

# Understanding the role of savings in building longer-term financial security: appendices

## Appendix 2: quantitative research methodology

We conducted new analysis of The UK Household Longitudinal Study, more commonly known as ‘Understanding Society’.<sup>1</sup> This is a large-scale, internationally recognised study, which provides vital evidence for scientists and policymakers on the causes and consequences of deep-rooted social problems. It is a longitudinal study, which has been tracking households annually since 2009. It covers the whole population, with boost samples to ensure it is representative of immigrant and ethnic minority groups, and its large sample enables sub-population groups to be examined.

The survey consists of several questionnaires and questionnaire modules, some of which are answered by a lead respondent (or ‘household reference person’) on behalf of the entire household and some of which are answered by each individual adult in the household.

Due to availability of the key variables that we were interested in, we conducted our analyses using different combinations of three waves of data:

- Wave 4: 2012/13
- Wave 8: 2016/17
- Wave 13: 2021/22

These were the only waves in which participants were asked detailed questions about the different types of savings and investment products that they hold and how much money they have within each type of account.

### Our research questions

Our research questions were as follows:

1. What level of savings helps to protect against financial difficulties, e.g. the risk of falling behind with payments or becoming over-indebted?
2. To what extent do basic forms of saving lead to use of other financial products or investments in future, thereby increasing longer-term financial security?

### Sample

We looked at a sample of people aged 18 to 65 in 2012 (aged 27 to 74 by 2021/22) and tracked them over time at three time points: 2012/13, 2016/17 and 2021/22. We followed just

---

<sup>1</sup> University of Essex, Institute for Social and Economic Research. (2023). Understanding Society: Waves 1-13, 2009-2022 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 18th Edition. UK Data Service. SN: 6614, DOI: <http://doi.org/10.5255/UKDA-SN-6614-19>

over 7,000 people across the three survey waves; of these, we had information on the amount their household was saving for a sample of just under 5,000.

We used weights provided by Understanding Society which are designed to correct for the probability of a household being invited to take part in the study in the first place, and then which correct for their likelihood of dropping out of the study over time (following the first wave). This helps to make the results more representative of the UK population. We used the *indinus\_lw* weighting variable, which is a longitudinal weight. All longitudinal weights were available only at the individual-level, as households are changeable over time (e.g. due to household members entering and leaving) and therefore cannot be tracked using a single weight over time.

## **Analytic approach**

StataNow/MP 19.5 was used for all analyses.

The Understanding Society data is provided in several different datasets, covering different waves and different questionnaire modules (for example, individual-level and household-level questionnaires). We began by merging these together into a single, long file format, where each row represents an individual adult respondent at a specific time point. For example, if someone responded to all three waves of interest then they would have three rows' worth of data. Where variables were already provided at the household-level, each individual in the household assumes the same household-level data (using a unique household ID for that household in a given wave). Where variables had been provided at the individual-level but were required at household-level, we derived new variables by aggregating up across all adult members of the household. Having derived all necessary variables, we then restructured the data from long format to wide format. This means that each row would represent one adult respondent, but containing data from all of the waves that they took part in; for example, one variable for their savings in 2012/13, one for 2016/17 and one for 2021/22.

Having filtered our sample to those within the appropriate age bracket and who had responded to all three waves, we performed a range of analyses. For each research question, we began by conducting a series of preparatory and formative analyses.

For RQ1, this included establishing which main outcome variables to use in our analyses, as we had identified a range of possible candidate variables. We conducted cross-tabulations between these variables, as well as correlation matrices and item response theory, to understand which variables captured unique aspects of financial difficulty. This led to us identifying three main elements of financial difficulty to be measured: being behind with bills, high levels of borrowing and poor subjective financial wellbeing. The two former elements were derived from a combination of other indicators of financial difficulty, as we outline in the next section.

For RQ1, we also conducted a series of cross-sectional analyses to understand how to categorise our 'total savings' variable, i.e. where to draw the different breaks between savings categories. This included concurrent regression analyses, where we explored the link between a household's *current* level of savings (rather than *previous*) and their current financial difficulty.

Our main analyses, however, focused on a longitudinal picture, exploring the relationship between previous levels of savings and subsequent likelihood of financial difficulty. This

included cross-tabulations and regression analyses, which allowed us to also control for other factors (including socio-demographics and prior financial difficulties). The regression analyses were binary logistic regressions, with the unit of analysis being individual adults. We used cluster robust standard errors to adjust for clustering caused by having multiple respondents per household. As previously stated, analyses were weighted using longitudinal weights.

We ran a series of regression models, each building on the last. Model 0 contained no other covariates, only our main independent variable (level of savings) and our dependent variable (financial difficulty). Model 1 added prior financial difficulty (from the previous two waves). Model 2 added household income quintile and respondent age group. We then ran a model with all variables shown in the 'Control variables' table below included, but then limited this to only significant ones (which we report as Model 3). For these final models, in addition to the odds ratios produced by the regression, we also used Stata's *margins* command to produce adjusted probabilities (or 'risks') of financial difficulties at different levels of savings (if all other variables in the regression were set at their 'average' level).

To explore whether results varied for different age or income groups, we conducted a series of regressions with sub-samples. For age, we had four different age bandings: 27-34 (in the 2021/22 wave); 35-44; 45-59; and 60-74. Due to the smaller sample sizes (ranging from n=371 for the 27-34 group to n=3107 for the 60-74 group) and the corresponding risk of over-specifying the regressions, we added fewer other variables to the models. In these analyses, we controlled only for previous financial difficulty and household income quintile (in addition to amount in savings in the previous wave). For income, we did the same but the sub-sample was based on income quintile and we controlled for age (not income).

As noted, for RQ1 we were primarily interested in total amount of savings held, but we also conducted analysis exploring if regularity of saving (or amount regularly saved) was differently associated with each form of financial difficulty. In addition, we conducted sensitivity analysis to check if the results obtained were likely to have been affected by the coronavirus pandemic and associated lockdown in 2020-21 (as our final wave of data was collected in 2021-22). This broadly gave comparable results to our main analyses.

For RQ2, we conducted formative analyses to understand how best to group the different savings product variables available in the data. In other words, what products are typically held together or as substitutes for one another? Using cross-tabulations, Principle Components Analysis (PCA) and Polychoric PCA, we tested an initial conceptualisation that we had developed and, as a result of the analyses, revised this slightly – in line with the 'RQ2 – main variables of interest' table below. Having re-conceptualised our savings products types, we conducted regression analysis to establish the extent to which holding each product type in 2021/22 was associated with having high subjective financial wellbeing ('living comfortably') at the same time.

Separately, we then conducted a series of regression analyses to explore the extent to which previously holding each product type (in 2012/13 and/or 2016/17) was associated with subsequently holding each other product type in 2021/22. In these analyses, we controlled for age cohort and income quintile. The outcome variable in each case was whether or not the respondent held a given product type (for example, National Savings products) in 2021/22. We conducted two sets of these regression models, one which included all respondents regardless of whether or not they had previously held the outcome product type in 2012/13 or 2016/17 and one which only included those who had not previously held the outcome product. In other words, the latter was predicting whether or not someone was starting to hold a given product type for the first time. The latter group of models had lower

sample sizes, depending on how common it was for each product to be held. For example, the full model with homeownership as the outcome variable had a sample size of 7,512; whereas the model which removed previous homeowners had just 1,364. This was the biggest decline in sample size as homeownership was the most common ‘product’ previously held.

We also sought to understand whether basic savings products lead to holding a more diverse portfolio of savings and investment products in future. In other words, we wanted to look not just at individual products held by someone, but at the combination of different products that they hold. Through a series of analyses (including latent class analysis), we created a segmentation of people based on their product-holding in 2021/22. We excluded basic savings products from the creation of this segmentation—as we were interested to see to what extent earlier basic savings would lead to being in different segments later on in someone’s life—but we include all of the other product categories previously looked at, with the exception of rental income (as this was very uncommon).

Similar to when using cluster analysis with continuous variables, latent class analysis (LCA) is a technique for identifying sub-groups within a sample based, in LCA, on the patterns of their answers to categorical questions (here, all being binary variables). In our case, this meant grouping people based on their product-holding (or not) in 2021/22. Through iterative analyses, the LCA revealed five optimal, broad groups, each with different probabilities of holding a given product type. In essence, these groups ran from least diverse to most diverse product-holding (as outlined in the main report). We used the LCA results to then manually assign people to groups. These groups were then used in further regression analyses and cross-tabulations to explore whether prior product-holding (e.g. holding basic savings) was associated with greater or lower likelihood of later being in each of the identified segments, and to understand the socio-demographic profile of each group.

## Variables used to answer the research questions

### Research question 1

Main variable of interest	
Variable	Descriptive / derivation
Total amount in savings	<p>Each adult in the household is asked about which type of savings products they hold, out of the following:</p> <ul style="list-style-type: none"> <li>- Savings or deposit accounts (with a bank, post office or building society)</li> <li>- National savings accounts (national savings bank or post office)</li> <li>- ISA – cash only</li> <li>- ISA – stocks and shares or PEPs</li> <li>- Premium bonds</li> <li>- Other types of savings account</li> </ul> <p>These questions (e.g. <a href="#">svacts1</a>) then lead participants to be asked how much they hold in each account type (e.g. <a href="#">svamt1</a>). It should be noted that saving within one’s current account was only asked about in wave 13 and therefore is not directly included within the earlier waves.</p> <p>From this, we are able to derive how much each individual holds in total within their savings. We then aggregate this to the household-level, giving the total amount in savings across all individuals in the household. We do this because the outcome variables we are interested in are primarily captured at the household-level and because we would generally expect that access to savings will generally be shared across individuals within the household.</p>

	<p>Where any of the individuals within the household either refused to respond to one of these questions or didn't know the answer, we also set the household-level variable as 'missing' (even if another individual had responded).</p> <p>Regardless of which survey wave the data was from, we convert the amounts into 2022 prices. We do this by identifying the specific year that the questionnaire was completed (as each wave takes place over two calendar years) and then using the <a href="#">Consumer Price Index</a> to convert into 2022 prices.</p> <p>The initial derived variable developed is a continuous one; however, for the purposes of our analyses we derived various categorical versions of this. During our formative stage of the analysis, we trialled various sensitivity analyses to determine the most theoretically and empirically-backed breakpoints for our savings categories. This led us to settle upon a final categorisation of: £0-199, £200-499, £500-999, £1,000-1,999, £2,000-4,999, £5,000-9,999, £10,000-49,999, and £50,000+. Those with missing savings data were also entered into our regression analyses under a separate 'missing' category. As well as a banded categorical variable, we also conducted some models using a series of binary 'threshold' variables, indicating whether the household had achieved a certain threshold of savings or not, e.g. more than £2,000 in savings.</p> <p>We added this total savings variable into our regression analyses in a variety of ways – to account for the longitudinal nature of the data. As we had two 'prior' values for this variable (from 2012/13 and 2016/17), we ran a range of combinations in our models. This included versions with only the 2012/13 total savings variable added; ones with only the 2016/17 variable; ones where we took the 'best/highest' value from the two years; and ones where we used the lowest savings category that they had held in 'both' years (for example, if they held £500 in one year and £1000 in another, they would assume the value of £500).</p>
<b>Other variables of interest</b>	
<b>Variable</b>	<b>Descriptive / derivation</b>
Number of months' income held in savings	As mentioned above, in the process of developing the 'total amount in savings' variable, we produced a continuous variable, giving the 'exact' amount the respondent reported they held in savings. We used this variable (at the household-level) to calculate how many months' income they had access to in savings. We simply divided the amount in savings by their monthly income and then used Stata's <i>floor</i> command to round down to the nearest integer. For example, if the calculation showed they had 4.6 months' income in savings, this would be rounded down to 4. We then used the recode function to create a categorical version with the following bands: less than 1 month's income; 1-2 months; 3-6 months; 7-12 months; more than a year's income.
Regular savings habit	<p>To explore not just whether the total amount in savings matters but also whether the savings habit matters, we look at the <a href="#">savreg</a> variable: respondent answers to the question 'Do you save on a regular basis or just from time to time when you can?'. We create a binary variable with 1 indicating they answered 'Regular basis' to the aforementioned question, while 0 indicates they either answered 'From time to time' or 'Other' or had previously said they do not save any money.</p> <p>Additionally, we consider the amount regularly saved, using a banded version of <a href="#">saved</a>, in 2022 prices.</p>
<b>Key outcome variables</b>	
<b>Variable</b>	<b>Descriptive / derivation</b>
Behind with bills?	<p>This variable indicates if a household was behind with any bills at the time of the survey or had been behind with their council tax or rent/mortgage in the last 12 months before the survey. It is derived from three variables:</p> <ul style="list-style-type: none"> <li>- Currently behind with any bills: <a href="#">xphsdba</a></li> <li>- Behind with council tax in past 12 months: <a href="#">xphsdct</a></li> <li>- Behind with rent/mortgage in past 12 months: <a href="#">xphsdb</a></li> </ul>

	<p>The derived variable equals 1 if any of the above variables indicate financial difficulty. It equals 0 if all of the above variables indicate no financial difficulty. The variable is set to missing if any of the above are missing, unless one of the other variables indicates financial difficulty.</p>
High level of borrowing?	<p>This variable indicates if anyone in the household has three or more types of borrowing (excluding mortgages or student loans) or if they have any current borrowing from private individuals (e.g. friends and family).</p> <p>The variable is derived from a set of 'debt' variables, asked at the individual-level, which show if the respondent currently owes money on:</p> <ul style="list-style-type: none"> <li>- Hire purchase agreements: <a href="#">debt1</a></li> <li>- Personal loan(s) from a bank, building society or other financial institution: <a href="#">debt2</a></li> <li>- Catalogue or mail order purchase agreement: <a href="#">debt4</a></li> <li>- DWP/SSA social fund loan: <a href="#">debt5</a></li> <li>- Any other loan(s) from a private individual: <a href="#">debt6</a></li> <li>- Overdrafts: <a href="#">debt7</a></li> <li>- Other debt not listed here: <a href="#">debt97</a></li> </ul> <p>We begin therefore by counting the number of debt types held (out of all of the above except for loans from a private individual) and identifying those with three or more. We then create a household-level version of this variable, indicating if anyone in the household either has three or more debts OR has at least one loan from a private individual (even if they do not have three or more debts). In doing so, we aim to capture both those who have been forced to take out multiple loans and also those who have turned to lending from private individuals (such as friends and family) because they are unable to access other more formal types of lending.</p>
Poor subjective financial wellbeing	<p>This variable indicates if anyone in the household, when asked 'How well would you say you yourself are managing financially these days? Would you say you are...', answered 'Finding it quite difficult' or 'Finding it very difficult'. It is derived through a simple recoding of <a href="#">finnow</a> into a binary, and then the creation of a household-level variable, indicating if anyone in the household suggested difficulty.</p>

## Research question 2

Main variables of interest	
Variable	Descriptive / derivation
Holds basic savings products	<p>Aggregated household-level version of <a href="#">svacts1</a> – whether or not they hold any savings or deposit accounts (with a bank, post office or building society).</p> <p>For this and each of the below variables, we then constructed a longitudinal version of the variable, which described whether the product had been held in neither of the previous waves (2012/13 or 2016/17), either of them, or both of them. This was the variable that we entered into the regression models.</p>
Holds NS&I products	<p>Aggregated household-level version of <a href="#">svacts2</a> (National Savings accounts), <a href="#">svacts5</a> (premium bonds) or <a href="#">nvestrt1</a> (National Savings Certificates or Bonds). The latter was already provided at the household-level.</p>
Holds Cash ISA(s)	<p>Aggregated household-level version of <a href="#">svacts3</a> (ISA – cash only).</p>
Holds other savings	<p>Aggregated household-level version of <a href="#">svacts97</a> (other types of savings account).</p>
Holds other investment products	<p>Aggregated household-level version of <a href="#">svacts4</a> (ISA – stocks and shares or PEPs), plus <a href="#">nvestrt3</a> (Company stocks or shares), <a href="#">nvestrt2</a> (Unit Trusts / Investment Trusts) and <a href="#">nvestrt97</a> (other investments, e.g. gilts, government or company bonds or securities, stock).</p>

Holds a private pension	This includes those who are either paying into a private pension or who are currently receiving income from a private pension. This is based on <a href="#">ppen</a> (whether currently a member of any persona pension scheme or currently contributing to a personal pension scheme) and <a href="#">benpen5</a> (whether currently receives any income from a private pension or annuity). Please note that our analysis focuses only on those aged 18 to 65 in 2012 (or 27 to 74 in 2021).
Homeowner	Derived from <a href="#">tenure_dv</a> , with those who own their home outright or as a mortgagor classified as homeowners.
Receives rental income	Derived from <a href="#">bensta7</a> (receives income in the form of rent on any other property (even if it only covers the mortgage)).

### *Control variables used for either research question*

Control variables	
Variable	Descriptive / derivation
Previous financial difficulty	For research question 1, we control for previous financial difficulty. This always uses the same variable as the outcome measure, but for the previous survey waves. For example, if our outcome measure is whether they had fallen behind with bills in 2021/22, we control for them previously having fallen behind with bills or not in 2012/13 and/or 2016/17. In each case, we assign a value of 1 to those who had fallen behind in at least one of the two previous waves.
Age group (in 2021/22)	Derived from <a href="#">age_dv</a> in 2021/22 wave. Age grouped as follows: 27-34; 35-44; 54-59; 60-74.
Sex	Binary version of <a href="#">sex_dv</a> , where 0 equals 'female' and 1 equals 'male'. Missing values and values that were inconsistent over time are excluded from the analysis.
Ethnic group	Derived from <a href="#">racel_dv</a> . Ethnic group grouped as follows: White British; Other White; Asian or Arab; Black; Mixed or other.
Migrant indicator	Derived from <a href="#">bornuk_dv</a> , indicating whether a respondent was born outside the UK (1) or not (0).
Household income quintile (in 2021/22)	Quintile of equivalised household income, in 2022 prices. Derived using <a href="#">fihhmnnet1_dv</a> (total household net income – no deductions), CPI indices, and <a href="#">ieqmoecd_dv</a> (modified OECD equivalence scale). Quintiles created within survey wave.
Worst household income quintile in previous waves	Takes the value of the lowest income quintile in 2012/13 or 2016/17.
Parental profession when respondent was 14	Derived from <a href="#">masoc10_cc</a> and <a href="#">pasoc10_cc</a> , which give mother and father's occupational class when the respondent was 14. We create a binary which equals 1 if either parent had an occupational classification between 111 and 299.
Illness or disability (in 2021/22)	Binary from <a href="#">health</a> , where 1 indicates if the respondent has any long-standing physical or mental impairment, illness or disability.
Household type (in 2021/22)	Derived from <a href="#">hhtype_dv</a> . Grouped as follows: Single working age adult; Single pensioner; Single parent; Working age couple, no children; Pensionable age couple, no children; Couple with children; Other household types.
Housing tenure (in 2021/22)	Derived from <a href="#">tenure_dv</a> (housing tenure) and <a href="#">livpar</a> (if respondent is living at home with their parent(s)).
Working status: retired (in 2021/22)	Household-level job status derived from <a href="#">jbstat</a> . To create the household-level variable, we count across all adults in the household to establish the number of retired adults and separately the number of adults out of work – which includes those unemployed, off sick, on furlough or laid off (but doesn't include those in full-time education, training or caring roles).
Working status: out of work (in 2021/22)	