Savings and Retirement Decisions in the UK

James Sefton, Justin van de Ven, & Martin Weale
National Institute of Economic and Social Research

1 Introduction

This paper describes a dynamic microsimulation model that has been developed to consider household savings and retirement behaviour. Microsimulation models were first used for economic analysis by Orcutt (1957), and are now commonly employed to undertake policy analyses in many countries around the world. The feature that distinguishes microsimulation models from their macro based counterparts is that each micro-unit (also referred to as agent) from a given population is individually represented.[1] This property facilitates an understanding of the influences that contribute to simulated outcomes, and makes microsimulation particularly useful for undertaking distributional analyses.

Microsimulation models are traditionally classified as either dynamic or static, depending upon how (and whether) the population is aged. Static microsimulation models, as their name suggests, determine the impact of counterfactual conditions upon a population of agents at a point in time. They usually consist of two parts; a reference database that details the characteristics of each agent in a population, and a procedure for calculating the impact on each agent of counterfactual conditions. Consequently, the range

---

[1] For macro-based models that study the impact of policy changes, see Dervis *et al.*(1982), Taylor (1990), and De Janvry *et al.* (1991). These are examples of Computable General Equilibrium models. Most micro-based models are constructed using a partial equilibrium framework. For examples of micro-based models that use a general equilibrium framework, see Meagher (1993), and Congneau and Robilliard (2000).

of policies that can be analysed by static microsimulation models is limited only by the degree of detail that is provided by the reference database used. Given the demographic and income characteristics of families, for example, static microsimulation models are often used to determine the impact effects of alternative benefits policies on the income distribution, and upon the budgetary cost of the transfer system.

Static microsimulation models 'age' a population by reweighting the reference database using statistical projections to reflect an alternative time period. In contrast, dynamic microsimulation models age each individual described by the reference database in response to stochastic variation and an accumulated history. For example, a dynamic microsimulation model that is designed to consider the effects of fiscal policy may generate characteristics that include marital status, parenthood, income, and mortality at annual intervals for each person described by a reference database. The income of each individual at any given year is often simulated based on characteristics such as the individual's past income, their demographic characteristics, and upon a stochastic term that accounts for unexplained variation. This type of procedure builds up a life history for each individual in a population, which significantly increases the range of questions that can be explored, relative to static models. Most dynamic microsimulation models are designed specifically to consider the intertemporal and long term effects of counterfactual conditions, rather than the impact effects with which static models are concerned.

Most microsimulation models that are currently in use are static. Prominent examples of these include, STINMOD (Australia; refer to the STINMOD Technical Series, NATSEM, Australia), POLIMOD (UK; see Redmond et al., 1998), EUROMOD (15 member states of the European Union; see Sutherland, 2001), TRIM2 (US; see Giannarelli, 1992), SPSP (Canada; refer to Statistics Canada), SWITCH (Ireland; see Callan et al., 1996), LOTTE (Norway, see Fjærli et al., 1995), and FASIT (Sweden; refer to the Swedish

Ministry of Finance).[2] Advances in computing power, analytical techniques, and the availability of increasingly detailed survey data have led to an increase in both the number and sophistication of dynamic microsimulation models. Some recent examples of these include ASPEN (US; see Basu et al., 1998), CORSIM (US; see Caldwell, 1997), DYNACAN (Canada; refer to Statistics Canada, based on DYNASIM, see Orcutt et al., 1976), HARDING (Australia; see Harding, 1993), MICROHUS (Sweden; see Andersson et al., 1992), and SESIM (Sweden; refer to the Swedish Ministry of Finance).

In addition to the static-dynamic dichotomy, microsimulation models can also be distinguished by the extent to which they incorporate agent specific behavioural responses. Given the ageing populations and reduced rates of economic growth observed in many industrialised countries, attention has been focused in recent years on the responsiveness of labour supply, savings, and fertility to alternative tax and benefit systems.[3] Behavioural response may be modelled using statistical projections estimated from survey data (see, for example, CORSIM), or an explicit consideration of how individual decisions are made. The latter of these methods usually involves assuming that reference units make their decisions to maximise an assumed objective (utility) function, subject to various practical constraints (such as the available funds that a household can spend).

The model described by this paper falls into the last of the categories described above. Specifically, household decisions regarding labour and consumption are simulated by assuming that the household maximises an intertemporal utility function, subject to a budget constraint. In our view, this approach is of particular importance for an analysis of the relationship between government fiscal policy, and household savings and retirement behaviour. This is because savings and retirement decisions depend crucially

---

[2] For useful surveys, refer to Zaidi and Rake (2001), Sutherland (1995), and Merz (1991).

[3] See Macunovich (1998), and Hotz *et al.* (1997) for surveys of the fertility literature, Auerbach (1997) on savings, and Debelle and Swann (1998) on trends in the Australian labour market.

upon individual expectations regarding their future economic situation. Any attempt to consider the effects of alternative pension policies on savings and retirement must consequently take into consideration the adaptive expectations of the population - a purpose for which regression models are ill-suited due to the limitations imposed by the nature of survey data.[4]

Most microsimulation models that currently exist generate a large number of characteristics for each individual agent to make a broad range of analyses possible. In contrast, the current model restricts agent heterogeneity to six characteristics to facilitate transparent analysis of simulated outcomes. Given the relatively few characteristics generated by the current model, the ability to include additional characteristics as required is a fundamental feature of the modular structure adopted.

An overview of the simulation procedure is provided in Section 2. Separate sections are devoted to providing detailed descriptions for core elements of the microsimulation model. Section 3 discusses simulation of household human capital, and Section 4 outlines the method of simulating household wealth, savings and labour supply. Concluding comments are made in Section 5.

2 The Current Model

A partial equilibrium dynamic microsimulation model has been constructed to explore household savings and retirement decisions in the UK. The decision unit in the model is the household. Each household is aged by annual increments, from 20 to 90 based upon the age of the household's reference person.[5] In every year, the household decides whether to work full-time, part-time or not at all (households are treated as having an aggregate labour

---

[4] This criticism is analogous to the criticism that Lucas raised with reference to Macroeconomic models in 1976.

[5] See The *Family Expenditure Survey 2000-2001 User Guide*, Vol. 1 for the definition of a household reference person.

supply), and how much to consume given its economic situation, under the constraint that its net worth must remain positive. We assume a broad definition for the economic situation of a household that includes the household's age, its size, the wealth that it has managed to accumulate, the interest rate, the level of means-tested income support available, and the wage that it can command for its labour. This wage rate evolves stochastically.

At age 65 the household is forced to retire if it has not already chosen to do so. In retirement the household pays for its consumption either out of its savings or from a state pension. This pension is characterised by its generosity (the replacement rate) and the degree to which it is withdrawn for every pound of private income earned (the taper rate).

Simulated households are described by 6 characteristics:

1. the number and age of household members
2. the human capital of the household
3. the labour supply of the household
4. household consumption
5. household wealth
6. time of death

Demographic Size and Composition: The size of each household varies with time to reflect the coupling of individuals, and the birth and aging of children who eventually leave home. Household size is, however, modelled in a pre-determined fashion, and consequently behavioural effects are not considered in this dimension. For a model of endogenous fertility, see Nerlove *et al.* (1984).

Human Capital: A household's labour income is equal to their human capital multiplied by their labour supply. The human capital of a household is simulated as a stochastic process using a regression toward the mean model that adjusts for a learning-by-doing effect. See Section 3 for a detailed description the model used to simulate human capital.

Labour Force Status, Consumption, and Wealth: Household decisions regarding labour supply, consumption and saving are endogenous to the model. As this is a fundamentally important aspect of the simulation model, a detailed description of the methods involved is provided in Section 4.

Household Mortality: Each household is selected to die, based upon an exogenously defined survival function that is calibrated from cross-sectional statistical mortality rates for 2000/01 in the UK.[6] The use of data from a single cross-section implies that temporal trends in mortality rates are not captured by the model. In this sense, the model generates a cohort of individuals who are born in 2000/01 and live for up to 90 years in a world that remains exactly as it was in their birth year. On this approach to microsimulation model calibration, see Harding (1993). Figure 1 displays the mortality rates exogenously imposed by the simulation model as a function of household reference person age.

3 Human Capital

The household is considered as a single unit when simulating human capital, labour supply, and consumption. Since a detailed description of the determinants that underlie household income lies outside the research agenda for which the microsimulation model has been constructed, a simple "regression-

---

[6] The mortality rates used are calculated using the proportion of female reference people by age recorded in the 2000/01 Family Expenditure Survey (FES), and mortality rates by age and sex recorded in the *Annual Abstract of Statistics*, Table 5.21, The Stationary Office. Mortality rates after the age of 84 are subject to manual adjustment.

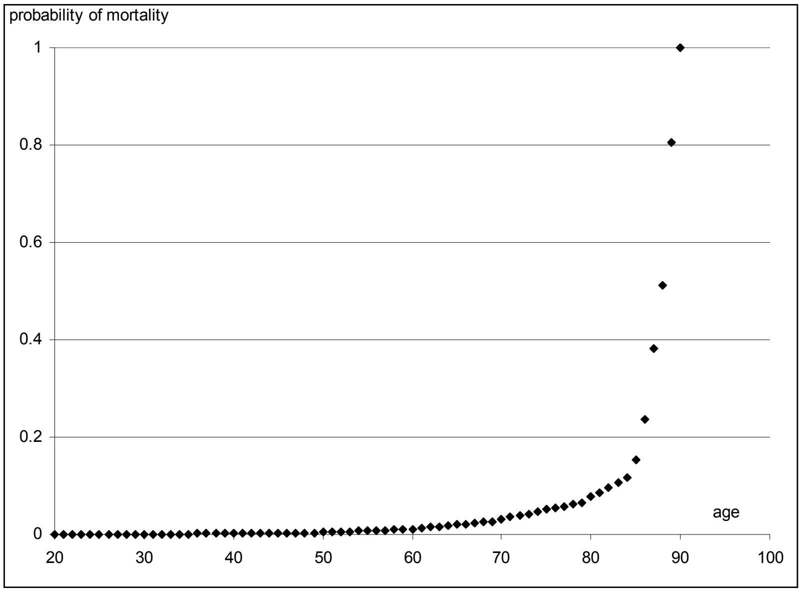


Figure 1: Mortality Probabilities by Age

toward-the-mean" model of human capital evolution has been adopted.[7]

Let h_{it} define the human capital of household i at age t. Defining m_t to be the geometric mean of all h_{it}, then, for a population of n individuals,

$$m_t = \sqrt[n]{\prod_{i=1}^{n} h_{it}} \tag{1}$$

The central assumption of the model is that the proportional variation of any household i's human capital from one year to the next deviates from the proportional variation of the respective mean by a random variable with a mean of zero. That is,

$$\frac{\dot{h}_{it}}{h_{it}} = \frac{\dot{m}_t}{m_t} + u_{it} \tag{2}$$

Defining $z_{it} = \log\left(\frac{h_{it}}{m_t}\right)$ and substituting into equation (2) obtains,

$$\dot{z}_{it} = u_{it} \tag{3}$$

---

[7] See van de Ven (1998) on the relationship between the regression toward the mean model and the model of income dynamics advocated by Mincer (1974).

Discretising equation (3) arrives at the following first order auto-regressive equation:

$$z_{it} = z_{i(t-1)} + u_{it} \tag{4}$$

Following Kalecki (1949), regression of human capital toward the mean implies that if $h_{it} > m_t$, then on average, $\frac{\dot{h}_{it}}{h_{it}} < \frac{\dot{m}_t}{m_t}$ and vice versa. When $h_{it} \gtrless m_t$ and $\beta < 1$, then $(1-\beta)\log\left(\frac{h_{it}}{m_t}\right) \gtrless 0$. Regression toward the mean is allowed for in the model by subtracting $(1-\beta)\, z_{i(t-1)}$ from the right-hand side of equation (4) to obtain,

$$z_{it} = \beta z_{i(t-1)} + u_{it} \tag{5}$$

The value of β consequently determines the variation of individual incomes relative to the geometric mean. When $\beta < 1$, regression toward the mean arises as described above. Regression away from the mean is characterised by $\beta > 1$, and when $\beta = 1$, the Gibrat process obtains.[8]

In the simulation model, equation (5) is extended to include a learning-by-doing effect, such that:

$$z_{it} = \beta z_{i(t-1)} + \theta\left(\bar{L}_{(t-1)} - L_{i(t-1)}\right) + u_{it} \tag{6}$$

where $L_t \in [0,1]$ is the proportion of time taken as leisure by household i in period t, and $\bar{L}_t$ is some threshold rate of leisure (taken to be the leisure enjoyed when a household works part-time). When a household takes less leisure than $\bar{L}_t$ (and hence supplies more labour), equation (6) implies that its income in subsequent periods will tend to increase relative to the population mean.

The model used to simulate human capital consequently depends on three elements; the disparity of the human capital distribution of 20 year-olds (σ_0); the relationship between mean human capital, m_t, and age; and the dynamic human capital variation characterised by equation (6).

---

[8] Equation (4) was first applied to income data by Gibrat (1931), and so in the above context is referred to as a Gibrat process, though more generally it is known as a Markov process. See Creedy (1985) for further details.

3.1 Calibration

Data

Data for calibrating the model are taken from the 2000/01 Family Expenditure Survey (FES), and Waves 6 and 10 of the British Household Panel Survey (BHPS). The 2000/01 FES provides income, expenditure, and demographic data for a nationally representative sample of 6115 households in Great Britain. Refer to the *Family Expenditure Survey 2000-2001 User's Guide* (published by the Office for National Statistics) for detailed information regarding the FES. The BHPS is a panel survey that provides annual data for a nationally representative sample of households in Great Britain. The tenth wave was undertaken in 1999/2000 and is the most recent wave to be made available to the public. See the *British Household Panel Survey User Manual* (published by the University of Essex) for detailed information regarding the BHPS.

Dispersion of human capital for 20 year-olds

The standard deviation of human capital for the base year of the simulated population is calibrated using survey data of household reference people taken from the FES. Given the small population of 20 year old household reference people recorded by the FES, the dispersion of human capital for the base year of the simulated population was inferred by taking a linear trend with age from older households. Specifically, the FES population was first divided into two subgroups, depending upon whether the household reference person continued full-time education beyond the age of 17. From each of the consequent subgroups, reference people who normally worked in excess of 30 hours per week were identified.[9] The log wages of these individuals were used to calculate standard deviations for five year age groups, which are displayed in Figure 2.

---

[9] See the following subsection for detailed descriptions of associated FES codes used.

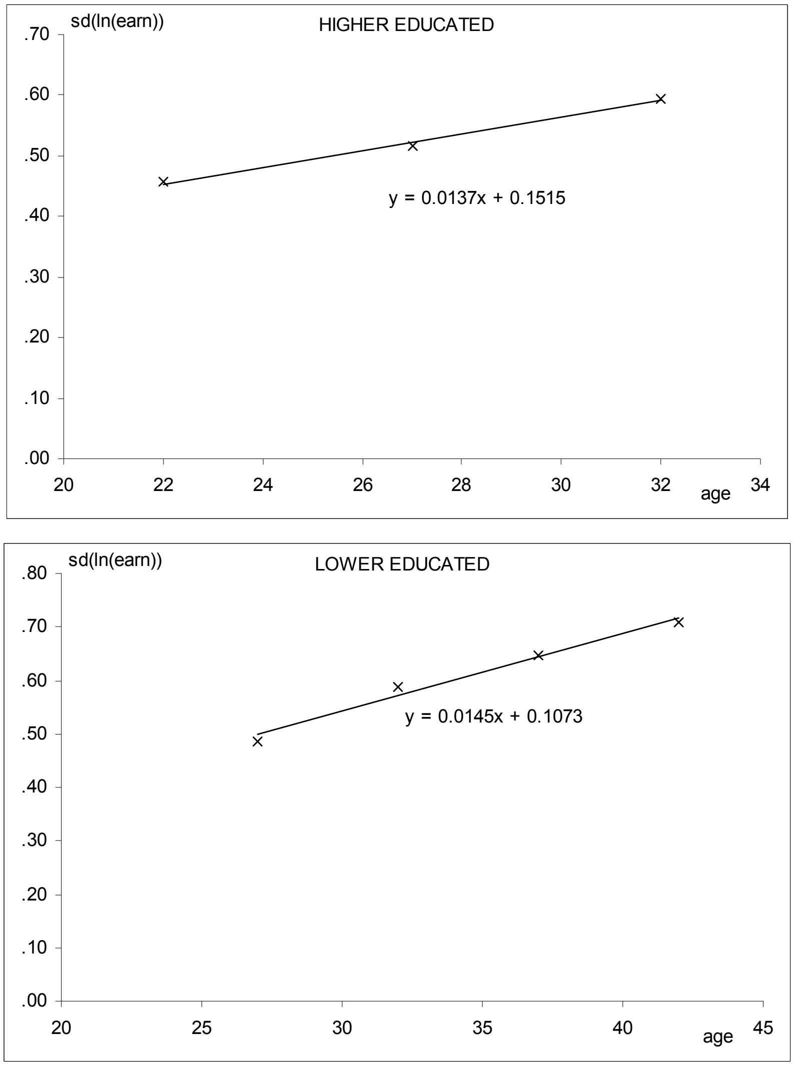


Figure 2: Standard Deviations of Income for Household Reference Individuals

The linear trends displayed in Figure 2 are used to infer the standard deviations from which the distributions of base year human capital are generated. These are:

$$\begin{aligned} 0.0137 * 20 + 0.1515 &= 0.3036 \text{ for higher educated} \\ 0.0145 * 20 + 0.1073 &= 0.3128 \text{ for lower educated} \end{aligned}$$

Average Human Capital

The profiles of average human capital are calibrated using a sample selection regression model of individual full-time employment wages. This model takes into consideration the fact that wages are only observed for individuals who are working, and that there is likely to be a relationship between the probability of an individual working and their wage rate. The regressions were undertaken using the "Sampsel" procedure in TSP; for full details refer to the "TSP 4.4 User's Guide" (see http://elsa.berkeley.edu/wp/tsp_user/tspugpdf.htm).

The sample selection model involves estimating two equations, a probit to identify individuals who are employed, and a (log) wage equation:

$$\begin{aligned} emp_i &= \pi_0 + \pi_1 age_i + \pi_2 age_i^2 + \pi_3 mar_i + \pi_4 chu5_i + \\ &\quad \pi_5 ch5p_i + \pi_6 na_i + \pi_7 totx_i + \varepsilon_{1i} \qquad (7) \\ \ln(w_i) &= \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 age_i^3 + \varepsilon_{2i} \qquad (8) \end{aligned}$$

where:

emp_i (a220+a221>30)	identifies when individual i works in excess of 30 hours per week
age_i (a005)	age of individual i
mar_i (a006= 1 or 2 or 3)	identifies when individual i is married
$chu5_i$ (a040+a041)	number of children under 5 years old
$ch5p_i$ (a042)	number of children 5 years or older
na_i (a049-chu5-ch5p)	the number of adults in i's household
$totx_i$ (p550tp)	total expenditure of i's household
w_i (p008+p011+p037+p047)	normal weekly wage of individual i

Bracketed terms identify FES variable codes

Equations (7) and (8) were estimated separately for higher and lower educated household reference individuals between the ages of 18 and 70 years taken from the 2000/01 FES (population sizes consist of 1071 and 4470 individuals respectively). Regression statistics are displayed in Table 1.

Table 1 indicates that all of the estimated coefficients are highly significant. Two variables were omitted from the regressions due to low statistical significance, one from each of the lower and higher educated subpopulations. The implied profiles of human capital with age are displayed in Figure 3

In addition to the profiles of human capital obtained from the sample selection model, Figure 3 also displays profiles estimated by standard Ordinary Least Squares (OLS). The profiles displayed in the figure indicate that the adjustment for sample selection has a small effect on the estimates obtained at low ages relative to high ages, which is consistent with the early retirement of individuals who had previously earned relatively high incomes. The effect is, however, surprisingly large for the lower educated population,

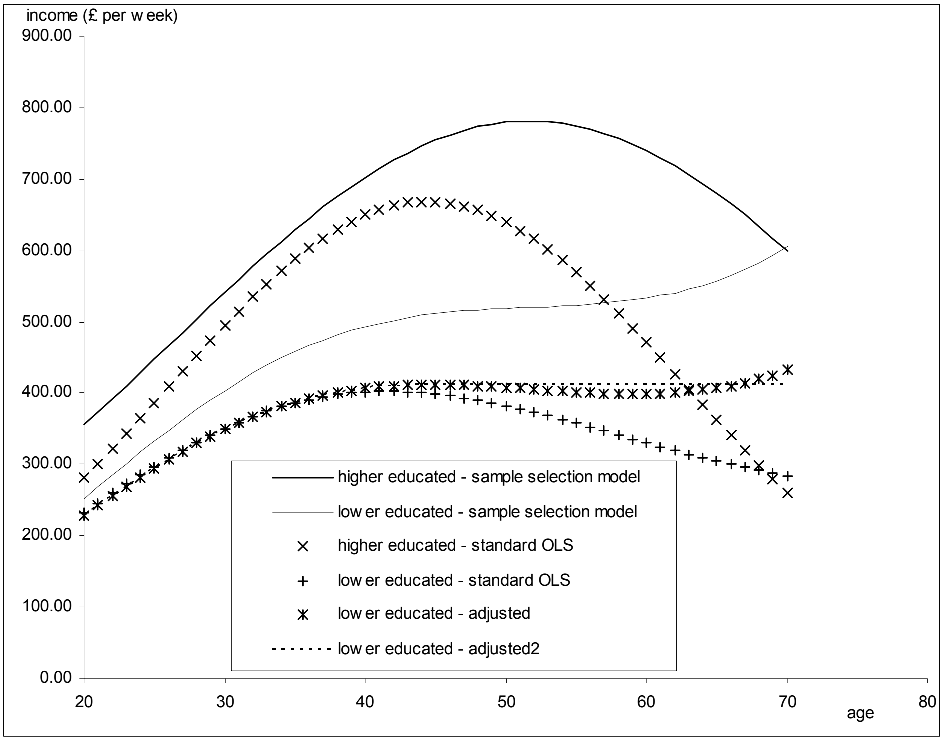
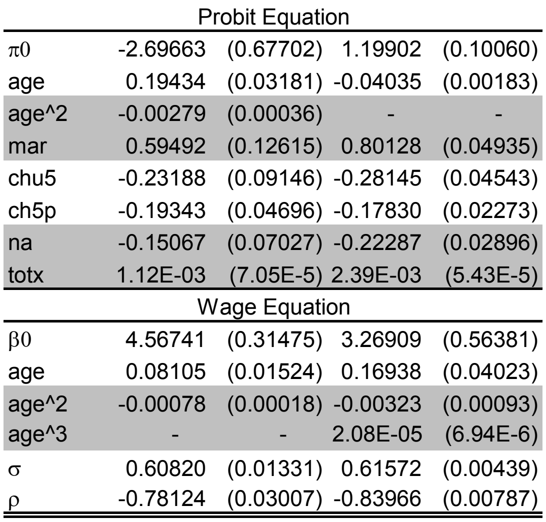


Figure 3: Profiles of Average Human Capital by Age and Education Status

Table 1: Sample Selection Model of Average Human Capital



	Higher Educated		Lower Educated	
	Probit Equation			
π0	-2.69663	(0.67702)	1.19902	(0.10060)
age	0.19434	(0.03181)	-0.04035	(0.00183)
age^2	-0.00279	(0.00036)	-	-
mar	0.59492	(0.12615)	0.80128	(0.04935)
chu5	-0.23188	(0.09146)	-0.28145	(0.04543)
ch5p	-0.19343	(0.04696)	-0.17830	(0.02273)
na	-0.15067	(0.07027)	-0.22287	(0.02896)
totx	1.12E-03	(7.05E-5)	2.39E-03	(5.43E-5)
	Wage Equation			
β0	4.56741	(0.31475)	3.26909	(0.56381)
age	0.08105	(0.01524)	0.16938	(0.04023)
age^2	-0.00078	(0.00018)	-0.00323	(0.00093)
age^3	-	-	2.08E-05	(6.94E-6)
σ	0.60820	(0.01331)	0.61572	(0.00439)
ρ	-0.78124	(0.03007)	-0.83966	(0.00787)

Standard Errors provided in parentheses

σ std deviation of the wage equation

ρ correlation between the errors of the probit and wage equations

a result that is attributable to the diminishing power of education as an identifier for the high human capital subgroup as age increases. Reducing the estimate obtained for β_1 by 15 per cent of its standard error alters the profile of human capital for the lower educated subpopulation to the series denoted "lower educated - adjusted" in Figure 3. The stylised form of this profile (which takes a constant value after the age of 43), defined as "lower educated - adjusted2" in Figure 3, is used in the microsimulation model.

Regression toward the mean (β) and temporal innovation (σ_ε)

The regression toward the mean parameter defined in equation (6) is calibrated using waves 6 and 10 of the BHPS data. A sample selection model was used, defined by the equations:

$$
\begin{aligned}
emp2_{it} &= \pi_0 + \pi_1 age_i + \pi_2 age_{it}^2 + \pi_3 mar_{it} + \pi_4 chu5_{it} + \\
&\quad \pi_5 ch5p_{it} + \pi_6 na_{it} + \pi_7 hhw_{it} + \varepsilon_{1it} \qquad (9) \\
z_{it} &= \beta_0 + \beta_1 z_{i(t-4)} + \varepsilon_{2it} \qquad (10)
\end{aligned}
$$

Table 2: Estimates of Regression Toward the Mean

	Higher Educated		Lower Educated	
	Probit Equation			
π0	-4.39528	(0.48317)	-3.76704	(0.49207)
age	0.23030	(0.02402)	0.19897	(0.02399)
age^2	-0.00287	(0.00028)	-0.00261	(0.00028)
mar	0.39456	(0.09059)	0.42233	(0.09438)
chu5	-0.05320*	(0.06459)	-0.39380	(0.07570)
ch5p	-0.22262	(0.03719)	-0.33964	(0.03748)
na	-0.01409*	(0.05726)	-0.00082*	(0.05247)
hhw	-7.49E-07	(2.86E-7)	-1.90E-07*	(4.04E-7)
	Wage Equation			
β0	-0.03306*	(0.03527)	-0.06658*	(0.04868)
z(t-4)	0.73405	(0.02208)	0.73474	(0.02637)
σ	0.30363	(0.00955)	0.31284	(0.00677)
ρ	0.34601	(0.11710)	0.16550*	(0.15277)

Standard Errors provided in parentheses
σ std deviation of the wage equation
ρ correlation between the errors of the probit and wage equations
* coefficients *not* significant at the 95% confidence interval

where $emp2_{it}$ is a dummy variable that identifies households in which z_{it} and $z_{i(t-1)}$ are greater than zero, hhw_{it} defines the wealth (including all financial and housing assets net of associated liabilities) of household i in wave 10 of the BHPS and all other variables take their previously stated definitions. The variable $z_{it} = \ln(y_{it}/m_t)$, where y_{it} defines the usual monthly income of the household reference person if the household reference person worked in excess of 30 hours per week, and m_t is the associated average income by age.[10] Regression statistics for the model defined by equations (9) and (10) are displayed in Table 2.

The coefficients obtained for the probit equation are similar to those obtained for the regression of mean human capital referred to in the previous subsection. The coefficients obtained for the wage equation indicate that the intercept is insignificant (consistent with the specification of equation (6)),

---

[10] The authors may be contacted for specifics regarding the coding, including the spss and tsp programmes used.

and that the regression toward the mean takes a value of 0.92562 for the higher educated subgroup, and 0.92583 for the lower educated ($\beta_1^{0.25}$). The estimates obtained for the standard deviation of the wage equation, σ in the table, are used to simulate temporal innovation of human capital evolution. The use of the fourth lagged period, implies that the standard deviation of temporal innovations is obtained by:

$$\sigma_\varepsilon = \frac{\sigma}{\sqrt{\beta^6 + \beta^4 + \beta^2 + 1}} \tag{11}$$

As in the case of the regression toward the mean coefficients, the estimates obtained for σ suggest that there is no significant difference between the temporal innovation of human capital for the lower and higher educated subpopulations. The fact that the estimate obtained for ρ for the lower educated population is not significant indicates that correcting for sample selection has little effect in this case.

4 Labour Force Status, Consumption, and Wealth

Household labour supply and consumption are generated by maximising expected utility subject to a budget constraint. Expected lifetime utility is described by the additvely separable function:

$$U = E\left(\sum_{i=t}^{T} u\left(\frac{c_t}{a_t}, L_t\right)\delta^{i-t}\right) \tag{12}$$

where $c_t \in R^+$ is household consumption, $a_t \in R^+$ is the household's equivalence scale, and $L_t \in [0,1]$ is household leisure at time t.[11] In the current specification of the model, labour choice is restricted to full-time employment, part-time employment, and not employed, such that $L_t \in (0.7, 0.85, 1)$. The equivalence scale is generated from household size using the McClements

---

[11] See, for example, Balcer and Sadka (1986), and Muellbauer and van de Ven (2003) on the use of this form of adjustment for household size in the utility function.

after housing costs specification.[12] The parameter δ is the discount factor (which is assumed to be constant) and T is the uncertain time of death. A Constant Elasticity of Substitution (CES) utility function is assumed, which is defined by:

$$u\left(C_t, L_t\right) = \frac{1}{\left(1 - \frac{1}{\gamma}\right)} \left(C_t^{\left(1-\frac{1}{\rho}\right)} + \alpha^{\frac{1}{\rho}} L_t^{\left(1-\frac{1}{\rho}\right)}\right)^{\frac{1-1/\gamma}{1-1/\rho}} \tag{13}$$

where γ is the inter-temporal elasticity of substitution and ρ is the elasticity of substitution between $C_t = c_t/a_t$ and L_t. The higher the value of ρ, the higher the proportional change between consumption and leisure for a given proportional change in prices. Similarly, the larger the value of γ, the higher the proportional substitution between consumption today and consumption tomorrow for a given change in interest rates. Wealth in any period, w_t, is constrained to be non-negative, and is given by:

$$w_{t+1} = w_t - c_t + y(w_t, x_t, t) \tag{14}$$

where $y\left(w_t, x_t, t\right)$ is the post-tax and benefit income obtained by a household of age t given wealth w_t, and labour income x_t. Post-tax and benefit income depends upon the exogenously imposed transfer policy (which is initially specified to reflect the existing tax and benefits system in the UK) and the real interest rate, R_t. Furthermore, labour income depends upon human capital, h_t, such that $x_t = h_t\left(1 - L_t\right)$.

Given equations (6) to (14), and values of w_t and h_t, it is possible to solve for the optimising consumption and labour decisions at any age t.[13] This allows the values of w_{t+1} and h_{t+1} to be inferred (from equations (6) and (14) respectively). Hence the optimising problem for consumption and

---

[12] For further details, see McClenents (1977) and *Households Below Average Income, A Statistical Analysis*, published by The Stationary Office.

[13] The solution method involves solving the implied set of Euler equations via backward induction, given that consumption and leisure in the final period, T, are known for any combination of h and w. The authors should be contacted for further details and associated software.

labour in period $t + 1$ can be solved. A complete life-history for individual household consumption and labour supply is generated by repeating this procedure until the (latest) death of the household in period T.

4.1 Calibration

A number of model parameters were calibrated to minimise three test statistics which relate the simulated output to labour force and consumption profiles derived from FES survey data. The model variables concerned are:

- the real rate of return, R
- the depreciation rate, δ
- the elasticity of substitution between consumption and leisure, ρ
- the intertemporal elasticity of substitution, γ
- the weight on leisure, α
- the learning by doing parameter, θ
- the wealth allocated to 20 year old households, w_0
- the proportion of the full-time wage earned by part-time workers, q.

Two of the three test statistics used to calibrate the model compare labour force trends, and the third statistic compares the simulated consumption profiles to the profiles implied by the regression statistics displayed in Table 3. We begin by describing the labour based test statistics.

One of the labour based statistics focuses solely upon profiles of full-time labour participation, while the other considers full-time, part-time and non- participation. Define $l_{ft,t}$ as the proportion of households at age t full-time employed, $l_{pt,t}$ as the proportion part-time employed, and $l_{ne,t}$ as the proportion not employed, where the simulated statistics, $\tilde{l}$, are compared with the associated statistics derived from survey data, l. Then the two labour based test statistics are defined, respectively, by:

$$
\begin{aligned}
l\_test1 &= \sum_{t=t_0}^{64} \left| l_{ft,t} - \tilde{l}_{ft,t} \right| &(15)\\
l\_test2 &= \sum_{t=t_0}^{64} \left| l_{ft,t} - \tilde{l}_{ft,t} \right| + \left| l_{pt,t} - \tilde{l}_{pt,t} \right| + \left| l_{ne,t} - \tilde{l}_{ne,t} \right| &(16)
\end{aligned}
$$

where $t_0 = 20$ for the lower educated, and 25 for the higher educated.[14]

The FES data referred to in the previous section were used to obtain the l statistics referred to in equations (15) and (16). Specifically, the average number of household reference people working full-time (in excess of 30 hours per week), part-time (between 1 and 30 hours per week) and not at all were calculated by age and education status (greater than 18 years full-time education identified as 'higher educated') from FES survey data. Figures 4, 5, and 6 display the proportions of the respective populations identified as full-time, part-time, and not employed respectively. The 'smoothed' series displayed in each of the figures are derived from the Hodrick-Prescott filter, and are used to calculate the labour based test statistics defined by equations (15) and (16). Also displayed in Figures 4 to 6 are the profiles of labour obtained using the parameter values adopted for the simulation model.

A statistical model of household consumption was required for the consumption based test statistic used to calibrate the model. Following analysis of alternative specifications, the FES data were used to estimate (via Ordinary Least Squares) the following regression of household consumption:

$$
\begin{aligned}
x_i = {} & \alpha_0 + \alpha_1 na_i + \alpha_2 nc_i + \alpha_3 nc_i^2 + \alpha_4 age_i + \alpha_5 age_i^2 \\
& + \alpha_{11} edn_i na_i + \alpha_{12} edn_i age_i + \alpha_{13} edn_i age_i^2 \\
& + \alpha_{14} ret_i + \alpha_{15} ret_i na_i + \alpha_{16} ret_i age_i + \varepsilon_i \qquad (17)
\end{aligned}
$$

---

[14] Absolute differences have been used to limit the influence of extreme values.

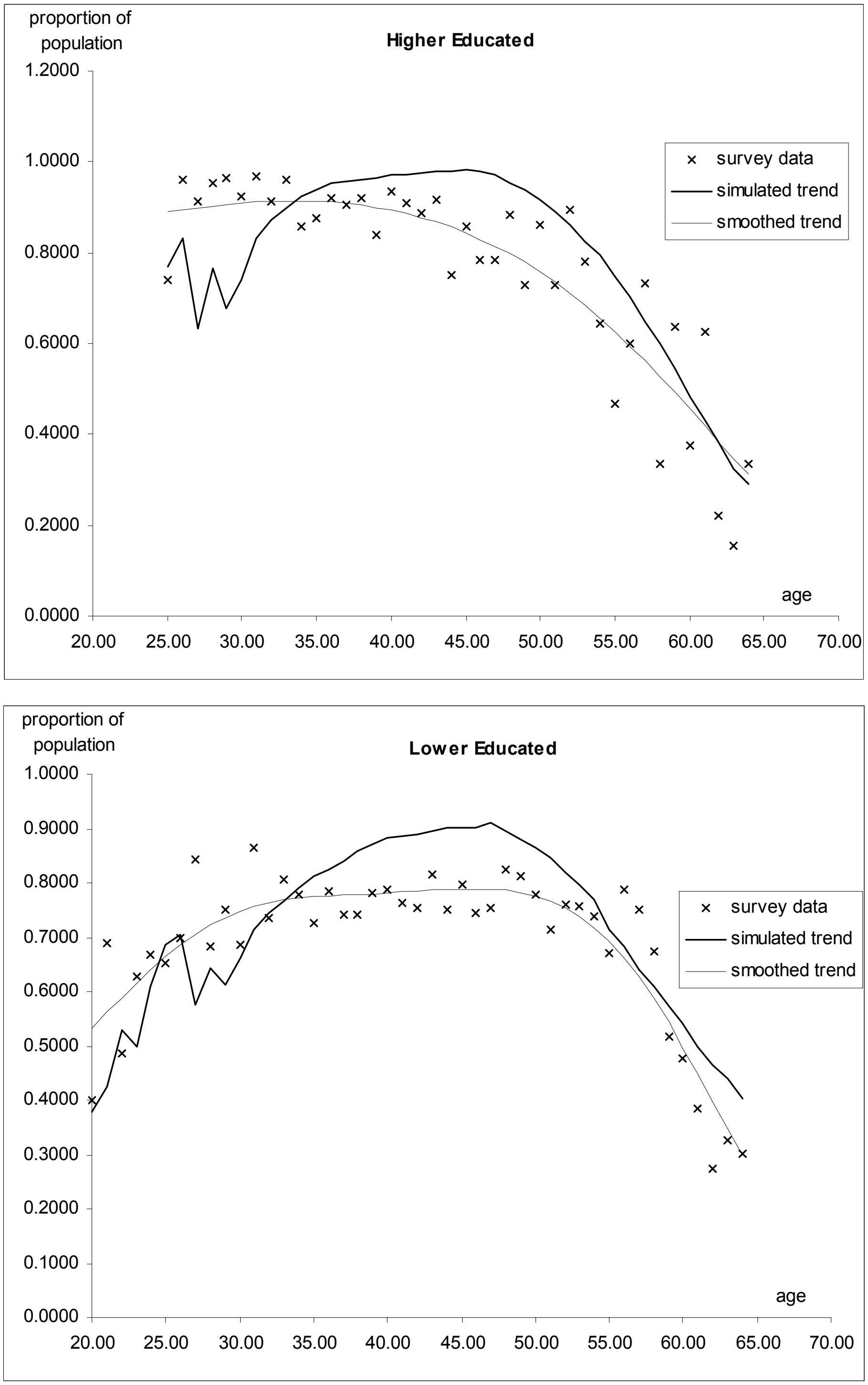
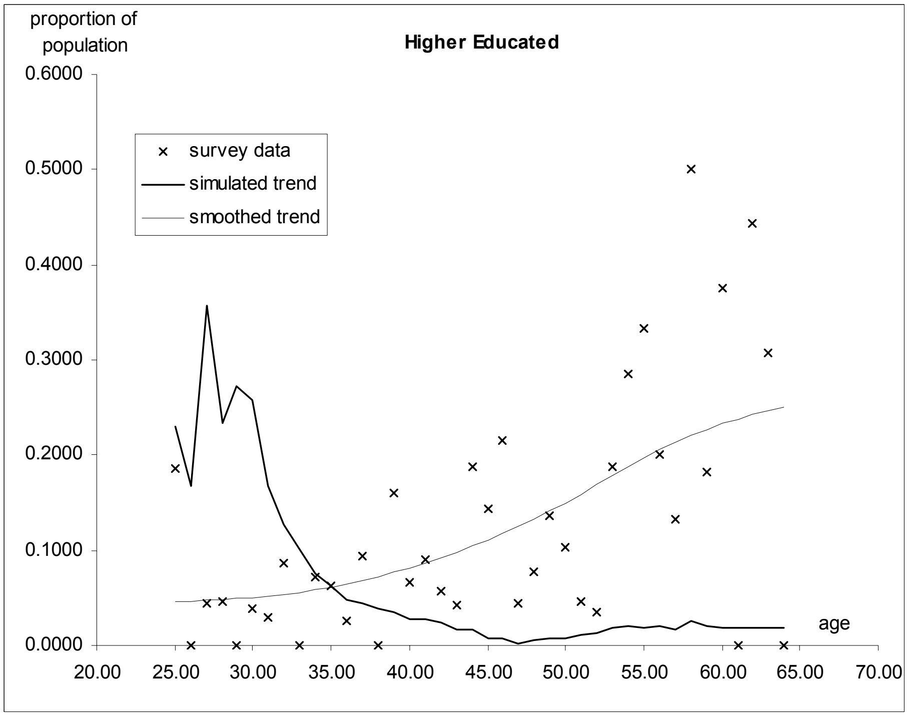


Figure 4: Average Full-time Employment by Age



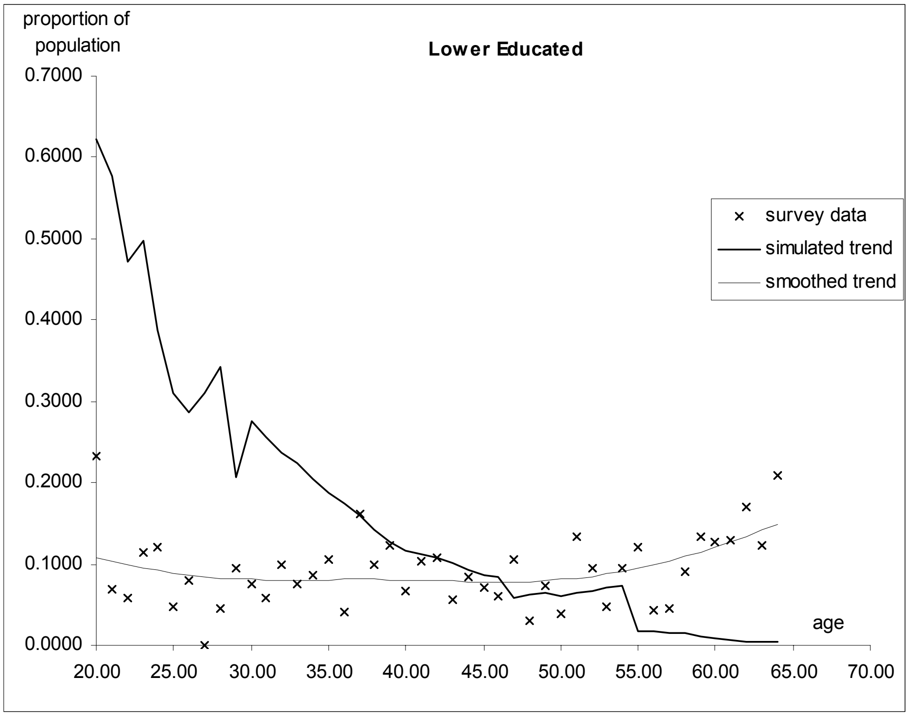


Figure 5: Average Part-time Employment by Age

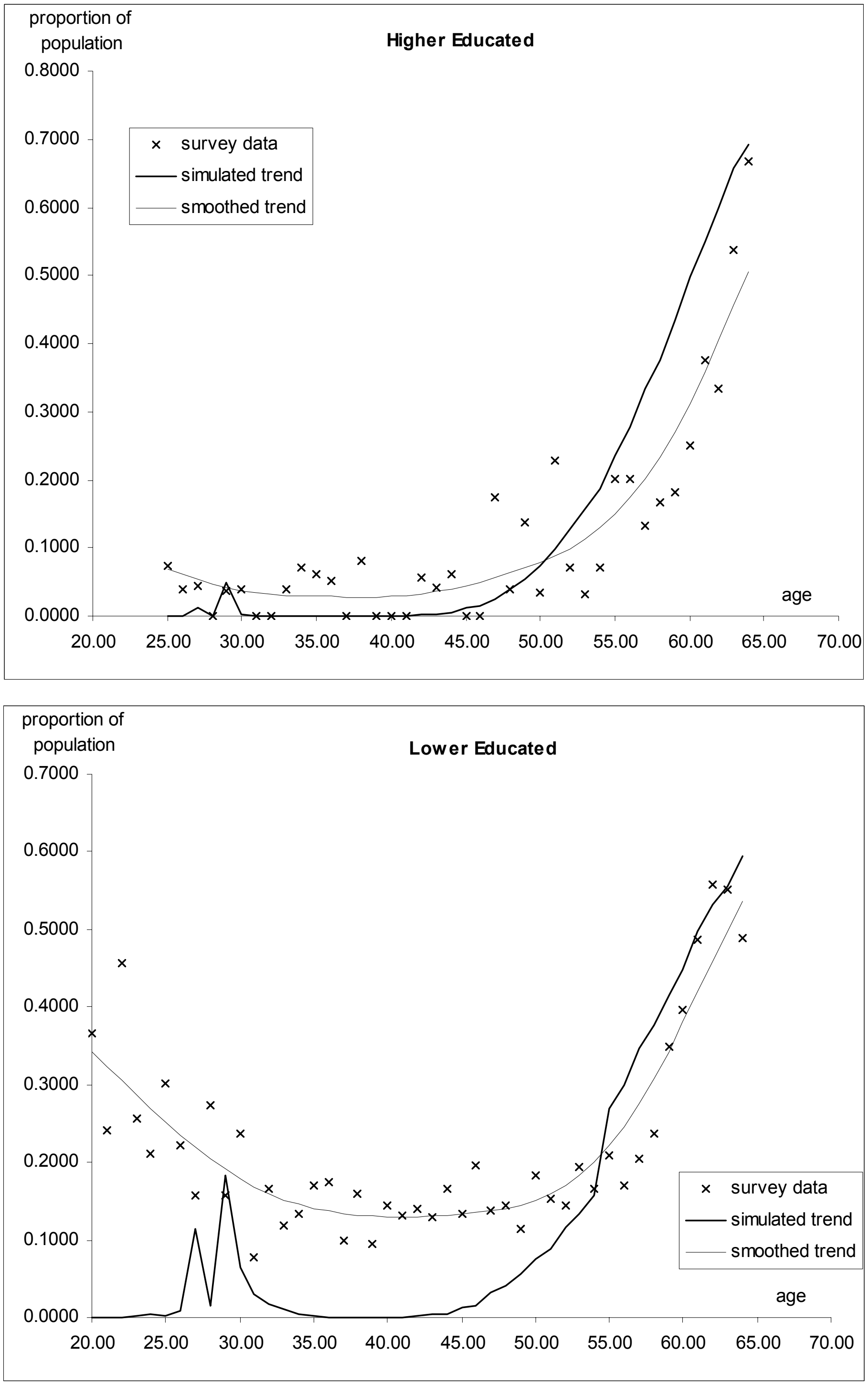


Figure 6: Average Not Employed by Age

Table 3: Total Household Expenditure

α0	-201.5520	(33.207)
na	155.0280	(5.342)
nc	53.6695	(9.342)
nc^2	-10.0958	(2.703)
age	13.4661	(1.529)
age^2	-0.1490	(1.69E-2)
edn*na	27.4911	(16.486)
edn*age	6.0746	(1.277)
edn*age^2	-0.0671	(1.53E-2)
ret	-331.4540	(79.544)
ret*na	-42.1201	(9.145)
ret*age	5.2702	(1.224)
R Squared	0.331294	
Eqn Std Error	237.423	

Standard Errors in parentheses

where x_i denotes total household expenditure (FES code p550tp), nc_i denotes the number of children in the household, and ret_i is a dummy variable that identifies household reference people over the age of 59 who are not employed. Regression statistics are displayed in Table 3.

The consumption based test statistic is the sum of absolute differences between the simulated consumption profiles, and the profiles implied by the regression statistics displayed in Table 3. Specifically, simulated household characteristics are used to calculate an estimate, $\hat{x}_{it}$ of consumption for each household i at each age t, based on equation 17 and the parameter values displayed in Table 3:

$$\begin{aligned}\hat{x}_{it} &= \hat{\alpha}_0 + \hat{\alpha}_1 na_{it} + \hat{\alpha}_2 nc_{it} + \hat{\alpha}_3 nc_{it}^2 + \hat{\alpha}_4 t_{it} + \hat{\alpha}_5 t_i^2 \\ &\quad + \hat{\alpha}_{11} edn_{it} na_{it} + \hat{\alpha}_{12} edn_{it} t_i + \hat{\alpha}_{13} edn_{it} t_i^2 \\ &\quad + \hat{\alpha}_{14} ret_{it} + \hat{\alpha}_{15} ret_{it} na_{it} + \hat{\alpha}_{16} ret_{it} t_i \qquad (18)\end{aligned}$$

The estimates of consumption are then aggregated over the entire life of each household, and the percentage at each age calculated:

$$p(\hat{x}_{it}) = \frac{\hat{x}_{it}}{\sum\limits_{j=20}^{90} \hat{x}_{ij}} \qquad (19)$$

Table 4: Simulation Model Calibration Statistics

	α	δ	θ	w0	R	q	l test 1	l test 2	c test
Base Case	0.034	0.98	0.015	200	1.06	0.45	1.8458	12.4362	91.9534
trial 1	0.032						2.1599	14.2191	90.3900
trial 2	0.036						2.9532	13.1589	91.4656
trial 3		0.99					4.2438	14.0474	118.7260
trial 4		0.96					2.8654	14.4821	50.0261
trial 5			0.016				1.9604	12.4859	91.4409
trial 6			0.014				1.9077	12.6757	90.5764
trial 7				250			1.8568	12.4131	92.0628
trial 8				300			1.9214	12.3499	91.7103
trial 9				150			1.8640	12.7350	91.6831
trial 10				400			1.9785	12.4593	90.9310
trial 11					1.08		5.7294	15.1673	133.4187
trial 12					1.04		3.9063	17.5984	58.8251
trial 13						0.35	1.9117	12.6550	91.8565
trial 14						0.55	2.6708	16.0703	86.8390

Parameter values set as described under the "base case" unless otherwise stated

Simulated population size = 500

γ = 0.5 and ρ = 0.4 set exogenously

The statistics $p\left(\hat{x}_{it}\right)$ are then compared to $p\left(x_{it}\right)$, where x_{it} denotes the consumption generated for household i at age t by the simulation model:

$$test\_c = \sum_{i=1}^{N} \sum_{t=20}^{90} \left| p\left(\hat{x}_{it}\right) - p\left(x_{it}\right) \right| \tag{20}$$

for simulated population size N.

Having observed that the relationships between the various model parameters and the three test statistics are highly non-linear, a trial-and-error search was used to identify the calibrated parameter values. A sample of the three test statistics, obtained for parameter values in the vicinity of those selected for the simulation model is presented in Table 4.

5 Conclusions

The microsimulation model described here uses a traditional economic framework to adjust for behavioural responses to counterfactual policy regimes. Incorporating these types of behavioural responses into a statistical regression model would require expectations regarding all of the exogenous variables for every conceivable future period of an individual's life to be included as exogenous variables in the regression equation. In practice, limitations of existing survey data make this impossible. The current framework does, however, make a number of assumptions that should be taken into consideration when interpreting the results obtained:

1. *Rational Expectations:* The model assumes that households have rational expectations which are consistent with the actual frequencies of the various variables generated. There is a growing body of research that suggests that individual may be myopic when making intertemporal decisions, and that hyperbolic discount factors may help to obtain a better reflection of reality.

2. *Homogenous preferences:* All households are assumed to posses the same preferences for consumption and leisure.

3. *Perfect foresight with regard to household size:* The omission of a stochastic element for the specification of household size, coupled with the assumption of rational expectations, implies that all individuals are able to perfectly predict the life-course demographic specification of their respective households. That demographic uncertainty exists, and is a factor that individuals take into consideration when making labour and consumption decisions, is a contention that is difficult to dispute.[15]

4. *Limited household heterogeneity:* Most existing microsimulation models include very many more household specific variables than the six that are generated by the present model. The list of characteristics included in the present model has been chosen on the basis of their relevance with regard to household savings and retirement decisions, and

---

[15] Some recent literature has, for example, suggested that female labour participation may depend partly upon risk of divorce.

has been kept deliberately short to facilitate analytical transparency. The specification of the model does, however, permit additional household heterogeneity to be included, should it be desired.

5. *Restrictive income generating procedure:* As in the case of the short list of simulated household characteristics, the income generating procedure has been kept deliberately simple to simplify analysis of observed trends. The simplifications inherent in the income generating procedure are achieved at the expense of descriptive power regarding the sources of household income heterogeneity. Since the project for which the model has been constructed is not concerned with describing the sources of labour income differences between households, the simplifications inherent in the model adopted for simulating household human capital seem justified.

6. *Differential Mortality:* There is some evidence to suggest that wealthy individuals tend to live longer than those less well-off. This could affect some of the budgetary implications of the analysis based upon data derived from the simulation model.

Some of these assumptions (such as supposing that decisions are made by rational agents) are endogenous characteristics of the simulation framework adopted. Others are easily relaxed. In selecting the current specification of the model, care has been taken to facilitate a transparent analysis of savings and retirement behaviour. Nevertheless, the model remains a work in progress, and any comments or suggestions made regarding its form are very much appreciated.

6 References

Andersson, I., Brose, P., Flood, L., Klevmarken, N.A., Olovsson, P. & Tasizan, A. (1992), "MICROHUS - A microsimulation model of the

Swedish household sector", *International Symposium on Economic Modelling*, Gothenburg.

Auerbach, A.J. (1997), *Fiscal Policy: Lessons from Economic Research.* London: MIT Press.

Balcer, Y. & Sadka, E. (1986), "Equivalence scales, horizontal equity and optimal taxation under utilitarianism", *Journal of Public Economics*, 29, pp. 79-97.

Basu, N., Pryor, R. & Quint, T. (1998), "ASPEN: A microsimulation model of the economy", *Computational Economics*, 12, pp. 223-41.

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

Callan, T., O'Donoghue, C. & O'Neill, C. (1996), "Simulating Welfare and Income Tax Changes: The ESRI Tax-Benefit Model", *The Economic and Social Research Institute.*

Cogneau, D. & Robilliard, A.S. (2000), "Growth, distribution and poverty in Madagascar: Learning from a microsimulation model in a general equilibrium framework", *Trade and Macroeconomics Division, International Food Policy Research Institute*, 61.

Creedy, J. (1985), *Dynamics of Income Distribution.* Oxford: Basil Blackwell.

de Janvry, A., Sadoulet, E. & Fargeix, A. (1991), "Politically feasible and equitable adjustment: Some alternatives for Ecuador", *World Development*, 19, pp. 1577-94.

Debelle, G. & Swann, T. (1998), "Stylised facts of the Australian labour market", *Reserve Bank of Australia Research Discussion Paper*, 9804.

Dervis, K., De Melo, J. & Robinson, S. (1982), *General Equilibrium Models for Development Policy.* Cambridge: Cambridge University Press.

Fjærli, J.A.E., Gravningsmyhr, H., Holmoy, A.M.K. & Lian, B. (1995), "The Norwegian microsimulation model LOTTE: Applications to personal and corporate taxes and social security benefits", *Microsimulation Unit Discussion Paper*, MU9504.

Giannarelli, L. (1992), *An Analyst's Guide to TRIM2.* Washington D.C.: Urban Institute Press.

Gibrat, R. (1931), *Les Inegalites Economiques.* Paris: Sirey.

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

Hotz, V.J., Klerman, J.A. & Willis, R.J. (1997), "The economics of fertility in developed countries". In M.R. Rosenzweig & O. Stark (Eds.), *Handbook of Population and Family Economics.* Oxford: Elsevier Science.

Kalecki, M. (1945), "On the Gibrat distribution", *Econometrica*, 13, pp. 161-70.

Lucas, R.J. (1976), "Econometric Policy Evaluation: A Critique". In K. Brunner & A. Meltzer (Eds.), *The Phillips Curve and Labor Markets.* Amsterdam: North-Holland.

Macunovich, D.J. (1998), "Fertility and the Easterlin hypothesis: An assessment of the literature", *Journal of Population Economics*, 11, pp. 52-111.

McClements, L. (1977), "Equivalence scales for children", *Journal of Public Economics*, 8, pp. 191-210.

Meagher, G.A. (1993), “Forecasting changes in the income distribution: An applied general equilibrium approach”. In A. Harding (Eds.), *Microsimulation and Public Policy.* Amsterdam: Elsevier.

Merz, J. (1991), “Microsimulation - A survey of principles, developments, and applications”, *International Journal of Forecasting*, 7, pp. 77-104.

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

Muellbauer, J. & van de Ven, J. (2003), “Equivalence scales and taxation”. *Mimeo.*

Nerlove, M., Razin, A. & Sadka, E. (1984), “Income distribution policies with endogenous fertility”, *Journal of Public Economics*, 24, pp. 221-30.

Orcutt, G. (1957), “A new type of socio-economic system”, *Review of Economics and Statistics*, 58, pp. 773-97.

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

Redmond, G., Sutherland, H. & Wilson, M. (1998), *The Arithmetic of Tax and Social Security Reform: A Users' Guide to Microsimulation Methods and Analysis.* Cambridge: Cambridge University Press.

Sutherland, H. (2001), “EUROMOD: An integrated European benefit-tax model”, *EUROMOD Working Paper*, EM9/01 - Final Report.

Sutherland, H. (1995), “Static microsimulation models in Europe: A survey”, *University of Cambridge Department of Applied Economics Working Paper*, 9523.

Taylor, L. (1990), *Socially Relevant Policy Analysis. Structural Computable General Equilibrium Models for the Developing World.* Cambridge, Mass.: MIT Press.

van de Ven, J. (1998), "A dynamic cohort microsimulation model", *University of Melbourne Department of Economics Research Paper*, 637.

Zaidi, A. & Rake, K. (2001), "Dynamic microsimulation models: A review and some lessons for SAGE", *SAGE Discussion Paper*, SAGEDP/02.