Financial derived variables user guide

Contents

1. INTRODUCTION

ELSA's financial derived variables are designed to make economic information quickly and easily available to ELSA users. This document describes how to use the financial derived variables. For more information on the imputation procedure, please refer to the document named Financial_Derived_Variables_and_Imputation_procedures". It is important to bear in mind when using data that imputation is a statistical process and imputed data should not be used without careful thought as to how any measurement error will affect any analysis.

Financial information is collected at a very detailed level in ELSA, but this detail can make the data somewhat unwieldy to use. The questionnaire contains hundreds of questions regarding respondents' economic situation (earnings, benefits, pensions, assets, debts, etc) and they are spread over several different modules (Work and Pensions (WP), Income and Assets (IA) and Housing (HO)). The raw data can therefore be difficult to work with. For example, a researcher wishing to know how much income each ELSA respondent (and/or their partner) received from the state pension would have to contend with over 20 different variables which record information relating to this source of income.

One reason that there are so many variables relating to each source of income is that if a respondent refuses to reveal, or does not know, how much of a particular income source they receive, further probing is carried out to see if the respondent is willing or able to give a range ("… between £X and £Y"). These follow-up questions are referred to as "unfolding brackets". The questions are designed to elicit a minimum and maximum number within which the value lies. The information from these brackets is recorded in around twenty different variables for each income source.

The financial derived variables are designed to make using the economic information in ELSA simpler and easier to use by:

- combining information from the continuous and bracketed values for each income or wealth source into a single variable
- transforming each value into a uniform period (where appropriate)
- imputing missing values where necessary so that we have information on almost every respondent[1]
- transforming the data so that any individual level measures of income or wealth are attributed to the respondent to whom they belong (and do not depend on which member of a couple answered the question)
- creating benefit unit level measures of income and wealth

[1] We do not impute values for respondents who do not answer any of the questions in the Income and Assets module.

- creating meaningful aggregates (e.g. total income, employment income, total wealth)
- creating quantiles of income and wealth

This User Guide starts by answering some questions which are commonly asked. It then goes on to describe the data in more detail.

2. FAQ

Which variable should I use?

This really depends on the research question. However, a good place to start when looking at the income and wealth variables is the Summary Variables and these may be sufficient for many users. Also have a look at Figure 1 and Figure 2 which show the structure of wealth and income and how the components fit together.

Table 5 may also be useful as it describes the components of income and wealth in the data and groups them into categories. It also details whether the measure is available at the household, benefit unit or individual level and whether it is a summary variable or a component of a summary variable.

The spreadsheet that accompanies this documentation may also be useful:

5050_Financial_Derived_Variable_Relationships.xls

The spreadsheet (which has one tab per wave) lists all the variables and their source variables and any formula used to create the derived variables. There are two columns which allow you to filter the number of variables shown in the spreadsheet. The filters apply to columns B and C and filter by the type of variable (whether it contains a value or whether it is an imputation flag for example) and by the category of income or wealth. For example, if you are interested in employment income, selecting only the "employment income" variables in the filter on column B shows you which variables are most relevant. If you are primarily interested in the actual value of the employment income (and not the corresponding imputation flags) you can further refine the filter by choosing "value" in the filter in column C.

Which private pension income variable should I use?

You may wish to read the section on naming conventions before reading the answer to this question.

There are two different measures of private pension income in the data because information about private pension income is collected in two different places in the ELSA survey instrument (once in the IA module and once in the WP module).

wppp_r_i and wppp_p_i (or the benefit unit equivalent - wppp_bu_i) are individual level, after tax measures of private pension income taken from the WP module (wppyr).

ppen_bu_i is a benefit unit level measure of before tax private pension income taken from the IA module (iappei)

wppp_bu_i and ppen_bu_i are alternative, similar, measures of the same income. Users should use one or the other but they should never be summed together.

Differences between "ppen" and "wppp"

- ppen_bu_i is measured before tax and wppp_bu_i is measured after tax so analysts should choose whichever is more appropriate in this respect.

- wppp__bu_i is taken from questions that each individual answers about their own income and are therefore available at the individual level. ppen_bu_i is taken from questions that are answered once per financial unit and so is only available for couples at the benefit unit level. If individual level private pension income is needed, analysts should use wppp_r_i (and wppp_p_i if they need partner's private pension income).

- However, one advantage of ppen_bu_i is that the variable is imputed from completely missing data less often. This is because where a respondent has a partner who does not respond to the survey, as long as the couple keep their finances together, the responding spouse will be able to report the total private pension income for the couple in the IA module. However, the WP module is asked once per respondent and so where a partner does not respond, data on that individual will be missing (and is therefore imputed).

In addition to wppp_bu_i and ppen_bu_i, there is one further variable which may be of interest to analysts looking for information about private pension income (see Table 5, Annuitised Income). The summary variable ppinc_bu_s (and the respondent and partner equivalents, ppinc_r_s and ppinc_p_s) sums together income from private pensions and other annuity income to form total annuitised income. In previous versions of the data (versions released prior to wave 5), instead of wppp_bu_i, wppp_r_i and wppp_p_i being used as the source variables for the summary measure of total annuitised income, the private pension income variable from IA was used (ppen_bu_i). Please see Section 10 for the reasons behind this change.

Which employment income variable should I use?

You may wish to read the section on naming conventions before reading the answer to this question.

There are two different variables relating to income from *main* employment in the data. This is partly because information about employment income is asked in two different places in the ELSA survey instrument (once in the IA module and once in the WP module).

thp_r_i and thp_p_i (or the benefit unit equivalent - thp_bu_i) are individual level, after tax measures of employment income taken from the WP module (wpthp)

sinc_bu_i is a benefit unit level of before tax employment income taken from the IA module (iasinc).

thp_bu_i and sinc_bu_i are alternative, similar, measures of the same income. Users should use one or the other but they should never be summed together.

Differences between "sinc" and "thp"

- sinc_bu_i is measured before tax and thp_bu_i is measured after tax so analysts should choose whichever is more appropriate in this respect.

- thp_bu_i is taken from questions that each individual answers about their own income and are therefore available at the individual level. sinc_bu_i is taken from questions that are answered once per financial unit and so is only available for couples at the benefit unit level. If individual level employment income is needed, analysts should use thp_r_i (and thp_p_i if they need partner's employment income).

- One advantage of sinc_bu_i is that the variable is imputed from completely missing data less often. This is because where a respondent has a partner who does not respond to the survey, as long as the couple keep their finances together, the responding spouse will be able to report employment income for the couple in the IA module. However, the WP module is asked once per respondent and so where a partner does not respond, data on that individual will be missing (and is therefore imputed).

In addition to thp_bu_i and sinc_bu_i, there is one further variable which may be of interest to analysts looking for information about employment income (see Table 5, Employment Income). The summary variable empinc_bu_s (and their individual level equivalents empinc_r_s and empinc_p_s) sums together income from main employment using thp_bu_i/thp_r_i/thp_p_i and income from subsidiary employment (oj_bu_i/oj_r_i/oj_p_i) to form *total* Employment income.

Which housing wealth variable should I use?

There are a number of different measures of housing wealth available in the data (see Table 5, Housing Wealth). Primary housing wealth is collected at the household level in

the survey but some of the measures are transformed into benefit unit level by attributing the wealth only to the individuals who are named on the property (and their spouses) and where owners are in different benefit units, dividing housing wealth accordingly. This avoids housing wealth being assigned to (for example) grown up children who may leave the family home or elderly parents who have moved in with their children. The variables nethw, grosshw and mgdebt are all measured at the benefit unit level.

However, for some purposes, it may be more appropriate for primary housing wealth to be assigned to all persons living within the household or the analyst may be more interested in the value of the house for purposes other than as a part of wealth. The variables hsval and hdebt are the value of the primary house and mortgage debt respectively and are assigned to all members of the household. For simple households, where all members of the household are named on the property and are in the same benefit unit and there are no people living outside the property named on the property, grosshw will be identical to hsval and mgdebt will be identical to hdebt. Therefore, users should use one or the other but they should never be summed together.

There is also a measure of secondary housing wealth (second homes or holiday homes) which is measured at the benefit unit level in the same way as other physical wealth.

Which variables in the questionnaire do the derived variables correspond to?

This is documented in the following spreadsheets:
5050_Financial_Derived_Variable_Relationships.xls

For example, looking at the Wave 1 tab in 5050_Financial_Derived_Variable_Relationships.xls, and scrolling down or searching for the variable name "spen_r_i", you can see from column E in the spreadsheet that the source variable is iapam or iappam. The information from these two variables has been combined with the appropriate period of receipt information, has been imputed where necessary and then assigned to the respondent or their spouse (depending on who answered the question) in order to create the variable spen_r_i.

Wherever possible, the source variable name is also included in the variable label (sometimes this is not possible because the number of source variables involved would make the variable label too long).

Is income measured net or gross (before or after tax)?

The ELSA survey generally collects income net of tax and therefore all our income measures are reported net of tax. For self employment drawings (from wave 2) respondents are asked to report a gross figure and for self-employment profit, respondents can report either a gross or a net figure. We convert all self-employment income into a net figure using the structure of the tax system and assuming that this

income is the main source of income (and so the full value of the personal allowance is applied).

What if I want a measure of gross income?

Some components of income are also collected gross of tax in the ELSA questionnaire. In particular, earnings from primary employment are collected gross of tax in the Work and Pensions module (questions wpotp and wpaotp). However, the financial derived variables dataset does not contain imputed values for this measure of earnings – interested users would need to decide for themselves how to deal with missing values.

Is there an individual measure of income or wealth?

The ELSA survey is primarily designed to measure income and wealth at the benefit unit level for couples who keep their finances together. However, some sources of income are collected in the survey at the individual level (income sources in WP) or collected separately for the individual and their partner (but reported once only by the person in the couple who is the financial respondent). Where a source of income is an individual level measure, it will contain an "_r" (for the respondent) or an "_p" (for the respondent's partner) suffix. Table 5 indicates which of the measures are available at the individual level. Broadly speaking, most income measures except asset income are available at the individual level but wealth measures are available only at the benefit unit level.

Over what time period is income measured?

Respondents are free to report their incomes over any time period but for the financial derived variables, all incomes are converted to weekly values.

I want a measure of earnings – which variable should I use?

Net (that is, after tax) income from primary employment is recorded at the individual level in "thp_r_i". Net income from subsidiary employment is recorded in "oj_r_i". Self employment income is recorded in netprof_r_i and sedraw_r_i depending on whether the individual keeps accounts (netprof_r_i) or not (sedraw_r_i). Corresponding measures of benefit unit earnings and partner earnings are also available.

Is there an equivalised measure of income?

Yes, for all the summary (aggregate) measures of income, there is a measure of equivalised total benefit unit income, which uses a commonly used "equivalence scale" to convert the total benefit unit income of multi-person benefit units into an equivalent figure for an "equally well-off" single individual. See the section on equivalisation below for further information.

Is there an equivalised measure of wealth?

No, because unlike for income, there is no commonly recognised way of adjusting wealth holdings to account for family size.

3. DATA DESCRIPTION

The Financial Derived variable datasets are at the individual level and can be merged into the Core ELSA dataset. The variables are labeled and the accompanying excel spreadsheet matches the derived variables to their source variables and provides further information specific to particular variables. There is one tab in the spreadsheet for each wave of data:

5050_Financial_Derived_Variable_Relationships.xls,

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[2]. However, all derivation of total income for the financial derived variables uses a measure of current income.

ELSA is designed to make the reporting of income and wealth sources as simple as possible for respondents. Where couples keep their finances together, it may not be meaningful to collect a separate amount of income or wealth for each member of a couple since the individual allocation may not be easily separated. For this reason, at the start of the survey, respondents who are part of a couple are asked whether they keep their finances completely separate. If they do not keep them completely separate, they are asked to identify which member of the couple is best placed to answer questions on income and assets and this person (the "financial respondent") answers all the questions in the Income and Assets (IA) module on behalf of the couple. However, all questions in the WP module are asked to each respondent regardless of whether they keep their finances together or separate. WP contains income questions relating to work and questions about business wealth. In most cases each person in a couple would be easily able to identify separately their own income or wealth. This means that the different sources of income and wealth are collected in one of three ways:

a) The question is asked once per respondent and each respondent reports their own income or wealth. All questions in WP are of this type.
b) The question is asked once per financial unit but the financial respondent reports their own ("how much do you receive?") and their partner's income ("how much does your partner receive?") or wealth separately. A subset of questions in IA are of this type.
c) The question is asked once per financial unit and the financial respondent reports the joint value of their and their partner's income or wealth ("how much do you and your partner receive?"). A subset of questions in IA are of this type.

Income sources that are asked once per financial unit but the financial respondent reports their own and their partner's income (type b) are sources where it is typically

[2] From Wave 5 onwards, income will be collected on a current basis only.

simple for the respondent to identify the recipient of the income (for example benefit receipt or state benefit receipt). Income or wealth sources of type c) are those where it is more difficult for couples who keep their finances together to separately identify who the income or wealth belongs to (for example any asset income where the asset is a joint one).

In the financial derived variables, we derive variables at the individual level wherever possible. However it is important to note that although income may be measured at the individual level, the value may have been reported by either the respondent or their partner for couples who keep their finances together. Users wishing to know which financial unit member answered questions from IA should use the variable called "iapid". Where it is not possible to measure a particular income or wealth component at the individual level (case (c) above), we construct a benefit unit level measure. This is a couple or a single person plus any dependent children they may have. Note that the "benefit unit" is different from a "financial unit". A couple who keep their finances separate will be defined as two financial units and each will answer the Income and Assets (IA) module 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.

Table 5 describes the components of income and wealth in the data and groups them into categories. It also details whether the measure is available at the household, benefit unit or individual level and whether it is a summary variable or a component of a summary variable.

4. A GUIDE TO NAMING CONVENTIONS

The variables in the financial derived variables dataset follow a standard set of naming conventions. Some examples are given at the end of this section. Each variable name starts with a main descriptive "stem" – for example, "sinc" (salary income). Following the stem name there will be one of four suffixes. These are described in Table 1.

Table 1. Description of first suffixes

Suffix	Description
_r	This source of income relates to the individual (respondent)
_p	This source of income relates to the partner of the respondent
_bu	This source of income relates to the benefit unit. All members of each benefit unit are assigned the same value for these income/wealth measures.
_hh	This source of income relates to the household. Each member of the same household is assigned the same value for these sources of income/wealth.

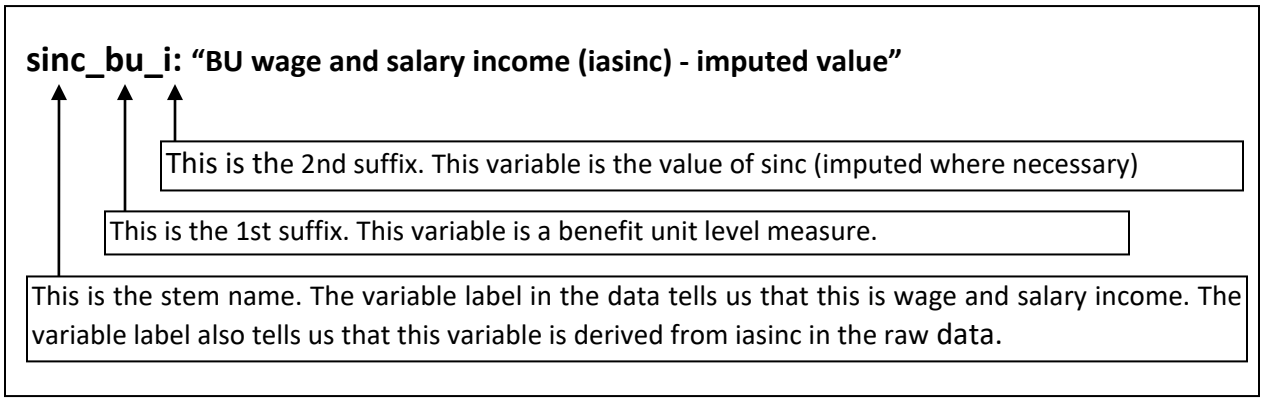
Following the first suffix, there is a second suffix. The second suffix is described in Table 2.

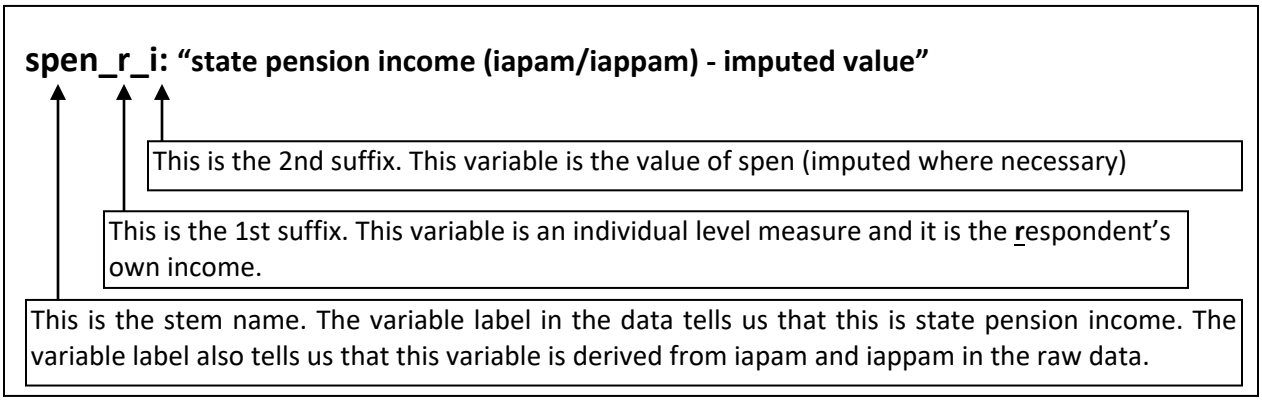
Table 2. Description of second suffixes

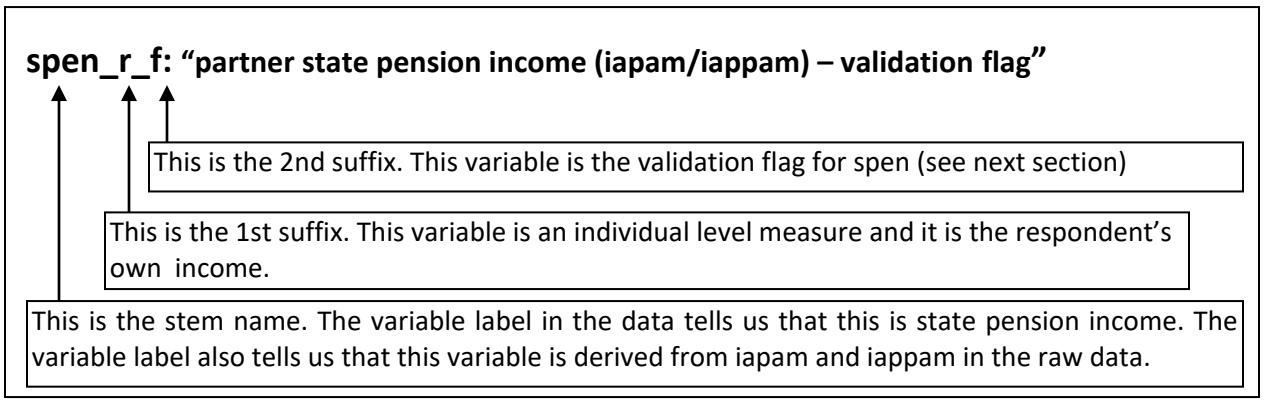
Suffix	Description
_i	These variables contain the (imputed) value of the income or wealth. For individuals who reported an exact figure for the component of income, the variable will contain that value. For individuals reporting a band or who had a missing value, the variable will contain the imputed value.
_t	These variables 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 between an upper and lower bound (e.g. between £100 and £200) **3 open band -** this type of income or wealth was imputed from a range bounded only by a lower bound, with 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 any indication of the 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 asset income. It occurs when 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.
_o	These variables store the number of observations used to impute the missing value. Note that when the number of observations is zero, an alternative imputation strategy was used and this variable is coded -2 (see "Financial_Derived_Variables_and_Imputation_procedures" for more details).
_f	These variables are validation flags. We do not impute for all individuals (see section 5 below on missing values). Where an imputed variable (an _i variable) is missing, the "_f" variable is given a value of 1 2 or 3 depending on why it is missing. The coding frame for the _f variables are as follows: **0 "valid"** **1 "non-sample member"**

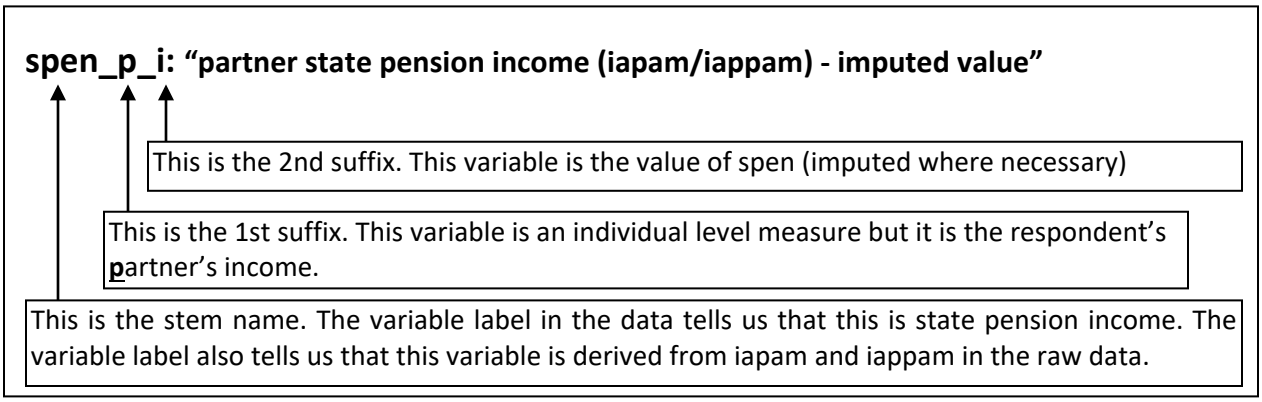
	2 "not imputed (see user guide)" **3 "institutional respondent"**
_s	These are variables that aggregate different components of income and wealth to derive summary measures. See the next section for more details.

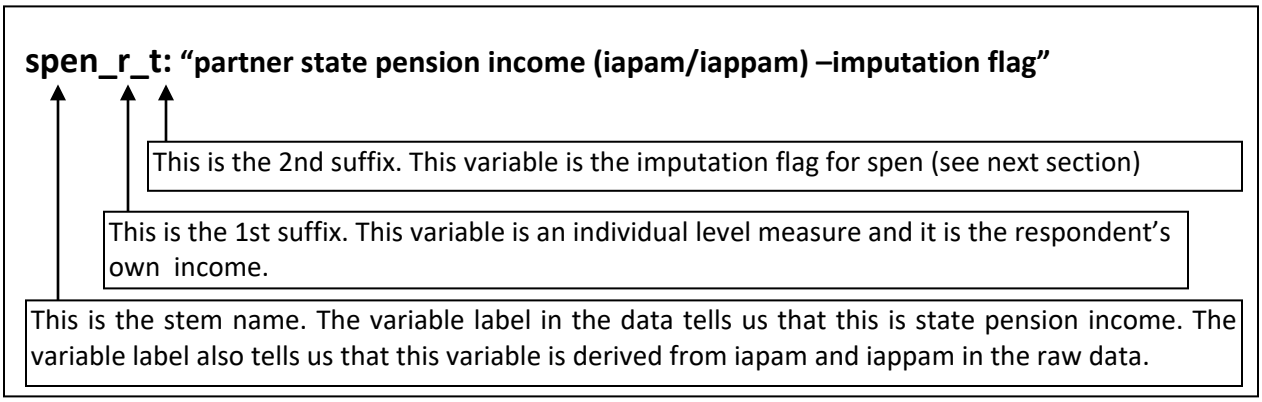
4.1. Examples of variable names

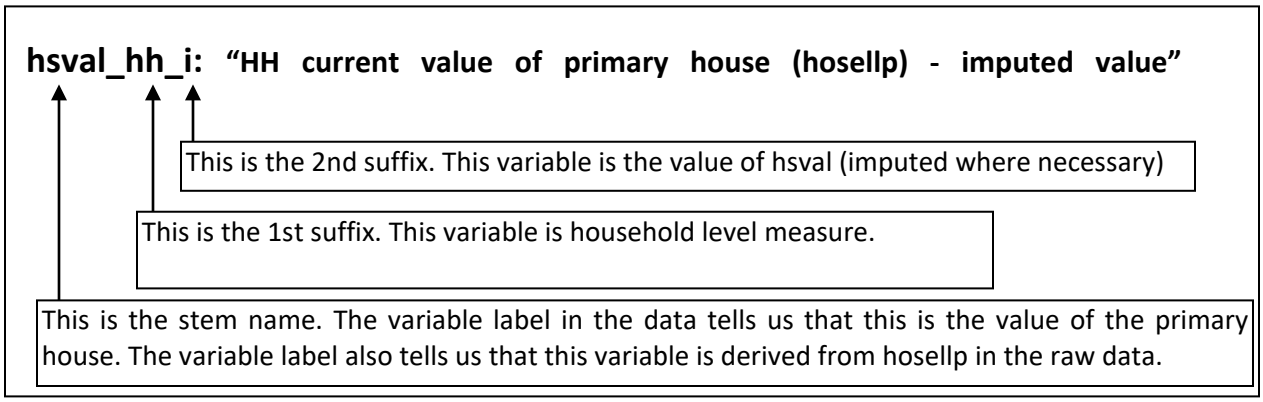







spen_p_i: "partner state pension income (iapam/iappam) - imputed value"
This is the 2nd suffix. This variable is the value of spen (imputed where necessary)
This is the 1st suffix. This variable is an individual level measure but it is the respondent's partner's income.
This is the stem name. The variable label in the data tells us that this is state pension income. The variable label also tells us that this variable is derived from iapam and iappam in the raw data.


spen_r_t: "partner state pension income (iapam/iappam) –imputation flag"
This is the 2nd suffix. This variable is the imputation flag for spen (see next section)
This is the 1st suffix. This variable is an individual level measure and it is the respondent's own income.
This is the stem name. The variable label in the data tells us that this is state pension income. The variable label also tells us that this variable is derived from iapam and iappam in the raw data.


hsval_hh_i: "HH current value of primary house (hosellp) - imputed value"
This is the 2nd suffix. This variable is the value of hsval (imputed where necessary)
This is the 1st suffix. This variable is household level measure.
This is the stem name. The variable label in the data tells us that this is the value of the primary house. The variable label also tells us that this variable is derived from hosellp in the raw data.

5. MISSING VALUES AND VALIDATION FLAGS

There is one observation for each ELSA respondent in each of the financial derived variables datasets. However, there are certain circumstances where individuals are given a system missing value for income and wealth measures. These cases are:

1) Where the extent of the non-response means that we do not have enough information to carry out imputation in a reliable way.

   Income and wealth from IA is not imputed for:

   a) Single people who do not answer IA
   b) 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)
   c) Separate finance couples where at least one partner did not answer IA

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

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

2) Individuals who are interviewed in an institution for whom imputation is more complex.

3) Non-Sample members for whom we do not calculate quantiles of income or wealth. Non-sample members are given a system missing value for their quantile but the value of their income and wealth is contained in the data.

4) Sample members who have a zero weight are treated the same as non-sample member. That is, we do not calculate quantiles of income or wealth. Sample members may have a zero weight either because they are in an institution, they no longer live in England or they are one of the respondents who should have been part of the Wave 3 refreshment sample but were only interviewed as a result of being an older or younger partner (see Wave 3 main User guide for more details of this latter group).

To identify each of these cases separately, each derived variable has an accompanying validation flag (an "_f" variable). The coding frame is as follows:

0 valid (i.e. non-missing)
1 non-sample member
2 not imputed
3 institutional respondent

Box 1 shows an example of a validation flag variable for Wave 4. The variable here is BU wage and salary income (sinc). There are 11,050 observations in total in the data but sinc_bu_i contains only 10755 values. Tabulating the corresponding validation flag (sinc_bu_f) reveals that 227 of the observations are not imputed (because we do not have sufficient information on them to carry this out reliably) and 68 respondents are in institutions.

Box 1. Example of a validation flag variable

```
. su sinc_bu_i

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   sinc_bu_i |      10755    307.6342    570.9892          0   15192.31

. tab sinc_bu_f

  BU wage and salary income
 (iasinc) - validation flag |      Freq.     Percent        Cum.
----------------------------+-----------------------------------
                      valid |     10,755       97.33       97.33
not imputed (see user guide) |        227        2.05       99.38
   institutional respondent |         68        0.62      100.00
----------------------------+-----------------------------------
                      Total |     11,050      100.00
```

6. USING THE IMPUTATION FLAGS

It is important to bear in mind when using data that imputation is a statistical process and imputed data should not be used without careful thought as to how any measurement error will affect any analysis. Longitudinal analyses of changes in income or wealth for example are very sensitive to any measurement error and so may be particularly sensitive to imputed values. In order that analysts can make informed decisions about the data they are using and in particular about how or whether to use any imputed values, each variable that contains imputed data has corresponding imputation flags. For all variables except the summary variables (see next section), the imputation flags have an "_o" and and "_t" suffix. "_t" variables tell us the type of information that the respondent reported (continuous, closed band, open band etc.) and so provides information on the extent of the imputation that was required. "_o" variables contain information on how many observations were available to carry out imputation from. Observations where there were very few observations available to impute from should be used with caution.

Box 2 shows an example of the "_t" imputation flag for spen_r_i (respondent state pension income). The tabulation shows that 45.18% of the sample reported a zero value for state pension income. A further 48.97% do receive the state pension and reported a continuous value when asked how much they receive. For 73 observations (0.66% of the

sample), the respondent gave a closed band (i.e both an upper and lower bound - for example, more than £100 but less than £200). For 27 observations, the respondent gave an open band (that is a lower bound, but no upper bound – for example, more than £500). For 212 observations, although the respondent reported that they did receive the state pension, they were unable to report how much they receive even after the unfolding bracket questions. For a further 40 observations, we do not know if they receive the state pension or not.

The more information that the respondent gives us, the more reliable is the imputation. Users may wish to exclude some imputed values from their analysis or include imputation dummy variables to flag imputed values in multivariate analysis.

Box 2. Example of an "_t" imputation flag

```
. tab spen_r_t

  state pension income (iapam/iappam) -
                        imputation flag |      Freq.     Percent        Cum.
----------------------------------------+-----------------------------------
          not imputed (see user guide)  |        227        2.05        2.05
              institutional respondent  |         68        0.62        2.67
                                  zero  |      4,992       45.18       47.85
                            continuous  |      5,411       48.97       96.81
                           closed band  |         73        0.66       97.48
                             open band  |         27        0.24       97.72
                      missing, positive |        212        1.92       99.64
                    missing completely  |         40        0.36      100.00
----------------------------------------+-----------------------------------
                                  Total |     11,050      100.00
```

6.1 Imputation flags for summary variables and the "_ni#" variables

Table 4. Description of summary variables in the financial derived 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, the "_t" imputation flag is the aggregate of the "_t" variables for the individual components. For example, employment income (empinc_bu_s) is the sum of income from main employment (thp_bu_i) and from other jobs (oj_bu_i). A respondent who gave a continuous value for main employment but who gave a closed band for other job income, would have a closed band for total employment income.

For the Summary Variables, in addition to the "_t" imputation flag, there is a further flag which tells us how many components were imputed and how. These variables are given an "_ni#" suffix where the # denotes a number from 2 to 4. A suffix of "ni2" denotes that the variable tells us the number of components that were imputed from a "closed band", "open band", "missing, positive" or "missing completely" (that is type 2 or higher

from the "_t" variables). A suffix of "ni3" denotes that the variable tells us the number of components that were imputed from an "open band" , "missing, positive" or "missing completely" (that is type 3 or higher from the "_t" variables). A suffix of "ni4" denotes that the variable tells us the number of components that were imputed from a "missing, positive" or "missing completely" (that is type 4 or higher from the "_t" variables).

Box 3 gives an example of how the "ni#" variables should be interpreted. The example is shown for ppinc (private pension income). Private pension income is the sum of 3 separate components of income (respondent annuity income, partner annuity income and benefit unit level private pension income. The maximum number of components that could be imputed is therefore 3.

The tabulation of ppinc_bu_ni2 tells us that 88.57% of the sample had no components imputed at all. 7.80% of the sample had one component imputed from a closed band or "worse". 45 observations had all three of their components of private pension income imputed from a closed band or worse. The tabulation of ppinc_bu_ni3 tells us the number of components imputed from open bands or worse. So 90.48% had none of their components imputed from an open band or worse (but may have had imputation from a closed band) and 6.02% had 1 component imputed from an open band or worse. The tabulation of ppinc_bu_ni4 tells us the number of components imputed from completely missing information or missing but positive information. The _ni#" variables can be useful when thinking about how much imputation is "acceptable". This will vary according the research being carried out. For example, one might choose to exclude observations where a large proportion of the components of a particular source of income have been imputed. They can also be helpful to include in multivariate regressions which include the summary income or wealth variables.

Box 3. Example of "_ni" variables (ppinc – private pension income)

```
. tab ppinc_bu_ni2

no. of imputed components of
 annuitised income (type>=2) |      Freq.     Percent        Cum.
-----------------------------+-----------------------------------
not imputed (see user guide) |        227        2.05        2.05
    institutional respondent |         68        0.62        2.67
                           0 |      9,787       88.57       91.24
                           1 |        862        7.80       99.04
                           2 |         61        0.55       99.59
                           3 |         45        0.41      100.00
-----------------------------+-----------------------------------
                       Total |     11,050      100.00

. tab ppinc_bu_ni3

no. of imputed components of
 annuitised income (type>=3) |      Freq.     Percent        Cum.
-----------------------------+-----------------------------------
not imputed (see user guide) |        227        2.05        2.05
    institutional respondent |         68        0.62        2.67
                           0 |      9,998       90.48       93.15
                           1 |        665        6.02       99.17
                           2 |         51        0.46       99.63
                           3 |         41        0.37      100.00
-----------------------------+-----------------------------------
                       Total |     11,050      100.00

. tab ppinc_bu_ni4

no. of imputed components of
 annuitised income (type>=4) |      Freq.     Percent        Cum.
-----------------------------+-----------------------------------
not imputed (see user guide) |        227        2.05        2.05
    institutional respondent |         68        0.62        2.67
                           0 |     10,221       92.50       95.17
                           1 |        447        4.05       99.21
                           2 |         50        0.45       99.67
                           3 |         37        0.33      100.00
-----------------------------+-----------------------------------
                       Total |     11,050      100.00
```

7. SUMMARY VARIABLES

There are a number of variables in the data which aggregate the various components of income and wealth to create summary measures. These are likely to be the variables of most use to a large number of analysts. The summary variables all have an "_s" suffix so they can be identified easily and are described in Table 3. For more information on the relationships between the summary variables and their components see Figure 1 and Figure 2. For more detailed information about the source variables for each of the income and wealth components, see the spreadsheets for each Wave:

5050__Financial_Derived_Variable_Relationships.xls

Table 3. Description of summary variables in the financial derived variables

Variable	Description
Measures of income	
Total income	
totinc_bu_s	Total benefit unit income. This is the sum of: employment income (empinc_bu) self-employment income (seinc_bu) state benefit income (beninc_bu) state pension income (spinc_bu) private pension income (ppinc_bu) asset income (assinc_bu) other income (othinc_bu). Each member of the benefit unit is assigned total benefit unit level income.
eqtotinc_bu_s	The equivalised version (adjusted for benefit unit size) of totinc_bu.
Employment income	
empinc_bu_s	Benefit unit income from employment
eqempinc_bu_s	The equivalised version of empinc_bu
empinc_r_s	Respondent's income from employment (individual level)
empinc_p_s	Partner's income from employment
Self employment income	
seinc_bu_s	Benefit unit income from self-employment
eqseinc_bu_s	The equivalised version of seinc_bu
seinc_r_s	Respondent's income from self-employment (individual level)
seinc_p_s	Partner's income from self-employment
Annuitised Income	
ppinc_bu_s	Benefit unit annuitized income (private pensions and other

	annuity)
eqppinc_bu_s	The equivalised version of ppinc_bu
ppinc_r_s	Respondent's annuitised income
ppinc_p_s	Partner's private pension income
State Pension Income	
spinc_bu_s	Benefit unit state pension income
eqspinc_bu_s	The equivalised version of spinc_bu
spinc_r_s	Respondent's state pension income (individual level)
spinc_p_s	Partner's state pension income
State Benefit Income	
beninc_bu_s	State benefit income (benefit unit level)
eqbeninc_bu_s	The equivalised version of beninc_bu
beninc_r_s	Respondent's state benefit income (individual level)
beninc_p_s	Partner's state benefit income
Asset income	
assinc_bu_s	Income from assets (benefit unit level)
eqassinc_bu_s	The equivalised version of assinc_bu
Other income	
othinc_bu_s	Benefit unit other income (mainly income from people outside the household such as child support payments)
eqothinc_bu_s	The equivalised version of eqothinc_bu
othinc_r_s	Respondent's other income (mainly income from people outside the household such as child support payments)
othinc_p_s	Partner's other income (mainly income from people outside the household such as child support payments)
Measures of wealth	
savings_bu_s	Total savings (benefit unit level) (money invested in "safe" assets such as bank accounts, savings accounts and cash ISAs)
invests_bu_s	Total investments (benefit unit level) (money invested in "risky" assets such as shares, bonds, stocks and shares ISAs or life insurance ISAs)
grossfw_bu_s	Gross financial wealth (benefit unit level) (savings + investments but not subtracting any financial debt)
debt_bu_s	Financial debt (benefit unit level) (credit cards, overdrafts, other private debt but not mortgages)
netfw_bu_s	Net financial wealth (benefit unit level). Gross financial wealth with financial debt subtracted.
netpw_bu_s	Net physical wealth (benefit unit level). (Physical wealth is second homes, farm or business property, works of art etc)
grosstotnhw_bu_s	Total gross non-housing wealth (benefit unit level). Sum of savings, investments, and physical wealth but not subtracting any financial debt.
nettotnhw_bu_s	Total net non-housing wealth (benefit unit level). Sum of

	savings, investments, and physical wealth after financial debt is subtracted.
grosshw_bu_s	Gross housing wealth (benefit unit level). The value of owner occupied primary housing *before* mortgage debt is subtracted. Housing wealth is assigned only to the person or people named on the property and their partner. The value of the mortgage is not subtracted.
nethw_bu_s	Net housing wealth (benefit unit level). The value of owner occupied primary housing *after* mortgage debt is subtracted. Housing wealth is assigned only to the person or people named on the property and their partner.
mgdebt_bu_s	Mortgage debt on primary housing assigned only to the person or people names on the property (and their partner)
nettotw_bu_s	Net total wealth (benefit unit level). The sum of savings, investments, physical wealth and housing wealth after financial debt and mortgage debt has been subtracted.
Deciles and quintiles of income and wealth	
yq5_bu_s	Quintiles of total income (totinc_bu)
yq10_bu_s	Deciles of total income (totinc_bu)
nfwq5_bu_s	Quintiles of net financial wealth (netfw_bu)
nfwq10_bu_s	Deciles of net financial wealth (netfw_bu)
tnhwq5_bu_s	Quintiles of net total non-housing wealth (nettotw_bu)
tnhwq10_bu_s	Deciles of net total non-housing wealth (nettotw_bu)
totwq5_bu_s	Quintiles of net total wealth (nettotw_bu)
totwq10_bu_s	Deciles of net total wealth (nettotw_bu)

Each of the summary variables has a set of variables that give information about the level of imputation carried out for each individual or benefit unit. The variables with an "_t" and the variables with an "_f" suffix are described in Table 2. The Summary Variables have additional imputation flags which accompany them. These are variables with an "ni" prefix. These are described in Section 6.1.

The quantiles are defined by lining up all sample members according to their income or wealth and dividing them up into 5 (quintiles) or 10 (deciles) equally sized groups. In cross-section, income and wealth tend to fall with age, so older individuals are more likely to be found in the lower income or wealth quantiles. For this reason, users may wish to redefine quantiles when working with sub-groups of the population or where comparisons need to make within age group.

8. IDENTIFIERS

In addition to the usual individual and household identifiers found in the main data, there are two additional identifiers. These are “coupid” and “fuid”.

“Coupid” uniquely identifies couples. Each member of a couple has the same coupid except where one member of that couple is in an institution. Single people have their own coupid. 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”. For couples who keep their finances together, only one member of the couple will have answered IA so their partner will also be assigned the same fuid.

9. 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[3] and 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.

10. NOTE FOR PREVIOUS USERS (earlier than Wave 5) OF THE FINANCIAL DERIVED VARIABLES

The first 4 waves of financial derived variables were substantially updated with the release of the Wave 5 data. The differences between the old and new versions are as follows:

1) Minor errors corrected and minor improvements made to imputation procedures.

2) A change in the naming conventions of variables. Table 4 shows the mapping between the old and new naming conventions. This change was implemented so that it is easier to identify variables of different types. For example in Stata if you want a list of all the variables containing imputed values, you can type “describe *_i”. In addition, summary variables have an “_s” suffix to make it easier to identify them.

Table 4. Mapping between old and new suffixes

[3] See HAGENAARS, A., DE VOS, K. and ZAIDI, A. (1994), *Poverty Statistics in the late 1980s*, Eurostat, Luxembourg.

Old suffix	New suffix	Description
i	_r_i	Imputed value – at individual level for respondent
i_p	_p_i	Imputed value – at individual level for respondent's partner
i_bu	_bu_i	Imputed value – at benefit unit level
i_hh	_hh_i	Imputed value – at household level
t	_r_t	Type of observation – at individual level for respondent
t_p	_p_t	Type of observation – at individual level for respondent's partner
t_bu	_p_bu	Type of observation – at benefit unit level
t_hh	_hh_t	Type of observation – at household level
o	_r_o	No. of observations used for imputation – at individual level for respondent
o_p	_p_o	No. of observations used for imputation – at individual level for respondent's partner
o_bu	_bu_o	No. of observations used for imputation – at benefit unit level
o_hh	_hh_o	No. of observations used for imputation – at household level

3) Variables which used to have an "ni" prefix now have an "ni" suffix instead

4) A change in the way that missing values are represented. Previously, missing values were represented by codes of -10, -3 and -2 with each code being a different type of missing. To prevent users from accidently using these negative missing codes as genuine values, all missing values are now denoted by a system missing (a "." in Stata for example). To differentiate between the different types of missing, each imputed variable now has a flag variable associated with it which takes the value of 0 if the imputed value is valid, and a code of 1/2/3 to distinguish between the other sorts of missing (see Section 5).

5) A change in variable name for gross and net housing wealth. Previously net housing wealth was "nhwealth" and gross housing wealth was "ghwealth". This has been changed to "nethw" and "grosshw" which better matches the naming conventions for the other summary wealth variables.

6) A change in the source variable for private pension income (ppinc_bu_s). This change also affects total income measures and quantiles of income. In previous versions, information on private pension income was taken from IA (from a variable called iappei). A switch has been made to the more detailed information on private pension income in WP (from a variable called wppyr). The advantage of using the WP variables is that private pension income can be identified at the individual level and we have information on whether the income is before or after tax. There are also now two additional private pension income variables showing income at the individual level (wppp_r_i, wppp_p_i).

7) The addition of new variables. These variables include benefit unit level mortgage debt on primary housing, take home pay, self employment profit and drawings, odd job income, value of businesses for partners and individual level measures of the summary income variables (where available).

Table 5. A summary of the income and wealth measures in the financial derived data

Stem name	Description	Ind level	BU level	HH level	Summary	Component of summary
Employment income						
empinc	Total employment income	✓	✓		✓	
thp	Take home pay	✓	✓			✓
oj	Income from subsidiary jobs	✓	✓			✓
sinc	Wage and salary income (gross)		✓			
Self employment income						
seinc	Total Self employment income	✓	✓		✓	
netprof	Net profit	✓	✓			✓
sedraw	Self employment drawings	✓	✓			✓
Annuitised income						
ppinc	Total annuitised income (after tax)	✓	✓		✓	
ppen	Gross private pension income		✓			
wppen	Net private pension income	✓	✓			✓
anin	Other annuity income	✓				✓
State Pension Income						
spinc	Total income from State Pensions	✓	✓		✓	
spen	Income from State Pensions	✓				✓
State Benefit income						
beninc	Total State Benefit Income	✓	✓		✓	
icb	Incapacity Benefit	✓				✓
esa	Employment and Support Allowance	✓				✓
sda	Severe Disablement Allowance	✓				✓
ssp	Statutory Sick Pay	✓				✓
attall	Attendance Allowance	✓				✓
dla	Disability Living Allowance	✓				✓
indinj	Industrial Injuries Allowance	✓				✓
war	War Pensions	✓				✓
invcare	Invalid Care Allowance (Wave 1 only)	✓				✓
carers	Carer's Allowance (Wave 2 onwards)	✓				✓
dptc	Disabled Person's Tax Credit (Wave 1)	✓				✓
oth1	Other health benefits	✓				✓
is	Income Support	✓				✓

Name	Description	Ind level	BU level	HH level	Summary	Component of summary
pc	Pension Credit (Wave 2 onwards)					
wftc	Working Families Tax Credit (Wave 1)	✓				✓
wtc	Working Tax Credit (Wave 2 onwards)	✓				✓
jsa	Job Seeker's Allowance	✓				✓
gall	Guardian's Allowance	✓				✓
widpen	Widow's Pension	✓				✓
cb	Child Benefit	✓				✓
ctc	Child Tax Credit (Wave 2 onwards)	✓				✓
oth2	Other non-health benefits	✓				✓
Asset income						
assinc	Total Asset Income		✓		✓	
savei	Income from Savings		✓			✓
tessai	Income from TESSAs		✓			✓
isai	Income from ISAs		✓			✓
prbondsi	Income from Premium Bonds		✓			✓
nsavi	Income from National Savings		✓			✓
pepi	Income from PEPs		✓			✓
sharesi	Income from Shares		✓			✓
trustsi	Income from Trusts		✓			✓
bondsi	Income from Bonds		✓			✓
othsavi	Other savings income		✓			✓
homei	Rental income from second homes		✓			✓
farmi	Rental income from farm and business property		✓			✓
Other income						
othinc	Total other income	✓	✓		✓	
othpay	Other income (such as maintenance)	✓				✓
Housing Wealth						
nethw	Net housing wealth		✓		✓	✓
grosshw	Gross housing wealth		✓		✓	✓
mgdebt	Mortgage debt		✓		✓	✓
hdebt	Mortgage debt (HH level)			✓	✓	
hsval	House value (HH level)			✓	✓	
home	Second homes (net)		✓			✓
Financial wealth						
netfw	Net financial wealth		✓		✓	✓
grossfw	Gross financial wealth		✓		✓	✓
savings	Total Savings (non-risky assets)		✓		✓	✓

invests	Investments (risky assets)		✓		✓	✓
Name	Description	Ind level	BU level	HH level	Summary	Component of summary
cashisa	Cash ISAs		✓			✓
tessa	TESSAs		✓			✓
prbonds	Premium Bonds		✓			✓
nsav	National Savings		✓			✓
pep	PEPs		✓			✓
shares	Shares		✓			✓
trusts	Trusts		✓			✓
bonds	Bonds		✓			✓
lisave	Life insurance Savings component		✓			✓
shisa	Shares ISAs		✓			✓
lisa	Life insurance ISAs		✓			✓
othsav	Other Savings		✓			✓
jntass	Joint assets (separate finance couples)		✓			✓
Financial debt						
debt	Total financial Debt		✓		✓	✓
ccard	Credit card debt		✓			✓
prdebt	Private debt		✓			✓
odebt	Other debt		✓			✓
Physical Wealth						
netpw	Net physical Wealth		✓		✓	✓
home	Second homes (net)		✓			✓
busv	Value of business		✓			✓
othbusv	Other business wealth		✓			✓
farm	Farm or business property		✓			✓
tothass	Other physical assets		✓			✓
jntass	Joint assets (separate finance couples)		✓			✓

Figure 1. A diagram of the structure of Total Income (Waves 2 to 4)

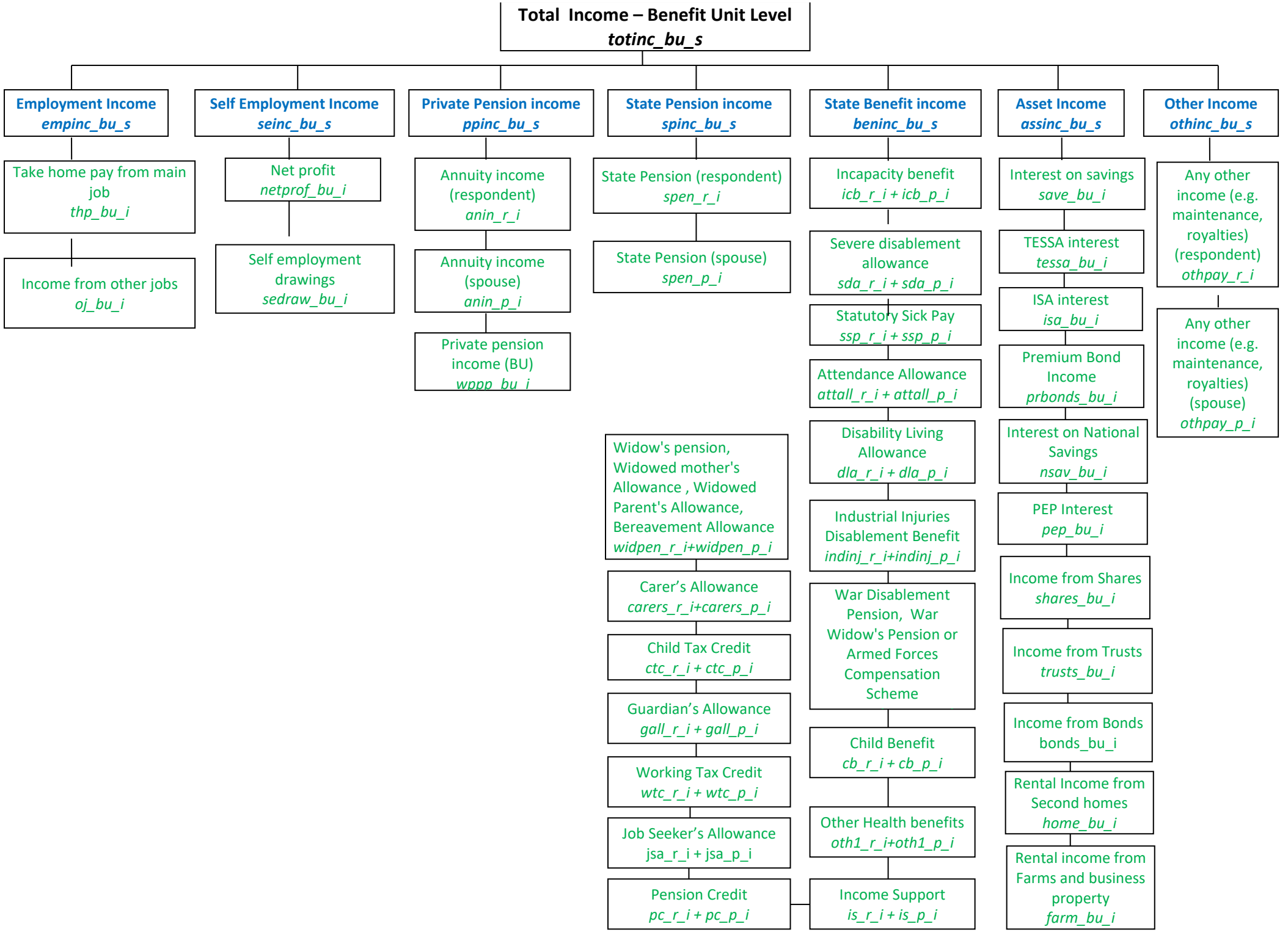


Figure 2. A diagram of the structure of Total (non-pension) Wealth (Waves 1 to 4)