This is a repository copy of *Modelling the composition of household portfolios: a latent class approach*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/193236/

Version: Published Version

Monograph:

Alzuabi, R., Brown, S., Harris, M.N. et al. (1 more author) (2022) Modelling the composition of household portfolios: a latent class approach. Working Paper. Sheffield Economic Research Paper Series, 2022019 (2022019). Department of Economics, University of Sheffield ISSN 1749-8368

Modelling the composition of household portfolios: A latent class approach

Raslan Alzuabi, Sarah Brown, Mark N. Harris, Karl Taylor

MODELLING THE COMPOSITION OF HOUSEHOLD PORTFOLIOS: A LATENT CLASS APPROACH[*]

Raslan Alzuabi[†], Sarah Brown[†#], Mark N. Harris[‡], and Karl Taylor[†#]

[†]Department of Economics, University of Sheffield
[‡]School of Accounting, Economics and Finance, Curtin University
[#]IZA Bonn

3rd November 2022

Abstract

We introduce the latent class modelling approach to the analysis of financial portfolio diversification at the household level. We explore portfolio allocation in Great Britain using household panel data based on a nationally representative sample of the population, namely the Wealth and Assets Survey. The latent class aspect of the model splits households into four groups, which serves to unveil a more detailed picture of the determinants of portfolio diversification than existing econometric approaches. Our findings reveal a pattern of class heterogeneity that conventional econometric models are unable to identify as the statistical significance as well as the direction of the effect of some explanatory variables varies across the four classes. When comparing our preferred latent class estimator to the commonly used approaches, we find that treating the population as a single homogeneous group may lead to biased parameter estimates and suggests that policy based on such models could be inappropriate or erroneous.

Keywords: Diversification; Financial portfolio; Household finance; Latent class model; Panel data;
JEL Classification: D14; G11; G40

1. Introduction and background

Over the last four decades, there has been considerable attention paid in the household finance literature to the composition of financial portfolios at the household level, exploring both the range of assets held as well as the amount of financial wealth allocated to distinct asset types. Such interest is not surprising given the significant increase in the number of financial products available over this time period, including shares and mutual funds, with varying degrees of risk and return associated with different asset types. Given that the composition of financial portfolios has implications for the exposure to financial risk faced by households, this remains an important area of research for both academics and policymakers.

In this context, many empirical studies, such as Bertaut (1998) and Shum and Faig (2006), have focused on the determinants of holding particular types of assets, with considerable interest in stockholding amongst U.S. households. This focus on risky asset holding has been explored in the context of the well-established 'stockholding puzzle', whereby very few households hold stocks despite the relatively high expected returns. In these studies, household characteristics such as age, gender, education, ethnicity and wealth, are found to be important determinants of portfolio composition, as are health status, the level of risk aversion and the planning horizon of the household. Similar studies have been undertaken for other countries including the Netherlands, Hochguertel et al. (1997), Australia, Cardak and Wilkins (2009) and Italy, Guiso et al. (1996). Although such studies have revealed some interesting insights relating to the determinants of stockholding at the household level, it is important to acknowledge that the focus on a particular type of asset reveals limited information on the diversification of household portfolios.[1]

In the early seminal contribution by Blume and Friend (1975), many private investors

---

[1] It is also important to acknowledge that in a number of studies, asset types are classified into different types such as safe, risky, and medium risk, *i.e.*, the focus lies beyond risky asset holding. For example, Rosen and Wu (2004) split the household's financial assets into four categories, namely; safe, risky, retirement and bonds. This asset classification has also been adopted by Berkowitz and Qiu (2006), Fan and Zhao (2009) and Borgan and Fertig (2013).

are found to hold undiversified portfolios of risky financial assets in contrast to the predictions of portfolio theory, proposed by Markowitz (1952), which indicates that, regardless of the degree of risk aversion, households should hold diversified portfolios. Similar evidence of a lack of diversification, even in the context of a sample of high income households, is reported by Kelly (1995), who explores the number of stocks held as a measure of diversification. In general, such theoretical predictions of portfolio theory are in stark contrast to the empirical observation that many households only hold a small number of asset types (e.g., Campbell, 2006; King and Leape, 1998).

One fundamental insight of portfolio theory asserts that, by holding a well-diversified portfolio, investors can reduce the idiosyncratic risk (i.e. the risk that is not compensated by the expected return) of their portfolio without sacrificing the return. However, underdiversification not only affects the asset-allocation and intertemporal consumption decisions of households, upon aggregation, it also distorts aggregate growth, which in turn amplifies social welfare losses, see Bhamra and Uppal (2019). Gaudecker (2015) argues that portfolio underdiversification ranks among those mistakes that are potentially most costly. Florentsen et al. (2019) analysed data related to stock market investors in Denmark and found that only 2% of their sample hold more than 20 stocks in their portfolio.[2] They estimated that underdiversification is costing the Danish population of stock holders 400 million US dollars annually as investors could eliminate 60% of their portfolio risk by moving from a portfolio with one randomly selected stock to a well-diversified portfolio.[3]

Given the implications of portfolio underdiversification, a number of studies have examined its determinants (e.g., Gaudecker, 2015; Goetzmann and Kumar, 2008; Calvet et al., 2007; Roche et al., 2013). Most of these studies report that the level of diversification is greater among older, wealthy, high-income, financially literate and educated investors. Other factors have also been examined in the literature. For example, Goetzmann and

---

[2]The availability of administrative data facilitates comprehensive analyses of investors' portfolio composition. For example, Calvet et al. (2007) used data from Swedish government records; and Grinblatt et al. (2012) used data from the Finnish Central Securities Depository (FCSD) Registry.

[3] Florentsen et al. (2019) is one of the few studies that measures the aggregate cost of underdiversification as they have access to data that contains information on all retail investors in Denmark.

Kumar (2008), using data related to retail investors at a major U.S. discount brokerage house, found that the level of underdiversification is correlated with three psychological biases: propensity to hold local stocks; sense of over-confidence; and trend-following behavior. Gaudecker (2015) shows that the largest losses from underdiversification are incurred by those who score low on financial literacy and those who do not seek advice from professionals or private contacts with their investments. On the other hand, Calvet et al. (2007) examine the efficiency of Swedish households' investment decisions and find that many Swedish households portfolios are well-diversified, with the performance of their portfolios outperforming the Sharpe ratio of their domestic stock index, which reflects the substantial share of international securities held through most Swedish mutual funds. Financial constraints are also found to be a significant determinant of the level of diversification (e.g., Roche et al., 2013; Liu, 2014).

As highlighted by Barasinska et al. (2009), although portfolio diversification has attracted the attention of academics for many decades, there is no commonly accepted approach to measuring the extent of diversification in household portfolios. Early contributions have explored portfolio diversification from the perspective of the number of different types of assets held. In this vein, Blume and Friend (1975) use the total number of securities held as a measure of diversification. Barasinska et al. (2009) refer to the number of asset types held in a portfolio as 'naive' diversification, with greater diversification associated with a larger number of asset types held. They relate this to the approach whereby individuals split their wealth evenly among available assets types, *i.e.*, the 1/n strategy, see Benartzi and Thaler (2001). The second measure explored by Barasinska et al. (2009) is based on grouping asset types according to the associated risk, specifically, low risk, moderate risk and high risk. According to this approach, a sophisticated investor categorises assets according to their risk and return and assigns them to one of these three classes. They find that the number of asset types held is negatively associated with the degree of risk aversion and that the propensity to hold complete portfolios decreases as risk aversion increases.

In terms of the econometric methods used in the existing literature on household portfolios, studies focusing on the holding of assets types have generally used standard models for limited dependent variables such as probit and logit frameworks, whereas those exploring asset shares have tended to use linear regression models or models that account for the truncated nature of the dependent variable such as tobit models. In contrast, our methodological contribution builds on Abreu and Mendes (2010), who recognise that an appropriate approach to modelling a portfolio diversification measure based on the number of asset types held is a count model given that it can only take non-negative integer values.[4] Using a Poisson model, they analyse a cross-sectional survey of 1,268 Portuguese investors and find that specific financial knowledge is positively associated with the number of assets in a financial portfolio.

In contrast to the existing literature, we explore portfolio allocation in Great Britain using panel data based on a nationally representative sample of the population. Interestingly, there is limited research on portfolio allocation in the UK, with much of the early literature based on US data (e.g., Goetzmann and Kumar, 2008; Ivković et al., 2008; Dimmock et al., 2016). In addition, many existing studies are based on cross-section data, such as the U.S. Survey of Consumer Finances (e.g., Kelly, 1995; Polkovnichenko, 2005) or based on samples of subgroups of the population (i.e. investors only), such as online brokers and administrative data, as in (Goetzmann and Kumar, 2008; Calvet et al., 2007; Grinblatt et al., 2012; Florentsen et al., 2019). Given that wealth is found to be an important determinant of asset holding and that wealth accumulates over the life cycle, the use of panel data appears to be particularly appropriate in this context. A related point is raised by Polkovnichenko (2005), who argues that one of the main limitations of the current empirical literature on portfolio diversification is that the samples used for empirical analysis are frequently not representative of the entire population. Our nationally representative sample of households in Great Britain does not suffer from such

---

[4]A number of papers in the empirical literature have also used the number of assets held as a measure of portfolio diversification (e.g., Ivković et al., 2008; Goetzmann and Kumar, 2008).

limitations.

In addition to our focus on a nationally representative sample of households and on panel data, we make an important methodological contribution to the literature on the diversification of household financial portfolios by introducing the latent class modelling approach to this area of research. Latent class modelling has been used extensively in other areas of economics including consumer behaviour (e.g., Chung et al., 2011) and health economics (e.g., Deb and Trivdei, 1997), but is yet to be widely applied to the household finance literature. One recent exception is Gerhard et al. (2018), who use a Finite Mixture Model (FMM) to explore whether psychological traits affect the level of household savings. The advantages of using the FMM approach in this application lie in its superiority in introducing unobserved heterogeneity by partitioning the sample endogenously into a number of homogeneous classes rather than relying on user-defined sub-samples, as in the existing literature.[5] They find evidence of two distinct classes and that accounting for latent heterogeneity when studying the drivers of savings behaviour is important as drivers differ between the two groups. More broadly in the finance literature, Durand et al. (2022) use a FMM to examine the capital structure decisions of firms in relation to adjustment towards target levels of leverage, which maximize a firm's value.

The latent class approach probabilistically divides the population into a set of homogeneous groups. Within each class, an appropriate statistical model applies, which in our case, following Abreu and Mendes (2010), is based upon a count specification as the number of asset types held can only take non-negative integer values. The latent class approach is arguably well-suited to the analysis of portfolio diversification given the potential for very diverse financial behaviour within a population. For example, in our data set, the number of different asset types held by households in Great Britain ranges from 0 to 21.

Such a latent class approach is advantageous, as it simultaneously introduces heterogeneity into the empirical framework and, *ex post*, allows the splitting of the popu-

---

[5]The received, minor, difference between FMM and latent class modelling is explained below.

lation into various sub-groups of households according to their portfolio diversification behaviour. *Ex post*, we can then examine sample statistics for each class by detailed asset type to evaluate the extent of portfolio diversification. Building on the heterogeneity afforded by the latent class approach, we take advantage of the panel data available to us to account for unobserved heterogeneity that is likely to be an important driver of household financial decisions.

2.   Methods

As stated above, the latent class approach involves probabilistically splitting the population into a finite number of homogeneous classes or types. Within each of these classes, the same statistical model applies, but the same explanatory variables are allowed to have different effects across the classes. This modelling approach starts from the premise that, although the classes are latent, *ex post*, researchers frequently label them according to the expected value within each class. Thus, finding evidence of the features of each class is an important outcome of the modelling approach. Our basic hypothesis is that there are distinct, but observed, types (or classes) of households with respect to their asset-holdings. Therefore, an appropriate approach here is based on the generic Latent Class Model (*LCM*) approach, which attempts to model this (e.g., McLachlan and Peel, 2000). Importantly, we have priors as to the drivers of these unobserved classes, so our generic approach will be based on the latent class modelling literature, but explicitly with predictors in the class equation(s).

Initially, for ease of exposition, assume cross-sectional data so that the overall density for household i $(i = 1, \ldots, N)$, $f(y_i|x_i, \beta)$, is assumed to be an additive mixture density of Q distinct sub-densities weighted by their appropriate mixing probabilities, π_{iq}. The outcome variable of interest, y_i (*i.e.*, the number of financial assets held), is driven by the $(k_x \times 1)$ vector of covariates in the model, x_i. Importantly, these will be allowed to have differing effects across the different q classes. β denotes all of the parameters of the

model. Hence, the corresponding mixed density will be

$$f(y_i|x_i) = \sum_{q=1}^{Q} \pi_{iq} \times f_q(y_i|x_i, \beta). \tag{1}$$

A restricted version of the approach, would have $\pi_{iq} = \pi_q$, whereas this would appear to be a mis-specified approach especially as we have priors as to the drivers of this process. The former is generally referred to as a FMM and the latter, a latent class model. So, here we allow π_{iq} to be a function of predictors (z_i), such that

$$\pi_{iq} = g(z_i, \gamma), \tag{2}$$

where g is an unspecified function and γ unknown parameters relating (z_i) to the class probabilities.

However, we first require an appropriate form for $f_q(y_i|x_i, \beta)$. Given the nature of the outcome variable, observable counts of the number of assets held, an appropriate form for f_q in equation (1), is one that respects the nature of this. Obvious examples here would be Poisson or Negative Binomial models/densities (Cameron and Trivedi, 1998). We note that the latter is normally employed to relax the restriction of the former that the conditional mean and variance are equal. In our approach, using the Poisson distribution for $f_q(y_i|x_i, \beta)$, once mixed as given by equation (1), the mixture Poisson density no longer embodies this restrictive assumption.

As is usual (Greene, 2018), we employ a multinomial logit (MNL) functional form for $g(z_i, \gamma)$ to model class membership:

$$prob(q_i = q) = \frac{\exp(z_i \gamma_q)}{\sum_{q=1}^{Q} \exp(z_i \gamma_q)} \tag{3}$$

An important part of the latent class approach concerns determining the appropriate number of classes, Q^*. A common approach is to use information criteria (IC) metrics; such as BIC/SC, Schwarz (1978), AIC, Akaike (1987), corrected AIC, $CAIC$, Bozdogan

(1987), and Hannon-Quinn, $HQIC$, Hannan and Quinn (1979). Such approaches can simultaneously be used to choose specifications including the choice of Q^* and cross-sectional versus panel variants (see below). As such, we use these metrics in determining our preferred approach.

The $Vuong$ test for non-nested models can also be used here (Vuong, 1989). As model size will vary considerably over different class models, the 'BIC' correction factor can be used here, as proposed in Vuong (1989). The standard $Vuong$ test (for example, Greene, 2018) for comparing two competing models $j = 1, 2$ is based on m_i, the individual differences in the two log-likelihoods, such that

$$m_i = ln \left(\frac{f_1 \left(y_i | \tilde{d}_i \right)}{f_2 \left(y_i | \tilde{d}_i \right)} \right), \tag{4}$$

where f_j are the respective likelihoods from the two $j = 1, 2$ competing models, and $\tilde{d}_i$ corresponds to all of the covariates in the model, $i.e.$, $\tilde{d}_i \in (x_i, z_i)$, for household i. The $Voung$ test is given by

$$V = \frac{\sqrt{n}\bar{m}}{s_m}, \tag{5}$$

where n is the sample size and $\bar{m}$ and s_m are the simple sample average and standard deviation of m_i, respectively. The test has a limiting standard normal distribution, with values of $|V| < 1.96$ being indeterminate, whereas large positive (negative) values favour model 1 (2). The BIC corrected version of this (Vuong, 1989) is given by

$$m_i^c = m_i + (k_2 - k_1) \frac{ln(n)}{2n}, \tag{6}$$

where k_j refers to the number of estimated parameters in model j. The $Vuong$ test is strictly a pairwise one, so with many potential competing models, it is possible to use the approach suggested in Durand et al. (2022) in that an appropriate model selection metric amongst all models is that model with the most favoured number of pairwise selections.

All of the expected values (EVs), predicted count probabilities, and partial effects,

within class are given by the usual Poisson expressions. For example, the class specific expected value EV_q is given by:

$$EV_q = \exp\left(x_i \beta_q\right) \tag{7}$$

Comparable overall quantities will be given by the probability-weighted average of the class ones, where it is usual to use the prior probabilities as the weights.

In predicting class membership, one could use the (prior) probabilities described above, or more favourably, the posterior ones, which additionally take into account the information on the observed outcome (Greene, 2018). These are defined as

$$prob(q_i = q|\tilde{d}_i, y_i) = \frac{f(y_i|q_i = q, \tilde{d}_i) \times prob(q_i = q|\tilde{d}_i)}{\sum\limits_{q=1}^{Q} f(y_i|q_i = q, \tilde{d}_i) \times prob(q_i = q|\tilde{d}_i) = L_i} \tag{8}$$

where L_i is the likelihood for the household, used in estimation. Note that for all secondary quantities of interest, standard errors of these can be obtained using the delta method, (Greene, 2018).

As noted above, our empirical analysis is based on panel data. Having repeated observations on each household allows us to better identify class membership. To this extent, we treat the model parameters $\theta \in (\beta, \gamma)$ non-parametrically as a random vector with discrete support, where the discrete outcomes define the classes. Thus, the class probabilities are constant for each household over time, and the joint density for the T_i observations for household i is given by

$$f_i(y_{i1,...,}y_{i,T_i}|z_i, x_{i1,...,}x_{i,T_i}) = \sum_{q=1}^{Q} \left\{ \pi_{iq} \prod_{t=1}^{T_i} f_{it}(y_{it}|z_i, x_{it}, \theta_q) \right\}, \tag{9}$$

and the corresponding log-likelihood is given by:

$$\ln L = \sum_{i=1}^{N} \ln\{ f_i(y_{i1,...,}y_{i,T_i}|z_i, x_{i1,...,}x_{i,T_i}) \}. \tag{10}$$

It is also possible to allow for random effects in nonlinear panel models to account for unobserved household heterogeneity (Matyas and Sevestre, 2008). These can be class-specific, but will be independent as households can only be in one class. Note that this complicates estimation, as the Q household effects need to be integrated out of the likelihood function. We use simulated maximum likelihood techniques, using 100 Halton draws.

3. Data

Our empirical analysis is based on data from the *Wealth and Assets Survey* (WAS), which is a biennial longitudinal household survey for Great Britain measuring the personal and economic well-being of individuals and households by assessing levels of assets, debt, savings and planning for retirement.[6] The WAS also provides information on a host of socio-demographic factors that we control for in our analysis, as detailed below. The survey started in 2006 and covers Great Britain: England; Wales; and Scotland. Our empirical analysis is based upon waves 2-5 of the survey, yielding 28,756 heads of household (N) and total observations (NT) equal to 46,424. Table I shows that 36% (19%) of household heads are observed once (three times or more) in the panel. The outcome variable of interest is the number of financial assets held, which is comprised of the following assets:[7] savings accounts; national savings accounts; Investment Savings Accounts (ISAs); fixed term investment bonds; unit trusts; employee shares and/or share options; shares; bonds and gilts; insurance products; endowment or regular premium policies; single premium policy; and other types of investment.[8] The minimum (maximum) number of financial assets held is 0 (21) and the distribution is shown in Figure 1, where 80% of households

---

[6]It should be acknowledged that the WAS over-samples wealthier households compared to other postal addresses. The reason for this is that, in general, other household surveys do not adequately capture the top part of the wealth distribution, see ONS (2012).

[7]In accordance with Abreu and Mendes (2010), we exclude current accounts from the definition of the number of financial assets held, where 98% of household heads have such an account.

[8]Within each of these broad asset categories, there are sub-categories which we distinguish between in defining the number of assets held. The full list of asset types is detailed in Table VI.

hold fewer than five financial assets. Clearly, the number of assets held is not continuous being characterised by kurtosis of 4.2 and the Shapiro-Wilk test for normality rejects the null at the 1% level.

Many empirical studies have explored the relationship between household portfolios and a wide range of household characteristics including socio-demographic characteristics such as age, education and health and financial characteristics such as net wealth, employment status and income. Young households, for example, with low levels of financial wealth have been found to hold undiversified portfolios comprising a small number of assets, see, Roche et al. (2013). Hence, our set of explanatory variables follows this literature.

In terms of the covariates, x_i, used to model the number of financial assets held (y_i) we include head of household characteristics such as: a quadratic in age; whether single, never married (other marital states form the reference category); educational attainment - whether degree level or above, or whether a qualification below degree level (no education is the omitted category); being in good health; whether employed; occupation - whether managerial or professional, intermediate, small employer and own account, lower supervisory and technical (semi routine and unemployed is the omitted category); having a defined benefit occupational pension; and whether the head of household is financially optimistic or financially pessimistic (no expected change in financial position is the reference category). In addition, a number of household characteristics are included: the natural logarithm of labour income; the natural logarithm of non-labour income; the natural logarithm of pension wealth; the natural logarithm of net wealth (defined as liquid assets plus house value minus the amount of unsecured and secured debt); the number of children in the household; and the number of adults in the household (excluding the household head).

In our panel data framework, given that the class probabilities are constant over time for each household, we follow the existing literature, e.g. Clark et al. (2005), Bago d'Uva and Jones (2009) and Greene (2018), and parameterise the model such that time-invariant

12

covariates, z_i, influence the probability of being in a particular class (q_i). Specifying the class membership equation with time invariant head of household controls in this way is akin to parameterising the household's fixed effect of being in each class. The covariates, z_i, which are specific to determining class membership, include the following head of household characteristics: cohort of birth - whether born 1945-55, whether born 1955-64 or whether born from 1965 onwards (pre 1945 is the omitted category); risk attitudes;[9] time preference;[10] whether their mother had post school education; whether their father had post school education; whether their mother was employed or self employed; whether their father was employed or self employed; whether they grew up in a single parent household; and the number of siblings when growing up.[11] The only control variable common to both the outcome and class membership equations, *i.e.*, appearing in both x_i and z_i, is gender.

Sample summary statistics are presented in Table I, where it can be seen that 58% of heads of household are male and their average age is 48. In terms of the controls for determining class membership, only 6% (9%) of the respondents' mothers (fathers) had post school education. Clearly, on average, the respondent's father was more likely to be employed than their mother, at 70% and 46%, respectively. Approximately 10% of household heads grew up in a single parent household. Turning to those covariates in the outcome y_i equation, *i.e.*, number of financial assets held, 49% of respondents are single, around 30% have at least degree level education, with approximately 15% having no qualifications. Labour income not surprisingly is higher than non-labour income, with means of £3,480 and £3,095, respectively. An equal proportion of household heads are financially optimistic or financially pessimistic about their finances for the coming year.

---

[9]This is a binary control constructed from the following question: *If you had a choice between a guaranteed payment of one thousand pounds and a one in five chance of winning ten thousand pounds which would you choose? 0=Guaranteed payment of £1,000, 1=One in five chance of £10,000.*

[10]Defined as a binary control constructed from the following question: *If you had a choice of receiving a thousand pounds today or one thousand one hundred pounds in 12 months which would you choose? 0=£1,000 today, 1=£1,100 next year.* Both risk attitudes and time preference are observed to be time invariant in the data. More generally, Schildberg-Horisch (2018) has recently argued that individual risk preferences appear to be persistent and moderately stable over time

[11]Childhood related questions are specific to when the respondent was around the age of 14.

4. Results

In terms of model comparison, we compare a range of latent class estimators using standard IC metrics to identify the preferred model. The models compared include a standard linear estimator, poisson and negative binomial count models, and the latent class approach (from two to five classes). Table II presents the summary IC, where Panel A shows the IC for the pooled models and Panel B the IC for the random effects models, where the longitudinal nature of the data is taken into account. All of the IC metrics favour the panel models *within* each type of estimator, e.g. panel linear versus pooled linear (OLS), and also *across* the alternative estimators. The panel 4-class MNL latent class model dominates all alternative specifications (see Panel B). Moreover, in terms of the latent class approach, the optimal structure is found to be four classes. The Vuong test reported in Panel C also confirms that the 4-class MNL model is the optimal latent class structure amongst the competing alternatives.

Tables III and A1 present the coefficients and partial effects, respectively, associated with the determinants of the number of financial assets held, evaluated at the sample means of the covariates. Although the MNL latent class approach does not impose any ordering on the expected values, we impose ordering on the classes *ex-post* according to the class expected values (EV). Hence, by definition, class 1 is characterised by a relatively low number of asset types held, at 1.61, and a relatively low level of diversification. In contrast, class 4 is characterised by a higher level of diversification, with an expected value of 4.06. The maximum number of financial assets held increases monotonically across classes 1 to 4, where the classes are ordered according to the ex-post EVs. One of the key features of the latent class approach is that the covariates are allowed to have different effects across the 4 classes, thereby unveiling a more detailed picture of the determinants of household portfolio composition than the modelling approaches employed in the existing literature.

This is supported by the results in Tables III and A1, which reveal a pattern of class heterogeneity that conventional econometric models are unable to identify. More

14

specifically, Table III shows that the magnitude, statistical significance, as well as the direction of the effect of some explanatory variables varies across the 4 classes. For example, the direction of the effect changes between the 4 classes for gender, marital status and labour income. Households with a male head in class 1, a class with a relatively low level of asset diversification, are associated with a lower number of financial assets than households with a female head by a factor of 0.886, whereas, for class 2, households with a male head have more financial assets by a factor of 1.061, ceteris paribus. However, the effect becomes negative again for class 3 and statistically insignificant for class 4. Households with a single head in the first three classes have more financial assets than households with a married head and the magnitude of the effect is similar across these classes, which is in line with the findings of Abreu and Mendes (2010), who argue that married investors are financially less well-informed.

The impact of labour income on the number of financial assets held also exhibits considerable heterogeneity in the effects across the 4 classes. Specifically, a 1% increase in labour income is associated with 1.058 more financial assets for households in class 2, whereas it is associated with 0.905 less financial assets for those in class 3, ceteris paribus. Such findings indicate that labour income influences diversification for the two middle classes, whereas it has no statistically significant for households in the lowest or the highest class of diversification. Findings of other empirical studies usually report a positive relationship between income and the level of diversification, (e.g., Calvet et al., 2009, 2007; Abreu and Mendes, 2010).

The other financial related variables have the expected impact across the 4 classes. However, the magnitude of the effect varies between classes. For example, a 1% increase in pension wealth is associated with 1.558 more financial assets for class 1, whereas the same increase is associated with only 1.156 more financial assets for class 4. On the other hand, net wealth has a more pronounced impact on households in the top 2 classes compared to those in the bottom 2 classes (i.e. those classes characterised by less diversification). Specifically, a 1% increase in net wealth is associated with only 1.019 more financial assets

for class 1, whereas the same increase is associated with 1.243 more financial assets for class 4. Having a head of household with a defined benefit pension is found to be associated with holding less financial assets across all classes, a result that might reflect the possibility that those who have a defined contribution pension are more exposed to the concept and implications of diversification than those with defined benefit pension schemes. In general, such findings not only highlight the importance of allowing parameter estimates to vary by class, but also the importance of distinguishing between different income sources.

The impact of the age of the head of household is only statistically significant for class 1. The findings for class 1 suggest that the older is the head of the household the lower is the number of financial assets held and the magnitude of the impact increases at a decreasing rate, as shown by the quadratic term. A number of empirical papers report that age is a significant determinant of underdiversification (e.g., Goetzmann and Kumar, 2008; Calvet et al., 2007; Roche et al., 2013). Roche et al. (2013) argue that young investors are more likely to be financially constrained as they generally have a low value of wealth to income ratio. Therefore, they hold underdiversified portfolios given that financial constraints are a significant determinant of portfolio diversification.

Education has the expected impact across the 4 classes with the impact being strongest for those in class 2. Specifically, in class 2, having a head of household with a degree or above is associated with holding more financial assets than those who have no education by a factor of 1.415. A similar pattern is found for the employment dummy, with the impact being stronger for those in class 2, which is as expected and accords with the existing literature (e.g., Dimmock et al., 2016; Calvet et al., 2007; Abreu and Mendes, 2010). Heads of household who are employed in lower supervisory and technical jobs have statistically significant coefficients only if they are in class 1 and the associated IRRs factors are the smallest compared to the other occupations. In contrast, those who are in managerial and professional occupations have the strongest IRRs factors and this is the only statistically significant occupation for households in class 4 (i.e. the most diversified class). This is as expected as those in these types of occupations arguably have more

financial knowledge and experience.

The number of children and the number of adults in the household also have the expected effects. Specifically, the number of children is negatively associated with the number of financial assets and the opposite is observed regarding the number of adults. These findings may reflect the presence of children in the household being associated with higher costs, whereas more adults may bring more financial and economic knowledge and/or an extra source of income into the household. Being financially optimistic has a positive impact on the number of financial assets held, but this impact is only statistically significant for heads of household in classes 1 and 2. On the other hand, being financially pessimistic is negatively associated with the number of financial assets. However, the impact is statistically insignificant across all classes.

To summarise, the most important factors that lead to more diversification, as observed in the coefficients associated with households in class 4, are heads of household with higher levels of education, being in managerial and professional occupations, and having high levels of net wealth and pension wealth. Moreover, in general, the economic magnitudes stemming from the effects of the covariates are non trivial given the size of the IRRs relative to the class specific EVs.

The ρ parameters in Table III show the degree of association of the panel structure of the data, i.e. the extent of the unobservable intra-household correlation in the data over time. This may be an indication of some persistence in portfolio allocation. Specifically, the overall average of this correlation is 0.22, which is statistically significant at the 1% level. This shows the importance of the longitudinal nature of the data in modelling the number of financial assets, particularly for those households with the least diversified portfolios.

To assess the factors that are correlated with the probability of belonging to a specific class, Table IV presents the partial effects of the class probabilities, evaluated at the sample means of the covariates. In general, the findings indicate that gender, the birth cohort controls, the measure of risk attitudes, time preference and childhood conditions

17

are mostly statistically significant, supporting a well-specified class membership equation. Furthermore, the table reports the average posterior probabilities across classes, which show that class 1 is the largest class of the four, containing 39 percent of the sample and the smallest class is class 4, which contains only 11 percent of the sample.

Table IV shows that there is a clear impact of risk attitudes on the probability of belonging to each class. Specifically, households with heads, who are more willing to take risks, are more likely to be in class 4, the class with a higher number of financial assets held, and less likely to belong to classes 1 and 2. Whilst the measure of time value is only statistically significant for class 1, where it has the expected sign. These findings are in line with the existing literature that examines the determinants of stock holding at the household level (e.g., Cardak and Wilkins, 2009; Shum and Faig, 2006), as those who are more likely to take risk are more likely to hold a higher number of risky financial assets. Similarly, households with a male head are around 20 percentage points more likely to be in class 4 than households with a female head, which also ties in with the existing literature, which reports differences in risk preference by gender, e.g., Guiso and Sodini (2013), with females generally found to be more risk averse.

Households with a head who grew up in a single parent family are around 9 percentage points less likely to belong to class 4 and 2 percentage points more likely to belong to class 1, compared to those household heads who did not grow up in a single parent household. However, the probability of being in class 4 is positively associated with the number of siblings the household head grew up with. The birth cohort controls are statistically significant across most of the 4 classes, with those who were born after 1965 being more likely to belong to class 1 and less likely to be in class 4 relative to those who were born before 1945 (the omitted category). However, it should be acknowledged that the pattern of the impact of the other birth cohorts controls is not clear.

Parental employment status of the household head is also a strong predictor of class membership. To be specific, having a mother (father) who was an employee or self employed when the household head was around 14 years old increases the probability of

being in the highest class of diversification by approximately 19 (11) percentage points. In contrast, the probabilities of belonging to the other 3 classes are lower for these heads of household. Although parental education is statistically significant for classes 2 and 3, the direction of the impact is not as clear as that for parental employment status.

The discussion so far illustrates how the latent class approach unveils differential partial effects across classes, with the approach essentially being used as a means to allow for more unobserved heterogeneity in the modelling approach. If this is the case, then focus may actually lie on the overall partial effects and whether or not there are any differences in overall effects across model variants. To explore this, in Table V, we compare the overall partial effects between a linear random effects (RE), negative binomial (NegBin) model and our preferred 4-class MNL estimator. The table also reports the AIC and BIC statistics to compare the overall statistical performance of these models. As mentioned above, both statistics reveal that, statistically, the latent class MNL estimator is the preferred approach for modelling household portfolio diversification.

Although the general pattern of results is broadly consistent across the three models, there are some substantive differences in terms of size and statistical significance for a number of explanatory variables.[12] Specifically, in contrast to the results from the 4-class MNL model, the linear RE and the NegBin models reveal positive and statistically significant gender and marital status effects on the number of financial assets held. Similarly, non labour income, being financially pessimistic and age are found to have negative and statistically significant effects according to the linear RE and the NegBin models. Furthermore, in terms of the size of the effect, in comparison to the latent class modelling approach, the linear model seems to overestimate the partial effects whereas these effects are underestimated according to the negative binomial model.

In general, given that the average partial effects of some controls in the linear and the NegBin models are attenuated by the most populated class, class 1, results based on these

---

[12]In general, the linear RE and the NegBin models results are roughly consistent with the results of class 1 from the latent class model, which accords with expectations since this class dominates in probabilistic terms (containing 39 percent of the sample).

models may not adequately reflect the determinants of diversification for the other groups, which suggests that policy based on such models could be inappropriate or erroneous.[13]

4.1. Analysis of ex-post statistics

Ex post, we are able to split the population into various sub-groups of households, *i.e.*, classes based upon ex-post EVs, and analyse their portfolio diversification behaviour. To be specific, we can examine the composition of portfolios within each class and also across classes. This allows us to explore questions such as whether it is the case that class 4 is characterised by a more diversified portfolio. Using the naive measure, *i.e.*, the number of financial assets held, see Barasinska et al. (2009), the EVs suggest that, on average, class 4 is characterised by more diversified portfolios as the number of asset types held is higher. However, in this section, we explore how this relates to asset shares and combinations of different types of assets. For example, a household in class 4 might hold five types of assets but 95% of the total value of the assets, on average, might be held in a single asset.

Table VI shows the mean EV for each class (as discussed above), and the proportion of households without any financial assets. The latent class approach allows for no asset holding across each class, and this is, indeed, a characteristic of our data, where the minimum EV is zero across all classes. Interestingly, the proportion of households reporting zero assets does not decrease monotonically across classes (*i.e.*, as the EVs increase), where it can be seen that class 1 (EV=1.61) and class 3 (EV=2.73) have approximately 22% and 15% reporting zero financial assets, respectively, with class 2 (EV=2.71) characterised by only 5.13% of households reporting zero assets. At the other extreme, the maximum values for the number of assets held are as follows: class 1=7; class 2=12; class 3=13; and class 4=21. Not only does the mean number of financial assets increase across classes so does the standard deviation indicating more diversified sub-groups as the *ex post* EV increases monotonically by class.[14]

---

[13]Table A1 in the Appendix shows partial effects by class, which also indicate that the overall partial effects of the 4-class MNL model reported in Table V are also attenuated by the largest class, class 1.

[14]Note that the EVs do not precisely match the average number of financial assets reported in Table

We also consider a Herfindahl-Hirschman Index (HHI), which can be used to measure portfolio diversification, e.g. see Ivković et al. (2008). It is defined as follows:

$$HHI_q = \sum_{s_q=1}^{N_q} \|\omega_{s_q}\|^2 \tag{11}$$

where HHI_q is the metric for class q and there are N_q assets in the class. The share of asset, s_q, in the household financial portfolio is class specific (q) and is given by ω_{s_q}. The index ranges from $1/N_q$ to unity; hence classes which are more diversified have a lower HHI_q. The final row of the summary information reported in Table VI shows that this is indeed the case. The HHI increases monotonically across the classes, consistent with class 1 (4) being the least (most) diversified.[15] For example, for class 4, the value, $HHI = 0.08$, is equivalent to a household portfolio comprising 12 financial assets. The HHI statistics support the conjecture that the LCM performs well in terms of ranking the extent of portfolio diversification across the different sub-groups (i.e. classes).

Table VI also reports the proportions held in each type of financial asset within and across classes. In terms of savings and deposit accounts, for class 1, only 41.31% of households have such an account, which is below the sample mean. The proportion of households holding cash ISAs increases monotonically across classes, with around 78% of households in class 4 having cash ISAs. Similarly, the percentage of households who have investment ISAs in their portfolio increases monotonically across classes and it is noticeable that over half of the population have such assets. Share ownership also increases monotonically across classes, where class 4 dominates in terms of the proportion of households having such assets in their portfolio: approximately 23% hold employee shares and just under 80% have shares in UK and foreign companies. The same pattern emerges for other asset types and suggests that diversification increases across the classes, *i.e.*, as the *ex post* EV increases.

VI. This is because the EV statistic is generated from the LCM estimator *ex post*, see equation (7), whilst the mean number of financial assets is a raw unconditional sample statistic.

[15]The overall mean $HHI = 0.18$ is lower than that found by Ivković et al. (2008) based upon data for US households, which indicates that financial portfolios in Great Britain are more diversified.

It is noticeable from the statistics shown in Table VI that households do not appear to split their financial wealth evenly among available asset types, which is at odds with the 1/n strategy (Benartzi and Thaler, 2001). Furthermore, for class 1, on average, the majority of financial wealth is held in more liquid assets, *i.e.*, savings and deposit accounts, compared to class 4, where there are similar orders of magnitude of the proportions of wealth held in liquid and illiquid assets (e.g. ISAs and shares).

In Table VII, again based upon ex-post analysis, we analyse the amount held in each type of asset and its proportion of financial wealth. The average amount of financial assets held by households over the period was £58,174. However, financial wealth is not evenly distributed across classes with class 1 - arguably the least diversified based upon the analysis of Table VI - having, on average, portfolios with a total value of £11,548 compared to class 4, where the average amount of financial assets held is £215,857. Focusing on savings accounts - one of the most liquid assets - although class 4 is found, on average, to have the highest monetary amount, in terms of the proportion of financial assets, it is the lowest at around 22%, which is below the sample mean. Indeed, for class 4, a higher proportion of financial assets is held in cash and investment ISAs than savings at just under 30%. It is particularly noticeable for the group of households with the most diversified portfolios, *i.e.*, class 4, that such households not only hold a higher monetary amount of each asset in comparison to other groups (*i.e.*, classes) but that, with the exception of savings accounts, each asset also constitutes a much higher percentage of the total amount held in financial assets. This is particularly apparent for more illiquid assets, such as fixed term investment bonds, unit trusts and shares.

Having explored a wide range of financial assets from very liquid to highly illiquid, the *LCM* approach would appear to be convincing in terms of splitting households into sub-groups based on their underlying level of financial diversification. These sub-groups were ordered monotonically into different classes by the *ex post* EVs, based upon a naive measure of diversification derived from the underlying number of financial assets. The analysis in this section has revealed that, across the different classes, the asset shares

and combinations of distinct asset types become considerably more diverse across the classes as the EV increases. This is consistent with the *LCM* providing information on heterogeneity in household financial behaviour for various sub-groups, some of whom are less likely to hold well-diversified portfolios (e.g. class 1).

There are potential implications for policymakers interested in promoting savings behaviour in less diversified sub-groups, e.g. class 1, where 59% of households in this group do not have a saving or deposit account. Policy interventions may target this identified sub-group or attempt to manipulate certain characteristics through interventions. However, policy aimed at targeting specific groups based on observed behaviour is potentially limited in that it can only use discerned behavioral differences and may overlook the latent heterogeneity in the data. Hence, acknowledging that latent heterogeneity across groups generally exists in terms of the impact socioeconomic characteristics have on financial behavior is potentially important. In the least diversified sub-group, incorporating an appreciation of the complexity of the relationships between behavioural traits such as risk attitudes, time preference (both of which are found to be statistically significant determinants of class 1 membership) and diversification in policy design could be of importance. Specifically, targeting this sub-group with interventions designed to improve financial literacy might be beneficial and more precise than basing an intervention on observable characteristics alone. Ultimately, it would also be potentially less resource intensive than implementing policy aimed at the population as a whole.

5. Conclusion

Recent theoretical work suggests that the composition of household portfolios should be of interest to policymakers. In particular, Bhamra and Uppal (2019) show that under-diversified household portfolios can lead to lower macroeconomic growth. Encouraging greater diversification of household portfolios may consequently result in benefits that are not just restricted to improving household welfare. In this paper, we make an important

methodological contribution to the literature on the diversification of household financial portfolios by applying the latent class modelling approach, based upon a count model specification, to panel data based on a nationally representative sample of households in Great Britain.

Given the extent to which wealth accumulates over the life cycle and the potential for very diverse financial behaviour within a population, the use of panel data and allowing the determinants of household portfolio composition to vary across different subgroups of the population seems to be a potentially important approach in order to fully understand the drivers of diversification of household portfolios. Our results confirm this and show that the statistical significance as well as the direction of the effect of some explanatory variables varies across the four classes supported by our data, advocating the use of a modelling approach that can reveal such a pattern of class heterogeneity.

Our key findings include revealing the considerable heterogeneity in the effect of labour income on the number of financial assets held, which indicates that labour income influences diversification for the two middle classes, whereas it has no statistically significant effect for households in the classes with the lowest or the highest level of diversification. Furthermore, our empirical analysis suggests that there are noticeable differences in the magnitude of the effects of some explanatory variables across the four classes. In particular, in relation to pension wealth, net wealth and being in managerial and profession occupations, the results show that these are the most important factors that are associated with more diversification, yielding interesting insights into the drivers of portfolio diversification.

The *ex post* analysis reveals that our modelling approach, which moves beyond the naive measure of diversification based upon the number of financial assets, is consistent with household portfolio diversification. To be specific, examining class specific heterogeneity through ex post summary statistics for detailed sub-categories of different types of assets held in terms of rates of holding, monetary amounts and the ratio of asset value to total household financial assets, reveals a pattern of results, which is consistent with

24

portfolio diversification increasing across the classes.

Moreover, the statistical performance of estimators typically used in the literature compared to our latent class approach, shows that the approach we adopt strongly dominates with regard to the information criteria metrics. This suggests that treating the population as a single homogeneous group when analysing household financial behaviour may lead to biased parameter estimates and that policy based on such models could be inappropriate.

Finally, splitting the population into different groups based upon observed behaviour and characteristics to implement policy targeted at specific groups of interest, may introduce investigator bias due to preconceived notions about how to categorise different sub-groups. The *LCM* approach does not suffer from this as it is based upon latent heterogeneity. In future research, applying this type of framework more generally in the household finance literature could aid theoretical developments, as complex patterns may emerge that require novel explanations, as well as appropriate policy response.

References

Abreu, M. and B. Mendes (2010). Financial literacy and portfolio diversification. *Quantitative Finance 10*(5), 515–528.

Akaike, H. (1987). Information measures and model selection. *International Statistical Institute 44*, 277–291.

Bago d'Uva, T. and A. Jones (2009). Health care utilisation in Europe: New evidence from the ECHP. *Journal of Health Economics 28*(2), 265–279.

Barasinska, N., D. Schafer, and A. Stephan (2009). Financial risk aversion and household asset diversification. DIW Discussion Papers 807, German Institute for Economic Research (DIW Berlin).

Benartzi, S. and R. H. Thaler (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review 91*(1), 79–98.

Berkowitz, M. K. and J. Qiu (2006). A further look at household portfolio choice and health status. *Journal of Banking and Finance 30*(4), 1201–1217.

Bertaut, C. C. (1998). Stockholding behaviour of U.S. households: Evidence from the 1983-1989 Survey of Consumer Finances. *Review of Economics and Statistics 80*(2), 263–275.

Bhamra, H. S. and R. Uppal (2019). Does household finance matter? Small financial errors with large social costs. *American Economic Review 109*(3), 1116–1154.

Blume, M. E. and I. Friend (1975). The asset structure of individual portfolios and some implications for utility functions. *Journal of Finance 30*, 585–603.

Borgan, V. and A. Fertig (2013). Portfolio choice and mental health. *Review of Finance 17*(3), 932–955.

Bozdogan, H. (1987). Model selection and akaike's information criteria (AIC): The general theory and its analytical extensions. *Psychometrika 52*, 345–370.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy 115*(5), 707–747.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2009). Fight or flight? portfolio rebalancing by individual investors. *Quarterly Journal of Economics 124*(1), 301–348.

Cameron, C. and P. Trivedi (1998). *Regression Analysis of Count Data.* Econometric Society Monographs. Cambridge, U.K.: Cambridge University Press.

Campbell, J. Y. (2006). Household finance. *Journal of Finance 61*(4), 1553–1604.

Cardak, B. and R. Wilkins (2009). The determinants of household risky asset holdings: Australian evidence on background risk and other factors. *Journal of Banking and Finance 33*(5), 850–860.

Chung, H., J. C. Anthony, and J. L. Schafer (2011). Latent class profile analysis: An application to stage sequential processes in early onset drinking behaviours. *Journal of the Royal Statistical Society, Series A 174*, 689–712.

Clark, A., F. Etile, F. Postel-Vinay, C. Senik, and K. Van der Straeten (2005). Heterogeneity in reported well-being: Evidence from twelve European countries. *Economic Journal 115*, 118–132.

Deb, P. and P. K. Trivdei (1997). Demand for medical care by the elderly: A finite mixture approach. *Journal of Applied Econometrics 12*, 313–336.

Dimmock, S. G., R. Kouwenberg, O. S. Mitchell, and K. Peijnenburg (2016). Ambiguity aversion and household portfolio choice puzzles: Empirical evidence. *Journal of Financial Economics 119*(3), 559–577.

Durand, R., W. Greene, M. Harris, and J. Khoo (2022). Heterogeneity in speed of adjustment using finite mixture models. *Economic Modelling 107*, 105713.

Fan, E. and R. Zhao (2009). Health status and portfolio choice: Causality and heterogeneity? *Journal of Banking and Finance 33*(6), 1079–1088.

Florentsen, B., U. Nielsson, P. Raahauge, and J. Rangvid (2019). The aggregate cost of equity underdiversification. *Financial Review 54*(4), 833–856.

Gaudecker, H.-M. V. (2015). How does household portfolio diversification vary with financial literacy and financial advice? *Journal of Finance 70*(2), 489–507.

Gerhard, P., J. Gladstone, and H. A (2018). Psychological characteristics and household saving behavior: The importance of accounting for latent heterogeneity. *Journal of Economic Behavior and Organization 148*, 66–82.

Goetzmann, W. N. and A. Kumar (2008). Equity portfolio diversification. *Review of Finance 12*(3), 433–463.

Greene, W. (2018). *Econometric Analysis* (8 ed.). Pearson.

Grinblatt, M., M. Keloharju, and J. T. Linnainmaa (2012). IQ, trading behavior, and performance. *Journal of Financial Economics 104*(2), 339–362.

Guiso, L., T. Jappelli, and D. Terlizzese (1996). Income risk, borrowing constraints, and portfolio choice. *American Economic Review 86*(1), 158–172.

Guiso, L. and P. Sodini (2013). Household finance: An emerging field. In *Handbook of the Economics of Finance*, Volume 2, pp. 1397–1532. Elsevier.

Hannan, E. and B. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B 41*, 190–195.

Hochguertel, S., R. Alessie, and A. Soest (1997). Saving accounts versus stocks and bonds in household portfolio allocation. *Scandinavian Journal of Economics 99*(1), 81–97.

Ivković, Z., C. Sialm, and S. Weisbenner (2008). Portfolio concentration and the performance of individual investors. *Journal of Financial and Quantitative Analysis 43*(3), 613–655.

Kelly, M. (1995). All their eggs in one basket: Portfolio diversification of U.S. households. *Journal of Economic Behavior and Organization 27*(1), 87–96.

King, M. A. and J. I. Leape (1998). Wealth and portfolio composition: Theory and evidence. *Journal of Public Economics 69*(2), 155–193.

Liu, H. (2014). Solvency constraint, underdiversification, and idiosyncratic risks. *Journal of Financial and quantitative Analysis 49*(2), 409–430.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 7*(1), 77–91.

Matyas, L. and P. Sevestre (Eds.) (2008). *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice.* Springer.

McLachlan, G. and D. Peel (2000). *Finite Mixture Models.* Canada: Wiley.

ONS (2012). Wealth and assets survey review report. Technical report, see https://www.ons.gov.uk/file?uri=/economy/nationalaccounts/uksectoraccounts/ methodologies/wealthandassetssurveywas//wasreviewfinalreporttcm77322814.pdf.

Polkovnichenko, V. (2005). Household portfolio diversification: A case for rank-dependent preferences. *Review of Financial Studies 18*(4), 1467–1502.

Roche, H., S. Tompaidis, and C. Yang (2013). Why does junior put all his eggs in one basket? A potential rational explanation for holding concentrated portfolios. *Journal of Financial Economics 109*, 775–796.

Rosen, H. S. and S. Wu (2004). Portfolio choice and health status. *Journal of Financial Economics 72*(3), 457–484.

Schildberg-Horisch, H. (2018). Are risk preferences stable? *Journal of Economic Perspectives 32*(2), 135–154.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics 6(2)*, 461–464.

Shum, P. and M. Faig (2006). What explains household stock holdings? *Journal of Banking and Finance 30*(9), 2579–2597.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*, 307–334.

Table I: Summary statistics

	MEAN	S.D.	MIN	MAX
Dependent variable				
Number of financial assets	2.676	2.34	0	21
Common covariates				
Male	0.583	0.49	0	1
Class covariates				
Born 1945 to 1954	0.247	0.43	0	1
Born 1955 to 1965	0.284	0.45	0	1
Born 1965 onwards	0.260	0.44	0	1
Risk attitude	0.225	0.42	0	1
Time preference	0.297	0.46	0	1
Mother post school education	0.064	0.25	0	1
Father post school education	0.091	0.29	0	1
Mother employee or self employed	0.459	0.50	0	1
Father employee or self employed	0.700	0.50	0	1
Single parent family when growing up	0.102	0.30	0	1
Number of siblings when growing up	1.666	1.72	0	8
Outcome covariates				
Age	47.927	11.35	20	65
Single	0.485	0.50	0	1
Degree or above	0.297	0.46	0	1
Qualification below degree	0.550	0.50	0	1
Very good health	0.331	0.47	0	1
Employee	0.675	0.47	0	1
Managerial and professional	0.451	0.50	0	1
Intermediate occupation	0.106	0.31	0	1
Small employers and own account	0.085	0.28	0	1
Lower supervisory and technical	0.084	0.28	0	1
Log labour income	6.561	4.51	0	13.69
Log non-labour income	4.967	4.11	0	10.01
Log pension wealth	8.120	5.10	0	16.92
Log net wealth	8.466	6.16	-13.41	16.80
Has a DB occupational pension	0.250	0.43	0	1
Number of children in household	0.512	0.90	0	5
Number of adults in household	1.923	0.85	0	5
Financially optimistic	0.286	0.45	0	1
Finanically pessimistic	0.309	0.46	0	1

Heads of household (N)	28,320
Number of times observed in panel:	
1 wave	36%
2 waves	45%
3 waves	10%
4 waves	9%
Observations (NT)	45,578

Table II: Summary IC measures (pooled and panel data variants)

Panel A: pooled estimates

	BIC	AIC	CAIC	HQIC
linear (OLS)	188,270.20	188,069.51	188,260.49	188,127.96
Poisson	187,043.52	166,842.88	166,901.19	167,033.78
NEGBIN	166,147.64	165,946.91	166,170.64	166,010.06
2-class	160,918.59	160,421.14	160,975.59	160,577.64
3-class MNL	160,915.79	160,112.89	161,129.28	160,365.48
4-class MNL	**160,881.17**	**159,755.68**	**161,029.94**	**160,075.95**
5-class MNL	161,074.88	159,772.81	161,241.88	160,155.75

Panel B: panel estimates (random effects)

	BIC	AIC	CAIC	HQIC
linear	181,005.42	180,787.21	180,974.24.49	180,841.64
Poisson	162,945.20	162,735.85	162,924.79	162,792.38
NEGBIN	162,737.80	162,596.00	162,920.74	162,110.26
2-class	160,797.79	160,196.84	160,770.75	160,358.83
3-class MNL	160,898.94	159,968.71	161,008.17	160,115.34
4-class MNL	**160,771.75**	**159,265.62**	**161,007.79**	**158,180.84**
5-class MNL	161,004.08	159,617.44	161,101.08	158,396.53

Panel C: Vuong tests for panel estimates (random effects)

Vuong test BIC: MNL(4) *vs* MNL(2)	39.21
Vuong test BIC: MNL(4) *vs* MNL(3)	28.47
Vuong test BIC: MNL(4) *vs* MNL(5)	15.58
Vuong BIC corrected m_i^c: MNL(4) *vs* MNL(2)	12.01
Vuong BIC corrected m_i^c: MNL(4) *vs* MNL(3)	11.99
Vuong BIC corrected m_i^c: MNL(4) *vs* MNL(5)	11.98

Table III: Coefficients and incident rate ratios for number of financial assets by class expected values

	EV=1.61 Class 1			EV=2.71 Class 2			EV=2.73 Class 3			EV=4.06 Class 4		
	$\beta_{q=1}$		IRR	$\beta_{q=2}$		IRR	$\beta_{q=3}$		IRR	$\beta_{q=4}$		IRR
Male	-0.121	(0.03)***	0.886	0.059	(0.03)*	1.061	-0.083	(0.02)***	0.920	-0.047	(0.03)	0.954
Age	-0.366	(0.09)***	0.693	-0.087	(0.11)	0.917	-0.126	(0.09)	0.881	0.047	(0.10)	1.048
Age2	0.036	(0.01)***	1.036	0.012	(0.01)	1.012	0.010	(0.01)	1.010	-0.008	(0.01)	0.992
Single	0.165	(0.03)***	1.180	0.067	(0.03)**	1.069	0.108	(0.02)***	1.113	-0.034	(0.02)	0.966
Degree or above	0.300	(0.05)***	1.349	0.347	(0.08)***	1.415	0.221	(0.03)***	1.247	0.140	(0.04)***	1.150
Qualification below degree	0.213	(0.04)***	1.238	0.252	(0.08)***	1.286	0.125	(0.03)***	1.133	0.092	(0.04)**	1.097
Very good health	-0.005	(0.03)	0.995	0.015	(0.03)	1.015	-0.007	(0.02)	0.993	-0.030	(0.02)	0.970
Employee	0.088	(0.04)**	1.091	-0.063	(0.05)	0.939	-0.036	(0.03)	0.964	0.038	(0.03)	1.039
Managerial & profession	0.250	(0.04)***	1.285	0.407	(0.05)***	1.503	0.136	(0.03)***	1.146	0.085	(0.03)***	1.088
Intermediate occupation	0.146	(0.04)***	1.157	0.257	(0.07)***	1.294	0.117	(0.03)***	1.124	0.037	(0.05)	1.037
Small employers & own account	0.150	(0.05)***	1.162	0.166	(0.07)**	1.180	0.087	(0.04)**	1.090	0.010	(0.04)	1.010
Lower supervisory & technical	0.113	(0.05)**	1.119	0.116	(0.08)	1.122	0.008	(0.04)	1.008	0.063	(0.04)	1.065
Log labour income	0.057	(0.04)	1.058	0.121	(0.05)**	1.129	-0.100	(0.03)***	0.905	-0.007	(0.03)	0.993
Log non-labour income	0.039	(0.03)	1.040	-0.017	(0.05)	0.983	-0.004	(0.02)	0.996	0.010	(0.03)	1.010
Log pension wealth	0.443	(0.03) ***	1.558	0.498	(0.04)***	1.646	0.182	(0.02)***	1.200	0.145	(0.03)***	1.156
Log net wealth	0.018	(0.00) ***	1.019	0.014	(0.00)***	1.014	0.162	(0.01)***	1.176	0.218	(0.01)***	1.243
Has a DB occ. pension	-0.162	(0.03)***	0.850	-0.253	(0.04)***	0.776	-0.050	(0.02)**	0.951	-0.022	(0.02)	0.979
Number of children	-0.078	(0.02)***	0.925	-0.051	(0.02)***	0.951	-0.058	(0.01)***	0.944	-0.017	(0.01)	0.983
Number of adults	0.062	(0.02)***	1.064	0.130	(0.02)***	1.139	0.012	(0.01)	1.012	0.015	(0.01)	1.022
Financially optimistic	0.107	(0.03)***	1.113	0.076	(0.04)**	1.079	-0.001	(0.02)	0.999	0.014	(0.02)	1.014
Financally pessimistic	-0.033	(0.03)	0.968	0.032	(0.04)	1.032	-0.028	(0.02)	0.973	-0.011	(0.02)	0.989
ρ	0.160	(0.03)***		0.159	(0.02)***		0.043	(0.04)		0.050	(0.16)	
$\bar{\rho}$	0.216	(0.06)***										

Notes: Observations (N)=28,320; (NT)=45,578. EV denotes ex-post expected value, i.e. the number of financial assets. The incidence rate ratio is given by $IRR = \exp(\beta_q)$. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table IV: Partial effects on prior class probabilities by class expected values

	EV=1.61 Class 1		EV=2.71 Class 2		EV=2.73 Class 3		EV=4.06 Class 4	
Male	-0.015	(0.01)	-0.191	(0.05)***	0.008	(0.01)	0.198	(0.05)***
Born 1945 to 1954	0.026	(0.02)*	-0.107	(0.04)***	0.036	(0.02)**	0.044	(0.05)
Born 1955 to 1965	-0.059	(0.02)***	0.084	(0.05)**	-0.030	(0.02)**	0.005	(0.04)
Born 1965 onwards	0.130	(0.04)***	-0.018	(0.04)	-0.044	(0.02)**	-0.068	(0.03)**
Risk attitude	-0.044	(0.02)***	-0.214	(0.07)***	-0.010	(0.02)	0.267	(0.05)***
Time preference	-0.029	(0.01)***	-0.029	(0.04)	0.027	(0.02)	0.032	(0.04)
Mother post school education	0.017	(0.01)	0.097	(0.05)**	-0.069	(0.03)***	-0.046	(0.05)
Father post school education	-0.001	(0.01)	-0.118	(0.04)***	0.145	(0.05)***	-0.026	(0.05)
Mother employee/self employed	-0.064	(0.02)***	-0.067	(0.06)	-0.055	(0.02)**	0.186	(0.06)***
Father employee/self employed	-0.001	(0.00)	-0.088	(0.04)**	-0.024	(0.01)**	0.112	(0.05)***
Single parent family growing up	0.021	(0.01)**	0.052	(0.04)	0.015	(0.01)	-0.088	(0.05)*
Number of siblings growing up	-0.014	(0.01)	-0.169	(0.03)***	0.018	(0.02)	0.164	(0.04)***
Posterior probabilities	0.3849		0.1820		0.3194		0.1138	

Notes: Observations (N)=28,320; (NT)=45,578. EV denotes ex-post expected value, i.e. the number of financial assets. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table V: Comparison of overall partial effects across models

	linear (RE)		NegBin		4-class MNL	
Male	0.064	(0.02)***	0.037	(0.01)***	-0.019	(0.05)
Born 1945 to 1954					-0.042	(0.02)**
Born 1955 to 1965					0.055	(0.02)***
Born 1965 onwards					-0.049	(0.01)***
Risk attitude					-0.008	(0.02)
Time preference					0.051	(0.03)*
Mother post school education					0.029	(0.02)
Father post school education					-0.118	(0.03)***
Mother employee/self employed					-0.025	(0.02)
Father employee/self employed					0.069	(0.02)***
Single parent family growing up					-0.015	(0.02)
Number of siblings growing up					-0.008	(0.02)
Age	-0.028	(0.01)***	-0.451	(0.06)***	-0.068	(0.15)
Age2	0.001	(0.00)***	0.049	(0.01)***	0.007	(0.02)
Single	0.302	(0.02)***	0.246	(0.02)***	0.051	(0.04)
Degree or above	0.895	(0.03)***	0.711	(0.03)***	0.551	(0.09)***
Qual. below degree	0.357	(0.03)***	0.411	(0.03)***	0.384	(0.09)***
Very good health	0.038	(0.02)*	0.018	(0.02)	-0.021	(0.04)
Employee	-0.113	(0.03)***	-0.045	(0.02)*	-0.019	(0.06)
Man. & prof. occ.	0.543	(0.06)***	0.621	(0.02)***	0.532	(0.06)***
Intermediate occ.	0.162	(0.06)**	0.430	(0.03)***	0.321	(0.08)***
Small employers occ.	0.077	(0.07)	0.385	(0.03)***	0.195	(0.09)**
Lower supervisory occ.	-0.085	(0.07)	0.189	(0.03)***	0.194	(0.09)**
Log labour income	0.090	(0.03)***	-0.018	(0.02)	0.099	(0.06)
Log non-labour income	-0.183	(0.03)***	-0.061	(0.02)***	-0.003	(0.05)
Log pension wealth	0.860	(0.02)***	0.846	(0.02)***	0.707	(0.06)***
Log net wealth	0.094	(0.01)***	0.159	(0.00)***	0.291	(0.02)***
Has DB occ. pension	-0.445	(0.02)***	-0.338	(0.02)***	-0.290	(0.05)***
Number of children	-0.028	(0.01)***	-0.127	(0.01)***	-0.082	(0.02)***
Number of adults	0.163	(0.01)***	0.095	(0.01)***	0.158	(0.03)***
Financially optimistic	0.121	(0.02)***	0.100	(0.02)***	0.096	(0.04)**
Financally pessimistic	-0.055	(0.02)***	-0.054	(0.02)***	0.012	(0.05)
AIC		180,787.21		162,596.00		159,265.62
BIC		181,005.42		162,737.80		160,797.79

Notes: Observations (N)=28,320; (NT)=45,578. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table VI: Ex-post summary statistics - percentage held

SUMMARY INFORMATION	Mean	EV=1.61 Class 1	EV=2.71 Class 2	EV=2.73 Class 3	EV=4.06 Class 4
% no financial assets	**15.31%**	21.78%	5.13%	15.13%	2.28%
Mean number of financial assets	**2.68**	1.28	2.70	3.01	6.91
Standard deviation number of financial assets	**2.34**	0.98	1.70	2.34	3.68
Minimum number of financial assets	**0**	0	0	0	0
Maximum number of financial assets	**21**	7	12	13	21
Herfindahl-Hirschman index	**0.18**	0.30	0.21	0.12	0.08

| TYPE OF FINANCIAL ASSET	Mean	EV=1.61 Class 1	EV=2.71 Class 2	EV=2.73 Class 3	EV=4.06 Class 4
A) Has saving or deposital account					
A1) Has a savings or deposit account	**56.42%**	41.31%	69.54%	59.97%	83.71%
A2) Has national savings easy access	**1.28%**	0.44%	1.27%	1.33%	4.81%
A3) Has an all-in one or offset account	**0.62%**	0.34%	0.85%	0.62%	1.72%
A4) Has funds saved with credit union	**0.36%**	0.15%	0.48%	0.39%	0.93%
A5) Has other savings or deposit account	**0.45%**	0.10%	0.41%	0.49%	1.66%
A6) Has overseas savings or deposits	**1.49%**	0.61%	2.30%	1.63%	3.35%
B) Has ISA account					
B1) Has a cash ISA	**45.04%**	24.59%	51.40%	53.41%	77.55%
B2) Has an investment ISA	**15.18%**	2.29%	11.65%	18.64%	57.43%
C) Has fixed-term investment bonds					
C1) Has fixed interest fixed term bond	**7.81%**	1.13%	3.97%	10.21%	29.07%
C2) Has variable return no capital guarantee	**2.00%**	0.21%	0.93%	2.46%	8.86%
C3) Has variable return capital guarantee	**1.47%**	0.18%	0.77%	1.77%	6.47%
D) Has unit investment bonds	**6.98%**	0.69%	3.92%	8.02%	33.34%
E) Has employee share or share options					
E1) Has employee shares	**6.54%**	1.65%	9.38%	6.71%	23.13%
E2) Has share options	**3.07%**	0.69%	5.27%	2.83%	11.90%
E3) Has both employee & share options	**1.29%**	0.08%	2.24%	0.95%	7.29%
F) Has other shares					
F1) Has shares in UK companies (listed & unlisted)	**15.52%**	1.06%	10.34%	18.83%	68.01%
F2) Has shares in listed UK companies	**15.04%**	0.98%	9.67%	18.19%	66.88%
F3) Has shares in unlisted UK companies	**1.33%**	0.09%	1.06%	1.29%	7.51%
F4) Has shares in foreign companies	**2.40%**	0.17%	1.97%	2.28%	13.62%
G) Has premium national savings bonds certificates					
G1) Has index-linked/fixed investment savings certificates	**1.86%**	0.07%	0.85%	1.71%	12.16%
G2) Has premium bonds	**23.02%**	7.34%	21.32%	28.66%	60.83%
G3) Has pensioners guaranteed income bonds	**0.74%**	0.09%	0.35%	0.82%	3.66%
G4) Has other national savings products	**0.62%**	0.11%	0.64%	0.71%	2.39%
H) Has government/corporate bonds & gilts					
H1) Has corporate bonds issued by a UK company	**0.75%**	0.06%	0.41%	0.66%	4.92%
H2) Has UK government local authority bonds or gilts	**0.67%**	0.03%	0.19%	0.62%	4.50%
H3) Has corporate bonds issued by a foreign company	**0.10%**	0.00%	0.04%	0.06%	0.84%
H4) Has government bonds issued by a foreign government	**0.17%**	0.01%	0.14%	0.10%	1.35%
I) Has life insurance, friendly society or endowment policies					
I1) Has an endowment or regular premium policy	**5.78%**	2.07%	4.42%	6.53%	19.89%
I2) Has a single premium, policy or investment bond	**0.87%**	0.23%	0.41%	0.96%	3.85%
I3) Has a friendly society tax exempt savings plan	**0.81%**	0.15%	0.48%	0.78%	4.39%
I4) Has an insurance policy which pays lump sum	**2.27%**	1.13%	1.97%	2.47%	6.44%
I5) Has other life insurance product	**16.21%**	11.40%	21.90%	16.69%	26.45%
J) Has other investments	**1.60%**	0.35%	1.22%	1.81%	6.47%

Notes: Observations (N)=28,320; (NT)=45,578. EV denotes ex-post expected value, i.e. the number of financial assets.

Table VII: Ex-post summary statistics - amounts & proportion of total assets

	Mean	EV=1.61 Class 1	EV=2.71 Class 2	EV=2.73 Class 3	EV=4.06 Class 4
TOTAL AMOUNT OF FINANCIAL ASSETS	**£58,174**	£11,548	£44,548	£70,036	£215,857
A) Savings Accounts					
Amount	**£13,979**	£4,312	£13,688	£16,664	£41,646
% of total amount of financial assets	**31.08%**	34.48%	38.28%	28.21%	22.39%
B) National Savings Accounts					
Amount	**£2,196**	£273	£1,078	£2,476	£10,745
% of total amount of financial assets	**4.24%**	3.53%	4.95%	4.45%	5.08%
C) ISA					
Amount	**£12,373**	£2,395	£7,808	£15,486	£45,114
% of total amount of financial assets	**24.10%**	17.83%	26.96%	27.31%	28.98%
D) Fixed Term Investment Bonds					
Amount	**£6,062**	£669	£3,168	£7,816	£23,971
% of total amount of financial assets	**3.62%**	1.05%	2.16%	5.16%	8.16%
E) Unit Trusts					
Amount	**£4,519**	£419	£3,143	£5,254	£20,462
% of total amount of financial assets	**1.84%**	0.41%	1.18%	2.38%	6.00%
F) Employee Shares & Share Options					
Amount	**£3,681**	£677	£2,937	£5,022	£10,236
% of total amount of financial assets	**2.66%**	1.42%	5.15%	2.60%	4.99%
G) Shares					
Amount	**£7,186**	£492	£5,846	£8,229	£32,828
% of total amount of financial assets	**3.26%**	0.50%	3.22%	4.24%	9.92%
H) Bonds and Gilts					
Amount	**£728**	£23	£609	£585	£4,889
% of total amount of financial assets	**0.23%**	0.04%	0.18%	0.28%	0.88%
I) Insurance Products					
Amount	**£3,090**	£1,190	£2,878	£3,514	£9,358
% of total amount of financial assets	**2.83%**	1.87%	2.75%	3.12%	5.51%
J) Endowment or Regular Premium Policy					
Amount	**£1,844**	£687	£1,772	£2,141	£5,351
% of total amount of financial assets	**1.54%**	0.84%	1.36%	1.82%	3.27%
K) Single Premium Policy					
Amount	**£391**	£148	£140	£426	£1,641
% of total amount of financial assets	**0.16%**	0.08%	0.08%	0.19%	0.42%
L) Other Investment					
Amount	**£2,126**	£265	£1,482	£2,423	£9,616
% of total amount of financial assets	**0.62%**	0.25%	0.59%	0.75%	1.62%

Notes: Observations (N)=28,320; (NT)=45,578. EV denotes ex-post expected value, i.e. the number of financial assets.

Figure 1: Number of financial assets
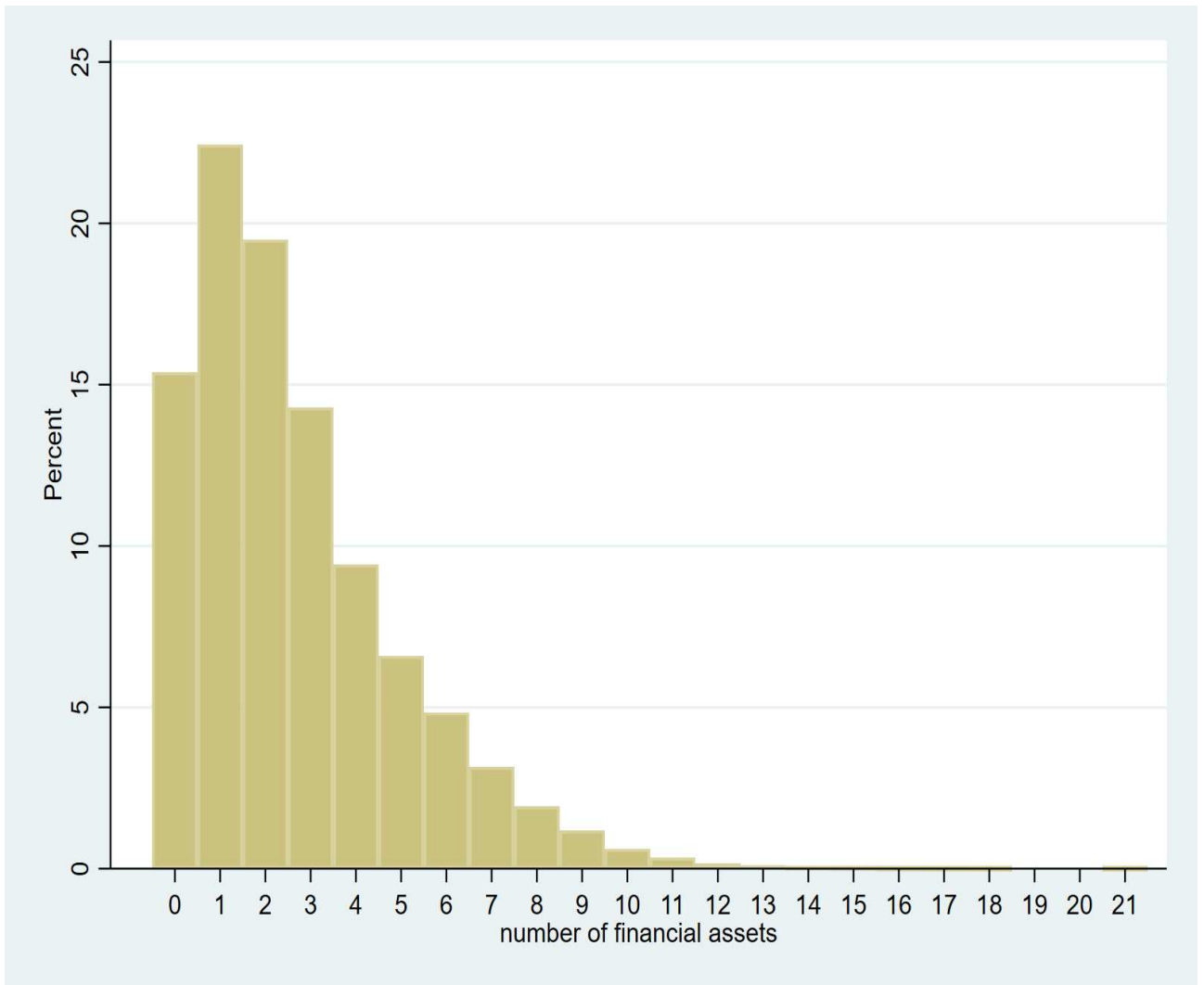
A. Appendix

Table A1: Partial effects by class expected values

	EV=1.61 Class 1		EV=2.71 Class 2		EV=2.73 Class 3		EV=4.06 Class 4	
Male	-0.179	(0.04)***	0.143	(0.08)*	-0.157	(0.05)***	-0.113	(0.08)
Age	-0.540	(0.14)***	-0.209	(0.27)	-0.238	(0.16)	0.115	(0.47)
Age2	0.053	(0.02)***	0.029	(0.03)	0.019	(0.02)	-0.018	(0.74)
Single	0.244	(0.04)***	0.160	(0.08)**	0.203	(0.04)***	-0.083	(0.52)
Degree or above	0.442	(0.07)***	0.833	(0.19)***	0.416	(0.06)***	0.338	(0.10)***
Qualification below degree	0.315	(0.06)***	0.604	(0.18)***	0.235	(0.06)***	0.224	(0.09)***
Very good health	-0.008	(0.04)	0.036	(0.07)	-0.014	(0.03)	-0.073	(0.05)
Employee	0.129	(0.06)**	-0.151	(0.13)	-0.069	(0.05)	0.093	(0.07)
Managerial & profession	0.369	(0.05)***	0.978	(0.12)***	0.257	(0.05)***	0.205	(0.07)***
Intermediate occupation	0.215	(0.07)***	0.618	(0.16)***	0.221	(0.06)***	0.089	(0.11)
Small employers & own account	0.221	(0.08)***	0.397	(0.18)**	0.163	(0.07)**	0.024	(0.10)
Lower supervisory & technical	0.166	(0.07)**	0.277	(0.20)	0.015	(0.07)	0.153	(0.10)
Log labour income	0.083	(0.06)	0.291	(0.13)**	-0.188	(0.05)***	-0.018	(0.07)
Log non-labour income	0.058	(0.05)	-0.041	(0.12)	-0.008	(0.05)	0.025	(0.06)
Log pension wealth	0.654	(0.05)***	1.196	(0.10)***	0.344	(0.05)***	0.351	(0.07)***
Log net wealth	0.027	(0.01)***	0.034	(0.01)***	0.306	(0.11)***	0.528	(0.11)***
Has a DB occ. pension	-0.239	(0.04)***	-0.608	(0.09)***	-0.094	(0.04)**	-0.053	(0.05)
Number of children	-0.116	(0.02)***	-0.121	(0.05)***	-0.109	(0.02)***	-0.041	(0.04)
Number of adults	0.092	(0.02)***	0.312	(0.05)***	0.023	(0.02)	0.052	(0.03)
Financially optimistic	0.157	(0.04)***	0.182	(0.09)**	-0.002	(0.04)	0.034	(0.05)
Financally pessimistic	-0.049	(0.04)	0.076	(0.10)	-0.052	(0.04)	-0.027	(0.06)

Notes: Observations (N)=28,320; (NT)=45,578. EV denotes ex-post expected value, i.e. the number of financial assets. Standard errors are shown in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.