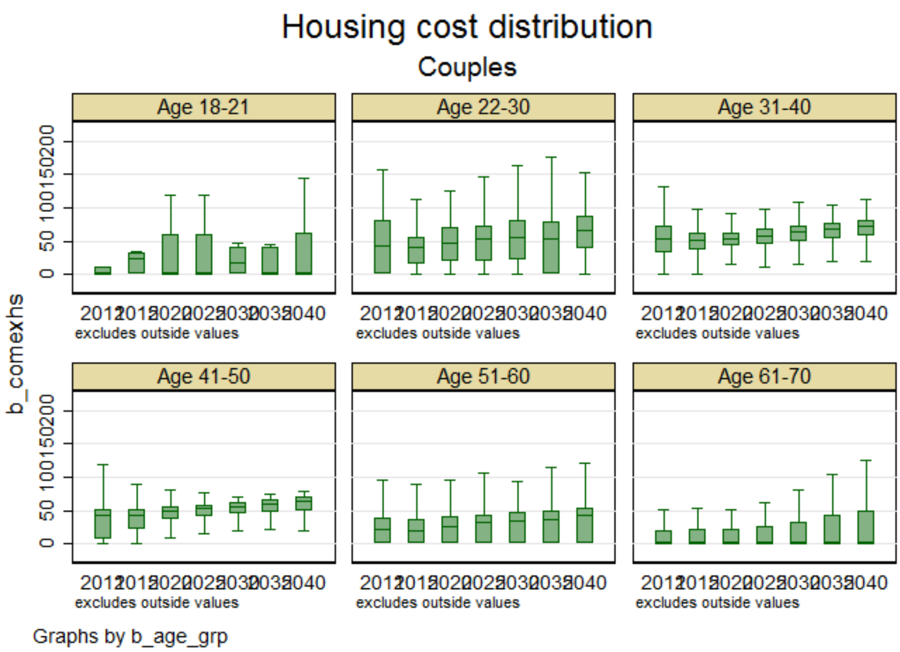
Appendix G.1.7 Committed expenditure

- The distribution of committed expenditure on childcare, analysed by number of children, is maintained over time.



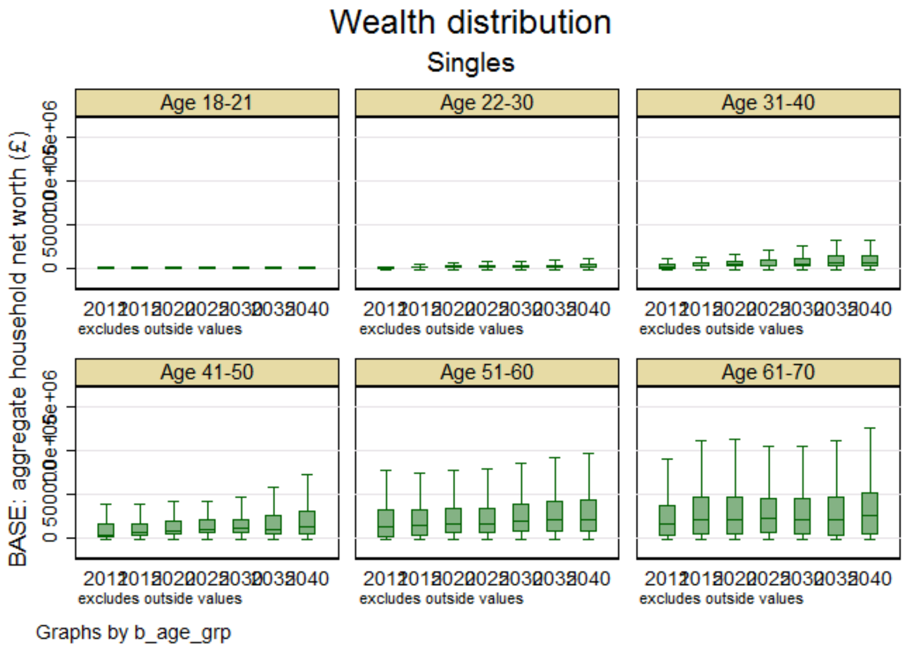
- The distribution of housing costs, analysed by age group, is broadly maintained over time.

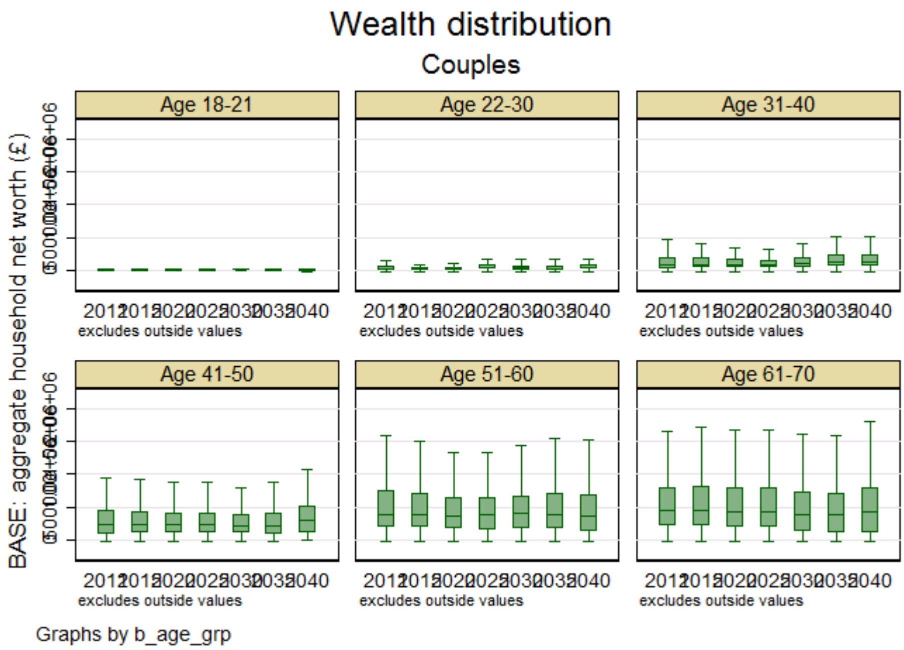




Appendix G.1.8 Wealth

The distribution in total household wealth, considered by age group and single/couples, is stable over time.

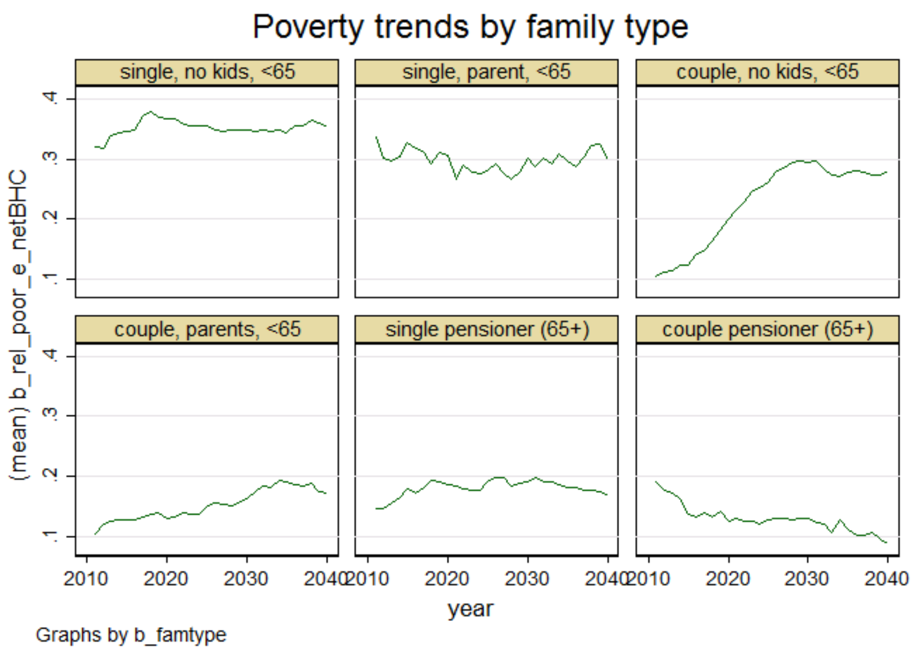




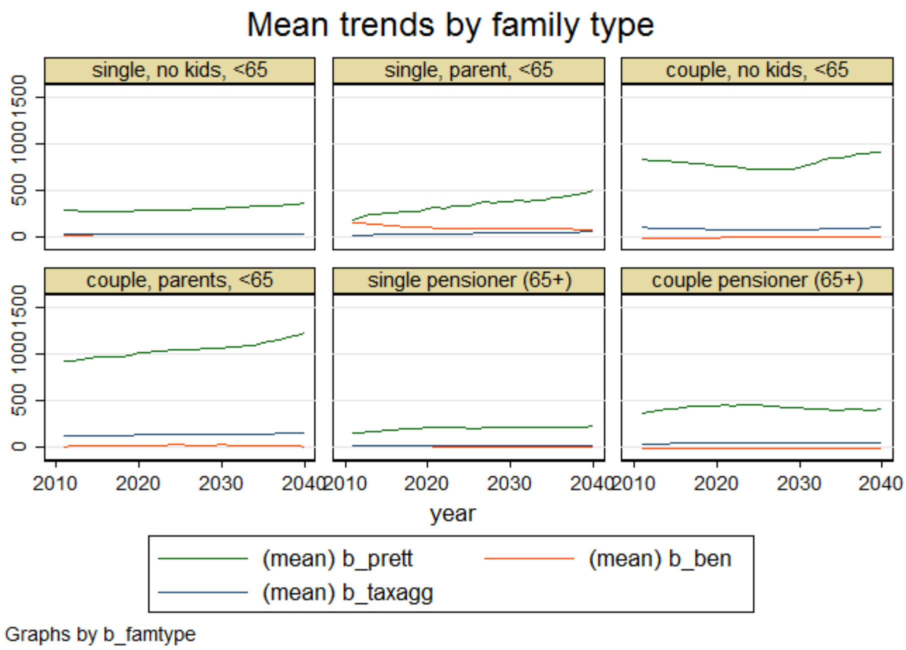
Appendix G.2 – Drivers of trends in poverty rates

In this secton we present the apparent drivers of the trends in poverty rates for selected family types. Specifically, we first display the trends in poverty rates by family type, and we then analyse the driving factors for groups where we find a marked trend.

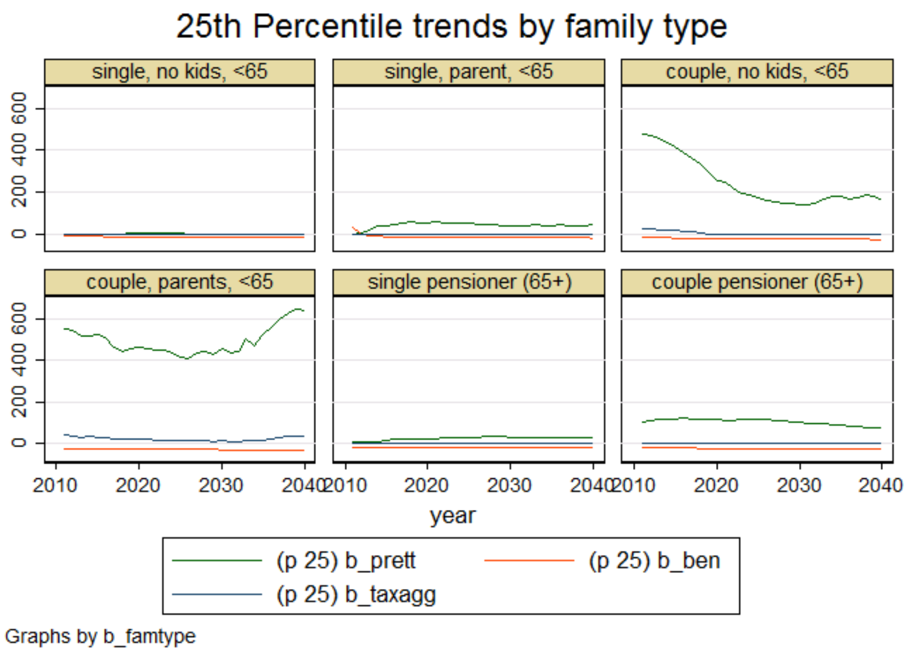
The graph below displays the relative poverty rate by family type over time. There is a notable increase in the poverty rate of couples without children and, to a lesser extent couples with children.



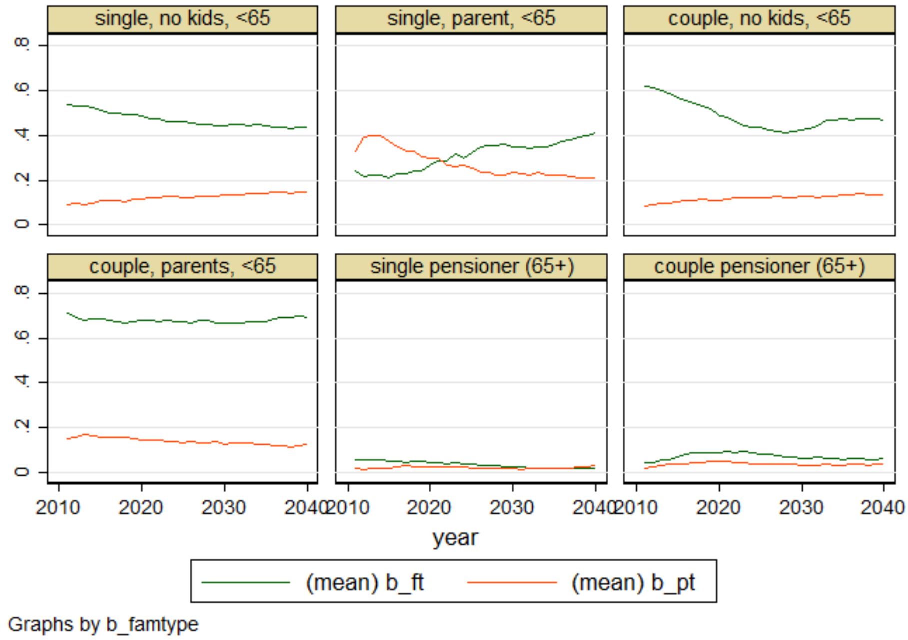
Looking at the trends in average gross market income, benefits and taxes (in the chart below) does not reveal any particular negative trend for this group.



However, trends in the 25th percentile of these variables by family type reveals a strong decrease in gross market income for couple families over the years where these families see an increase in poverty rates. Additional analysis (not shown) confirms this decline is driven by a decline in labour income.



The same trend for childless couples can be seen when looking at full-time and part-time employment rates by family type.



Part of the decline in labour supply by couples may be due to a relative decline in their earnings potential. As we see below, while all points of the earning potential for other family types grow over time, the lower tail of the potential earnings distribution for couple families appears to remain flat over the simulated period.