

# **Time Series and Cross-Sectional Analysis and Modelling in the Monitoring of UK Labour Market**

**Marta F Haworth**  
**Office for National Statistics**  
**United Kingdom**

**David Caplan**  
**Office for National Statistics**  
**United Kingdom**

## **Introduction**

This paper<sup>1</sup> discusses the methodological issues arising for the UK Office for National Statistics (ONS) in developing an integrated system of labour market statistics drawing together data from both surveys and administrative data. It describes recent changes to the presentation of labour market statistics in the UK and looks in particular at the use of the UK labour Force Survey (LFS). It also describes the programme of current methodological work linked to the LFS which is designed to enhance the quality of information available on the UK labour market.

## **New Improved ONS Labour Market Statistics**

UK labour market statistics are of vital importance to a very wide range of users. They have a direct impact on decision-making, on the formulation of macro-economic and social policies and on policy evaluation. In recent years, there has been a great deal of public debate and criticism of official labour market statistics in the UK. Following a wide-ranging public consultation, in early 1998 the ONS launched a major programme of improvements, the aim of which was to develop a more integrated and more user-relevant statistical picture of the UK labour market, bringing together data from all available sources. This covers household and employer surveys, censuses and administrative records. Strong methodological underpinning is of crucial importance to strengthening public confidence. The methodological work discussed in this paper is a strategically important part of the development of improved ONS labour market statistics. Much greater prominence is now given to the UK Labour Force Survey (LFS) than previously.

## **The UK Labour Force Survey (LFS)**

The UK LFS is a continuous survey designed to survey around 60,000 households each quarter and produce publishable UK level estimates quarterly. A sample rotation scheme is used in which four fifths of the households selected in a particular week are included in the survey 13 weeks later. Altogether, a household is surveyed 5 times at 13-week intervals. This design was introduced in 1992 and also conforms to the

---

<sup>1</sup> **Note:** *The views expressed in this paper are of the authors and not those of the ONS.*

requirements of the 1998 European Council Regulation governing continuous EU labour force surveys yielding quarterly results.

### *Monthly Reports from the UK Labour Force Survey*

There is a very strong user demand for timely estimates of short-term changes, particularly for employment and unemployment and for sub-categories of economic inactivity. The design of the LFS permits production of monthly estimates, but the volatility of these estimates is very high as they are based on a smaller sample which is geographically clustered and there is no sample in common between two consecutive months. However, the design does mean that estimates can be produced for any period of 13 consecutive weeks using an unclustered sample of 60,000 households. Thus, it is possible to produce each month an estimate of the average unemployment in the three months up to and including the latest reference month. This approach has been implemented in the UK since April 1998 (previously the survey results have only been produced as averages for consecutive quarters). The new monthly system of (overlapping) moving average estimates have the same sampling standard errors as the quarterly estimates. The estimates of change from the previous 13 weeks also have the same standard errors as the estimates of change between consecutive quarters.

Within the constraint of the established design, the UK is thus using the continuous survey as a source of monthly series constructed from three months moving averages. The estimates are inevitably subject to significant lags. Sampling variability, particularly for estimates of changes in levels of unemployment but also for employment and inactivity is relatively large. Sampling variability is an even more serious issue at regional and small area levels. In order to overcome these problems, the ONS has set up new methodological work on trend estimation and construction of model-based estimates for small areas. Another important area of methodological work is the construction of longitudinally linked data sets and analyses, which are needed in response to strong user demand for data on gross flows. For the LFS, this means exploiting the panel element of the repeated cross-sectional survey and dealing with the bias issues involved.

### **Linking Data from Different Sources**

The new improved system of labour market statistics brings together data from all available sources in order to produce an integrated statistical picture of the labour market. The assessment is presented in a monthly Integrated First Release for the media and users more generally.

### *The Integrated First Release*

The release replaces the previously separate releases with statistics from administrative data, employer surveys and the LFS. In addition to the overall picture of economic activity, employment and unemployment, there is an analysis of inactivity and different

types of labour market attachment. The presentation recognises that no single indicator can capture the full complexity in the labour market.

It covers:

- labour market structure;
- economic activity and inactivity; those wanting and not wanting a job;
- employment;
- unemployment;
- earnings and productivity;
- working patterns and hours;
- vacancies;
- labour disputes; and
- sub-national and international comparison

#### *Sub-national statistics*

There is very strong user interest in regional and small area labour market statistics. In response to this, new integrated regional releases draw on both administrative and survey data, including the LFS. A great deal of user attention focuses on small area unemployment rates.

The ONS has undertaken a public consultation and a review of the basic approach for calculating unemployment rates for small areas. This has concentrated on the geographical areas for which unemployment rates should be calculated and on the basis for calculating them. User responses indicated that at least two approaches are needed:

- residence-based rates, derived from LFS data on unemployment measured according to the ILO definition divided by the economically active population residents in the area (the sum of the ILO unemployed residents in the area plus employed residents).
- Workplace-based rates derived from the administrative count of claimants of unemployment-related benefits divided by a measure of jobs using employer survey data on those working in the area, together with the number of self-employed, those on government supported training and employment programmes, the armed forces and the number of claimants.

#### **The Methodological Programme**

There are three main components of the major new programme: – trend analysis, small area estimation and construction of longitudinally linked data sets and analyses. The first two components seek to deal with the LFS sampling variability issue. The aim of trend analysis address is to aid interpretation of movements in direct survey estimates especially given the relatively large standard errors of LFS estimates of changes in the

main aggregates such as employment, inactivity and particularly unemployment. It is also seeking to respond to user need for detecting underlying short-term movements and turning points. The small area estimation project is intended to reduce the errors in direct survey estimates by constructing model-based estimators linked to LFS data. The longitudinal methodological work is seeking to exploit the panel element of the LFS for estimating changes in gross flows between key labour market status categories over time. These three areas of current methodological work are discussed further below.

## **Trend Estimation**

### *ONS Standard Trend Estimation: Levels*

The labour market statistics First Release was the first to use the standard ONS method for estimating short-term trends in levels (Kenny & Knowles, 1997). The recommended approach is to apply a Henderson moving average, augmented by two stages of outlier detection and ARIMA modelling. For labour market statistics, this is applied to the seasonally adjusted series of LFS employment and ILO unemployment (Hastings, 1999). The choice of Henderson filter to apply to the series depends on the I/C (irregular to cycle) ratio. (The I/C ratio compares the movement attributable to the irregular part of the series to the movement attributable to the trend.) The trend estimates are presented in the Integrated First Release together with margins of uncertainty for the latest three periods. These “trumpets” are derived by looking at the effect on current trend estimates of the likely future range of observations.

### *Survey Estimates and Trend Analysis*

In parallel with the publication of headline trend series for employment and unemployment rates using the standard approach, the ONS is also investigating the properties of moving averages constructed from quarterly LFS data and the trend series derived from the monthly survey observations. (The monthly release of latest three-month averages is, of course, not the same as releasing monthly estimates.)

The general approach is as follows: Let  $Y_t$  be the population value of the characteristic of interest, for example total unemployment for month  $t$ . Each month is defined as the 4 or 5 weeks that approximate it. Let  $X_t$  denote the value that would be obtained if the series of values  $Y_t$  was seasonally adjusted. When data for month  $t$  have been collected the estimate of average monthly change that is highlighted is  $(\tilde{y}_{t-1} - \tilde{y}_{t-4})/3$  which is unbiasedly estimating  $[(Y_t - Y_{t-3})/3 + (Y_{t-1} - Y_{t-4})/3 + (Y_{t-2} - Y_{t-5})/3]/3$ . An alternative is to use the change in the overlapping 3 month average,  $\tilde{y}_{t-1} - \tilde{y}_{t-2}$ , which unbiasedly estimates  $(Y_t - Y_{t-3})/3$  as an indication of current direction of the series. The corresponding changes in the seasonally adjusted series can also be examined. Analysis of the overlapping change will indicate changes earlier, but has higher standard errors than the estimate obtained from the non-overlapping 3 month average.

An additional method of analysing trends to calculate the smoothed trend cycle estimate produced by X-11 (i) using monthly estimates from the LFS, (ii) the monthly published series of non-overlapping three months averages. These estimates may assist in interpreting the underlying trends and changes. The usefulness and reliability, in particular the expected degree of revision, of these estimates are being evaluated.

### *Estimating Underlying Changes*

In responding to user demand for an improved assessment of short-term changes the ONS has drawn on recent research by Professor A Harvey (Cambridge University, UK) in exploratory analytical work on estimating underlying changes over time. This uses time series models in state space form. The latest underlying changes are computed using the Kalman filter. The methodology allows linking the LFS three months averages data on unemployment measured according to the ILO definition with the monthly counts of claimants of unemployment-related benefits. The underlying change is defined as the slope in the stochastic trend component. The early indications from the analyses to date are encouraging and suggest that this approach yield valuable additional information, which helps in interpretation of headline series. The models effectively provide a way of weighting the survey data in constructing estimates of change and as such have an intuitive appeal and presentational advantages. Measures of precision of the estimates can also be constructed.

### **Small Area estimation**

The ONS methodological work on small area estimation is designed to produce more robust estimates than those produced directly from the LFS. The aim is to use model-based estimation to reduce the variability of estimates of key labour market series, with initial work focusing on unemployment. Three approaches are currently being investigated. Multilevel modelling and SPREE (structure preserved estimation) use information from administrative data on the count of those claiming unemployment related social security benefits to generate cross-sectional, LFS-linked estimates. A third approach uses time series analysis.

### *Multilevel Modelling*

An approach based on multilevel modelling is being used to estimate unemployment at the level of local authority district (LAD), of which there are around 400 in great Britain (Wang et al, 1999). LADs are just one of a range of possible geographies which could have been chosen for analysis. they were selected because data are fairly easily obtainable at the LAD level and there is customer interest in data at this level.

The general model applied is of the form

$$y_{ij} = \beta_{0j} + x_{ij}^t \beta_{ij} + \varepsilon_{ij}$$

The  $y_{ij}$  are the proportions of unemployed in LADs  $j=1$  to  $n$  in various subcategories  $i=1$  to  $m_j$ . The betas are parameters which are estimated and the  $x$ s are covariates. A model of this form with randomness in the  $\beta_j$  allows for small area effects. Various assumptions about distributions and independence are required when fitting such a model.

Work to date has demonstrated that it is necessary to transform the data and a log model has been found to produce reasonable results. The covariates used have been information on the numbers of people claiming unemployment related benefits. Data on both the response and explanatory variables has been split into two sex categories, each of which has been split into three age groups. Interaction effects have also been found to be significant in the model.

There are a number of practical issues with this type of modelling. The model has to be fitted using the estimated values from the LFS. These are measured with potentially large error. Further, for unemployment, the estimates for some sub-categories are zero as no unemployed people will have been sampled. This presents a problem when fitting a log model and an imputation step is required.

Users require an estimate of the precision of the estimates derived from the model. An approach to this has been developed. However, the current prediction intervals are rather larger than expected. It is not clear whether this represents a problem with the model or with the estimation of the error ranges. What is clear is that although the model fits well and satisfies the assumptions, its performance will always be dependent on the quality of the basic data - the problem with small sample sizes at small areas can not entirely be overcome by adopting a modelling approach.

#### *Structure Preserving Estimation (SPREE)*

This method makes use of auxiliary information from the count of claimants of unemployment-related benefit – (“the claimant count”). The basic assumption is that the unemployment distribution in a small area can be explained by a log-linear model with interaction effects, which are constant over the small areas. The combination of variables in the model has been determined from exploratory analysis of the probability of concordance between claimants status and ILO unemployment status of individuals, using the data set resulting from a UK record linkage matching study (see Pease (1997)).

Data from the claimant count are obtained for the small areas cross-classified by the chosen combination of variables - sex by age group. Control totals of ILO unemployment from the LFS are obtained for the same cross-classification for a higher geography such as regions. An iterative proportional fitting algorithm is used to obtain estimates of ILO unemployment for the small areas. Experience to date suggests that the association structure is stable but the results for particular areas can be seriously biased.

The advantage of the SPREE approach is that it is very easy to implement and can produce quick results. However, there is uncertainty over the reliability of the results - variance estimation is difficult and the results are thought to be biased.

#### *Time series modelling approach*

This follows the approach developed by Professor Danny Pfefferman of the Hebrew University of Jerusalem and used on US, Australian, and Canadian monthly LFS data (Pfefferman, D., Bell, P. and Signorelli, D (1996); Pfeffermann, D., Feder, M., and Signorelli, D. (1997). This method utilises the finding that LFS sampling errors for small areas are likely to be large and strongly autocorrelated due to the rotational design.

Estimates of these survey error autocorrelations are calculated and used to identify models. The ONS is working with a relatively short series (24 quarters from Spring 1992) which is likely to reduce the reliability of the results. Models identified for the survey errors and for the population values of interest can be combined and arranged in state space form. Kalman filter equations can be used to derive the estimates for any month using all data up to and including that month. Because of the practical difficulties with short runs of data, this is considered only as a possible long-term development, although properties of time estimates derived by the main method over time may be studied.

#### **Longitudinal analyses of LFS data**

Because the same sample households are interviewed in five consecutive quarters, the LFS is potentially a rich source of data on flows constructed by linking together the information on the same persons and households across consecutive observation. This could cover:

- the numbers of people moving in both directions between employment, unemployment and economic inactivity;
- the kinds of entry-level of jobs taken by those previously unemployed;
- the effectiveness of different jobsearch strategies; and
- the destinations of people leaving government training schemes.

The process of following an individual from quarter to quarter in the LFS is relatively straightforward. However, there are potentially serious risks of distortion in the results. Such problems have been identified in other countries' labour force surveys. The ONS is therefore undertaking a programme of methodological work to address these problems, in order to be able to make available longitudinally linked data sets and analyses available regularly for general use (Tate, P., 1999).

The key results to date are:

- A methodology has been developed which compensates satisfactorily for the effect of attrition non-response bias on the LFS. It has been established that this does not introduce distortion into the distribution of labour force flows.
- However, initial findings of investigations into the effects of response error (e.g. individuals reporting a status change when the status has remained the same) suggest that this is likely to affect the longitudinal datasets, probably in the direction of an upward bias in estimates of gross flows between different economic activity categories. Some simulation work is planned for the short-term and further methodological research for the longer term.
- While these investigations continue, longitudinal datasets incorporating compensation for non-response bias are being produced on an experimental basis, and are being made available on request to experienced LFS users with whom ONS is working to develop and assess improved methods of analysing and presenting the data.

## **Other Developments**

### *ONS analytical work*

These improvements in ONS labour market statistics are supported by a great deal of work in improving the methods of data collection, getting more out of the data, and reconciling the different sources.

### *Reconciliation of labour market statistics*

Explaining differences between sources is an aid to understanding and improving the interpretation of labour market developments. It is also the first step towards investigating the feasibility of labour accounts, in which all labour market data are brought together into one coherent system. Work on reconciling the sources includes comparing employment estimates from the LFS with data from employer surveys and assessing the quality of LFS estimates of claimants of unemployment-related benefits using record linkages, as well as comparing the LFS and New Earnings Survey (NES) estimates of earnings. (Pease (1997), Wilkinson (May 1998 and Dec 1998), Stuttard, Tiwana, Partington (1998).

### *Improving data from employer surveys on earnings, jobs and business*

The ONS is also reviewing the methodology underpinning the calculation of the underlying rate of change in earnings and is evaluating the feasibility of different conceptual approaches. Other developments involving employment survey data are also taking place. The monthly and quarterly surveys collecting employer data or jobs have been merged with the monthly and quarterly surveys collecting turnover data. The collection of both employment and turnover data from the same businesses, coupled with an increase in the proportion of the business population covered by the survey,

means that a new range of statistical products can now be produced, notably improved productivity statistics.

### **Conclusion**

The ONS developmental programme is of great importance to raising public confidence in these vital and very high profile statistics. The methodological work plays a key strategic role in this programme. The ONS experience indicates that extensive user education and documentation has to be provided to support the programme and gain user confidence in the results.

## References:

Compton, S., (1998) Estimating and Presenting Short-Term Trends. Published in Economic Trends, **April 1998, Issue No. 533**

Harvey, A., and Chia-Hui Chung (1999) Report on Estimating the Underlying Change in UK Unemployment – technical report by Professor Andrew Harvey of the Faculty of Economics and Politics, University of Cambridge.

Hastings, D., (1999) – Trend Estimation of Labour Force Survey Data. Published in Labour Market Trends (August 1999).

Kenny, P., & Knowles, J., (1997): An investigation into trend estimation methods. A technical report available from Methods and Quality Division, ONS.

Pease, P., (1997) LFS estimates of claimants of unemployment related-benefits: results of an ONS record linkage study published in Labour Market Trends (Nov. 1997).

Pease, P., (1997) Comparison of sources of employment data. Published in Labour Market Trends (December 1997).

Pfeffermann, D., Bell, P. and Signorelli, D. (1996). Labour Force trend estimation in small areas. *Proceedings of the Bureau of Census Annual Research Conference*.

Pfeffermann, D., Feder, M., and Signorelli, D. (1997). Estimation of Autocorrelations of Survey Errors with Application to Trend Estimation in Small Areas, *Journal of Business and Economic Statistics*.

Stuttard, N., Tiwana, H., Partington, J., (1998) Industry comparisons of employment data. Published in Labour Market Trends (October 1998)

Tate, P. (1999) – Longitudinal data from the Labour Force Survey, published in Labour Market Trends (July 1999)

Wang, S., Douglas, A., Caplan, D., and Chambers, R (1999) – Small Area Estimation of Unemployment in Great Britain - Paper to be presented to International Association of Sample Surveys conference on Small Area Estimation, Riga (Aug 1999).

Wilkinson, D., (1998) – Towards reconciliation of NES and LFS earnings data. Published in Labour Market Trends (May 1998).

Wilkinson, D., (1998) – How are the low paid? Published in Labour Market Trends (December 1998).