Proceedings of the 22nd GSS Methodology Symposium 2017

Methodology: Insight; Innovation; Implementation; Impact

**Increasing frequency and improving timeliness of unemployment estimates from the UK Labour Force Survey**

Duncan Elliott[[1]](#footnote-1)

**Abstract**

Unemployment is estimated from data collected in the Labour Force Survey (LFS). While data is collected continuously the survey design is structured in such a way as to provide quarterly estimates. These quarterly estimates are published each month as rolling quarterly estimates and have been assessed to be of sufficient quality to be designated as National Statistics. ONS also publish monthly estimates of unemployment, but these are designated Experimental Statistics due to concerns over the quality of these data. Due to the sample design of the LFS, monthly estimates of change are volatile as there is no sample overlap. A state space model can be used to develop improved estimates of monthly change, accounting for aspects of the survey design, and so providing increased frequency and a slight improvement in timeliness. Additional sources of information related to unemployment could be used within such a framework to help improve timeliness further. In this paper an overview of the modelling approach is provided and issues associated with regular publication of results are discussed.

Key Words: time series; state space models; unemployment; repeated surveys.

**1. Introduction**

**1.1 UK Labour Market Publications**

The Office for National Statistics (ONS) publishes regular monthly information on the labour market. The headline Labour Force Survey estimates for unemployment are rolling-quarterly estimates. The unemployment variable is a time averaged stock calculated over a rolling three month period for the number of people classified as unemployed. These rolling-quarterly estimates have sufficient quality to be designated as National Statistics.

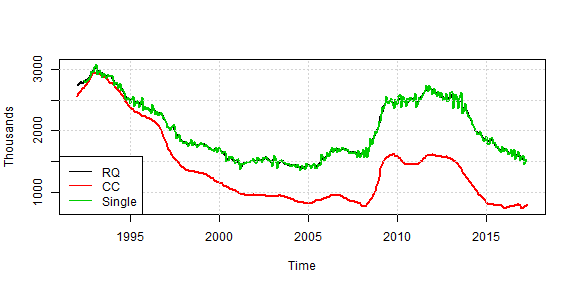
There is an additional experimental publication of monthly estimates of unemployment but these are not of sufficient quality to be designated National Statistics and do not form part of the main headline figures. These experimental monthly estimates are published in a separate article and are classified as Experimental Statistics. The main purpose of this publication is to facilitate understanding of some of the movements observed in the rolling-quarterly estimates.

The time difference between the publication date and the mid-point of the latest reference period for the rolling quarterly data is approximately three months. For example on 14 June 2017 rolling quarterly estimates referred to the period February to April 2017. The headline figures typically include the level of unemployment, the change in the level of unemployment between the most recent three month period and the previous (non-overlapping) three month period and also the same three month period a year ago. For example, the publication on 14 June 2017 stated that “between November 2016 to January 2017 and February to April 2017 the number of unemployed people fell” and that “there were 1.53 million unemployed people” with “50,000 fewer than for November 2016 to January 2017” and “145,000 fewer than for a year earlier”. These figures are often picked up directly in the media.

The experimental statistics do not get the same amount of coverage in the media as they are published in a separate article and are compared to the headline rolling-quarterly data. Figures 1.1.1 and 1.1.2 plot some of the available time series on a monthly basis. Included in these figures are estimates of the claimant count, which is administrative data on the number of people claiming benefits for not being in work, as counted on the second Thursday of the month. In both figures the single month estimates are considerably more volatile than the rolling-quarterly and claimant count data. However, figure 1.1.2 shows the single month estimates are more timely, and the claimant count more timely still.

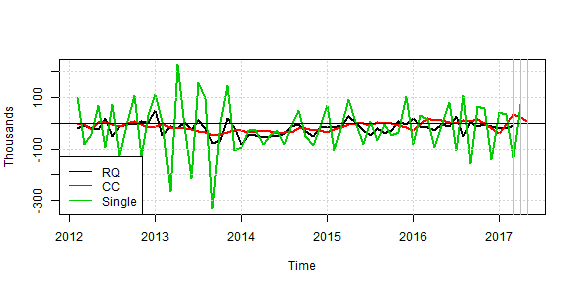
**Figure 1.1.1**

**Seasonally adjusted estimates of UK unemployment for ages 16+ and claimant count February 1992 – March 2017 published in June 2017**



**Figure 1.1.2**

**Monthly difference of seasonally adjusted estimates of UK unemployment for ages 16+ and claimant count February 2012 – March 2017 with vertical lines indicating the midpoints of final reference period for each series as published in June 2017.**



**1.2 Issues to address**

Two key dimensions of quality for important regular economic indicators such as unemployment are timeliness and accuracy. Two issues with the current publication of unemployment estimates are the timeliness of the rolling-quarterly data and the accuracy of the experimental monthly data and whether the timeliness of this could be improved to match that of the claimant count.

In order to address these issues a time series model has been developed to improve the timeliness and accuracy in particular of monthly changes in unemployment. The time series model developed accounts for the survey design and potentially allows the incorporation of administrative data similar to Harvey and Chung (2000).

**2. Time Series Model for Labour Force Survey variables**

A state space model for unemployment is developed that accounts for particular features of the Labour Force Survey. The key features that are addressed are described in section 2.1 with the model presented in section 2.2 and outputs from the model discussed in section 2.3.

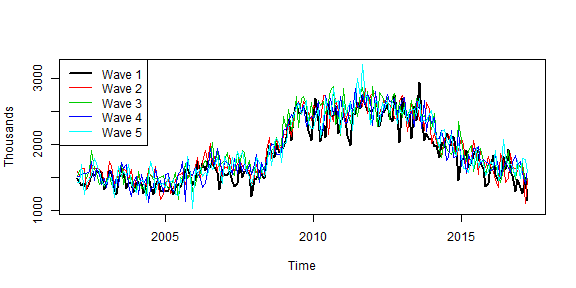
**2.1 Features of the Labour Force Survey**

Respondents to the Labour Force Survey (LFS) are interviewed five times (waves) at three month intervals. This means that there is approximately an eighty per cent overlap between consecutive quarters or between months at a three month lag. For monthly estimates there is no sample overlap between consecutive months which means that accuracy of monthly change is poor compared to estimates of quarterly change where the covariance due to sample overlap improves the accuracy of estimates of change. There is in theory still some overlap at an annual lag which should help to improve accuracy of estimates of change over a year. Unfortunately response rates have been declining over time. With declining response rates we can expect a decline in accuracy of estimates, which is important to incorporate into the time series model.

It is possible to create monthly wave specific estimates of total unemployment as shown in figure 2.1.1. These estimates use similar methods to the rolling quarter estimates but due to the smaller sample size within a month simpler calibration groups are required for estimation. As can be seen these estimates are volatile and there may also be, on average, level differences which could be due to bias due to the mode of collection (typically wave 1 interviews are face to face while subsequent waves are telephone interviews).

**Figure 2.1.1**

**Monthly wave specific estimates of UK unemployment ages 16 plus.**



**2.2 Proposed model**

The proposed model is a multivariate time series model using monthly wave-specific time series and their estimated

sampling errors as inputs. Estimates of correlation between these wave-specific time series at three month lags (due to the sample overlap) are also incorporated into the model. The model assumes that unemployment has a stochastic unobserved population process () that follows a Basic Structural Model as shown in equations 1-9. The survey error for each wave-specific estimate () includes a wave-specific bias term () and an autocorrelated sampling error () which is modelled as an autoregressive process of order (*i*-1) allowing for changing variance over time by standardising the error terms estimated in the model using the design-based estimates of the variance of the design-based estimators ().

(1)

(2)

(3)

(4)

(5)

(6)

(7)

(8)

(9)

The are estimated using the pseudo survey error autocorrelation method described by Pfeffermann *et al* (1998). Equations 1-9 are put into a state space framework with hyperparameters (variance terms in equations 3,4,5,7 and 9) estimated using maximum likelihood. The Kalman filter and smoother can be used to estimate filtered and smoothed unobserved components. Estimation is done using the dlm package (Petris, 2010) in R (R Core Team, 2017).

**2.3 Model outputs**

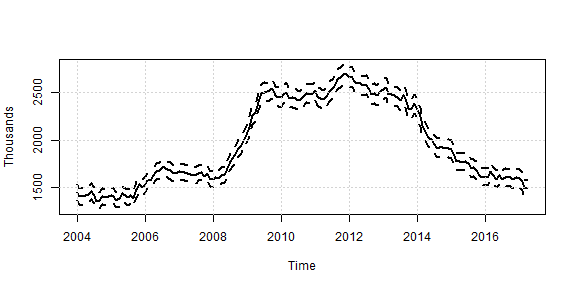
Estimates of unobserved components are elements of the state vector (*αt*) in the state space framework. These include the average level of the series at time *t* (*Lt*), an estimate of monthly change in average level (*Rt*), seasonality (*St*), wave-specific bias () and wave-specific error terms (). For each of these components it is possible to calculate filtered estimates for each time point that only use data up until that time point (*αt|t*) and also smoothed estimates that used all available data (*αt|T*). Where *T* indicates that last available time point of observed data.

Figure 2.3.1 shows the filtered average level of the series for UK unemployment with a ninety-five per cent confidence interval. This is clearly an improvement in the accuracy of the current single month estimates even just observing the volatility of the series but the gain in accuracy especially for estimates of monthly change is clearly demonstrated in figure 2.3.2.

The model is useful as it provides many different outputs and allows in-depth analysis of the time series. However, when publishing key headline data it is important not to confuse the main message by overwhelming users with too much information. Clearly different users will want different amounts of detail and therefore if these model-based outputs were to be used for replacing current monthly estimates it is important to consider how they are published.

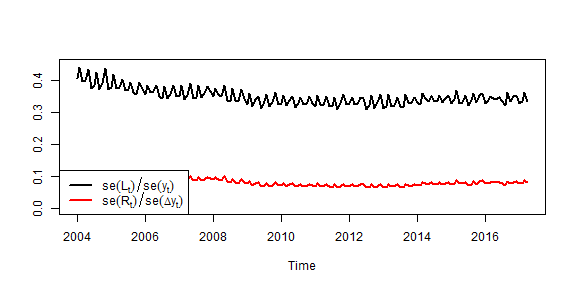
**Figure 2.3.1**

**Filtered trend of UK Unemployment ages 16 plus with 95% confidence interval**

****

**Figure 2.3.2**

**Standard errors of filtered trend and slope relative to design based standard errors for single month estimates of UK Unemployment ages 16 plus**

****

**3. Publishing issues**

The model-based estimates provide greater accuracy, especially of monthly and also improve the timeliness compared to the rolling quarterly data. Models have also been tested that incorporate claimant count in a relatively flexible way similar to the flexibility of the bivariate model discussed by Harvey and Chung (2000) which could provide monthly model-based estimates of unemployment as timely as the claimant count. This is not presented here. As briefly outlined in section 2.3 there are many possible outputs from the model which could be interesting to analyse which raises the questions of what should be published and how?

**3.1 Available guidance**

There are various publications that provide guidance on publishing, for example “Making data meaningful: a guide to presenting statistics” (UNECE, 2009), “Data and metadata reporting and presentation handbook” (OECD, 2007), “Methodology of short-term business statistics: interpretation and guidelines” (Eurostat, 2006), “ESS guidelines on seasonal adjustment” (Eurostat, 2015) and “Web dissemination strategy for Official Statistics” (ONS, 2011). In general these guidelines focus, especially for monthly and quarterly data, on how to present seasonally adjusted data, assuming that the seasonal adjustment has been performed using a method such as the X-11 method or SEATS both of which are available in X-13ARIMA-SEATS and JDemetra+ software. The main focus is on publishing levels, period-on-period changes of the seasonally adjusted time series. The guidance however does not cover the sorts of alternative outputs available from a state space model. For example, should you focus on smoothed or filtered estimates, and which unobserved components are of most interest to which types of user?

The well known dimensions of quality for statistical outputs are useful to consider in the context of different outputs from the state space model. The dimensions of quality include relevance, accuracy and reliability, timeliness and punctuality, accessibility and clarity, and coherence and comparability. It is also important to consider the trade-off between dimensions, user needs and perceptions, cost and respondent burden, confidentiality, transparency and security.

Current publications provide a lot of additional information as well as the focus on the latest level, period on period, and annual changes of the seasonally adjusted series, time series charts of longer spans of data are provided as are standard errors and coefficients of variation, revisions information, quality reports on dimensions of quality and datasets of unadjusted and seasonally adjusted time series.

**3.2 Quality issues for model based outputs**

For model based outputs as with their design-based counterparts it is important to consider the relevance of outputs for different types of users. For example, categorising users into expert analysts, information foragers, and inquiring citizens, we might expect expert analysts to have more interest in many of the outputs from the model, whereas an enquiring citizen might want to know what it the best (optimal) estimate of the average level or change from a particular period. However, here there are trade-offs between dimensions of quality. For example, the optimal estimate for all time points, given *T,* is the smoothed estimate. However, every time you get a new time point you can have revisions. Users are used to revisions from seasonal adjustment and other planned revisions therefore this may not be a problem. Alternatively the filtered estimate is optimal at time point *t,* using information up to time point *t.* Unless there are revisions to the model, hyperparameters or input data, filtered estimates will not be revised, which may be a desirable property. There are many other quality considerations which there are not space to discuss here, but a few other considerations include whether to improve timeliness by having a model that incorporates administrative data, whether and how to create coherence between model-based and design-based estimates and other identity constraints in LFS datasets.

**3.3 Proposed outputs**

It is proposed that outputs include methods papers on input data and the time series model that make it explicit what the input data are and the assumptions made in the model. In addition a clear revisions policy for published time series and the model including, model specification, hyperparameters, and revisions to input data. The Netherlands publish the filtered trend components, as the series appears to behave similarly to a seasonally adjusted time series and therefore for consistency with other European time series it is proposed to publish the filtered trend estimate. However, further investigation of smoothed outputs should be considered as it is the smoothed estimates that are published by the US Bureau of Labor Statistics although the use of the model-based estimates is different.

Standard errors for the filtered trend estimates should also be published regularly to help users understand the accuracy of estimates. Other outputs of the model and the input data should be made available on request. Clearly as other models are developed, for example for smaller area estimates this may then cause some issues in terms of confidentiality.

**4. Conclusions**

A time series model has been developed to improve the timeliness, periodicity and accuracy of UK unemployment. The model provides a wealth of information but there must be careful consideration of what information is published and how it is published. This example provides a useful case study for publishing considerations of time series model-based estimates which are likely to become more common place with the eagerness to incorporate alternative administrative and big data sets into estimation methods.

**References**

Eurostat, (2006). Methodology of short-term business statistics: interpretation and guidelines URL <http://ec.europa.eu/eurostat/documents/3859598/5874981/KS-BG-02-002-EN.PDF/226d148f-e04b-4d9e-bbad-9d5f40ab3a27?version=1.0>

Eurostat, (2015). ESS guidelines on seasonal adjustment URL <http://ec.europa.eu/eurostat/documents/3859598/6830795/KS-GQ-15-001-EN-N.pdf>

Harvey, A. and Chung, C. (2000) Producing monthly estimates of unemployment and employment according to the international labour office definition (with discussion). Journal of the Royal Statistical Society Series A, 160, 5-46.

Giovanni Petris (2010). An R Package for Dynamic Linear Models. Journal of Statistical Software, 36(12), 1-16. URL <http://www.jstatsoft.org/v36/i12/>

OECD, (2007). Data and metadata reporting and presentation handbook URL

<https://www.oecd.org/std/37671574.pdf>

ONS, (2011). Web dissemination strategy for Official Statistics URL <https://www.statisticsauthority.gov.uk/archive/national-statistician/ns-reports--reviews-and-guidance/national-statistician-s-guidance/web-dissemination-strategy-for-official-statistics.pdf>

Pfeffermann, D., Feder, F., and Signorelli, D. (1998). Estimation of autocorrelations of survey errors with application to trend estimation in small areas. Journal of Business & Economic Statistics, 16:339-348.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL <https://www.R-project.org/>

UNECE, (2009). Making data meaningful: a guide to presenting statistics. URL <http://www.unece.org/stats/documents/writing/>

1. Duncan Elliott, Office for National Statistics, Duncan.elliott@ons.gov.uk [↑](#footnote-ref-1)