LFS User Workshop

Martina Helme / Simon Woodsford

ONS Social Survey Operations – Labour Force Survey Research Team

20 May 2021

Session 2

Considerations when using the data

- Datasets - scope/limitations

- What to know about weights

- What to know about variables

- Data linking/pooling

Simon Woodsford / Martina Helme

Dataset overview

- **LFS person level** data - quarterly
- **LFS household level** data - now all quarters
- **LFS longitudinal** data – 2 quarters and 5 quarters
- **APS person level** data - quarterly, covering 12 months
- **APS household level** data – annually
- **APS 3 year pooled** data – annually

- Government level files
- Public files (now only end user licence)

- Data accessible from UK Data Archive/SRS

Office for **National Statistics**

LFS Person level data

Purpose → Personal level analysis at a national or regional level

Frequency → Quarterly – official released around 6 weeks after the end of data collection period, available through UKDS/SRS several weeks later

Time periods → Normally released for calendar quarters only (e.g. Jan to March)

Availability → AJ92 to AJ01 (2007 weight)

JS01 to AJ11 (2014 weight)

JS11 to OD19 (2018 weight)

JM20 to current (2020 weight)

Weights → PWT (person analysis) and PIWT (earnings analysis)

Limitations → Not for analysis below regional level

No earnings information for self employed

Some variables are specific to certain quarters

Not for Well-Being, Sexual Orientation, Family/Household analysis

Office for **National Statistics**

LFS longitudinal data for 2 quarters

Purpose → Conducting analysis on flows over 3 months

Frequency → Quarterly – official released around 6 weeks after the end of data collection period, available through UKDS/SRS several weeks later

Time periods → Two calendar quarters only (e.g. JM to AJ)

Availability → Q197-Q297 to Q101 –Q201 (2007 weight)
Q201-Q301 to Q211-Q311 (2014 weight)
Q311-Q411 to Q319-Q419 (2018 weight)

Q419-Q120 to current (2020 weight)

Weights → LGWT

Limitations → Only available for person level analysis
Only includes respondents who responded in both quarters
Only includes variables asked in both quarters

Doesn't contain all variables available on LFS person datasets

Office for **National Statistics**

LFS longitudinal data for 5 quarters

Purpose → Conducting analysis on flows over 12 months

Frequency → Quarterly – official released around 6 weeks after the end of data collection period, available through UKDS/SRS several weeks later

Time periods → Five calendar quarters only (e.g. JM to AJ)

Availability → Q1'05-Q1'06 to Q2'10-Q2'11 (2010 weight)
Q3'10-Q3'11 to Q2'11 to Q2'12 (2014 weight)
Q2'12 –Q3'13 to Q4'18 to Q4'19 (2018 weight)

Q1'19 –Q1'20 onwards (2020 weight)

Weights → LGWT

Limitations → Only available for person level analysis
Only includes respondents who responded in all 5 quarters

As a result sample size drastically reduced
Doesn't contain all variables available on LFS person datasets

Office for **National Statistics**

APS Person level data

Purpose → Personal level analysis below regional level or for breakdowns on certain characteristics

Frequency → Quarterly – official released around 11 weeks after the end of data collection period, available through UKDS/SRS several weeks later

Time periods → Rolling 12 months (e.g. January to December, April to March)

Availability → JD04 to JD11 (2014 weights)

A11M12 to JD19 (2018 weights)

A19M20 onwards (2020 weights)

Weights → PWTA (A05M onwards, before then PWAPS, PWLFS)

PIWTA (for earnings analysis) – JD only from 2012-17, all periods from 2018 onwards

SIDWT (for Sexual Orientation analysis) – 2011 onwards

NPWT (for Personal Well-being analysis) – 2011 onwards

Limitations → Some LFS quarterly specific variables not available on APS

Office for **National Statistics**

APS longitudinal data for 2 years

Purpose → Personal level analysis on flows below regional level over 12 months

Frequency → Annual

Time periods → January to December for two consecutive years

Availability → JD12 - JD13 to JD18 - JD19 (2018 weight)

JD19 - JD20 (2020 weight)

Weights → LGWT

Limitations → Only at person not household level
Only for a sub-set of variables
Only includes respondents who have responded in both years

Office for **National Statistics**

APS 3-year pooled person level data

Purpose → Personal level analysis at lower level geographies and for certain topics where achieved sample is much smaller

Frequency → annual – normally released around July every year

Time periods → January to December covering 3 years

Availability → JD13 to JD15 (2016 weight)
JD14 - JD16 to JD15 - JD17 (2017 weight)
JD16 - JD18 to JD17 to JD19 (2018 weight)

JD18 to JD20 (2020 weight)

Weights → PWTA**C, SIDWT**C and NPWT**C

Limitations → Point in time estimates and not to be used for any time series analysis

Office for **National Statistics**

LFS Household level data

Purpose → household/family level analysis at a national or regional level

Frequency → Quarterly - Officially released around 8 weeks after the end of data collection period, available through UKDS/SRS several weeks later

Time periods → Pre 2004 – Apr-Jun periods only;
Between 2004 and 2011 Apr-Jun and Oct-Dec only;
Quarterly from AJ12 onwards

Availability → AJ96 to AJ01 (2007 weight)
AJ02 to AJ11 (2014 weight)
OD11 to OD19 (2018 weight)
JM20 to current (2020 weight)

Weights → PHHWT

Limitations → Not suitable for household earning analysis as no earnings household weight

Office for **National Statistics**

APS Household level data

Purpose → Household and Family level analysis below regional level or for breakdowns on certain characteristics

Frequency → Annually - released around July every year

Time periods → January to December

Availability → JD04 to JD05 (2010 weight)
JD06 to JD11 (2014 weight)
JD12 to JD19 (2018 weight)

JD20 (2020 weight)

Weights → PHHWTA

Limitations → Some LFS variables are not available on APS datasets, particularly LFS quarterly specific variables.
Not suitable for household earning analysis as no earnings household weight

Office for **National Statistics**

What to know about LFS/APS weights

- LFS/APS collects data on a representative sample of UK population

- Grossing weights are created to estimate for the rest of the UK population we don't sample and the survey design

- Historically weighting groupings have been by age, sex and geography

- Since the pandemic we've added Tenure and CoB adjustments to the weighting process

- Historically each individual represents around 800 people of similar characteristics (LFS person weights)

- Sum of the weights make up the eligible population in the UK

- Critical to make sure you use the appropriate weight for the analysis you are doing

Office for **National Statistics**

What to know about variables

- Datasets contain questionnaire variables; derived variables; geography variables; system variables, all can be used for analysis

- Details of all variables can be found in UG2, UG3 and UG4

- Not all variables will be on all datasets

  - Well-Being; Sexid; Smoking; Veterans; General Health; Quality of work – APS only

  - Family and Household variables – Household datasets only

  - Quarterly specific variables – LFS datasets only

  - Reduced variables list on Longitudinal datasets

- Understand the variables you are using (e.g. periods; type; routing)

- Datasets will include 'missing values'

  - -9 means not routed to that specific question

  - -8 means respondent refused to answer question or answered don't' know

Office for **National Statistics**

Data Pooling

- It isn't advised to carry out any dataset pooling yourself

- Pitfalls of doing this include no suitable weights and duplicate cases

- The LFS and APS datasets available should allow you to carry out all permitted analysis

- Caution with any data linking across waves

  - ✓ A unique case identifier (not available on EUL datasets)

  - ✓ Sex

  - ✓ DoB

- Name and address information not available on any LFS/APS datasets

- Pseudo identifiers available on LFS UKDS datasets

Office for **National Statistics**

Which dataset to use when?

- Analysis on sexual orientation → APS person (SIDWT)
- Analysis on wellbeing → APS person (NPWT)
- Analysis on earnings below reg. level → APS Person (PIWTA)
- Analysis on other topics below reg. level → APS Person (PWTA)

- Analysis on earnings at or above reg. level → LFS Person (PIWT)
- Analysis on other topics at or above reg. level → LFS Person (PWT)

- Longitudinal analysis below regional level → APS Person 2yr (LGWT)
- Longitudinal analysis at or above regional level → LFS 2Q or 5Q (LGWT)

- Analysis at hhold/family level below reg. level → APS Hhold (PHHWTA)
- Analysis at hhold/family level at or above reg. level → LFS Hhold (PHHWT)

Office for **National Statistics**

How to get started…

- Always refer to LFS/APS user guides first:
  - ➢UG1 – LFS (and some APS) Methodology
  - ➢UG2 to UG4 – Questionnaire and Derived Variables
  - ➢UG5 – Classifications (SIC/SOC)
  - ➢UG 6 - APS
  - ➢UG8 – Household and family variables
  - ➢UG9 – Eurostat
  - ➢UG10 – Advice on which dataset to use
  - ➢ UG11 – Longitudinal

- Key things depending on your analysis:
  - ➢Use the correct dataset
  - ➢Use the correct weight
  - ➢Use the correct variables

Office for **National Statistics**