Financial derived variables and imputation procedures

Introduction

This document describes the imputation procedures and any assumptions made in the derivation of the Financial Derived Variables.

See "User_Guide_Financial_Derived_Variables" for details on how to use the financial derived variables.

Financial information and unfolding brackets

A common feature of all financial variables in ELSA is the use of unfolding brackets. Each financial variable in ELSA is collected by initially requesting an exact answer and then following up with a series of what are referred to as "unfolding brackets". Unfolding brackets operate by asking respondents who are unable or refuse to give an exact answer, a series of follow up questions designed to elicit a minimum and maximum number defining a range or "closed band" within which the value lies. Unlike conventional brackets, respondents are not presented with a list of ranges within which to place themselves. Instead, on entering the unfolding brackets, respondents are asked to say whether they have more, less or about a particular value. This question is repeated using different values (which will be a lower or higher number depending on the answer to the preceding question). The procedure stops at the point when either an upper and lower bound is provided, the respondent refuses or says "don't know", or the respondent places themselves in the top bracket. The unfolding bracket questions are randomly ordered for each respondent such that any possible anchoring effects from the procedure are averaged across the distribution, and the bracket values are selected on the basis of the density of the underlying financial variable.

Unfolding brackets significantly reduce the number of observations for which we have no information at all on any one source of income or wealth. Nevertheless, because refusals and "don't know" are accepted at any stage in the unfolding brackets, the data that we have are of variable type. Respondents provide either an exact amount, a "closed band" where both an upper and a lower bound are reported, an "open band" where respondents are able to provide a lower bound but cannot provide an upper bound (or alternatively place themselves in the top bracket which is open ended by design), missing but positive and missing completely. Missing but positive values arise when the respondent reports that they have a particular type of income or asset but are unable or refuse (even after proceeding through the unfolding brackets) to report how much they have. In a small number of cases, values are completely missing and this arises when respondents refuse to report or don't know whether they have a particular type of income or asset. Completely missing values also arise in a small number of cases where respondents did not complete the housing module and/or the income and assets module (partial interviews).

Where we have banded or missing information, imputation is used to obtain a value. The method of imputation is described in the remainder of this document.

Units of observation for imputation purposes
Financial information in ELSA is recorded in a number of different modules: Work and Pensions (WP), Income and Assets (IA) and Housing and Consumption (HO). WP is asked once per individual, IA is asked once per financial unit and HO is asked once per household. Single people and couples who report that they keep their finances together are defined as a single financial unit. However, couples who report that they keep their finances separate are defined as two separate financial units. Within IA, for couples who keep their finances together (joint finance couples), some variables are identified separately for respondent and their spouse (so the financial respondent is asked separately "How much do you receive?" and "How much does your spouse receive?") and some variables are collected jointly (so the financial respondent is asked "How much do you and your husband/wife/partner receive?"). This means that there are four different types of variables:

a) Variables asked once per individual
b) Variables asked once per financial unit and joint finance units are asked jointly how much each of them receive/have
c) Variables asked once per financial unit but joint finance units are asked separately how much each of them receive/have
d) Variables asked once per household

For types a) and c), we impute values for separately for each individual.

Type b) is imputed at the *benefit unit* level (i.e. once per benefit unit). A benefit unit is defined as a single person or a couple (regardless of whether they keep their finances separate or together). So for a couple who keep their finances separate, even though each of them answered IA separately, we impute a joint amount of each variable of type b). This is because there are insufficient observations of separate finance couples in order to impute separately for each member of these couples. We obtained a joint upper and lower bound of each variable of type b) by summing the upper and lower bounds of each member of the couple.

Type d) are variables relating to housing wealth. We attribute housing wealth to the person/people whose name the accommodation is in and their spouse. If there is more than one person whose name the accommodation is in and those people are in different benefit units, the housing wealth is divided equally across those benefit units.

Imputation
The imputation procedure that we use is the conditional hot-deck. For each missing or banded case, this involves choosing a random observation from all observations with matching characteristics in a number of dimensions (the conditioning variables) and,

where we have banded information, with income or wealth within the same range. The level of wealth or income from the observation that is picked at random is then assigned to the missing or banded case.

For variables that are imputed at the benefit unit level, the conditioning variables that we use are broad age band (50 to state pension age, state pension age to 75 and 75+; for couples the age of the male is used or for single-sex couples, the age of the oldest is used), and benefit unit type (couple, single man or single woman). For variables that are imputed at the individual level, the conditioning variables we use are age and gender. The conditioning variables are used in addition to the upper and lower bounds provided by the respondent. For a small number of variables we use additional or alternative characteristics that are predictive of the missing values. These variables are as follows:

Take home pay: Income from employment is collected in two different places in the ELSA survey instrument. In addition to age and gender, for those who have a completely missing value for take home pay, we impute this using quantile of gross annual benefit unit level earnings.

Current value of main home: (hsval_hh_i, nethw_bu_s, grosshw_bu_s) This is imputed using number of rooms, region and the highest educational level of the householders (where a householder is defined as someone whose name the accommodation is in). Where number of room is missing, the conditioning variables are household size, region and highest educational level of the householders.

Current mortgage debt: (hdebt_h_i, mgdebt_bu_i) This is imputed using tertile of current house value (imputed where necessary) and number of years left to pay on the mortgage. For those who do not report the number of years left to pay on the mortgage, mortgage debt is imputed conditional on the maximum age of the householders. Note that we impute mortgage debt unconditionally on the type of mortgage. Endowment mortgage debt is collected via two questions: the outstanding balance of the mortgage not including the value of the endowment and how much the endowment policy is worth. We compute an upper and lower bound for the net mortgage debt and impute unconditionally on the type of mortgage.

The conditioning variables are used for imputation of missing values, closed bands and open bands except where the open band arises due to the respondent reporting that they have/receive an amount in the top bracket which is open by design. In these cases, because the brackets are designed so that only around 1 per cent of the distribution falls into the top brackets, there are very few observations from which to impute and so we use unconditional hot-deck imputation for these cases.

In the few cases where there are no observations from which to impute missing values, for closed bands, the mid-point of the band is used and for open bands, twice the lower bound is used.

Special cases

There are two types of observations where imputation is more complicated. These are:

a) Benefit units with one non-responding member
b) Benefit units where one or both partners are in institutions (applicable from wave 2 onwards only)

The imputation procedures for these special cases are described in this section.

a)   Benefit units with one non-responding partner

Ultimately what we wish to derive is benefit level income and wealth. In some benefit units, only one member of that benefit unit responds to the survey and one problem is how to deal with non-responding spouses.

Most of the income information is collected at the "financial unit" level. This means that if a couple reports that they keep their finances jointly, the responding member of that couple will answer some questions about income and assets on behalf of the couple. In these benefit units, non-responding spouses do not pose a problem when imputing variables that are collected at the financial unit level.

The majority of benefit units where there is a non-responding spouse, do keep their finances together. However, some variables (namely, those in WP) are asked at the individual level. In order to measure total income of the benefit unit, we need to measure the earnings of both members of that unit. Information on earnings is collected at two points in the ELSA survey. Detailed information is collected in WP at the individual level and then a summary question is asked in IA at the financial unit level. In the derivation of total income, the detailed information on earnings collected at the individual level is used. For benefit units where one member did not respond, even if that unit keep their finances together we have missing information on earnings for the non-responding spouse. In wave 1, we do not know the labour market status of the non-responding spouse. However, the age of the non-responding spouse is recorded and the approach we take in imputing earnings for the non-responder is to assume that non-responders over state pension age have zero earnings. In wave 2 onwards, we do know the employment status of the non-responding spouse as this information is collected in the household grid from the other responding members of the household. For those under state pension age, we use quantiles of the summary measure of earnings (joint earnings of the couple in the past year) recorded in IA as an additional conditioning variable for the purposes of imputing earnings of the non-responding spouse

For couples with one non-responding member who keep their finances separate the responding member of the couple answers all questions on income and assets on their

own behalf and not on behalf of the couple. Because of the lack of information on the non-responding member of these couples, we do not impute income or asset measures for these couples. See below for other cases where we do not impute.

> b) *Benefit units where one or both partners are in institutions (applicable from wave 2 onwards only)*

From Wave 2 onwards, ELSA respondents were followed into institutions (although data is only available on institutions respondents from Wave 3 onwards). This raises additional issues for imputation and these are described here.

We do not derive income and wealth for anyone who was interviewed in an institution. The conceptual issues surrounding allocating these individuals' income and wealth (especially where they have a partner who remains in the private household sector) are complex, and we do not attempt to impute financial variables for these individuals – their values are set to missing for all derived variables. Researchers interested in the institutional sector can use the raw data from these individuals to derive their own income and wealth measures.

There are two identifiers for couples, each of which treats couples in institutions differently. The variable "coupid" is a couple identifier which splits couples where one member (or both members) is in an institution. This means that if there is a couple where one member is in an institution and one member is still living at home, they will each have a different "coupid". The variable "coupid_inst" is a couple identifier which links couples where one member (or both members) is in an institution. Partners of individuals in institutions (who themselves remain in the private household sector) who report that they keep their finances separate from their partner are treated as single individuals for the purposes of deriving financial variables. For couples where one member is in an institution who report that they keep their finances together, we assign the partner living in the private household sector all the income reported by them for themselves and their partner. This is based on the assumption that despite the fact that one of them is in an institution, the partner remaining at home does benefit from the partner's income. Where imputation is required for joint finance couples where one of them is in an institution, we do not condition on whether they are a member of couple. Given the additional complex issues surrounding couples where one member is in an institution, careful consideration should be given to the treatment of these couples for the purposes of analysis.

Derivation of quantiles

Quantiles are weighted and are derived only for sample members since this is the representative part of the ELSA sample. Non-sample members are assigned a missing value.

Cases where income and wealth is not imputed

There are some cases for whom we do not impute income and wealth. These are cases where the extent of the non-response means that we do not have enough information to carry out imputation in a reliable way. These cases are assigned a missing value. These cases are:

Income and wealth from IA is not imputed for:
1) Single people who do not answer IA
2) Joint finance couples where neither partner answered IA (either because they did not respond to the whole survey or because they did not answer the IA module)
3) Separate finance couples where at least one partner did not answer IA

Income from WP is not imputed at the benefit unit level for:

1) Separate finance couples where one partner did not respond.