ELSA Financial Derived Variables

Zoe Oldfield

Institute for Fiscal Studies
Overview

What are they?

What information do they contain?

How to use them?
The ELSA questionnaire

• Includes hundreds of questions regarding respondents’ economic situation (earnings, benefits, pensions, assets, debts, etc...)

• ... spread over several different modules:
  – Work and Pensions (WP)
  – Income and Assets (IA)
  – Housing (HO)

• Lots of flexibility offered to respondent in the way that they answer the questions
What are the financial derived variables?

• We’ve combed through ELSA’s many financial questions and
• transformed income flows into uniform periods (weekly)
• imputed amounts where we have incomplete or missing values
• converted amounts so they relate to a meaningful unit of observation
• added components together into economically meaningful aggregates (employment income, benefit income etc.)
• ...so that you don’t have to.
Which files are we talking about?

<table>
<thead>
<tr>
<th>Data files</th>
<th></th>
</tr>
</thead>
</table>
| Wave_X_financial_derived_variables | •Contains only information relating to financial variables  
•One observation per ELSA respondent |
| Wave_X_IFS_derived variables | •Contains other derived variables that are commonly used by IFS (and other) researchers  
•One observation per ELSA respondent |
<table>
<thead>
<tr>
<th>Documentation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>User_Guide_Financial_Derived_Variables.pdf</td>
<td>Basic user guide</td>
</tr>
<tr>
<td>User_Guide_IFS_Derived_Variables.pdf</td>
<td>Basic user guide</td>
</tr>
<tr>
<td>Financial_Derived_Variables_Relationships.xlsx</td>
<td>Detailed information about the source variables for each financial variables (one tab for each wave)</td>
</tr>
<tr>
<td>IFS Derived Variables Description.xlsx</td>
<td>Detailed information on how each non-financial derived variables is created</td>
</tr>
<tr>
<td>Financial_Derived_variables_and Imputation_procedures.pdf</td>
<td>A more in-depth description of the way that we impute missing data</td>
</tr>
</tbody>
</table>
## Which files are we talking about?

<table>
<thead>
<tr>
<th>Documentation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>User_Guide_Financial_Derived_Variables.pdf</td>
<td>Basic user guide</td>
</tr>
<tr>
<td>User_Guide_IFS_Derived_Variables.pdf</td>
<td>Basic user guide</td>
</tr>
<tr>
<td>Financial_Derived_Variables_Relationships.xlsx</td>
<td>Detailed information about the source variables for each financial variables (one tab for each wave)</td>
</tr>
<tr>
<td>IFS_derived_variables_description.xlsx</td>
<td>Detailed information on how each non-financial derived variables is created</td>
</tr>
<tr>
<td>Financial_Derived_variables_and_Imputation_procedures.pdf</td>
<td>A more in-depth description of the way that we impute missing data</td>
</tr>
<tr>
<td>Documentation</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>User_Guide_Financial_Derived_Variables.pdf</td>
<td>Basic user guide</td>
</tr>
<tr>
<td>User_Guide_IFS_Derived_Variables.pdf</td>
<td>Basic user guide</td>
</tr>
<tr>
<td>Financial_Derived_Variables_Relationships.xlsx</td>
<td>Detailed information about the source variables for each financial variables (one tab for each wave)</td>
</tr>
<tr>
<td>IFS_derived_variables_description.xlsx</td>
<td>Detailed information on how each non-financial derived variables is created</td>
</tr>
<tr>
<td>Financial_Derived_variables_and_Imputation_procedures.pdf</td>
<td>A more in-depth description of the way that we impute missing data</td>
</tr>
</tbody>
</table>
Q. What do we end up with?

A. Still a lot of variables to deal with!

But, the variables can be used as they are without further processing and once you know the naming conventions and the structure of the variables, it should (hopefully) be easier to work with!
Three categories of variable

• Single components of income/wealth
  – income support, incapacity benefit, ESA...
  – Take home pay, odd job income, private pension income, other annuity income...
  – ISAs, second homes, other assets...

• Aggregate measures of income/wealth
  – Total income, state benefit income, employment income, self employment income..
  – Total wealth, housing wealth, financial wealth..

• Auxiliary variables
  – Variables designed to give you extra information about each variable: why it’s missing, what the underlying quality of the variable is
General naming convention

Every variable in the financial derived variables is named in the following way:

\[ \text{stem}_1^{\text{st}} \text{suffix}_2^{\text{nd}} \text{suffix} \]
General naming convention

Every variable in the financial derived variables is named in the following way:

stem_1^{st}suffix_2^{nd}suffix

What the income or wealth component is
General naming convention

Every variable in the financial derived variables is named in the following way:

stem$_{1}^{st}$suffix$_{2}^{nd}$suffix

What the income or wealth component is

Who the income or wealth relates to
General naming convention

Every variable in the financial derived variables is named in the following way:

stem_1\textsuperscript{st}suffix_2\textsuperscript{nd}suffix

What the income or wealth component is

Who the income or wealth relates to

What type of variable it is
Example – state pension (respondent)

- `spen_r_i`
- State pension
- Of the respondent
- Value including imputed data
Example – state pension (partner)

\[
\text{spen}_p\_i
\]

- State pension
- Of the spouse
- Value including imputed data
1\textsuperscript{st} suffix can be:

<table>
<thead>
<tr>
<th>suffix</th>
<th>Relates to...</th>
</tr>
</thead>
<tbody>
<tr>
<td>_r</td>
<td>The respondent</td>
</tr>
<tr>
<td>_p</td>
<td>The partner of the respondent</td>
</tr>
<tr>
<td>_bu</td>
<td>The benefit unit.</td>
</tr>
<tr>
<td>_hh</td>
<td>The household</td>
</tr>
</tbody>
</table>
1\textsuperscript{st} suffix can be:

<table>
<thead>
<tr>
<th>suffix</th>
<th>Relates to...</th>
</tr>
</thead>
<tbody>
<tr>
<td>_r</td>
<td>The respondent</td>
</tr>
<tr>
<td>_p</td>
<td>The partner of the respondent</td>
</tr>
<tr>
<td>_bu</td>
<td>The benefit unit.</td>
</tr>
<tr>
<td>_hh</td>
<td>The household</td>
</tr>
</tbody>
</table>
1st suffix can be:

<table>
<thead>
<tr>
<th>suffix</th>
<th>Relates to...</th>
</tr>
</thead>
<tbody>
<tr>
<td>_r</td>
<td>The respondent</td>
</tr>
<tr>
<td>_p</td>
<td>The partner of the respondent</td>
</tr>
<tr>
<td>_bu</td>
<td>The benefit unit.</td>
</tr>
<tr>
<td>_hh</td>
<td>The household</td>
</tr>
</tbody>
</table>
What is a “benefit unit”?

• Technically we are referring to a unit that is used for the assessment of state benefits
• It is a couple or a single person (plus any dependent children they may have)
• A “household” (all individuals living together) can have multiple benefit unit
What if they don’t know the amount?

- Respondent is asked a series of follow up questions designed to extract as much information as possible
- Of the form “is it more or less than £X?”
- Respondent can say “don’t know” at any stage
- We end up with:
  - Closed band (between £X and £Y)
  - Open band (more than £Y)
  - “Missing positive”
  - “Missing completely”
2\textsuperscript{nd} suffix can be:

<table>
<thead>
<tr>
<th>suffix</th>
<th>What it means</th>
</tr>
</thead>
<tbody>
<tr>
<td>_i</td>
<td>These variables contain the (imputed if necessary) value of the income or wealth</td>
</tr>
</tbody>
</table>
| \_t    | These variables indicate what information the respondent gave us: 0 – zero (no income/wealth) 1 – continuous answer 2 – closed band 3 – open band 4 – missing positive 5 – missing completely  

*Used only for components of income or wealth*
| \_o    | The number of observations that the imputation was based on |
| \_f    | Validation flag – gives the reason why a variable is missing (i.e. not imputed) |
| \_s    | Indicates that the variable is an aggregate (summary) measure of income or wealth |
| \_ni2  | Number of components of a summary measure of income/wealth that were of type 2 (closed band), 3 (open band) or 4 (missing positive) or more  

*Used only for aggregate (summary) measures of income or wealth*
### 2nd suffix can be:

<table>
<thead>
<tr>
<th>suffix</th>
<th>What it means</th>
</tr>
</thead>
<tbody>
<tr>
<td>_i</td>
<td>These variables contain the (imputed if necessary) value of the income or wealth</td>
</tr>
</tbody>
</table>
| _t     | These variables indicate what information the respondent gave us:  
0 – zero (no income/wealth)  
1 – continuous answer  
2 – closed band  
3 – open band  
4 – missing positive  
5 – missing completely  
   *Used only for components of income or wealth* |
| _o     | The number of observations that the imputation was based on |
| _f     | Validation flag – gives the reason why a variable is missing (i.e. not imputed) |
| _s     | Indicates that the variable is an aggregate (summary) measure of income or wealth |
| _ni2 _ni3 _ni4 | Number of components of a summary measure of income/wealth that were of type 2 (closed band), 3 (open band) or 4 (missing positive) or more  
   *Used only for aggregate (summary) measures of income or wealth* |
### 2nd suffix can be:

<table>
<thead>
<tr>
<th>suffix</th>
<th>What it means</th>
</tr>
</thead>
<tbody>
<tr>
<td>_i</td>
<td>These variables contain the (imputed if necessary) value of the income or wealth</td>
</tr>
</tbody>
</table>
| _t     | These variables indicate what information the respondent gave us:  
0 – zero (no income/wealth)  
1 – continuous answer  
2 – closed band  
3 – open band  
4 – missing positive  
5 – missing completely  

*Used only for components of income or wealth* |
| _o     | The number of observations that the imputation was based on |
| _f     | Validation flag – gives the reason why a variable is missing (i.e. not imputed) |
| _s     | Indicates that the variable is an aggregate (summary) measure of income or wealth |
| _ni2   | Number of components of a summary measure of income/wealth that were of type 2 (closed band), 3 (open band) or 4 (missing positive) or more  

*Used only for aggregate (summary) measures of income or wealth* |
### 2nd suffix can be:

<table>
<thead>
<tr>
<th>suffix</th>
<th>What it means</th>
</tr>
</thead>
<tbody>
<tr>
<td>_i</td>
<td>These variables contain the (imputed if necessary) value of the income or wealth</td>
</tr>
<tr>
<td>_t</td>
<td>These variables indicate what information the respondent gave us:</td>
</tr>
<tr>
<td></td>
<td>0 – zero (no income/wealth)</td>
</tr>
<tr>
<td></td>
<td>1 – continuous answer</td>
</tr>
<tr>
<td></td>
<td>2 – closed band</td>
</tr>
<tr>
<td></td>
<td>3 – open band</td>
</tr>
<tr>
<td></td>
<td>4 – missing positive</td>
</tr>
<tr>
<td></td>
<td>5 – missing completely</td>
</tr>
<tr>
<td></td>
<td>Used only for components of income or wealth</td>
</tr>
<tr>
<td>_o</td>
<td>The number of observations that the imputation was based on</td>
</tr>
<tr>
<td>_f</td>
<td>Validation flag – gives the reason why a variable is missing (i.e. not imputed)</td>
</tr>
<tr>
<td>_s</td>
<td>Indicates that the variable is an aggregate (summary) measure of income or wealth</td>
</tr>
<tr>
<td>_ni2</td>
<td>Number of components of a summary measure of income/wealth that were</td>
</tr>
<tr>
<td>_ni3</td>
<td>of type 2 (closed band), 3 (open band) or 4 (missing positive) or more</td>
</tr>
<tr>
<td>_ni4</td>
<td>Used only for aggregate (summary) measures of income or wealth</td>
</tr>
</tbody>
</table>
What measures of income and wealth are available?

• In total there are:
  – 45 components of income
  – 31 components of wealth
  – 7 aggregate (summary) measures of income

• Best place to find out what those are is Table 5 in the User_guide_Financial_derived_variables_.pdf
  – Tells you whether available individual/benefit unit/household level
  – Tells you what category of income or wealth it is
  – Tells you whether it’s a component or summary measure of income or wealth
What’s included in the FDVs?

Summary measures of income

- **eqtotinc_bu_s** = Total income adjusted for family size (‘equivalised’)
- **totinc_bu_s** = Unadjusted (‘unequivalised’) total income
- Income quintiles and deciles also available:
  - **yq5_bu_s** = Equivalised income quintiles
  - **yq10_bu_s** = Equivalised income deciles
- Users may wish to recalculate quintiles/deciles when working with subgroups
  - To do this you rank individuals according to their income (within subgroup) and split into equal groups
What’s included in the FDVs?

Summary measures of income

Total income (totinc_bu_s)
What’s included in the FDVs?

Summary measures of income

Total income (totinc_bu_s)

- Employment (empinc_bu_s)
- Private pensions (ppinc_bu_s)
- Benefits (beninc_bu_s)
- ‘Other’ (othinc_bu_s)

Self-employment (seinc_bu_s)

State pension (spinc_bu_s)

Assets (assinc_bu_s)
What’s included in the FDVs?

Summary measures of wealth

Total (non-pension) wealth

$\text{nettotw\_bu\_s}$

$\text{(totwq5\_bu\_s)}$
What’s included in the FDVs?

Summary measures of wealth

- Total (non-pension) wealth
  \[ \text{nettotw}_{bu_s} \]
  \[ \text{(totwq5}_{bu_s} \]

- Net non-housing wealth
  \[ \text{nettotnhw}_{bu_s} \]
  \[ \text{(tnhwq5}_{bu_s} \]

- Net primary housing wealth
  \[ \text{nethw}_{bu_s} \]
What’s included in the FDVs?

Summary measures of wealth

- Total (non-pension) wealth
  \( \text{nettotw\_bu\_s} \)
  \( (\text{totwq5\_bu\_s}) \)

- Net non-housing wealth
  \( \text{nettotnhw\_bu\_s} \)
  \( (\text{tnhwq5\_bu\_s}) \)

- Net primary housing wealth
  \( \text{nethw\_bu\_s} \)

- Gross primary housing wealth
  \( (\text{grosshw\_bu\_s}) \)

  less

- Primary housing debt
  \( \text{mgdebt\_bu\_s} \)
What’s included in the FDVs?

Summary measures of wealth

- Total (non-pension) wealth: \( \text{nettotw}_{bu} \) (\( \text{totwq5}_{bu} \))
  - Net non-housing wealth: \( \text{nettotnhw}_{bu} \) (\( \text{tnhwq5}_{bu} \))
  - Net financial wealth: \( \text{netfw}_{bu} \) (\( \text{nfwq5}_{bu} \))
  - Net primary housing wealth: \( \text{nethw}_{bu} \)
  - Net physical wealth: \( \text{netpw}_{bu} \)

- Gross primary housing wealth: \( \text{grosshw}_{bu} \)
  - Primary housing debt: \( \text{mgdebt}_{bu} \)

© Institute for Fiscal Studies
What’s included in the FDVs?

Summary measures of wealth

- Total (non-pension) wealth
  $\text{nettotw} \_ \text{bu}_s$
  $(\text{totwq5} \_ \text{bu}_s)$

- Net non-housing wealth
  $\text{nettnhw} \_ \text{bu}_s$
  $(\text{tnhwq5} \_ \text{bu}_s)$

- Net primary housing wealth
  $\text{nethw} \_ \text{bu}_s$

- Net financial wealth
  $\text{netfw} \_ \text{bu}_s$
  $(\text{nfwq5} \_ \text{bu}_s)$

- Net physical wealth
  $\text{netpw} \_ \text{bu}_s$

- Gross primary housing wealth
  $(\text{grosshw} \_ \text{bu}_s)$

- Gross financial wealth
  $\text{grossfw} \_ \text{bu}_s$

- Financial debt
  $\text{debt} \_ \text{bu}_s$

© Institute for Fiscal Studies
What's included in the FDVs?

Summary measures of wealth

- Total (non-pension) wealth
  \[ \text{nettotw\_bu\_s} \]
  \[ (\text{totwq5\_bu\_s}) \]

- Net non-housing wealth
  \[ \text{nettotnhw\_bu\_s} \]
  \[ (\text{tnhwq5\_bu\_s}) \]

- Net primary housing wealth
  \[ \text{nethw\_bu\_s} \]

- Net financial wealth
  \[ \text{netfw\_bu\_s} \]
  \[ (\text{nfwq5\_bu\_s}) \]

- Net physical wealth
  \[ \text{netpw\_bu\_s} \]

- Gross primary housing wealth
  \[ \text{grosshw\_bu\_s} \]

- Gross financial wealth
  \[ \text{grossfw\_bu\_s} \]

- Financial debt
  \[ \text{debt\_bu\_s} \]
Which measure of income or wealth should I use?

• This really depends on the research question

• Good place to start is the summary variables
  – Remember these end in “_s”

• Also look at:
  – Table 5 in the FDV user guide
  – Figures 1 and 2 in the FDV user guide
  – Financial-derived_Variable_Relationships.xls (and filter by category of income)
Which variables in the questionnaire do the derived variables correspond to?

• This is documented in: Financial_Derived_Variable_Relationships.xls

• It’s also noted in the variable description

  trusts_bu_i: “BU Trusts (iauit) - value (incl. imputed values)”

This variable is based on iauit
Is there an individual measure of income?

• The short answer is no
  – ELSA is designed to measure income at the benefit unit level
  – All summary measures relate to the benefit unit level

But...

• Some components of income are collected at the individual level
  – These are measures which are easily assignable to an individual (state benefits, employment income etc.)
Important things to remember

• All income flows are measured weekly
• There are equivalised and unequivalised measures of aggregate income..
• ..but not wealth
• The unit of observation is the individual but measures of income/wealth (generally) relate to the benefit unit.
Example household

- Age 85
- Age 55
- Age 57
- Age 25
- Age 27
Example household

Benefit unit 1

Age 85

Age 55  Age 57

Age 25  Age 27
Example household

Benefit unit 1

Benefit unit 2

Age 85

Age 55

Age 57

Age 25

Age 27

Age 57
Example household

Benefit unit 1

Age 55
Age 57

Benefit unit 2

Age 85
# Example household

<table>
<thead>
<tr>
<th>Benefit unit</th>
<th>Age</th>
<th>totinc_bu_s (total income)</th>
<th>£</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>57</td>
<td>£300</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>55</td>
<td>£300</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>85</td>
<td>£150</td>
<td></td>
</tr>
</tbody>
</table>
## Example household

<table>
<thead>
<tr>
<th>Benefit unit 1</th>
<th>Total income</th>
<th>Take home pay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 57</td>
<td>£300</td>
<td>£220</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Benefit unit 1</th>
<th>Total income</th>
<th>Take home pay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 55</td>
<td>£300</td>
<td>£50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Benefit unit 2</th>
<th>Total income</th>
<th>Take home pay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 85</td>
<td>£150</td>
<td>£0</td>
</tr>
</tbody>
</table>
## Example household

<table>
<thead>
<tr>
<th>Benefit unit</th>
<th>Age</th>
<th>Totinc_bu_s (total income)</th>
<th>thp_r_i (take home pay)</th>
<th>thp_bu_i (take home pay)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>57</td>
<td>£300</td>
<td>£220</td>
<td>£270</td>
</tr>
<tr>
<td>1</td>
<td>55</td>
<td>£300</td>
<td>£50</td>
<td>£270</td>
</tr>
<tr>
<td>2</td>
<td>85</td>
<td>£150</td>
<td>£0</td>
<td>£0</td>
</tr>
</tbody>
</table>
Non-financial derived variables

- Purpose: centralise the derivation of some of the more complex variables and reduce duplication of effort across the research community
- Includes additional identifiers, children variables, family type, labour market participation, summary pension variables, summary health, expectations, expenditure
Much more in the documentation

- More detail on how the variables were imputed
- More detail on how the variables were derived
- How to use the auxiliary variables
- More detail on the way the data are collected
- Information about the pension wealth variables
- Please come and ask me if you have questions
Funding from the NIA and the ESRC (Centre for Microeconomic Analysis of Public Policy at IFS) is gratefully acknowledged. Data from the English Longitudinal Study of Ageing (ELSA) were made available by the UK Data Archive. ELSA was developed by a team of researchers based at University College London, University of Manchester, the Institute for Fiscal Studies and the National Centre for Social Research, with funding provided by the National Institute of Aging in the United States, and a consortium of UK government departments co-ordinated by the Office for National Statistics.

Thank you!