CeMPA Working Paper Series

CeMPA WP 06/23

SimPaths: An open-source microsimulation model for life course analysis

Patryk Bronka

Justin van de Ven

Daniel Kopasker

Srinivasa Vittal Katikireddi

Matteo Richiardi

March 2023

CeMPA

Centre for Microsimulation and Policy Analysis

SimPaths: An open-source microsimulation model for life course analysis*

Patryk Bronka[a], Justin van de Ven[a], Daniel Kopasker[b], Srinivasa Vittal Katikireddi[b], Matteo Richiardi[a]

[a] Centre for Microsimulation and Policy Analysis, University of Essex

[b] MRC/CSO Social & Public Health Sciences Unit, University of Glasgow

Date: 11 Dec 2023.

Abstract

SimPaths is a family of models for individual and household life course events, all sharing common components. The framework is designed to project life histories through time, building up a detailed picture of career paths, family (inter)relations, health, and financial circumstances. It builds upon standardised assumptions and data sources, which facilitates adaptation to alternative countries – versions currently exist for the UK and Italy, and are under development for Hungary, Poland and Greece. Careful attention is paid to model validation, and sensitivity of projections to key assumptions. The modular nature of the SimPaths framework is designed to facilitate analysis of alternative assumptions concerning the tax and benefit system, sensitivity to parameter estimates and alternative approaches for projecting labour/leisure and consumption/savings decisions. Projections for a workhorse model parameterised to the UK context are reported, which closely reflect observed data throughout a validation window between the Financial crisis (2011) and the Covid-19 pandemic (2019).

Keywords: Dynamic microsimulation, static-dynamic microsimulation linkage, population ageing, open source

JEL codes: C51, C61, C63, H31

---

* Model development has been supported by the National Institute for the Analysis of Public Policies (Italy), the Health Foundation (UK), the National Institute for Health Research (UK), the UKRI Economic and Social Research Council (UK, grant number ES/W001543/1) and the Joint Programming Initiative (JPI More Years Better Lives). The authors bear full responsibility for the study, including any omissions or errors made.

The authors would like to thank Andy Baxter (MRC/CSO Social & Public Health Sciences Unit, University of Glasgow) for his contributions to the SimPaths project.

1 Introduction

The demographic transition driven by ageing of the baby boom generation has profound implications for diverse aspects of OECD societies, including the functioning and financing of the welfare state. As baby boomers (those born in the two decades after WWII) move out of work and into retirement, the shares of national populations of working age are projected to decline, which will reduce public tax receipts at the same time as needs in terms of health care and social assistance are projected to rise.

The old-age dependency ratio (the population aged 65 and over, relative to the population aged 20 to 64) of all OECD countries doubled from 14% in 1950 to 30% in 2020 and is projected to double again to 59% by 2075.[1] Although increases in age dependency ratios are anticipated in all OECD countries by 2075, there is substantial cross-country variation. Korea is an outlier in this series, projected to rise from the lowest dependency ratio in 1950 (6%) to the highest in 2075 (79%). EU countries also feature prominently in the transition, accounting for eight of the ten countries with the highest projected age dependency ratios in the OECD by 2075.

The current rises in age dependency ratios are driven by unprecedented declines in fertility and rises in life expectancy. The OECD average total fertility rate fell by more than half from 3.3 children per woman in 1960 to 1.6 children in 2020.[2] During this same period, the total fertility rate in EU countries fell from 2.6 to 1.5, and from 6.0 to 0.8 in Korea. Similarly, average life expectancy at birth in OECD countries increased from 68.1 years in 1960 to 80.5 years in 2020, from 69.7 to 79.9 years in EU countries, and in Korea from 58.7 years in 1970 to 80.5 years in 2020.[3]

These remarkable shifts in fertility and life expectancy have a pervasive bearing on social and private organisation. From partner relations to education decisions, labour market participation to housing demand, changing gender roles, caring needs, and healthcare provisions; few aspects of modern life are left unaffected. With longer lives, inequalities in income, wealth and health also have more time to compound. In short, OECD countries are passing through a period of social revolution.

Many current trends are now well established, displaying predictable patterns over time. The influence that these trends have on margins of concern are also often predictable. For example, an older population implies a greater prevalence of age pensions in payment, detracting from the public purse. Yet, to move beyond basic postulations, numerical analyses are required. This is particularly true when attempting to take into consideration multiple inter-related temporal trends.

Most numerical approaches used to anticipate the scale and scope of population ageing provide limited detail for exploring distributional effects at a given point in time, longitudinal effects over individual life courses, and implications for financing of the welfare state. The European Commission and OECD, for example, both adopt a cohort methodology to project the scale and effects of population ageing.[4] These methods are based on assumptions concerning cohort-average effects for employment, fertility, health, and mortality. Such cohort averages, however, are ill-suited for exploring fiscal flows associated

---

[1] OECD (2023), Old-age dependency ratio (indicator). doi: 10.1787/e0255c98-en (accessed on 29 March 2023).

[2] OECD (2023), Fertility rates (indicator). doi: 10.1787/8272fb01-en (accessed on 29 March 2023). Total fertility describes the number of children that would be born to each woman if she were to live to the end of her child-bearing years and give birth to children in alignment with the prevailing age-specific fertility rates.

[3] OECD (2023), Life expectancy at birth (indicator). doi: 10.1787/27e0fc9d-en (accessed on 29 March 2023).

[4] See Carone (2005), Scherer (2002) and Burniaux *et al*. (2003).

with the welfare state, which crucially depend upon distributional differences within (as well as between) cohorts.

Interest in within-cohort variation and heterogeneity in life course trajectories has motivated the development of dynamic microsimulation models, especially during the last three decades. In dynamic microsimulation models, the characteristics of each micro unit (individuals in our case) are projected through time from a starting point usually derived from cross-sectional survey (micro)data. Temporal projections are based on biological, institutional, or behavioural rules. Examples of biological rules are ageing and death. Examples of institutional rules are tax and benefits systems. Examples of behavioural rules are any choices that the units can make, for instance related to education, household composition, fertility, labour supply, lifestyle and health behaviour, savings, and investments.

The output from a dynamic microsimulation can usefully be conceptualised as a database that reports evolving information for the population of interest.[5] In a dynamic microsimulation structure, individuals can be linked, so that partner and household characteristics complement individual state variables. New individuals can enter the simulation at later periods, for instance as the result of immigration or fertility. The rules for updating the simulated population include parameters whose values are either exogenously assumed (e.g. tax-benefit parameters) or estimated from available survey data.

Use of dynamic microsimulation methods has grown substantially during the last three decades, benefitting from the increasing availability of high-quality microdata, analytical advances, and increases in computing power.[6] Despite the emergence of generic software packages (openM++, JAS-mine, LIAM2, MODGEN, GENESIS)[7], bespoke analytical frameworks continue to be (re)implemented in the literature. Each independent research group has typically developed its own model code, which is often maintained as a proprietary asset. This imposes considerable developmental overhead on prospective entrants to the literature and limits external validation of reported results.

One way to mitigate developmental costs and facilitate external validation is to publish all research materials as open source. This approach is being actively promoted by the European Commission in its "open access" requirements for funded research, which extend to peer-reviewed publications and research metadata.

This paper describes a novel open-source framework for dynamic microsimulation modelling, which we refer to as SimPaths. All source code is freely available for download without registration, alongside evolving, increasingly detailed documentation. The framework incorporates many state-of-the-art features which are rarely combined in dynamic models.

First, SimPaths generates data for a diverse range of life course domains – education, work, family life and health – explicitly modelling the dynamic feedback effects between them.

Second, SimPaths is linked to an underlying tax-benefit model, which provides a realistic description of the impact of taxes and benefits at both the individual and population level. The detailed tax-benefit

---

[5] As Guy Orcutt, the father of dynamic microsimulation, put it, "I thought, 'If you could represent a real population with a real sample, why couldn't we represent a theoretical population with a synthetic sample? Why couldn't we have a real sample representation of the real population at the start, and then move forward in time according to behavioral relationships applied to micro entities?'" (interview with Duo Qin, 1988, reported in Cheng, 2020).

[6] See O'Donoghue and Dekkers (2018) for a review, and O'Donoghue (2021) for an applied companion study.

[7] openM++: https://openmpp.org. JAS-mine: Richiardi and Richardson (2017). LIAM2: de Menten *et al.* (2014). MODGEN: https://www.statcan.gc.ca/en/microsimulation/modgen/modgen. GENESIS: Gillman (2017).

description that reflects prevailing public policy is important for any evaluation of the funding and distributional implications of population ageing for the welfare state.

Third, SimPaths features rich behavioural models over the principal economic margins of decision making (time-use and savings), where projected choices depend not only on individual characteristics, but also on the influence of fiscal incentives on future expectations.

Fourth, from an architectural perspective, SimPaths is built following a highly modular approach. This facilitates switching between alternative methods for projecting behaviour to allow for sensitivity and robustness analysis. The model is written in Java, using the JAS-mine library (Richiardi and Richardson, 2017).

Fifth, SimPaths is built with an eye to facilitate adaptation to different countries. This is achieved by decoupling the dynamic structure from the tax-benefit model, so that alternative tax-benefit systems can be easily interchanged within the model. Furthermore, care has been taken to describe model dynamics that can be estimated on a single standardised data source for European Union countries (Statistics on Income and Living Conditions, EU-SILC).

The remainder of the paper is structured as follows. Section 2 discusses dynamic microsimulation and how it differs from other projection methods, placing SimPaths in the context of contemporary microsimulation models. Section 3 presents the architecture behind SimPaths. Section 4 discusses model validation. Section 5 concludes.

2 Background

2.1 Econometric modelling from the late 1930s

Econometric modelling was made possible by two developments that occurred during the early 1900s (Klein, 2004). First, there was a shift in focus in economics in favour of mathematical methods, both in the use of statistical evaluation (econometrics; Frisch) and theoretical development (Keynes, Hicks, Marshall, following Jevons, Menger, Walras). Second, the economic shocks of the great depression of the 1930s and the Second World War prompted interest in the development of public statistics. The first sets of National Accounts were published immediately following the end of the Second World War, with the UK publishing in 1946 (covering the period 1938 to 1945), and the US in 1947. The start of modern economic modelling can be traced to attempts to make sense of the newly available data using contemporary economic methodologies.

Pioneering work was conducted by Jan Tinbergen, who is credited with producing the first dynamic macroeconomic model describing an open economy in 1936 (Dhaene and Barten, 1989). This model, which was constructed for the Netherlands, was comprised of 22 equations in 31 variables, and was parameterised using regression techniques to describe correlations described by survey data. Tinbergen followed up this work by constructing a similar model for the US in 1939, and yet another for the UK (published in 1951).

Tinbergen's work has been recognised as “a major force in the transformation of economics from a discursive discipline into a model-building discipline” (Solow, 2004, p. 159). Although Tinbergen's work related closely to that of Keynes, Keynes was sceptical of the value gained through econometric analysis. As he stated in a review of Tinbergen's work: "The method is one neither of discovery nor of criticism. It is a means of giving quantitative precision to what, in qualitative terms, we know already"

(Keynes, 1939, p. 560). Keynes' scepticism has been cited as an explanation for the general lack of interest in econometric modelling in the UK that persisted until the 1960s (Ball, 1963).

Keynes was especially sceptical of the usefulness of empirical formulations not based on theory. As he noted: "the main *prima facie* objection to the application of the method of multiple correlation to complex economic problems lies in the apparent lack of any adequate degree of uniformity in the environment." (Keynes, 1939, p. 567). In this regard Keynes' critique of Tinbergen's approach can be understood as anticipating later developments in the literature (discussed below). Nevertheless, macro-econometric models continue to be developed and maintained through to the present, primarily for forecasting purposes.[8]

2.2 Microsimulation from the late 1950s

Increasingly detailed data and the rise in computing power from the 1930s permitted inclusion of more detailed descriptions of the macroeconomy. A particularly popular direction of research was the inclusion in econometric models of a larger number of equations describing the economy, following the basic approach developed by Tinbergen.

In contrast to this literature, Orcutt (1957) advocated re-specification of economic models in terms of simulated micro-units, which could be aggregated up to macroeconomic aggregates. The motivation for this approach was the observation that non-linearities in the effects of policy on micro-units complicate projections of analyses specified at the aggregate level. Where this is true, then it is necessary to take the distribution of the population into account to obtain valid forecasts. Furthermore, microsimulation models usually make no assumptions concerning the existence of a general equilibrium, as is common of other approaches for modelling the macroeconomy.

Orcutt's seminal work prompted significant interest in the development of microsimulation models during the 1960s and 1970s. Throughout this period, microsimulation models were generally produced using similar econometric techniques as employed by the macroeconomic literature following Tinbergen.

2.3 Rise of structural methods from the 1980s

The growing influence of econometric modelling on the policy reform process motivated critical appraisal of the approach. This line of enquiry culminated in recognition of the limitations of reduced-form specifications for forecasting the effects of policy counterfactuals, especially from the mid-1970s (Conant and Ashby, 1970; Lucas, 1976; Campbell, 1979; Goodhart, 1984).

The primary critique levelled at econometric models focussed on the plausibility of the premise that estimated parameters governing simulated temporal evolution are independent of the policy environment. Econometric models suffered a conspicuous blow after unemployment and inflation were observed to increase in tandem, as anticipated by Lucas' critique of the Philips curve. This led to a shift from the 1980s in macroeconomic modelling in favour of specifications that are plausibly structurally invariant to the evolving policy environment. The literature is broadly comprised of Computable General Equilibrium (CGE) models where uncertainty is ignored, and Dynamic Stochastic General

[8] See, for example, Project LINK: https://www.un.org/development/desa/dpad/project-link.html

Equilibrium (DSGE) models that allow for forward-looking behavioural responses in context of uncertainty (for related reviews see Ghaith *et al.*, 2021, and Christiano *et al.*, 2018).

The period between 1980 and 2000 also saw a decline in academic interest in the development of dynamic microsimulation models. Most of the developmental activity of dynamic microsimulation models during this period was sponsored by government organisations interested primarily in economic forecasts. This work focussed upon the development of dynamic microsimulation models based on econometric methods, supported by increases in computing technology and the increasing availability of high-quality survey microdata (see reviews by Li and O'Donoghue, 2013, and O'Donoghue and Dekkers, 2018).

Of the 60 dynamic microsimulation models reviewed by Li and O'Donoghue (2013), for example, 16 are reported to implement structural behavioural specifications. Furthermore, the authors note that, of those studies that do allow for structural behavioural modelling, "responses are often limited to labour market simulations" (p. 26). Although the authors describe some models designed to project labour supply responses to the tax-benefit system, and others that project retirement responses to the social security system, they describe none that include consumption/savings responses, concluding that there remains "limited implementation of life-cycle models in microsimulation" (p.26).

2.4 Structural dynamic microsimulation from the 2000s

Structural dynamic microsimulation models replace the econometric descriptions used to project behaviour in traditional microsimulation models with structural descriptions based on microeconomic theory; see Appendix B for a brief formal sketch. The most common approach to include structural descriptions of behaviour in microsimulation models ignores the influence of uncertainty, following the seminal studies by Barbara Bergmann (Bergmann et al., 1977) and Gunnar Eliasson (Eliasson et al., 1976; Eliasson, 1977). This can be done either by omitting forward-looking expectations – for example, by focussing on single period employment decisions, see the review by Li and O'Donoghue (2013) – or adopting a specification for utility that has a closed-form solution in context of uncertain expectations (e.g. Pylkkänen, 2002).

Most academic attention during the last two decades has focussed on descriptions of behaviour capable of reflecting precautionary behavioural responses to risk. This focus is driven in part by recognition of the potential importance of precautionary motives underlying behaviour (e.g. Deaton, 1991, Browning and Lusardi, 1996, Carroll, 1997), and in part to test alternative behavioural hypotheses following the seminal study by Gourinchas and Parker (2002).

At least two factors have limited the use of structural microsimulation models that account for precautionary motives. First, these models require numerically demanding Dynamic Programming methods, and the required computing technology has only been widely available from the early 2000s (e.g. Rust, 2008).[9] Related to this point, even where sufficient computing power is available, implementation of Dynamic Programming methods complicates model development considerably.

Second, the late 2000s saw a rise in dissatisfaction with rational agent models. This dissatisfaction motivated a shift in the literature in favour of Agent Based Models (ABMs). ABMs in economics typically replace the assumption of perfect rationality with a form of bounded rationality, in which

[9] There is also a related literature exploring discrete choice dynamic programming; see Keane et al. (2011)

agents are assumed to adapt their behaviour to their environment via a learning process, including the possibility of learning through interactions.

A bifurcation can consequently be observed in the recent literature. Studies focussed on (public) policy alternatives tend to adopt a microsimulation approach, motivated by the need to capture policy detail. In contrast, ABMs tend to be used to conduct theoretical experiments, including the development of general (dis)equilibrium frameworks.[10] In a sense, this shift in focus of ABMs on development of micro-founded projections of general equilibrium frameworks echoes the early work by Barbara Bergmann and Gunnar Eliasson, who developed "general" models comprised of individuals, households and firms, but where behaviour was assumed to follow highly simplified decisions rules (Richiardi, 2014).

2.5 Current modelling frameworks

Dynamic microsimulation models require significant resources to develop and maintain, and are consequently most commonly developed within policy institutions (e.g. government departments), or form part of the modelling infrastructure of research institutions (e.g. Statistics Canada, NATSEM, Urban Institute, CeMPA, GenIMPACT).[11] This is a marked departure from the common academic practice framed upon 'one model – one paper'.[12] It also presents challenges to assessment of prevailing best-practices, as models are often proprietary, and developers often have few incentives to publicly document them.

This section reviews a selection of microsimulation models that satisfy three conditions: there is evidence that the model is in current active use; the model is publicly documented; and the model focuses on life-course dynamics of people. These filters identify seven examples for discussion. The condition on "active use" is a particularly important, as it excludes the majority of examples discussed in previous surveys (O'Donoghue and Dekkers, 2014, 2018, Li and O'Donoghue, 2013, Harding, 2023, O'Donoghue, 2001, and Klevmarken, 1997).[13]

DYNASIM (Favreault et al., 2015) projects a representative sample of the US population forward in time, simulating demographic events such as births, deaths, marriages, divorces, and health status, and economic events such as labor force participation, earnings, hours of work, and retirement. The model is under development at the Urban Institute and evolved from the original work of Orcutt (1976). The model simulates home and financial assets, living arrangements, and includes a detailed calculation of tax and benefit entitlements. In recent years the scope of the model has been considerably expanded to

[10] Most microsimulation models assume a partial equilibrium that omits macro-economic feedback effects; see Maitino et al. (2020) for an exception.

[11] Statistics Canada: https://www.statcan.gc.ca/en/microsimulation/index; NATSEM: https://www.canberra.edu.au/about-uc/faculties/busgovlaw/our-centres/about-natsem, see also Schofield et al. (2023); Urban Institute: https://www.urban.org/research/data-methods/data-analysis/quantitative-data-analysis/microsimulation; CeMPA: www.microsimulation.ac.uk; GenIMPACT: https://www.mq.edu.au/research/research-centres-groups-and-facilities/centres/genimpact-centre-for-economic-impacts-of-genomic-medicine (all websites accessed Dec 7, 2023).

[12] With exceptions: models are sometimes used for different applications, and small tweaks to a model often lead to related publications.

[13] The interest in life-course events is interpreted as excluding the numerous models that focus exclusively on health - for a review, see Schofield et al. (2018).

cover health-related outcomes, including disability, chronic conditions, and projections of health insurance coverage, premium costs, and out-of-pocket medical spending.

MOSART (Andreassen et al., 2020) is a life course model based on administrative data for the entire Norwegian population, which projects birth, death, migration, marriage, divorce, educational activities, labour force participation, retirement, income and wealth based on estimated transition probabilities. The model first became operative in 1990 and is used by Statistics Norway and the Norwegian government for projections and policy analyses related to the pension system.

IrpetDin (Maitino et al., 2020) and T-DYMM (Conti et al., 2024) are two models calibrated to Italian data. IrpetDin is estimated on two different samples: the whole of Italy, and the Tuscany region. It simulates death, ageing, marriage, fertility, divorce, leaving parental home, migration, secondary school enrolment and graduation, university enrolment and graduation, labour force participation, employment status, income, health status, pensions, social assistance for old people and retirees, disability and long-term care. Education is modelled in more detail compared with other microsimulation models, while endogenous projections of labour supply are matched with external projections of labour demand, coming from an auxiliary macro model.

T-DYMM, developed at the Italian treasury, is comprised of a demographic module (fertility, mortality, immigration and emigration, education, exit from parental home, marriages, divorces), a labour market module (employment), a pension module (public and private pensions), a wealth module (home ownership and income from other assets), and a tax-benefit module. Employment distinguishes between self-employment and dependent employment, contract type (open-ended vs. fixed term), part-time vs. full time, and public vs. private sector. All transitions are modelled as reduced-form probabilities.

A more limited focus on the labour market characterises SLAMM, a microsimulation model for Slovakia (Štefánik and Miklošovič, 2020). The microsimulation model projects labour supply, and iscoupled with an external input-output model that projects sectoral employment levels, with wages endogenously adjusting to ensure market closure.

The LifeSim model by Skarda et al. (2020) projects developmental, economic, social and health outcomes from birth to death for each child in the Millennium Birth Cohort (MCS) in England. The model controls for a large number of individual characteristics and behaviour, including human capital development in childhood (social skills, cognitive skills, teenage smoking), and has a focus on mental and physical health, and well-being. All transitions are governed by reduced form probabilities, while life course income profiles are adjusted for individual shifters, such as disability. Taxes and benefits are modelled using stylised functions.

microWELT (Spielauer et al., 2020) reproduces demographic projections for Austria, Spain, Finland, and the UK, by simulating fertility, mortality, education, partnership formation and dissolution, and migration. These projections are then used to re-weight cross-sectional microdata generated by the EUROMOD tax-benefit model (Sutherland and Figari, 2013), allowing computation of a number of National Transfer Account (NTA) indicators.

DYNASIM, MOSART, IrpetDin, and T-DYMM are all proprietary models. SLAMM code is available upon request, while LifeSim and microWELT are open-source.

Many characteristics are common to the models discussed above. Most evolve forward in time a representative cross-section of the population (LifeSim is cohort-based). Most simulate events on a discrete yearly basis (microWelt is cast in continuous time). Most include demographic events, events related to family composition, health events, and economic events (SLAMM is limited to education and economic activity). Most give at least some consideration to tax and benefit policies (SLAMM is again an exception).

Differences between models mostly relate to the respective analytical focus, technical implementation, and econometric specification. In this regard, DYNASIM stands out for its comprehensiveness in both economic and health-related outcomes, while IrpetDin and SLAMM are noticeable for their interaction with a macro module.

Relative to the models discussed above, the main innovations of SimPaths are: (i) focus on facilitating new entrants to the field via open-source coding and associated documentation (in common with LifeSim and microWELT); (ii) externalisation of the tax-benefit component to a third-party dedicated tax-benefit static model (see Section 3.9 below); and (ii) use of a structural model of individual decision-making, rather than simple transition probabilities.

The workhorse version of SimPaths employs a structural model of labour supply where households choose hours worked by each component of the benefit unit in order to maximise the trade-off between leisure and income, in each simulated period. An advanced version extends this behavioural model to take into account intertemporal considerations along the income-leisure and consumption-savings margins (Section 3.8.2). The advantage of a structural behavioural modelling, with respect to reduced-form transition probabilities, is to allow for a direct and transparent channel through which policy incentives affect individual decision making. This brings SimPaths closer to the discrete choice, dynamic programming tradition (e.g. Blundell et al., 2016, 2021).

3 Model description

SimPaths is a fully open-source structural dynamic microsimulation framework, designed to facilitate experimentation with alternative model assumptions. It is coded in Java using the JAS-mine simulation libraries (Richiardi and Richardson, 2017). SimPaths models are currently estimated for the United Kingdom and Italy, and are under development for Hungary, Poland, and Greece. In Section 4 we describe validation of the UK model parameterisation.

SimPaths implements a hierarchical architecture where individuals are organised in benefit units (for fiscal purposes), and benefit units are organised in households.[14] The model projects data at yearly intervals, reflecting the yearly frequency of the survey data used to estimate model parameters. The model is composed of eleven modules: (i) Ageing, (ii) Education, (iii) Health, (iv) Family composition, (v) Social care, (vi) Investment income, (vii) Labour income, (viii) Disposable income, (ix) Consumption, (x) Mental health (2), and (xi) Statistical display. Each module is composed of one or more processes; for example, the ageing module contains ageing, mortality, child maturation, and population alignment processes. Empirical specification of dynamic processes makes extensive use of cross-module characteristics (state variables).[15]

---

[14] A benefit unit is comprised of a single adult or adult couple and their dependent children. There can be households comprised of a single benefit unit, and benefit units comprised of a single individual. A household can be comprised of multiple benefit units only in case of *adult children* continuing to live at their parental home.

[15] For example, lagged employment-related characteristics are not defined for students, but a valuable source of information for other simulated individuals.

Simulated modules and processes are organised as displayed in Figure 1 and

Table 1. In each simulated year, agents are first subject to the ageing process, followed by population alignment. The alignment process adjusts the simulated population to match official population projections distinguished by gender, age (single-year brackets[16]), and geographic region at NUTS1 level[17], which ensures that simulated output remains representative of the country's population.

The education module determines transitions into and out of student status. Students are assumed not to work and therefore do not enter the labour supply module. Individuals who leave education have their level of education re-evaluated[18] and can become employed.

The health module projects an individual's health status, comprising both self-rated general health and mental healthmetrics (based on a clinically validated measure of psychological distress using a Likert scale and a caseness indicator), and determines whether an individual is long-term sick or disabled (in which case, he/she is not at risk of work and may require social care).

The family composition module is the principal source of interactions between simulated agents in the model. The module projects the formation and dissolution of cohabiting relationships and fertility. Where a relationship forms, then spouses are selected via a matching process that is designed to reflect correlations between partners' characteristics observed in survey data. The proportion of the population in a cohabiting relationship is, by default, aligned to the population aggregate in the years for which observational data is available, to account for changes in household structure introduced by the population alignment.

Females in couples can give birth to a (single) child in each simulated year, as determined by a process that depends on a range of characteristics including age and presence of children of different ages in the household. In case of divergence from the officially projected number of newborns, fertility rates are adapted by an alignment process to match population projections for new-born children distinguished by gender, region, and year.

The social care module projects provision and receipt of social care activities for people in need of help due to poor health or advanced age. The module is designed to distinguish between formal and informal social care, and the social relationships associated with informal care. The social care module accounts for the time cost incurred by care providers with respect to informal care, and the financial cost incurred by care receivers with respect to formal care.

The investment income module projects income from investment returns and (private) pensions. The approach taken to project these measures of income depends upon the model variant considered for analysis. Where consumption/savings decisions are simulated using a structural behavioural framework, then asset income is projected based on accrued asset values and exogenously projected rates of return.

---

[16] Up to age 99, and bundled together for centenarians.

[17] Ongoing work is aimed at disaggregating outcomes at NUTS3 level.

[18] Students are assumed to have a "Low" level of education until they leave school for the first time, when it is re-evaluated. Individuals who return to full time education can only improve their level of education.

Alternatively, computational burden of model projections can be economised by proxying non-labour income without explicitly projecting asset holdings.

The labour income module projects potential (hourly) wage rates for each simulated adult in each year and their associated labour activity. Given potential wage rates, hours of paid employment by all adult members of a benefit unit are generated. Labour (gross) income is then determined by multiplying hours worked by the wage rate.

The disposable income module uses information concerning disability, relationship status and fertility, social care, investment income and labour income to evaluate taxes and benefits and disposable income for each projected benefit unit in each year. The model includes alternative methods for projecting employment status, some of which involve interactions between the labour income and disposable income modules to identify preferred combinations of labour supply and disposable income. An alignment routine is used to match projected rates of employment against population aggregates, to correct for biases in the labour supply model.

Given disposable income and household demographics, the consumption module projects measures of benefit unit expenditure. Where the model projects wealth, then a simple accounting identity is used to track the evolution of benefit unit assets through time. A regression-based homeownership process predicts if the primary residence is owned by either of the responsible adults in a benefit unit, in which case the benefit unit is considered to own its home.

A secondary subjective-wellbeing process adjusts estimates obtained by the primary process to account for the effect of exposure to labour market transitions, such as moving in and out of employment and/or poverty.

At the end of each simulated year, SimPaths generates a series of year specific summary statistics. All of these statistics are saved for post-simulation analysis, with a subset of results also reported graphically as the simulation proceeds.

Figure 1: Structure and order of processes modelled in SimPaths

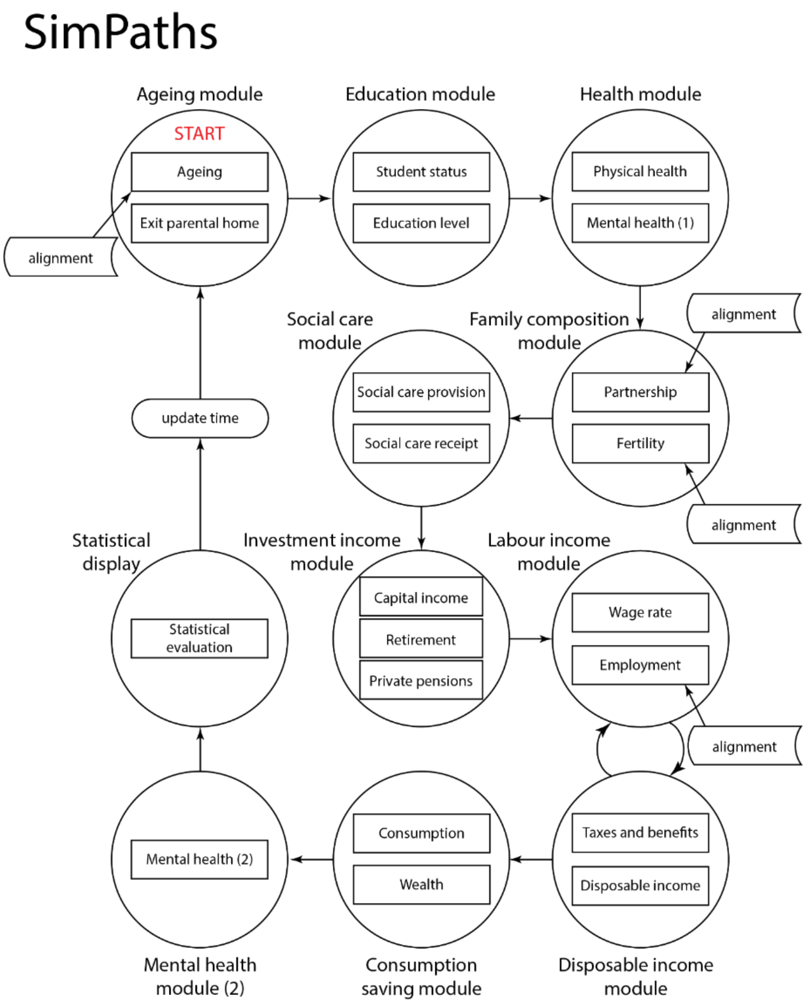


Table 1: List of modules and estimated processes

Module	Process
Ageing	Age increases.
	Probability of leaving the parental home for those who have left education. (Students stay in the parental home).
Education	Probability of remaining in education for those who have always been in education without interruptions.
	Probability of returning to education for those who had left school.
	Level of education for those leaving education.
Health	Self-rated health status for those in continuous education.
	Self-rated health status for those not in continuous education (out of education or returned having left education in the past).

	Probability of becoming long-term sick or disabled for those not in continuous education.
	(Mental Health (1)) Level of psychological distress on GHQ-12 Likert scale and binary case-based indicator of psychological distress.
	(Mental Health (2)) Effect of exposure to employment-state transitions, household income change, and poverty for individuals aged 25 – 64 on psychological distress.
Family composition	Probability of entering a partnership for those in continuous education.
	Probability of entering a partnership for those not in continuous education.
	Probability of partnership break-up.
	Probability of giving birth to a child.
Social care	Probability of needing care for individuals over an age threshold.
	Probability of receiving care for individuals under an age threshold with a disability or long-standing illness or over the age threshold, distinguished by formal, partner, son, daughter, and other providers.
	Hours of care for those in receipt of care, and financial cost for those receiving formal care.
	Probability of providing informal social care.
	Hours of informal social care, among those providing care.
Investment income	Probability of retiring for single individuals.
	Probability of retiring for partnered individuals.
	Probability of receiving capital income for those in continuous education.
	Probability of receiving capital income for those not in continuous education.
	Amount of capital income for those in continuous education.
	Amount of capital income for those not in continuous education.
	Amount of pension income for those who are retired and were not retired in the previous year.
Labour income	Heckman corrected wage equation; females not employed last period.
	Heckman corrected wage equation; males not employed last period.
	Heckman corrected wage equation; females employed last period.
	Heckman corrected wage equation; males employed last period.
	Hours worked, single males.
	Hours worked, single females.
	Hours worked, single male adult children.
	Hours worked, single female adult children.
	Hours worked, males with dependent partner.
	Hours worked, females with dependent partner.
	Hours worked, couples.
Disposable income	Benefit recipiency indicator.
	Amount of disposable income.
Consumption & saving	Consumption.
	Home ownership.
	Savings and assets.
Statistical display	Evaluate summary statistics for simulated population.

3.1 Demographics

3.1.1 Ageing

The first simulated process in each period increments the age of each simulated person by one year. Any dependent child that reaches an exogenously assumed "age of independence" (18 years-of-age in the parameterisation for the UK) is extracted from their parental benefit unit and allocated to a new benefit unit. Individuals are then subject to a risk of death, based on age, gender and year specific probabilities that are commonly reported as components of official population projections. Death is simulated at the individual level but omitting single parent benefit units (to avoid the creation of orphans).

Alignment

Population alignment is performed to adjust the number of simulated individuals to national population projections by age, gender, region, and year. Alignment proceeds from the youngest to the oldest age described by national population projections. Each age is considered in two discrete steps. First, within each age-gender-region-year subgroup, the simulated number of individuals is compared against the associated population projection. Regions with too few simulated individuals (relative to the respective target) are partitioned from those with too many. Net "domestic migration" is then projected by moving individuals from regions with too many simulated people to those with too few, until all options for (net) domestic migration are exhausted. All migratory flows are simulated at the benefit unit level, with reference to the youngest benefit unit member.

Following domestic migration, remaining disparities between simulated and target population sizes are adjusted to reflect international immigration (if the simulated population is too small), or emigration and death (if the simulated population is too large). Like domestic migration, international migration is simulated net of opposing flows[19] and at the benefit unit level with reference to the youngest benefit unit member. Death is simulated in preference to international emigration for population alignment for all ages above an exogenously imposed threshold (65 for the UK).

Except for the distinction between age, gender, region, and year, all transitions simulated for population alignment are randomly distributed. This means that the model does not reflect, for example, the higher incidence of international emigration among prior international immigrants. Furthermore, the model projects international immigration by cloning existing benefit units (e.g. Duleep and Dowhan, 2008) without taking into consideration any systematic disparities between the domestic and migrant populations, including with regard to their respective financial circumstances.

3.1.2 Leaving parental home

Individuals who have recently attained the assumed age of independence and were moved to separate benefit units (see 3.1.1) are evaluated to determine if they leave their parental home. Any individual still in education is assumed to remain a member of their parental household.[20] For mature children not in education, the probability of leaving their parental home is based on a probit model conditional on gender, age, level of education, lagged employment status, lagged household income quintile, region,

[19] That is, only immigration or emigration is projected for each population subgroup, not both.

[20] In the simulation, this is represented by a household comprised of the parental benefit unit, and one, or more, benefit units representing adult children.

and year (to reflect observed time trends). Mature children who are projected to remain in their parental homes may leave in any subsequent year.

3.2 Education

3.2.1 Student status

Individuals leave continuous full-time education during an exogenously assumed age band (16 to 29 for the UK). The probability of leaving continuous full-time education within this age band is described by a probit model conditional on gender, age, mother's education level, father's education level, region, and year.[21]

Individuals who are not in education may re-enter education within another exogenously assumed age band (16 to 45 for the UK). In this case, the probability of re-entering education is described by a probit model conditional on gender, age, lagged level of education, lagged employment status, lagged number of children in the household, lagged number of children aged 0-2 in the household, mother's and father's education levels, region, and year.

Students are considered not to work. Those who return to education can leave again in any subsequent year.

3.2.2 Educational level

Individuals who cease to be students are assigned a level of education based on an ordered probit model that conditions on gender, age, mother's and father's education level, region, and year. The education level of individuals who exit student status after re-entering education may remain unchanged or increase but cannot decrease.

3.3 Health

3.3.1 Physical health

Physical health status is projected on a discrete 5-point scale, designed to reflect self-reported survey responses (between "poor" and "excellent" health). Physical health dynamics are based on an ordered probit, distinguishing those still in continuous education. For continuing full-time students, the ordered probit conditions on gender, age, lagged benefit unit income quintile, lagged physical health status, region, and year. The same variables are considered for individuals who have left continuous education, with the addition of education level, lagged employment status, and lagged benefit unit composition.

3.3.2 Long-term sick and disabled

Any individual aged 16 and above who is not in continuous education can become long-term sick or disabled. The probability of being long-term sick or disabled is described by a probit equation defined with respect to lagged disability status, prevailing and lagged physical health status, gender, age, education, income quintile, and lagged family demographics.

[21] Conditioning on parental education introduces a correlation between socio-economic position across generations and facilitates investigation of intergenerational inequality.

3.3.3 Psychological distress 1 (baseline level and caseness)

In each simulation cycle, a baseline level of psychological distress for individuals aged 16 and over is determined using the 12-item General Health Questionnaire (GHQ-12). Two indicators of psychological distress are computed: a Likert score, between 0 and 36, estimated using a linear regression model; and a dichotomous indicator of the presence of potentially clinically significant common mental disorders[22] is obtained using a logistic regression model. Both specifications are conditional on the lagged number of dependent children, lagged health status, lagged mental health, gender, age, level of education, household composition, region, and year.

3.3.4 Psychological distress 2 (impact of economic transitions and exposure to the Covid-19 pandemic)

The baseline measures of the level and caseness of psychological distress described above are modified by the effects of economic transitions and non-economic exposure to the Covid-19 pandemic. Fixed effects regressions are used to estimate the direct impact of transitions from employment to non-employment, non-employment to employment, non-employment to long-term non-employment, non-poverty to poverty, poverty to non-poverty, and poverty to long-term poverty, as well as changes in growth rate of household income, a decrease in household income, and non-economic effect of the exposure to Covid-19 pandemic in years 2020 and 2021. The effects of economic transitions are estimated on pre-pandemic data to ensure validity in other periods. The non-economic effects of the pandemic are estimated using a multilevel mixed-effects generalized linear model. Further details of the estimation procedure are provided in Kopasker et al., 2023.

3.4 Family composition

3.4.1 Partnerships and cohabitation

Individuals aged 18 and over who do not have a partner may decide to enter a partnership based on the outcome of a probit model. For students, the probit conditions on gender, age, lagged household income quintile, lagged number of (all) dependent children, lagged number of children aged 0-2, lagged self-rated health status, region, and year. For non-students, the probit conditions on the same set of variables as for students, expanded to include level of education and lagged employment status.

Individuals who enter a partnership are matched using either a parametric or non-parametric process, focussing exclusively on opposite-sex relationships. In the (default) parametric matching process, the model searches through the pools of males and females identified as cohabiting in each simulated period to minimise the distance between individual expectations, in terms of partner's ideal earnings potential and age, and individual characteristics of each individual in the matching pool. The matching procedure prioritises matching individuals within regions, but if the sufficient quantity and / or quality of matches cannot be achieved, matching is performed nationally. In contrast, the non-parametric process uses an iterative proportional fitting procedure to replicate the distribution of matches observed in survey data between different types of individuals, where a type is defined as a combination of gender, region, education level, and age.

Partnership dissolution is modelled at the benefit unit level with the probability described by a probit model conditional on female partner's age, level of education, lagged personal gross non-benefit

[22] Individuals scoring four points or more on a 0-12 scale are classified as positive cases.

income, lagged number of (all) children, lagged number of children aged 0-2, lagged self-rated health status, lagged level of education of the spouse, lagged self-rated health status of the spouse, lagged difference between own and spouse's gross, non-benefit income, lagged duration of partnership in years, lagged difference between own and spouse's age, lagged household composition, lagged own and spouse's employment status, region, and year.

Alignment

The matching processes for new relationships outlined above fails to identify matches for all individuals flagged as entering a partnership by the related probit equations. This tends to bias the simulated population, resulting in an under-representation of partner couples. An alignment process is consequently used to match the rate of incidence of partner couples to survey targets. The alignment process works by adjusting the intercept of the probit relationships governing relationship formation, increasing the intercepts where the incidence of couples is too low.

3.4.2 Fertility

Females aged 18 to 44 can give birth to a child whenever they are identified in a partnership. The probability of giving birth is described by a probit model conditional on a woman's age, benefit unit income quintile, lagged number of children, lagged number of children aged 0-2, lagged health status of the woman, lagged partnership status for those in continuous education. For those not in continuous education, the probability of giving birth is described by a probit model conditional on a woman's age, the fertility rate of the UK population, benefit unit income quintile, lagged number of children, lagged number of children aged 0-2, lagged health status of the woman, lagged partnership status, lagged labour market activity status, level of education, and region. The inclusion of the overall fertility rate allows the model to capture fertility projections for future years, whereas the overall change in projected fertility is distributed across individuals according to their observable characteristics.

Alignment

The number of projected births is aligned to the number of newborns supplied by the official projections used for population alignment. The alignment procedure randomly samples fertile women and adjusts the outcome of the fertility process until the target number of newborns has been met.

3.5 Social care

3.5.1 Receipt of social care

The model distinguishes between individuals aged above and below an age threshold when projecting receipt of social care. This reflects the relatively high prevalence of social care received by older people, for whom more detailed information is often reported by publicly available data sources.

Receipt of social care among older people

For individuals aged above an exogenously defined threshold (65 years in the UK), the model begins by considering whether an individual is in need of care. This is simulated as a probit equation that varies by gender, education, relationship status, whether care was needed in the preceding year, self-reported health, and age. The probability of receiving care is projected using a similar set of explanatory variables. Where an individual is identified as receiving care, a multinomial logit equation is used to determine if the individual receives: i) only informal care; ii) formal and informal care; or iii) only formal care. This multinomial logit varies by education, relationship status, and age band in addition to a lag dependent variable.

For individuals projected to receive informal care, a multi-level model is used to distinguish between alternative care providers. The first level considers whether a partner provides informal care, for individuals with partners. For individuals who receive social care from their partner, the second level uses a multinomial logit to consider whether they also receive care from a daughter, a son, or someone else (other). For individuals in receipt of informal care who do not have a partner caring for them, another multinomial logit is used to select from six potential alternatives that allow for up to two carers from “daughter”, “son”, and “other”. Log-linear equations are then used to project the number of hours of care received from each identified carer. Finally, hours of formal care are converted into a cost, based on the year-specific mean hourly wages for all social care workers.

Receipt of social care among younger people

Receipt of social care among individuals under the exogenously assumed age threshold is simulated using a more stylised approach to that described for older people, reflecting the less detailed data available for parameterisation. In this case, the model focusses exclusively on informal social care for individuals simulated to be long-term sick or disabled. At the time an individual is projected to enter a disabled state, a probit equation is used to identify whether the individual receives informal social care. This identification is assumed to persist for as long as the person remains disabled.

If an individual under age 65 is identified as receiving social care, then the time of care received is described by a log-linear equation.

3.5.2 Provision of social care

The model is adapted to project provision of social care by informal sector providers; the characteristics of formal sector providers of social care are beyond the current scope of the model. The approach adopted for simulating receipt of social care described above identifies the incidence and hours of informal social care that individuals are projected to receive. In the case of people over the assumed age threshold, it also identifies the relationship between those in receipt of informal social care and their informal care providers, and the persistence of those care relationships. These details consequently provide much of the information necessary to simulate provision of informal social care, in addition to the receipt of care.

Nevertheless, the data sources for starting populations considered for SimPaths – with the notable exception of partners – generally omit social links that are implied to exist between informal social care providers and those receiving care. Specifically, links between adult children and their parents, and the wider social networks that often supply informal social care services are generally not recorded. The method that is used to project informal provision of social care is designed to accommodate limitations of the simulated data in a way that broadly reflects projection of social care receipt discussed above.

Specifically, the model distinguishes between four population subgroups with respect to provision of informal social care: (i) no provision; (ii) provision only to a partner; (iii) provision to a partner and someone else; and (iv) provision but only to non-partners. For people who are identified as supplying informal care to their partner via the process described in Section 3.5.1, a probit equation is used to distinguish between alternatives (ii: provision only to partner) and (iii: provision to a partner and someone else). Similarly, for the remainder of the population, another probit equation is used to distinguish between alternatives (i) and (iv). A log linear equation is then used to project number of hours of care provided, given the classification of who care is provided to.

3.6 Retirement

Simulation of retirement varies slightly depending on the accommodation of forward-looking expectations (see Section 3.8.2). In both cases, retirement is possible for any adult above an assumed age threshold (50 in the parametrisation for the UK). When forward-looking expectations are implicit, entry to retirement is based on a probit model that controls for gender, age, education, lagged employment status, lagged (benefit unit) income quintile, lagged disability status, indicator to distinguish individuals in excess of state pension age (accounting for changes in the state pension age), region, and year. For couples, characteristics of the spouse (employment status, reaching retirement age) also affect the probability of retirement. When forward-looking expectations are explicit, then entry to retirement is considered to be a control variable. Retired individuals may receive pension income, as described in Section 3.7.

3.7 Investment income

Investment income in SimPaths is comprised of capital income and private (non-public, personal, or occupational) pensions. The methods used to project these sources of income vary depending on whether wealth is included in the set of characteristics projected by the model. Wealth is omitted from the simulation by default but is tracked when discretionary consumption and employment decisions are simulated to reflect forward-looking behavioural incentives (described in Section 3.8.2).

3.7.1 Capital income

Wealth implicit

When wealth is not projected by the model, then the incidence of capital income among the simulated population aged 16 and over is based on probabilities described by a logit regression equation that varies by age, lagged health, lagged gross employment and capital income, region and year. For individuals not in continuous education, the list of explanatory variables for the logit equation also includes education status, lagged employment status, and lagged household composition.

For individuals simulated to be in receipt of capital income, the amount of capital income is described by linear regression models that condition on gender, age, lagged health status, lagged gross employment income, lagged capital income, region, and year for individual in continuous education. Individuals not in continuous education are also distinguished by their level of education, lagged employment status, and lagged household composition.

Wealth explicit

When wealth is explicitly projected by the model, then capital income is the product of net asset holdings and an assumed rate of return. The rate of return varies by year, and by the value of benefit unit net wealth, $w_{i,t}$, as described by:

$$r_{i,t} = \begin{cases} r_{a,t} & \text{if } w_{i,t} \geq 0 \\ r_{dl,t} + (r_{du,t} - r_{dl,t})\varphi_{i,t} & \text{otherwise} \end{cases} \quad (1)$$

where *i* denotes the benefit unit and *t* time. $1 \geq \varphi_{i,t} \geq 0$ denotes the (bounded) ratio of benefit unit debt to full-time potential earnings. Assuming $r_{du,t} \geq r_{dl,t}$ reflects a ‘soft constraint’ where interest rates increase with indebtedness.

3.7.2 Private pensions

Private pensions are projected for adults identified as having retired in the model. The projection of retirement is described in Section 3.6.

Wealth implicit

When wealth is implicit in the model, then private pension income is projected using a linear regression model that conditions on age, level of education, lagged household composition, lagged health status, lagged private pension income, region, and year for individuals who continue in retirement. For individuals entering retirement, the probability of receiving private pension income is first determined using a logit model that conditions on having reached the state pension age, level of education, lagged employment status, lagged household composition, lagged health status, lagged hourly wage potential, region, and year. The amount of pension income is projected using a linear regression model conditional on the same observed characteristics.

Wealth explicit

When the simulation projects wealth explicitly, then an assumed fraction of benefit unit wealth at the time of retirement is converted into a life annuity, or joint-life annuity for adult couples. Annuity rates in the model are actuarily fair, given (cohort specific) mortality rates and an assumed internal rate of return.

3.8 Labour income

3.8.1 Wage rates

Hourly wage rates are simulated for each adult in the model based on Heckman-corrected regressions stratified by gender and lagged employment status (distinguishing between employed and not-employed individuals) that include as explanatory variables, part-time employment identifiers, age, education, student status, parental education, relationship status, presence of children, health, and region. For individuals observed in employment in the previous year, lagged (log) hourly wage rates are also included as an explanatory variable.

3.8.2 Employment decisions

Two alternative methods for projecting employment decisions can be considered by the model. These alternatives are both designed to reflect the influence of financial incentives on behaviour and are distinguished by whether they reflect forward-looking expectations.

Expectations implicit

The default specification of SimPaths projects labour supply using a non-forward-looking random utility model. This approach is common in the associated literature (see review by Li and O'Donoghue, 2013), and has the advantage that it limits computational burden.

The method projects labour supply as though employment decisions are made to maximise within-period benefit unit utility over a discrete set of labour/income alternatives (by default, 5 alternatives for individuals, and 25 for couples). Given any labour alternative, labour income is computed by combining hours of work with the respective hourly wage rate, projected as discussed in Section 3.8.1. The utility of the benefit unit is calculated using a quadratic utility function and takes as arguments benefit unit disposable income (see Section 3.9) and the number of hours worked by adult members.

Alignment

The estimated utility of single men, single women, and couples is adjusted to align the aggregate employment rate to the employment rate observed in the data in the validation period. The final adjustment value is used in the subsequent periods, for which no data is available. This procedure accounts for the existence of unemployment in the real economy and the fact that labour supply decisions simulated using the random utility model assume no constraints on labour demand in the economy.

Expectations explicit

The model can be directed to project labour and discretionary consumption to reflect forward-looking expectations for behavioural incentives. As for the implicit expectations case, the unit of analysis is the benefit unit. Incentives are translated into behaviour via an assumed intertemporal utility function. By default, the model adopts a nested constant elasticity of substitution (CES) utility function as described by equation (2), although the model is designed to facilitate experimentation with alternative specifications.

$$U_{i,t} = \frac{1}{1-\gamma}\left\{u\left(\hat{c}_{i,t}, l_{i,t}\right)^{1-\gamma} + E_{i,t}\left[\sum_{j=t+1}^{\infty} \delta^{j-t}\left(\varphi_{i,j} u\left(\hat{c}_{i,j}, l_{i,j}\right)^{1-\gamma} + \left(1-\varphi_{i,j}\right) Z\left(w_{i,j}\right)^{1-\gamma}\right)\right]\right\} \quad (2)$$

$$u\left(\hat{c}_{i,t}, l_{i,t}\right) = \left[\hat{c}_{i,t}{}^{1-1/\varepsilon} + \alpha^{1/\varepsilon} l_{i,t}{}^{1-1/\varepsilon}\right]^{\frac{1}{1-1/\varepsilon}} \quad (3)$$

$$Z\left(w_{i,j}\right) = \zeta_0 + \zeta_1 w_{i,j}^{+} \quad (4)$$

where subscripts i denotes benefit unit and t time. $u\left(\hat{c}_{i,t}, l_{i,t}\right)$ represents within period utility derived from equivalised discretionary consumption ($\hat{c}$) and time spent in leisure (l). $Z(w)$ represents the warm-glow model of bequests, derived from non-negative net wealth at death (w^{+}). E is the expectations operator and φ the probability of survival of the benefit unit reference person, which varies by gender, age and year. $\gamma > 0$ is the coefficient of relative risk aversion, $\varepsilon > 0$ the elasticity of substitution between equivalised consumption and leisure, α the utility price of leisure, and δ the constant exponential discount factor.

Each adult is considered to have three alternative labour supply options, corresponding to full-time, part-time and non-employment. Labour supply and discretionary consumption are projected as though they maximise the assumed utility function, subject to a hard constraint on net wealth and assumed agent expectations. Expectations are "substantively rational" in the sense that uncertainty is characterised by the random draws that underly dynamic projection of modelled characteristics. As no analytical solution to this problem exists, numerical solution methods are employed as is now standard in the dynamic programming literature (see e.g. van de Ven, 2017).

The model proceeds in two discrete steps. The first step involves solution of the lifetime decision problem for any potential combination of agent specific characteristics, with solutions stored in a look-up table. The second step uses the look-up table as the basis for projecting labour supply and discretionary consumption. Technical details of the numerical solution method are provided in Appendix C.

3.9 Disposable income

Disposable income is simulated by matching each simulated benefit unit in each projected period with a *donor* benefit unit reported by a tax-benefit reference database, following the procedure described by van de Ven *et al*. (2022). The database stores details of taxes and benefits alongside associated demographic and private income characteristics for a sample of benefit units. This database could be populated from a wide range of sources. The approach was originally formulated to draw upon output data derived from the UK version of EUROMOD (UKMOD), a static tax-benefit microsimulation model (see Richiardi *et al*., 2021), and then extended to accommodate projections from any EUROMOD country.

The matching procedure for benefit units applies coarsened exact matching over a number of discrete-valued characteristics, followed by nearest-neighbour matching on a set of continuous variables. The nearest neighbour matching is performed with respect to Mahalanobis distance measures evaluated over multiple continuous valued characteristics.

The default set of discrete value characteristics considered for matching includes age of the benefit unit reference person, relationship status, numbers of children by age, hours of work by each adult member, disability status, and informal social care provision. Similarly, the default set of continuous value matching characteristics includes original (pre-tax and benefit) income, second income (to allow for income splitting withing couples), and formal childcare costs.

Having matched a simulated benefit unit to a donor, disposable income is imputed via one of two methods. For benefit units with original income above a “poverty threshold”, disposable income is imputed by multiplying original income of the simulated benefit unit by the ratio of disposable to original income of the donor unit. For benefit units below the considered poverty threshold, disposable income is set equal to the (growth adjusted) disposable income of the donor.

Finally, adjustments to account for public subsidies for the costs of (formal) social care are evaluated separately from the database approach described above, based on internally programmed functions. This is done because public subsidies for social care are not always included in database sources (e.g. tax-benefit models) considered for analysis.

3.10 Consumption and savings

3.10.1 Non-discretionary expenditure

The model can project two forms of non-discretionary benefit unit expenditure: formal social care costs and formal childcare costs. As described in Section 3.5, social care costs are projected based on projections of hours of formal social care received and assumed hourly wage rates for social care workers.

Childcare costs are projected using a double-hurdle model, characterised by a probit function describing the incidence of formal childcare costs and a linear least-squares regression equation describing the value of childcare costs when these are incurred. Both equations include the same set of explanatory variables describing the number and age of dependent children in a benefit unit, the relationship status and employment status of adults in the benefit unit, whether any adult in the benefit unit is higher educated, region, and year.

3.10.2 Discretionary consumption

As discussed in Section 3.8.2, the model can be directed to project employment and discretionary consumption jointly to reflect forward-looking expectations for behavioural incentives. The projection of discretionary consumption varies depending on whether forward-looking expectations are chosen to be explicit or implicit within a simulation.

Expectations implicit

Yearly equivalised disposable income is calculated by adjusting disposable income (see Section 3.9) to account for benefit unit demographic composition via the modified OECD scale. Equivalised consumption is set equal to equivalised disposable income for retired individuals, and to disposable income adjusted by a fixed discount factor to account for an implicit savings rate otherwise. The assumed savings rate, in turn, influences simulated capital income (see Section 3.7.1).

Expectations explicit

As discussed in Section 3.8.2, the model evaluates solutions to the lifetime decision problem in the form of a look-up table when directed to reflect forward-looking expectations for behavioural incentives. In the case of discretionary consumption, the look-up table stores the ratio of consumption to "cash on hand", where cash on hand is the sum of net wealth, disposable income, and available lines of credit. This ratio has the advantage that it is bounded between zero and one, which facilitates the computational routines and consideration of selected policy counterfactuals.

3.10.3 Assets accumulation

Net wealth is the key transition mechanism that balances intertemporal behavioural incentives when forward-looking expectations are treated explicitly by the model. In this case, dynamic evolution of wealth in most periods is described by the accounting identity:

$$w_{i,t} = w_{i,t-1} + y_{i,t} - c_{i,t} - \bar{c}_{i,t} \tag{5}$$

where $w_{i,t}$ denotes the net wealth of benefit unit i in period t, $y_{i,t}$ disposable income, $c_{i,t}$ discretionary consumption, and $\bar{c}_{i,t}$ non-discretionary expenditure. The only departures from equation (5) are at the time of retirement if $w_{i,t} > 0$, when a fixed fraction of net wealth is converted into a fixed life annuity (see Section 3.7.2), or if there is a change in relationship status. In context of relationship formation, the wealth of each new partner is aggregated to obtain the wealth of the new benefit unit. In context of relationship dissolution due to separation, each spouse is assumed to take half the wealth of their preceding benefit unit. Relationship dissolution due to spouse death has no effect on benefit unit with, reflecting the implicit assumption that all wealth of the deceased passes to their surviving spouse.

Home ownership

Although net wealth is not disaggregated in the model, the incidence of home ownership is reflected, as this is used as an input to for projection of psychological distress (Section3.3.3 – 3.3.4). Home ownership is evaluated at the benefit unit level, by considering if at least one of the adult occupants is classified as a homeowner. At the individual level, home ownership is determined using a probit regression model conditional on gender, age, lagged employment status, education level, lagged self-rated health, lagged benefit unit income quintile, lagged gross personal non-employment non-benefit income, region, lagged household composition, lagged spouse's employment status, and a time trend.

3.11 Assessing simulated uncertainty

Uncertainty regarding a model's projections arise for a variety of reasons (Bilcke *et al*., 2011; Creedy *et al*., 2007):

(i) Input data; due to sampling or measurement errors in initial survey populations.

(ii) Model structure[23]; referring to the validity of the modelling approach adopted.
(iii) Model specification; concerning the choice of the covariates and the functional forms used, and in particular the crucial assumption that any regularity observed in the data will persist into the future.
(iv) Model parameters; concerning the statistics imprecision of parameter estimates and/or exogenously derived parameters.
(v) Montecarlo variation; concerning sensitivity of simulated aggregates of interest to the set of random draws used to project diversity among simulated agents.

Studies based on microsimulation methods frequently ignore these sources of uncertainty, which is a recognised source of critique (e.g. Goedemé et al., 2013). This omission can generally be attributed to the observation that "the calculation of confidence intervals around model results that account for all sources of error remains a major challenge" (Mitton et al., 2000).

The first source of uncertainty listed above (i) should decline with the increasing availability of high-quality survey data, and in any case is generally beyond the scope of expertise of data analysts. Sources (ii) and (iii) that focus on model specification can be explored using established statistical techniques based on in-sample and out-of-sample measures of fit.

SimPaths accounts for parameter uncertainty (iv) by including routines that facilitate bootstrapping parameter estimates, based on estimated point values and covariance matrices. This involves repeated simulations, each based on a different random draw for model parameters. Similarly, Montecarlo variation (v) can be explored by conducting repeated simulations each based on fresh set of random draws or by arbitrarily scaling-up the simulated population size. These methods can be used to generate a distribution of model outcomes, around central projections.

4 Estimation and validation

The workhorse version of the model includes two of the three frontier features of dynamic microsimulation models discussed above. It employs a structural random utility specification to project labour supply, and the household budget constraint is derived from a tax-benefit calculator.[24]

4.1 Data

SimPaths uses three types of data as input:

1. The *initial population* to be evolved over time.
2. *Donor populations* used to impute the effects of tax and benefit policy.
3. Estimated *parameters* governing transition probabilities assumed by the model.

The model has been designed to draw the initial population from data reported by the UK Household Longitudinal Study (UKHLS). The UKHLS, (sometimes referred to as Understanding Society), is the successor to the British Household Panel Survey, and is the principal general-purpose panel survey

[23] Sometimes referred to as "methodological uncertainty".

[24] Other versions under development replace the cross-sectional utility maximisation driving labour supply choices with an intertemporal maximisation procedure determining both labour supply and consumption/saving behaviour, hence incorporating expectations and life-cycle decision making. General equilibrium feedback is still absent.

administered in the UK. Multiple initial populations are derived from the UKHLS, corresponding to different years of data reported by the survey (from 2011 to 2017), and used for model validation (see below).

The donor populations for tax and benefit imputations are derived from UKMOD and are based on data reported by the Family Resources Survey (FRS). These data include a wide range of benefit unit characteristics in addition to tax and benefit payments. SimPaths imputes tax and benefit payments from these data by matching simulated individuals to individuals described by donor populations.

Parameters for the UK have been estimated on UKHLS data, Waves 1 to 8, and FRS (labour supply and social care, various years). Estimates are reported in Appendix A.

4.2 Validation

We validate the workhorse version of SimPaths parameterised to the UK. This version of the model uses static labour supply optimisation and includes alignment to population projections, cohabitation rates, and aggregate employment rates. The validation was undertaken by comparing simulated and observed data, starting with observations reported for 2011, and then at annual intervals to 2019. This sample window avoids complications associated with the 2008 Financial crisis on the one hand, and the Covid-19 pandemic on the other. This validation window overlaps the sample frame used to estimate model parameters (2009-2017).

Validation is always motivated by the need to increase confidence in the model (National Research Council, 2012). This, in turns, depends on the research questions that the model is designed to address, which should ultimately determine what validation tests the model has to pass. SimPaths is currently being used for a number of different research projects, with more applications being evaluated: a discussion of all the different research questions involved is therefore outside the scope of the present paper. Consequently, we opt here for a generic evaluation of how "realistic" the full set of model outcomes are, under a baseline parameterisation. Given the large number of state variables in the model, such a broad validation strategy spans multiple dimensions, covering both cross-sectional (evolution of summary statistics of variables over time) and longitudinal measures (transitions between states), referring both to individual variables and to their joint distribution (e.g. correlations).[25] For the sake of

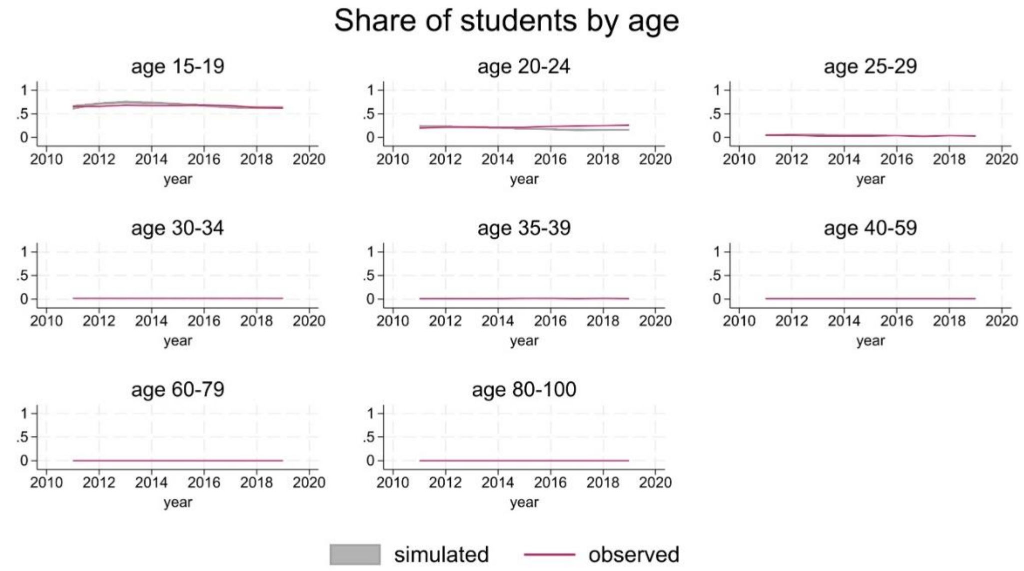
[25] Details of the process undertaken to arrive at the validation reported here are too numerous to recount succinctly, and so the current section focusses on the outcome of the process rather than the process itself, and the lessons learnt. A quick aside may, however, provide the reader with some appreciation for the issues involved. Empirical estimations of the equations that govern evolution of relationship status in the model were evaluated using pooled data from the UKHLS. From this basis, it was found that SimPaths tended to understate proportions of population projected to be in a relationship during the validation window. After further analysis, it was found that the discrepancy between the model and survey data was attributable to the interaction between the matching method used to identify partners from within the pool of simulated individuals identified as entering a relationship, and the probit relationships governing the incidence of relationship transitions. This type of mis-match reflects an issue that underlies any effort to identify parameters outside of a given model's structure. Ideally, all model parameters should be evaluated together and endogenous to the model of interest. This is the case, for example, in related two-stage econometric methods including Simulated Minimum Distance (Lee & Ingram, 1991), Method of Simulated Moments (Stern, 1997), Indirect Estimation (Gourieroux, et al, 1993), and Efficient Method of Moments (Gallant and Tauchen, 1996). In practice that is often not possible, which generates a source of model mis-identification. In the current context, an alignment method was implemented to account for the model mis-match. The alignment method adjusts the intercept of the estimated probit equations governing the incidence of entering a relationship until the model reflects summary statistics for the proportions of the population observed

brevity, we discuss here only a selection of cross-sectional measures, presented in graphical form for ease of visualisation, leaving validation of longitudinal measures to a future exercise. For each simulated series, 95% confidence bands are displayed, computed based on the uncertainty assessment strategy outlined in Section 3.11. The simulated confidence bands are shown against the weighted means of the corresponding variables computed on the UKHLS data.

4.2.1 Education

SimPaths reproduces the distribution of students by age accurately (Figure 2).

Figure 2: Student status



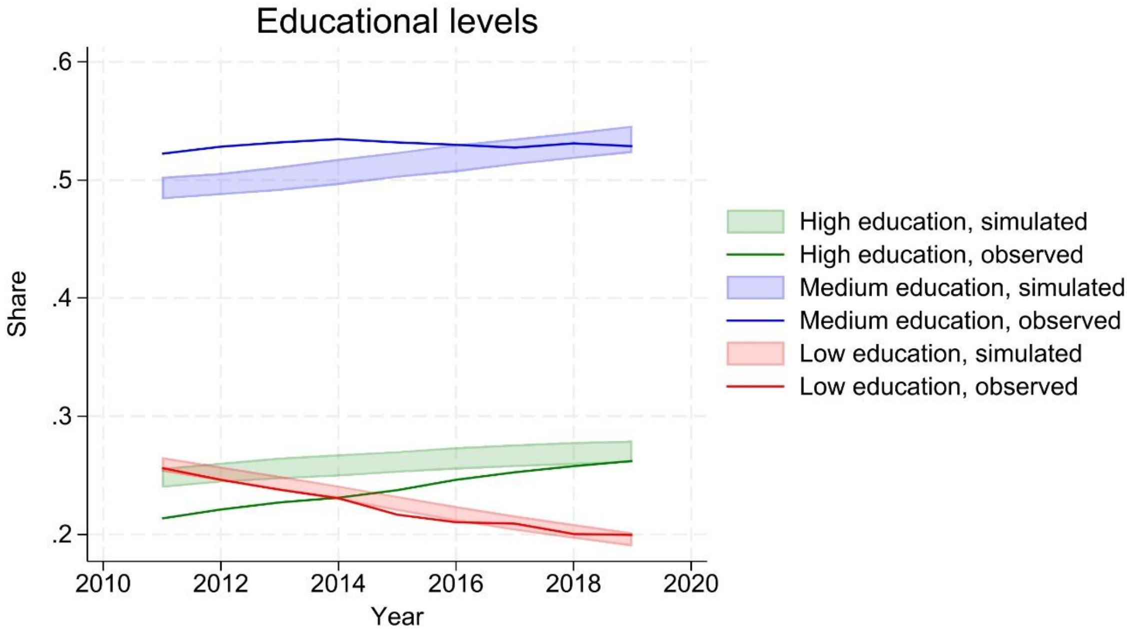
Simulated educational attainments show convergence between the simulated and observed share of the population with high education, starting from a higher simulated level (

---

in survey data to be in a relationship. For years beyond the validation period, the intercept is then kept constant at the last calibrated level.

Figure 3). This implies a slower increase than observed in the data. This is partly attributable to a conservative choice about continuation of estimated trends in the projections. More in details, specifications assume a linear time trend. This is motivated by the relatively short length of the estimation panel, which would not support a more flexible modelling of the time trend. However, extrapolating a linear trend is problematic, as it will eventually lead to implausible levels of the variable of interest. In projections, SimPaths stops the estimated linear trends after a given calendar year. The default option – adopted in this and other processes - is to stop any estimated time trend at the end of the estimation sample (2017). Data shows that the trend towards increasing educational levels is continuing beyond 2017.

Figure 3 : Educational attainment



4.2.2 Health

SimPaths distinguishes between a general health score (Likert scale 1-5), and a psychological distress score (Likert scale 0-36). Projected distribution of general health by age and gender is in line with observations (

Figure 4 and Figure 5).

Figure 4: General health score, men

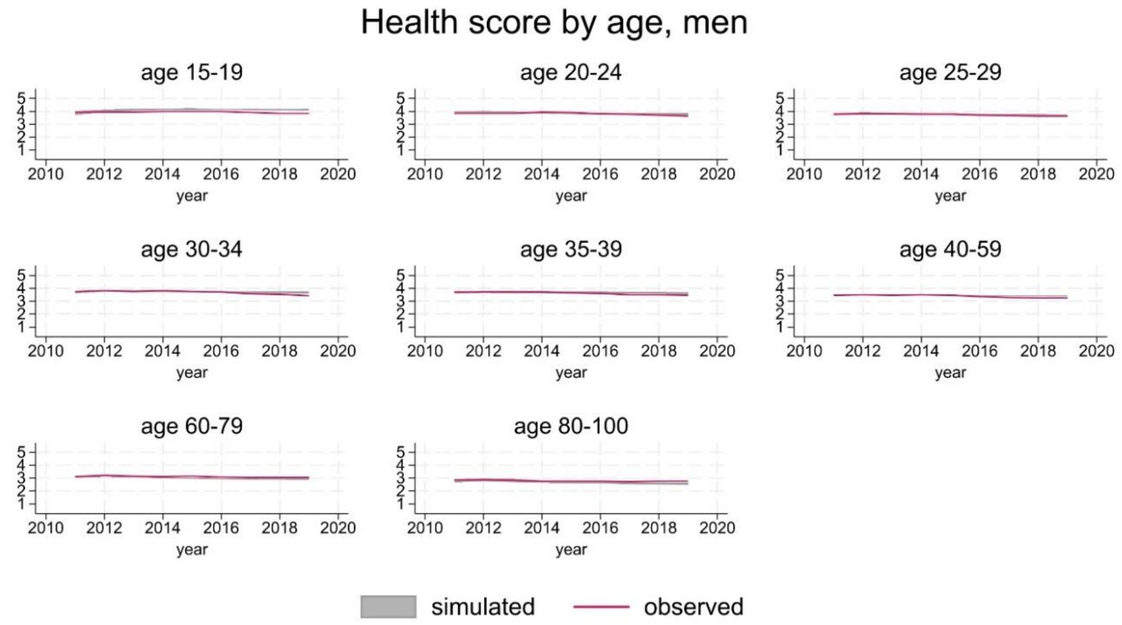
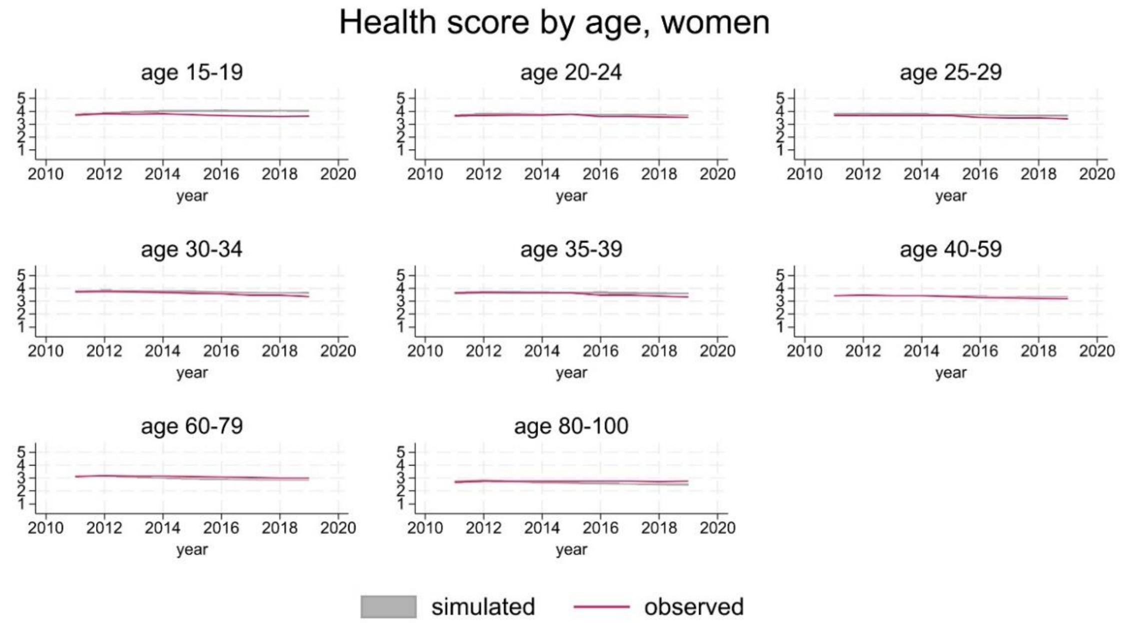


Figure 5: General health score, women



For ease of interpretation, we report caseness of psychological distress (see Section 3.3.3), in addition to the score (Figure 6). Distributions by age and gender are substantially in line with the observations, considering the volatility implied by the level of prevalence of psychological distress in the population (Figures 7 and 8).

Figure 6: Psychological distress

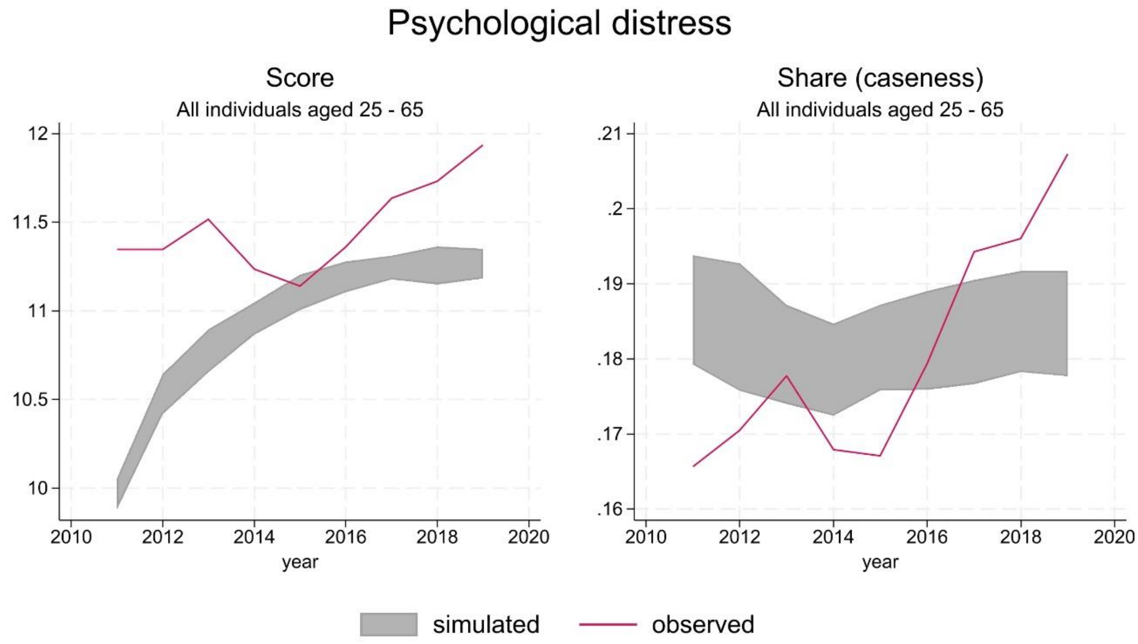
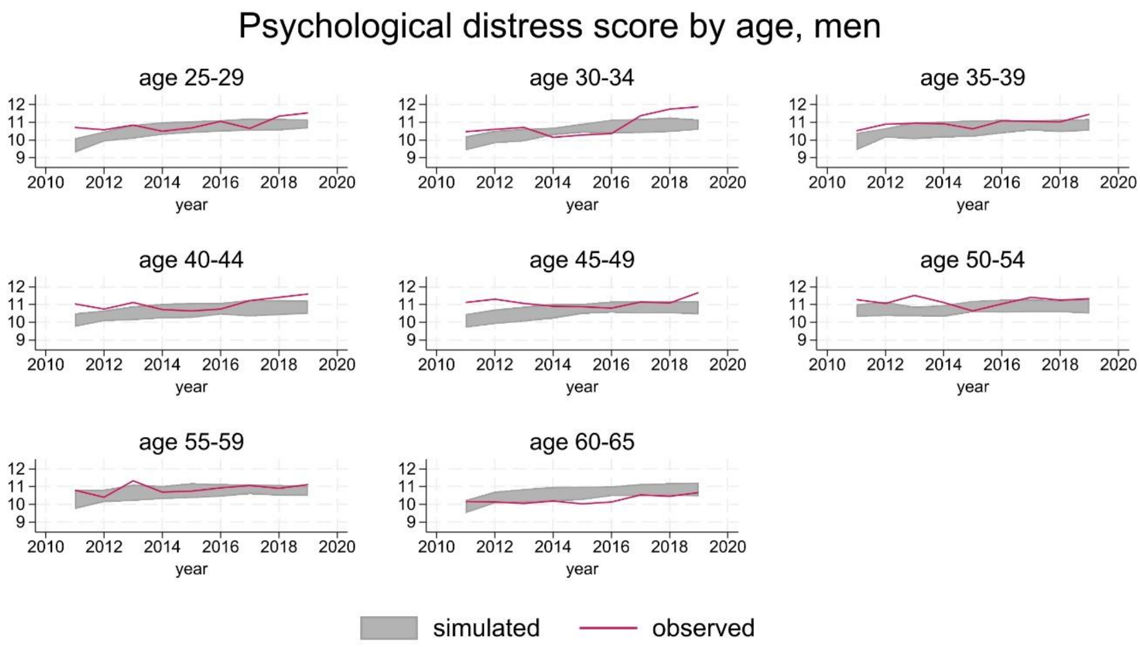
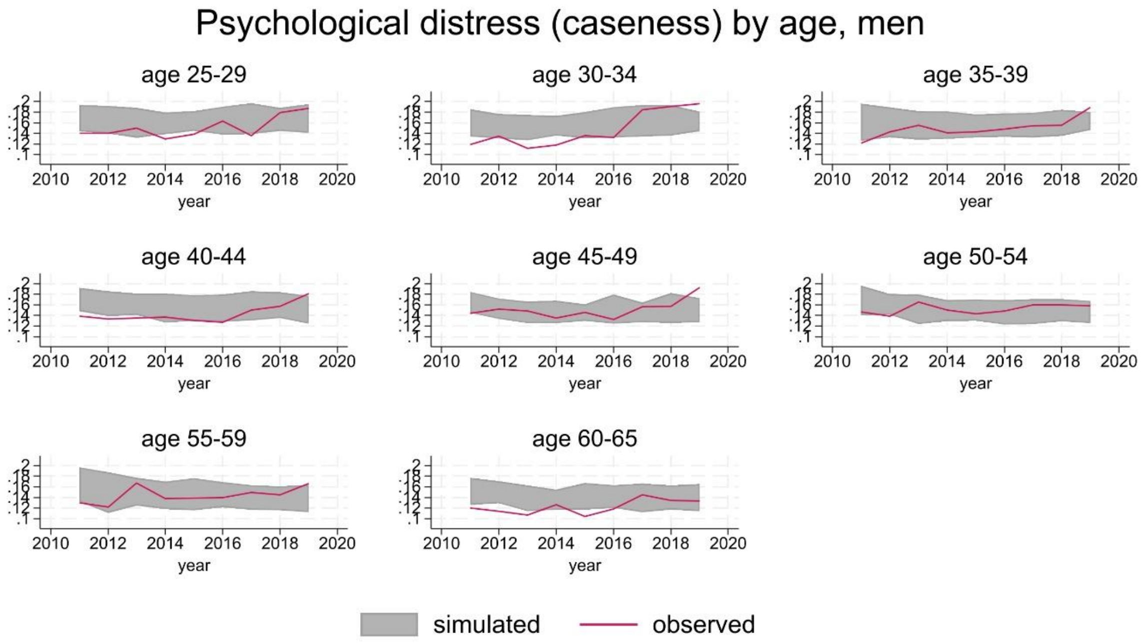


Figure 7: Psychological distress, men

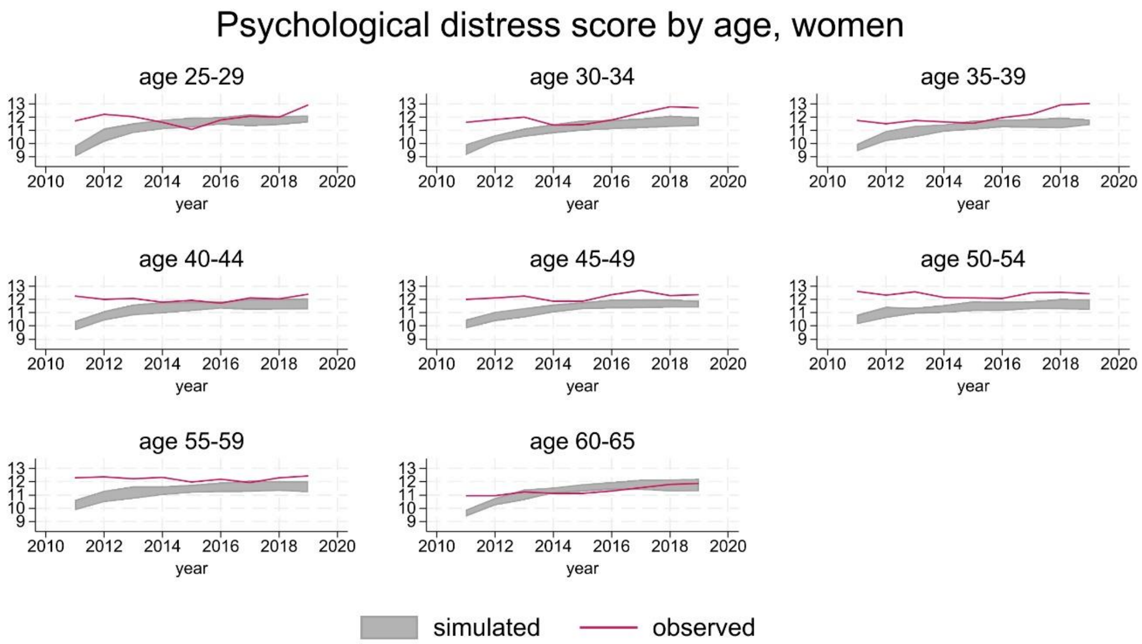


(a) Score

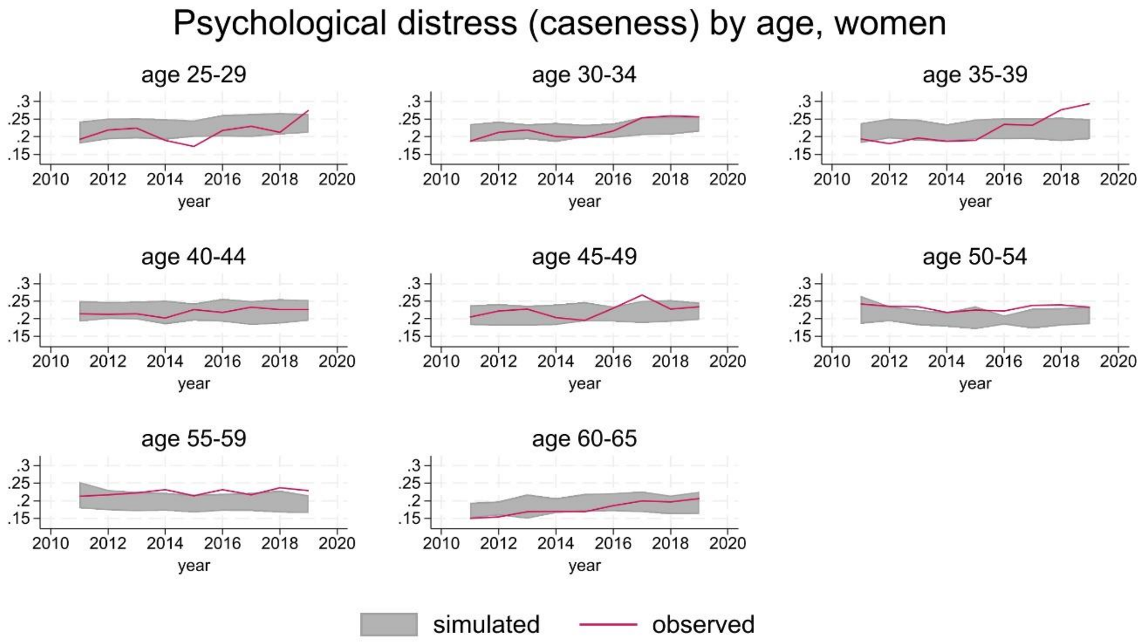


(b) Caseness

Figure 8: Psychological distress, women



(a) Score



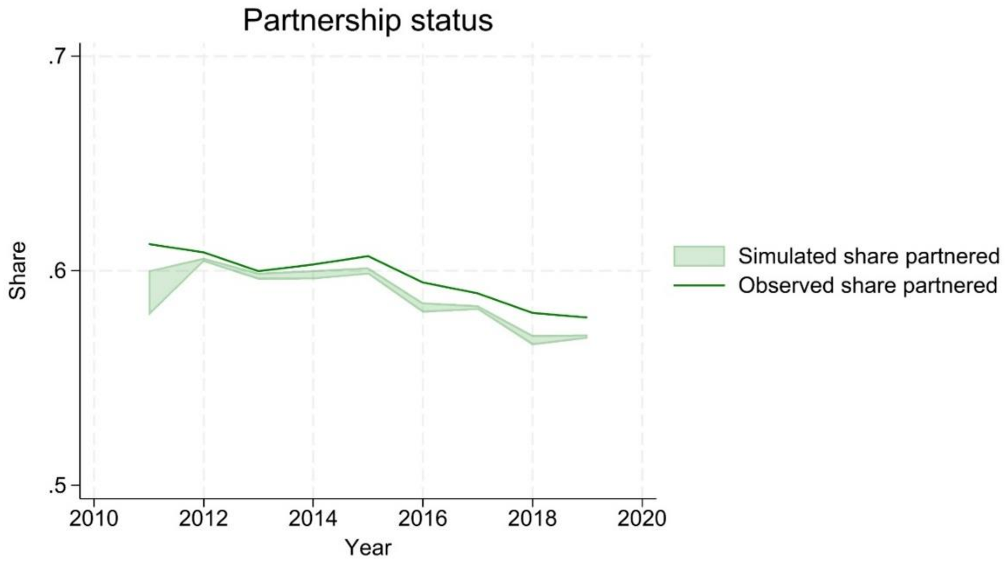
Psychological distress (caseness) by age, women
age 25-29
age 30-34
age 35-39
age 40-44
age 45-49
age 50-54
age 55-59
age 60-65
.3
.25
.2
.15
2010
2012
2014
2016
2018
2020
year
simulated
observed

(b) Caseness

4.2.3 Household structure

Projections correctly reproduce a declining share of partnered households (Figure 9), although the simulated series is slightly below the observed one.

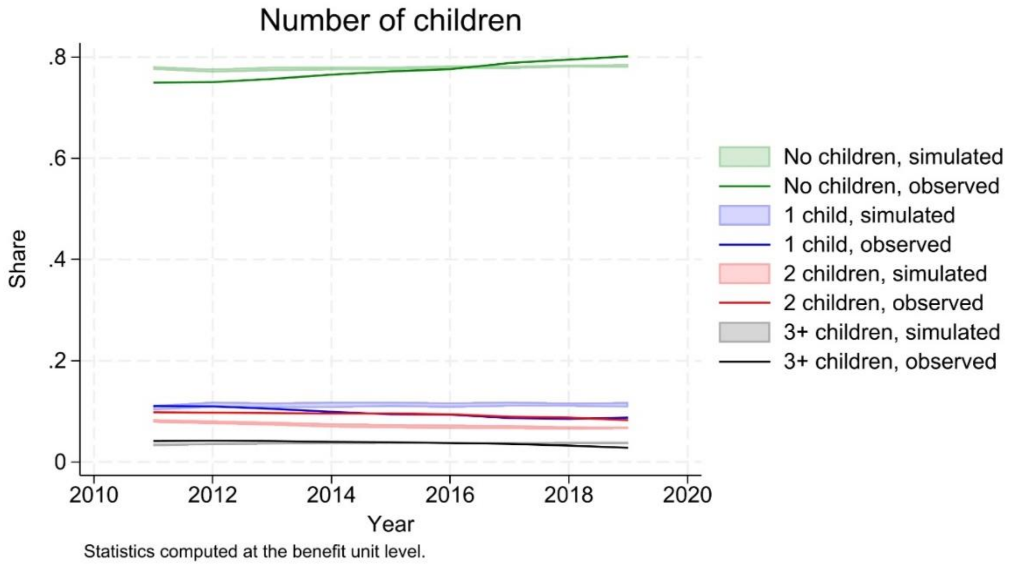
Figure 9: Partnership status



The simulations also reproduce, with some approximation, the distribution of benefit units by number of children (

Figure **10**).

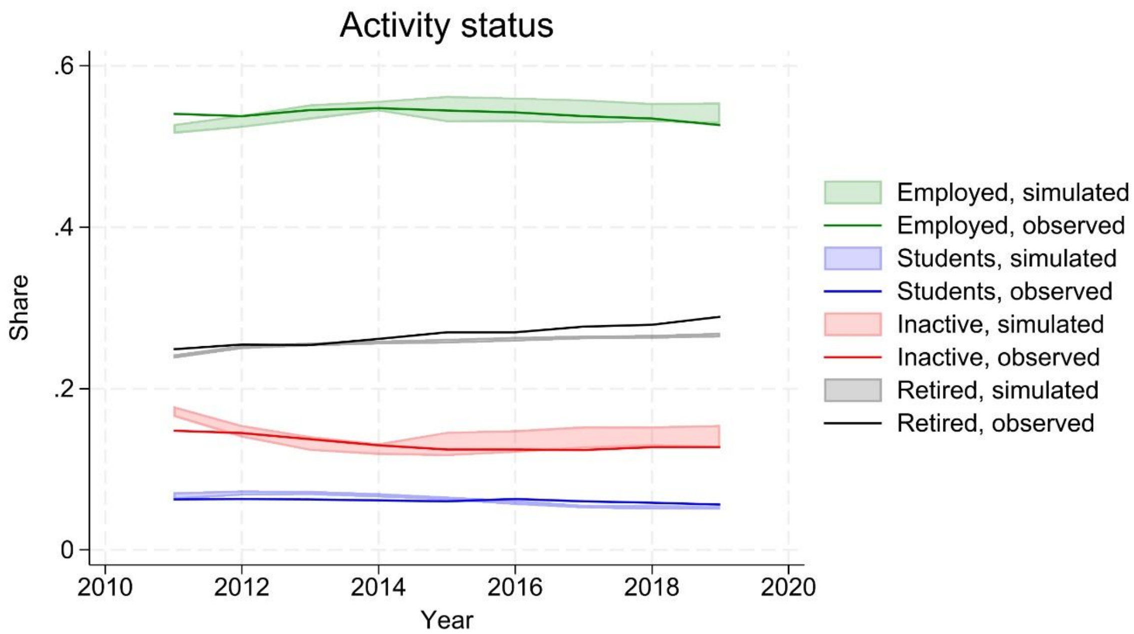
Figure 10: Number of children



4.2.4 Activity status, employment and wages

As discussed above (Section 3.8.2), we calibrate the labour supply model to align to observed employment rates. This is done by modifying the estimated parameters, rather than the simulated outcomes, resulting in a non-perfect hitting of the target. The other possible activity statuses on the other hand (in education, inactivity, retirement) are not aligned. Figure 11 shows that the simulated activity statuses broadly follow observed data, with a slight under-projection of pensioners.

Figure 11: Activity status



While projections are broadly aligned to aggregate employment figures, the distribution of employment by individual characteristics is freely determined by the model.

Figure 12 and

Figure **13** show that group-specific employment rates are substantially in line with the data, replicating the gender and age gradient and showing little trend over time. The main discrepancies are limited to young men (20-24 age group), where simulations over-predict employment rates.

Figure 12: Employment rates, men

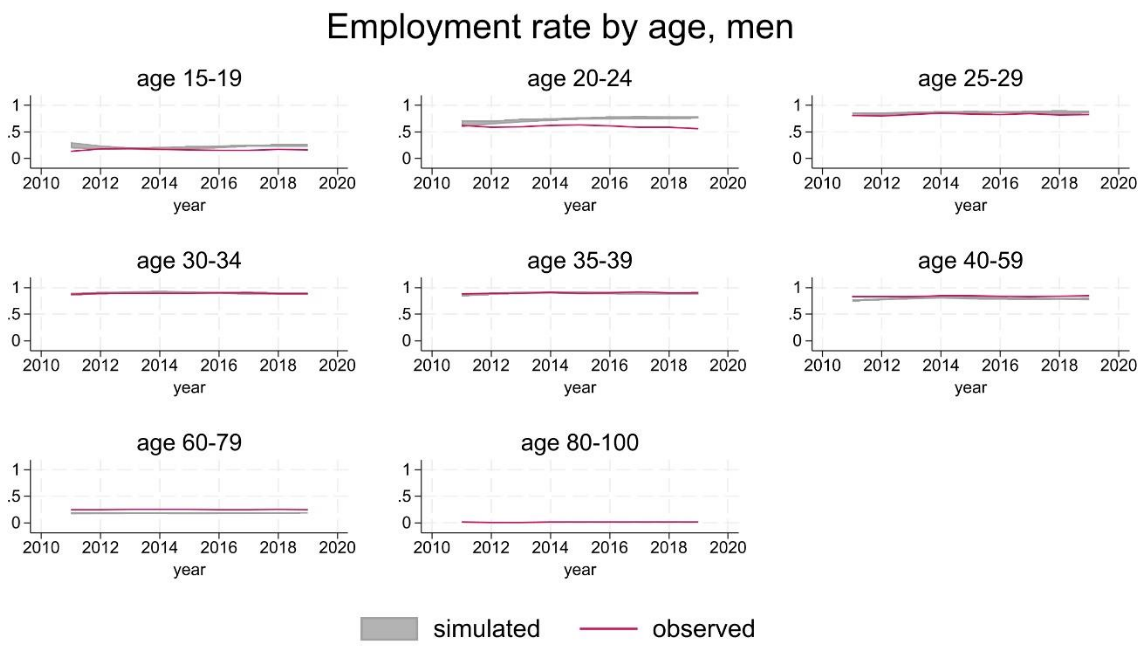
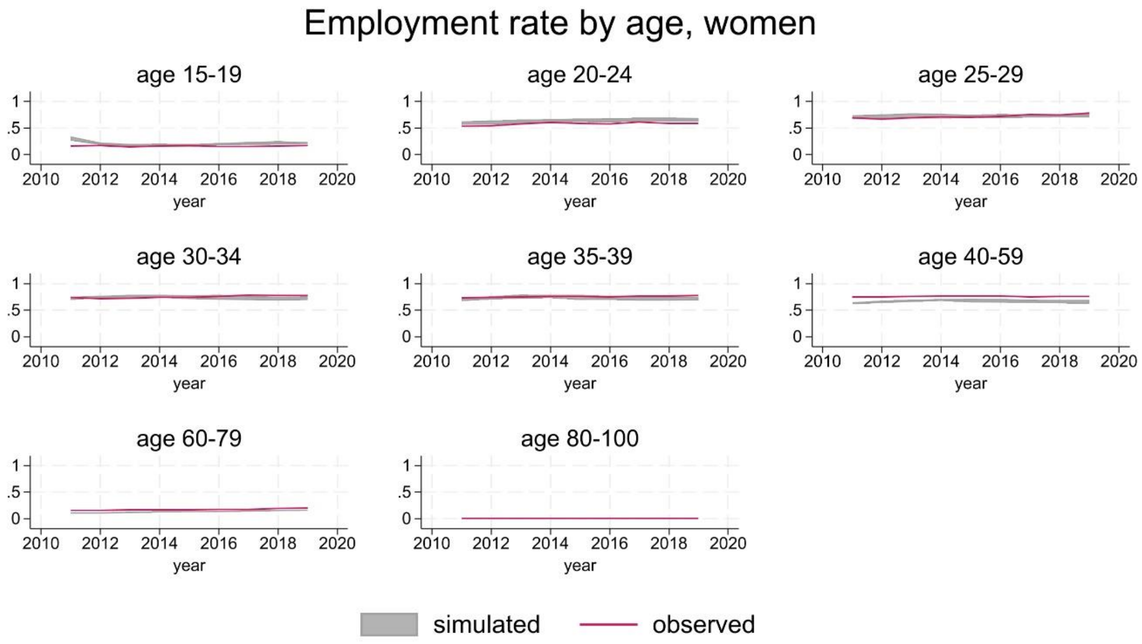


Figure 13: Employment rates, women



The growth in real wages is captured by the model to a good extent (Figure 14), while simulated and observed distributions are also largely overlapping (

Figure 15).

Figure 14: Real wages, trend

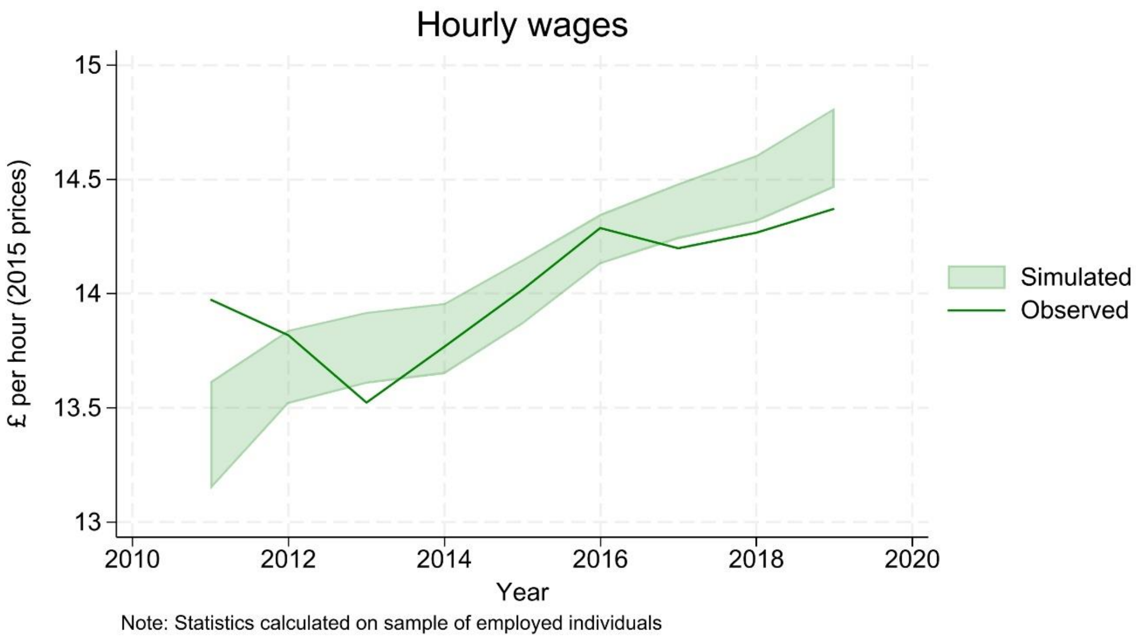
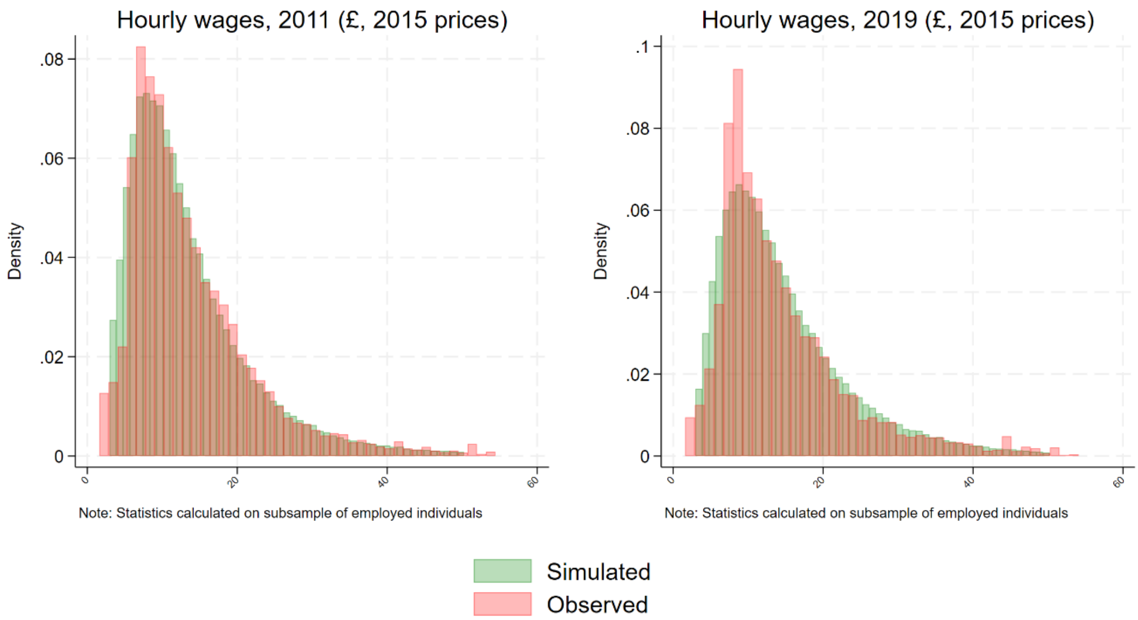
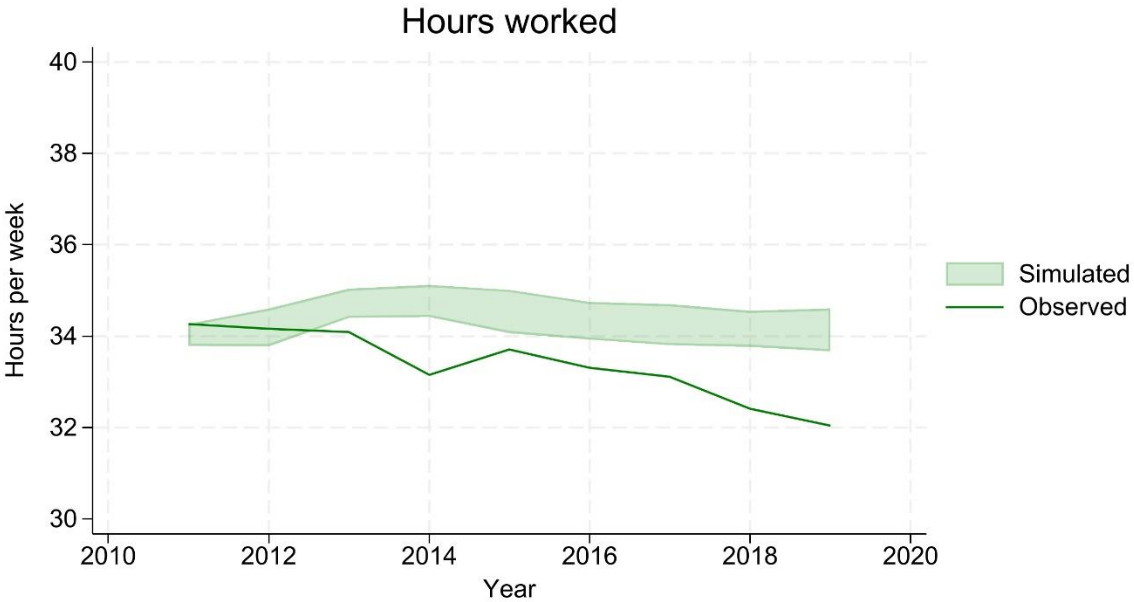


Figure 15: Real wages, distribution



Finally, the model struggles a bit in replicating the observed downward trend in hours worked (Figure 16). This is potentially due to the fact that the underlying random utility model of labour supply is estimated on one cross section of data only (2017). Sensitivity analysis shows that estimating the model on previous years results in broadly constant coefficients, which is consistent with the assumed structural nature of the model. However, the data seems to suggest that preferences might have indeed changed slightly over time.

Figure 16: Hours worked



Note: Statistics calculated on sample of employed individuals

4.2.5 Gross income

The model is able to replicate well both the trend and the distribution of individual gross income (

Figure **17** and Figure 18).

Figure 17: Gross income, trend

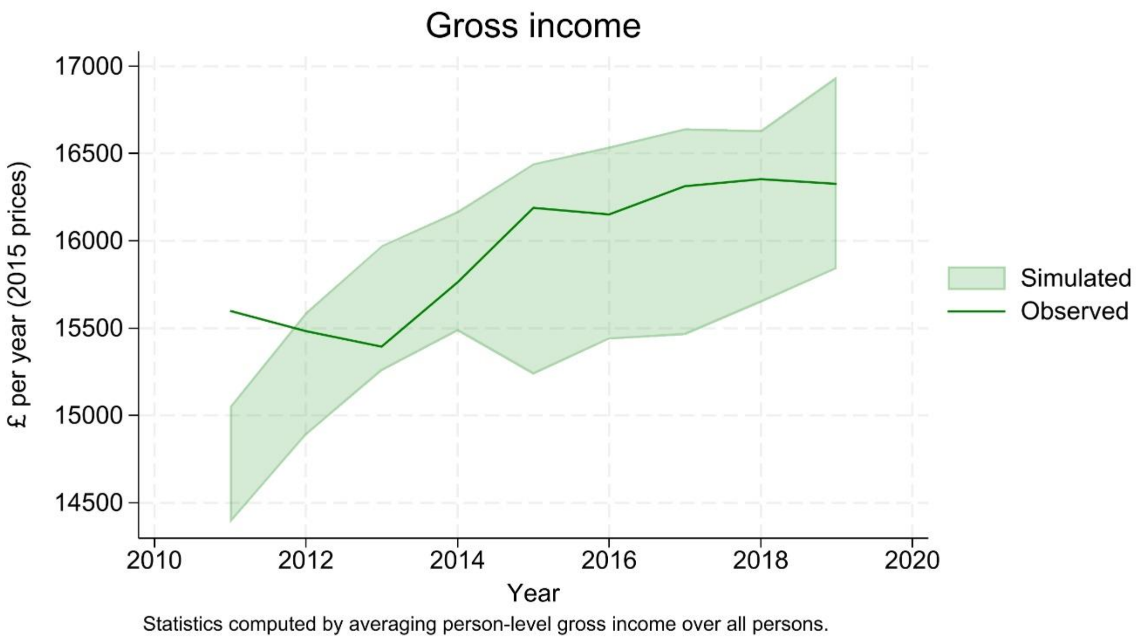
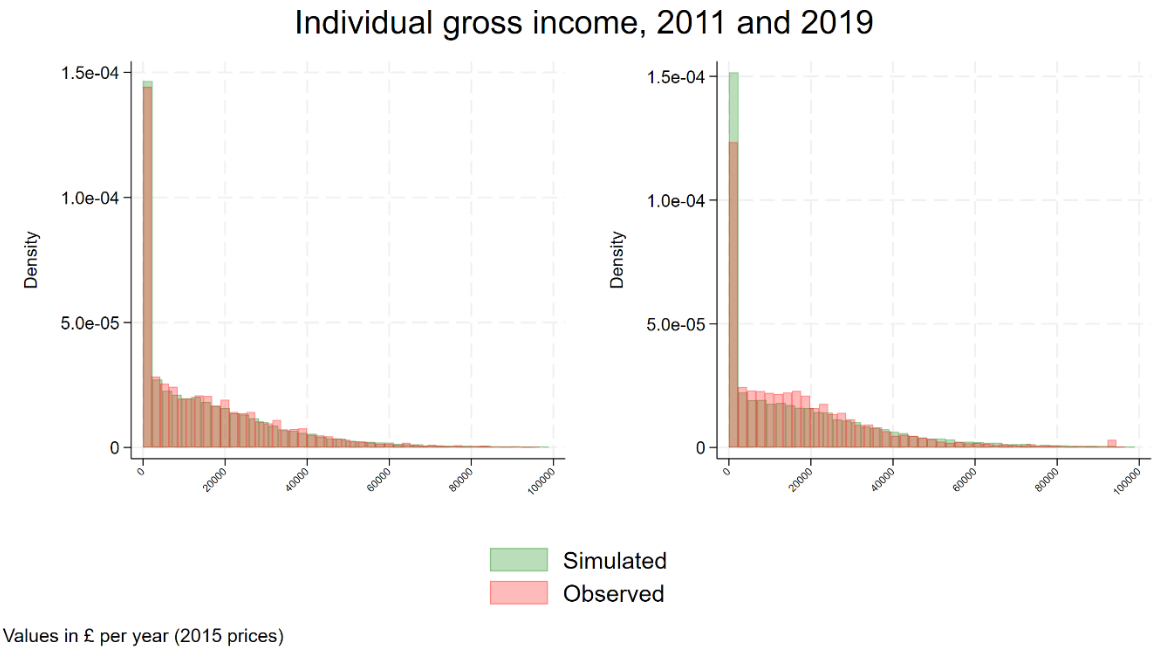


Figure 18: Gross income, distribution



Projected contributions of different income sources (labour, pension, capital and miscellaneous) by age groups along the income distribution also mimic the observed ones, both in levels (Figure 19) and in shares (

Figure **20**).

Figure 19: Income sources, value

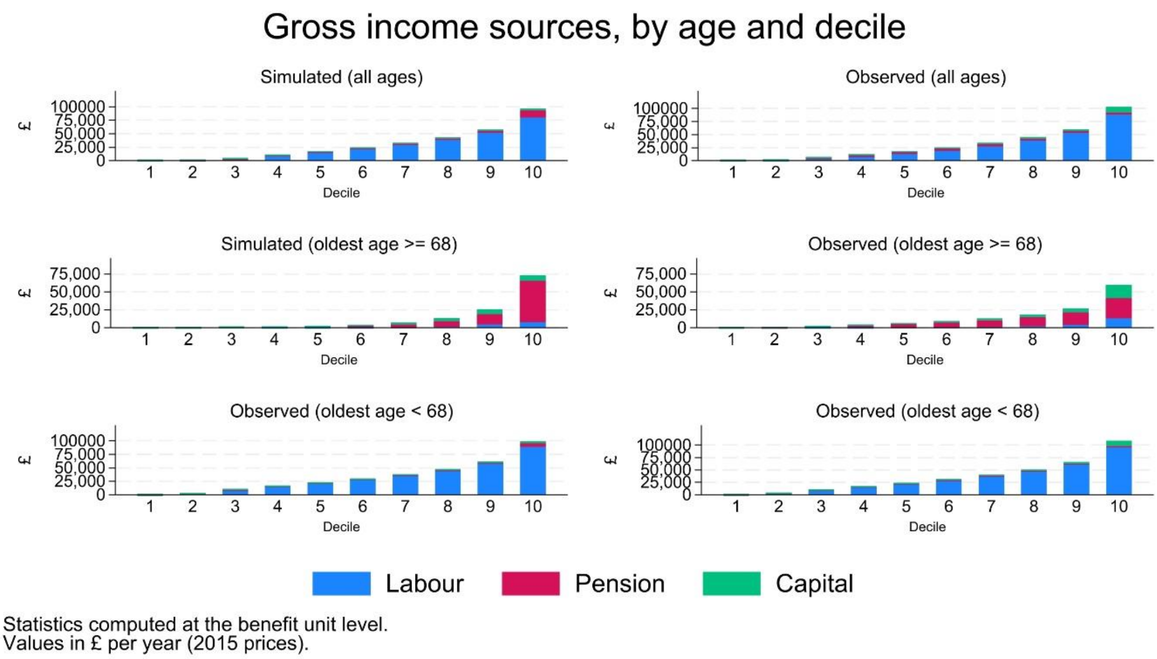
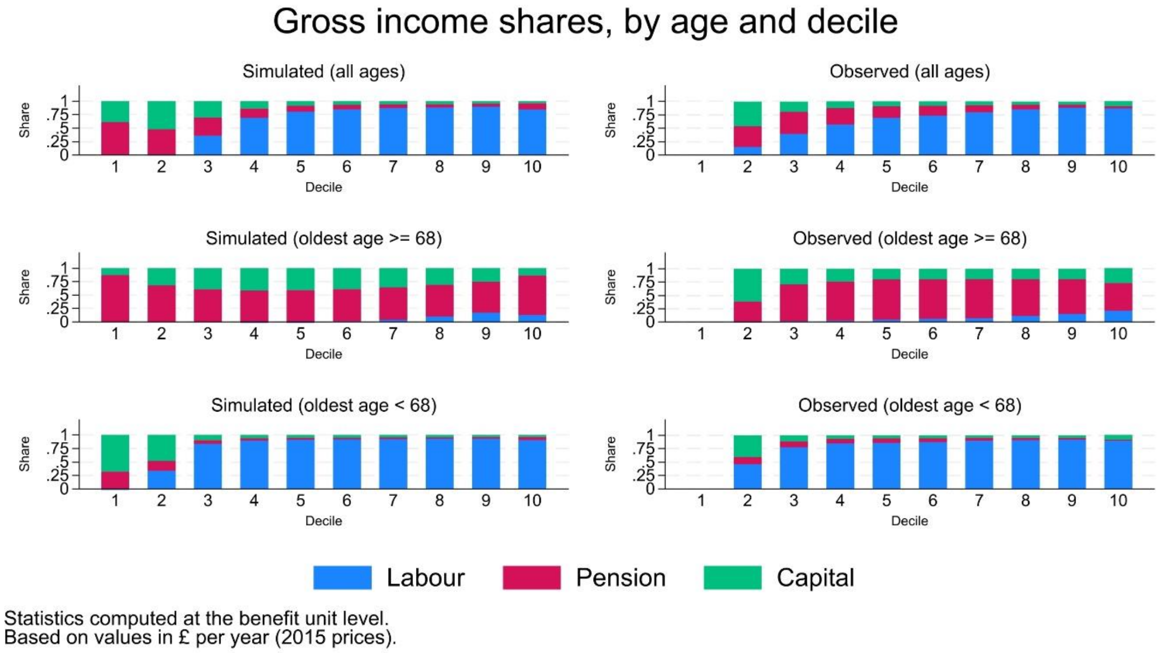


Figure 20: Income sources, share



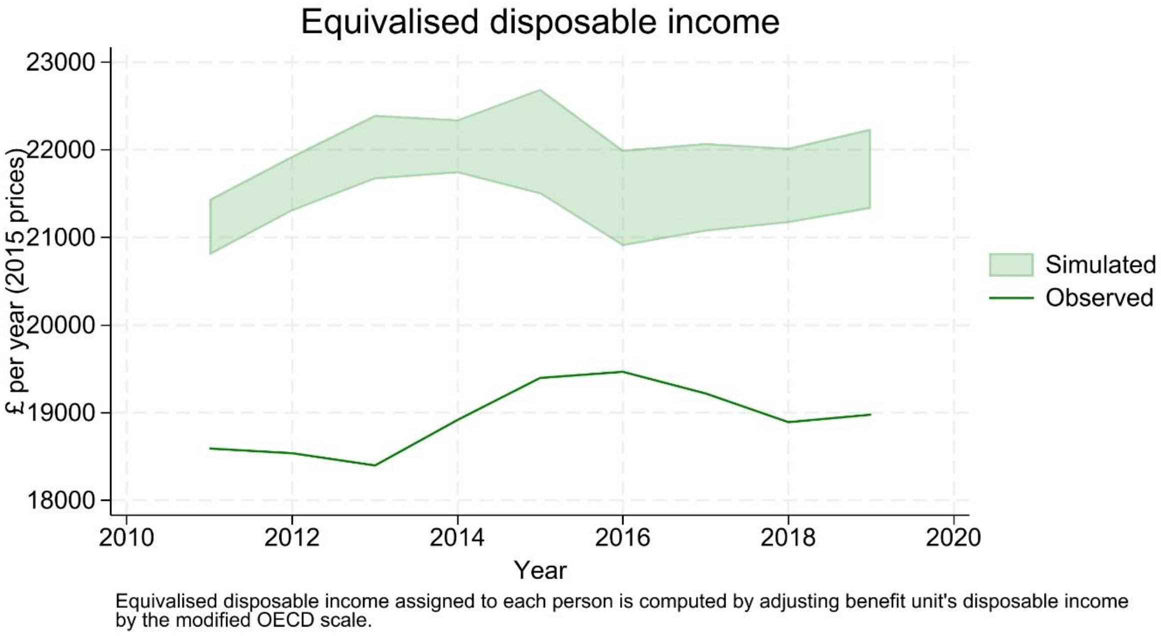
Labour income, computed by multiplying simulated hours worked by simulated wages, is obviously the main source of income for individuals below retirement age, while pension income is the main source for individuals above, on average. Both are projected with a fair level of accuracy (results not shown, but available on request). Capital income, on the other hand, is under-estimated in the simulations (average simulated values of around £1,200 - in 2015 prices - against observed values of around £1,700). However, the limited relevance of this source of income for the vast majority of the population – reflected in its small average value – limits the consequences of inadequate model specification.

4.2.6 Net income

Gross income is transformed into net income by means of the procedure described in Section 3.9. Results displayed in Figure 21 point to a slight over-estimation of disposable income (around 10%), possibly due to the fact that not all the characteristics relevant to the tax-benefit system can be simulated and controlled for in the matching procedure.[26]

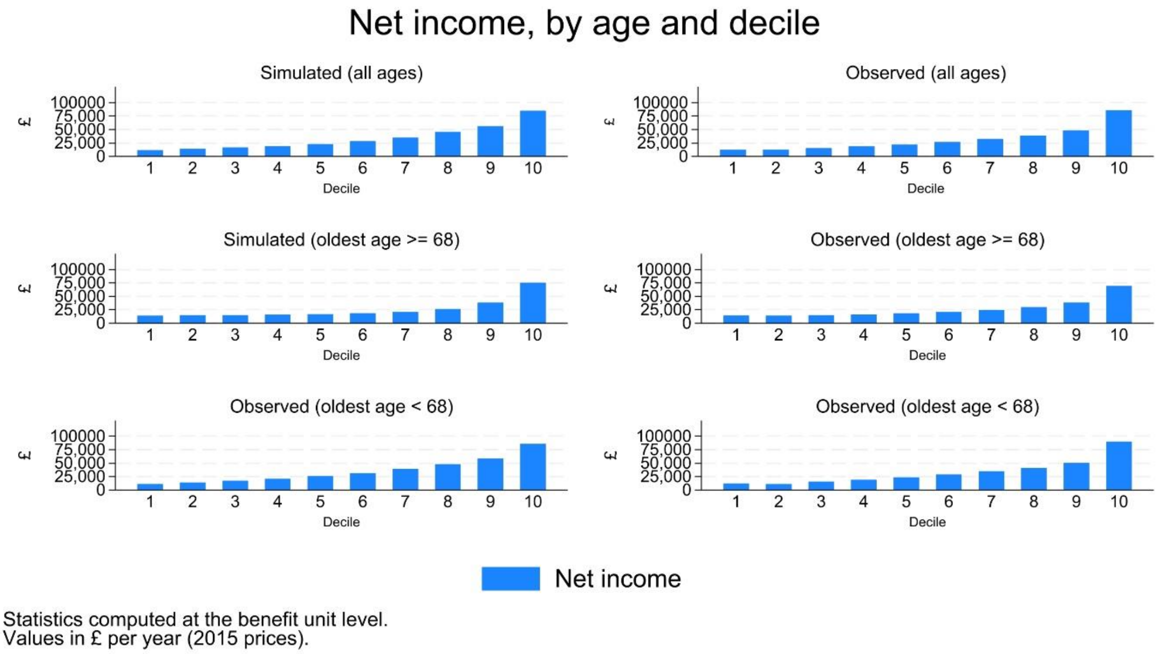
Figure 21: Equivalised disposable income

[26] Figures for disposable income, where disposable income is not adjusted for the benefit unit composition, follow a similar pattern.



The distribution of simulated disposable income however looks remarkably similar the observed one, both for the working age population, and for the population above retirement age (Figure 22).

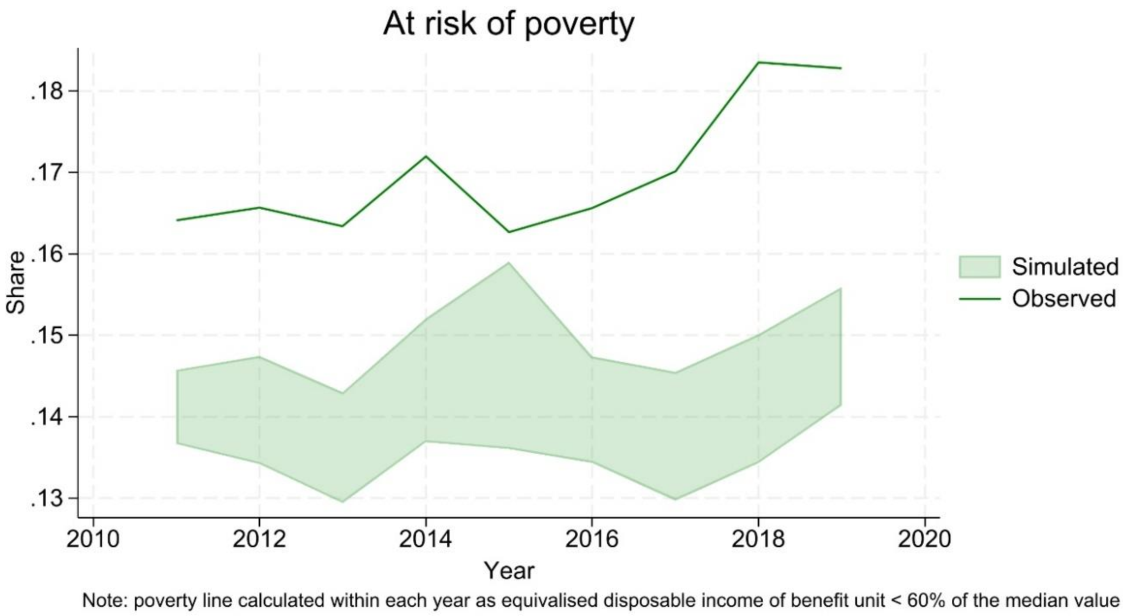
Figure 22: Disposable income, distribution



4.2.7 Poverty and inequality

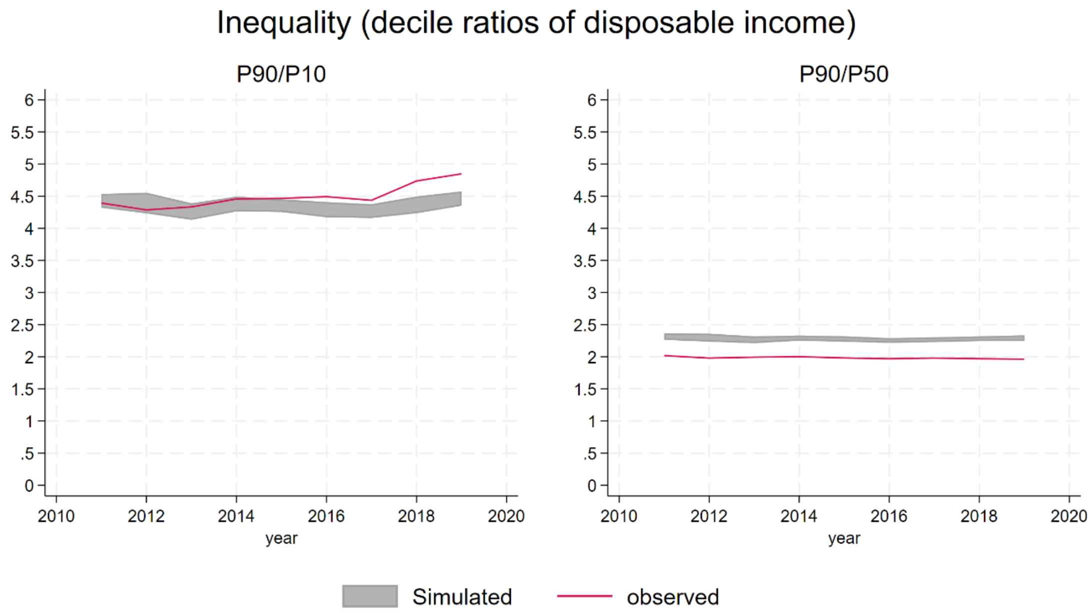
Biases in the simulation of disposable income translate into an under-estimation of poverty rates (Figure 23), although the error is small (around 2.5 percentage points), and the trends broadly comparable.

Figure 23: Poverty



Income inequality however, as measured by percentile ratios, is very much aligned with observed measures (Figure 24).

Figure 24: Inequality

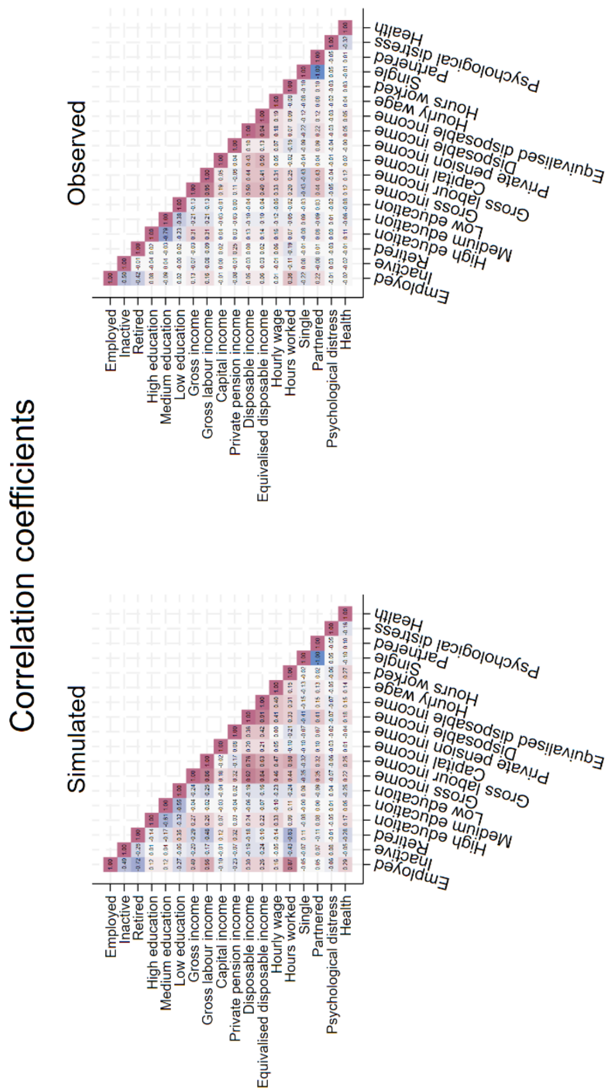


4.2.8 Correlations

Maintaining the cross-sectional perspective of the previous sections, we conclude with an assessment of pairwise correlations between the main outcome variables.

compares simulated and observed correlation coefficients. The main features of the data are reproduced by the model, from the most trivial (positive correlation between various income measures) to less straightforward ones (positive correlation between being partnered and labour income). Health is more positively correlated with income in the simulations than in the data, possibly because we force disabled people to drop off the labour market. The negative correlation between general health and psychological distress is faithfully reproduced, as well as the very tenuous negative correlation between psychological distress and income on the one hand, and psychological distress and being partnered on the other.

Figure 25: Correlations



5 Conclusion

In this paper we have introduced SimPaths, a novel dynamic microsimulation model of individual life course trajectories covering demographic, family, health and work-related events, with parameterisation to the United Kingdom. The innovative features of the model lie in the flexible integration with an external tax-benefit model, and in the structural behavioural modelling at the core of individual decision making over important margins of interest. Extensive validation over the period 2011-2019 shows that the model is able to replicate well the joint evolution of individual characteristics.

SimPaths is fully open source with increasingly detailed documentation, which sets it apart from most comparable models in the existing literature. The objective of the developmental team in this regard is to facilitate, and thereby encourage researcher entry into a field that we believe presents extensive opportunities to further understanding of a wide range of practically important phenomena. From theoretical descriptions of individual decision making, sustainability of personal and public finances, and the nature of burdens and opportunities to which people are and will be subject during the prospective half century, there is much of interest to explore.

References

Altonji, J.G., Martins, A.P., Siow, A. (2002), "Dynamic factor models of consumption, hours and income", *Research in Economics* 56(1): 3-59.

Andreassen, L., Fredriksen, D., Gjefsen, H.M., Halvorsen, E., Stølen, N.M. (2020), „The dynamic cross-sectional microsimulation model MOSART“, *International Journal of Microsimulation* 13(1): 92-113.

Barigozzi, M., Pellegrino, F. (2023), “Multidimensional Dynamic Factor Models”, paper available at https://arxiv.org/pdf/2301.12499.pdf

Bilcke, J., Beutels, P., Brisson, M., Jit, M. (2011), “Accounting for Methodological, Structural, and Parameter Uncertainty in Decision-Analytic Models: A Practical Guide”, *Medical Decision Making* 31(4): 675-692.

Blundell, R., Costa Dias, M., Meghir, C., and Shaw, J. (2016), “Female Labor Supply, Human Capital, and Welfare Reform” *Econometrica* 84: 1705–1753.

Blundell, R., Costa Dias, M., Goll, D., Meghir, C. (2021), “Wages, Experience, and Training of Women over the Life Cycle”, *Journal of Labor Economics* 39(S1): S275-S315.

Browning, M., Ejrnæs, M. (2009), “Consumption and children”, *Review of Economics and Statistics* 91: 93-111.

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

Burniaux J.M., Duval, R., Jaumotte, F. (2003), “Coping with ageing: a dynamic approach to quantify the impact of alternative policy options on future labour supply in OECD countries”, OECD Economic Department WP 371.

Campbell, D. (1979), “Assessing the impact of planned social change”, *Evaluation and Program Planning* 2: 67-90.

Carone G. (2005), “Long-term labour force projections for the 25 EU Member States: A set of data for assessing the economic impact of ageing”, DG ECFIN, European Economy, Economic Papers No. 235.

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

Cheng, C.-T. (2020), “Guy H. Orcutt’s Engineering Microsimulation to Reengineer Society”, *History of Political Economy* 52(S1): 191-217.

Christiano, L.J., Eichenbaum, M.S. and Trabandt, M. (2018), “On DSGE models”, *Journal of Economic Perspectives* 32: 113-140.

Conant, R.C., Ashby, W.R. (1970), “Every good regulator of a system must be a model of that system”, *International Journal of Systems Science* 1: 89-97.

Conti, R., Bavaro, M., Boscolo, S., Fabrizi, E., Puccioni, C., Tedeschi, S. (2024), “The Italian Treasury Dynamic Microsimulation Model (T-DYMM): Data, Structure and Validation”, *International Journal of Microsimulation*, forthcoming.

Creedy, J., Kalb, G., Kew, H. (2007), “Confidence intervals for policy reforms in behavioural tax microsimulation modelling”, *Bulletin of Economic Research* 59(1): 37-65.

De Menten, G., Dekkers, G., Bryon, G., Liégeois, and O‘Donoghue, C. (2014), “LIAM2: a new open source development tool for discrete-time dynamic microsimulation models”, *Journal of Artificial Societies and Social Simulation* 17, doi: 10.18564/jasss.2574.

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

EC - European Commission (2020a). “The 2021 Ageing Report. Economic & Budgetary Projections for the EU Member States (2019-2070)”, DG ECFIN, European Economy, Institutional Papers No 148.

EC - European Commission (2020b), “The 2021 Ageing Report. Underlying Assumptions and Projection Methodology”, DG ECFIN, European Economy, Institutional Papers No 142.

Favreault, M., Smith, K.E., and Johnson, R.W. (2015); “The Dynamic Simulation of Income Model (DYNASIM)”; Urban Institute research report; September.

Gallant, A.R., Tauchen, G.E. (1996), “Which moments to match?”, *Econometric Theory* 12: 657-681.

Gourieroux, C., Monfort, A., Renault, E. (1993), “Indirect inference”, *Journal of Applied Econometrics* 8: S85-S118.

Ghaith, Z., Kulshreshtha, S., Natcher, D. and Cameron, B.T. (2021), “Regional computable general equilibrium models: a review”, *Journal of Policy Modelling* 43: 710-724.

Gillman, M.S. (2017), “GENESIS - The GENEric SImulation System for modelling state transitions”, *Journal of Open Research Software* 5, doi: 10.5334/jors.179.

Goedemé T, Van den Bosch K, Salanauskaite L, Verbist G (2013), “Testing the Statistical Significance of Microsimulation Results: A Plea”, *International Journal of Microsimulation* 6(3): 50-77.

Goodhart, C. (1984), “Problems of monetary management: the UK experience”, in: Goodhart, C. (ed.), *Monetary Theory and Practice*, Macmillan, London.

Gourinchas, P., Parker, J. (2002), “Consumption over the life cycle”, *Econometrica* 70: 47–89.

Harding, A. (2023), “Challenges and Opportunities of Dynamic Microsimulation Modelling”, presentation at the 1st General Conference of the International Microsimulation Association, Vienna, *International Journal of Microsimulation*, 16(2): 1-13.

Kean, M.P., Todd, P.E., Wolpin, K.I. (2011), “The structural estimation of behavioral models: discrete choice dynamic programming methods and applications”, *Handbook of Labor Economics*, 4 Part A, ch. 4: 331-461.

Keys, R.G. (1981), “Cubic convolution interpolation for digital image processing”, *IEEE Transactions on Acoustics*, *Speech, and Signal Processing* 29: 1153-1160.

Klevmarken, N.A. (1997), “Behavioral Modeling in Micro Simulation Models. A Survey”, Working Paper 1997:31, Department of Economics, Uppsala University, republished as Klevmarken, N.A. (2022), “A Brief Survey of Behavioral Modeling in Micro Simulation Models”, *International Journal of Microsimulation* 15(1): 78-88.

Klevmarken, N.A., Lindgren, B. (2008), *Simulating an Ageing Population: a microsimulation approach applied to Sweden*, Emerald Publishing Limited, Bingley.

Kopasker, D., Bronka, P., Thomson, R. M., Khodygo, V., Kromydas, T., Meier, P., Heppenstall, A., Bambra, C., Lomax, N., Craig, P., Richiardi, M., & Katikireddi, S. V. (2023), “Evaluating the influence of taxation and social security policies on psychological distress: A microsimulation study of the UK”, University of Gasgow, mimeo.

Lee, B.S., Ingram, B. (1991), “Simulation estimation of time-series models”, Journal of Econometrics 47: 197-205.

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

Li, J., O’Donoghue, C., Dekkers, G. (2014), “Dynamic Models”, in: O’Donoghue, C. (ed.), *Handbook of Microsimulation Modelling*, Emerald Group Publishing, Bingley.

Lucas, R. J. (1976), “Econometric policy evaluation: a critique”, in: Brunner K., Meltzer, A. (eds.), *The Phillips Curve and Labor Markets*, North-Holland, Amsterdam.

Maitino, M.L., Ravagli, L., Sciclone, N. (2020), “IrpetDin. A Dynamic Microsimulation Model for Italy and the Region of Tuscany”, *International Journal of Microsimulation* 13(3): 27-53.

Mitton, L., Sutherland, H., Weeks, M. (2000), *Microsimulation Modelling for Policy Analysis. Challenges and Innovations*, Cambridge University Press, Cambridge.

National Research Council (2012), *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*, The National Academies Press, Washington, DC.

O'Donoghue, C. (2021), *Practical Microsimulation Modelling*, Oxford University Press, Oxford.

O'Donoghue, C., Dekkers, G. (2014), "Dynamic microsimulation", in: O'Donoghue, C. (ed.), *Handbook of Microsimulation Modelling*, Contributions to Economic Analysis 293: 305-343, Emerald Group Publishing Limited, Leeds.

O'Donoghue, C. and Dekkers, G. (2018), "Increasing the impact of dynamic microsimulation modelling", *International Journal of Microsimulation* 11(1): 61-96.

Orcutt, G.H., Caldwell, S., and Wertheimer, R. (1976), "Policy Exploration through Microanalytic Simulation", Urban Institute, Washington DC.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. (2007), *Numerical Recipes: the art of scientific computing*, Cambridge University Press, Cambridge.

Pylkkänen, E. (2002), "Modelling wages and hours of work", in: Pylkkänen, E. (Ed.), *Studies on Household Labor Supply and Home Production,* Goteborgs Universitet, Gothenburg.

Richiardi, M. (2014), "The missing link: AB models and dynamic microsimulation", in: Leitner S., Wall, F. (eds), *Artificial Economics and Self Organization*, Springer, Lecture Notes in Economics and Mathematical Systems, vol. 669, Berlin.

Richiardi, M., Collado, D., Popova, D. (2021), "UKMOD – A new tax-benefit model for the four nations of the UK", *International Journal of Microsimulation* 14(1): 92-101.

Richiardi, M. and He, Z. (2021), "No one left behind: The labour supply behaviour of the entire Italian population", Centre for Microsimulation and Policy Analysis, mimeo.

Richiardi, M., Richardson R.E. (2017), "JAS-mine: a new platform for microsimulation and agent-based modelling", *International Journal of Microsimulation* 10, pp. 106-134.

Rust, J. (2008), "Dynamic programming". In: Durlauf, S.N., Blume, L.E. (eds.), *New Palgrave Dictionary of Economics*. Palgrave Macmillan, New York.

Scherer P. (2002), "Age of withdrawal from the labour force in the OECD countries", Labour Market and Social policy Occasional Papers, No.49, DELSA.

Schofield, D.J., Zeppel, M.J.B., Tan, O., Lymer, S., Cunich, M.M., Shrestha, R.N. (2018), "A brief, global history of microsimulation models in health: Past applications, lessons learned and future directions", *International Journal of Microsimulation* 11(1): 97-142.

Schofield, D.J. (2023), "Editorial. Special Issue in honour of Vale Emerita Professor Ann Harding AO", *International Journal of Microsimulation* 16(2): i-iii.

Spielauer, M., Horvath, T. and Fink, M. (2020), "microWELT: A dynamic microsimulation model for the study of welfare transfer flows in ageing societies from a comparative welfare state perspective", WIFO Working Papers 609/2020.

Skarda, I., Asaria, M., Cookson, R. (2021), "LifeSim: A Lifecourse Dynamic Microsimulation Model of the Millennium Birth Cohort in England", *International Journal of Microsimulation* 14(1): 2-42.

Štefánik, M., Miklošovič, T. (2020), "Modelling foreign labour inflows using a dynamic microsimulation model of an ageing country – Slovakia", *International Journal of Microsimulation* 13(2): 102-113.

Stern, S. (1997), "Approximate solutions to stochastic dynamic programs", *Econometric Theory 13*: 392-405.

Sutherland, H., Figari F. (2013), "EUROMOD: The European Union tax-benefit microsimulation model", *International Journal of Microsimulation* 6(1): 4-26.

van de Ven, J. (2011), “A structural dynamic microsimulation model of household savings and labour supply”, *Economic Modelling* 28: 2054-2070.
van de Ven, J. (2017), “SIDD: An adaptable framework for analysing the distributional implications of policy alternatives where savings and employment decisions matter”, *Economic Modelling* 63: 161-174.
van de Ven J., Bronka P., Richiardi M. (2022), “Dynamic simulation of taxes and welfare benefits by database imputation”, CeMPA Working Paper 3/22.

Appendix A Estimates

The estimates for the utility functions used in the labour supply model are separately described in Richiardi and He (2021).

The estimates for the psychological distress models are separately described in Kopasker et al. (2023).

Estimation sample is UKHLS waves “a”-“h”, unless differently specified.

Table A.1: Process E1a: Probability of being in education.
Sample: Individuals aged 16-29 in continuous education.

Probit In education	(1) Coef.	(2) s.e.
Gender = 1, Male	-0.02	0.03
Age	-1.23***	0.07
Age Squared	0.03***	0.00
Mother's Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.10**	0.04
Mother's Educational Attainment: 3 Category = 3, Other/No Qualification	-0.12**	0.06
Father's Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.30***	0.04
Father's Educational Attainment: 3 Category = 3, Other/No Qualification	-0.30***	0.05
Region = 1, North East	-0.02	0.09
Region = 2, North West	-0.26***	0.07
Region = 3, Yorkshire and the Humber	-0.13*	0.07
Region = 4, East Midlands	-0.22***	0.07
Region = 5, West Midlands	-0.13*	0.07
Region = 6, East of England	-0.09	0.07
Region = 8, South East	-0.21***	0.06
Region = 9, South West	-0.24***	0.07
Region = 10, Wales	-0.20**	0.09
Region = 11, Scotland	-0.28***	0.08
Region = 12, Northern Ireland	-0.07	0.08
Year	-0.12***	0.01
Constant	16.65***	0.74
Observations	11,841	
R2	0.134	
Chi2	1160	
Log-likelihood	-5413	

*** p<0.01, ** p<0.05, * p<0.1

Table A.2: Process E1b: Probability of being in education.
Sample: Individuals aged 16-35 not in continuous education.

Probit In education	(1) Coef.	(2) s.e.
Gender = 1, Male	-0.12***	0.04
Age	-0.36***	0.02
Age Squared	0.00***	0.00
Lagged Educational Attainment: 3 Category = 2 Other Higher/A-level/GCSE,	0.08*	0.05
Lagged Educational Attainment: 3 Category = 3 Other/No Qualification,	-0.38***	0.10
Lagged Employment Status: 3 Category = Student,	1.78***	0.05
Lagged Employment Status: 3 Category = Not employed,	0.26***	0.04
Lagged Number of Children in Household,	0.08***	0.02
Lagged Number of Children aged 0-2 in Household,	-0.23***	0.06
Mother's Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.20***	0.05
Mother's Educational Attainment: 3 Category = 3, Other/No Qualification	-0.24***	0.06
Father's Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.15***	0.05
Father's Educational Attainment: 3 Category = 3, Other/No Qualification	-0.27***	0.06
Region = 1, North East	0.05	0.09
Region = 2, North West	-0.15**	0.07
Region = 3, Yorkshire and the Humber	-0.18**	0.08
Region = 4, East Midlands	-0.30***	0.08
Region = 5, West Midlands	-0.15**	0.07
Region = 6, East of England	-0.24***	0.07
Region = 8, South East	-0.20***	0.06
Region = 9, South West	-0.41***	0.08
Region = 10, Wales	-0.06	0.09
Region = 11, Scotland	0.08	0.07
Region = 12, Northern Ireland	-0.17*	0.10
Year	-0.01	0.01
Constant	4.60***	0.32
Observations	51,525	
R2	0.393	
Chi2	3151	
Log-likelihood	-5278	

*** p<0.01, ** p<0.05, * p<0.1

Table A.3: Process E2: Educational attainment.
Sample: Respondents from Process 1a who have left education.

Ordered probit Educational attainment: High, Medium, Low	(1) Coef.	(2) s.e.
Gender = 1, Male	-0.03	0.03
Age	1.43***	0.10
Age Squared	-0.03***	0.00
Mother's Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.13***	0.04
Mother's Educational Attainment: 3 Category = 3, Other/No Qualification	-0.26***	0.06
Father's Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.04	0.04
Father's Educational Attainment: 3 Category = 3, Other/No Qualification	-0.15***	0.06
Region = 1, North East	0.04	0.10
Region = 2, North West	0.12*	0.07
Region = 3, Yorkshire and the Humber	0.02	0.07
Region = 4, East Midlands	0.08	0.07
Region = 5, West Midlands	-0.02	0.08
Region = 6, East of England	0.20***	0.07
Region = 8, South East	0.17***	0.06
Region = 9, South West	0.20***	0.07
Region = 10, Wales	-0.09	0.09
Region = 11, Scotland	0.09	0.09
Region = 12, Northern Ireland	0.04	0.09
Year	0.02**	0.01
/cut1	15.70***	1.05
/cut2	19.72***	1.10
Observations	11,431	
R2	0.304	
Chi2	1076	
Log-likelihood	-4663	

*** p<0.01, ** p<0.05, * p<0.1

Table A.4: Process H1a: Self-rated Health Status.
Sample: Individuals aged 16-29 in continuous education.

Ordered probit Self-rated health status, categories 1 to 5	(1) Coef.	(2) s.e.
Gender = 1, Male	0.15***	0.04
Age	0.27**	0.11
Age Squared	-0.01**	0.00
Lagged Annual Household Income Quintile = 2,	-0.04	0.05
Lagged Annual Household Income Quintile = 3,	-0.06	0.07
Lagged Annual Household Income Quintile = 4,	0.13	0.11
Lagged Annual Household Income Quintile = 5,	-0.01	0.16
Lagged Self-rated Health = 1,	-2.62***	0.32
Lagged Self-rated Health = 2,	-2.17***	0.10
Lagged Self-rated Health = 3,	-1.58***	0.06
Lagged Self-rated Health = 4,	-0.83***	0.05
Region = 1, North East	-0.00	0.10
Region = 2, North West	0.11	0.08
Region = 3, Yorkshire and the Humber	0.02	0.08
Region = 4, East Midlands	-0.01	0.08
Region = 5, West Midlands	0.18**	0.08
Region = 6, East of England	-0.06	0.08
Region = 8, South East	0.06	0.07
Region = 9, South West	0.05	0.08
Region = 10, Wales	0.14	0.10
Region = 11, Scotland	0.07	0.09
Region = 12, Northern Ireland	0.22**	0.10
Year	-0.01	0.01
/cut1	-1.02	1.12
/cut2	0.12	1.11
/cut3	1.33	1.11
/cut4	2.71**	1.11
Observations	4,549	
R2	0.135	
Chi2	996.2	
Log-likelihood	-4747	

*** p<0.01, ** p<0.05, * p<0.1

Table A.5: Process H1b: Self-rated Health Status.
Sample: Individuals aged 16 or older not in continuous education.

Ordered probit Self-rated health status, categories 1 to 5	(1) Coef.	(2) s.e.
Gender = 1, Male	-0.00	0.01
Age	-0.01***	0.00
Age Squared	0.00***	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.10***	0.01
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.18***	0.01
Lagged Employment Status: 3 Category = Student,	0.10***	0.03
Lagged Employment Status: 3 Category = Not employed,	-0.11***	0.01
Lagged Annual Household Income Quintile = 2,	0.11***	0.02
Lagged Annual Household Income Quintile = 3,	0.16***	0.02
Lagged Annual Household Income Quintile = 4,	0.21***	0.02
Lagged Annual Household Income Quintile = 5,	0.32***	0.02
Lagged Self-rated Health = 1,	-3.80***	0.03
Lagged Self-rated Health = 2,	-2.61***	0.02
Lagged Self-rated Health = 3,	-1.70***	0.01
Lagged Self-rated Health = 4,	-0.87***	0.01
Lagged Household Type: 4 Category = Couple with children,	0.04***	0.01
Lagged Household Type: 4 Category = Single without children,	-0.02*	0.01
Lagged Household Type: 4 Category = Single with children,	0.04*	0.02
Region = 1, North East	-0.01	0.02
Region = 2, North West	-0.01	0.02
Region = 3, Yorkshire and the Humber	0.02	0.02
Region = 4, East Midlands	0.03	0.02
Region = 5, West Midlands	0.01	0.02
Region = 6, East of England	0.00	0.02
Region = 8, South East	0.01	0.02
Region = 9, South West	0.02	0.02
Region = 10, Wales	0.03	0.02
Region = 11, Scotland	0.01	0.02
Region = 12, Northern Ireland	0.02	0.02
Year	-0.01***	0.00
/cut1	-4.40***	0.05
/cut2	-3.13***	0.05
/cut3	-1.86***	0.05
/cut4	-0.44***	0.05
Observations	118,011	
R2	0.237	
Chi2	34000	
Log-likelihood	-132736	

*** p<0.01, ** p<0.05, * p<0.1

Table A.6: Process H2b: Probability of becoming long-term sick or disabled.

Sample: Individuals aged 16 or older not in continuous education.

Probit Probability of long-term sickness or disability	(1) Coef.	(2) s.e.
Self-rated Health	-0.48***	0.02
Gender = 1, Male	0.03	0.03
Age	0.11***	0.01
Age Squared	-0.00***	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	0.12***	0.04
Educational Attainment: 3 Category = 3, Other/No Qualification	0.22***	0.05
Lagged Annual Household Income Quintile = 2,	-0.50***	0.04
Lagged Annual Household Income Quintile = 3,	-0.58***	0.04
Lagged Annual Household Income Quintile = 4,	-0.69***	0.05
Lagged Annual Household Income Quintile = 5,	-0.86***	0.05
Lagged Self-rated Health,	-0.10***	0.02
Lagged Long-term Sick or Disabled,	1.87***	0.04
Lagged Household Type: 4 Category = Couple with children,	-0.18***	0.04
Lagged Household Type: 4 Category = Single without children,	0.16***	0.04
Lagged Household Type: 4 Category = Single with children,	-0.14**	0.06
Region = 1, North East	0.16**	0.08
Region = 2, North West	0.13*	0.07
Region = 3, Yorkshire and the Humber	0.13*	0.07
Region = 4, East Midlands	-0.03	0.08
Region = 5, West Midlands	-0.07	0.07
Region = 6, East of England	0.04	0.07
Region = 8, South East	0.05	0.07
Region = 9, South West	0.11	0.07
Region = 10, Wales	0.11	0.08
Region = 11, Scotland	0.15**	0.07
Region = 12, Northern Ireland	0.18**	0.08
Year	-0.02**	0.01
Constant	-2.55***	0.23
Observations	118,011	
R2	0.635	
Chi2	7485	
Log-likelihood	-6269	

*** p<0.01, ** p<0.05, * p<0.1

Table A.7: Process P1a: Probability of leaving the parental home.
Sample: All non-student respondents living with a parent.

Probit Probability of leaving parental home	(1) Coef.	(2) s.e.
Gender = 1, Male	0.16***	0.02
Age	-0.08***	0.00
Age Squared	0.00***	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.07***	0.03
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.24***	0.04
Lagged Employment Status: 4 Category = Student,	-0.27***	0.05
Lagged Employment Status: 4 Category = Sick / disabled,	-0.05	0.07
Lagged Employment Status: 4 Category = Not employed,	-0.17***	0.04
Lagged Annual Household Income Quintile = 2,	-0.12***	0.04
Lagged Annual Household Income Quintile = 3,	-0.17***	0.04
Lagged Annual Household Income Quintile = 4,	-0.28***	0.04
Lagged Annual Household Income Quintile = 5,	-0.49***	0.05
Region = 1, North East	0.11*	0.06
Region = 2, North West	-0.04	0.05
Region = 3, Yorkshire and the Humber	0.01	0.05
Region = 4, East Midlands	0.03	0.05
Region = 5, West Midlands	-0.03	0.05
Region = 6, East of England	0.01	0.05
Region = 8, South East	0.08*	0.05
Region = 9, South West	0.01	0.05
Region = 10, Wales	-0.01	0.07
Region = 11, Scotland	-0.15***	0.06
Region = 12, Northern Ireland	-0.04	0.08
Year	-0.01	0.01
Constant	0.72***	0.12
Observations	117,942	
R2	0.182	
Chi2	3090	
Log-likelihood	-14292	

*** p<0.01, ** p<0.05, * p<0.1

Table A8: Process U1a: Probability of entering a partnership.
Sample: All single respondents aged 16 and older, in continuous education.

Probit Probability of entering partnership	(1) Coef.	(2) s.e.
Gender = 1, Male	0.04	0.15
Age	0.71*	0.40
Age Squared	-0.01	0.01
Lagged Annual Household Income Quintile = 2,	-0.41*	0.21
Lagged Annual Household Income Quintile = 3,	-0.19	0.27
Lagged Annual Household Income Quintile = 4,	-0.43	0.37
Lagged Annual Household Income Quintile = 5,	0.00	0.00
Lagged Number of Children in Household,	0.16	0.24
Lagged Number of Children aged 0-2 in Household,	0.15	0.38
Lagged Self-rated Health,	0.00	0.09
Region = 1, omitted	-	-
Region = 2, North West	0.53	0.41
Region = 3, Yorkshire and the Humber	0.30	0.44
Region = 4, East Midlands	0.30	0.41
Region = 5, West Midlands	-0.19	0.50
Region = 6, East of England	-0.18	0.49
Region = 8, South East	0.08	0.47
Region = 9, South West	-0.60	0.54
Region = 10, omitted	-	-
Region = 11, Scotland	0.17	0.43
Region = 12, omitted	-	-
Year	-0.20***	0.05
Constant	-8.29**	4.12
Observations	1,759	
R2	0.181	
Chi2	85.50	
Log-likelihood	-210.6	

*** p<0.01, ** p<0.05, * p<0.1

Table A.9: Process U1b: Probability of entering a partnership.
Sample: All single respondents aged 16 and older, not in continuous education.

Probit Probability of entering a partnership	(1) Coef.	(2) s.e.
Gender = 1, Male	-0.06**	0.03
Age	-0.01*	0.01
Age Squared	-0.00**	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.04	0.03
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.07	0.05
Lagged Employment Status: 5 Category = Student,	-0.37***	0.08
Lagged Employment Status: 5 Category = Not employed,	-0.15***	0.04
Lagged Annual Household Income Quintile = 2,	-0.22***	0.05
Lagged Annual Household Income Quintile = 3,	-0.36***	0.05
Lagged Annual Household Income Quintile = 4,	-0.40***	0.05
Lagged Annual Household Income Quintile = 5,	-0.57***	0.05
Lagged Number of Children in Household,	-0.23***	0.02
Lagged Number of Children aged 0-2 in Household,	-0.20***	0.05
Lagged Self-rated Health,	-0.00	0.01
Region = 1, North East	-0.09	0.08
Region = 2, North West	0.00	0.06
Region = 3, Yorkshire and the Humber	0.05	0.07
Region = 4, East Midlands	0.01	0.06
Region = 5, West Midlands	0.03	0.06
Region = 6, East of England	-0.02	0.07
Region = 8, South East	-0.04	0.06
Region = 9, South West	0.04	0.06
Region = 10, Wales	0.10	0.08
Region = 11, Scotland	-0.04	0.07
Region = 12, Northern Ireland	-0.41***	0.11
Year	-0.01	0.01
Constant	-0.48**	0.19
Observations	93,826	
R2	0.136	
Chi2	1385	
Log-likelihood	-8151	

*** p<0.01, ** p<0.05, * p<0.1

Table A.10: Process U2: Probability of partnership break-up.
Sample: Female member of a couple aged 16 or older, not in continuous education.

Probit Probability of partnership dissolution	(1) Coef.	(2) s.e.
Age	-0.02	0.02
Age Squared	0.00	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	0.04	0.09
Educational Attainment: 3 Category = 3, Other/No Qualification	0.14	0.14
Lagged Personal Non-benefit Gross Income,	-0.07	0.07
Lagged Personal Non-benefit Gross Income Squared,	0.01	0.01
Lagged Number of Children in Household,	0.01	0.05
Lagged Number of Children aged 0-2 in Household,	-0.02	0.09
Lagged Self-rated Health,	-0.05	0.04
Spouse's Educational Level: 3 Category = 2, Other Higher/A-level/GCSE	0.26***	0.10
Spouse's Educational Level: 3 Category = 3, Other/No Qualification	0.24*	0.14
Lagged Spouse's Self-rated Health,	-0.06	0.04
Lagged Differential Personal Non-Benefit Gross Income,	-0.02	0.02
Lagged Number of Years in Partnership,	-0.02***	0.00
Lagged Difference in Age between partners in a couple,	0.02***	0.01
Household Type: 4 Category = Couple with children,	-0.05	0.12
Lagged Couple Employment Status = Employed, spouse not,	0.23	0.17
Lagged Couple Employment Status = Not employed, spouse employed,	-0.29*	0.15
Lagged Couple Employment Status = Both not employed,	0.24	0.23
Region = 1, North East	0.03	0.20
Region = 2, North West	0.03	0.16
Region = 3, Yorkshire and the Humber	-0.11	0.18
Region = 4, East Midlands	0.10	0.17
Region = 5, West Midlands	-0.16	0.19
Region = 6, East of England	-0.05	0.17
Region = 8, South East	0.04	0.16
Region = 9, South West	0.11	0.16
Region = 10, Wales	0.00	0.22
Region = 11, Scotland	-0.09	0.20
Region = 12, Northern Ireland	-0.05	0.36
Year	0.00	0.02
Constant	-1.42**	0.58
Observations	50,347	
R2	0.131	
Chi2	160.6	
Log-likelihood	-723.8	

*** p<0.01, ** p<0.05, * p<0.1

Table A.11: Process F1a: Probability of giving birth to a child.
Sample: Women aged 18-44 in continuous education.

Probit Probability of giving birth	(1) Coef.	(2) s.e.
Age	0.01	0.06
Lagged Annual Household Income Quintile = 2,	-0.22	0.20
Lagged Annual Household Income Quintile = 3,	-0.03	0.28
Lagged Annual Household Income Quintile = 4,	-0.82**	0.35
Lagged Annual Household Income Quintile = 5,	0.00	0.00
Lagged Number of Children in Household,	0.14	0.25
Lagged Number of Children aged 0-2 in Household,	0.32	0.71
Lagged Self-rated Health,	0.07	0.12
Marital Status = Single never married,	-0.29	0.41
Marital Status = Previously partnered,	1.29	0.97
Constant	-2.51*	1.39
Observations	1,525	
R2	0.0518	
Chi2	20.41	
Log-likelihood	-77	

*** p<0.01, ** p<0.05, * p<0.1

Table A.12: Process F1b: Probability of giving birth to a child.
Sample: Women aged 18-44 not in continuous education.

Probit Probability of giving birth to a child	(1) Coef.	(2) s.e.
Age	0.17***	0.02
Age Squared	-0.00***	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	0.05	0.04
Educational Attainment: 3 Category = 3, Other/No Qualification	0.10	0.07
UK Fertility Rate	0.21***	0.01
Lagged Employment Status: 3 Category = Student,	-0.23***	0.09
Lagged Employment Status: 3 Category = Not employed,	0.16***	0.04
Lagged Annual Household Income Quintile = 2,	-0.08	0.06
Lagged Annual Household Income Quintile = 3,	-0.11*	0.06
Lagged Annual Household Income Quintile = 4,	-0.12**	0.06
Lagged Annual Household Income Quintile = 5,	-0.01	0.06
Lagged Number of Children in Household,	-0.12***	0.02
Lagged Number of Children aged 0-2 in Household,	0.17***	0.04
Lagged Self-rated Health,	0.06***	0.02
Lagged Marital Status = Single never married,	-0.62***	0.05
Lagged Marital Status = Previously partnered,	-0.24***	0.08
Region = 1, North East	-0.18**	0.09
Region = 2, North West	-0.19***	0.06
Region = 3, Yorkshire and the Humber	-0.26***	0.07
Region = 4, East Midlands	-0.14**	0.07
Region = 5, West Midlands	-0.06	0.07
Region = 6, East of England	-0.20***	0.07
Region = 8, South East	-0.12**	0.06
Region = 9, South West	-0.21***	0.07
Region = 10, Wales	-0.06	0.08
Region = 11, Scotland	-0.23***	0.07
Region = 12, Northern Ireland	-0.13	0.09
Constant	-16.37***	0.96
Observations	25,646	
R2	0.113	
Chi2	907.5	
Log-likelihood	-4800	

*** p<0.01, ** p<0.05, * p<0.1

Table A.13: Process HO1: Probability of being a homeowner.
Sample: Individuals aged 16 and above.

Probit Probability of owning a house	(1) Coef.	(2) s.e.
Gender	-0.09***	0.03
Age	0.07***	0.01
Age Squared	-0.00***	0.00
Employment Status: 5 Category = 2, Student	-0.43***	0.13
Employment Status: 5 Category = 3, Not Employed	-0.06*	0.03
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.07**	0.04
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.40***	0.05
Lagged Self-rated Health	0.09***	0.01
Lagged Adjusted Annual Household Income Quintile = 2	0.47***	0.05
Lagged Adjusted Annual Household Income Quintile = 3	0.89***	0.05
Lagged Adjusted Annual Household Income Quintile = 4	1.30***	0.05
Lagged Adjusted Annual Household Income Quintile = 5	1.67***	0.06
Lagged Inverse Hyperbolic Sine Gross Non-employment or Benefit Personal Income	0.05***	0.00
Region = 1, North East	0.34***	0.08
Region = 2, North West	0.60***	0.06
Region = 3, Yorkshire and the Humber	0.40***	0.07
Region = 4, East Midlands	0.45***	0.07
Region = 5, West Midlands	0.43***	0.07
Region = 6, East of England	0.26***	0.06
Region = 8, South East	0.26***	0.06
Region = 9, South West	0.37***	0.06
Region = 10, Wales	0.58***	0.08
Region = 11, Scotland	0.31***	0.07
Region = 12, Northern Ireland	0.57***	0.09
Lagged Household Type: 4 Category = 2, Couples with Children	0.40***	0.03
Lagged Household Type: 4 Category = 3, Single with No Children	0.26	0.32
Lagged Household Type: 4 Category = 4, Single with Children	0.28	0.26
Lagged Spouse's Labour Force Status: 3 Category = 2, Student	-0.36***	0.11
Lagged Spouse's Labour Force Status: 3 Category = 3, Not Employed	-0.08**	0.03
Year	-0.00	0.00
Constant	-3.61***	0.16
Observations	75,437	
R2	0.256	
Chi2	3376	
Log-likelihood	-29043	

*** p<0.01, ** p<0.05, * p<0.1

Table A.14: Process R1a: Probability of retiring.
Sample: Non-partnered individuals aged 50+ who are not yet retired.

Probit Probability of retiring	(1) Coef.	(2) s.e.
Gender = 1, Male	0.21***	0.04
Age	0.43***	0.04
Age Squared	-0.00***	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.08	0.06
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.12*	0.06
Reached State Pension Age = 1, Yes	0.78***	0.07
Lagged Employment Status: Not employed,	0.99***	0.07
Lagged Adjusted Annual Household Income Quintile = 2,	0.43***	0.07
Lagged Adjusted Annual Household Income Quintile = 3,	0.30***	0.09
Lagged Adjusted Annual Household Income Quintile = 4,	0.39***	0.09
Lagged Adjusted Annual Household Income Quintile = 5,	0.50***	0.09
Lagged Long-term Sick or Disabled,	-0.03	0.06
Region = 1, North East	0.01	0.11
Region = 2, North West	0.04	0.09
Region = 3, Yorkshire and the Humber	-0.09	0.10
Region = 4, East Midlands	-0.13	0.10
Region = 5, West Midlands	-0.08	0.10
Region = 6, East of England	-0.15	0.09
Region = 8, South East	-0.15*	0.08
Region = 9, South West	0.00	0.09
Region = 10, Wales	0.01	0.11
Region = 11, Scotland	-0.07	0.09
Region = 12, Northern Ireland	-0.02	0.11
Year	-0.01	0.01
Constant	-17.60***	1.35
Observations	29,065	
R2	0.491	
Chi2	1559	
Log-likelihood	-2246	

*** p<0.01, ** p<0.05, * p<0.1

Table A16: Process R1b: Probability of retiring.
Sample: Partnered individuals aged 50+ who are not yet retired.

Probit Probability of retiring	(1) Coef.	(2) s.e.
Gender = 1, Male	0.01	0.03
Age	0.52***	0.04
Age Squared	-0.00***	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.12***	0.03
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.23***	0.04
Reached State Pension Age = 1, Yes	0.43***	0.05
Lagged Employment Status: Not employed,	0.76***	0.05
Lagged Adjusted Annual Household Income Quintile = 2,	0.15**	0.07
Lagged Adjusted Annual Household Income Quintile = 3,	0.22***	0.07
Lagged Adjusted Annual Household Income Quintile = 4,	0.27***	0.07
Lagged Adjusted Annual Household Income Quintile = 5,	0.43***	0.07
Lagged Long-term Sick or Disabled = 1,	-0.06	0.06
Spouse of Pension Age = 1, Yes	0.07*	0.04
Spouse's Labour Force Status: 3 Category = Student,	0.04	0.34
Spouse's Labour Force Status: 3 Category = Not employed,	0.33***	0.03
Spouse's Disability Status = 1,	-0.10	0.07
Region = 1, North East	0.32***	0.08
Region = 2, North West	0.18***	0.06
Region = 3, Yorkshire and the Humber	0.28***	0.07
Region = 4, East Midlands	0.22***	0.07
Region = 5, West Midlands	0.13**	0.07
Region = 6, East of England	0.10	0.06
Region = 8, South East	0.03	0.06
Region = 9, South West	0.07	0.06
Region = 10, Wales	0.19**	0.08
Region = 11, Scotland	0.15**	0.07
Region = 12, Northern Ireland	0.11	0.09
Reached State Pension Age = 1#Lagged Employment Status: Not employed	0.62***	0.07
Year	0.00	0.01
Constant	-20.34***	1.15
Observations	27,198	
R2	0.286	
Chi2	3328	
Log-likelihood	-5561	

*** p<0.01, ** p<0.05, * p<0.1

Table A17: Process I3a selection: Probability of receiving capital income.
Sample: Individuals aged 16 - 29 who are in continuous education.

Logit Probability of receiving capital income	(1) Coef.	(2) s.e.
Gender = 1, omitted	-	-
Age	0.74	0.65
Age Squared	-0.02	0.02
Lagged Self-rated Health,	0.14	0.10
Lagged Inverse Hyperbolic Sine Gross Employment Personal Income,	-0.04	0.03
Lagged Inverse Hyperbolic Sine Gross Capital Income,	0.19***	0.03
Region = 1, North East	-0.10	0.43
Region = 2, North West	-0.22	0.39
Region = 3, Yorkshire and the Humber	0.12	0.40
Region = 4, East Midlands	0.66*	0.39
Region = 5, West Midlands	-0.06	0.37
Region = 6, East of England	0.56	0.41
Region = 8, South East	0.25	0.32
Region = 9, South West	0.69*	0.37
Region = 10, Wales	0.39	0.41
Region = 11, Scotland	0.43	0.48
Region = 12, Northern Ireland	-0.07	0.50
Year	-0.03	0.05
Constant	-8.55	6.67
Observations	950	
R2	0.0859	
Chi2	71.52	
Log-likelihood	-653.8	

*** p<0.01, ** p<0.05, * p<0.1

Table A18: Process I3b selection: Probability of receiving capital income.
Sample: Individuals aged 16+ who are not in continuous education.

Logit Probability of receiving capital income	(1) Coef.	(2) s.e.
Gender = 1, omitted	-	-
Age	-0.01*	0.01
Age Squared	0.00***	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.34***	0.04
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.68***	0.05
Lagged Employment Status: 3 Category = Student,	-0.91***	0.22
Lagged Employment Status: 3 Category = Not employed,	-0.06	0.07
Lagged Household Type: 4 Category = Couples with children,	-0.39***	0.04
Lagged Household Type: 4 Category = Singles without children,	-0.02	0.04
Lagged Household Type: 4 Category = Singles with children,	-0.44***	0.17
Lagged Self-rated Health,	0.12***	0.01
Lagged Inverse Hyperbolic Sine Gross Employment Personal Income L =1,	0.00	0.01
Lagged Inverse Hyperbolic Sine Gross Capital Income L = 1,	0.35***	0.01
Inverse Hyperbolic Sine Gross Employment Personal Income L = 2,	0.07***	0.01
Lagged Inverse Hyperbolic Sine Gross Capital Income L = 2,	0.22***	0.01
Region = 1, North East	-0.03	0.09
Region = 2, North West	0.10	0.07
Region = 3, Yorkshire and the Humber	0.17**	0.07
Region = 4, East Midlands	0.37***	0.07
Region = 5, West Midlands	0.28***	0.07
Region = 6, East of England	0.31***	0.07
Region = 8, South East	0.37***	0.07
Region = 9, South West	0.34***	0.07
Region = 10, Wales	0.13	0.09
Region = 11, Scotland	-0.02	0.07
Region = 12, Northern Ireland	0.02	0.09
Year	0.02***	0.01
Constant	-2.63***	0.22
Observations	43,310	
R2	0.298	
Chi2	6594	
Log-likelihood	-24199	

*** p<0.01, ** p<0.05, * p<0.1

Table A19: Process I3a: Amount of capital income.
Sample: Individuals aged 16 - 29 who are in continuous education and receive capital income.

Linear regression Amount of capital income	(1) Coef.	(2) s.e.
Gender = 1, omitted	-	-
Age	1.33**	0.66
Age Squared	-0.03*	0.02
Lagged Self-rated Health,	0.09	0.12
Lagged Inverse Hyperbolic Sine Gross Employment Personal Income,	0.04	0.04
Lagged Inverse Hyperbolic Sine Gross Capital Income,	0.33***	0.05
Region = 1, North East	-0.38	0.46
Region = 2, North West	0.58	0.46
Region = 3, Yorkshire and the Humber	-0.06	0.35
Region = 4, East Midlands	0.16	0.39
Region = 5, West Midlands	0.29	0.41
Region = 6, East of England	0.29	0.42
Region = 8, South East	0.16	0.30
Region = 9, South West	0.33	0.38
Region = 10, Wales	0.83***	0.31
Region = 11, Scotland	-0.09	0.41
Region = 12, Northern Ireland	-0.08	0.51
Year	-0.09*	0.05
Constant	-11.62*	6.81
Observations	656	
R-squared	0.25	
R2	0.250	
RMSE	2.007	

*** p<0.01, ** p<0.05, * p<0.1

Table A20: Process I3b: Amount of capital income.
Sample: Individuals aged 16+ who are not in continuous education and receive capital income.

Linear regression Amount of capital income	(1) Coef.	(2) s.e.
Gender = 1, omitted	-	-
Age	0.04***	0.01
Age Squared	-0.00***	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	0.05	0.04
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.09*	0.05
Lagged Employment Status: 3 Category = Student,	0.46**	0.23
Lagged Employment Status: 3 Category = Not employed,	0.25***	0.08
Lagged Household Type: 4 Category = Couples with children,	0.02	0.05
Lagged Household Type: 4 Category = Singles without children,	-0.14***	0.04
Lagged Household Type: 4 Category = Singles with children,	0.60**	0.29
Lagged Self-rated Health,	0.04***	0.01
Lagged Inverse Hyperbolic Sine Gross Employment Personal Income L = 1,	0.02*	0.01
Lagged Inverse Hyperbolic Sine Gross Capital Income L = 1,	0.32***	0.01
Lagged Inverse Hyperbolic Sine Gross Employment Personal Income L = 2,	0.01	0.01
Lagged Inverse Hyperbolic Sine Gross Capital Income L = 2,	0.19***	0.01
Region = 1, North East	-0.27***	0.10
Region = 2, North West	-0.15**	0.07
Region = 3, Yorkshire and the Humber	-0.23***	0.08
Region = 4, East Midlands	-0.34***	0.07
Region = 5, West Midlands	-0.12	0.08
Region = 6, East of England	-0.29***	0.07
Region = 8, South East	-0.21***	0.07
Region = 9, South West	-0.25***	0.07
Region = 10, Wales	-0.16*	0.09
Region = 11, Scotland	-0.11	0.08
Region = 12, Northern Ireland	-0.07	0.11
Year	-0.03***	0.01
Constant	1.74***	0.24
Observations	21,567	
R-squared	0.37	
R2	0.374	
RMSE	1.775	

*** p<0.01, ** p<0.05, * p<0.1

Table A21: Process I4b: Amount of pension income.
Sample: Retired individuals who were retired in the previous year.

Linear regression Amount of pension income	(1) Coef.	(2) s.e.
Age	-0.20***	0.02
Age Squared	0.00***	0.00
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.18***	0.03
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.29***	0.03
Lagged Household Type: 4 Category = Couples with children,	0.01	0.20
Lagged Household Type: 4 Category = Singles without children,	0.01	0.02
Lagged Household Type: 4 Category = Singles with children,	-0.02	0.19
Lagged Self-rated Health,	0.03***	0.01
Lagged Inverse Hyperbolic Sine Gross Private Pension Income L = 1,	0.66***	0.01
Lagged Inverse Hyperbolic Sine Gross Private Pension Income L = 2,	0.26***	0.01
Region = 1, North East	0.02	0.05
Region = 2, North West	0.03	0.04
Region = 3, Yorkshire and the Humber	-0.00	0.05
Region = 4, East Midlands	0.04	0.05
Region = 5, West Midlands	0.03	0.05
Region = 6, East of England	0.04	0.04
Region = 8, South East	0.05	0.04
Region = 9, South West	-0.02	0.04
Region = 10, Wales	0.02	0.05
Region = 11, Scotland	-0.01	0.05
Region = 12, Northern Ireland	-0.05	0.05
Growth Rate	-0.42	1.01
Year	-0.01	0.01
Constant	9.04***	1.13
Observations	26,750	
R-squared	0.82	
R2	0.820	
RMSE	1.423	

*** p<0.01, ** p<0.05, * p<0.1

Table A22: Process I5a: Probability of receiving private pension income.
Sample: Retired individuals who were not retired in the previous year.

Logit Probability of receiving private pension income	(1) Coef.	(2) s.e.
Gender = 1, omitted	-	-
Above State Pension Age	0.28	0.19
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-0.32*	0.17
Educational Attainment: 3 Category = 3, Other/No Qualification	-0.69***	0.20
Lagged Employment Status: 3 Category = Employed,	0.66***	0.17
Lagged Employment Status: 3 Category = Not Employed,	0.00	0.00
Lagged Household Type: 4 Category = Couples with children,	-0.23	0.32
Lagged Household Type: 4 Category = Singles without children,	-0.05	0.16
Lagged Household Type: 4 Category = Singles with children,	-0.55	1.17
Lagged Self-rated Health,	0.08	0.06
Lagged Hourly Wage Potential,	-0.00	0.00
Region = 1, North East	0.55	0.38
Region = 2, North West	0.44	0.33
Region = 3, Yorkshire and the Humber	0.55	0.34
Region = 4, East Midlands	0.29	0.35
Region = 5, West Midlands	0.47	0.34
Region = 6, East of England	0.40	0.32
Region = 8, South East	0.10	0.30
Region = 9, South West	0.25	0.33
Region = 10, Wales	0.33	0.39
Region = 11, Scotland	0.08	0.35
Region = 12, Northern Ireland	0.14	0.37
Growth rate	-1.51	5.03
Year	-0.04	0.04
Constant	1.70	4.79
Observations	1,202	
R2	0.0488	
Chi2	60.37	
Log-likelihood	-809.3	

*** p<0.01, ** p<0.05, * p<0.1

Table A23: Process I5b: Amount of private pension income.
Sample: Retired individuals who were not retired in the previous year and receive private pension income.

Linear regression Amount of private pension income	(1) Coef.	(2) s.e.
Gender = 1, omitted	-	-
Above State Pension Age	-398.11***	102.91
Educational Attainment: 3 Category = 2, Other Higher/A-level/GCSE	-331.87***	100.65
Educational Attainment: 3 Category = 3, Other/No Qualification	-676.12***	118.93
Lagged Employment Status: 3 Category = Employed,	126.78	110.02
Lagged Employment Status: 3 Category = Not Employed,	0.00	0.00
Lagged Household Type: 4 Category = Couples with children,	-129.88	142.53
Lagged Household Type: 4 Category = Singles without children,	-184.97*	102.33
Lagged Household Type: 4 Category = Singles with children,	-252.97	177.48
Lagged Self-rated Health,	92.07**	36.11
Lagged Hourly Wage Potential,	15.46***	3.84
Region = 1, North East	-358.45	286.18
Region = 2, North West	-426.01	287.05
Region = 3, Yorkshire and the Humber	-140.91	314.69
Region = 4, East Midlands	136.49	382.75
Region = 5, West Midlands	-337.23	303.80
Region = 6, East of England	-118.07	294.96
Region = 8, South East	-194.72	300.50
Region = 9, South West	-284.52	320.28
Region = 10, Wales	-199.00	301.17
Region = 11, Scotland	-174.80	318.36
Region = 12, Northern Ireland	64.55	316.53
Growth rate	5,333.02*	2,733.35
Year	-3.02	23.74
Constant	-4,182.81	2,735.74
Observations	713	
R-squared	0.28	
R2	0.281	
RMSE	1006	

*** p<0.01, ** p<0.05, * p<0.1

Table A24: Probit regression estimates for receipt of informal social care services among people aged 16 to 64 with a long-term illness or disability.

Sample: Pooled data reported by FRS at annual intervals between 2015/16 and 2019/20, and 2021/22, individuals between age 16 and 64 with a long-term illness or disability.

	Coef.	s.e.	p>z
Education Level (Ref = High)			
Medium	0.0018	0.0009	0.036
Low	-0.0231	0.0013	0.000
Gender (Ref = Women)			
Men	0.0937	0.0008	0.000
under age 25	0.3368	0.0013	0.000
Region (Ref = London)			
North East	0.2579	0.0022	0.000
North West	0.2259	0.0017	0.000
Yorkshire and the Humber	0.1577	0.0019	0.000
East Midlands	0.2917	0.0020	0.000
West Midlands	0.1143	0.0019	0.000
East of England	0.1945	0.0020	0.000
South East	0.1999	0.0019	0.000
South West	0.2308	0.0019	0.000
Wales	-0.0191	0.0021	0.000
Scotland	0.1728	0.0018	0.000
Northern Ireland	0.2750	0.0024	0.000
Constant	-0.7291	0.0015	0.000
Observations	7,248		
Pseudo R2	0.0098		

Note: Robust standard errors reported. Long term illness or disability identified as code 9 of variable empstati.

Table A25: Linear least squares regression estimates for hours of informal care per week received by people aged 16 to 64 years, with a long-term illness or disability, and in receipt of some informal social care.

Sample: Pooled data reported by FRS at annual intervals between 2015/16 and 2019/20, and 2021/22, individuals between age 16 and 64 with a long-term illness or disability.

	Coef.	s.e.	p>z
Education Level (Ref = High)			
Medium	0.064	0.0014	0.000
Low	0.077	0.0020	0.000
Gender (Ref = Women)			
Men	-0.039	0.0013	0.000
Age (Ref = under age 25)			
25 to 39	-0.308	0.0022	0.000
40+	-0.568	0.0018	0.000
Region (Ref = London)			
North East	-0.008	0.0032	0.010
North West	0.046	0.0027	0.000
Yorkshire and the Humber	0.066	0.0030	0.000
East Midlands	-0.202	0.0031	0.000
West Midlands	0.022	0.0030	0.000
East of England	-0.148	0.0032	0.000
South East	-0.154	0.0030	0.000
South West	-0.251	0.0031	0.000
Wales	-0.033	0.0033	0.000
Scotland	-0.001	0.0029	0.724
Northern Ireland	-0.086	0.0035	0.000
Constant	4.213	0.0028	0.000
Observatioms	2,265		
RMSE	1.1671		
R-squared	0.0359		

Note: Robust standard errors reported. Long term illness or disability identified as code 9 of variable empstati.

Table A26: Probit regression estimates for “in need of care” for people aged 65+.

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and over without missing variables.

	Coef.	s.e.	p>z
Gender (Ref = Women)			
Men	-0.041	0.0242	0.089
Education Level (Ref = High)			
Medium	0.058	0.0318	0.067
Low	0.071	0.0344	0.038
partner	0.160	0.0265	0.000
need care (lag)	2.323	0.0260	0.000
Self-rated health (Ref = Excellent)			
Very good	0.016	0.0556	0.770
Good	0.203	0.0545	0.000
Fair	0.508	0.0568	0.000
Poor	0.965	0.0724	0.000
Age group (Ref = 65-66)			
67-68	-0.374	0.0472	0.000
69-70	-0.258	0.0446	0.000
71-72	-0.298	0.0442	0.000
73-74	-0.185	0.0472	0.000
75-76	-0.215	0.0502	0.000
77-78	-0.054	0.0522	0.296
79-80	-0.047	0.0582	0.415
81-82	-0.019	0.0596	0.754
83-84	0.051	0.0617	0.411
85+	0.167	0.0583	0.004
Region (Ref = London)			
North East	0.040	0.0767	0.606
North West	0.072	0.0624	0.250
Yorkshire and the Humber	0.085	0.0651	0.190
East Midlands	0.050	0.0657	0.449
West Midlands	0.067	0.0647	0.300
East of England	0.047	0.0622	0.451
South East	-0.021	0.0600	0.720
South West	0.092	0.0618	0.137
Wales	0.176	0.0653	0.007
Scotland	0.112	0.0628	0.075
Northern Ireland	0.191	0.0651	0.003
Constant	-1.542	0.0802	0.000
Observations	21,723		
Proportion positive	0.3766		
Pseudo R2	0.5063		

Note: Robust standard errors reported. "Need care" defined as requiring assistance with any one of the activities of daily living reported by the UKHLS (including instrumental activities).

Table A27: Probit regression estimates for receipt of social care for people aged 65+. Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, limited to individuals aged 65 and over without missing variables.

	Coef.	s.e.	p>z
Gender (Ref = Women)			
Men	-0.100	0.0284	0.000
Education Level (Ref = High)			
Medium	0.026	0.0387	0.497
Low	0.082	0.0407	0.045
partner	0.201	0.0312	0.000
receive care (lag)	2.296	0.0323	0.000
Self-rated health (Ref = Excellent)			
Very good	0.124	0.1012	0.219
Good	0.498	0.0988	0.000
Fair	0.916	0.0995	0.000
Poor	1.423	0.1071	0.000
Age group (Ref = 65-66)			
67-68	-0.250	0.0564	0.000
69-70	-0.121	0.0539	0.024
71-72	-0.128	0.0528	0.016
73-74	-0.070	0.0549	0.202
75-76	-0.030	0.0591	0.611
77-78	0.059	0.0610	0.335
79-80	0.141	0.0628	0.025
81-82	0.205	0.0660	0.002
83-84	0.289	0.0657	0.000
85+	0.542	0.0631	0.000
Region (Ref = London)			
North East	0.041	0.0920	0.659
North West	0.022	0.0737	0.768
Yorkshire and the Humber	0.030	0.0769	0.699
East Midlands	0.037	0.0789	0.643
West Midlands	0.123	0.0753	0.103
East of England	0.074	0.0733	0.315
South East	-0.001	0.0725	0.989
South West	0.048	0.0729	0.506
Wales	0.177	0.0769	0.021
Scotland	0.134	0.0742	0.071
Northern Ireland	0.268	0.0764	0.000
Constant	-2.376	0.1227	0.000
Oservations	21,723		
Proportion positive	0.2116		
Pseudo R2	0.5372		

Note: Robust standard errors reported. "Receive care" defined as reported receipt of help with at least one of the activities of daily living reported by the UKHLS in the week preceding the survey.

Table A28: Multinomial logit regression estimates for formal and informal social care of population aged 65 and over in receipt of some care (reference group: only informal care).

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and over receiving social care without missing variables.

	Coef.	s.e.	p>z	Coef.	s.e.	p>z
	formal and informal care			*only formal care*		
Population share	0.2057			0.1227		
Education Level (Ref = High)						
Medium	-0.292	0.1570	0.063	-0.387	0.1950	0.047
Low	-0.416	0.1533	0.007	-1.145	0.1938	0.000
partner	-0.576	0.1050	0.000	-1.687	0.1460	0.000
care market (lag, ref = none)						
informal only	-1.244	0.1160	0.000	-2.543	0.2109	0.000
formal and informal	2.987	0.1364	0.000	0.777	0.2076	0.000
only formal	1.607	0.2781	0.000	4.191	0.2431	0.000
aged 85 and over	0.258	0.1295	0.046	-0.006	0.1761	0.974
Region (Ref = London)						
North East	-0.020	0.3503	0.955	-1.156	0.5184	0.026
North West	0.021	0.2964	0.944	-0.197	0.3457	0.569
Yorkshire and the Humber	0.456	0.2991	0.128	-0.118	0.3707	0.750
East Midlands	0.081	0.3118	0.796	0.345	0.3586	0.336
West Midlands	0.124	0.3065	0.686	0.044	0.3583	0.901
East of England	0.769	0.2929	0.009	0.359	0.3368	0.286
South East	0.493	0.2940	0.093	0.094	0.3353	0.779
South West	0.445	0.2892	0.124	0.143	0.3363	0.671
Wales	0.093	0.2918	0.751	-0.272	0.3481	0.434
Scotland	0.321	0.2875	0.264	-0.310	0.3440	0.368
Northern Ireland	0.534	0.2881	0.064	0.017	0.3273	0.960
Constant	-1.128	0.2862	0.000	-0.267	0.3131	0.394
Observations	5,726					
Share of "only informal care"	0.6716					
Pseudo R2	0.4481					

Note: Robust standard errors reported. "lag" defined as preceding year.

Table A29: Probit regression estimates describing incidence of partners providing social care for people aged 65 and over receiving care and with a partner.

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and over receiving social care, with a partner, and without missing variables.

	Coef.	s.e.	p>z
Gender (Ref = Women)			
Men	0.254	0.0864	0.003
care from partner (lag)	1.446	0.0971	0.000
formal care received	-0.301	0.1025	0.003
aged 85 and over	-0.548	0.1142	0.000
Region (Ref = London)			
North East	0.190	0.3080	0.538
North West	-0.047	0.2286	0.837
Yorkshire and the Humber	-0.154	0.2354	0.514
East Midlands	-0.106	0.2416	0.661
West Midlands	-0.303	0.2281	0.184
East of England	-0.043	0.2497	0.862
South East	0.235	0.2435	0.334
South West	0.121	0.2535	0.633
Wales	-0.251	0.2330	0.282
Scotland	0.108	0.2485	0.665
Northern Ireland	-0.329	0.2318	0.156
Constant	0.825	0.2017	0.000
Observations	3,176		
Proportion positive	0.9186		
Pseudo R2	0.2505		

Note: Robust standard errors reported. "lag" defined as preceding year.

Table A30: Multinomial logit regression estimates for receipt of supplementary care for population aged 65 and over who receive care from their partner (reference group: none).

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and over receiving social care from their partner and without missing variables.

	Coef.	s.e.	p>z
Daughter			
Population share	0.1048		
Supplementary carer (lag, ref = none)			
Daughter	5.253	0.2482	0.000
Son	2.345	0.6135	0.000
Other	2.479	0.6058	0.000
Care from partner (lag)	1.087	0.7086	0.125
Constant	-4.752	0.7263	0.000
Son			
Population share	0.0406		
Supplementary carer (lag, ref = none)			
Daughter	2.305	0.5646	0.000
Son	5.988	0.3731	0.000
Other	3.424	0.6542	0.000
Care from partner (lag)	1.419	0.8477	0.094
Constant	-5.889	0.8788	0.000
Other			
Population share	0.0238		
Supplementary carer (lag, ref = none)			
Daughter	1.332	1.0583	0.208
Son	2.999	0.7267	0.000
Other	6.108	0.4798	0.000
Care from partner (lag)	16.038	0.5285	0.000
Constant	-20.810	0.6080	0.000
Observations	1998		
Share of "none"	0.8309		
Pseudo R2	0.5285		

Note: Regression considers four alternatives for supplementary carers: none (reference), daughter, son, and other. Robust standard errors reported. "lag" defined as preceding year.

Table A31: Multinomial logit regression estimates for informal carer(s) for population aged 65 and over who receive care but not from a partner (reference group: daughter only).

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and receiving social care but not from a partner and without missing variables.

	Coef.	s.e.	p>z	Coef.	s.e.	p>z
	Daughter and son			*Daughter and other*		
Population share	0.0822			0.0924		
Carer(s) (lag, ref: none)						
Daughter only	-2.279	0.3566	0.000	-1.701	0.3164	0.000
Daughter and son	3.415	0.3473	0.000	-2.708	1.0562	0.010
Daughter and other	-0.955	0.6524	0.143	3.162	0.3449	0.000
Son only	2.537	0.5140	0.000	-0.147	0.6953	0.833
Son and other	2.944	1.4254	0.039	1.149	1.4277	0.421
Other only	-0.285	1.0008	0.776	0.757	0.6439	0.240
Constant	-1.533	0.1756	0.000	-1.586	0.1931	0.000
	Son only			*Son and other*		
Population share	0.1640			0.0513		
Carer(s) (lag, ref: none)						
Daughter only	-4.261	0.5518	0.000	-2.628	0.6440	0.000
Daughter and son	-0.152	0.4764	0.750	0.488	0.8075	0.545
Daughter and other	-3.164	1.0421	0.002	-1.710	1.0677	0.109
Son only	4.475	0.4313	0.000	2.982	0.5800	0.000
Son and other	4.226	1.0790	0.000	7.554	1.0474	0.000
Other only	0.400	0.5718	0.484	1.446	0.7086	0.041
Constant	-0.784	0.1372	0.000	-2.216	0.2696	0.000
	Other only					
Population share	0.2492					
Carer(s) (lag, ref: none)						
Daughter only	-4.145	0.4039	0.000			
Daughter and son	-1.396	0.7752	0.072			
Daughter and other	-1.607	0.6581	0.015			
Son only	-0.606	0.7058	0.391			
Son and other	1.213	1.3403	0.365			
Other only	3.771	0.4380	0.000			
Constant	-0.264	0.1181	0.025			
Observations	2,232					
Share of "daughter only"	0.3609					
Pseudo R2	0.5311					

Note: Regression considers six possible alternatives: none daughter only (reference), daughter and son, daughter and other, son only, son and other, and other only. Robust standard errors reported. "lag" refers to preceding year.

Table A32: Linear least squares regression estimates for log hours of informal care per week provided by partner to people aged 65 and over.

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and receiving social care from a partner and without missing variables.

	Coef.	s.e.	p>z
Gender (ref = Women)			
Men	0.144	0.070	0.041
Education Level (ref = High)			
Medium	0.056	0.109	0.606
Low	0.288	0.109	0.009
Supplementary carer (ref = none)			
Daughter	0.355	0.127	0.005
Son	0.280	0.153	0.067
Other	0.522	0.161	0.001
Formal market	0.264	0.096	0.006
Self-rated health poor	0.659	0.085	0.000
Region (Ref = London)			
North East	0.314	0.254	0.217
North West	0.024	0.193	0.901
Yorkshire and the Humber	0.131	0.200	0.513
East Midlands	-0.053	0.198	0.791
West Midlands	-0.267	0.194	0.168
East of England	-0.014	0.187	0.940
South East	-0.128	0.197	0.516
South West	-0.177	0.189	0.348
Wales	-0.012	0.187	0.950
Scotland	-0.090	0.191	0.637
Northern Ireland	-0.026	0.199	0.897
Constant	1.641	0.189	0.000
Observations	1,626		
RMSE	1.2093		
R-squared	0.1179		

Note: Robust standard errors reported.

Table A33: Linear least squares regression estimates for log hours of informal care per week provided by daughter to people aged 65 and over.

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and receiving social care from a partner and without missing variables.

	Coef.	s.e.	p>z
Gender (ref = Women)			
Men	-0.053	0.088	0.549
Education Level (ref = High)			
Medium	-0.236	0.193	0.224
Low	-0.198	0.186	0.286
Supplementary carer (ref = none)			
Partner	-0.282	0.095	0.003
Son	-0.002	0.094	0.985
Other	-0.124	0.089	0.166
Formal market	0.176	0.091	0.055
Self-rated health poor	0.305	0.091	0.001
Region (Ref = London)			
North East	-0.389	0.233	0.094
North West	0.012	0.225	0.959
Yorkshire and the Humber	-0.075	0.243	0.759
East Midlands	-0.204	0.219	0.353
West Midlands	0.013	0.199	0.948
East of England	-0.361	0.201	0.073
South East	-0.329	0.202	0.104
South West	-0.084	0.209	0.688
Wales	0.061	0.206	0.766
Scotland	-0.057	0.202	0.777
Northern Ireland	0.023	0.203	0.909
Constant	1.982	0.234	0.000
Observations	894		
RMSE	0.9889		
R-squared	0.0570		

Note: Robust standard errors reported.

Table A34: Linear least squares regression estimates for log hours of informal care per week provided by son to people aged 65 and over.

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and receiving social care from a partner and without missing variables. Explanatory variables describe characteristics of person in receipt of care.

	Coef.	s.e.	p>z
Gender (ref = Women)			
Men	-0.039	0.109	0.723
Education Level (ref = High)			
Medium	-0.293	0.244	0.232
Low	-0.080	0.228	0.727
Supplementary carer (ref = none)			
Partner	-0.255	0.124	0.039
Daughter	-0.070	0.097	0.470
Other	-0.145	0.098	0.141
Formal market	-0.045	0.110	0.681
Self-rated health poor	0.340	0.116	0.004
Region (Ref = London)			
North East	0.245	0.453	0.589
North West	0.031	0.207	0.882
Yorkshire and the Humber	-0.017	0.220	0.937
East Midlands	-0.056	0.257	0.828
West Midlands	-0.146	0.205	0.476
East of England	-0.255	0.210	0.225
South East	-0.291	0.192	0.130
South West	-0.230	0.226	0.309
Wales	-0.207	0.211	0.327
Scotland	0.177	0.254	0.487
Northern Ireland	0.191	0.203	0.349
Constant	1.892	0.283	0.000
Observations	547		
RMSE	0.9513		
R-squared	0.0760		

Note: Robust standard errors reported.

Table A35: Linear least squares regression estimates for log hours of informal care per week provided by others to people aged 65 and over.

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and receiving social care from a partner and without missing variables. Explanatory variables describe characteristics of person in receipt of care.

	Coef.	s.e.	p>z
Gender (ref = Women)			
Men	0.076	0.086	0.378
Education Level (ref = High)			
Medium	0.072	0.147	0.626
Low	0.239	0.147	0.105
Supplementary carer (ref = none)			
Partner	-0.186	0.093	0.047
Daughter	0.006	0.086	0.944
Son	-0.088	0.098	0.366
Formal market	0.113	0.094	0.234
Self-rated health poor	0.285	0.089	0.001
Region (Ref = London)			
North East	-0.604	0.310	0.052
North West	-0.717	0.281	0.011
Yorkshire and the Humber	-0.536	0.279	0.056
East Midlands	-0.418	0.300	0.164
West Midlands	-0.572	0.293	0.051
East of England	-0.859	0.295	0.004
South East	-0.642	0.281	0.023
South West	-0.536	0.313	0.087
Wales	-0.401	0.277	0.149
Scotland	-0.276	0.285	0.334
Northern Ireland	-0.432	0.296	0.145
Constant	1.760	0.261	0.000
Observations	585		
RMSE	0.8472		
R-squared	0.0934		

Note: Robust standard errors reported.

Table A36: Linear least squares regression estimates for log hours of formal care per week provided to people aged 65 and over.

Sample: Pooled data reported by waves "g", "i", and "k" of UKHLS, individuals aged 65 and receiving social care from a partner and without missing variables.

	Coef.	s.e.	p>z
Gender (ref = Women)			
Men	0.234	0.078	0.003
Education Level (ref = High)			
Medium	-0.015	0.108	0.890
Low	0.183	0.109	0.093
Informal carer	0.196	0.071	0.005
Self-rated health poor	0.306	0.087	0.000
Region (Ref = London)			
North East	0.016	0.272	0.954
North West	-0.010	0.199	0.961
Yorkshire and the Humber	-0.141	0.211	0.504
East Midlands	0.168	0.224	0.453
West Midlands	0.048	0.210	0.820
East of England	-0.062	0.199	0.754
South East	-0.159	0.190	0.402
South West	-0.044	0.194	0.822
Wales	-0.240	0.187	0.199
Scotland	-0.009	0.190	0.964
Northern Ireland	0.094	0.189	0.617
Constant	1.293	0.179	0.000
Observations	1,026		
RMSE	0.9433		
R-squared	0.0681		

Ote: Robust standard errors reported.

Table A37: Probit regression estimates for the incidence of informal care to non-partners among people aged 18 and over who supply informal care to their partners.

Sample: Pooled data reported between 2015 and 2020 by waves "f" to "l" of the UKHLS, individuals aged 18 and over with partners to whom they provide informal care and without missing variables.

	Coef.	s.e.	p>z
Gender (Ref = Women)			
Men	-0.100	0.0463	0.031
Education Level (Ref = High)			
Medium	0.006	0.0641	0.922
Low	-0.118	0.0715	0.100
care for partner (lag, Ref = no care)			
care only for partner	-0.135	0.0566	0.017
care for partner and non-partner	1.236	0.0688	0.000
care only for non-partner	1.253	0.0897	0.000
Self-rated health (Ref = Excellent)			
Very good	0.001	0.1030	0.995
Good	-0.005	0.0991	0.956
Fair	-0.033	0.1009	0.746
Poor	-0.007	0.1146	0.953
Age group (Ref = 18-19)			
20-24	0.472	0.4815	0.327
25-29	0.344	0.2273	0.130
30-34	0.592	0.1996	0.003
35-39	0.781	0.1789	0.000
40-44	0.641	0.1701	0.000
45-49	0.775	0.1502	0.000
50-54	0.741	0.1434	0.000
55-59	0.590	0.1422	0.000
60-64	0.436	0.1384	0.002
65-69	0.275	0.1370	0.045
70-74	0.181	0.1346	0.180
75-59	0.164	0.1402	0.243
80-84	-0.031	0.1475	0.832
85+		(omitted)	
Constant	-1.373	0.1868	0.000
Observations	6,355		
Proportion positive	0.2057		
Pseudo R2	0.2115		

Note: Robust standard errors reported. "lag" defined as preceding year. Regional dummy variables generally not significant, and omitted from table for brevity (available upon request).

Table A38: Probit regression estimates for the incidence of informal care to non-partners among people aged 18 and over who do not supply informal care to a partner.

Sample: Pooled data reported between 2015 and 2020 by waves "f" to "l" of the UKHLS, individuals aged 18 and over who do not provide informal care to a partner and without missing variables.

	Coef.	s.e.	p>z
Gender (Ref = Women)			
Men	-0.139	0.0112	0.000
Education Level (Ref = High)			
Medium	0.099	0.0128	0.000
Low	0.007	0.0181	0.714
Care for partner (lag, Ref = no care)			
care only for partner	0.259	0.0561	0.000
care for partner and non-partner	1.514	0.0744	0.000
care only for non-partner	1.806	0.0119	0.000
Self-rated health (Ref = Excellent)			
Very good	0.043	0.0193	0.024
Good	0.063	0.0195	0.001
Fair	0.082	0.0223	0.000
Poor	-0.007	0.0293	0.815
Partner	-0.107	0.0123	0.000
Age group (Ref = 18-19)			
20-24	0.106	0.0476	0.026
25-29	0.173	0.0482	0.000
30-34	0.216	0.0475	0.000
35-39	0.320	0.0459	0.000
40-44	0.342	0.0447	0.000
45-49	0.434	0.0437	0.000
50-54	0.534	0.0433	0.000
55-59	0.526	0.0431	0.000
60-64	0.483	0.0437	0.000
65-69	0.395	0.0439	0.000
70-74	0.255	0.0448	0.000
75-59	0.106	0.0482	0.028
80-84	0.005	0.0537	0.927
85+	-0.188	0.0639	0.003
Constant	-1.902	0.0473	0.000
Observations	167,458		
Proportion positive	0.1355		
Pseudo R2	0.3021		

Note: Robust standard errors reported. "lag" defined as preceding year. Regional dummy variables generally not significant, and omitted from table for brevity (available upon request).

Table A39: Probit regression estimates for the incidence of informal care among people aged 18 and over who do not have a partner.

Sample: Pooled data reported between 2015 and 2020 by waves "f" to "l" of the UKHLS, individuals aged 18 and over who do not have a partner and without missing variables.

	Coef.	s.e.	p>z
Gender (Ref = Women)			
Men	-0.093	0.0193	0.000
Education Level (Ref = High)			
Medium	0.109	0.0233	0.000
Low	0.025	0.0308	0.421
Care for partner (lag, Ref = no care)			
care only for partner	0.400	0.1061	0.000
care for partner and non-partner	1.198	0.1898	0.000
care only for non-partner	1.778	0.0202	0.000
Self-rated health (Ref = Excellent)			
Very good	-0.008	0.0333	0.807
Good	0.038	0.0333	0.260
Fair	0.076	0.0369	0.040
Poor	-0.012	0.0442	0.788
Age group (Ref = 18-19)			
20-24	0.110	0.0483	0.023
25-29	0.191	0.0537	0.000
30-34	0.261	0.0581	0.000
35-39	0.351	0.0578	0.000
40-44	0.423	0.0556	0.000
45-49	0.472	0.0517	0.000
50-54	0.499	0.0503	0.000
55-59	0.446	0.0491	0.000
60-64	0.453	0.0510	0.000
65-69	0.361	0.0515	0.000
70-74	0.291	0.0522	0.000
75-59	0.156	0.0563	0.005
80-84	0.025	0.0609	0.681
85+	-0.160	0.0689	0.021
Constant	-1.922	0.0581	0.000
Observations	61,235		
Proportion positive	0.1353		
Pseudo R2	0.2956		

Note: Robust standard errors reported.

Table A.40: Multinomial logit regression estimates for the incidence of informal care provision among people aged 18 and over with a partner. Sample: Pooled data reported between 2015 and 2020 by waves "f" to "l" of the UKHLS, individuals aged 18 and over who have a partner and without missing variables.

	only care for partner (4.9%)			care for partner and other (1.3%)			only care for other (13.0%)		
	Coef.	s.e.	p>z	Coef.	s.e.	p>z	Coef.	s.e.	p>z
Gender (Ref = Women)									
Men	-0.028	0.046	0.550	-0.194	0.075	0.010	-0.336	0.026	0.000
Education Level (Ref = High)									
Medium	0.366	0.057	0.000	0.410	0.096	0.000	0.157	0.029	0.000
Low	0.632	0.069	0.000	0.415	0.118	0.000	-0.059	0.042	0.160
Care for partner (lag, Ref = no care)									
care only for partner	4.707	0.055	0.000	4.601	0.110	0.000	0.317	0.133	0.018
care for partner and non-partner	4.549	0.120	0.000	6.771	0.134	0.000	2.742	0.129	0.000
care only for non-partner	0.404	0.099	0.000	2.561	0.113	0.000	3.198	0.026	0.000
Self-rated health (Ref = Excellent)									
Very good	0.045	0.094	0.632	0.094	0.157	0.550	0.155	0.045	0.001
Good	0.191	0.092	0.038	0.218	0.152	0.152	0.157	0.045	0.001
Fair	0.522	0.099	0.000	0.611	0.159	0.000	0.140	0.052	0.007
Poor	0.606	0.122	0.000	0.722	0.190	0.000	-0.026	0.075	0.732
Age group (Ref = under 35)									
35-44	0.069	0.123	0.574	0.292	0.213	0.171	0.296	0.055	0.000
45-54	0.251	0.116	0.030	0.572	0.192	0.003	0.626	0.052	0.000
55-64	0.651	0.112	0.000	0.554	0.192	0.004	0.701	0.052	0.000
65+	1.203	0.108	0.000	0.472	0.191	0.013	0.199	0.053	0.000
Constant	-5.068	0.162	0.000	-6.623	0.257	0.000	-3.274	0.076	0.000

Note: Robust standard errors reported. "lag" defined as preceding year. Regional dummy variables generally not significant, and omitted from table for brevity. Observations: 112,579. Pseudo R2: 0.3560. Reference group is people not providing social care. Population shares reported in brackets.

Table A41: Linear least squares regression estimates for log hours of informal care per week provided by people aged 18 and over.

Sample: Pooled data reported by waves "f" to "l" of UKHLS, individuals aged 18 and over supplying some social care and without missing variables. See table A.17 for further details.

	Coef.	s.e.	p>z
Gender (Ref = Women)			
Men	-0.260	0.0179	0.000
Education Level (Ref = High)			
Medium	0.250	0.0208	0.000
Low	0.523	0.0285	0.000
Self-rated health (Ref = Excellent)			
Very good	0.011	0.0328	0.739
Good	0.172	0.0331	0.000
Fair	0.329	0.0367	0.000
Poor	0.553	0.0477	0.000
Social care provided (Ref = care only for partner)			
care for partner and non-partner	-0.205	0.0502	0.000
care only for non-partner	-1.272	0.0278	0.000
Partner	-0.234	0.0219	0.000
Age group (Ref = 18-19)			
20-24	0.165	0.0913	0.070
25-29	0.279	0.0936	0.003
30-34	0.526	0.0926	0.000
35-39	0.597	0.0888	0.000
40-44	0.564	0.0864	0.000
45-49	0.309	0.0837	0.000
50-54	0.223	0.0818	0.006
55-59	0.196	0.0811	0.016
60-64	0.152	0.0812	0.062
65-69	0.065	0.0820	0.427
70-74	0.068	0.0833	0.414
75-59	0.071	0.0874	0.415
80-84	0.068	0.0946	0.474
85+	-0.072	0.1086	0.506
Constant	2.704	0.0933	0.000
Observations	31,490		
RSME	1.2789		
R2	0.1783		

Note: Robust standard errors reported.

Table A42: Heckman-corrected wage equation estimated on the sample of women who were not in employment in the previous year.

Log of hourly wage	(1) Wage equation coef.	(2) Wage equation s.e.	(3) Selection equation coef.	(4) Selection equation s.e.
Age	0.02	0.02	0.10***	0.01
Age squared	-0.00	0.00	-0.00***	0.00
Educational Level: 3 Category = Medium	0.01	0.08	-0.36***	0.08
Educational Level: 3 Category = Low	0.06	0.25	-1.15***	0.14
Medium education#Age	-0.01***	0.00	0.00	0.00
Low education#Age	-0.01***	0.00	0.01***	0.00
In Education: Binary	0.01	0.11	-0.63***	0.05
Mother's Educational Level: 3 Category = Medium	-0.04*	0.03	-0.01	0.04
Mother's Educational Level: 3 Category = Low	-0.06	0.05	-0.22***	0.04
Father's Educational Level: 3 Category = Medium	-0.08***	0.02	0.03	0.03
Father's Educational Level: 3 Category = Low	-0.05*	0.03	0.05	0.04
Partnered	0.13***	0.03	-0.12***	0.03
Has children	-0.01	0.11	-0.54***	0.03
Long-term Sick or Disabled	-0.09	0.38	-1.55***	0.11
Self-rated Health = 2	-0.05	0.09	0.29***	0.06
Self-rated Health = 3	-0.05	0.12	0.53***	0.06
Self-rated Health = 4	-0.01	0.14	0.66***	0.06
Self-rated Health = 5	0.05	0.15	0.67***	0.06
Government Office Region = North East	-0.08	0.06	-0.18***	0.06
Government Office Region = North West	-0.04	0.03	-0.06	0.04
Government Office Region = Yorkshire and the Humber	-0.11***	0.04	-0.10**	0.05
Government Office Region = East Midlands	-0.03	0.04	-0.03	0.05
Government Office Region = West Midlands	-0.05	0.04	-0.07	0.05
Government Office Region = East of England	0.04	0.03	-0.04	0.05
Government Office Region = London	0.07*	0.04	-0.11***	0.04
Government Office Region = South West	-0.01	0.04	0.06	0.05
Government Office Region = Wales	-0.03	0.05	-0.06	0.06
Government Office Region = Scotland	0.00	0.04	0.05	0.05
Government Office Region = Northern Ireland	-0.07	0.05	-0.08	0.06
Works part-time	0.08***	0.02		
Growth	2.09***	0.47		
Lagged Employment Status = Not Employed			-0.14***	0.04
lambda	-0.15	0.25		
Constant	0.01	0.75	-1.89***	0.16
Observations	22,171		22,171	

*** p<0.01, ** p<0.05, * p<0.1

Table A43: Heckman-corrected wage equation estimated on the sample of men who were not in employment in the previous year.

Log of hourly wage	(1) Wage equation coef.	(2) Wage equation s.e.	(3) Selection equation coef.	(4) Selection equation s.e.
Age	0.06***	0.02	0.05***	0.01
Age squared	-0.00*	0.00	-0.00***	0.00
Educational Level: 3 Category = Medium	-0.20	0.16	-0.55***	0.10
Educational Level: 3 Category = Low	-0.56	0.43	-1.43***	0.16
Medium education#Age	-0.00	0.00	0.01***	0.00
Low education#Age	-0.00	0.01	0.02***	0.00
In Education: Binary	-0.30	0.26	-0.72***	0.06
Mother's Educational Level: 3 Category = Medium	-0.03	0.04	0.04	0.05
Mother's Educational Level: 3 Category = Low	-0.15***	0.04	-0.04	0.05
Father's Educational Level: 3 Category = Medium	0.01	0.03	0.00	0.04
Father's Educational Level: 3 Category = Low	0.04	0.04	-0.01	0.05
Partnered	0.22**	0.09	0.27***	0.04
Has children	-0.04	0.03	0.01	0.04
Long-term Sick or Disabled	-0.65	0.66	-1.69***	0.12
Self-rated Health = 2	0.17	0.15	0.32***	0.08
Self-rated Health = 3	0.32	0.21	0.56***	0.08
Self-rated Health = 4	0.40*	0.24	0.66***	0.08
Self-rated Health = 5	0.41*	0.25	0.69***	0.08
Government Office Region = North East	-0.16*	0.09	-0.23***	0.07
Government Office Region = North West	-0.18**	0.08	-0.24***	0.06
Government Office Region = Yorkshire and the Humber	-0.20**	0.09	-0.27***	0.06
Government Office Region = East Midlands	-0.16**	0.08	-0.20***	0.06
Government Office Region = West Midlands	-0.14**	0.07	-0.16***	0.06
Government Office Region = East of England	-0.02	0.06	-0.10*	0.06
Government Office Region = London	-0.05	0.07	-0.20***	0.05
Government Office Region = South West	-0.17***	0.05	-0.09	0.06
Government Office Region = Wales	-0.21**	0.10	-0.25***	0.07
Government Office Region = Scotland	-0.16*	0.09	-0.22***	0.07
Government Office Region = Northern Ireland	-0.16**	0.08	-0.17**	0.08
Works part-time	0.09***	0.02		
Growth	1.37**	0.57		
Lagged Employment Status = Not Employed			0.11**	0.05
lambda	0.53	0.43		
Constant	-0.72	0.95	-1.13***	0.19
Observations	12,989		12,989	

*** p<0.01, ** p<0.05, * p<0.1

Table A44: Heckman-corrected wage equation estimated on the sample of women who were in employment in the previous year.

Log of hourly wage	(1) Wage equation coef.	(2) Wage equation s.e.	(3) Selection equation coef.	(4) Selection equation s.e.
Lagged log hourly wage	0.68***	0.00		
Age	0.02**	0.01	0.17***	0.01
Age squared	-0.00**	0.00	-0.00***	0.00
Educational Level: 3 Category = Medium	-0.07***	0.02	-0.10	0.07
Educational Level: 3 Category = Low	-0.11*	0.06	-0.87***	0.17
Medium education#Age	-0.00***	0.00	0.00	0.00
Low education#Age	-0.00*	0.00	0.01***	0.00
In Education: Binary	-0.04	0.04	-0.50***	0.07
Mother's Educational Level: 3 Category = Medium	-0.02**	0.01	0.09***	0.03
Mother's Educational Level: 3 Category = Low	-0.04***	0.01	0.09**	0.04
Father's Educational Level: 3 Category = Medium	-0.02***	0.01	0.02	0.03
Father's Educational Level: 3 Category = Low	-0.02***	0.01	-0.04	0.03
Partnered	0.01***	0.00	-0.01	0.02
Has children	-0.02**	0.01	-0.23***	0.02
Long-term Sick or Disabled	-0.06	0.15	-1.78***	0.11
Self-rated Health = 2	0.01	0.02	0.37***	0.06
Self-rated Health = 3	0.03	0.03	0.45***	0.06
Self-rated Health = 4	0.05*	0.03	0.50***	0.06
Self-rated Health = 5	0.06**	0.03	0.48***	0.06
Government Office Region = North East	-0.03***	0.01	0.13**	0.05
Government Office Region = North West	-0.03***	0.01	-0.02	0.03
Government Office Region = Yorkshire and the Humber	-0.03***	0.01	0.01	0.04
Government Office Region = East Midlands	-0.04***	0.01	0.04	0.04
Government Office Region = West Midlands	-0.02**	0.01	0.04	0.04
Government Office Region = East of England	-0.00	0.01	0.10***	0.04
Government Office Region = London	0.02***	0.01	-0.08**	0.03
Government Office Region = South West	-0.02**	0.01	0.01	0.04
Government Office Region = Wales	-0.03***	0.01	0.05	0.05
Government Office Region = Scotland	-0.01	0.01	0.06	0.04
Government Office Region = Northern Ireland	-0.03***	0.01	0.09*	0.05
Works part-time	0.02***	0.00		
Growth	0.36***	0.10		
lambda	0.02	0.14		
Constant	0.13	0.21	-2.29***	0.15
Observations	40,572		40,572	

*** p<0.01, ** p<0.05, * p<0.1

Table A45: Heckman-corrected wage equation estimated on the sample of men who were in employment in the previous year.

Log of hourly wage	(1) Wage equation coef.	(2) Wage equation s.e.	(3) Selection equation coef.	(4) Selection equation s.e.
Lagged log hourly wage	0.69***	0.00		
Age	0.02***	0.01	0.11***	0.01
Age squared	-0.00**	0.00	-0.00***	0.00
Educational Level: 3 Category = Medium	-0.08***	0.03	-0.35***	0.08
Educational Level: 3 Category = Low	-0.23***	0.07	-1.13***	0.15
Medium education#Age	-0.00	0.00	0.01***	0.00
Low education#Age	0.00	0.00	0.02***	0.00
In Education: Binary	-0.09	0.08	-0.78***	0.08
Mother's Educational Level: 3 Category = Medium	-0.00	0.01	-0.03	0.04
Mother's Educational Level: 3 Category = Low	-0.03***	0.01	-0.05	0.04
Father's Educational Level: 3 Category = Medium	-0.02***	0.01	0.03	0.03
Father's Educational Level: 3 Category = Low	-0.03***	0.01	-0.02	0.04
Partnered	0.06***	0.01	0.21***	0.03
Has children	-0.00	0.01	-0.09***	0.03
Long-term Sick or Disabled	0.03	0.19	-1.80***	0.13
Self-rated Health = 2	-0.02	0.02	-0.03	0.08
Self-rated Health = 3	0.01	0.02	0.07	0.08
Self-rated Health = 4	0.03*	0.02	0.11	0.08
Self-rated Health = 5	0.04**	0.02	0.10	0.08
Government Office Region = North East	-0.05***	0.01	-0.00	0.06
Government Office Region = North East	-0.05***	0.01	-0.01	0.04
Government Office Region = Yorkshire and the Humber	-0.05***	0.01	-0.14***	0.04
Government Office Region = East Midlands	-0.05***	0.01	-0.04	0.04
Government Office Region = West Midlands	-0.03***	0.01	-0.05	0.04
Government Office Region = East of England	-0.02*	0.01	-0.03	0.04
Government Office Region = London	-0.02	0.01	-0.15***	0.04
Government Office Region = South West	-0.05***	0.01	-0.02	0.04
Government Office Region = Wales	-0.07***	0.01	-0.08	0.05
Government Office Region = Scotland	-0.04***	0.01	0.04	0.05
Government Office Region = Northern Ireland	-0.07***	0.01	0.04	0.06
Works part-time	0.12***	0.01		
Growth	0.33***	0.12		
lambda	0.06	0.18		
Constant	0.12	0.19	-0.52***	0.17
Observations	33,567		33,567	

*** p<0.01, ** p<0.05, * p<0.1

Appendix B Agent focussed simulations; a formal sketch

Mathematically, dynamic microsimulation models are Markov chains, where at each time *t* an agent $i \in \{1, \dots, N\}$ is fully described by some state variables $\boldsymbol{x}_{i,t} \in \mathbb{R}^K$. The evolution of her (vector of) state variables is specified by the difference equation:

$$\boldsymbol{x}_{i,t+1} = \boldsymbol{f}_i(\boldsymbol{x}_{i,t}, \boldsymbol{x}_{-i,t}, \boldsymbol{\theta}, \boldsymbol{P}_t, \boldsymbol{\xi}_{i,t}) \quad (1)$$

where $\boldsymbol{\theta}$ is a vector of behavioural parameters, $\boldsymbol{P}_t$ are time-varying environmental parameters (including current and announced or expected future policies), and $\boldsymbol{\xi}_{i,t}$ are stochastic disturbances. Individual outcomes can also depend on the state variables of other agents $\boldsymbol{x}_{-i,t}$, for instance their partners or children.

Structural modelling, in this context, refers to the parameters $\boldsymbol{\theta}$ governing behaviour – for instance those describing utility functions – being policy-invariant. Expectations about the future are accommodated in the notation as they can be expressed as a function of the state variables $\boldsymbol{x}$ and the policy parameters $\boldsymbol{P}$. Realism in the policy description requires $\boldsymbol{P}$ to be a fairly detailed mapping from real-world policy environment. Finally, the notation can easily be generalised from partial equilibrium approaches – where there are only specific types of agents in the economy (say, individuals but not firms) – to general equilibrium approaches – where there are more agent types *i,j,h,…* each defined by their own state variables $\boldsymbol{x}_{i,t}, \boldsymbol{x}_{j,t}, \boldsymbol{x}_{h,t}$ ... possibly depending on the state variables of all other agents of any type.

In this context, interaction between different life domains is simply defined as lagged variables pertaining to one domain having a causal impact on the evolution of other domains. Consider for instance health (h) and employment (e) and suppose their law of motion is specified as follows: [27]

$$h_{i,t+1} = h(h_{i,t}, e_{i,t}, \dots, \boldsymbol{\theta}_h, \boldsymbol{P}_t, \boldsymbol{\xi}_{i,t}) \quad (2)$$

$$e_{i,t+1} = e(e_{i,t}, h_{i,t}, \dots, \boldsymbol{\theta}_e, \boldsymbol{P}_t, \boldsymbol{\xi}_{i,t}) \quad (3)$$

Health status at time *t* affects both health and employment outcomes at time *t*+1, and similarly for employment status at time *t.* The structure is similar to micro-level Dynamic Factor Models (Altonji et al., 2022; Barigozzi and Pellegrino, 2023), with the added flexibility associated to the algorithmic nature of the simulation approach.

[27] The example easily generalises to more domains.

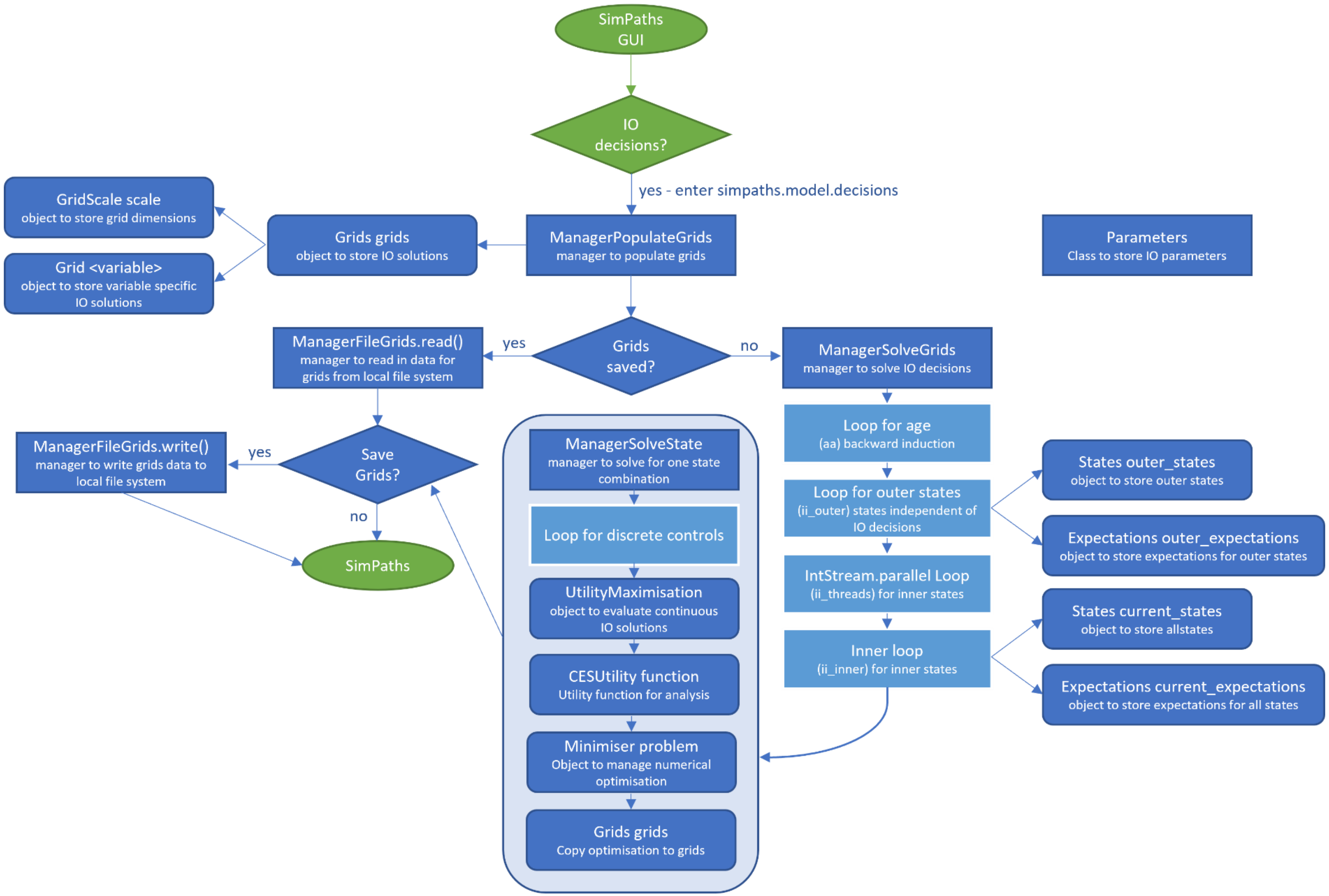
Appendix C Dynamic programming methods in SimPaths

Projections based on the DP approach proceed in two discrete stages. In the first stage, the model evaluates a look-up table that describes utility maximising decisions for all feasible simulated combinations of individual specific circumstances (the model state-space). In the second stage, starting from data for a reference population cross-section, the model projects panel data at discrete intervals over the simulated time-horizon. These panel data are generated using statistical descriptions for the intertemporal evolution of individual specific characteristics and the behavioural descriptions evaluated in the first stage. Importantly, the utility maximising decisions evaluated in the first stage are based on the same statistical descriptions for intertemporal evolution as are used in the second stage of the projection. It is this feature that makes the projected decisions 'rational'.

The key to the DP approach is that it seeks to obtain a complete solution to the lifetime decision problem at each point in time – that is for all feasible combinations of characteristics at a point in time – before it proceeds to consideration of earlier time periods. This systematic consideration of the entire state-space is important because it does not impose any conditions on statistical out-turns associated with earlier time periods.

Solution of the lifetime decision problem – the 'first stage' of the simulation noted above – is evaluated by a dedicated Java package in SimPaths: simpaths.model.decisions. This appendix provides technical detail of that program package.

Figure C.1: Flow chart of Java package to evaluate solutions to dynamic programming problem



Green elements are parts of SimPaths not involved in solution of lifetime decision problem. Manager methods in dark blue rectangles with square corners – these provide logic to organise the computations. Objects are denoted by dark blue rectangles with rounded corners. Light blue rectangles denote computational loops.

Figure C.1 displays a schematic of the *decisions* package, which proceeds as follows:

1. The user chooses to implement intertemporal optimising (IO) decisions via the SimPaths GUI (Graphical User Interface).
    a. SimPaths routes work to the simpaths.model.decisions package
2. ManagerPopulateGrids
    a. This class is responsible for creating and populating the look-up table used to simulate IO decisions – this table is referred to as "the grids".
3. ManagerFileGrids
    a. Reads and writes data for the grids to and from the file system
4. ManagerSolveGrids
    a. This class is responsible for managing evaluation of the IO solutions and storing these in the grids.
    b. The solution proceeds via a series of concentric loops.
    c. In the inner-most loop,
        i. the state combination is defined by object: States current_states
        ii. expectations are defined by object: Expectations expectations
5. ManagerSolveState
    a. This class manages numerical optimisation of all control variables for a given combination of state characteristics, as supplied by ManagerSolveGrids.
        i. Search over discrete control variables (labour options) is conducted via an outer set of loops in the ManagerSolveState class.
        ii. Search over continuous control variables (consumption) is passed to a dedicated UtilityMaximisation object.
6. UtilityMaximisation
    a. This class defines:
        i. the function to optimise
            1. Defined as the CESUtility object by default
        ii. upper bounds for the control variables
        iii. lower bounds for the control variables
        iv. a set of control variables to start the numerical search
    b. The optimisation problem is then passed to a generic Minimiser object for evaluation.
7. Minimiser
    a. Is instantiated with the factors defined by the UtilityMaximisation class
    b. Runs numerical optimisation routines via a call to the minimise() method
8. Minimiser.minimise()
    a. Passes the optimisation problem to:
        i. the brent() method if optimising over a single continuous control
            1. e.g. consumption only
        ii. the powell() method if optimising over 2 or more continuous controls
            1. e.g. consumption and portfolio allocation
9. CESUtility
    a. Is accessed by Minimiser via the IEvaluation interface, to facilitate testing of alternative utility specifications.
    b. The CESUtility object is instantiated with Expectations and Grid objects supplied by the UtilityMaximisation class
        i. The Expectations object describes expectations reflecting all but the decisions described by the continuous controls over which the Minimiser object conducts its search.

   ii. The Grid object records valueFunction solutions obtained via preceding age-specific loops evaluated by the ManagerSolveGrids class.

c. Calls to the CESUtility.evaluate(double[] args) method returns a (real number) variable describing (minus) the expected lifetime utility associated with the set of continuous control variables listed in the "args" array.
   i. This result is generated by combining within-period utility, with expected utility, via an intertemporal CES function.
      1. The within-period measure of utility is a simple CES function of consumption and leisure time associated with the prevailing combination of control variables (consumption and employment)
      2. Expected utility is evaluated by:
         a. identifying the set of expected states in the immediately succeeding period associated with the prevailing set of control variables (based on the Expectations object)
         b. identifying the value function outcome associated with each set of expected states, via a call to the Grid.interpolation(States) method for the valueFunction attribute.
         c. aggregating up the measures of the value function, by weighting each by its associated probability
d. Calls to the Grid.interpolateAll(States states, boolean solution_call) method return a (real number) variable, by interpolating over the respective Grid object.
   i. The interpolation begins by identifying a grid slice for all continuous states associated with the combination of discrete states described by the "states" object supplied to the method.
      1. The (Boolean) "solutionCall" variable is used to determine whether the birth year state is considered to be a discrete or continuous state for the interpolation routine.
   ii. Interpolation over the set of continuous states described by the "states" object supplied to the interpolateAll method is evaluated by the interpolateContinuous method.
      1. The interpolateContinuous method implements a linear spline interpolation

Evaluation of solutions to the dynamic programming problem are organised by a series of "manager" classes, which are described at further length below.

C.1 ManagerPopulateGrids

ManagerPopulateGrids is the highest-level manager class in the decisions package, providing the entry and exit point of the package. ManagerPopulateGrids instantiates the "grids" object that stores solutions to the lifetime decision problem. The manager then organises for the "grids" object to be populated, either by delegating solution of the lifetime decision problem to the ManagerSolveGrids Class, or delegating reading from the file system to the ManagerFileGrids class. Finally, ManagerPopulateGrids organises for the populated grids object to be saved to the file system, via another reference to ManagerFileGrids.

C.2 ManagerSolveGrids

ManagerSolveGrids is called by ManagerPopulateGrids if new solutions to the IO problem are required. ManagerSolveGrids organises solutions to the IO problem using four concentric loops.

The first loop (aa) proceeds backward from the last potential age in life, to the first age at which an individual is considered to enter the model as a responsible adult of a benefit unit. This backward iterating loop allows the solution to proceed via backward induction.

All state characteristics other than age are divided into two groups, considered in either an “inner” or “outer” loop. Outer loop characteristics are treated in the first loop following age (ii_outer). These characteristics are predominantly comprised of discrete variables that are exogenous to IO decisions (control variables). Consideration of these variables in a separate loop is useful because it allows their state combinations and associated expectations to be evaluated once and re-used for all of the state combinations considered within the inner set of loops.

The “inner” states are divided into chunks that are iterated over by a parallel loop (IntStream.parallel) to make use of multi-core processing. Inner states are grouped into chunks helps to economise the computational overhead associated with creation and destruction of worker threads.

Combinations of states are recorded by ManagerSolveGrids in objects of the States class. State combinations identified in the outer grid are stored in the object outerStates, and these are used to initialise state combinations identified in the inner loops: States currentStates. A similar approach is used to manage state expectations, via objects of the Expectations class.

C.3 ManagerSolveState

A solution needs to be obtained for utility maximising decisions at each grid ordinate visited via the loop structure of ManagerSolveGrids. This solution is obtained for an assumed utility function, and expectations consistent with the intertemporal dynamics used to project states (individual characteristics) through time.

The code starts from a prevailing set of individual specific characteristics, as supplied by the ManagerSolveGrids class. Each potential discrete decision (control variable, e.g. labour alternative) is considered in turn. For each discrete decision, numerical methods are used to optimise expected lifetime utility with respect to the set of continuous decision variables (e.g. consumption). A preferred set of decision variables is then identified as that with the highest overall measure of expected lifetime utility.

Expected lifetime utility is evaluated in two components. The first and most straight-forward is (current) within-period utility, which is evaluated as a CES function of current period consumption and leisure (the corollary of employment). The second component is expected utility for all periods following the current period. Expected lifetime utility at age A+1, from age A, is evaluated as a weighted sum of a discrete set of alternative possibilities calculated previously by the solution routine. This is made possible by the following features of the solution method:

- Starting with the maximum potential age, and iteratively solving backwards through time.
- Assumption of a von Neumann Morgenstern utility function.
- Use of the Gaussian quadrature to approximate summation over continuous normal distributions via a discrete set of weights and abscissae.
- Use of linear interpolation for approximating off-grid solutions (Keys, 1981).

The numerical optimisation method is based on value function calls rather than first order conditions as the value function is not guaranteed to be smooth or concave, and the computational overhead associated with evaluating first order conditions can outweigh advantages of zero-search algorithms.[28] Brent’s method is

[28] This observation was obtained using a model structure that included alternative solutions procedures; see van de Ven (2011).

used to search over a single (continuous) dimension, and Powell's method to search over multiple dimensions (see Press *et al*., 2007).

C.3.1 Dimensionality of the grids object

Key features assumed for each of the states considered for analysis are listed here.

- *Scale* describes the scale used to describe the respective state in the decision grids.
- *Loop* indicates the loop structure (inner/outer) used to represent the characteristic when solving the IO problem
- *Endogenous* indicates whether or not evolution of the respective state is permitted to depend upon IO decisions (control variables)
- *Uncertain* indicates whether or not the respective state is considered to evolve stochastically when solving the IO problem
- *Dynamics* summarises the intertemporal dynamics assumed to solve the IO problem.

The order of the list reflects the assumed grid structure, as set out in the Grids class.

- Net wealth
  - Scale: Continuous, adjusted logarithmic
  - Loop: inner
  - Endogenous: yes
  - Uncertain: no
  - Dynamics: Follows an accounting identity, where wealth in next period is equal to wealth in current period plus disposable income less consumption.
- Wage potential
  - Scale: Continuous, adjusted logarithmic
  - Loop: inner
  - Endogenous: yes
  - Uncertain: yes
  - Dynamics: Based on estimated latent wage equation.
- Private pension
  - Scale: Continuous, adjusted logarithmic
  - Loop: inner
  - Endogenous: yes
  - Uncertain: no
  - Dynamics: Assumed fraction of net wealth converted to a fixed life annuity upon retirement.
- Health status
  - Scale: Continuous indicator variable
  - Loop: outer
  - Endogenous: no
  - Uncertain: yes
  - Dynamics: Based on a linear regression equation
- Birth cohorts (year of birth)
  - Scale: Discrete for IO solutions, continuous for projections
  - Loop: outer
  - Endogenous: no
  - Uncertain: no
  - Dynamics: none
- Retirement status

- Scale: Discrete, distinguishing between those in and out of retirement
- Loop: outer
- Endogenous: yes
- Uncertain: no
- Dynamics: Entry to retirement is non-reversible, and occurs in the first period of non-employment beyond a "minimum age of retirement"

- Disability status
  - Scale: Discrete, distinguishing those affected by disability
  - Loop: outer
  - Endogenous: no
  - Uncertain: yes
  - Dynamics: Based on an estimated probit regression
- Region
  - Scale: Discrete
  - Loop: outer
  - Endogenous: no
  - Uncertain: no (ignored)
  - Dynamics: none (ignored)
- Student status
  - Scale: Discrete
  - Loop: outer
  - Endogenous: no
  - Uncertain: yes
  - Dynamics: Based on an estimated probit regression
- Education attainment
  - Scale: Discrete
  - Loop: outer
  - Endogenous: no
  - Uncertain: yes
  - Dynamics: Education assigned at transition from student status and otherwise remains invariant.
- Number and age of dependent children
  - Scale: Discrete number of 'birth ages', with discrete number of children permitted per birth age
  - Loop: outer
  - Endogenous: no
  - Uncertain: yes
  - Dynamics: Scaled to reflect fertility probabilities described by estimated probit regressions
- Cohabitation status
  - Scale: Discrete, distinguishing single/couple
  - Loop: outer
  - Endogenous: no
  - Uncertain: yes
  - Dynamics: Based on estimated probit regressions
- Gender
  - Scale: Discrete, distinguishing male/female
  - Loop: outer
  - Endogenous: no

- Uncertain: no
- Dynamics: none

- Age
  - Discrete: Annual increments
  - Loop: first (before both outer loop, which is before inner loop)
  - Endogenous: no
  - Uncertain: no
  - Dynamics: age next period equals age this period + 1

C.4 ManagerFileGrids

There are a wide range of methods available for reading and writing data to disk available in Java. Some of the available approaches are legacy methods that have been superseded by newer ones. Nevertheless, there is no single method that is most efficient to apply in all contexts, which complicates design. In the current context, we seek the quickest method for reading and writing large double formatted arrays. For our use case, two methods currently stand out:

- BufferedOutputStream with byte arrays
- FileChannel with direct byte buffer

Of these two methods, FileChannel was selected for the ManagerFileGrids class.