

Labour Force Survey User Meeting

A Labour Force Survey User Group Meeting was held at the Royal Statistical Society in London on Friday 26 November 2004. Barry Werner (Labour Market Division, Office for National Statistics) and Angela Dale (ESDS Government) chaired the meeting.

LFS Reweighting and The Impact on Micro Dataset Users

Stephen Hicks (Labour Market Division, Office for National Statistics)

Stephen Hicks explained the LFS reweighting process following the 2001 Census. He illustrated the differences between estimates derived from microdata and the interim revised series which are used in ONS publications. The problem originated from the 2001 Census showing that previous UK population estimates are around 1 million too high. The LFS sample is weighted to the UK household population so the LFS data had to be reweighted to keep results in line with the latest population estimates. To avoid a lengthy process an 'interim re-weighting project' was established to provide results in line with latest population estimates on a timely basis.

LFS headline estimates were calculated as usual from the micro datasets (weighting to old population estimates). Adjustment factors (new population estimates divided by old) were calculated and applied to produce interim adjusted estimates (headline Labour Market Statistics estimates only). Interim reweighting took place in five stages over the period November 2002 to October 2004.

The reweighting process has some implications for users of microdata. Estimates drawn directly from microdata are not in line with latest population estimates so they will differ to the interim adjusted series published in ONS outputs (e.g. monthly Labour Market Statistics First Release). Although the differences in UK headline labour market estimates are small, for sub groups and smaller geographies within UK the relative differences can be large.

To allow microdata users to see the extent of the differences for a wide range of LFS variables a special dataset for Summer 2001 was produced with two weighting variables. The dataset shows the differences between interim adjusted estimates (published in March 2004) and estimates directly from micro datasets (in line with population estimates published in Feb/Mar 2003). Estimates of aged 16+ populations for Summer 2001 taken directly from the current LFS micro datasets are around 200,000 lower than the interim adjusted estimates published in March 2004. The differences can be seen in headline levels, headline rates, age, sex, GOR and country. It is advised that the following footnote is added to any tables stating that they are based directly on microdata "*The levels and rates published in this analysis are derived from LFS microdata and will not match the interim-reweighted levels published in the ONS Labour Market Statistics First Release and LFS Quarterly Supplement tables, as they are not based on the most up-to-date population estimates.*" A 'LFS Reweighting Toolkit', which contains the Summer 2001 LFS Toolkit Microdata and a full explanation of the reweighting, can be found at www.statistics.gov.uk/downloads/theme_labour/Toolkit_article11.pdf.

Comparing labour market data from the 2001 Census with the LFS ***Dan Heap (Office for National Statistics)***

Dan Heap presented an overview of the differences between labour market data taken from the 2001 Census and that from the Spring 2001 Labour Force Survey. The LFS covers England, Scotland and Northern Ireland whereas the Census covers England and Wales, Scotland and Northern Ireland separately. The LFS has an achieved sample of 60,000 UK households but the Census has 100% coverage and a 98% response rate. The LFS is a voluntary sample survey compared with the compulsory Census. The LFS mainly uses telephone interviews and is carried out 'quarterly' at 13-week intervals. The Census uses self-completion questionnaires and is carried out every 10 years. Data in the LFS is available at national, regional and local authority level but the Census can provide data for very small geographies (e.g. ward level). The necessary simplicity of the Census questions limits its scope and the lack of an interviewer prevents confirming information. There are also some differences in question wording between the two sources. The Census uses different coding methods to the LFS for occupation and industry. The LFS suffers from sampling variability.

Comparisons were made between LFS-based and Census-based labour market indicators such as economic activity, economic inactivity, occupation and industry, hours worked, region, age, ethnicity and qualifications. For example, the largest differences in Occupation were found in Professional Occupations (+9.7%) and Personal Services (+8.9%). In Industry, the largest differences were found in Other Services (+14.6%), Agriculture and fishing (-13.8%) and Manufacturing (+13.2%). There was a 4.6% difference between the LFS and the Census for Hours Worked; the LFS may exclude overtime but the Census includes overtime.

In general, the Census represents a point in time measure of the labour market. The LFS is a more reliable source of national and regional data which uses international conventions and is capable of doing long time series. However, the LFS suffers from sampling variability.

An Introduction to the Annual Population Survey ***Tim Burrell (Office for National Statistics)***

Tim Burrell explained that the LFS aims to provide better quality data on key social and socio-economic variables such as housing, employment, ethnicity, education and health, particularly at a local level. This will improve monitoring of key variables between Censuses for a range of policy purposes and will provide local area information as part of the Neighbourhood Statistics Programme (NeSS).

January 2004 marks the start of the Annual Population Survey Boost Sample (APS) with approximately 65,000 achieved household interviews a year. When combined with results from the Labour Force Survey and English Local Labour Force Survey, the survey will provide enhanced annual data for England covering a target sample of at least 500 economically active persons for each Unitary Authority/Local Authority District and at least 450 economically active persons in Greater London Boroughs, where these targets are not already met by the LFS and sample boosts in England, Wales and Scotland.

The APS Boost will focus on key Census-type topics which are also included in the existing LFS. It will not include all topics covered by the LFS and boosts, resulting in a shorter average interview length and, hopefully, higher levels of response.

The plan is to provide a single annual local area database, including the interviews from the new survey, to replace the existing local area LFS database. Various cuts of the dataset will then be produced from the 'master' database to meet different customer requirements, as under current arrangements. The inclusion of the new interviews in the annual local area database will enable ONS to provide enhanced local area information on the National Statistics web site and via NOMIS. These data will also be included as part of the Annual Local Area Labour Force Survey (ALALFS) publication. As now, special tables will be available from the ONS-LFS Sub-Regional Table Service. Databases with a reduced set of variables and/or geographical restrictions will also be made available to departments on request, subject to approval by the ONS Microdata Release Panel.

The proposal is for the first new annual local area database to cover the period January-December 2004 and for subsequent annual databases to apply to calendar years. This fits in with the switch to calendar quarter databases for the main LFS already announced, from January 2005. It is hoped that the first new annual database and results for the Neighbourhood Statistics web site will be available in June 2005. Further information about the APS is available on the ONS web site.

Current Developments in the LFS

Alex Clifton-Fearnside (Office for National Statistics)

(Presented by Andrea Grabham, Office for National Statistics)

(1) Forecasting of early LFS estimates: currently the LFS is published on average every three months, using data for 13 consecutive weeks. It takes about six weeks after the end of the survey period to publish the results so it is less timely than benefit-related measures such as claimant count. One of the key recommendations of the 2002 LFS Quality Review is "Early" forecast of LFS estimates. This project is led by ONS Methodology. It uses 11 weeks of LFS data for each quarter and modelled data for the final two weeks. The aim is to publish provisional estimates one month earlier and revised estimates at the standard time. This involves a complex Box Jenkins model, using ARIMA regression. It has been tested on grossed weekly LFS data and the initial results are very positive. However, it needs to be tested on weekly weighted data independent of LFS weighting procedures and is still in the development stage.

(2) Single month LFS estimates: one of the recommendations of the LFS Quality Review was that ONS should investigate the production of estimates using one month of LFS data only. ONS developed an analytical tool, in response to movements in the rolling 3-month LFS data, which is based on applying interim LFS methods to ungrossed LFS data. This is a good indicator of movements in the key LFS series. The initial approach is very crude (aged 16+ and gender). A weighting factor is calculated in each month by dividing the total population by the sample population. The factor is then applied to monthly ungrossed estimates, e.g. employment, which are seasonally adjusted by using single

months. The improved methodology uses a 6 x 2 weighting matrix (gender and six age-bands) as for Interim LFS estimates. Weighting factors are calculated as the total population divided by the sample population for each cell. Factors are applied to monthly ungrossed employment, unemployment etc. by age and sex. It is seasonally adjusted and constrained. There is consistency in the single month estimates despite different weighting methods which illustrates a good indicator of quality. The work is published as experimental analysis only. Regional single month estimates and estimates of employment by sex and age are being developed. Trend estimates will also be produced.

(3) *Calendar Quarters*: The LFS is moving to Calendar Quarters from January 2006 as part of a Eurostat requirement for all EU countries (the survey as well as the results). The move will have no effect on the rolling monthly aggregate series. ONS are developing plans to deal with issues such as a consistent back series. The move raises weighting issues because of the differences between Calendar Quarter and Seasonal Quarter weights. Also, it is not possible to produce back series for quarter specific questions, or for 1984-1991 data. There are also resource issues, particularly computing resources.

(4) *Quarterly Population Estimates*: Quarterly Population Estimates have been developed by ONS using available administrative data (births, deaths and migration) and some forecasting of components (migration). The estimates are awaiting the agreement of the National Statistician to publish as experimental statistics. This is for England and Wales only as there are issues to be resolved before a UK series is available. The LFS population is currently based on a straight line between mid-year estimates. If the LFS sample is representative of the population then the LFS sample is seasonal. ONS are starting a project to analyse the effect on the LFS aggregate series. There is no plan to use Quarterly Population Estimates whilst they are experimental. There are significant resource issues in implementing this at a time when ONS is committed to re-engineering.

Outline for a ONS microdata access and release strategy ***Paul Jackson and Richard Laux (Office for National Statistics)***

A draft policy for ONS microdata access and release was agreed by the Statistical Policy Committee (SPC) in July 2004. The policy distinguishes between non-disclosive microdata that can be released as a statistical product and disclosive (confidential) microdata that can be accessed for statistical purposes. The policy outlines four possible solutions for best access:

- *licensed use via an archive*: this involves non-disclosive social survey data being deposited at the UK Data Archive (UKDA) as a matter of course. Suitable for low-risk microdata only. However, the full potential of the UKDA is not realised and it could offer more flexibility than the supply of non-disclosive data. Thus two additional levels of access can be used. First, professional users who want more detail than is currently on a dataset could apply for the data and sign an additional agreement without ONS approval. Users who want possibly disclosive data (such as date of birth or small areas) would follow the same procedure but ONS would have to specifically agree the request. In addition, the user and their institution would agree to additional conditions

- *off-site, access agreements*: Suitable for medium risk microdata. Confidential data leaves ONS but with meaningful control
- *off-site, remote access*: Suitable for high-risk microdata. Confidential data does not leave direct ONS control. Access is either by remote job submission or by direct query. This is not a popular option for expert research users
- *on-site, laboratory/secondment*: Suitable for high-risk microdata. Confidential data does not leave ONS. It involves access within a 'safe centre' where the organisational and technical environment is wholly and directly owned and controlled by ONS. This is an inconvenient option for many users unless there are labs distributed around the country.

A Microdata Release Panel has been established to regulate all release/access to ONS microdata and a Microdata Access and Release Strategy is in preparation. The microdata access and release strategy may include the following:

- present microdata access as a single corporate service via the ONS web site
- provide a spectrum of solutions based on a matrix of *user type, purpose and disclosiveness of data*
- make microdata access a horizontal specialism to include a single 'microdata access team' drawn from existing ONS services to deliver user screening, advice training, output checking and a single point of contact for users
- build a remote access lab and on-site data labs into the design for the Corporate ONS Repository of Data (CORD) and Metadata (CORM)
- seek funding support from the wider research community.

Born in Britain: How are native ethnic minorities faring in the British labour Market?

David Blackaby (University of Wales Swansea), D.G. Leslie (Manchester Metropolitan University), P.D. Murphy (University of Wales Swansea) and N.C. O'Leary (Manchester Metropolitan University)

David Blackaby explained how he and colleagues used the LFS data to explain two problems that Britain's ethnic minorities face in the labour market: finding employment and having lower earnings than similar white individuals. Three major factors may explain the above differences:

- ethnic minorities may have lower levels of characteristics that increase earnings and employment, such as education
- for any given characteristic, earnings and employment may be lower for ethnic minorities. This will be referred to as discrimination
- while all ethnic minority groups may have experienced discrimination, their response to it may have been different.

Decomposition analysis can be used to explain the differences in earnings and employment due to characteristics of individuals e.g. sex or age. Factors influencing employment and earnings can be controlled for and the differences can then be examined to see if discrimination is at play. QLFS data from 1993 (quarter 3) to 2000 (quarter 4) were pooled to obtain a sample of non-white *British born* ethnic minorities. Foreign-born cases were excluded to control for factors such as poor command of language, overseas schooling etc. The sample was restricted to economically active male employees. Separate earnings functions were estimated for each ethnic group. Employment and wage

differences were decomposed into the amount due to characteristic differences and discrimination. The approaches of Neumark (1988) and Oaxaca and Ransom (1994) were adopted. The conclusions of the research included:

- ethnic differences in the labour market cannot be explained solely as a characteristics problem
- for British born ethnic groups, particularly Black and Pakistani groups, employment and earnings disadvantage is particularly severe. This may partly reflect a taste for isolation, rather than just a consequence of discrimination
- additional discrimination legislation may be required as well as policies to encourage certain ethnic groups to engage more fully with the dominant culture.

Identifying the British

John MacInnes (University of Edinburgh and Univeristat Autonoma de Barcelona)

John MacInnes used QLFS pooled data (2000 quarter 1 to 2004 quarter 4) to examine the problems with ethnic group and national identity questions. Nationality and ethnicity are not capable of definition because the same terms mean different things to different respondents. For example, in the LFS data only 37% of those born in Britain identified themselves as British. Within the LFS questionnaire the term 'British' has seven different meanings including pertaining to UK state, pertaining to mainland British state (excluding Ireland), white or pertaining to cultural or other background qualifying group defined by skin colour. This means that the data analyst cannot be certain which of these meanings the respondent had in mind when they answered the question. Questions and context can influence responses; 89% of respondents in Britain answered 'British' to the ethnicity question but only 37% answered British to the national identity question. Questionnaire design, such as the order of categories and the presence of a 'tick box' for separate countries such as Wales, can also influence response. John's conclusions were that Censuses and surveys should ask about ethnicity but they do crystallise categories so the analysis and presentation of 'ethnic' results should be better monitored.

Population at risk measures and entrants to unemployment

Ray Thomas (Open University and Campion Research Fellow, RSS)

Ray Thomas discussed population at risk measures and entrants to unemployment. Population at risk (PAR) unemployment rates measure the proportion of a duration group remaining unemployed. PAR exit-from-unemployment rates measure the proportion of a duration group exiting from unemployment. Conventional rates indicate unemployment as a social or economic condition. PAR rates indicate the labour market response to unemployment. PAR rates can be applied to any definition of unemployment but the usefulness of PAR statistics depends upon the quality of statistics. The LFS measures ILO unemployment but does not measure entry to unemployment or short term unemployment. Ray suggested that the LFS needs a question on unemployment addresses to *all* respondents. For example 'Have you been unemployed at any time during the past six months'. Ray also suggested that the LFS needs a less behavioural definition of unemployment.

Measuring labour market participation for ethnic minority groups

Jo Lindley (University of Sheffield), Angela Dale (University of Manchester) and Shirley Dex (University of Cambridge)

Angela Dale explained that most ethnic minority groups have higher unemployment rates than their white counterparts, even after controlling for qualifications and social and economic differences. The ILO definition of unemployment omits people who would like to work but are not actively seeking work and it includes the employed who are claiming unemployment benefit and are also looking for full-time work. These two factors are known to impact differently on men and women but the differences between ethnic groups are not known. Data on all men and women aged 19-65 were used from the 1992-2003 LFS (Spring quarter) to explore why women are consistently less likely to be recorded as ILO unemployed than white men and whether or not this applies across all ethnic groups. Results from logit models for those who were ILO unemployed but would like to work were as follows:

- Black Caribbean men were more likely to be seeking employment than their white counterparts
- men with children were less likely to be seeking work than those without
- both men and women with a partner not in the labour force were more likely to be seeking work than those with a partner in the labour force
- women with young children were less likely to be seeking work than those without.

The following conclusions were made:

- the ethnic unemployment gap increases for men when 'hidden' unemployment is included
- the ethnic unemployment gap decreases for Black Caribbean/Black Other women but increases for other groups
- the gender gap decreases for all groups
- home and family are the most important reasons for not seeking among women who want to work
- sickness is the most important reason for not seeking among men who want to work.