

## Financial derived variables user guide (Financial derived variables.dta)

### INTRODUCTION

This document describes how to use the derived income and asset variables. For more information on the imputation procedure, please refer to the document named “Financial derived variables and imputations procedures”.

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.

Each financial variable in ELSA will therefore have a set of raw variables associated with it containing both the continuous and the bracketed information. For more details on this, see the Questionnaire. These variables are used to create a single variable that, after imputation, contains a continuous variable for every (non-proxy) respondent.

These variables are then used to derive summary variables for income and wealth.

### HOW TO USE THE DERIVED VARIABLES

The derived financial variables are in a dataset called “Financial derived variables.dta”.



The variables are labelled and the accompanying excel spreadsheet (“Variable relationships for financial DVs.xls”) matches the derived variables to their source variables and provides further information specific to particular variables.

All monetary income flows are expressed as weekly equivalents and are *current*. ELSA collects information about current income (i.e. income in the last month) and, in order to obtain a longer term picture, income in the last year. All derivation of total income uses a current measure of income.

Many of the variables are derived at the “benefit unit level”. This is a couple or a single person with any dependent children they may have. These variables are suffixed “\_bu”. Variables which have no suffix refer to the individual. Variables which have a “\_p” suffix refer to the partner of that individual.

Note that the “benefit unit” is different to a “financial unit”. A couple who keep their finances separate will be defined as two financial units and each will answer IA on their own behalf. For couples that keep their finances separate, we combine the information reported by each member of the couple to obtain a benefit unit definition of income and wealth.

It should be noted that because IA is asked once per “financial unit” the income and wealth information for any individual may have been reported by that individual OR their partner (see User Guide for the Wave 1 Core Dataset for details). Users wishing to know which financial unit member answered questions from IA should use the variable called “iapid”.

## **TYPES OF VARIABLES**

A) variables ending in “i” (before the \_bu/\_p suffix), contain the imputed values.

B) Variables ending in “t” describe the type of imputation that took place

**0 zero** - value of zero reported so no imputation took place

**1 continuous** - exact value reported so no imputation took place

**2 closed band** - this type of income or wealth was imputed from an upper and lower bound (e.g. between £100 and £200)

**3 open band** - this type of income or wealth was imputed from a lower bound but no upper bound (e.g. £100 or more)

**4 missing, positive** - either (or both) members of the benefit unit reported that they have this type of income or wealth but they cannot (or refuse to) give us a value even after entering the unfolding brackets

**5 missing completely** – neither member of the benefit unit can tell us if they have this type of income or wealth

**7 missing, has asset** – this code is unique to assets. It occurs when the either member of the benefit unit reports that they have a particular asset but they cannot tell us any information about the income earned from that asset.

**8 missing, has some type of ISA** – this code is unique to ISAs. It occurs when the benefit unit reports that they have an ISA but they cannot tell us which type (cash, shares, or life insurance) of ISA it is.

C) Variables ending in “o” store the number of observations used to impute the missing value. Note that when the number of observations is zero, an alternate imputation strategy was used (see “Derived variables and imputations procedures” for more details).

D) Summary variables. These are variables that sum different components of income and wealth.

E) Variables beginning with “ni”. These are the imputation flags which accompany the summary variables. Because the summary variables use many different components of

income and wealth and any of those components could have been imputed in some way, these variables describe how many of the components were imputed. Variables of the form *nivarname2\_BU* (where *varname* is the summary variable such as *totinc*) report how many components had a “type” of greater or equal to 2 (i.e. closed band or worse). Variables of the form *nivarname3\_BU* report how many components had a “type” of greater or equal to 3 (i.e. open band or worse). Variables of the form *nivarname4\_BU* report how many components had a “type” of greater or equal to 4 (i.e. missing, positive or worse).

## **IDENTIFIERS**

In addition to the usual identifiers found in the main data, there are two additional identifiers. These are “cupid” and “fuid”. “Cupid” is an identifier which uniquely identifies couples. Each member of a couple has the same cupid. Single people have their own cupid. Note that where someone is a member of a couple, but one member of that couple did not respond to the main survey, there will only be one observation for that couple in the ELSA data. “fuid” uniquely identifies financial units. Each respondent who answered IA has a unique “fuid”.

## **MISSING VALUES**

There is one observation for each ELSA respondent in “Financial derived variables.dta”. However, there are 122 observations for whom we do not derive income and asset variables. Of these 99 are sample members and 23 are not. All of them have spouses who did not respond to the survey. Variables for these observations take the value -3. The 99 sample members are members of couples who keep their finances separate. Because the spouses of these sample members did not respond, we have no information on that spouse. This would make imputation for the benefit unit very unreliable so we do not attempt it. We also do not carry out imputation for the 22 non-sample members whose spouses (who are sample members) do not respond.

Note that it is possible to impute missing values for couples where one member did not respond but who keep their finances together as the responding member of the couple answers IA on behalf of the couple.

The second type of missing value is reserved for the quantiles of income and wealth. Quantiles are calculated only for sample members and are weighted. Non-sample members are given a value of -2.

## **EQUIVALISATION**

Equivalisation is a way in which income can be adjusted to take account for differing household size. The equivalence scale used here is an OECD equivalence scale<sup>1</sup> and

---

<sup>1</sup> See HAGENAARS, A., DE VOS, K. and ZAIDI, A. (1994), *Poverty Statistics in the late 1980s*, Eurostat, Luxembourg.

assigns a weight of 0.5 to second adults and dependent children aged 14 and over and a weight of 0.3 to children under 14 years of age.

## **ABBREVIATIONS**

a/c	account
DLA	disability living allowance
DPTC	disabled persons tax credit
ind inj benefit	industrial injuries disablement benefit
inv care all	invalid care allowance
ISA	individual savings account
JSA	job seeker's allowance
MIG	minimum income guarantee
PEP	personal equity plan
SDA	severe disablement allowance
SSP	statutory sick pay
TESSA	tax exempt special savings account
WFTC	working families tax credit

### **Note to users of old versions of SPSS (11.5 and earlier)**

Those supplied with this study in SPSS format should note that versions of SPSS prior to 12.0 will not support long (> 8 character) variable names, of which there are many in the 'financial\_derived\_variables.sav' data file.

SPSS abbreviates variable names as follows:

Where the first 8 characters of a long variable name are unique, SPSS will abbreviate the variable name to that. Where duplicates exist it will use the first six characters, followed by an \_1 to indicate the first duplicated name, \_2 for the second etc (assuming there are <10 duplicates). If there are 10 or more duplicates on the first 8 characters, SPSS will use the first 6 characters followed by \_1 to \_9 for the first nine encountered, followed by the first five characters and \_10 to \_99, followed by the first four characters and \_100 to \_999, etc.

If you are using a version of SPSS prior to 12, the following list provides a look-up table to compare the abbreviated names that will be present when you view the data in older versions of SPSS, compared with how they will be referenced in the documentation. SPSS does not inform the user it has abbreviated names upon opening a data file with long variable names.

Abbreviated variable names in *financial\_derived\_variables.sav*

<b>Abbreviated variable name [SPSS 11.5 and prior versions]</b>	<b>Full name [in SPSS 12 and the STATA and TAB versions]</b>
assinc_b	assinc_bu
attalli_	attalli_p
attallo_	attallo_p
attallt_	attallt_p
beninc_b	beninc_bu
bondsi_b	bondsi_bu
bondsii_	bondsii_bu
bondsio_	bondsio_bu
bondsit_	bondsit_bu
bondso_b	bondso_bu
bondst_b	bondst_bu
cashisai	cashisai_bu
cashisao	cashisao_bu
cashisat	cashisat_bu
ccardi_b	ccardi_bu
ccardo_b	ccardo_bu
ccardt_b	ccardt_bu
empinc_b	empinc_bu
eqassinc	eqassinc_bu
eqbeninc	eqbeninc_bu
eqempinc	eqempinc_bu
eqothinc	eqothinc_bu
eqppinc_	eqppinc_bu
eqseinc_	eqseinc_bu
eqspinc_	eqspinc_bu

eqtotinc	eqtotinc_bu
farmii_b	farmii_bu
farmio_b	farmio_bu
farmit_b	farmit_bu
ghwealth	ghwealth_bu
grossfw_	grossfw_bu
grosstot	grosstotnhw_bu
hdebt_i_b	hdebt_i_bu
hdebto_b	hdebto_bu
hdebtt_b	hdebtt_bu
homeii_b	homeii_bu
homeio_b	homeio_bu
homeit_b	homeit_bu
hsvali_b	hsvali_bu
hsvalo_b	hsvalo_bu
hsvalt_b	hsvalt_bu
indinji_	indinji_p
indinjo_	indinjo_p
indinjt_	indinjt_p
invcar_1	invcarei_p
invcar_2	invcareo_p
invcar_3	invcaret_p
invests_	invests_bu
jntassi_	jntassi_bu
jntasso_	jntasso_bu
jntasst_	jntasst_bu
lilumpi_	lilumpi_p
lilumpo_	lilumpo_p
lilumpt_	lilumpt_p
lisavei_	lisavei_bu
lisaveo_	lisaveo_bu
lisavet_	lisavet_bu
netpro_1	netprofi_bu
netpro_2	netproft_bu
nettotnh	nettotnhw_bu
nettotw_	nettotw_bu
nfwq10_b	nfwq10_bu
nhwealth	nhwealth_bu
niassi_1	niassinc3_bu
niassi_2	niassinc4_bu
niassinc	niassinc2_bu
nibeni_1	nibeninc3_bu
nibeni_2	nibeninc4_bu
nibeninc	nibeninc2_bu
nidebt2_	nidebt2_bu
nidebt3_	nidebt3_bu
nidebt4_	nidebt4_bu
niempi_1	niempinc3_bu
niempi_2	niempinc4_bu
niempinc	niempinc2_bu
nighwe_1	nighwealth3_bu
nighwe_2	nighwealth4_bu
nighweal	nighwealth2_bu
nigros_1	nigrossfw3_bu

nigros_2	nigrossfw4_bu
nigros_3	nigrosstotnhw3_bu
nigros_4	nigrosstotnhw4_bu
nigrossf	nigrossfw2_bu
nigrosst	nigrosstotnhw2_bu
niinve_1	niinvests3_bu
niinve_2	niinvests4_bu
niinvest	niinvests2_bu
ninetfw2	ninetfw2_bu
ninetfw3	ninetfw3_bu
ninetfw4	ninetfw4_bu
ninetpw2	ninetpw2_bu
ninetpw3	ninetpw3_bu
ninetpw4	ninetpw4_bu
ninett_1	ninettotnhw3_bu
ninett_2	ninettotnhw4_bu
ninett_3	ninettotw2_bu
ninett_4	ninettotw3_bu
ninett_5	ninettotw4_bu
ninettot	ninettotnhw2_bu
ninhwe_1	ninhwealth3_bu
ninhwe_2	ninhwealth4_bu
ninhweal	ninhwealth2_bu
niothi_1	niothinc3_bu
niothi_2	niothinc4_bu
niothinc	niothinc2_bu
nippin_1	nippinct3_bu
nippin_2	nippinct4_bu
nippinct	nippinct2_bu
nisavi_1	nisavings3_bu
nisavi_2	nisavings4_bu
nisaving	nisavings2_bu
niseinc2	niseinc2_bu
niseinc3	niseinc3_bu
niseinc4	niseinc4_bu
nispinc2	nispinc2_bu
nispinc3	nispinc3_bu
nispinc4	nispinc4_bu
nitoti_1	nitotinc3_bu
nitoti_2	nitotinc4_bu
nitotinc	nitotinc2_bu
nsavii_b	nsavii_bu
nsavio_b	nsavio_bu
nsavit_b	nsavit_bu
odebti_b	odebti_bu
odebto_b	odebto_bu
odebtt_b	odebtt_bu
othbus_1	othbusvi_bu
othbus_2	othbusvt_bu
othinc_b	othinc_bu
othpayi_	othpayi_p
othpayo_	othpayo_p
othpayt_	othpayt_p
othsavi_	othsavi_bu

othsavii	othsavii_bu
othsavio	othsavio_bu
othsavit	othsavit_bu
othsavo_	othsavo_bu
othsavn_	othsavn_bu
prbond_1	prbondsit_bu
prbond_2	prbondsii_bu
prbond_3	prbondsio_bu
prbondsi	prbondsi_bu
prbondso	prbondso_bu
prbondst	prbondst_bu
prdebt_	prdebt_bu
prdebt_1	prdebt_1_bu
prdebt_2	prdebt_2_bu
prdebt_3	prdebt_3_bu
saveii_b	saveii_bu
saveio_b	saveio_bu
saveit_b	saveit_bu
savings_	savings_bu
sedrawi_	sedrawi_bu
sedrawt_	sedrawt_bu
sharesi_	sharesi_bu
sharesii	sharesii_bu
sharesio	sharesio_bu
sharesit	sharesit_bu
shareso_	shareso_bu
sharest_	sharest_bu
shisai_b	shisai_bu
shisao_b	shisao_bu
shisat_b	shisat_bu
tessai_b	tessai_bu
tessai_	tessai_bu
tessaio_	tessaio_bu
tessait_	tessait_bu
tessao_b	tessao_bu
tessat_b	tessat_bu
tnhwq10_	tnhwq10_bu
tnhwq5_b	tnhwq5_bu
tothassi	tothassi_bu
tothasso	tothasso_bu
tothasst	tothasst_bu
totinc_b	totinc_bu
totwq10_	totwq10_bu
totwq5_b	totwq5_bu
trustsi_	trustsi_bu
trustsii	trustsii_bu
trustsio	trustsio_bu
trustsit	trustsit_bu
trustso_	trustso_bu
trustst_	trustst_bu
widpeni_	widpeni_p
widpeno_	widpeno_p
widpent_	widpent_p

## **Financial derived variables and imputation procedures**

### **INTRODUCTION**

This document describes the imputation procedures and any assumptions made in the derivation of the income and asset variables. See “Financial DVs user guide.doc” for details on how to use to income and asset 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 spouse 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 c). 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 c) 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:

*Current value of main home:* (hsvali\_bu) 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\_bu) 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.

## **BENEFIT UNITS WITH ONE NON-RESPONDING MEMBER**

Ultimately what we wish to derive is benefit level income and wealth. In some benefit units (468), only one member of that benefit unit responds to the survey and one problem is how to deal with non-responding spouses. 23 of these couples are made up of one sample member and one non-sample member and it is the sample member who responds to the survey. Because it is the sample member who we are interested in for analysis, we do not impute income for the responding non-sample member.

In the remaining 445 couples the responding member of the couple is a sample member. 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 (346) 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. 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. 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.

In addition to the 346 couples with one non-responding member who keep their finances together, there are 99 couples who keep their finances separate. In these cases, 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.

In total, there are 122 cases where we do not impute income and wealth measures. Of these, 23 are individuals who are not sample members but who have non-responding spouses who are sample members. The remaining 99 are couples where one member does not respond and they keep their finances separate. These cases are identified by a value of -3 for the derived variables.