https://doi.org/10.1140/epjds/s13688-022-00317-x

EPJ.org

EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE Open Access

A generative model for age and income distribution

Fatih Ozhamaratli[1*], Oleg Kitov[2] and Paolo Barucca[1]

*Correspondence:
fatih.ozhamaratli.19@ucl.ac.uk
[1] University College London, Gower Street, WC1E 6BT, London, UK
Full list of author information is available at the end of the article

Abstract

Each individual in society experiences an evolution of their income during their lifetime. Macroscopically, this dynamic creates a statistical relationship between age and income for each society. In this study, we investigate income distribution and its relationship with age and identify a stable joint distribution function for age and income within the United Kingdom and the United States. We demonstrate a flexible calibration methodology using panel and population surveys and capture the characteristic differences between the UK and the US populations. The model here presented can be utilised for forecasting income and planning pensions.

Keywords: Income dynamics; Agent based model; Pension system

1 Introduction

A universal element of societies is the emergence of hierarchical organisation structures within professions. People develop work experience through time and manage to obtain jobs of increasing responsibility and increasing level of income with time. Hence, it is a natural property of income distribution to be correlated with work experience and age; nevertheless, most income models do not study the relationship between income and age, and consequently between income distribution and demographic changes. This paper introduces a model of income, dependent on age-specific model parameters and random shocks. The model contributes to the understanding of the relationship between age and income and its dynamics.

Our aim is to compare the estimated parameters in the UK and the US age and income distribution to find out similar characteristics of age and income across states, as well as the contrasting differences. A simple age and income model is fundamental for the development of a sustainable pension system. The model focuses on the age and income relationship and further factors, such as occupational levels, are not considered. The model is estimated via panel survey data from the UK and population survey data from the USA. The data from panel surveys track the same individuals for the duration of the survey, and the population survey is repeated with different people each wave. The results reflect a clear income-age relationship in the UK and US, a clear structure of the joint distribution characterised by rapidly increasing income at younger ages, followed by income levels stabilising near mean income but spreading till retirement. At this point, the income

Springer

decreases and concentrates around mean retirement income. The paper demonstrates a flexible methodology to estimate parameters from population surveys, as well as panel surveys. The paper provides a simple generative model to evolve age-income population for simulation and forecasting purposes, which can constitute the foundation for future studies of financially sustainable pension systems by providing a benchmark for capturing age and income relationship. The purpose is to have a baseline model simple enough for isolating age and income relationship of income dynamics. Such a model will serve for investigating the properties of a sustainable and balanced pension system. The mean and standard deviation statistics from the panel and population surveys on Fig. 6 and Fig. 1 from observed panel data and simulation results reflect a clear relationship between age and income. More complex models, which investigate additional factors, and profile heterogeneity of income dynamics are out of the scope of our work.

Previous research on income have been conducted, and the research focused on investigating and explaining wage dynamics. Champernowne explicitly introduces a first-order Markov process to model the time-evolution of wages [1]. Following Markov process path, the validity of the first-order Markov assumption is tested by Shorrocks [2]. Following research introduces a second-order Markov process, yet neither of these works links individual wage dynamics to time-evolution of the distribution of wages [3]. A different approach focused on poverty, which deals with modelling individual data using linear regression and transitioning to poverty (probit model) [4]. A more comprehensive model incorporating various factors is developed to estimate transition probability in wage quintiles conditioned on various regressors, including education, experience and age [5], furthermore study both intra- and inter-group inequality. The persistence of the low pay state and factors affecting the low pay probability are expressed with a generalized regression model. For modelling low income transitions the previous research use British Panel Data for the ′90s, focus on the transition probability and state dependence for the poverty status [6] and define poverty transition equation as coarse-grained dynamics. Inequality and upward mobility between quintiles considering gender effects are investigated [7]. The previous models in literature either incorporate numerous external variables, distribution characteristics and functions, such as innovation constants or they are limiting their scope to the investigation of dependence on a single variable [8, 9]. A more recent article by Guvenen investigated a model for which focal variables are the human capital consisting of education, work experience, and idiosyncratic shocks [10], following research modelled male income for studying the impact of labour income taxation policy on inequality [11] The referred life-cycle model's distribution characteristics of the pre-tax income arise from the differences in the individual's ability to learn new things and idiosyncratic shocks. Previous research tried to capture the income dynamics with Markov Models, linear autoregressive models, or by relying on econometric toolset such as covariance matrices. We investigate a generative model with an empirical distribution for sustainable age and income relationship in a population; we achieve this via an income evolution model with an age-dependent parameter, estimated from previous population and panel surveys.

The previous research [12] presents a two-class distribution, majority of population is described by the exponential function and small fraction of population of higher income individuals are described by power law [13]. The BHPS and IPUMS CPS data are topcoded and not suitable for studying the power law at the top, but the majority of the population as reflected on Fig. 3 is consistent with empirically well-established exponential distribution of income [14].

Although there are models that incorporate indirectly the age as years of experience in job for studying income dynamics. There are no studies, to the best of our knowledge, studying the joint distribution of age and income in the scope of income evolution.

In contrast to previous research, our study introduces a dynamic model that describes the income-dependent only on age and previous income. This paper investigates the stationary property of the income distribution dependent on age. We provide a model in which the mean and variance of income given age are preserved at any time point.

2 Methods

We introduce a simple model which focuses on age and income relationship and differs from recent literature by not incorporating other variables such as occupational level, level of education and skill coefficients. The model is stationary, i.e. the mean and variance of income given age are preserved in time. The model is utilised to represent observed panel data for gaining empirical insights regarding age dependant, income dynamics and mobility. The calibrated model can be utilised as a simple generative model to evolve an age-income population for simulation purposes and it provides a theoretical background for studies focusing on ageing and pension income of the population. We initially assume the following model, by which $\mu(\cdot)$ and $\sigma(\cdot)$ represent a function of age, income, and individual-specific additional parameters θ_i or λ_i, for the sake of generality. $\mu(\cdot)$ is a function capturing mean income characteristics, and $\sigma(\cdot)$ captures the variational characteristics of the income. We consider the following individual income stochastic process for an economic agent i characterised at each time step t by a given age a_i and income y_i:

$$y_{i(t+1)} = \mu(a_{i(t+1)}, y_{it}|\boldsymbol{\theta}_i) + \sigma(a_{i(t+1)}, y_{it}|\boldsymbol{\lambda}_i)\eta_{it}. \tag{1}$$

The characterising insights on Fig. 1 from the panel data lead to the assumption that the probabilistic step at time t depends only on the age and income of the preceding step.

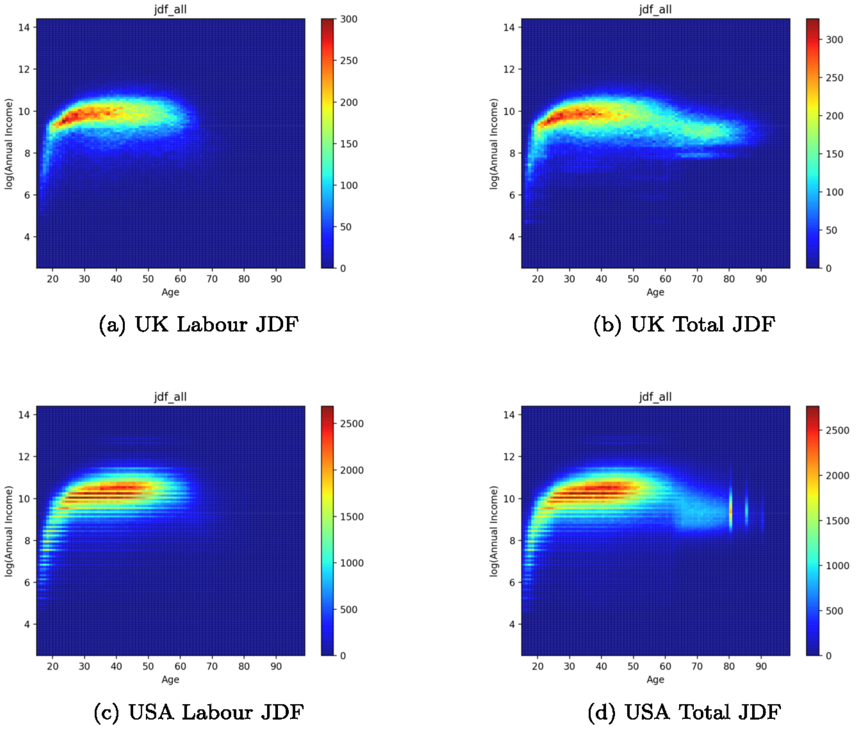
2.1 Defining income and age dynamics

Earnings of individual i at the time step t is denoted as $Y_{i,t}$ and its logarithm is $y_{i,t}$. The parameters that describe the income process are: age-dependent persistence parameter q_a, age-dependent mean μ_a and age-dependent standard deviation σ_a. The income shock process consists of independent random shock η_t^i which is normally distributed with mean zero and variance 1, and it is applied to σ_a, the model can be defined as follows:

$$y^i_{a+1,t+1} = q_a y^i_{a,t} + \mu_a + \sigma_a \eta^i_t. \tag{2}$$

Averaging income y, for individuals who are a years old, gives $\bar{y}_a$,which denotes the average income for age group a across all individuals i and periods t. Assuming that the age-dependent income profiles are stationary, we can average incomes $y^i_{a,t}$ across individuals and time to get the following equation:

$$\bar{y}_{a+1} = q_a \bar{y}_a + \mu_a, \tag{3}$$



(a) UK Labour JDF

(b) UK Total JDF

(c) USA Labour JDF

(d) USA Total JDF

Figure 1 All Years Pooled Age and Income Joint Distribution Function for UK and USA between Ages 15–100 where labour and total income heatmap reflect unique shape. Income after retirement can be observed on Total Income JDFs which include pension and capital income, the data from USA is top-coded which is reflected by stripes of concentrated concentration around ages 80, 85 and 90

where $\bar{y}_a$ denotes the average income for age group a, taken across all individuals i and periods t. The following equation can find the estimator for μ_a:

$$\mu_a = \bar{y}_{a+1} - q_a \bar{y}_a. \tag{4}$$

The income data from different waves are inflation adjusted to isolate effects of economic growth.

2.2 Data

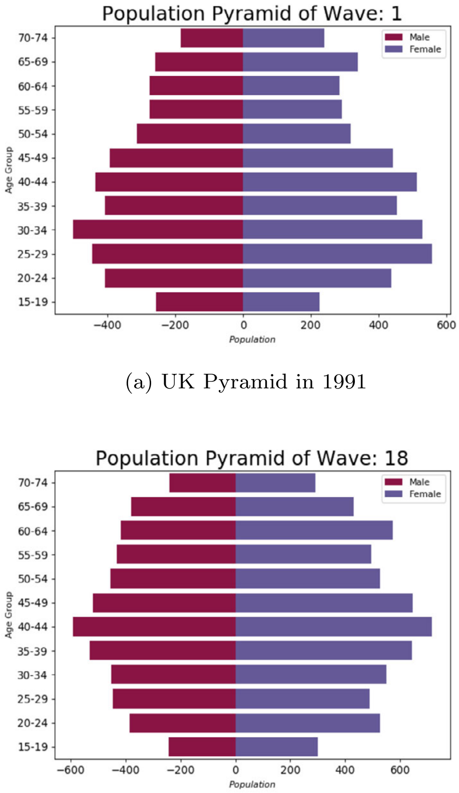
The British Household Panel Survey (BHPS) [15] from the UK and The Current Population Survey (IPUMS CPS) [16] from the USA are used for estimating the parameters of the model in Eq. (2) and comparing the results of simulated data and surveys. The BHPS is a Panel Survey conducted between 1991–2008. For our model we focus on labour income data, which captures wage, salary or self-employment income. To investigate population characteristics, we also incorporate other income sources and call it "Total Income", which additionally captures the transfers, pensions, grants, aids, state-benefits, dividends, capital income and rents. BHPS provides individuals specific longitudinal weights for ensuring the representativeness of the population. Two types of weights are provided with BHPS. The first wave is weighted for adjusting population marginals at the households and post-stratified to the population age by sex marginals. Consecutive waves are re-weighted to take into account sample attrition, variables such as address change, household region, age, sex, race, employment status, income total and composition, educational qualifications [17]. Panel Survey is conducted via questionnaires with tracked individuals of the initial sample. Detailed explanation of the relevant variables from BHPS dataset can be foınd on the Appendix. There is an extension to the sample population in 1999. For the USA, IPUMS CPS is used, which is annually conducted with different samples each year. In contrast with BHPS, the Labour Income from IPUMS CPS does not include self-employed income, and the weights are cross-sectional.

Income distribution, age distribution and income-dependent age distribution from the surveys are utilised for parameter estimation and further analysis. q_a, μ_a, σ_a are the key parameters estimated according to the proposed model. Following investigation and interpretation of the estimated parameters, these parameters are used to simulate the population's income transitions. The simulation is initialised using the panel data from the Wave 1, and the income evolution function on Eq. (2) is applied transitively in an iterative approach to the data for simulating successive waves. The simulated data is plotted, interpreted and finally compared with the observed data.

Figure 2 reflects the Population Pyramid in the UK, and how the shape evolved over the 18 years considered. The population pyramid of USA can be found in the Appendix. The UK population sample from BHPS has a relatively balanced population with a slight weight towards younger cohorts initially in 1991, which denotes Wave 1. The UK population gradually got older, and the population pyramid reflects mass's shift towards older generations, this shift happened gradually over the years. The US population from CPS reflects a young population in Wave 1 with a notable skew towards younger cohorts, after 17 years the US population loses this property towards younger cohorts and gets significantly older. Both the UK and US population get older and reflect a trend towards an ageing population, which will significantly impact the pension system.

The shape of the population pyramid and its evolution with time from the panel survey reflect an ageing community [18]. JDFs of Total and Labour Income in the UK and USA reflect that, there is a sharp increase between the ages 15–20, which can be interpreted as the beginning of the work-life, transitioning from part-time work to full-time work, and graduation from higher or vocational education. The most significant difference of the UK Total JDF in contrast to the UK Labour JDF is the tail section corresponding to the retired population, which denotes the significant percentage of individuals older than 55. The tail section is relatively concentrated, which can be explained by the state pension benefit levels and mandatory social security system. The US population reflects a surprisingly sparse older cohort for the Total Income data, and the most significant difference to the UK is the relatively lower income levels compared to the wage income. In the US population higher variance spread to a wider band, which might be caused by a non-standard retirement system not supported by strong state pension benefit and mandatory pension schemes during employment.

The comparison for the model simulation and observed data shows common characteristics, as the joint distribution of age and income in logarithmic scale is presented in Fig. 1: an initial sharp rise between the entrance to the graph on 16 years old, the amount of 16 years old includes pocket money, allowance and part-time or internship jobs. There is a steep increase in mean and variance between the initial income and income at the age of 20. The increase is sharper for the mean in comparison to the variance. The population's mass has similar characteristics with near 23k GBP annual income, and for ages between



(a) UK Pyramid in 1991

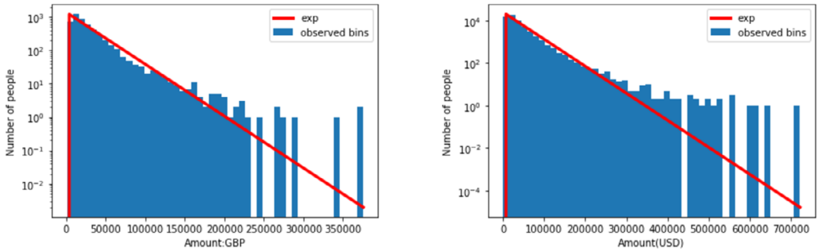
(b) UK Pyramid Wave in 2008

Figure 2 Population Pyramid for the UK Income between Ages 15–100 reflecting changes of age distribution in 18 years, which reflects mostly an aging population

20–45. Between ages 65–75, there is a significant decrease in income and after 75, the income converges to a certain mean. The data and the models provide an essential tool to tackle problems related to an ageing population and shocks introduced by technological and political changes.

In compliance with the literature [12], the population is divided as two-classes, and the majority of the population covering low and middle income follows Boltzmann–Gibbs distribution. The observed and simulated populations from the UK and USA are reflecting exponential characteristics for low and mid-income individuals, log-linear PDF plots reflect similar PDF characteristics on Fig. 3.

In the following sections we will focus on labour income and employed labour population. Total Income covers all of the income streams including transfer income such as pocket-money, labour income, capital income, and pension income; these different streams might be governed by varying dynamics non-uniform across the type of income; so we decided to focus on labour income, which involves the broadest section of the population; with most significant impact. The only other primary source of income in terms of gross value is the capital income, which might be significantly affected by other factors such as inter-generational shifts, market conditions and global financial state. In order to focus on labour income dynamics, the other income sources are left out of our modeling.



(a) PDF of UK Labour Income

(b) PDF of USA Labour Income

Figure 3 PDF Plots of the labour income bins of population and corresponding exponential distributions after 18 Simulated Waves, which confirms the exponential behaviour for low and mid-income individuals

3 Data processing and calibration

BHPS provides a vast amount of socio-economical data for each individual and household participating in the study. The columns of income, age data, the individual's statistical weight representative of the British population and overall survey- with the individual's intra-wave unique identifier mostly suffice for this paper's scope. PID, Wxage12, wFIYR, Wxrwght fields of BHPS are used for each wave.

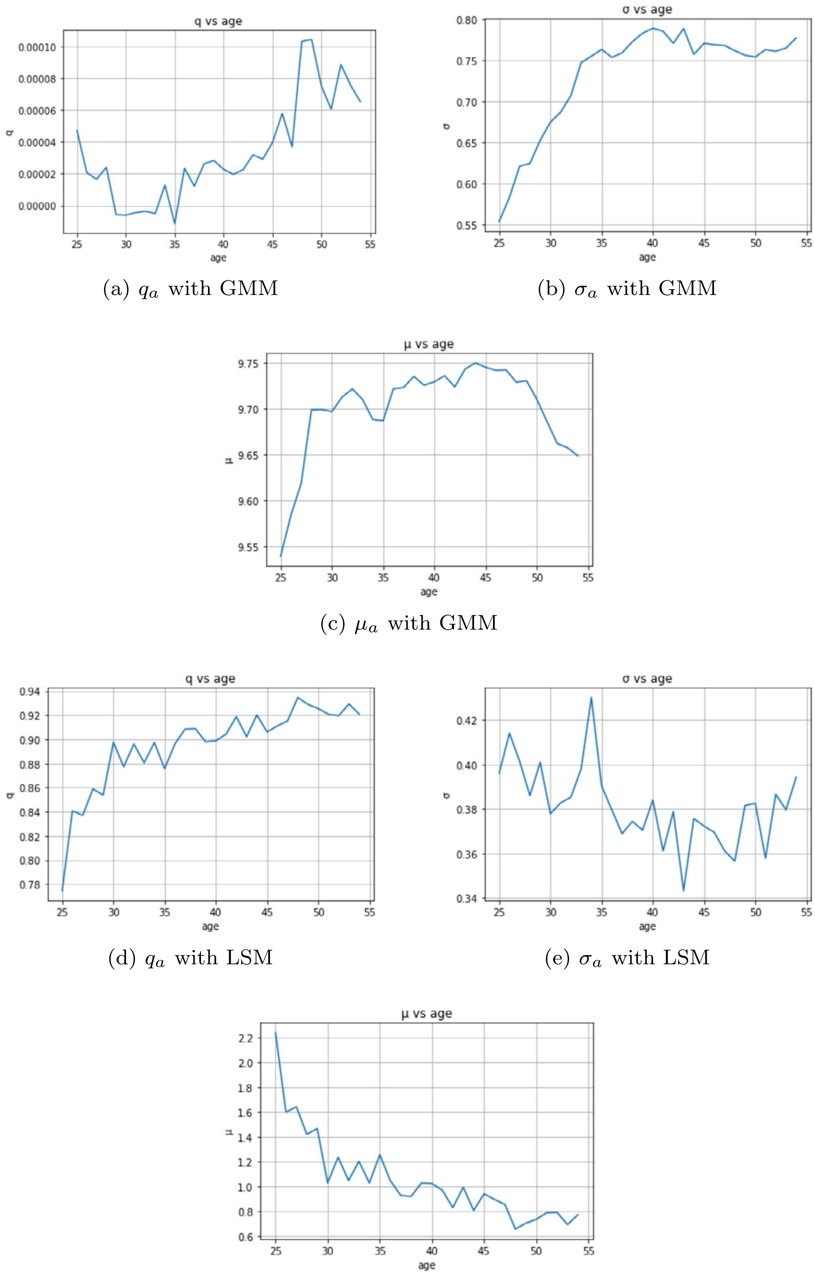
The Income variable xfiyr is each individual's annual income, including labour income, benefits, pensions, transfer income, and investment income. Participants were asked according to annual income in the reference year from September in the year prior to the interview until September in which interviewing begins [17]. The income figures are adjusted for inflation, as part of pre-processing. During the dataset preparation, a floor wage is determined to exclude in labour income, which denotes to excluding part-time and short-employment income. The income data is inflation-adjusted and transformed into log-domain.

IPUMS harmonizes the CPS and provides IPSUM CPS micro-data. The IPSUM CPS includes a large spectrum of topics such as demographics and employment, as well as supplemental studies such as the Annual Social and Economic Supplement (ASEC). Each individual can be identified by "CPSIDP", "INCTOT" and "INCWAGE" correspond to the total income and wage income, and "ASECWT" denotes the weights derived from ASEC Supplement. The data set is topcoded, and specific codes are used for labelling missing and incorrect data. The ages over 80, 90 and 99 are top-coded till 2004, and after 2004, the top-coding bins are determined as ages 80, 85, 90 and 99 by the panel data collectors [16]. Although this dataset contains high-income individuals, there is top-coding applied, so individuals with very high income are not included.

3.1 Fitting distributions

Estimating the income evolution function parameters is the most critical part of the research, and the decision depends on various factors such as the type of data, bias, and assumptions. Various techniques are investigated, leading to different results, with each having unique strengths and weaknesses.

The first method investigated is Generalized Method of Moments (GMM), which presumes that the first three moments of the income evolution functions provide the necessary information for approximating the underlying generative process. The equations of the first three moments of the income evolution equation can be solved for the parameters



(a) q_a with GMM

(b) σ_a with GMM

(c) μ_a with GMM

(d) q_a with LSM

(e) σ_a with LSM

(f) μ_a with LSM

Figure 4 q_a, σ_a and μ_a Plots for UK Labour Income, the parameters are estimated with two different methodologies of Generalized Method of Moments and Least Squares Method. The LSM utilises longitudinal data, which provides capability to estimate substantially high q values, which represent the persistence level of an individual's income, GMM capture much lower q values, due to its incapability of utilising longitudinal data

q_a, σ_a, μ_a. Both of the BHPS from the UK and the IPUMS CPS from the USA can be used for estimation with Generalized Method of Moments (GMM) with first three moments as reflected on Fig. 4 and on the Appendix.

The second method utilises the micro-data from the longitudinal surveys, which tracks the individual for consecutive years. The parameters are approximated to fit the income evolution function using least squares minimisation for the individuals participating in the studies for consecutive years. The BHPS from the UK is a suitable micro-data consisting of a panel survey, and the survey tracks income of the same individuals over the years.

3.2 Estimation for generalized method of moments (GMM)

The three moments of the age income evolution function are utilised to find a polynomial; afterwards, the equation is solved for q_a, σ_a and μ_a; at this point, a observed solution for parameters is found, but the relationship captures only the dynamics of the first three moments. Calculations can be found in the Appendix. The statistical variables such as $\bar{y}_a$ are found for each wave and than averaged across waves for finding a one set of stationary variables, which can be used to estimate q_a, σ_a, μ_a. The details and derivations for the GMM estimation technique can be found in the Appendix.

3.3 Estimation of least squares for micro data

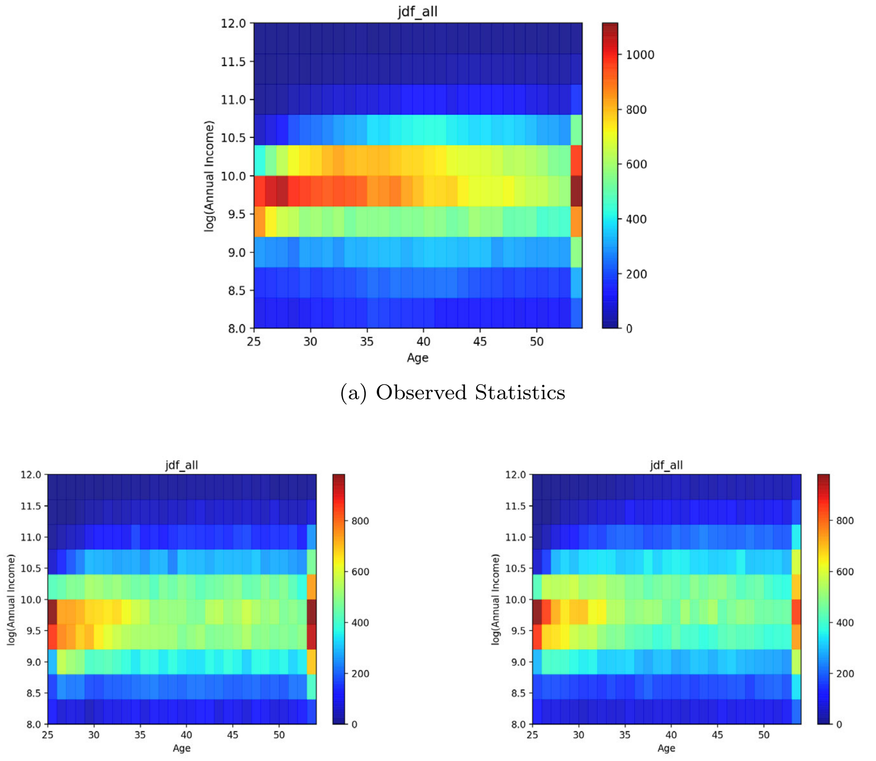
The Least Square Method requires that an individual's income for two consecutive years be existent in the dataset, this restriction is fulfilled by the BHPS, a panel survey, but the CPS IPUMS population survey does not satisfy this condition. The income data from two consecutive years per agent is used to estimate age-specific parameters, which characterise the income evolution function at Eq. (2). LSM tries to estimate parameters by fitting the data to the income evolution function.

4 The generative model

The model can also be used for simulation and forecast, tracking income trajectories of the individuals, providing a bench table for observing the stylized facts and complex properties of the income dynamics. Following the estimation of model parameters, the model is bootstrapped with data from Wave 1 for initialising the simulation. Each individual from Wave 1 is initialised as an agent in our model. According to Age Income Evolution Dynamics Eq. (2), the income of an agent is transitively updated at each consecutive wave update. η_{it} provides the random feed, which introduces variability for the income evolution of the agents. At each wave update, a new generation of agents consisting of 25 years old individuals from the initial wave are injected. Following each wave, distributions corresponding to the state of the simulated population are calculated. A full calibration of the model is shown in the Appendix.

5 Parameters

The optimized performance of these three methods are compared and discussed in the following sections. No boundaries are explicitly imposed by LSM estimation. The μ, σ and q variables are independent of each other, but the estimation process or data itself can introduce a slight dependence. The GMM estimation technique results in minimal q values near 0, so the estimated parameters approximately resemble an auto-regressive model. However, despite near 0 negligible q values, the q plot has a distinct shape with an increasing trend with a small decrease between 25 and 30, has very different characteristics depending on the estimation method. The GMM estimation method results in minimal q and the μ reflects the characteristics of $\bar{y}$, which is in compliance with this estimation method's nature. The μ value increases at first and then plateaus and slightly decreases near retirement. On the contrary LSM estimation mainly characterises the income with an increasing q parameter, so the μ parameter has limited effect and reflects a



(a) Observed Statistics

(b) Simulation Statistics with GMM Estimation

(c) Simulation Statistics with LSM

Figure 5 UK Labour Data Observed and Simulation All Years Pooled JDF between ages 25 and 55, the stripe of concentration at last column is due to concatenation of last two ages to fit the plot

decreasing trend. σ values reflect a distinct trend of initially decreasing values with a spike around the age of 34 followed by a stable decrease and noisy plateau with a minor increase towards 55. The LSM with bootstrap is the most accurate estimation method and reflects the characteristics of the model clearly.

5.1 GMM

GMM estimation technique approximates the μ_a values to be consistently around 10 and the q_a values are around 0 with an initial sine-like wave followed by a steady increase. The σ_a values are around 0.8 and have a positive trend. q_a values display a positive trend as well. The $\bar{y}_a$ and std$(y)_a$ plots of the simulation is similar to the observed data, but the standard deviation plot is particularly noisy. The JDF of the simulation on Fig. 5 is sparse, consistent; but not highly concentrated around mean. Both of these methods depends on assumptions about the dynamics of the income evolution function. The GMM method assumes that the first three moments of the equation are sufficient for estimating the parameters because they provide a solvable system. However individual characteristics in an age group such as different income levels and clusters within are lost during the moment estimation.

5.2 LSM by individual transitions

In order to use LSM for approximating the parameters, one needs the individual income transitions in consecutive years, thus identifying the same individual in consecutive cohorts is necessary and the panel studies such as BHPS satisfy this condition. The age-dependent income evolution function is fitted with individual income transitions of consecutive years, which results in consistent parameter plots and the $\bar{y}_a$ plot of the simulation reflect similar shape with the observed data on Fig. 6. The JDF of the simulation on Fig. 7 is able to reflect the dispersion among various clusters better because unlike the other methods heavily depending on the statistics such as mean and standard deviation of the entire age group, the LSM utilises individual-level microdata.

The 95% confidence interval with 2000 bootstrap samples for the estimated parameters from UK microdata by LSM can be found on Fig. 8. It is evident from the plots of $\bar{y}_a$ and $\mathrm{std}(y_a)$ for the observed and simulated data that the model can capture the characteristics of the income conditional on age distribution. A close investigation of simulations on models calibrated with UK Labour Income Data suggests that the GMM is most successful for reflecting the outcomes with similar mean and standard deviation characteristics of all waves after simulation with 18 waves that were simulated with the parameters q_a, σ_a, μ_a estimated by the GMM. But LSM reflects the individual trajectories, and JDF more accurately. The results showing the performance of GMM method can be found on the Appendix.

A general analysis of the comparison of joint distribution of age and inflation-adjusted income results in the following plots for weighted observed data and simulated data in Fig. 5: JDF of the simulated UK Labour data is in parallel with the expectations for GMM Estimation method, consistent and stable, resembling a similar shape but not concentrated for the heat regions with intense concentrations on Fig. 6. The main difference between the observed and simulated JDFs is the concentration of the mass of the population between 23 and 50.

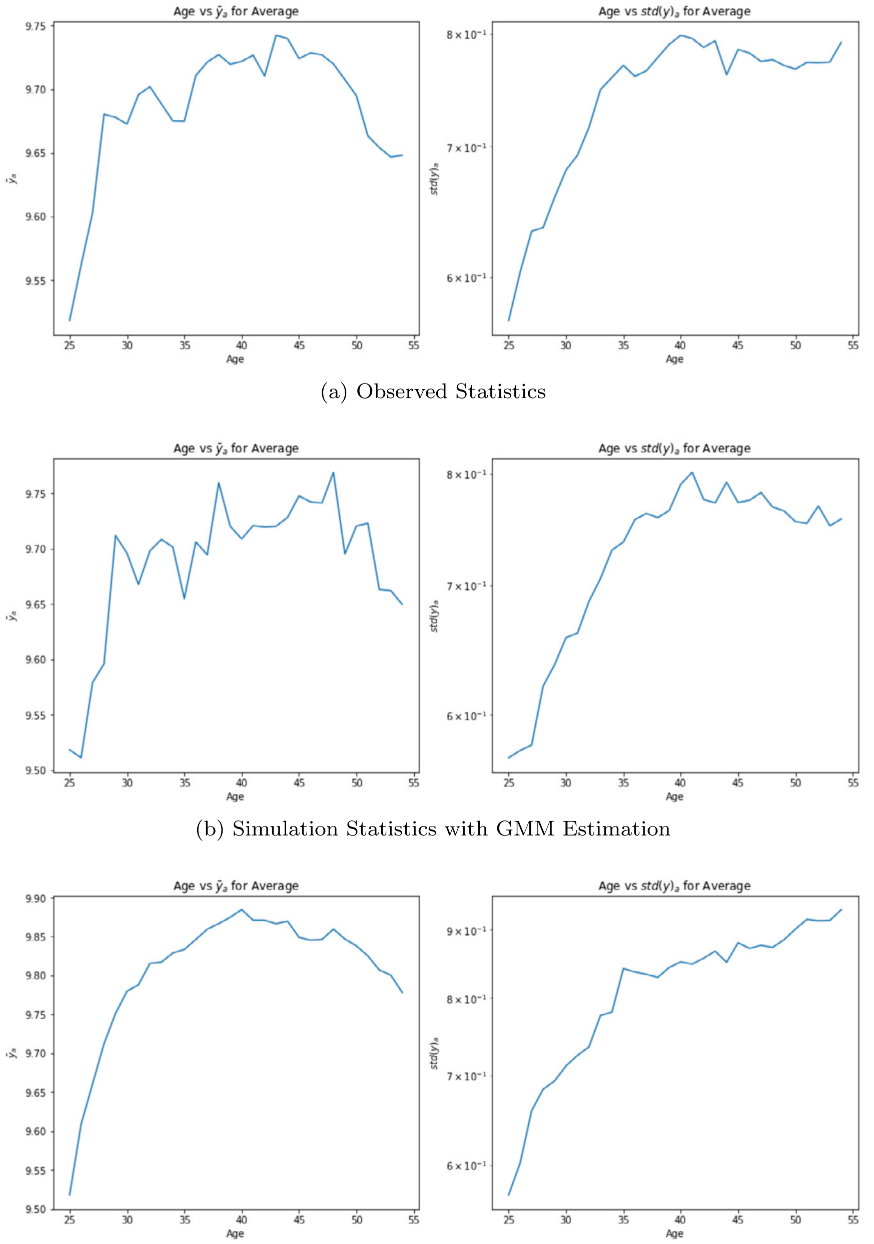
5.3 Wave-specific analysis

The population from wave-1 is used for bootstrapping the simulation and the weights of the individuals are not incorporated to the simulation, because the income evolution Eq. (2) is the focus of this paper, and the main purpose is not the perfect representativeness of the initial wave. The new agent injection on 1999 by panel survey is reason of difference in the UK simulation and observed JDF plots. Although the simulation's initial state is bootstrapped as the unweighted dataset, starting from the second wave, the JDF of the simulated population resembles the characteristics of the JDF from the panel survey with the weighted population, which reflects that the model is successfully capturing income evolution dynamics.

5.4 A simple pension system

A financially sustainable pension system can be characterised by the balance between inflow and outflow of funds. The specifics and stability of pension system is out of the scope of this paper, and needs case specific detailed modelling. For a general demonstration, we assume simple inflow and outflow dynamics(Eq. (5) and Eq. (6)), which are derived to represent statistical properties of the savings and consumption. Figure 9 reflects the imbalance between inflow and outflow, which results in a deficit.

Pension is assumed to be annually £16,368, in light of the median net income before housing costs for all pensioners from DWP Pensioners Income Series in 2008/2009 [19]. Constant alpha for pension saving rate is selected to be 0.0775 to 0.2, in light of OECD Pension Report statistics [20].



(a) Observed Statistics

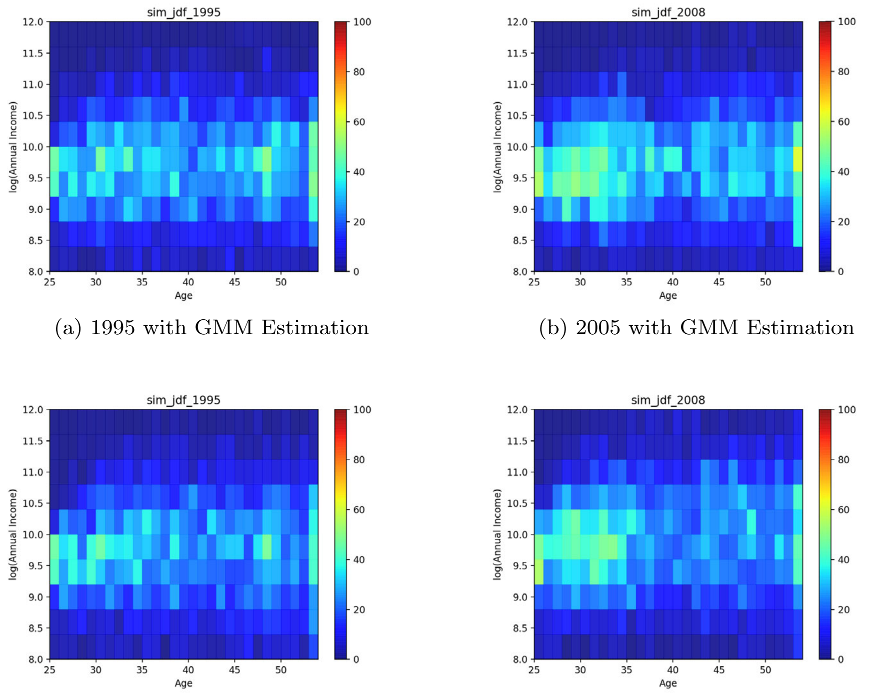
(b) Simulation Statistics with GMM Estimation

(c) Simulation Statistics with LSM Estimation

Figure 6 UK Labour Data Observed and Simulation Statistics representing average and standard deviation of the survey data and simulated population, which reflect similar shape with the statistics from survey data, further insight can be found on PDF plots

Outflow O_t in a given year t is characterised by constant annual pension amount p, and count of people above 65 $c_{a>65}$ is assumed to be pensioner counts.

$$O_t = pc_{a>65}. \tag{5}$$



(a) 1995 with GMM Estimation

(b) 2005 with GMM Estimation

(c) 1995 with LSM Estimation

(d) 2005 with LSM Estimation

Figure 7 JDF Plots of Simulation for UK Labour Income

Inflow I_t in a given year t is characterised by constant pension contribution rate α and total labour income of individuals y_a

$$I_t = \alpha \sum_i (y_{a \leq 65}). \tag{6}$$

The amounts are adjusted for inflation and reflect the 2009 levels. The inflow and outflow plots from our simplified generalisation of the pension system reflect a deficit.

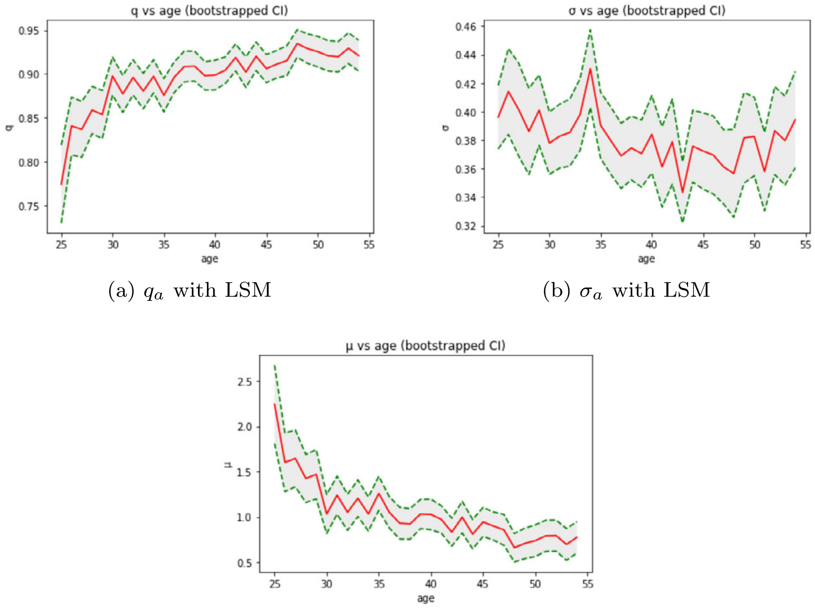
6 Discussion

The income evolution Eq. (2) of the proposed model consists of the parameters q_a, μ_a, σ_a: the persistence coefficient for the respective age group q_a, determines the rate of persistence at a given age.

Age-dependent mean income parameter μ_a expresses the expected age-specific income evolution mean for the next income and behaves such that if the mean parameter is high the persistence q_a is most likely to be lower. If the mean parameter μ_a is lower, the persistence parameter is higher which signals a potential widening of the income gap for the population.

σ_a captures the variability of the individuals according to conditional distribution and incorporates randomness of the shocks.

The social safety nets, basic pension incomes and the Defined Benefit Pension plans are financed via the working population; the ever-growing unbalance towards ageing cohorts needs careful forecasting and planning. The demographic shift will impact the economy's functioning in general, introducing a heavy burden to welfare states financing the health

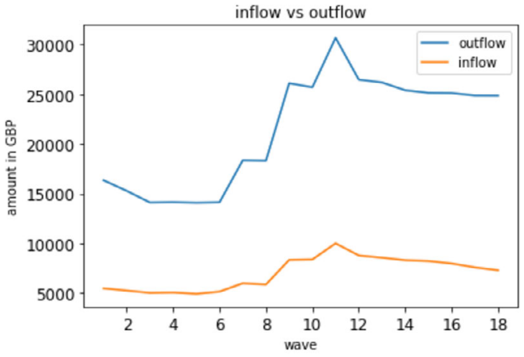


(a) q_a with LSM

(b) σ_a with LSM

(c) μ_a with LSM

Figure 8 q_a, σ_a and μ_a Confidence Interval for UK Data LSM Estimation reflecting bootstrapped parameter estimation values and the robustness of the estimation method



(a) Inflow & Outflow

Figure 9 UK Inflow Outflow Plot of our Simple Pension System reflecting inflow of pension savings(assuming savings rate of 0.0775) from contributors who are still in workforce, and outflow of funds to pensioners(assuming weekly pension of £308)

and pension of the retired population, which will reflect society as taxes and benefit cuts. The best course of action is forecasting the changes and planning in advance for the future.

6.1 Interpretation

The q_a persistence estimated by GMM reflects that UK population reflects an initially high q_a value in youth, followed by relative decrease, and then a consistent increase. The q values estimated by GMM fluctuate around 0 and minimal. The income persistence variable of individuals is not captured by GMM, which does not utilise panel survey's tracked individual income micro-data each year.

LSM calibration methodology uses longitudinal information regarding evolution of an individual's income, this property of LSM makes it suitable to provide higher q values, which captures the persistence. The GMM calibration methodology does not utilise longitudinal data, which makes it applicable with survey data but results in lower q values. This tells us that we are still able to reproduce consistently the data year by year almost disregarding the past year data. Our analysis suggests that LSM is a more reliable calibration model as using longitudinal information appears to be crucial to capture the income evolution dynamics. Furthermore the model has power to capture some heterogeneity, if the parameters are fit with LSM; and the income evolution function can preserve the bootstrapped wave heterogeneity due to persistence parameter q and the randomness injected with σ can provide outlier behaviour after multiple waves of income evolution function updates.

The LSM results in a consistently increasing q_a value by the UK model, with a significant jump between ages 25–30, which corresponds with a μ_a plot consistently decreasing with a significantly sharper decrease between ages 25–30. q_a and μ_a corresponding each other in an inverse proportion, especially by significant changes, especially by LSM. There are various examples of parameter effects that can be observed from BHPS dataset.

One example of q_a effect is the upward mobility of age-group between 25 and 30, which is reflected by the increasing q_a values and sharp increase observed on the joint-distribution plot. This effect can be due to finishing higher education and internships, in addition few years of experience, which results in a widening of income scissors. This change in mobility is healthy for the economy and does not represent a negative effect. One assumption should be researched further; if either this initial difference in mobility might limit of people with lower income for upward mobility.

An example of the σ_a mobility is the age group of 30–35, which is reflected by a locally sharp increase of σ_a values. Such mobility reflects a bidirectional movement of income for individuals, and such a variation might arise from the short-time employment, interruption of employment for education, temporary jobs and most importantly this mobility might be caused by the initial differentiation according to the education of individuals such as higher education or vocational education. This window represents an increase in the variation of the income.

In general, the shape of the distribution can be explained in three periods; the first period is the introduction to employment and teenagers, which represents income from part-time and temporary jobs at the beginning and start of full-time employment it sharply increases on Fig. 1.

The age group of 25–55 denotes the main productive era of the economic life, and the income reflects a high dispersion. All of the factors and random shocks act together and result in dispersed but a consistent distribution. Mobility wise this era provides opportunities for upward mobility and possesses downward mobility risks. At the end of this period, income tends to decrease slightly, which reflects a decrease in productivity. Another limiting factor is the minimum wage and state benefits, which introduces a lower bound envelope for the mass. Income sources and affecting factors of individuals in this era vary greatly, which results in the widest dispersion in the entire life-span. Some of the factors are education, social strata, adaptability to innovation, total-work hours per week, experience and expertness, seniority of the jobs and ageism. The third and final era represents the exit from the workforce and retirement, and temporary or part-time jobs for low-income old individuals. The income decreases gradually as the number of individuals exiting workforce increases with time, the income stabilises, and variation decreases significantly. Income in this era is relatively low, and the source is usually pension benefits, state support or temporary jobs. This model's outcomes can be used for various purposes; the most apparent fields for drawing consequences and planning are the works on inequality and mobility depending on age. Characteristics of workforce entrance, work-efficiency of individuals per age, the structure of the society, pension system, income stability, and the taxation system are the most obvious fields.

In the paper, two main estimation techniques are investigated, and the corresponding results from the simulated waves are presented. The first estimation method investigated is GMM Estimation. The income regions appear smoothed and spread. The second estimation method investigated is LSM, it utilises the microdata and is suitable for capturing an agent's income evolution. The JDFs from the simulated waves have the most similar mean characteristics to the observed data.

The LSM evidently performs better by utilizing longitudinal microdata; the GMM estimation method can be applied to both population and panel surveys, provides feasible distributions but with unrealistic modeling of an agent's individual income trajectory.

6.2 Conclusions

We demonstrated (1) a clear income-age relationship, which is reflected by the data from BHPS and IPSUM CPS, as well as simulations. (2) a clear structure of the joint age-income distribution in both the UK and USA. (3) a flexible methodology to estimate parameters from population surveys, as well as panel surveys. (4) a simple generative model to evolve the age-income population with real constraints for evaluating general policy scenarios, that is agnostic about occupation levels.

The model can be interpreted as delivering a premise that the information of an individual's experience and education can be encapsulated by income. Although in early career, the income dynamics are governed by the initial difference at the level of education and profession; the main dynamics governing income transitioning can be reduced to the relationship between income and age, which collectively encapsulate education and experience. These premises can be leveraged for developing simplified models for evaluating mobility, inequality, welfare state, and pensions.

The proposed model focuses on the evolution of age and income population and the paper successfully demonstrates a simple model that can be calibrated for age and income that can be used as a backbone for forecasting income and planning pensions. Understanding the dynamics and having the ability to forecast the age and income population is the key to the design of financially sustainable pension systems.

There are different dimensions for the future work: one of the dimensions is injecting random shocks to the distributions itself, which can be in the form of new population injection or withdrawal, as well as tuning the η_{it} with various means for simulating a global or regional shock, such as pandemics or mass migration. Stress-testing the age and income distribution for different labour market scenarios could lead to relevant policy implications. The second dimension for future work is modifying the simulation system to estimate parameters on the fly, and provide a more adaptive and granular version of the simulation system. The third dimension for future work is incorporating data encompassing more years and more countries and with a higher resolution in time to investigate the role of multiple economic factors for short, medium and long time horizons.

Appendix

A.1 Variables of BHPS dataset

Table 1 Description of Variables from the British Household Panel Survey spanning years 1991–2008

Column Name	Description
pid	Unique ID Describing an Individual
wAGE12	Age of Individual on 1st of December
wFIYR	Total annual income including labour income, benefits, pensions, transfer income, and investments
wFIYRL	Annual labour income
wXRWGHT	A cross-sectional respondent weight
wave	BHPS Wave Number between 1–18, Wave 1 denotes to 1991

A.2 Model calibration

We can define the mean and standard deviation of income at a given age a as following:

$$\big(\bar{y}_a, \mathrm{std}(y_a)\big), \tag{7}$$

$$\big\langle y^i_{a,t}\big\rangle = \bar{y}_a. \tag{8}$$

The standard deviation and mean has the following relation with the squared average of incomes:

$$\big\langle \big(y^i_{a,t}\big)^2\big\rangle - (\bar{y}_a)^2 = \big(\mathrm{std}(y_a)\big)^2. \tag{9}$$

η_{it} has characteristics of the standard normal distribution:

$$\langle \eta_{it}\rangle = 0, \tag{10}$$

$$\big\langle \eta^2_{it}\big\rangle = 1, \tag{11}$$

$$\big\langle \eta^3_{it}\big\rangle = 0. \tag{12}$$

Squaring both sides of income evolution equation (2) results in following distribution:

$$\big(y^i_{a+1,t+1}\big)^2 = \big(q_a y^i_{a,t} + \mu_a + \sigma_a \eta^i_t\big)^2. \tag{13}$$

Equation (9) can be formalized as:

$$(\bar{y}_a)^2 + \big(\mathrm{std}(y_a)\big)^2 = \big\langle \big(y^i_{a,t}\big)^2\big\rangle. \tag{14}$$

Placing Eq. (14) for $a + 1$ and Eq. (13) results in following equation:

$$(\bar{y}_{a+1})^2 + \big(\mathrm{std}(y_{a+1})\big)^2 = \big\langle \big(q_a y^i_{a,t} + \mu_a + \sigma_a \eta^i_t\big)^2\big\rangle. \tag{15}$$

Expanding the right side of the equation results in:

$$(\bar{y}_{a+1})^2 + \big(\mathrm{std}(y_{a+1})\big)^2 = \big\langle \big(q_a y^i_{a,t}\big)^2 + \big(\mu_a + \sigma_a \eta^i_t\big)^2 + 2\big(q_a y^i_{a,t}\big)\big(\mu_a + \sigma_a \eta^i_t\big)\big\rangle, \tag{16}$$

$$\begin{aligned}
&(\bar{y}_{a+1})^2 + \big(\mathrm{std}(y_{a+1})\big)^2 \\
&\quad = \big\langle \big(q_a y^i_{a,t}\big)^2 + \big(\mu_a + \sigma_a \eta^i_t\big)^2 + 2\big(q_a y^i_{a,t}\big)\big(\mu_a + \sigma_a \eta^i_t\big)\big\rangle && (17) \\
&\quad = \big\langle \big(q_a y^i_{a,t}\big)^2 + (\mu_a)^2 + \big(\sigma_a \eta^i_t\big)^2 + 2(\mu_a \sigma_a \eta_{it}) + 2\big(q_a y^i_{a,t} \mu_a + \big(q_a y^i_{a,t}\big)\sigma_a \eta^i_t\big)\big\rangle. && (18)
\end{aligned}$$

Averaging the equation by using Eq. (14), Eq. (10) and Eq. (11).

$$(\bar{y}_{a+1})^2 + \big(\mathrm{std}(y_{a+1})\big)^2 = (q_a)^2\big((\bar{y}_a)^2 + \big(\mathrm{std}(y_a)\big)^2\big) + (\mu_a)^2 + (\sigma_a)^2 + 2q_a\mu_a\bar{y}_a. \tag{19}$$

A.3 Deriving the update equations

For clarity $(\bar{y}_a)^2 + (\mathrm{std}(y_a))^2$ is expressed as $(\Delta_a)^2$, The number of parameters can be reduced to 2 using the third parameter of Eq. (19) by expressing μ_a as $\bar{y}_{a+1} - q_a\bar{y}_a$ according to Eq. (4):

$$(\Delta_{a+1})^2 = (q_a)^2(\Delta_a)^2 + (\mu_a)^2 + (\sigma_a)^2 + 2q_a\mu_a\bar{y}_a, \tag{20}$$

$$(\Delta_{a+1})^2 = (q_a)^2(\Delta_a)^2 + (\bar{y}_{a+1} - q_a\bar{y}_a)^2 + \sigma_a^2 + 2q_a(\bar{y}_{a+1} - q_a\bar{y}_a)\bar{y}_a \tag{21}$$

unpacking Δ:

$$\begin{aligned}
(\bar{y}_{a+1})^2 + \big(\mathrm{std}(y_{a+1})\big)^2 &= q_a^2\big((\bar{y}_a)^2 + \big(\mathrm{std}(y_a)\big)^2\big) + (\bar{y}_{a+1})^2 + (q_a\bar{y}_a)^2 \\
&\quad - 2(\bar{y}_{a+1}q_a\bar{y}_a) + \sigma_a^2 + 2q_a\bar{y}_a\bar{y}_{a+1} - 2(q_a)^2(\bar{y}_a)^2
\end{aligned} \tag{22}$$

expressions at both sides of the equation cancel each other and simplify as follows:

$$\big(\mathrm{std}(y_{a+1})\big)^2 = q_a^2\big(\mathrm{std}(y_a)\big)^2 + (\sigma_a)^2 \tag{23}$$

solving in quadratic equation form:

$$0 = q_a^2\big(\mathrm{std}(y_a)\big)^2 + (\sigma_a)^2 - \big(\mathrm{std}(y_{a+1})\big)^2 \tag{24}$$

for $(-(\sigma^a)^2((\tilde{\sigma}^a)^2 - (\sigma^{a+1})^2)) > 0$ and $(\sigma^a)^2 > 0$, $\tilde{q}$ values can be solved as follows:

$$\tilde{q}^a_1 = \frac{\sqrt{-(\sigma^a)^2((\tilde{\sigma}^a)^2 - (\sigma^{a+1})^2)}}{(\sigma^a)^2}, \tag{25}$$

$$\tilde{q}^a_2 = \frac{-\sqrt{-(\sigma^a)^2((\tilde{\sigma}^a)^2 - (\sigma^{a+1})^2)}}{(\sigma^a)^2}. \tag{26}$$

Following equations are used in the method of GMM:

Using unnormalized unstandardized third moment of the Equation (2)

$$E\big[(y_{a+1})^3\big] = E\big[(q_a y_a + \mu_a + \sigma_a\eta)^3\big]. \tag{27}$$

Expanding the cube equation

$$E\big[(y_{a+1})^3\big]$$

$$\begin{aligned} &= E\bigl[(q_a y_a)^3 + (\mu_a)^3 + \bigl(\sigma_a \eta^a\bigr)^3 + (6 q_a y_a \mu_a \sigma_a \eta) + 3(q_a y_a)^2 \sigma_a \\ &\quad + 3(q_a y_a)^2 \mu_a + 3(\mu_a)^2 q_a y_a + 3(\mu_a)^2 \sigma_a \eta + 3(\sigma_a)^2 \mu_a + 3(\sigma_a)^2 q_a y_a\bigr]. \end{aligned} \tag{28}$$

Using Eq. (12), $(\sigma_a \eta^a)^3$, η^3 equals zero

$$\begin{aligned} E\bigl[(y_{a+1})^3\bigr] = {}& (q_a)^3 E\bigl[(y_a)^3\bigr] + (\mu_a)^3 + 3(q_a)^2 \mu_a E\bigl[(y_a)^2\bigr] + 3(\mu_a)^2 q_a E[y_a] \\ & + 3(\sigma_a)^2 \mu_a + 3(\sigma_a)^2 q_a E[y_a], \end{aligned} \tag{29}$$

$$\begin{aligned} E\bigl[(y_{a+1})^3\bigr] = {}& (q_a)^3 E\bigl[(y_a)^3\bigr] + (\mu_a)^3 + 3\mu_a\bigl((q_a)^2 E\bigl[(y^a)^2\bigr] + (\sigma_a)^2\bigr) \\ & + 3 q_a E[y_a]\bigl((\mu_a)^2 + (\sigma_a)^2\bigr). \end{aligned} \tag{30}$$

Expressing $(\sigma_a)^2$ from Eq. (23) in terms of q_a

$$\begin{aligned} & E\bigl[(y_{a+1})^3\bigr] \\ &\quad = (q_a)^3 E\bigl[(y_a)^3\bigr] + (\mu_a)^3 + 3\mu_a\bigl((q_a)^2 E\bigl[(y_a)^2\bigr] + \bigl(\mathrm{std}(y_{a+1})\bigr)^2 - (q_a)^2\bigl(\mathrm{std}(y_a)\bigr)^2\bigr) \\ &\quad\quad + 3 q_a E[y_a]\bigl((\mu_a)^2 + \bigl(\mathrm{std}(y_{a+1})\bigr)^2 - (q_a)^2\bigl(\mathrm{std}(y_a)\bigr)^2\bigr). \end{aligned} \tag{31}$$

Replacing $E[y_a] = \bar{y}_a$ and $E[(y_a)^2] = (\mathrm{std}(y_a))^2 + (\bar{y}_a)^2$ from Eq. (9)

$$\begin{aligned} & E\bigl[(y_{a+1})^3\bigr] \\ &\quad = (q_a)^3 E\bigl[(y_a)^3\bigr] + (\mu_a)^3 + 3\mu_a (q_a)^2 (\bar{y}_a)^2 \\ &\quad\quad + 3\mu_a\bigl(\mathrm{std}(y_{a+1})\bigr)^2 + 3 q_a \bar{y}_a (\mu_a)^2 + 3\mu_a \bar{y}_a\bigl(\mathrm{std}(y_{a+1})\bigr)^2 - 3(q_a)^3 \bar{y}_a\bigl(\mathrm{std}(y_a)\bigr)^2. \end{aligned} \tag{32}$$

Expressing μ_a from Eq. (4) in terms of q_a

$$\begin{aligned} & E\bigl[(y_{a+1})^3\bigr] \\ &\quad = (q_a)^3 E\bigl[(y_a)^3\bigr] + (\bar{y}_{a+1} - q_a \bar{y}_a)^3 + 3(\bar{y}_{a+1} - q_a \bar{y}_a)\bigl(\mathrm{std}(y_{a+1})\bigr)^2 \\ &\quad\quad + 3 q_a \bar{y}_a (\bar{y}_{a+1} - q_a \bar{y}_a)^2 + 3 q_a \bar{y}_a\bigl(\mathrm{std}(y_{a+1})\bigr)^2 - 3(q_a)^3 \bar{y}_a\bigl(\mathrm{std}(y_a)\bigr)^2. \end{aligned} \tag{33}$$

Expressing in the form of cubic polynomial equation of q_a

$$\begin{aligned} 0 = {}& (q_a)^3\bigl(E\bigl[(y_a)^3\bigr] - (\bar{y}_a)^3 - 3\bar{y}_a\bigl(\mathrm{std}(y_a)\bigr)^2\bigr) \\ & + (q_a)^2\bigl(3\bar{y}_{a+1}(\bar{y}_a)^2 - 6(\bar{y}_a)^2 \bar{y}_{a+1}\bigr) \\ & + (q_a)\bigl(3(\mu_{a+1})^2 \bar{y}_a - 3\bar{y}_a\bigl(\mathrm{std}(y_{a+1})\bigr)^2 + 3\bar{y}_a(\bar{y}_{a+1})^2 + 3\bar{y}_a\bigl(\mathrm{std}(y_{a+1})\bigr)^2\bigr) \\ & + (\bar{y}_{a+1})^3 + 3\bar{y}_{a+1}\bigl(\mathrm{std}(y_{a+1})\bigr)^2 - E\bigl[(y_{a+1})^3\bigr]. \end{aligned} \tag{34}$$

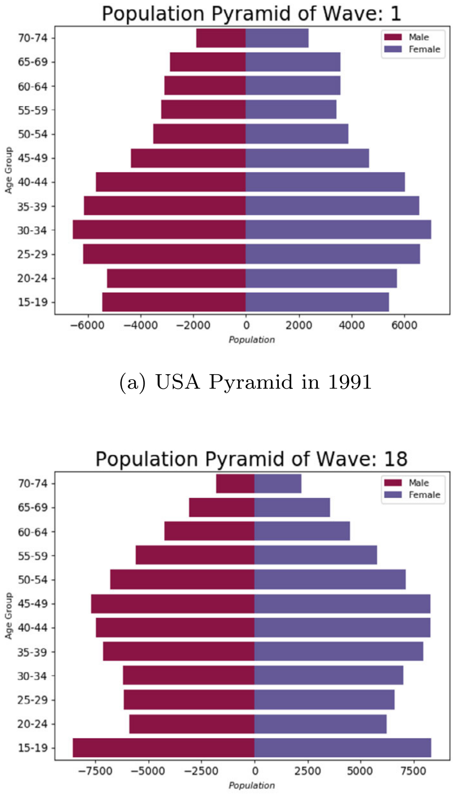
This equation can be solved for q_a corresponding each age group. Cardano solution for cubic equations guarantees single real root to exist, the other two complex roots that Cardano solution provides are not used. Both of the $\sigma_a = \mathrm{std}(y_a)$ and GMM estimation techniques can use the following equations for determining the μ_a and σ_a: For $(q_a)_1$ and $(q_a)_2$ according to Eq. (4):

$$\mu_a = \bar{y}_{a+1} - q_a \bar{y}_a. \tag{35}$$

The σ_a^2 can also be expressed in terms of q_a, using Eq. (4)):

$$\sigma_a^2 = (\Delta_{a+1})^2 - q_a^2(\Delta_a)^2 - \left(\bar{y}_{a+1} - q_a\mu^2\right)^2 - 2q_a(\bar{y}_{a+1} - q_a\bar{y}_a)\bar{y}_a. \tag{36}$$

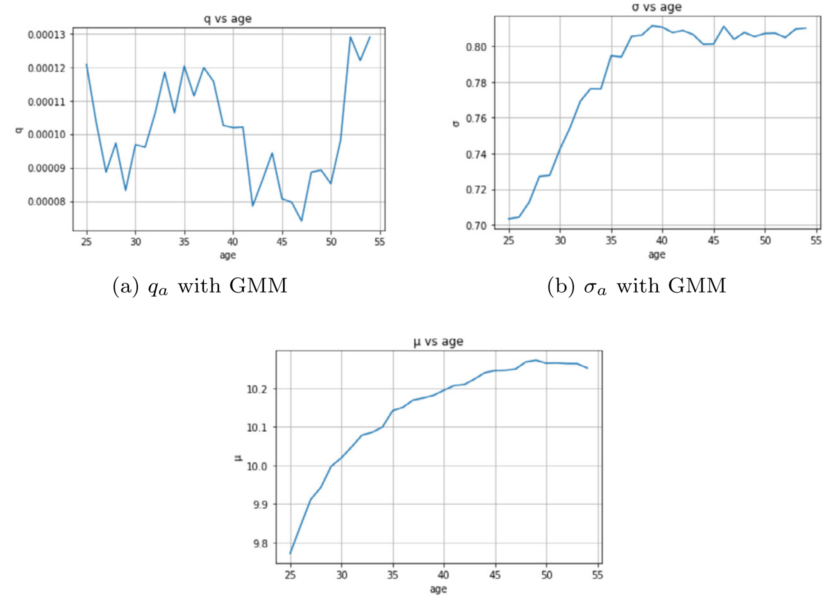
A.4 Supplementary plots of the USA



(a) USA Pyramid in 1991

(b) USA Pyramid in 2008

Figure 10 Population Pyramid for the USA Income between Ages 15–100 reflecting changes of age distribution in 18 years, which reflects mostly an aging population

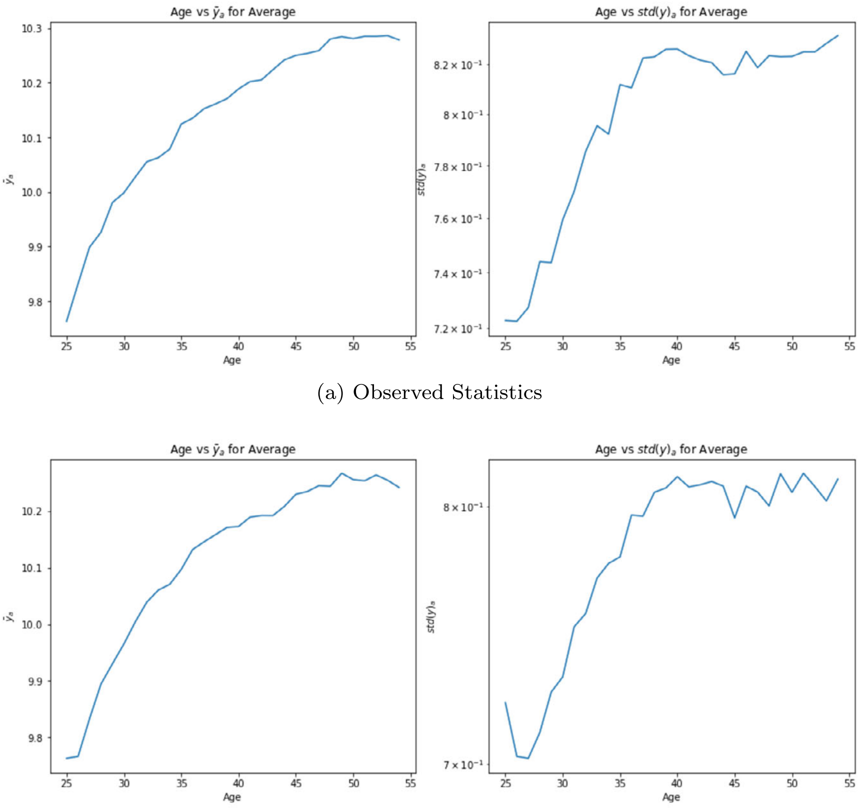


(a) q_a with GMM

(b) σ_a with GMM

(c) μ_a with GMM

Figure 11 q_a, σ_a and μ_a Plots for USA Labour Income

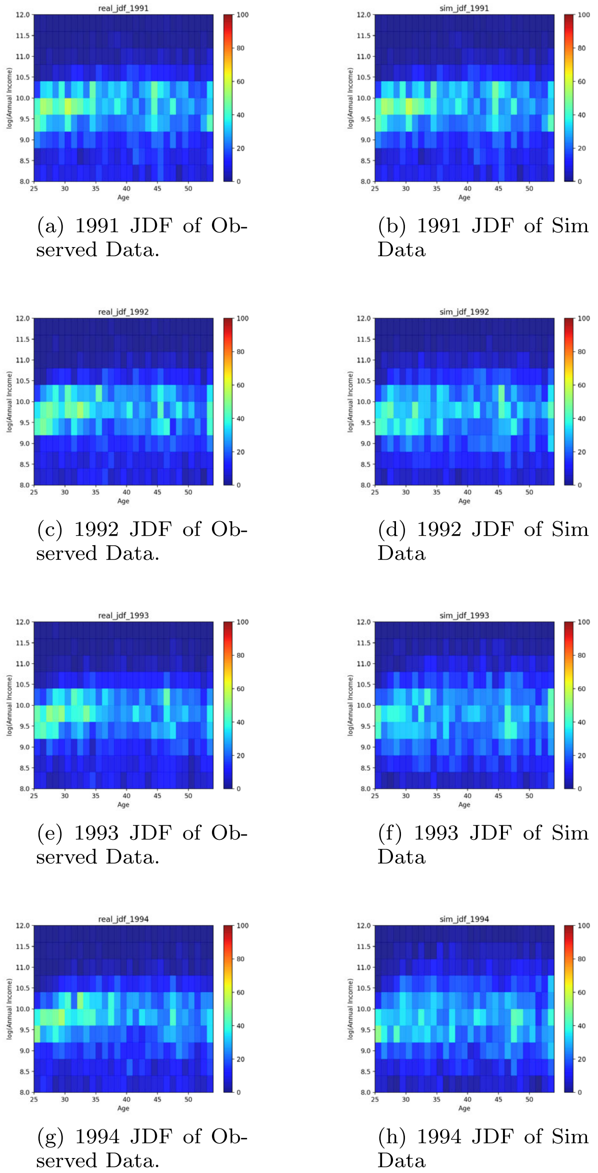


(a) Observed Statistics

(b) Simulation Statistics with GMM

Figure 12 USA Labour Data Observed and Simulation Statistics

A.5 BHPS—JDFs of age and income for observed and simulated data(LSM)



(a) 1991 JDF of Observed Data.

(b) 1991 JDF of Sim Data

(c) 1992 JDF of Observed Data.

(d) 1992 JDF of Sim Data

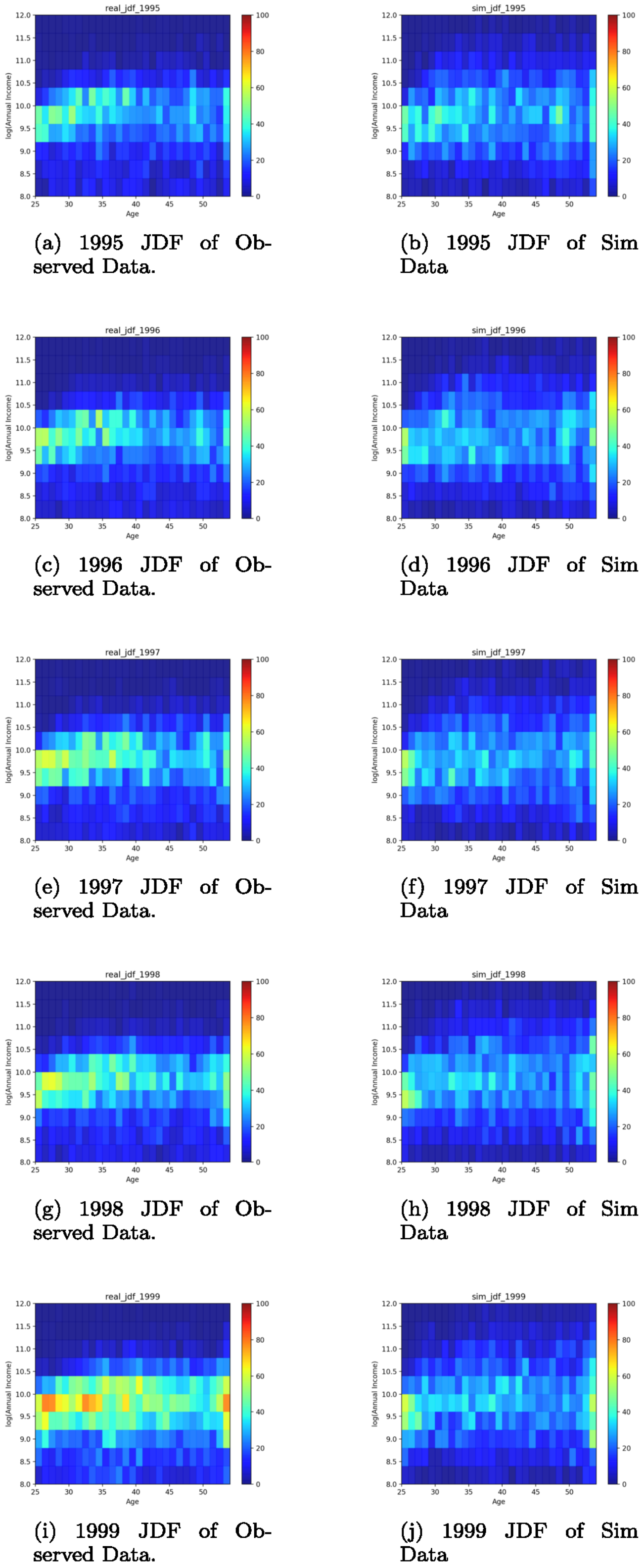
(e) 1993 JDF of Observed Data.

(f) 1993 JDF of Sim Data

(g) 1994 JDF of Observed Data.

(h) 1994 JDF of Sim Data

Figure 13 JDF for 1991–1994



(a) 1995 JDF of Observed Data.

(b) 1995 JDF of Sim Data

(c) 1996 JDF of Observed Data.

(d) 1996 JDF of Sim Data

(e) 1997 JDF of Observed Data.

(f) 1997 JDF of Sim Data

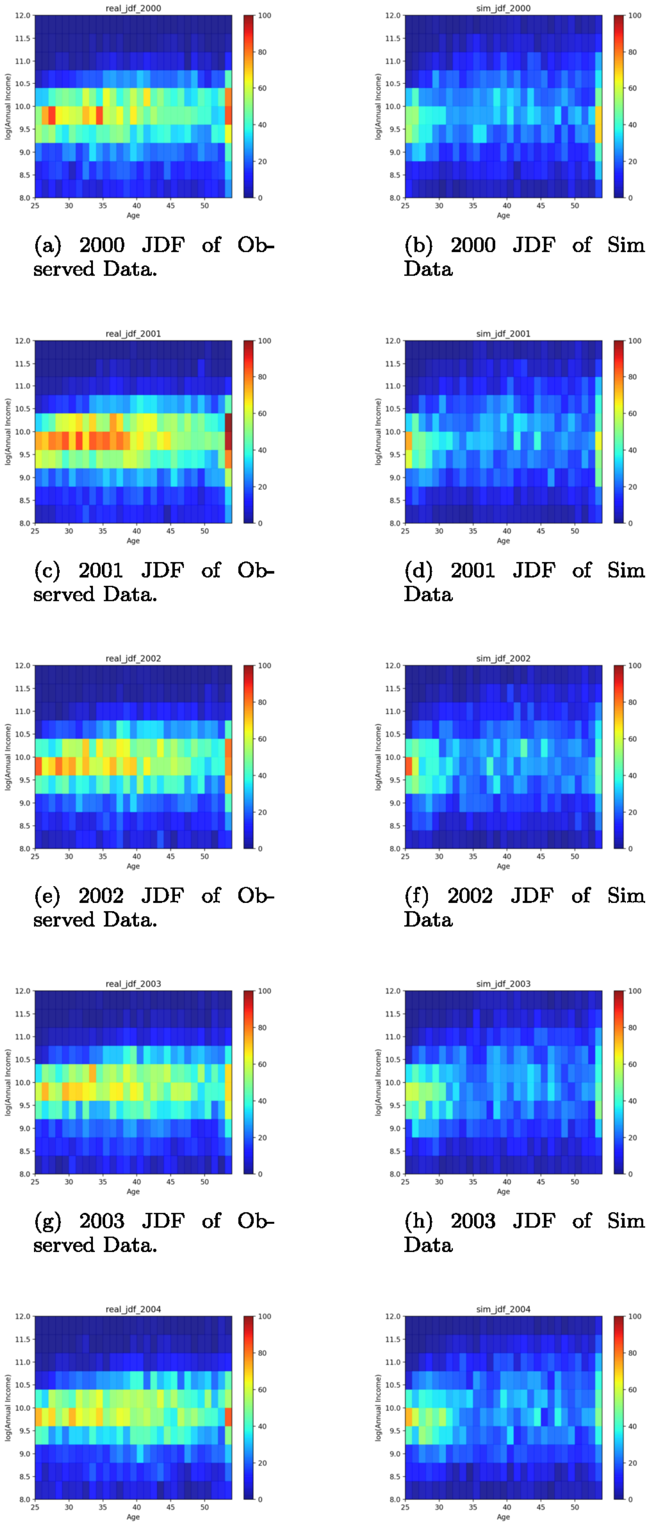
(g) 1998 JDF of Observed Data.

(h) 1998 JDF of Sim Data

(i) 1999 JDF of Observed Data.

(j) 1999 JDF of Sim Data

Figure 14 JDF for 1995–1999



(a) 2000 JDF of Observed Data.

(b) 2000 JDF of Sim Data

(c) 2001 JDF of Observed Data.

(d) 2001 JDF of Sim Data

(e) 2002 JDF of Observed Data.

(f) 2002 JDF of Sim Data

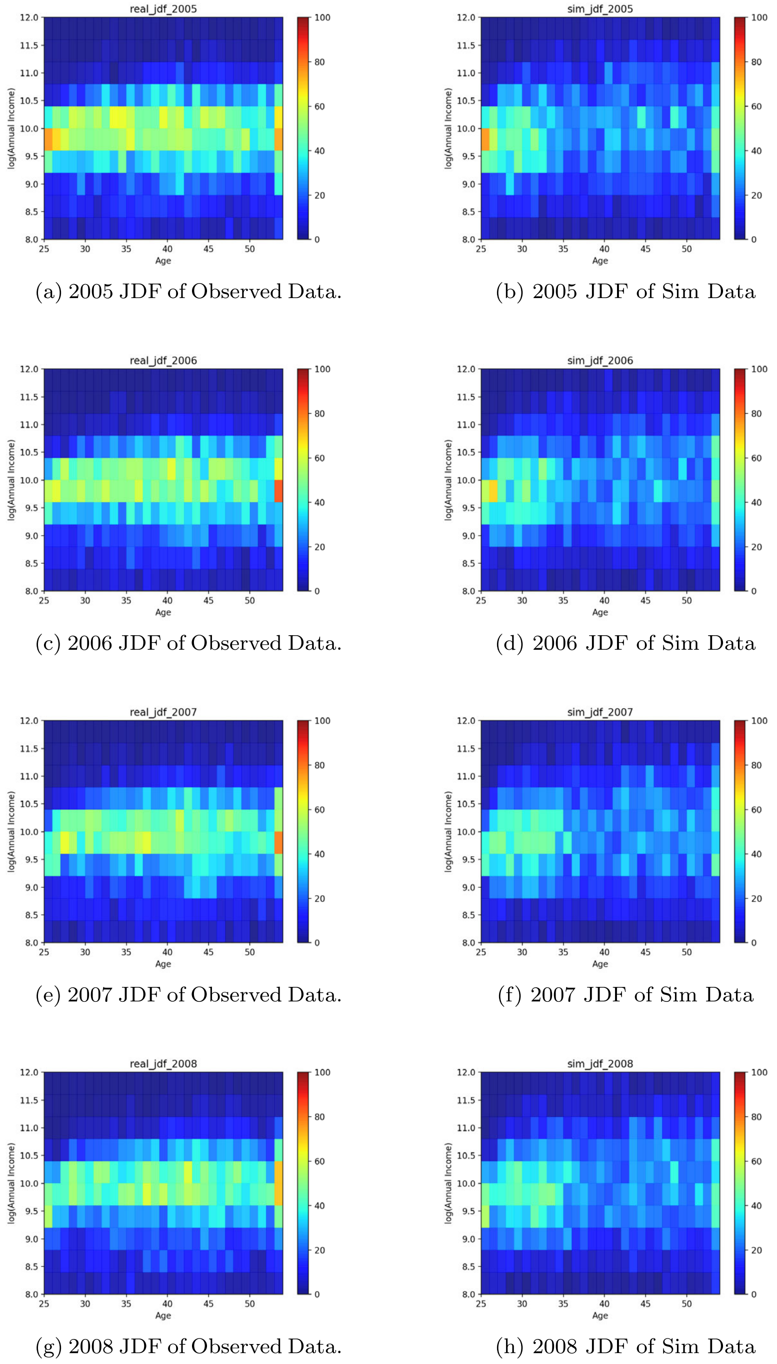
(g) 2003 JDF of Observed Data.

(h) 2003 JDF of Sim Data

(i) 2004 JDF of Observed Data.

(j) 2004 JDF of Sim Data

Figure 15 JDF for 2000–2004



(a) 2005 JDF of Observed Data.

(b) 2005 JDF of Sim Data

(c) 2006 JDF of Observed Data.

(d) 2006 JDF of Sim Data

(e) 2007 JDF of Observed Data.

(f) 2007 JDF of Sim Data

(g) 2008 JDF of Observed Data.

(h) 2008 JDF of Sim Data

Figure 16 JDF for 2005–2008

Acknowledgements
The authors acknowledge the valuable ideas and feedback for experiment design of J. Doyne Farmer during Early Stage of Research.

Funding
No funding is used for the presented research.

Availability of data and materials
The BHPS dataset analysed during the current study is available in the UK Data Service repository with restricted use terms, https://doi.org/10.5255/UKDA-SN-5151-2. The IPUMS CPS dataset analysed during the current study is available in the Minneapolis, MN: IPUMS repository with restricted use terms, https://doi.org/10.18128/D030.V7.0. The materials generated during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
FO contributed at all stages of the research, particularly developing the income dynamics model, running the experiments, running the analysis and wrote the paper. PB conceptualized the problem, developed income dynamics model, interpreted the results and wrote the paper. OK conceptualized the problem and interpreted the results. All authors read and approved the final manuscript.

Author details
[1]University College London, Gower Street, WC1E 6BT, London, UK. [2]The Old Schools, University of Cambridge, Trinity Ln, CB2 1TN, Cambridge, UK.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 27 July 2021 Accepted: 12 January 2022 Published online: 29 January 2022

References

1. Champernowne DG (1953) A model of income distribution. Econ J 63(250):318–351. https://doi.org/10.2307/2227127
2. Shorrocks AF (1976) Income mobility and the Markov assumption. Econ J 86(343):566–578. https://doi.org/10.2307/2230800
3. Shorrocks A (1978) Income inequality and income mobility. J Econ Theory 19(2):376–393. https://doi.org/10.1016/0022-0531(78)90101-1
4. Lillard LA, Willis RJ (1976) Dynamic aspects of earnings mobility. Report 0898-2937. National Bureau of Economic Research
5. Buchinsky M, Hunt J (1999) Wage mobility in the United States. Rev Econ Stat 81(3):351–368
6. Cappellari L, Jenkins SP (2004) Modelling low income transitions. J Appl Econom 19(5):593–610
7. Kopczuk W, Saez E, Song J (2010) Earnings inequality and mobility in the United States: evidence from social security data since 1937*. Q J Econ 125(1):91–128. https://doi.org/10.1162/qjec.2010.125.1.91
8. Firpo S, Fortin NM, Lemieux T (2009) Unconditional quantile regressions. Econometrica 77(3):953–973
9. Firpo S, Fortin NM, Lemieux T (2011) Occupational tasks and changes in the wage structure. IZA Discussion Papers
10. Guvenen F (2009) An empirical investigation of labor income processes. Rev Econ Dyn 12(1):58–79. https://doi.org/10.1016/j.red.2008.06.004
11. Guvenen F, Kuruscu B, Ozkan S (2014) Taxation of human capital and wage inequality: a cross-country analysis. Rev Econ Stud 81(2):818–850
12. Drăgulescu A, Yakovenko VM (2001) Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. Phys A, Stat Mech Appl 299(1–2):213–221. https://doi.org/10.1016/S0378-4371(01)00298-9
13. Ludwig D, Yakovenko VM (2021) hysics-inspired analysis of the two-class income distribution in the USA in 1983–2018. arXiv:2110.03140 [physics, q-fin, stat]. Accessed 2021-12-07
14. Tao Y (2021) Boltzmann-like income distribution in low and middle income classes: evidence from the United Kingdom. Phys A, Stat Mech Appl 578:126114. https://doi.org/10.1016/j.physa.2021.126114
15. University of Essex, Institute for Social and Economic Research (2018) BHPS British Household Panel Survey: Waves 1–18, 1991–2009. UK Data Service. https://doi.org/10.5255/UKDA-SN-5151-2
16. Flood S, King M, Rodgers R, Ruggles S, Warren JR (2020) Integrated Public Use Microdata Series, Current Population Survey: Version 7.0. Minneapolis, MN: IPUMS. https://www.ipums.org/projects/ipums-cps/d030.V7.0
17. Taylor MF, Brice J, Buck N, Prentice-Lane E (2018) British Household Panel Survey User Manual Volume A: Introduction. Technical Report and Appendices. Colchester: University of Essex
18. Office for National Statistics: Population Estimates for UK, England and Wales, Scotland and Northern Ireland. National Statistics (2018). https://data.gov.uk/dataset/849f9984-dfe4-46a5-8162-c5dee3f19ea4/population-estimates-for-uk-england-and-wales-scotland-and-northern-ireland
19. Evans J, Robinson H (2010) The Pensioners' Incomes Series 2008-09. Pensions, Department for Work and Pensions (2010). http://statistics.dwp.gov.uk/asd/index.php?page=pensioners_income_arc#PI_Prev
20. Holzmann R, Stiglitz JE (2001) New ideas about old age security: toward sustainable pension systems in the 21st century. World Bank