Inheritances, expectations and intergenerational mobility*

Pascale Bourquin[†] Robert Joyce[‡] David Sturrock[§]

February 16, 2022

Abstract

Over the past 40 years, inheritances have grown as a share of national income in many advanced economies and looks likely to grow further in the coming decades. Inheritances are both highly unequally distributed across individuals and, for a given individual, uncertain in both their size and timing. What are the implications of this source of lifetime income risk for the distributions of consumption and wealth over the lifecycle and for the consumption and wealth mobility? This paper gives a quantitative answer to these questions. We estimate a model of wealth accumulation over the lifecycle in the presence of a parental wealth and inheritance process and us it to decompose the impact of inheritances on consumption and wealth inequality. Inheritances have a minor impact on wealth inequality but are of growing importance for several measures of intergenerational mobility.

Keywords: Wealth inequality, inheritances

*This work was funded by the Nuffield Foundation and the Economic and Social Research Council (ESRC) through the ESRC Centre for the Microeconomic Analysis of Public Policy (CPP) (grant reference ES/M010147/1)). We thank James Banks, Richard Blundell and Eric French for helpful comments.
Correspondence to david.sturrock@ifs.org.uk.

[†]Institute for Fiscal Studies

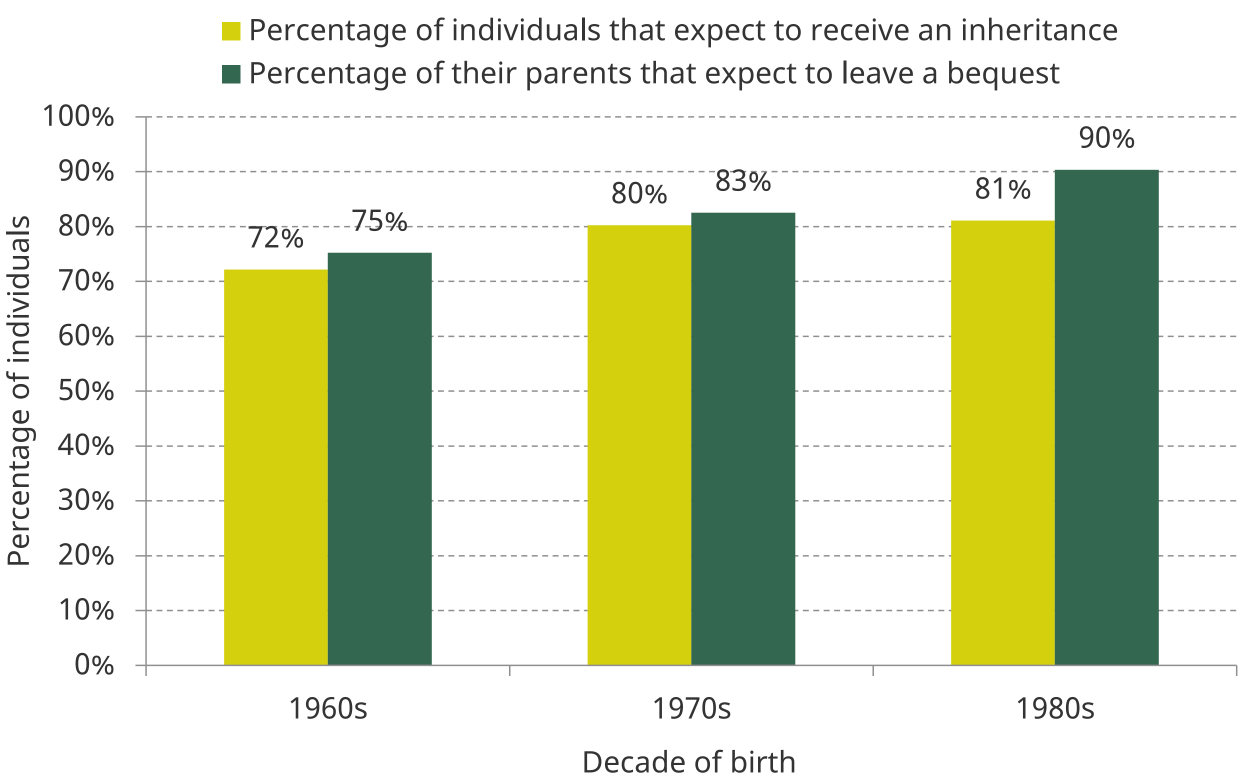
[‡]Institute for Fiscal Studies

[§]University College London and Institute for Fiscal Studies

1 Introduction

The past 40 years have seen rising private wealth-to-income ratios drive increasing flows of intergenerational wealth transfers in many advanced economies (Alvaredo et al., 2017). In the UK, intergenerational transfers of wealth have risen from 4.8% of national income in 1977 to 8.2% of national income in 2006, making them equivalent in size to annual private pension payments (Atkinson, 2018). This trend is set to continue, and the significance of inheritances in lifetime income for those of working-age today is expected to rise further due to historically slow rates of earnings growth and the increasing share of wealth held by older households. Such a trend is expected by households. While 40% of those born in the 1930s received an inheritance 80% of those born in the 1970s report expecting to inherit (Bourquin et al., 2020).

Figure 1: Inheritance and parental bequest expectations by birth cohort



Source: Bourquin et al. (2020)

Inheritances are not only growing in size but are both highly unequally distributed across individuals and, for a given individual, uncertain in both their size and timing. Amongst those born in the 1930s and 1940s, while just over half of households did not inherit anything, 13% of households inherited more than £100,000 per person (Hood and Joyce, 2017). There are likely to be similar inequalities in the inheritances to be received in the coming decades. While 20% of individuals born in the 1980s have parents who have wealth of less than £10,000 per-heir, 25% have parents holding more than £250,000 per-heir. Given the increasing importance of this source of income relative to earnings from work and the inequality in the size of inheritances, this raises the question: what are the implications of inheritances for the distributions of consumption and wealth over the lifecycle and for intergenerational mobility?

This paper gives a quantitative answer to this question for the 1960s, 1970s and 1980s birth cohorts in the UK. Our approach is to employ a lifecycle model that features a realistic parental wealth and inheritance process and is calibrated to match microdata on the distributions of wealth and consumption and the reduced-form response of households to the receipt of inheritances. In the model, households face uncertainty about their own earnings, the future evolution of their parents' wealth and the timing of their parents' death. When a parental household dies, their after-tax estate is split between their heirs and received as an inheritance. Households make consumption and savings decisions in the awareness of their parents' level of current wealth and the uncertainty over the future inheritance they may receive. We study the effect of inheritances on consumption and wealth inequality over the lifecycle by using the model to simulate counterfactuals in which inheritances are equalised or eliminated. We can decompose the impact of inheritances by studying counterfactual simulations that isolate the effects of heterogeneity across individuals and uncertainty within individuals in the size and timing of inheritances and by separating out the impact of behavioural responses.

The contribution of this paper is to quantify the implications of inheritances for inequalities over the entire lifecycle in a rich environment that allows for behavioural responses to the receipt, and expectation of receipt, of inheritances. Given that inheritances may be anticipated to some extent, but remain uncertain in their size and timing until received, an analysis of their contribution to inequality and intergenerational mobility must seek to account for the processes generating inheritances and expectations of these processes on the part of potential heirs. To give one example, if those who expect to receive an inheritance reduce their saving rate in anticipation of it, then an analysis that fails to account for this behavioural response may overstate the impact of inheritances on wealth inequality later in the lifecycle but fail to capture an effect on consumption inequality earlier in the lifecycle. Secondly, given the important secular trends in the economic conditions, including the rising share of wealth held by older generations and slowing rates of cohort-upon-cohort growth in earnings, any analysis must contend with these important non-stationarities in the economic environment and how they may affect expectations and behaviour. We use rich microdata on household wealth at older ages to estimate stochastic processes for the evolution of wealth that capture variation in the distribution of shocks to wealth that are experienced by households of different types and at different points of the wealth distribution. In conjunction with estimated parental survival survival curves, our estimated wealth processes are embedded within a lifecycle model of consumption, savings and idiosyncratic earnings risk.

A number of papers have attempted to quantify the contribution of inherited wealth to the level and distribution of wealth and lifetime income (see Kotlikoff and Summers (1981); Kotlikoff and Kotlikoff (1988); Modigliani (1988); Gale and Scholz (1994) for the case of the US and Elinder et al. (2018) in the case of Sweden). These papers have typically compared inherited amounts to the stock of lifecycle savings. In the UK, Karagiannaki (2017) uses the British Household Panel Survey to compare the flow of reported inheritances over a 10-year period to the wealth of inheritors. Crawford and Hood (2016) uses data from the English Longitudinal Study of Ageing on lifetime receipt of inheritances and gifts and concludes that they have a negligible impact on the distribution of wealth, once public pension wealth is included (and are slightly equalising otherwise). These papers take the savings decisions of households as given and look at the distribution of wealth for a particular cohort, or set of cohorts, typically later in life. We go beyond this by first considering behavioural changes due to inheritance receipt and, as then makes sense once allowing for dynamics effects, by considering the effects on the distributions of wealth and consumption across the whole lifecycle.

In examining behavioural responses to the receipt and expectation of receipt of inheritances, we contribute to a reduced-form literature on this topic. Holtz-Eakin et al. (1993) provided the first evidence of the so-called 'Carnegie effect' whereby inheritors reduce their labour supply after receipt of an inheritance. Elinder and Erixson (2012) finds similar effects using Swedish data. We contribute by providing the first such estimates for the UK and by placing them in the context of a lifecycle model that can generate these reduced form responses. If our ultimate concern is with the contribution of inheritances to inequality in the lifetime distributions of income, consumption and wealth, then going beyond the time horizon of the years surrounding the inheritance receipt to examine the whole lifecycle as we do is crucial.

Two other papers have embedded inheritances within a lifecycle model and studying the implications for the distributions of consumption and wealth. De Nardi (2004) and De Nardi and Yang (2016) calibrate general-equilibrium overlapping generations models with bequests. The former paper shows the voluntary and accidental bequests are important for generating the distribution of estates and that a bequest motive is important in rationalising the distribution of savings rates. The latter paper considers the effects of bequest taxation. While these two papers model overlapping generations and the effect of savings rates on capital accumulation, allowing them to consider the general equilibrium effects of estate tax policies, for example, this comes at the cost of more stylised set up that is confined to studying the steady state of the model. While our model is in partial equilibrium, we can examine the implications of long-term trends in the distributions of earnings and wealth within and between generations. Given the pronounced trends outline above, these are potentially very important to incorporate.

Finally, Kindermann et al. (2020) examines the implications of the taxation of bequests for the tax revenue raised from heirs in a lifecycle model that includes consumption, savings and labour supply responses and calibrates this to match the German economy and estimates of the labour supply response to wealth shocks. This paper finds that labour supply responses to the receipt of inheritances imply a positive fiscal externality of increasing inheritance tax of around 8 cents on the dollar. While their model includes some of the features that we examine, their main object of interest is the fiscal externality rather than the effect of inheritance on inequality. Perhaps in reflection of these different aims, our model features a much more rich environment in terms of the evolution of parental wealth and the expectations of heirs.

This paper is organised as follows. Section 2 sets out the reduced-form estimates of the effect of inheritances on wealth and labour supply that motivate our analysis and serve later as validation of our lifecycle model. Section 3 sets out the lifecycle model and Section 4 sets out the identification and estimation of the parameters of the earnings and wealth processes in the model. Section sets out the parameterisation of other features of the lifecycle model and how it is solved. Section 6 discusses the planned results and Section **??** sets out next steps.

2 Reduced-form evidence

In this section, we set out the reduced-form analysis of the receipt of inheritances that both motivates our lifecycle model analysis and is used to validate the predictions of the model. Our aim is to examine the short-run effect of receiving an inheritance on household wealth and labour supply. We implement an event-study type research design where we exploit variation in the timing of the receipt of inheritances, conditional on covariates, to identify the effect of inheritance receipt.

We draw on data from the UK Office for National Statistics' Wealth and Assets Survey (WAS), a biennial panel survey that has run over 2-year periods from 2006-08 onwards. This survey elicits information about individuals' and households' levels of wealth and earnings. In each wave from the second wave of the data onwards, WAS asks individuals if they have received an inheritance over the past two years and, if so, the value of the inheritance that they received. We treat the reporting of the receipt of an inheritance over the past 2 years as an event. The sample we use is all observations of individuals or couples whose average year of birth is between 1950 and 1979. These individuals (couples) therefore have an (average) age of between 27 and 58 in the first wave of our data.

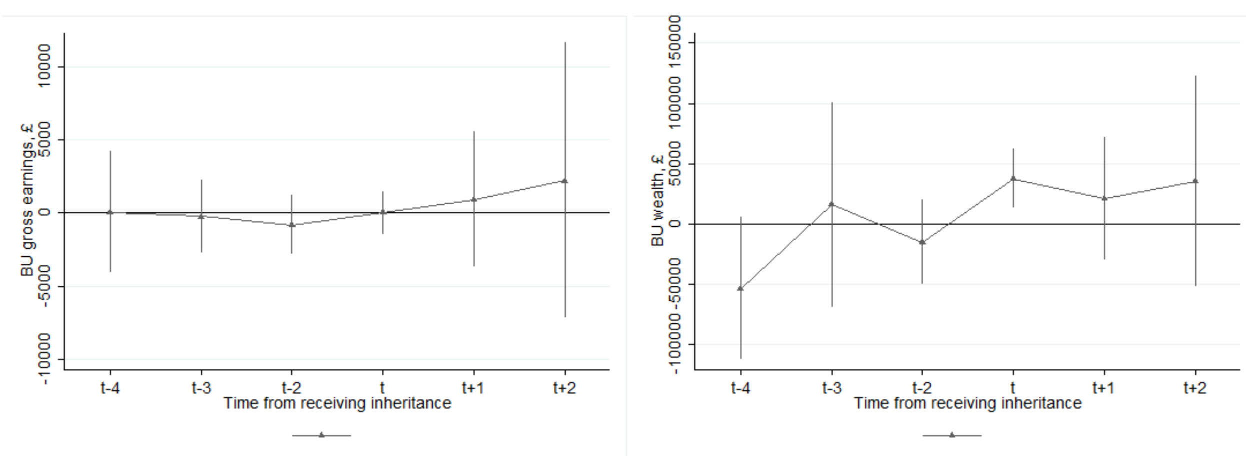
For each household, we can construct a series of indicator variables for whether they reported receiving an inheritance at certain number of leads or lags. The outcomes that we examine are total household earnings, household labour supply (measured as an indicator for at least one person being in work) and wealth (the sum of net housing wealth, household private pension wealth and household net financial assets).

The relationship that we estimate is set out in Eq. (1). Our outcome is denoted $y_{i,t}$. The variables of the form $I_{i,t-\tau}$ are a series of indicator variables for having received an inheritance τ periods ago. Our coefficients of interest are β_τ, the effect of receiving an inheritance τ waves ago. We exclude the first lead (i.e. β_{-1} is set to zero) such that the interpretation is the effect of inheritance receipt relative to the period before receipt. We include a vector of controls $X_{i,t}$ consisting of 10-year birth cohort, age in 5-year categories, education and marital status. We include also a dummy variable is dummy for whether a household has ever reported an inheritance α^I. In this sense our setup is non-standard because it includes household that have never received an inheritance. We do this due to the gain in precision in our estimates, given the modest sample of household that receive an inheritance. Clearly, the inclusion of these households has the potential to bias our results if there is a different relationship between our outcomes and the control variables amongst those who do not inherit as compared to those who do inherit and these covariates are systematically correlated with both inheritance receipt and our outcomes of interest within the inheriting group. Finally, $\epsilon_{i,t}$ is the idiosyncratic error term. We estimate a probit for labour supply, OLS for wealth and earnings. We cluster standard errors at the household level and weight observation using the given sample weights.

$$y_{i,t} = \alpha^I + \sum_{\tau=-4}^{\tau=+3} \beta_\tau I_{i,t-\tau} + \sum_{t=1}^{t=5} \gamma_t + X_{i,t} + \epsilon_{i,t} \tag{1}$$

Figure 2 shows our results in graphical form. We plot the coefficients on the number of waves to or since receipt of an inheritance, with $t-1$ excluded. Our analysis lacks power but the estimated coefficients on annual earnings are negligible (£200 or less at each time horizon) whereas the effects on wealth are large. In the first observation after receipt, wealth is £40,000 higher on average and this effect is significant at the 5% level. The estimated effects are of a similar magnitude at longer time horizons after receipt, although not statistically significantly different from zero. We can see that precision falls at longer time horizons and in part this may be due to the smaller samples that remain in the panel for multiple years.

Figure 2: Estimated effects of receipt of inheritance on earnings and wealth, relative to period before receipt of inheritance



From this initial analysis, we take that there is evidence of a reaction to receipt on inheritances on the savings margin but no evidence of a response on the labour supply margin. This justifies our modelling choices in the next section.

3 Model

In this section we set out the heterogeneous agents lifecycle model that we employ. The agent is a household that is endowed with a level of education (either attains compulsory schooling attendee only i.e. GCSEs; completed high school i.e. achieved A-levels, or completed higher education) and begins life at age 26, faces survival risk from age 65 and dies at latest at age 110. In the case where a household is a couple, this education level can be interpreted as the highest educational level of the two couple members and household death as the death of the second member of the couple to die. The household receives exogenously given earnings from work and a state pension in retirement and makes consumption (equivalently, savings) decisions in each period.

Into this standard setting we introduce that each household has two parental households. The parental households have a level of wealth whose value evolves according to an exogenous process. The timing of the death of the parental households is uncertain and when they die they may leave an inheritance which is split equally between their heirs.

We now set out more detail on each aspect of the model.

3.1 Preferences and demographics

The household is assumed to have constant relative risk aversion utility, defined as a the following function of consumption, $c_{i,t}$, within each period:

$$u_t(c_{i,t}) = \theta_t \frac{(c_{i,t}/\theta_t)^{1-\gamma} - 1}{1-\gamma} \tag{2}$$

where θ_t is the equivalisation factor. The household faces mortality risk in each year from age 65 onwards. The unconditional probability of survival to age t is given by S_t and the probability of survival to t conditional on survival to $t-1$ is given by s_t. The household receives warm glow utility from bequests, b_i according to the function (following De Nardi (2004))

$$\phi(b_i) = \phi_1 \frac{(b_i + \phi_2)^{(1-\gamma)}}{1-\gamma} \tag{3}$$

where ϕ_1 determines the strength of the bequest motive and ϕ_2 the degree to which bequests are a luxury. Lifetime utility if therefore given by

$$U_i = E_0 \left[\sum_{t=0}^{t=110} S_{t-1} \beta^t [s_t u_t(c_{i,t}) + (1-s_t)\phi(b(a_t))\right] \tag{4}$$

where β is the discount factor.

3.2 Exogenous processes

Earnings

The household receives earnings in each period from age 26 to the latest retirement age, K. Earnings household i in period t are denoted $e_{i,t}$. Earnings can be zero, representing voluntary or involuntary unemployment or early retirement. We denote the household's binary employment status as $E_{i,t}$. Log earnings conditional on working are the sum of an education-specific deterministic age component, $f^e_{ed}(age_{i,t})$ and a persistent stochastic component, $\eta_{i,t}$:

$$\begin{aligned} ln(e_{i,t}) &= f^e_{ed}(age_{i,t}) + \eta_{i,t} \qquad E_{i,t} = 1 \\ e_{i,t} &= 0 \qquad\qquad\qquad\qquad\quad E_{i,t} = 0 \end{aligned}$$

Following the framework of Arellano et al. (2017), we assume that the stochastic component follows a first order Markov process. The persistent earnings component is drawn from a distribution that varies with age and with its lagged value and is given by the series of conditional quantile functions $Q_t(\cdot|\eta_{i,t-1})$. Employment status is assumed to be drawn from a distribution that varies by age and lagged employment status and lagged persistent earnings component. The draws of the persistent component and employment status are independent across time and independent of each other. These conditions can be written as

$$\eta_{i,t} = Q_t(u_{i,t}|\eta_{i,t-1}) \tag{5}$$

$$E_{i,t} = \mathbb{1}\{v_{i,t} > \bar{v}(E_{i,t-1}, \eta_{i,t-1})\} \tag{6}$$

$$u_{i,t} \overset{iid}{\sim} U(0,1) \tag{7}$$

$$v_{i,t} \overset{iid}{\sim} U(0,1) \tag{8}$$

$$u_{i,t} \perp\!\!\!\perp v_{i,s}, \;\; \forall \;\; t, s \tag{9}$$

Number of living parental households

Each household may start life with up to two living parental households. These parental households face a probability of death in each year that is a function of their age and the (child) household's education level. We assume that the age gap between living parents and their children is constant, conditional on education, such that parental survival is equivalently a function of the household's age. The realisation of the parental households' mortality is draw independently between the two households (if there are two). We denote the unconditional survival curve for the parental household j of household i with the set of probabilities $\{S_t^{p_j}\}_{t=0}^{t=111}$. These assumptions then define a law of motion for the number of parents parents, $P_t \in \{0, 1, 2\}$, denoted

$$P_{t+1} = \Phi_t^{age_{i,t}, ed_i}(P_t).$$

Parental wealth

Living parental households have a level of wealth that evolves according to an exogenous stochastic process. We denote the level of wealth of parental household j of household i at time t as $a_{i,t}^{p_j}$. As in the case of earnings, log parental wealth is assumed to be the sum of a deterministic component which is a function of the parental household's age and education level, $f^a_{ed}(age^{p_j}_{i,t})$ and a stochastic component, $\mu^j_{i,t}$:

$$ln(a^{p_j}_{i,t}) = f^a_{ed}(age^{p_j}_{i,t}) + \mu^j_{i,t}$$

The stochastic component is assumed to follow a first order Markov process as given by a series of conditional quantile functions:

$$\mu^j_{i,t} = Q_t(v^j_{i,t}|\mu^j_{i,t-1}) \tag{10}$$

$$v^j_{i,t} \overset{iid}{\sim} U(0,1) \tag{11}$$

Inheritances

The inheritances received by a household depend on the processes for parental wealth and parental survival in the following way. When a parental household dies, their wealth is left as a bequest, which is taxed according to the function $b(\cdot)$ and split between their heirs, n^{ed}_k, to yield an inheritance. The number of heirs of parental household is a function of the level of education of the (child) household. Denoting the inheritance received by household i from parental household j as $H^j_{i,t}$, we can write the level of inheritance received as a function of the above processes as follows:

$$H^j_{i,t} = \begin{cases} 0 & \text{if } P_t = P_{t-1} \\ b(a^j_{i,t})/n^{ed}_k \text{ w.p. } 0.5 & \text{if } P_t = P_{t-1} - 1 \\ b(a^j_{i,t})/n^{ed}_k & \text{if } P_t = P_{t-1} - 2 \end{cases} \tag{12}$$

Initial conditions

Initial levels of the persistent earnings shock and persistent parental wealth shocks are drawn from an education-specific joint distribution.

3.3 Choices and constraints

In each period the household chooses its level of consumption or, equivalently, its level of net saving, $z_{i,t}$. Saving is into a single, risky asset whose returns are drawn independently over time. Agents may borrow but must repay all borrowing with certainty by age 75 (they face the implied natural borrowing constraint). Inheritances received add to the household's assets.

A government levies a labour tax on income less net savings, according to the tax function $T_t(\cdot)$, and provides a public pension, $sp_t(e_{i,K-1})$, that is a function of education and earnings in period $K-1$.

This can be formalised through the following budget constraints:

$$y_{i,t} = e_{i,t} + sp_{i,t} + H^1_{i,t} + H^2_{i,t} \tag{13}$$

$$c_{i,t} = T_t(y_{i,t} - z_{i,t}) \tag{14}$$

$$a_{i,t+1} = (a_{i,t} + z_{i,t})(1 + r_{t+1}) \tag{15}$$

$$a_{i,t+1} \geq \underline{\mathrm{a}} \tag{16}$$

where $y_{i,t}$ is gross income in period t, $a_{i,t}$ is the level of assets held in time t, r_{t+1} is the net rate of return on assets held from time t to time $t+1$ and $\underline{\mathrm{a}}$ is the natural borrowing limit for period $t+1$.

3.4 Timing and household problem

We now formally define the household problem. We first clearly state the within-period timing assumtions. The timing of events each period is as follows:

1. The household may die, potentially leaving a bequest

2. If the household does not die, their earnings of evolve

3. Any remaining parental households have their wealth evolve

4. One or more of the remaining parental households may die, resulting in an inheritance

5. Household makes their consumption/savings choice

Before formally defining the solution to the household problem, we note that the household problem can be re-written in a way that will be convenient when solving the model by defining the variable "cash in hand" as the sum of assets and income: $M_{i,t} = e_{i,t} + sp_{i,t} + H^1_{i,t} + H^2_{i,t} + a_{i,t}$. The solution to the household problem solves the following Bellman equation for the household problem

$$\begin{aligned} V_t(\eta_t, M_t, P_t, \mu^1_t, \mu^2_t; ed) = max_{c_t}\big\{u(c_t;\theta_t)+ \\ \beta s_{t+1}\int V_{t+1}(\eta_{t+1}, M_{t+1}, P_{t+1}, \mu^1_{t+1}, \mu^2_{t+1}; ed) \\ dF(\eta_{t+1}, M_{t+1}, P_{t+1}, \mu^1_{t+1}, \mu^2_{t+1} | \eta_t, M_t, P_t, \mu^1_t, \mu^2_t; ed) \\ + (1 - s_{t+1})\beta\phi(b(a_{t+1}))\big\} \end{aligned}$$

subject to the budget constraints and laws of motion for earnings, parental wealth and number of parental households. The state variables of this problem are: time (equivalently age), education, the persistent component of earnings, cash in hand, the number of living parental households and the levels of the persistent component of wealth for the parental households. This problem has no analytical solution and must be solved numerically.

3.5 Discussion of modelling assumptions

Before moving on, we note some key assumptions made in our modelling setup. Intergenerational wealth transfers in our model take the form of transfers at the end of life of the parental household only. The interpretation here is that parents do not make wealth transfers while alive and within couples, wealth is transferred to any surviving spouse at death, before being transferred to children after the death of the second member of the couple. Further, we note that estates are split evenly between children. Clearly, in many cases some transfers are made during life and some inheritances may also be made to others than a surviving spouse or children and estates split unevenly. However, these assumptions are very much in line with the trends and norms in the UK and US.[1]

The initial levels of earnings and parental wealth are drawn from a joint distribution. This can be thought of as capturing intergenerational correlation in ability or economic status (which drives earnings and wealth accumulation). The realisation of shocks in the model are independent meaning that, conditional on the initial level of earnings and education, there is no correlation between the household's earnings and the inheritances they receive. In other words, there are no compensatory bequests in the sense of parents' bequests responding to how their child's life turns out. As discussed, the prevailing norm is that parents do not provide compensatory bequests through the way they split their estate (though Boar (2018) for example, finds that parents wealth accumulation does respond to the earnings risk faced by their children). Vice versa, an child's level of earnings does not respond to their realisation of inheritance. This could be relaxed in further work through a labour supply choice, as will be discussed, though in such a model we would likely still assume no effect on wages.

4 Identification and estimation of exogenous processes

In this section, we set out the indentification of the exogenous processes for household earnings and parental wealth that feed into the model of household behaviour. We also set out the data and procedure used to estimate the model parameters.

[1]Nolan (2020) finds that approximately 90% of the annual flow of intergenerational transfers in the UK is in the form of inheritances as opposed to inter-vivos transfers. Crawford and Mei (2018) document that 86% of respondents to the ELSA 'End of Life' module who were homeowners and died with a surviving spouse gave their main home entirely to their partner and 82% of those holding 'other' assets (the closest category to financial and physical wealth) who had a surviving spouse bequeathed them the entirety of this wealth. Among those with no surviving spouse, 75% of those with housing wealth bequeathed at least some of it to their children, and of those with non- housing wealth, 60% gave the entirety of this wealth to their children. Menchik (1980) and Wilhelm (1996) document that equal division of estates between children is very much the norm.

4.1 Earnings process

4.1.1 Empirical specification and identification

The earnings model set out above places minimal restrictions on the form of the deterministic earnings component and the process for the persistent earnings component. In order to take this process to the data, we will must specify the form of the deterministic component, conditional quantile functions and initial conditions in a way can be estimated.

Deterministic component

The deterministic component of earnings in our model is a function of education, age and time. In the description of our model above, we abstracted from differences in earnings across cohorts. As the estimation of our earnings model will require us to pool data from multiple cohorts in order to have observations of households of all working ages, we allow for differences by 10-year cohort in the deterministic component of earnings. We specify the deterministic function as

$$f_t^{c,ed}(age_{i,t}) = \beta_{ce} + \sum_{a=16}^{64} \gamma_{ae} \mathbb{1}\{age_{i,t} = a\} + \delta_t + \sum_{k=1}^{45} DP_{it}^k \tag{17}$$

where β_{ce} is a 10-year birth cohort and education-specific intercept, $\{\gamma_{ae}\}$ is an education-specific set of individual year-of-age effects, δ_t is a linear time trend and DP^k are a series of time dummies constrained to sum to zero.[2] We are allowing for levels of earnings that differ by the interaction of 10-year birth cohort and education level. The age profile of earnings is flexibly given by single year of age dummies that vary by education level. Time effects are assumed to take the form of common deviations around a linear trend. This specification differs from Arellano et al. (2017), which includes only a set of age dummies.

The identification of the parameters of type of specification with cross-sectional data is

[2] This restriction, proposed in Deaton and Paxson (1994), means we do not have collinearity of age, time and cohort controls.

standard. We can separately identify the cohort and age effects for each education group so long as we have cohort overlap, given the assumption of a constant age profile across cohorts, within education groups. The time trends are separately identified from the age and cohort effects given the Deaton-Paxson restriction, which imposes a particular form on the effects of time. The assumption around the commonality of the age profile across cohorts may seem strong, but produces a close fit for each cohort in the first stage of the estimation process.

Stochastic component

We follow the empirical specification of Arellano et al. (2017) and specify the quantile functions as the sum of a set of products of low-order hermite polynomial functions of age, lagged value of the persistent earnings component and vary over a grid of the shock distribution. We also allow for a transitory component to earnings as in the setting of Arellano et al. (2017). We denote this $\epsilon_{i,t}$. This component can be interpreted as in part measurement error. It is for this reason that we discard this part of the earnings process when feeding it into our model. This transitory component is specified as an age-varying quantile function, with an empirical specification set out in an analogous way to that for the persistent component.[3] We make the addition, compared to Arellano et al. (2017), of allowing the quantile functions to vary fully flexibly by education group. Arellano et al. (2017) show that the parameters of an earnings process of the form set out are identified given panel data with at least 4 periods.

Probabilities of employment

We specify the probabilities of being in employment as a function of education, age and the persistent stochastic component of earnings. As with the level of earnings, these probabilities are in practice allowed to vary also by 10-year birth cohort. The empirical specification that

[3] We also follow Arellano et al. (2017) in other aspects of the empirical specification not discussed here for brevity. See the appendix for further details.

we implement allows these probabilities to vary non-parametrically by age, education, 10-year cohort and by a fixed number of quantiles of the lagged persistent component.

4.1.2 Data and sample

We now set out the data used to estimate this model. We draw upon data from Family Expenditure Survey (FES) and its successor surveys (the Expenditure and Food Survey and the Living Costs and Food Survey), which are cross-sectional surveys covering the years 1968 to 2018. We will refer to this collections of surveys as "the FES". We also use the UKHLS, a household panel survey running 1991 to 2018. The "British Household Panel Survey" waves of the data cover 1991 to 2008 and these survey waves are annual, with fieldwork in September to December of each year. The "Understanding Society" waves cover 2009 to 2018 and are rolling 2-year periods. In all surveys, individuals are interviewed at approximately 12 month intervals and data can be treated as annual.

We use two different datasets for the following reason. The FES covers a longer time period, enabling us to more precisely disentangle age, time and cohort effects. However, to estimate the stochastic processes, we require panel data and must use the UKHLS. Since the estimation of the deterministic profile only requires cross-sectional data, we use the FES to estimate the parameters of this function. We then strip out the deterministic component of earnings from the UKHLS data using the same method and estimate the stochastic process for earnings using this data.

We are interested in modelling processes for total household (or, more accurately, "benefit unit", corresponding to the fiscal unit i.e. a couple or single individual) pre-tax earnings. We drop dependent children from our sample. When modelling benefit unit earnings, we define a couple's age as the mean of their ages and their level of education as the highest achieved by the couple. The 10-year birth cohort variable is defined based on the mean birth year of the couple. We keep observations from the 1930s through to the 1980s birth cohorts. We define individual education using a three-way education categorisation based on the highest qualifications achieved by the individual: low: up to and including GCSEs (or no qualifications for those born before 1958), mid: A-level or equivalent (or GCSEs for those born before 1958), high: higher education degree.

When selecting a balanced panel of benefit units as required for the estimation of the stochastic components, we define a benefit unit as the same benefit unit in another period if it has the same members (excluding dependent children).[4]

The measure of earnings that we use is annual gross real earnings (2020 prices), including self-employment income.[5]

4.1.3 Estimation

To estimate the parameters of the deterministic component of earnings, we pool all observations of benefit unit earnings from the FES for those born between the 1930s and 1980s and run OLS estimation of the empirical counterpart to Eq. (17):

$$ln(e_{it}) = \sum_{c=30s}^{80s} \sum_{e=1}^{3} \beta_{ce} cohort_{it} \times ed_{it} + \sum_{a=16}^{64} \sum_{e=1}^{3} \gamma_{ae} age_{it} \times ed_{it} + \delta_t + \sum_{k=1}^{40} DP_{it}^{k} + \upsilon_{i,t} \quad (18)$$

The second step is to use obtain the residuals from the equivalent OLS regression on the UKHLS. We then take all non-overlapping sets of 6 consecutive observations of a benefit unit. We estimate the parameters of the empirical specification of the conditional quantile functions by using the EM algorithm. A description of the algorithm is given in the appendix and in Arellano et al. (2017). Note that in this step we pool all observations of a given education group together (combining cohorts) and estimate sets of parameters separately by education group. We assume here that the process for the stochastic component of earnings varied by education group but not by cohort.

[4]When selecting balanced panels of both individuals and benefit units, we enforce that a panel cannot run across the 'seam' between the BHPS and Understanding Society surveys. When selecting balanced panels we also drop observations where consecutive interviews are less than 9 months or more than 15 months apart (this is 3% of interviews, causing us to drop 14% of otherwise usable subpanels).

[5]We use the IFS' Household's Below Average Incomes "before housing costs" variant of the consumer price index (estimated based on the RPI for years before 1997-98) to convert nominal values to real terms.

The final step is to estimate the employment probabilities. Our approach here is somewhat ad-hoc since the existing literature on earnings process estimation does not yet model non-employment jointly with the evolution of earnings in the form set out above. We pool all observations in the UKHLS of benefit units observed for two consecutive periods conditional on being in employment in the first period. We then calculate, conditional on age, education, 10-year birth cohort and quantile of lagged earnings the proportion of observations that are in employment in the second observation. Given the observed employment rate for each group at each age, we can then back out the implied probability of being in employment in the following period conditional on being out of work in the previous period, for each quantile of the persistent component of earnings.

We show assessments of model fit in the appendix.

4.2 Parental wealth process

We now set out our empirical specification, identification and estimation of the process for parental households' wealth. This takes a similar form to the process for earnings.

4.2.1 Empirical specification and identification

In the model set out in section 3, wealth is the sum of a deterministic component that is a function of education and age, and a stochastic component whose process is given by age-varying conditional quantile functions. We set out the empirical specification given to these two components here.

Deterministic component

We specify the function that defines the deterministic component of wealth as the sum of an education and cohort-specific intercept, an education-specific fourth order polynomial in age and time effects that are the sum of a linear trend and a series of time effects constrained to sum to zero (i.e. we again make the Deaton and Paxson (1994) restriction).

$$f^{a}_{ed}(age^{p_j}_{i,t}) = \alpha_{ce} + \phi^{1}_{ae} age^{p_j}_{i,t} + \phi^{2}_{ae}(age^{p_j}_{i,t})^2 + \phi^{3}_{ae}(age^{p_j}_{i,t})^3 + \phi^{4}_{ae}(age^{p_j}_{i,t})^4 + \delta_t + \sum_{k=1}^{7} DP^{k}_{it} \quad (19)$$

We note that this function depends on the (child) level of education. One might expect a more natural specification to be that the function vary by parental household education. We make this specification as parental households' level of education is not a state variable of our lifecycle model. Intuitively, the intercept captures the average level of parental wealth for households of a particular level of education and the age coefficients capture the average age trend of parental wealth (in parental age) within that group. We note also that as parental age is not a state variable of our lifecycle model, when we feed the parental process in the model we will assume a constant age gap between parents and children conditional on birth cohort and education level and hence assume away heterogeneity in the parental wealth process that comes from heterogeneity in this age gap.

The identifying assumptions include those analogous to in the case of earnings. We require that, conditional on education, the age-profile of parental wealth is the same across cohorts. One threat to this might be if different cohorts of parental households draw down on their wealth at different rates on average. This might be expected if, for example, the parents of later-born households expect to live longer and so have a more gradual path of decline of wealth at older ages. To some extent, changes in the average rate of wealth drawdown across cohorts are allowed for to the extent that these are captured by a different educational composition of the households in the cohort. Given we are examining parental households at older ages, differential mortality according to wealth means that we will observe a selected sample (only those who survive) in the cross-section. We can augment the above specification in two ways to deal with this. First, we can use data only from balanced panels of observations and secondly, we can interact the intercepts with a set of dummy variables indicating the first age at which an individual was observed. Under the assumptions that

(1) level but not the age profile of wealth varies systematically by age of death and (2) we observe each cohort at some age before there has been any differential mortality by wealth level then we will recover the common age profile and intercept terms for each cohort and education group. These assumptions of course rule out that parental households might draw down their wealth faster in response to news about the resolution of uncertainty over the timing of their death.

Stochastic component

Here we follow the same empirical specification as used for the household earnings process with the same identifying assumptions applying. We assume that the process for parental wealth varies by child education level but does not vary across cohort and is not affected by survivor bias.

4.2.2 Data and sample

We draw on data from the English Longitudinal Study of Ageing (ELSA) a biennial household panel that began in 2002-03. Theer are currently 9 waves available. We use the measured level of household non-pension wealth (the overwhelming majority of pension wealth for the cohorts we examine is non-bequeathable defined benefit pension wealth) which consists of the sum of net property wealth (including second homes), business, physical and net financial assets.

ELSA contains information about the number and year of birth of all children of sample members. This allows us to use ELSA where the level of observation is the 'child'. ELSA does not contain information on the educational attainment of the children of sample members. We therefore impute child education in a 2-step procedure. This procedure draws on two datasets that have data linking parents and their children and contain parent characteristics and child educational attainment. First, we use the UKHLS. The UKHLS follows household 'split offs' of original sample members, including those who are originally children in a household and leave home to form their own household at older ages. There is also a wealth module in selected waves of the UKHLS. This allows us to determine, for each percentile of the distribution of parental wealth, the distribution across child education levels. Specifically, we make a non-parametric estimation, for each parental wealth percentile, of the percentage of children with each education level. Here the unit of observation is again the 'child'. We make rankings of parental wealth levels based on the parental wealth observation from when parental household was aged closest to 50 (this is approximately the starting age of parents in our model).[6] Secondly, we use estimates from the Longitudinal Study of Linked Censuses (LS) of the relationship between parental characteristics (including their education, social class, housing tenure and region, but not including wealth) and child education. These estimates are taken from probit regressions estimated by Bourquin et al. (2020) where the outcome variable is the child education level and the explanatory variables are parent characteristics. These models were estimated separately for each parent birth-decade and child birth-decade combination.

We use these two sets of estimates to impute child education within ELSA in the following way. Firstly, within parental wealth ranks in ELSA (where rankings were calculated based on a household's first wealth observation - which will be that closest to age 50 - and within wave and cohort), we rank observations according to the predicted probability that they are high-educated, using the estimated coefficients from the models estimated using the LS in Bourquin et al. (2020). Second, we assign the first X% to be highly educated, the next Y% to be mid-educated and the remaining 1-X-Y% to be low educated, according to the proportions estimated for that percentile using the UKHLS.

In our sample, we include observations of benefit units born between 1950 and 1989 that have wealth information and the required covariates. From these, we select all observations that are observed for 6 consecutive waves or more.

[6] We make these rankings separately by parent decade of birth and wave in order to account for age and time effects.

4.2.3 Estimation

Estimation of the deterministic component of the wealth process is by means of OLS estimation of the below equation.

$$ln(a)_{it} = \sum_{c=50s}^{80s} \sum_{e=1}^{3} \sum_{p=1}^{4} \alpha_{ce} cohort_{it} \times ed_{it} \times p_{i,t} + \sum_{e=1}^{3} \phi_{ae} f(age_{it}) \times ed_{it} + \delta_t + \sum_{k=1}^{7} DP_{it}^{k} + \upsilon_{i,t} \quad (20)$$

where $f(age_{it})$ is a quartic in age. Note that the cohort-and education-specific intercepts are interacted with a series of dummy variables denoted by $p_{i,t}$. These dummies record which wave a household is first observed in the survey (given that we restrict to observations that are present for at least 6 waves and there are 9 waves of ELSA, this means that observations are first observed in either wave 1, 2, 3 or 4). This estimation using only observations of households with positive levels of wealth (around 90% of observations). Note that the interpretation of the estimated age profile of log wealth is the expected level of log wealth, conditional on wealth being positive.

The second stage estimation uses the residuals from the first stage estimation. For households with negative levels of wealth, assign them a level of log wealth of zero and assign them a residual equal to the negative of the predicted level of log wealth from the estimated first stage relationship. In effect, we are bottom-coding the wealth distribution at zero for use in the second stage. This creates a mass point of low negative levels of the residual component. This means that in our second stage estimation, we cannot model any dynamics within those who have negative wealth levels and any heterogeneity in the evolution of their subsequent levels of wealth, if positive. Given that debts are not heritable, this is not a concern for us.

Model fit

Figure 3 shows an assessment of the within-sample fit of the wealth model based on its ability to reproduce cross-sectional patterns in the data. We show, for the 6-wave balanced panels beginning in wave 1 and wave 4, the comparison between actual wealth and the simulated levels of parental wealth from the estimated model. We show the 25th, median and 75th percentile, for each education group for the 1970s and 1980s birth cohorts.

Next, we show the fit of the model to the persistence of wealth. This is a key input into our model: we might expect that wealth is highly persistent from year to year for most households, but that households are exposed to a small probability of a large changes in their wealth, for example from long-term care costs or large idiosyncratic gains or losses in house value or business wealth etc. Figure 4 shows a measure of the persistence of wealth for both the data and the estimated model. This measure of persistence is the average derivative of the conditional quantile function for the stochastic component of wealth with respect to its lagged value. For the model, this is estimated value of

$$\rho_t(\tau, \eta_{t-1}) = E\left[\frac{\partial Q_t(\tau|\eta_{t-1})}{\partial \eta_{t-1}}\right] \tag{21}$$

This can be seen as measuring the extent to which a marginal increase in wealth 'persists' when hit by a shock from position τ in the shock distribution. This average derivative is a function of both the shock and the initial level of wealth. In the RHS panels of Figure 4, we plot the value of this average derivative as a function of the quantile of the shock to earnings (τ_{shock}) and the quantile of the lagged value of wealth (τ_{init}). The left panel shows the counterpart to this object in the data. It is obtained by running a series of quantile regressions where the outcome variable is the wealth residual from the first stage of the estimation. We the explanatory variables are a set of low-order interactions of hermite polynomials of the same form as used in the specification of the model.

We can see from this figure that the model captures the main features of the data including the higher persistence of wealth for low-wealth households hit by positive shocks. There is particularly low persistence for high-wealth households hit by negative shocks. These shocks have the interpretation of shocks that 'wipe out' the history of wealth. We can see that amongst the low-educated households, the sizeable minority of households that have their wealth bottom coded leads to a very low estimated persistence in the data for low wealth households. This estimate is spurious in the sense that it is driven by an artificial lack of variation amongst low wealth households. However, again, as we are not primarily concerned with modelling low or very negative wealth and the model fits the other parts of the distribution of persistence well, this is not a concern.

5 Parameterising and solving the model

This sections sets out the parameterisation of some elements of the model and how it is solved numerically.

5.1 Parameterisation

There are several other components that must be parameterised in the model. Here, we briefly outline the data sources used for each.

Tax and benefit system

Taxation of labour incomes in the UK is at the individual level. Given that we are using a household model, we therefore estimate a household level tax function. We do this by using the FES data from 1968 to 2018 , which includes measures of pre-tax income and post-tax income. Using these, we estimate for each year, the following tax function using nonlinear least squares:

$$T_t(y_{i,t}) = \psi_t^a + \psi_t^b (y_{i,t})^{\psi_t^c} \tag{22}$$

For future, years we assume that tax system remains unchanged in its 2018 form.

Estate tax

The estate tax is of a form designed to capture the features of the UK inheritance tax over the relevant period. The UK inheritance tax is set at a 40% rate on the value of estates over a threshold. There are a number of exemptions and additional allowances. The most relevant of these in the vast majority of cases of these is for owner-occupied housing. As we model only total wealth, we assume that a certain share of total wealth is held as housing and apply the inheritance tax system, assuming that the remainder of wealth attracts no exemptions.

Public pensions

Public pensions in the UK are based on an individual's full history of earnings and employment.[7] We are restricted to include in our model a public pension that is a function of the model's state variables. In this case, the relevant variables are education (relevant here as proxy for lifetime average earnings) and earnings in period $K-1$, the period before the pension is received.

Our approach to construct a household level state pension function is as follows. First, we simulate our household earnings process 10,000 times for each of our cohort and education groups. We then assign household earnings within members of a couple by using the mean shares of earnings of received by the first and second-highest earning members of couples by using shares estimated from the FES.[8] We then use a pensions calculator which calculates entitlements for each simulated individual. Couples have their entitlements recombined to give a household level of pension income. We then estimate, separately for each education and cohort group, entitlements as a linear function of final period earnings.

Our estimated functions capture the main features of the UK state pension system: higher

[7] Entitlement can also be gained for some other activities including the receipt of out of work benefits but we abstract form this.

[8] We estimate a the share of earnings amongst couples as cubic function of age interacted with education plus the interaction of cohort and education.

educated individuals receive higher entitlements but the system is progressive. For the later cohorts, the system becomes less related to earnings but more generous for lower earners, reflecting reforms to the UK state pension system over time.

Survival curves

We estimate individual survival curves for each cohort and education group. We combine UK Office for National Statistics data for estimated survival curves that are specific to year of birth and sex with mortalty data from ELSA to create survival curves that vary also by individual education level.[9]

With these individual survival curves, we are able to construct household survival curves for each 10-year birth cohort and education group by taking the observed distribution of birth years and education levels for both individuals and couples and assuming independence of timing of death within couples. This gives us a set of survival curves for the final member of the household that we use for both households and their parental households.

The joint distribution of the initial level of earnings and parental wealth

The joint distribution of the initial level of earnings and parental wealth is an important input into our model. We draw on the UKHLS which contains information on linked parent-child pairs. We take all observations where parental wealth is observed and child earnings are observed (while in their 20s). We rank parental wealth as described in section 4.2.2. We rank child earnings by removing age effects from each earnings observation and the taking the mean level of observed earnings from all of the child's earnings observations and ranking children within cohort according to this earnings rank. We then non-paramertically estimate the joint distribution of quantiles of initial earnings and parental wealth within each education group.

[9]We follow the method set out in Sturrock and O'Dea (2020).

Behavioural parameters

We parameterise the coefficient of relative risk aversion and the bequest preference parameters using existing estimates from the literature. The coefficient of relative risk aversion is set to 3, following Scholz et al. (2006) and Crawford and O'Dea (2020). We take the marginal propensity to consume from final-period wealth implied by the parameter governing the strength of the bequest motive in Lockwood (2018). We also take the 'threshold' parameter from that paper and convert it to 2018 pounds.

5.2 Solving the model

We solve the model numerically. In order to do this, we need to make our earnings and parental wealth into discretised processes with Markov transition matrices. We do this by following the method of De Nardi et al. (2020). We simulate the processes for the stochastic components of earnings/wealth a large number of times, divide the distributions of earnings/wealth at each age into a certain number of quantiles and calculate the transition probabilities between quantiles in the simulations. We set the level of earnings/wealth for that quantile as equal to the median simulated level of earnings/wealth in each quantile. This grid of levels of earnings/wealth and transition probabilities defines a discrete Markov process.[10] We note that the periodicity of the wealth process is 2 years. Therefore every other year is defined by a transition matrix equal to the identity matrix i.e. the quantile of wealth changes only every other year.

We construct a grid of values for cash-in-hand based on the borrowing constraint and the maximum attainable level of cash in hand at each age. We then begin in the terminal period and, for each combination of the state variables, we calculate the optimal choice of end-of-period assets, taking into account the possible evolution of the household's own mortality, earnings, parental mortality, parental wealth, and the rate of return. We solve the model

[10] We use 8 quantiles of the following sizes for both earnings and wealth: 5%, 10%, 10%, 25%, 25%, 10%, 10%, 5%.

recursively, working back until the initial period.

5.3 Validating the model

Figure 5 shows a comparison of the age profiles of median levels of wealth simulated by our model with those from the Wealth and Assets Survey.

We can further validate our model predictions against the reduced form evidence set out in section 2.

6 Results

In this section we set out our results from the model, examining consumption and wealth inequality and intergenerational mobility.

6.1 Effects on consumption inequality

We first use the model to simulate the distribution of consumption both with and without the presence of inheritances. In the case without inheritances, households neither expect nor receive inheritances.

Figure 6 shows the simulated effect of inheritances on the 80:20 ratio of consumption. We see that while that largest differences in the levels of consumption due to inheritances come later in life, once inheritances are received, there is some anticipatory behaviour. There are relatively modest impacts on the 80:20 ratio at all ages for each of the birth cohorts we examine. We do see a modest decline in consumption inequality at older ages as a result of inheritances. Notably, at younger ages, for the 1980s-born generation, inheritances mildly increase consumption inequality while then switching to decrease it at older ages. This is because those households who are further up the lifetime income and consumption distributions are more willing and able to react to the anticipation of receiving an inheritance by increasing their consumption. Lower lifetime income households are more constrained by precautionary motives from spending the inheritances that they anticipate in advance of its receipt.

6.2 Effects on wealth inequality

Next, we make an equivalent set of simulations for the distribution of wealth. Figure 7 shows the simulated effect of inheritances on the 80:20 ratio of wealth.

At each age, the total effect of inheritances on wealth can be thought of as the sum of two factors: the direct effect of inheritances received by that age (and the returns to these if saved) and the indirect effect of the change in behaviour induced by the expectation and receipt of inheritances. Across the distribution, the effect of inheritances before around age 40 is too small to be seen but it does decrease wealth. This suggests that anticipatory effects dominate at these ages. At older ages, once the bulk of inheritances are received, wealth is substantially higher as a result of inheritances.

The effects of inheritances on wealth levels early in life are larger in absolute terms for those higher up the distribution. This is the result of the larger anticipatory consumption response to inheritances that we saw for households with higher lifetime income, who are also the households that tend to be higher up in the wealth distribution. But while these anticipatory effects, where households save less in order to consume, and so build up less wealth, are larger in absolute terms further up the distribution – and indeed larger as a percentage of the inheritance received – they are smaller as a percentage of wealth than the effects further down the wealth distribution.

A significant amount of previous research has examined the contribution of past inheritances to current wealth inequality in the UK (Crawford and Hood, 2016; Karagiannaki, 2017; Nolan et al., 2020). In crude terms, these investigations have compared inheritances reported as received over some fixed period (indexed with inflation or an assumed rate of return) with wealth held at the end of that period. The effect of inheritances on wealth inequality is assessed by computing measures of inequality with and without subtracting the present value of inheritances received. This would be equivalent to our method of assessing the effect of inheritances on inequality only in the case where households did not change their behaviour as a result of expecting to receive and receiving an inheritance. Our analysis suggests that behavioural effects could have a substantive impact on levels of wealth held, and that these anticipatory effects differ across the distribution of lifetime income and therefore wealth. As these anticipatory effects are larger for the lower parts of the distribution, this suggests that an analysis of wealth inequality that does not account for them will tend to overstate how much inheritances reduce wealth inequality (or understate how they increase it).

6.3 Effects on intergenerational mobility and equality of opportunity

We combine the effects of inheritances at different ages (weighting by the probability that individuals survive to each older age) to calculate the effect on the distribution of lifetime equivalised consumption. The effects are small: inequality in lifetime consumption as measured by the Gini coefficient is decreased by 2.7% for the 1960s cohort, falling to 2.4% for the 1980s cohort. The main finding is that we do not expect inheritances to have a substantial effect on overall lifetime consumption inequality in any of these cohorts. Arguably though, it is not so much overall inequality of this kind that is of most interest or concern when it comes to inheritances – rather it is social mobility or, loosely, inequality by family background.

A way of seeing concretely how parental wealth background is set to become a more important driver of differences between people is to decompose inequality into a component measuring inequality within groups of people who have parents with the same level of wealth and that between groups of people with different levels of parental wealth. We can examine how much of inequality is accounted for by this between parental wealth groups component versus that within parental wealth groups. Lee and Seshadri (2019) suggest that the extent to which inequality is driven by differences between individuals with different 'background characteristics' (of which parental background can be seen as one), as opposed to differences between individuals with similar background characteristics, is a measure of the degree of inequality of opportunity. We can also ask what contribution inheritances make to the share of inequality that is between parental wealth groups, and how this is set to change across cohorts. This can be interpreted as a measure of the contribution of inheritances to inequality in opportunity by parental background and how this may be expected to change across cohorts.

Table X shows the results of such an analysis for the Theil index of inequality, which is used because it can be decomposed in the way required. This shows that the share of inequality in lifetime consumption that is between parental wealth deciles is 16% for the 1960s cohort when there are no inheritances and rises to 22% as a result of including inheritances. This means that 6 percentage points, or 24% of the within-group share can be attributed to the effect of inheritances. This effect grows as inheritances become a more important part of lifetime income. For the 1980s cohort, inheritances are responsible for 8 percentage points, or 33% of the within-group share of inequality. We can describe this as meaning that inheritances are projected to grow from accounting for about a quarter of inequality in living standards by parental background for those born in the 1960s to accounting for about a third of inequality of living standards by parental background for those born in the 1980s.

7 Conclusion

Recent increases in wealth among older generations, combined with sluggish working-income growth over an extended period, mean that the growing magnitude of inheritances – not just in absolute terms but in proportion to young people's other economic resources – is set to continue.

Our work supports previous research suggesting that the implications of inheritance for standard measures of inequality between rich and poor are, perhaps counterintuitively, small.

But the implications for what is happening to inequality between people from different family backgrounds – that is, loosely, social mobility – are much starker. Our work also underlines the importance of accounting for anticipatory behaviour in the context of inheritance receipt and its impact on inequalities.

References

Alvaredo, F., B. Garbinti, and T. Piketty (2017). On the Share of Inheritance in Aggregate Wealth: Europe and the USA, 1900–2010. Economica 84(334), 239–260.

Arellano, M., R. Blundell, and S. Bonhomme (2017). Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework. Econometrica 85(3), 693–734.

Atkinson, A. B. (2018). Wealth and inheritance in Britain from 1896 to the present. Journal of Economic Inequality 16(2), 137–169.

Boar, C. (2018). Dynastic Precautionary Savings. pp. 1–76.

Bourquin, P., R. Joyce, and D. Sturrock (2020). Inheritances and Inequality Within Generations. Technical report.

Crawford, R. and A. Hood (2016). Lifetime Receipt of Inheritances and the Distribution of Wealth in England. Fiscal Studies 37(1), 55–75.

Crawford, R. and C. O'Dea (2020). Household portfolios and financial preparedness for retirement. Quantitative Economics 11(2), 637–670.

De Nardi, M. (2004). Wealth inequality and intergenerational links. Review of Economic Studies 71(3), 743–768.

De Nardi, M., G. Fella, and G. Paz-Pardo (2020). Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare. Journal of the European Economic Association 18(2), 890–926.

De Nardi, M. and F. Yang (2016). Wealth inequality, family background, and estate taxation. Journal of Monetary Economics 77, 130–145.

Deaton, A. S. and C. H. Paxson (1994). Saving, Growth, and Aging in Taiwan. In Studies in the Economics of Aging, pp. 331–362. University of Chicago Press.

Elinder, M. and O. Erixson (2012). The Impact of Inheritances on Heirs' Labor and Capital Income. The B.E. Journal of Economic Analysis & Policy 12(1), 1–35.

Elinder, M., O. Erixson, and D. Waldenström (2018). Inheritance and wealth inequality: Evidence from population registers. Journal of Public Economics 165, 17–30.

Gale, W. G. and J. K. Scholz (1994). Intergenerational Transfers and the Accumulation of Wealth. Journal of Economic Perspectives 8(4), 145–160.

Holtz-Eakin, D., D. Joulfaian, and H. S. Rosen (1993). The Carnegie Conjecture: Some Empirical Evidence. The Quarterly Journal of Economics 108(2), 413–435.

Hood, A. and R. Joyce (2017). Inheritances and inequality across and within generations. IFS Briefing Note BN192, 1–22.

Karagiannaki, E. (2017). The Impact of Inheritance on the Distribution of Wealth: Evidence from Great Britain. Review of Income and Wealth 63(2), 394–408.

Kindermann, F., L. Mayr, and D. Sachs (2020). Inheritance taxation and wealth effects on the labor supply of heirs. Journal of Public Economics 191, 104127.

Kotlikoff, L. J. and L. J. Kotlikoff (1988). Intergenerational Transfers and Savings. The Journal of Economic Perspectives 2(2), 41–58.

Kotlikoff, L. J. and L. H. Summers (1981). The Role of Intergenerational Transfers in Aggregate Capital Accumulation. Journal of Political Economy 89(4), 706–732.

Lockwood, L. M. (2018). Incidental Bequests : Bequest Motives and the Choice to Self-Insure Late-Life Risks. American Economic Review 108(9), 2513–2550.

Modigliani, F. (1988). The Role of Intergenerational Transfers and Life Cycle Saving in the Accumulation of Wealth. Journal of Economic Perspectives 2(2), 15–40.

Scholz, J. K., A. Seshadri, and S. Khitatrakun (2006). Are Americans Saving "Optimally" for Retirement? Journal of Political Economy 114(4), 607–643.

A Details of earnings process estimation and model fit

A.1 Further details of empirical specification

We use the same specification of the order of the hermite polynomials for each component of the process as used by Arellano et al. (2017) (tensor products of polynomials of degree 3 (in η_{t-1}) and 2 (in age) for the transition of η, and polynomials of degree 2 in age for η_0 and ϵ). Again following Arellano et al. (2017), we use quantile regressions of earnings on their lagged values and age to set initial parameter values. We use the same variances for the random walk proposals in the Metropolis-Hastings sampler as Arellano et al. (2017), which yields an acceptance rate of around 0.20-0.25.

A.2 Further details of estimation

The estimation procedure is based on quantile regressions corresponding to the restrictions embodied in the empirical specification. We begin with a balanced panel (in our case $T = 6$) dataset of observations of log of gross earnings and age. Log earnings have been purged of cohort, age and time effects in the initial regression step. Denoting the parameter vector θ, the posterior density of $(\eta_{i,0}, ..., \eta_{i,T})$ given $(y_{i,0}, ..., y_{i,T})$, the ages at which the individuals is observed and θ, as $f(\eta_i|y_i, age_i; \theta)$, we carry out the following steps:

1. **Select initial values for the parameter vector, $\hat{\theta}^{(0)}$.** These are selected by running a series of quantile regressions. For the parameters for the distribution of η_0, we estimate quantile regressions of observed earnings on hermite polynomials in age. For the distribution of ϵ_i we estimate quantile regressions of observed earnings on a different set of hermite polynomials in age. For the distribution of η_t given η_{t-1} we estimate quantile regressions of observed earnings on hermite polynomials in lagged earnings and age (and their products).

2. **Draw $\{\eta_i\}_{i=1}^{N}$ from $f(\eta_i|y_i, age_i; \hat{\theta}^{(s)})$, using the current parameter vector $\hat{\theta}^{(s)}$.**

A set of permanent components $\{\eta_i\}_{i=1}^N$ clearly implies a set of transitory components $\{\epsilon_i\}_{i=1}^N$, given observed earnings.

3. **Update the parameter vector to** $\hat{\theta}^{(s+1)}$ by computing, for each $l = 1, ..., L$:

$$(a_{1,l}^{Q(s+1)}, ..., a_{K,l}^{Q(s+1)}) = \underset{(a_{1,l}^Q, ..., a_{K,l}^Q)}{\text{argmin}} \sum_{i=1}^N \sum_{t=1}^T \rho_{\tau_l}\big(\eta_{i,t} - \sum_{k=1}^K a_{k,l}^Q(\tau)\varphi_k(\eta_{i,t-1}, age_{i,t})\big) \tag{23}$$

$$(a_{1,l}^{\epsilon(s+1)}, ..., a_{K,l}^{\epsilon(s+1)}) = \underset{(a_{1,l}^\epsilon, ..., a_{K,l}^\epsilon)}{\text{argmin}} \sum_{i=1}^N \sum_{t=0}^T \rho_{\tau_l}\big(y_{i,t} - \eta_{i,t} - \sum_{k=1}^K a_{k,l}^\epsilon(\tau)\varphi_k(age_{i,t})\big) \tag{24}$$

$$(a_{1,l}^{\eta(s+1)}, ..., a_{K,l}^{\eta(s+1)}) = \underset{(a_{1,l}^\eta, ..., a_{K,l}^\eta)}{\text{argmin}} \sum_{i=1}^N \rho_{\tau_l}\big(\eta_{i,0} - \sum_{k=1}^K a_{k,l}^\eta(\tau)\varphi_k(age_{i,0})\big) \tag{25}$$

$$\hat{\lambda}_-^{Q(s+1)} = -\frac{\sum_{i=1}^N \sum_{t=1}^T \mathbb{1}\{\eta_{i,t} < \sum_{k=1}^K a_{k,l}^Q(\tau)\varphi_k(\eta_{i,t-1}, age_{i,t})\}}{\sum_{i=1}^N \sum_{t=1}^T \big(\eta_{i,t} - \sum_{k=1}^K a_{k,l}^Q(\tau)\varphi_k(\eta_{i,t-1}, age_{i,t})\big)\mathbb{1}\{\eta_{i,t} < \sum_{k=1}^K a_{k,l}^Q(\tau)\varphi_k(\eta_{i,t-1}, age_{i,t})\}} \tag{26}$$

and analogously for the other tail parameters, where $\rho_\tau(u) = u(\tau - \mathbb{1}\{u \le 0\})$ is the check function.

4. **Iterate steps 2 and 3 for** $s = 1, ..., S$.

5. **Set the final parameters as the mean values over the last** $\tilde{S}$ **iterations**: $\theta = \frac{1}{\tilde{S}} \sum_{s=S-\tilde{S}+1}^S \hat{\theta}^{(s)}$

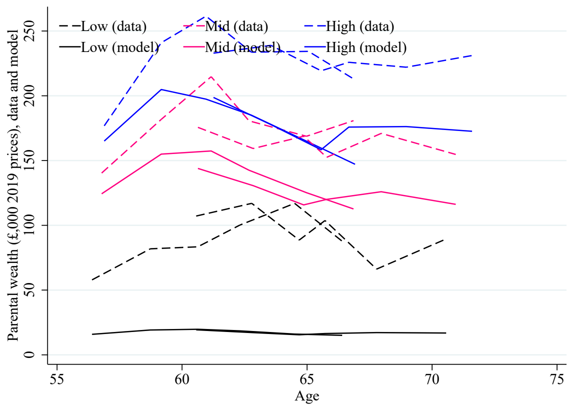
The draws from the posterior distribution of η_i are made using the Metropolis-Hastings algorithm.

A.3 Fit of earnings model

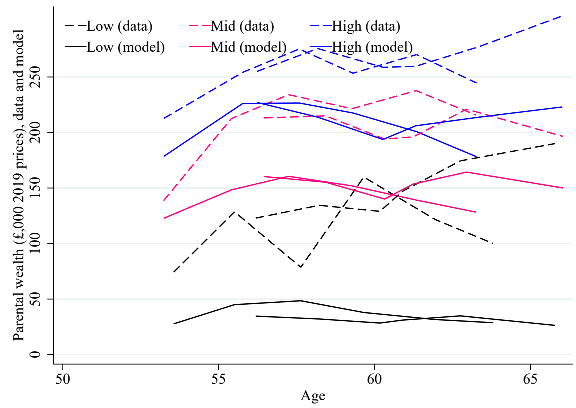
Figure 8 shows densities of the arc-percentage changes in earnings from our estimated model and in the BHPS data as an assessment of model fit. Higher order changes see a larger difference between the model and data in the density of changes at -2, 0 and 2. Given the good fit of the model conditional on positive earnings, even at higher order lags, this suggests that the divergence in the mass at zero is attributable to too few benefit units being simulated to remain at zero earnings over longer time horizons. The model's overestimation of the mass at -2 and 2 at longer time horizons represents the corresponding over-estimation of entry and exit from employment at these horizons.

Figures 9, 10 and 11 show the cross-sectional distribution of simulated and actual earnings by age and education group for benefit units in the 1960s, 1970s and 1980s birth cohorts.

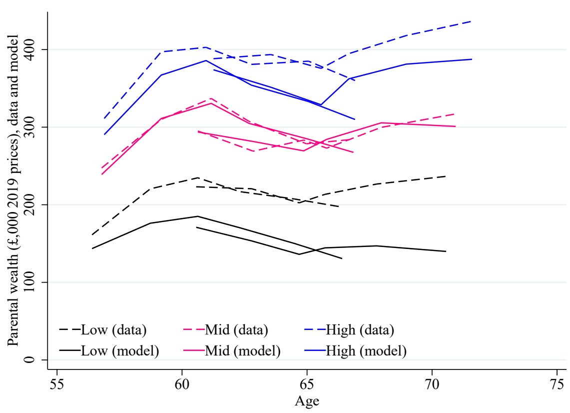
Figure 3: Comparison of distribution of simulated and actual parental wealth by education group and wave for two-balanced panel samples



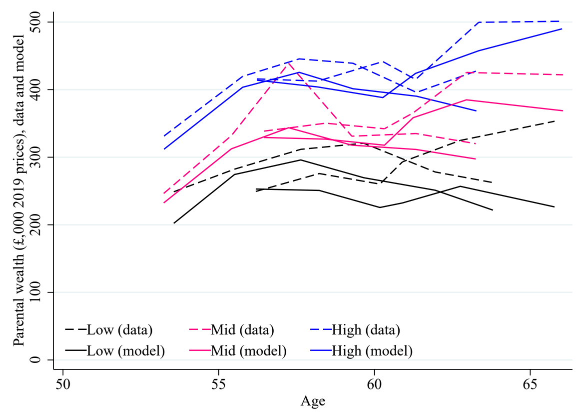
(a) 1970s: 25th percentile



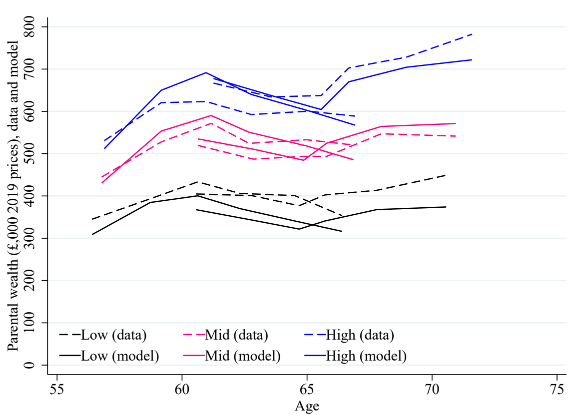
(b) 1980s: 25th percentile



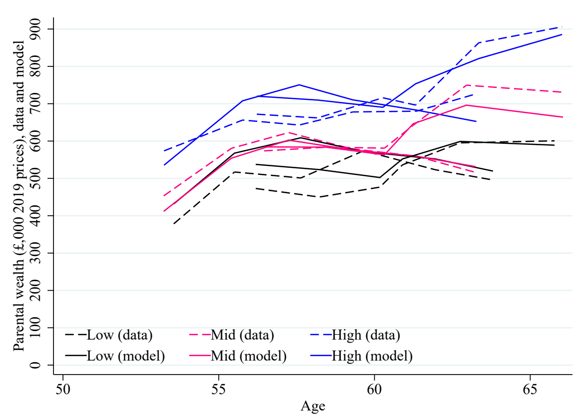
(c) 1970s: 50th percentile



(d) 1980s: 50th percentile

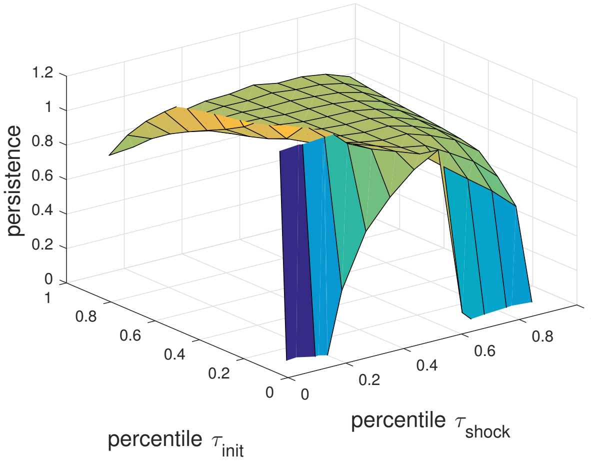


(e) 1970s: 75th percentile

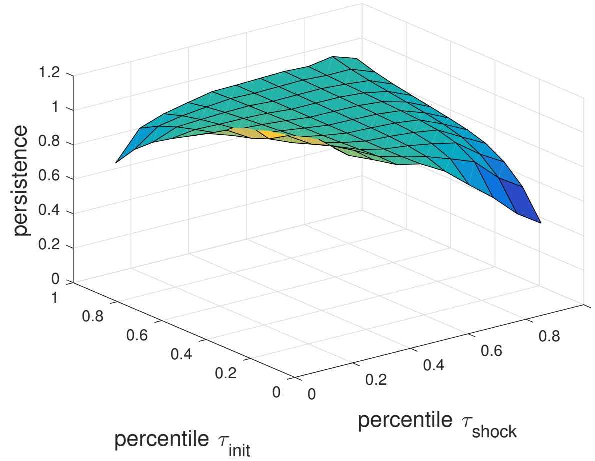


(f) 1980s: 75th percentile

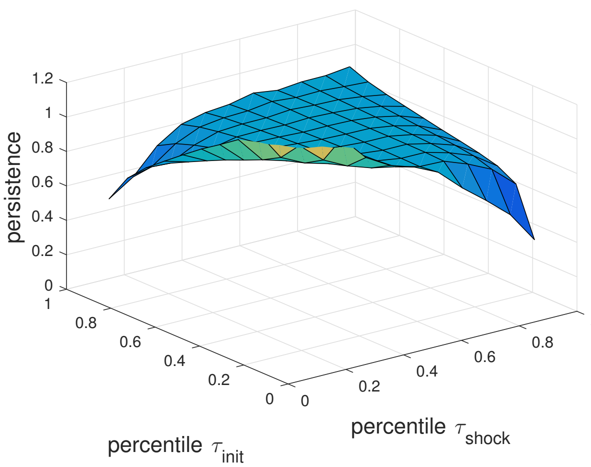
Figure 4: Comparison of persistence of earnings in data and estimated model



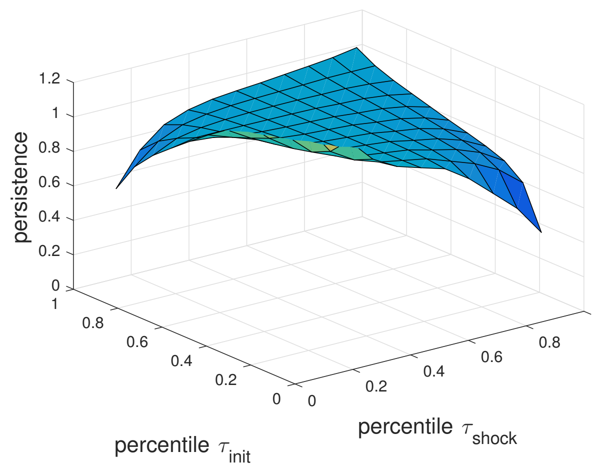
(a) Low-educated: data



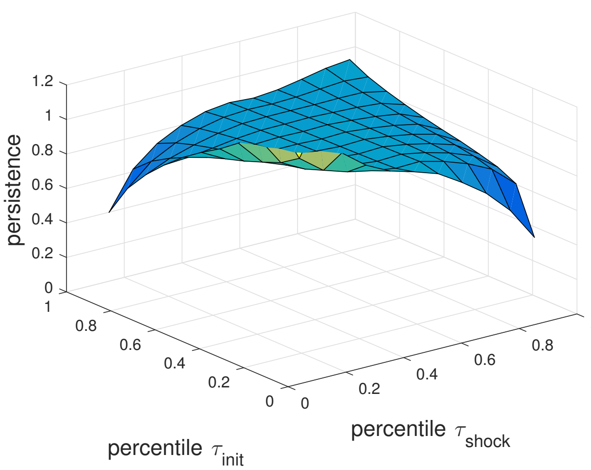
(b) Low-educated: model



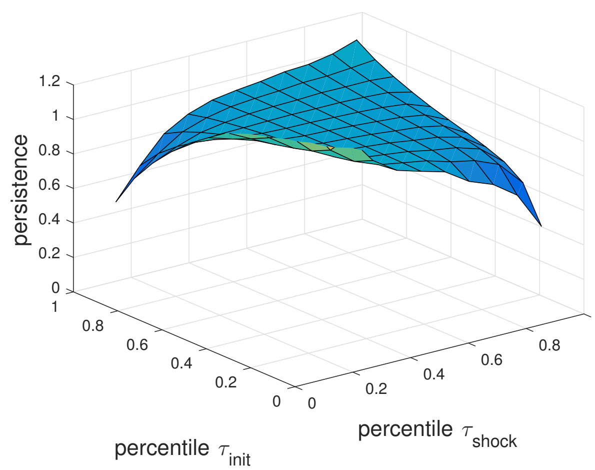
(c) Mid-educated: data



(d) Mid-educated: model



(e) High-educated: data



(f) High-educated: model

Figure 5: Inheritance and parental bequest expectations by birth cohort

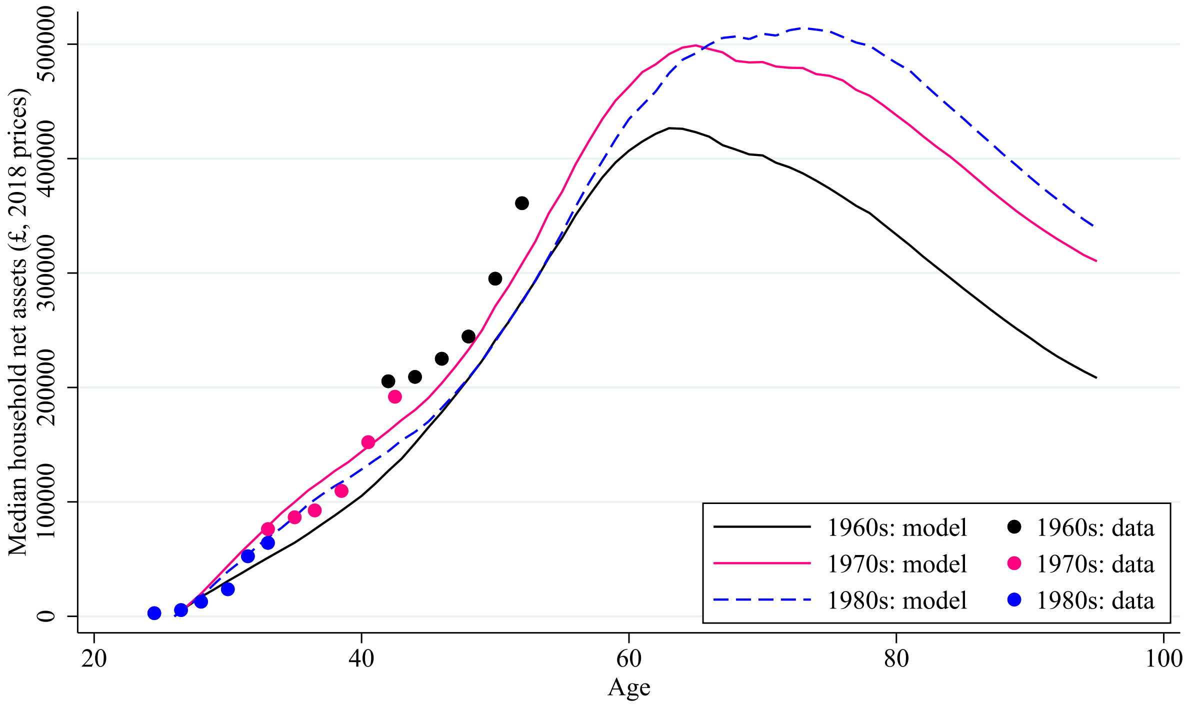
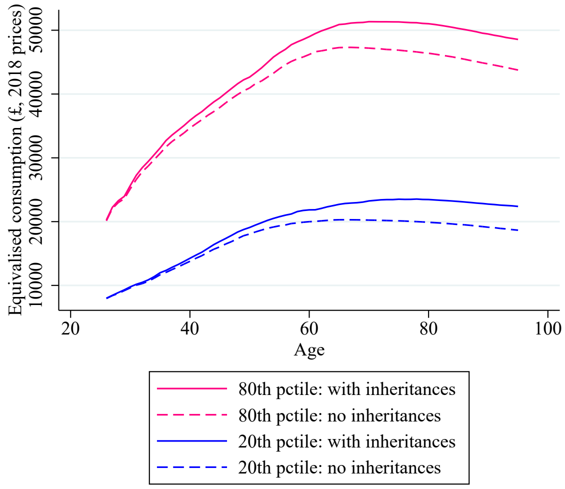
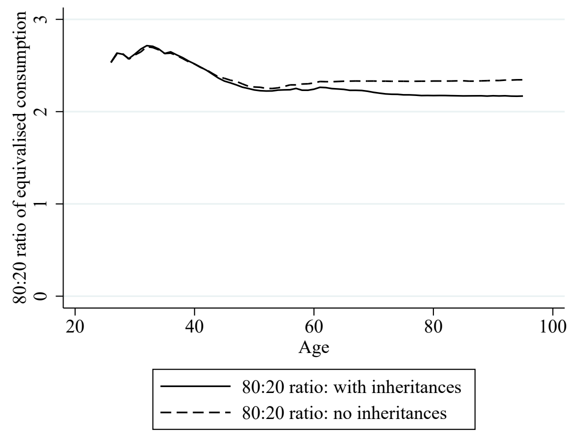


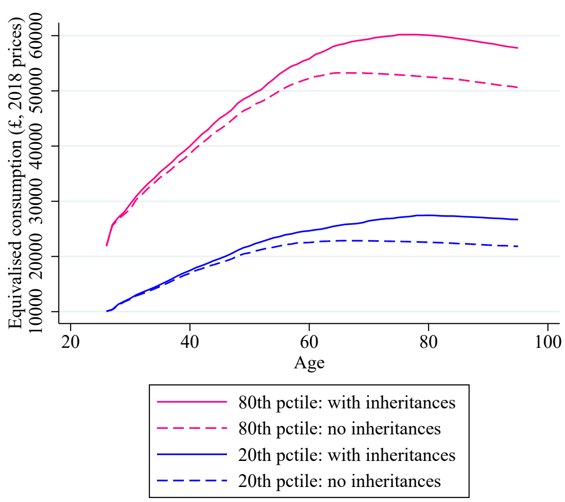
Figure 6: Simulated effects of inheritances on the 80:20 ratio of consumption



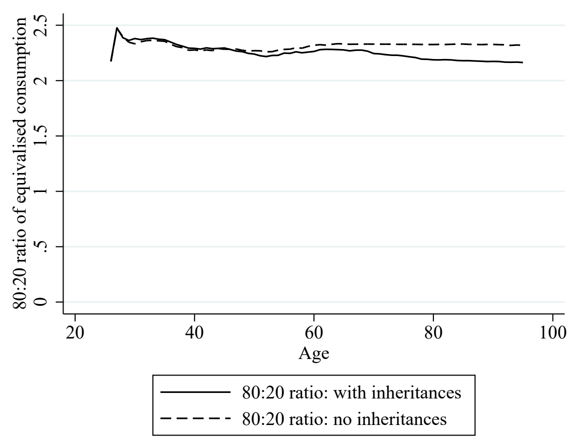
(a) 1960s



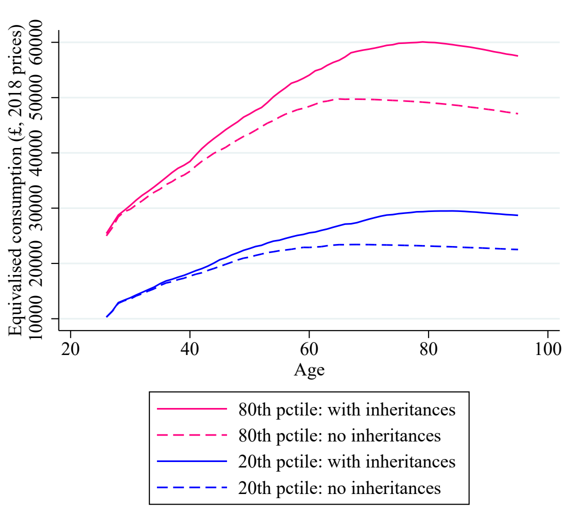
(b) 1960s

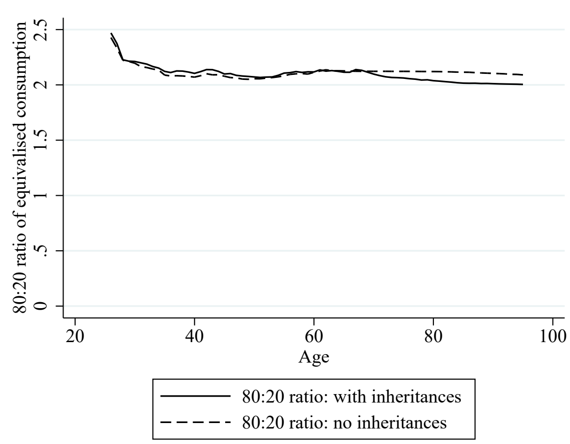


(c) 1970s



(d) 1970s

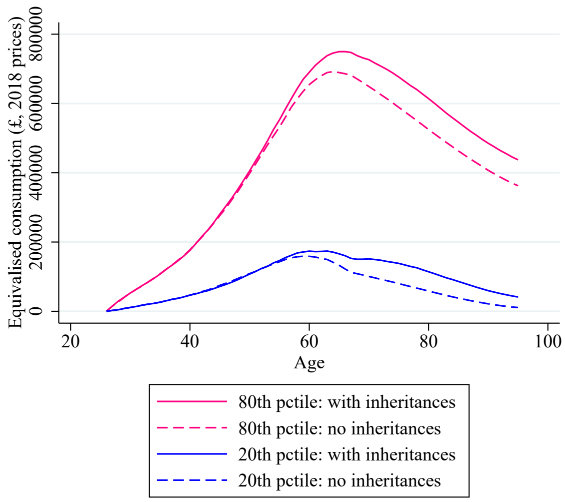




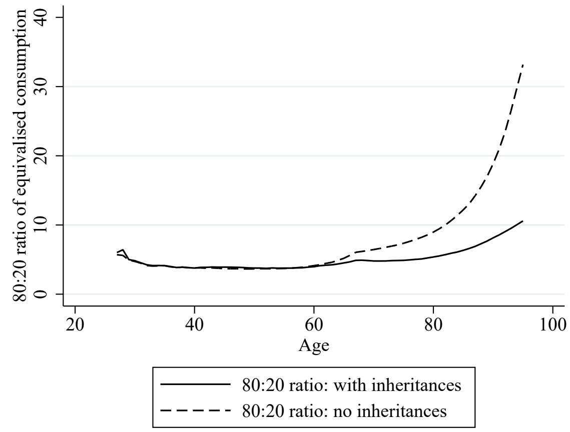
(e) 1980s

(f) 1980s

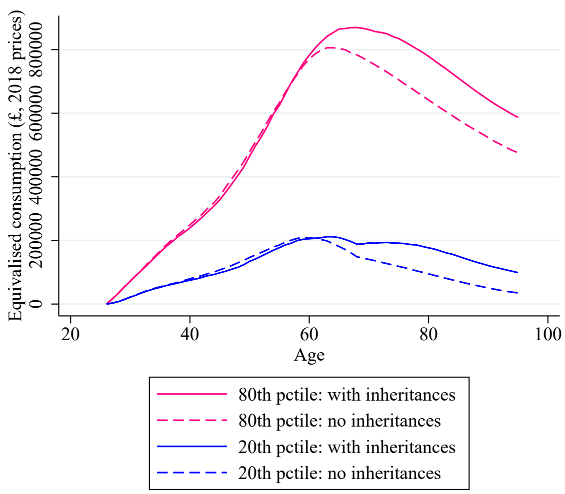
Figure 7: Simulated effects of inheritances on the 80:20 ratio of wealth



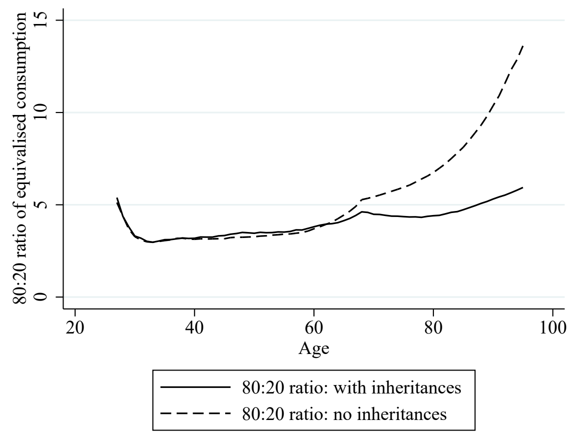
(a) 1960s



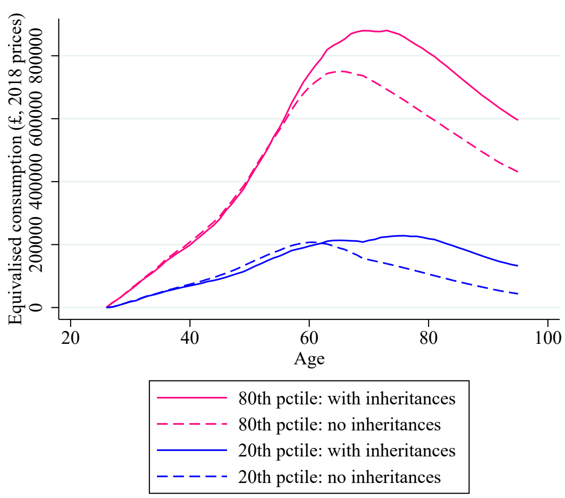
(b) 1960s



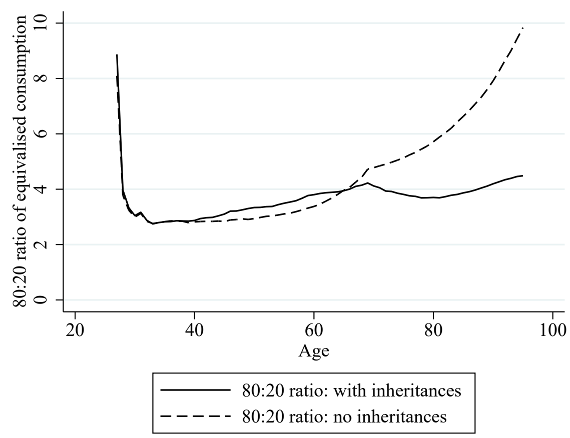
(c) 1970s



(d) 1970s

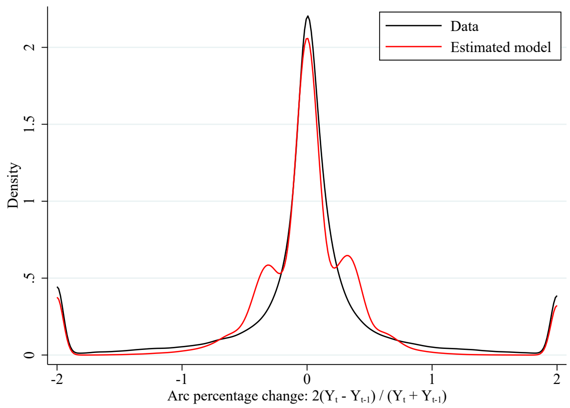


(e) 1980s

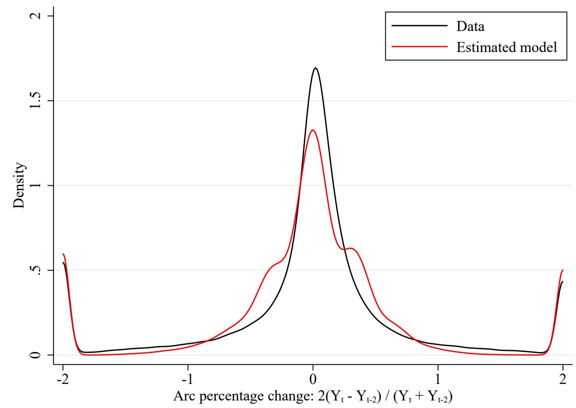


(f) 1980s

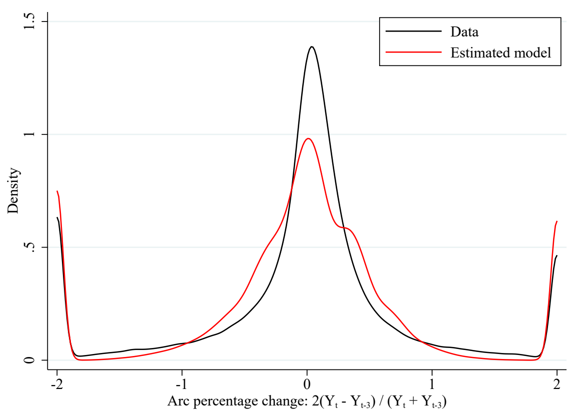
Figure 8: Simulated and actual arc-percentage change in earnings for changes over the time horizon of 1 to 5 years



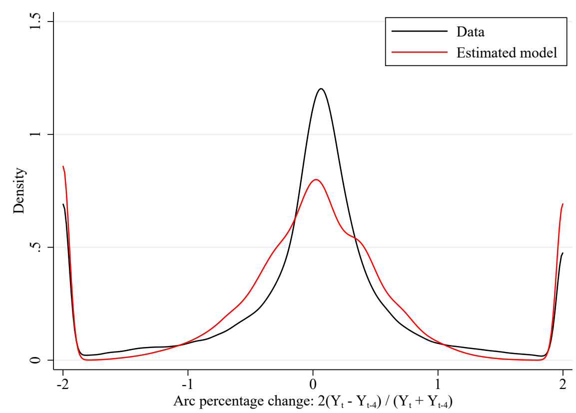
(a) 1-year change



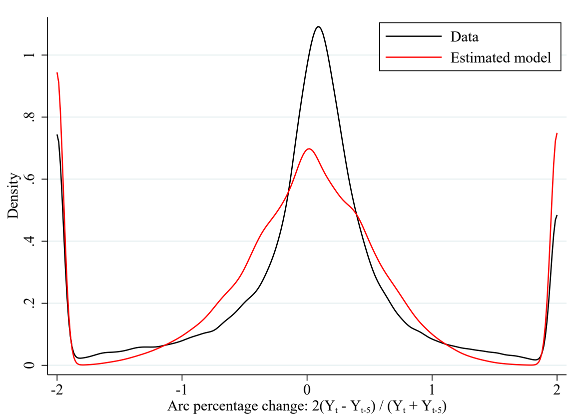
(b) 2-year change



(c) 3-year change

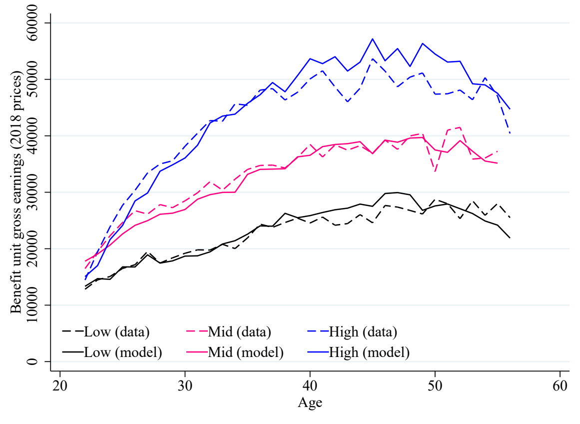


(d) 4-year change

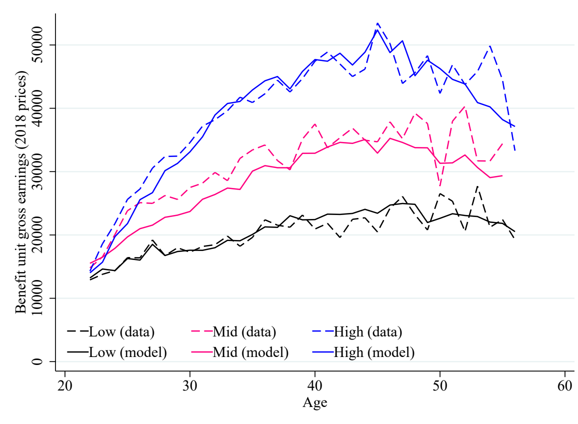


(e) 5-year change

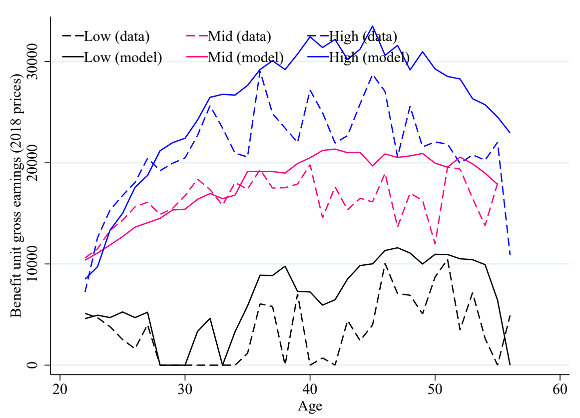
Figure 9: Comparison of distribution of simulated and actual earnings by age and education group for benefit units in the 1960s birth cohort



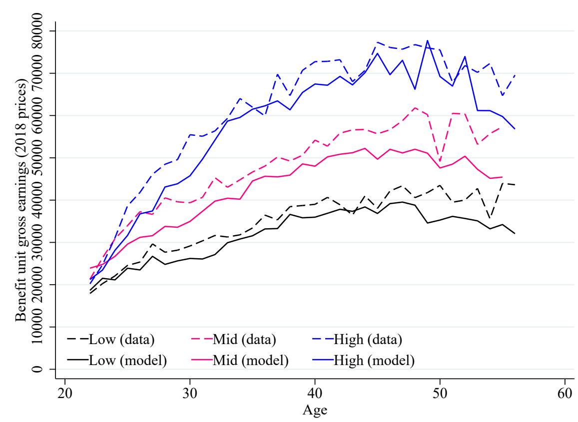
(a) Mean



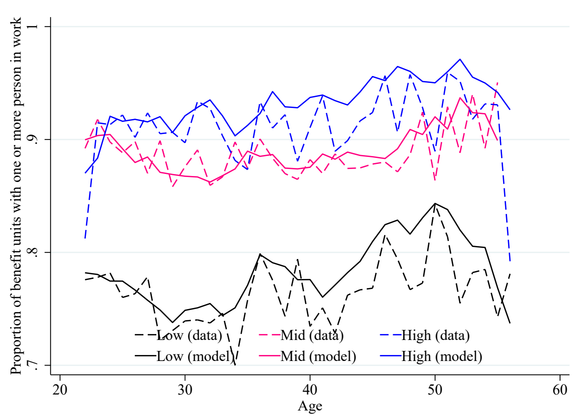
(b) Median



(c) 25th percentile

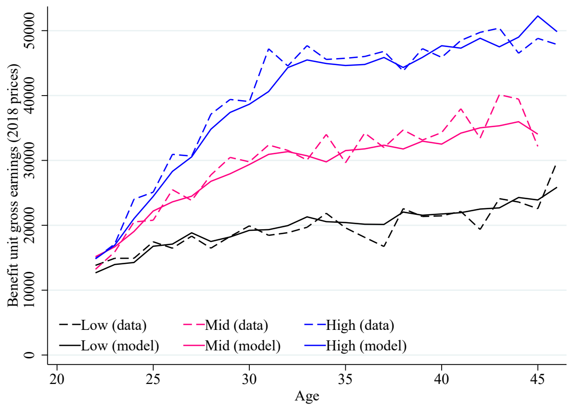


(d) 75th percentile

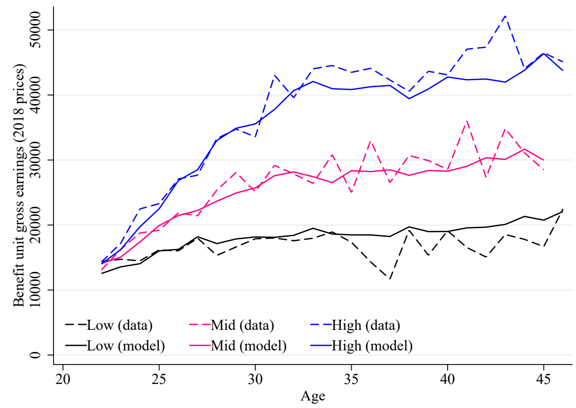


(e) Employment rate

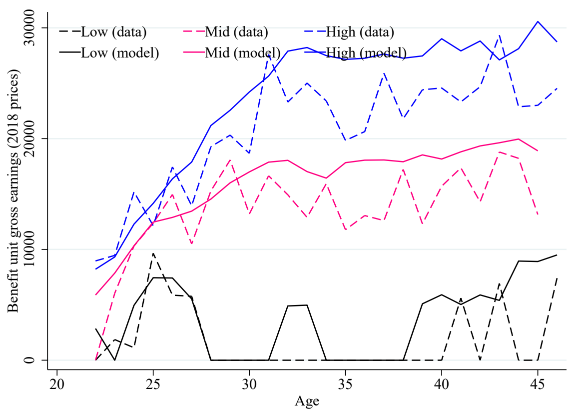
Figure 10: Comparison of distribution of simulated and actual earnings by age and education group for benefit units in the 1970s birth cohort



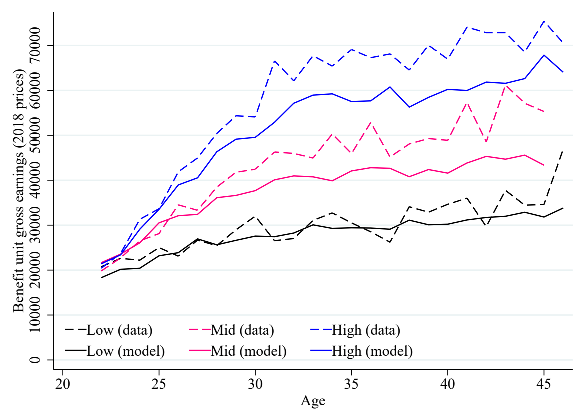
(a) Mean



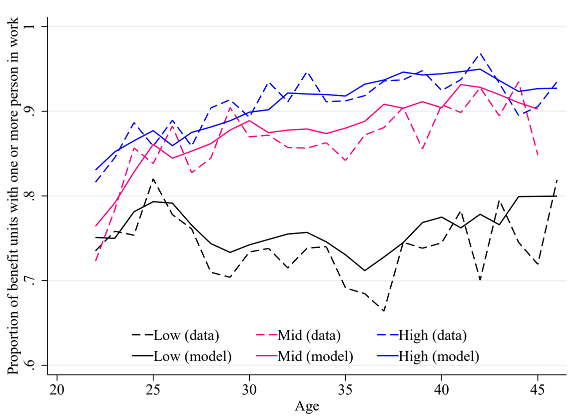
(b) Median



(c) 25th percentile

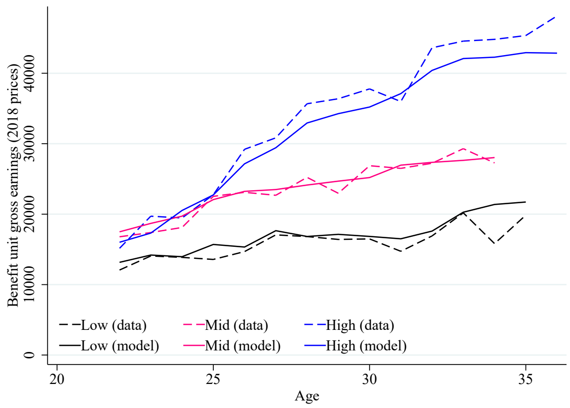


(d) 75th percentile

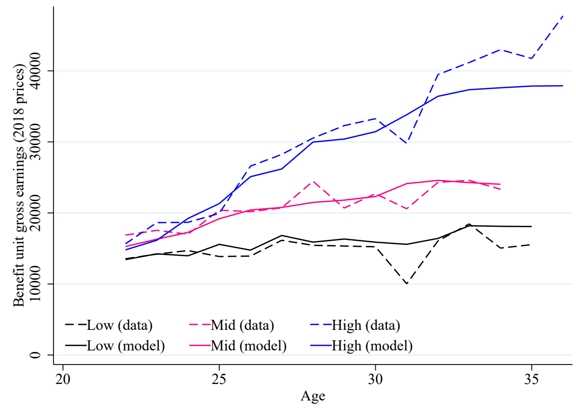


(e) Employment rate

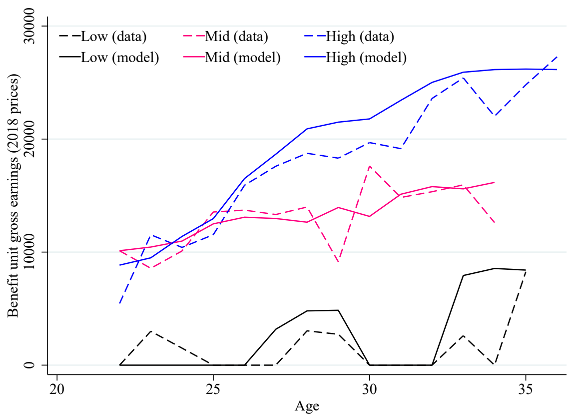
Figure 11: Comparison of distribution of simulated and actual earnings by age and education group for benefit units in the 1980s birth cohort



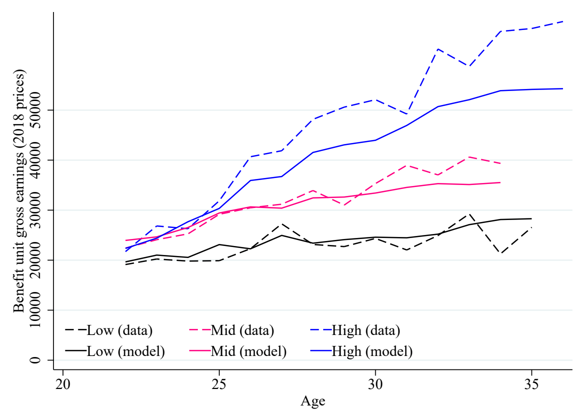
(a) Mean



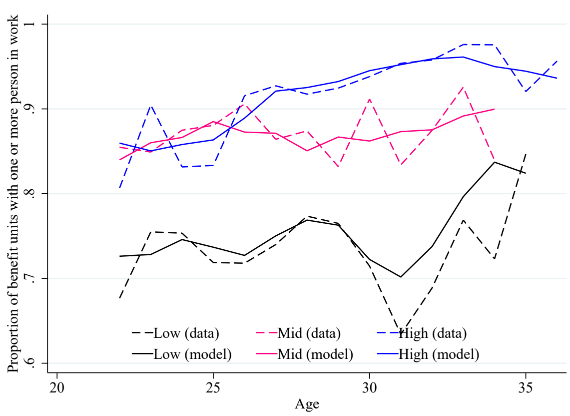
(b) Median



(c) 25th percentile



(d) 75th percentile



(e) Employment rate