State Space Model for the UK Labour Force Survey

Duncan Elliott[1], Ping Zong[2]
[1,2] Office for National Statistics, Cardiff Road, Newport. NP10 8XG United Kingdom

Abstract
The UK Labour Force Survey (LFS) publishes a single estimate for each rolling quarter based on a rotating wave design. The purpose of this research is to propose and empirically assess a State Space Model for LFS wave-specific rolling quarterly data with two distinct features. First, the state vector for UK LFS unemployment will not only consider the rolling quarter but will take all relevant monthly periods into account for all state variables in order to capture the characteristics of the rolling quarterly data. Second, the survey error in wave *i* at time *t* is correlated with the survey error in wave (*i*+1) at time *t*+3 for all waves due to the rotating panel survey design. With the development of such a model we aim to deal with potential discontinuities as a result of changes to the survey design and improve estimates of LFS unemployment. Additional improvements include the use of administrative data on unemployment benefits, and the potential for a monthly unemployment estimate.

Key Words: state space model, rotating survey, unemployment

1. Introduction

The UK Labour Force Survey (LFS) publishes seasonally adjusted estimates of unemployment based on rolling quarterly data each month. The identification of trend and seasonal components using X-12-ARIMA does not explicitly account for the rotating wave design that may impose a particular correlation structure on the data. Moreover, there is a timeliness issue in publication of data due to a lag between the date of publication and the period to which the most recent data refer. The aim of this work is to develop a structural time series model that appropriately deals with the correlation structure due to survey design that could also and moves towards more timely estimates and improved estimates of change at the latest time point.

1.1 Published data

ONS publish a Labour Market Statistical Bulletin monthly. This includes estimates of unemployment rates that represent the average unemployment rate over a three month (rolling quarterly) period. The headline figures are the seasonally adjusted estimates of the level of the unemployment rate, the change in that level from the previous quarter and the change in the level from the same quarter for the previous year. The published unemployment rate at month *t* is

$$SA(z_{t-3}) = SA(\Theta_{t-3} + \epsilon_{t-3})$$

where
$z_t =$ the rolling quarterly sample estimate of the unemployment rate at time *t*

Θ_t = signal associated with the rolling quarterly sample estimate at time *t*
ϵ_t = noise associated with the rolling quarterly sample estimate at time *t*
$SA(\ldots)$ =seasonally adjusted estimate of (...)

$$z_t = \frac{1}{3}\sum_{h=-1}^{h=1} y_{t+h} \quad (1)$$

$$y_t = \theta_t + \varepsilon_t$$

where y_t, θ_t, and ε_t are unobserved monthly equivalents of the unemployment rate, signal and noise respectively. Note that observed monthly estimates are possible but have not been used in this research. However, they will be of interest in future developments of this work. The observed data that we use are rolling quarterly estimates of unemployment.

The Labour Market Statistics publication also includes the Claimant Count. This is administrative data on the number of people claiming benefits for being unemployed. The Claimant Count is monthly data and is a more timely data source than the unemployment estimates but is not an estimate of unemployment as given by the International Labour Organisation definitions (ONS, 2011). For month *t* publications seasonally adjusted estimates of Claimant Count are available for month *t*-1.

Figure 1 shows the published seasonally adjusted time series for the unemployment rate for all aged 16 to 64 and the seasonally adjusted rate of the Claimant Count.

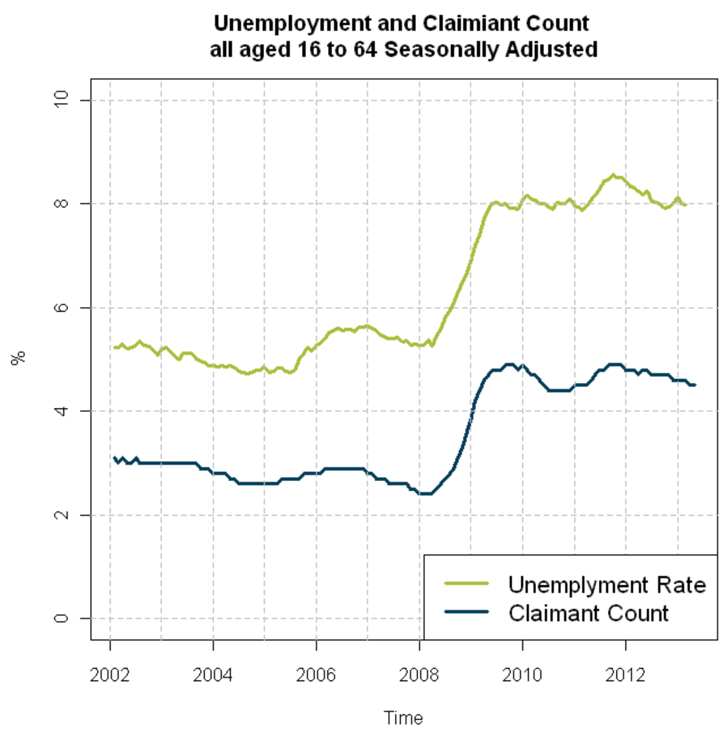

Figure 1: Seasonally adjusted unemployment rate for all aged 16 to 64 and seasonally adjusted rate of the Claimant Count.

1.2 Labour Force Survey Sample Design and Estimation

Comprehensive information on the methods used in the UK Labour Force Survey (LFS) can be found in ONS (2011). Here we briefly summarise some aspects of the methods relevant to the development of our model.

The LFS is a quarterly survey with a rotating panel design. A survey period is comprised of 13 weeks. Once selected respondents are included for five successive survey periods (waves) and then drop out of the sample. Therefore, in each survey period 20 per cent of the sample is replaced. The main purpose of this design is to improve the accuracy of quarterly and annual estimates of change due to sample overlap (see for example Steel, 1997 and Harvey and Chung, 2000). Figure 2 illustrates the wave and cohort structure. For example, those households in cohort 5 are first surveyed during a 13 week period in Jan-Mar 2012, that is to say they would be in wave 1 in this period. This cohort of households is surveyed again in Apr-Jun 2012 and is therefore in wave 2 during this period. In the period Apr-Jun 2012 cohort 1 will drop out of the sample whereas cohort 6 will join the sample. Therefore, in any two consecutive quarters there is approximately an 80 per cent overlap, 60 per cent over lap at a two quarter lag and so on. After a lag of five quarters the two samples will be independent. In practice such an overlap is not achieved due to non-response.

	Cohort											
Period	1	2	3	4	5	6	7	8	9	10	11	12
Jan-Mar 2012	W5	W4	W3	W2	W1							
Apr-Jun 2012		W5	W4	W3	W2	W1						
Jul-Sep 2012			W5	W4	W3	W2	W1					
Oct-Dec 2012				W5	W4	W3	W2	W1				
Jan-Mar 2013					W5	W4	W3	W2	W1			
Apr-Jun 2013						W5	W4	W3	W2	W1		
Jul-Sep 2013							W5	W4	W3	W2	W1	
Oct-Dec 2013								W5	W4	W3	W2	W1

Figure 2: Illustration of the wave and cohort structure

The sample is implicitly stratified by geography due to the geographic ordering of the sampling frame. In general respondents in the first wave are interviewed face-to-face (Computer Assisted Personal Interviews) and those in waves two to five are interviewed by Computer Assisted Telephone Interviews. This may give rise to mode effects whereby responses are influenced by the mode of interview, as discussed in the Dutch case by Brakel and Krieg (2009).

Response rates in wave 1 tend to be higher than in the subsequent waves and there has over time been a steady drop in response (ONS, 2011: 32). Imputation for non-responders is done by rolling forward their response from the previous wave. However, this is only done once. If non-response continues then the record is removed from the dataset. Attrition and the reasons for non-response make it difficult to assess the nature of differences in wave specific estimates as to whether this is due to mode effects or attrition.

As the LFS is a continuous survey it is possible to obtain rolling quarterly estimates. Calibration weighting is used in estimation with calibration groups based on different geographical classifications, age and sex (ONS, 2011: 64). The allocation of the quarterly sample and dividing up of interviewer areas into 13 stints means that while the sample in a 13 week period is not geographically clustered for any shorter period is.

Steel (1997), discusses options for monthly estimates of unemployment. Although his recommendations for a redesign of the survey were not subsequently followed, it is possible to obtain monthly estimates from the current survey. Monthly estimates from the current sample design have no sample overlap, leading to reduced accuracy of month-on-month changes in unemployment. Month-on-month changes are also affected by geographical clustering due to interviewer stints, with the two consecutive months based on samples from different interviewer stints. ONS does currently publish monthly estimates of unemployment but these are not classified as “National Statistics” as they are not considered to be based on a sufficiently robust methodology (Chandler *et al*, 2011).

From the sample data we can also obtain wave specific estimates of unemployment, again using calibration weighting with calibration groups based on different geographical classifications, age and sex. Rolling quarterly wave specific estimates are used in the multivariate state space model described in section 2.2, while only quarterly estimates are used for estimating the survey error autocorrelation described in section 4.

Unemployment data is presented as seasonally adjusted. Seasonal adjustment of unemployment estimates is performed on the monthly series of rolling quarterly data using the X-11 algorithm in X-12-ARIMA (USCB, 2011). The ARIMA model used for extending the series and in prior adjustment accounts for the autocorrelation in the observed data due to the sample design and rolling quarterly estimates but does not explicitly model the sample error autocorrelation discussed in section 4. Pfeffermann *et al* (1998) find that ignoring the autocorrelation due to the sample design can cause some problems, especially for trend estimation in small areas.

2. State Space Model

We consider univariate and multivariate models for the rolling quarterly and wave specific estimates of unemployment. The main purpose of the multivariate model is to better model the survey error autocorrelation and allow for testing of wave effects in the data that could possibly be due to mode effects and/or attrition. This may lead to improved rolling quarterly unemployment estimates, and possibly improved monthly estimates of unemployment, in particular for estimates of change, as documented in Harvey and Chung (2000).

2.1 Univariate model

We assume that the unobserved monthly sample estimate of unemployment (y_t) is the unobserved population estimate (Y_t) plus sample error (e_t) and that the population estimate is comprised of trend (L_t), seasonal (S_t) and irregular (I_t) components.

$$y_t = Y_t + e_t$$

$$Y_t = L_t + S_t + I_t$$

where

$$L_t = L_{t-1} + R_{t-1} + w_t^L \qquad w_t^L \sim N(0, \sigma_L^2)$$

$$R_t = R_{t-1} + w_t^R \qquad w_t^R \sim N(0, \sigma_R^2)$$

$$S_t = -\sum_{i=1}^{11} S_{t-i} + w_t^S \qquad w_t^S \sim N(0, \sigma_S^2)$$

$$I_t = w_t^I \qquad w_t^I \sim N(0, \sigma_I^2)$$

$$e_t = \phi e_{t-3} + w_t^e \qquad w_t^e \sim N(0, \sigma_e^2)$$

From (1)

$$z_t = \frac{1}{3}\sum_{h=-1}^{1} L_{t+h} + S_{t+h} + I_{t+h} + e_{t+h}$$

Our general state space representation is

$$z_t = F\alpha_t + v_t \qquad v_t \sim N(0, \sigma_v^2) \qquad (2)$$

$$\alpha_t = G\alpha_{t-1} + w_t \qquad (3)$$

where

$$E(w_t) = 0$$
$$cov(w_t, w_s) = \begin{cases} 0 & for \quad t \neq s \\ W & for \quad t = s \end{cases}$$

We consider two alternative models in the univariate framework. In model A we assume $\sigma_v^2 = 0$ and estimate the monthly irregular component (I_t) in the state vector

$$\alpha_t = (L_{t+1}, L_t, L_{t-1}, R_{t+1}, S_{t+1}, \ldots, S_{t-9}, I_{t+1}, I_t, I_{t-1}, e_{t+1}, e_t, e_{t-1})'$$

and following similar notation to Brakel and Krieg (2009) where $\mathbf{0}_i$ is a vector of length *i* where all elements are zero.

$$W = Diag(\sigma_L^2, \mathbf{0}_2, \sigma_R^2, \sigma_S^2, \mathbf{0}_{10}, \sigma_I^2, \mathbf{0}_2, \sigma_e^2, \mathbf{0}_2)$$

In the second model (B) we do not estimate the monthly irregular component in the state vector and estimate the variance of the irregular component as σ_v^2. The observation errors are therefore a rolling quarterly population irregular term. The state vector for model B is

$$\alpha_t = (L_{t+1}, L_t, L_{t-1}, R_{t+1}, S_{t+1}, \dots, S_{t-9}, e_{t+1}, e_t, e_{t-1})'$$

and

$$W = Diag(\sigma_L^2, \mathbf{0}_2, \sigma_R^2, \sigma_S^2, \mathbf{0}_{10}, \sigma_e^2, \mathbf{0}_2)$$

For both models $\sigma_e^2 = 3var(e_t)(1-\phi^2)$, where $var(e_t)$ is the design based variance of the rolling quarterly estimate of the unemployment rate, which we assume is constant over time. This assumption will be relaxed in future work. Other variances are estimated as hyperparameters in the model via maximum likelihood.

2.2 Multivariate model

The multivariate model follows a similar structure to that of the univariate model. Here the observations are wave specific rolling quarterly estimates of unemployment (z_t^j) for waves $j = 1, \dots, 5$.

$$z_t^j = \frac{1}{3}\sum_{h=-1}^{h=1} y_{t+h}^j$$

We assume that the unobserved monthly wave specific sample estimate of unemployment (y_t^j) is the unobserved monthly population estimate (Y_t) plus monthly wave specific sample error (e_t^j).

$$y_t^j = Y_t + e_t^j$$

where

$$e_t^j = \phi^j e_{t-3}^{j,t} + w_t^{e,j} \qquad\qquad w_t^{e,j} \sim N(0, \sigma_{e,j}^2) \qquad (4)$$

We assume the wave specific sample error in (4) for wave *j* at time *t* is correlated with the sample error at *t*-3 where $e_{t-3}^{j,t}$ is the error from that sample associated with wave *j* at time *t*. For example, $e_t^3 = \phi^3 e_{t-3}^{3,t} + w_t^{e,3}$, where $e_{t-3}^{3,t} = e_{t-3}^2$.

The general representation of the state space model from (2) and (3) holds, but now $\mathrm{z_t} = (z_t^1, z_t^2, z_t^3, z_t^4, z_t^5)'$ and $v_\mathrm{t} = (v_t^1, v_t^2, v_t^3, v_t^4, v_t^5)'$ where $v_t^j \sim N(0, \sigma_{\mathrm{v},j}^2)$. Similarly to the univariate model the irregular component of the population (I_t) can either be estimated in the state vector or as a rolling quarterly version with the innovations in the observations. These are described as model C and model D respectively.

In model C we assume $\sigma_{v,j}^2 = 0$. The state vector and the variance-covariance matrix of the transition equation for model C are

$$\alpha_t = (L_{t+1}, L_t, L_{t-1}, R_{t+1}, S_{t+1}, \dots, S_{t-9}, I_{t+1}, I_t, I_{t-1}, \mathbf{e}_t)'$$

$$W = Diag(\sigma_L^2, \mathbf{0}_2, \sigma_R^2, \sigma_S^2, \mathbf{0}_{10}, \sigma_I^2, \mathbf{0}_2, \sigma_{e,1}^2, \sigma_{e,2}^2, \sigma_{e,3}^2, \sigma_{e,4}^2, \sigma_{e,5}^2, \mathbf{0}_{13})$$

where

$$\mathbf{e}_t = (e_{t+1}^1, e_{t+1}^2, \dots, e_{t+1}^5, e_t^1, e_t^2, \dots, e_t^5, e_{t-1}^1, e_{t-1}^2, \dots, e_{t-1}^5)$$

In model D we assume $\sigma_{v,j}^2 = \sigma_v^2$. The state vector and the variance-covariance matrix of the transition equation for model D are

$$\alpha_t = (L_{t+1}, L_t, L_{t-1}, R_{t+1}, S_{t+1}, \dots, S_{t-9}, \boldsymbol{e}_t)'$$

$$W = Diag(\sigma_L^2, \mathbf{0}_2, \sigma_R^2, \sigma_S^2, \mathbf{0}_{10}, \sigma_{e,1}^2, \sigma_{e,2}^2, \sigma_{e,3}^2, \sigma_{e,4}^2, \sigma_{e,5}^2, \mathbf{0}_{13})$$

For models C and D, we assume $\sigma_{e,j}^2 = 15var(e_t)(1 - (\phi^j)^2)$, where $var(e_t)$ is the design based variance of the rolling quarterly estimate of the unemployment rate, which we assume is constant over time. This assumption will be relaxed in future work, and better estimates of the wave specific design based variances will be used rather than relying on the assumption that all waves are as accurate as one another and that the variance of the published unemployment rate is a fifth that of the wave specific estimates.

3. Simulation

Harvey and Chung (2000) estimate a state space model for unemployment in the UK. They use data that includes one quarter per year for the period 1984 to 1991, quarterly data from 1992 to 1998 and rolling quarterly data from 1998 to 1999. As we are using a span of data from February 2002 to January 2013, a short simulation study was performed to test the approach of attempting to identify monthly components from rolling quarterly estimates, rather than for example using purely quarterly estimates. We simulate monthly time series (y_t) where

$$y_t = L_t + e_t$$

$$L_t = L_{t-1} + R_{t-1}$$

$$R_t = R_{t-1} + u_t \qquad u_t \sim N(0, \sigma_u^2)$$

$$e_t = \phi e_{t-3} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

We also calculate a rolling quarterly series (z_t) from the monthly data.

$$z_t = \frac{1}{3}\sum_{h=-1}^{1} y_{t+h}$$

The aim of the simulation is to estimate the monthly trend and variances σ_ε^2 and σ_u^2 under different scenarios. The first is to use a state space model to estimate the monthly trend (L_t) from monthly observations (y_t) where

$$\alpha_t = (L_t, R_t, e_t)'$$

$$W = Diag(0, \hat{\sigma}_u^2, \hat{\sigma}_\varepsilon^2)$$

Where $\hat{\sigma}_u^2$ and $\hat{\sigma}_\varepsilon^2$ are maximum likelihood estimates of the level and error variance. This scenario is the benchmark scenario, as it should provide the best estimates.

The second scenario is to estimate the monthly trend using the above model but with rolling quarterly observations. The third scenario is to estimate the monthly trend with rolling quarterly observations and the state vector and transition error variance-covariance matrix defined as

$$\alpha_t = (L_{t+1}, L_t, L_{t-1}, R_t, e_{t+1}, e_t, e_{t-1})'$$

$$W = Diag(\mathbf{0}_3, \hat{\sigma}_u^2, \hat{\sigma}_\varepsilon^2, \mathbf{0}_2)$$

The fourth scenario is to use the above model where the observations are quarterly rather than rolling quarters (ie missing observations for $t = 2,3,5,6, \ldots$). That is to say interpolation of the monthly trend.

We simulate 500 time series of length 100 where

$$L_0 = 8, R_0 = 0.1, e_0 = 0, \sigma_\varepsilon^2 = 0.2, \sigma_u^2 = 0.02$$

We estimate the variances with maximum likelihood and the unobserved components of the state vector (α_t) with Kalman smoothing using the DLM package (Petris, 2010) in R (R Development Core Team, 2010).

From the simulations we calculate the mean absolute per cent error (MAPE) of the estimated trend component, and the mean square error (MSE) of $\hat{\sigma}_u^2$ and $\hat{\sigma}_\varepsilon^2$. Table 1 shows the ratios of these performance measures under different scenarios to that of scenario 1. Therefore, values less than one indicate that the scenario performs better than the benchmark (scenario 1).

Table 1: Simulation results

	$\mathrm{MAPE}(L_t)$	$\mathrm{MSE}(\sigma_\varepsilon^2)$	$\mathrm{MSE}(\sigma_u^2)$
Scenario(2):Scenario(1)	0.86	31.76	17.79
Scenario(3):Scenario(1)	1.00	1.07	1.02
Scenario(4):Scenario(1)	1.19	12.01	1.40

Table 1 shows that scenario three performs best providing a good estimate of the trend and also variances of the slope and error, demonstrating that using the model for scenario three we can extract a monthly trend from observed rolling quarterly data with reasonable estimates of the transition equation variances.

4. Estimation of Survey Error Autocorrelation

Due to the survey design with overlapping sample in consecutive quarters we can expect survey error autocorrelation. In particular we can expect to see high correlation between certain wave specific estimates as discussed briefly in section 2.2. For estimates based on the wave 1 sample there should be no survey error autocorrelation as this is the first time that they have been surveyed. Wave 2 estimates at quarter t (month *t*) should be correlated with wave 1 estimates at quarter *t*-1 (month *t*-3). Wave 3 estimates at quarter t (month *t*) should be correlated with wave 2 estimates at quarter *t*-1 (month *t*-3) and also wave 1 estimates at quarter *t*-2 (month *t*-6) and so on. When calculating survey error autocorrelation for the unemployment rate we may expect to find quarter t (month *t*) is correlated with quarter *t*-1 (month *t*-3), quarter *t*-2 (month *t*-6), quarter *t*-3 (month *t*-9) and quarter *t*-4 (month *t*-12) with the correlation highest at lag one and decaying to zero for lag 5 and greater.

Using wave specific quarterly estimates of the unemployment rate we estimate the autocorrelations and partial autocorrelations using the pseudo-error approach described by Pfeffermann *et al* (1998). We assume that $cor(z_t, z_{t-i}) = cor(y_t, y_{t-3i})$, for $i = 1,2,3,4$, for example the correlation at lag 1 in the quarterly data is the same as the correlation at lag three in the unobserved monthly data, and $cor(y_t, y_{t-ki}) = 0$, for $k = 1,2$ as there is no sample overlap in the monthly data at these lags. We also assume $cor(z_t, z_{t-i}) = cor(y_t, y_{t-3i}) = 0$, for $i \geq 5$.

Table 2 shows the estimated autocorrelations by wave and for averages of waves 1 to 5, waves 2 to 5 and waves 3 to 5. The reason for estimating the autocorrelations for each wave and also averages over a number of waves is that the average for wave one to five is required for the univariate state space model. In the multivariate model we could use wave specific estimates of the autocorrelation, or we could assume a similar autoregressive process for certain waves, for example waves 3 to 5. The shaded area of the table indicates lags where there is sample overlap and * indicates significance at the five per cent level.

For wave 1 we find significant correlation at lag 3 which does not conform to our expectations. There is some possibility for a geographical correlation, but this would be more likely to be observed at lag 1 (as we are working with quarterly data). For wave 2 we find evidence of correlation at lag 1 but not at higher lags, as would be expected. For wave 3 the correlation at lag 2 is not significantly different from zero. This may be due to the effects of attrition and the way in which imputation is performed in the survey (non-responders in wave 2 having their response imputed from their wave 1 response and assuming continued non-response the removal of the unit in wave 3). Significant correlation is found at lag 2 for both waves 4 and 5 but at no greater lags. We therefore find some evidence that autocorrelation is decaying to zero as the lags increase.

Estimates of the partial autocorrelations are obtained by solving the Yule-Walker equations (for example, Wei, 2006). For some of the wave specific estimates this results in some very large and unreliable estimates. Figure 3 plots the estimated partial autocorrelations for the average of waves 3 to 5 and the average of waves 1 to 5, which both provided reasonable estimates. Both show significant correlation at lag 1 and zero after. This gives some justification for assuming that the survey errors follow an autoregregressive process of order one.

Table 2: Autocorrelation by wave and overall

	Lag								
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Wave 1	-0.116	0.108	-0.395*	-0.262	-0.213	0.215	-0.226	-0.122	0.157
Wave2	0.651*	-0.019	0.089	-0.223	-0.067	0.070	0.008	-0.050	-0.120
Wave3	0.533*	0.198	0.008	0.053	0.015	-0.076	0.011	-0.119	-0.034
Wave4	0.701*	0.365*	0.132	0.264	0.090	0.027	0.010	0.175	0.022
Wave5	0.419*	0.448*	0.167	0.157	0.024	0.189	-0.061	-0.039	0.220
Av W1:5	0.461*	0.213	0.009	0.000	-0.025	0.073	-0.043	-0.031	0.036
Av W2:5	0.551*	0.230	0.073	0.041	0.004	0.051	-0.015	-0.017	0.017
Av W3:5	0.532*	0.297*	0.073	0.115	0.026	0.044	-0.018	-0.007	0.052

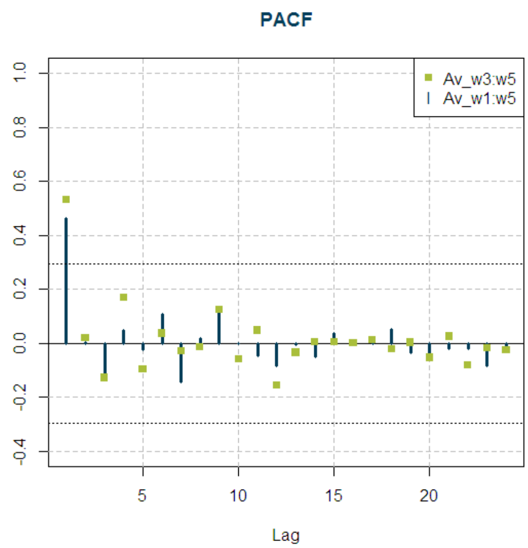

Figure 3: Survey Error Partial Autocorrelations overall and the average for waves 1 to 3

It is interesting to note that the estimated correlation coefficient for the quarterly unemployment rate series at lag 1 is close to the 0.5 value found in Harvey and Chung (2000).

5. Results

The results of fitting the models described in section 2 are presented below. For each model we assume different scenarios providing some alternative model parameters. There are two main reasons for doing this. First, to demonstrate what happens in instances where the survey error autocorrelation is not addressed and second as the published design based variance estimates of the unemployment rate were found in some instances to lead to an estimated monthly irregular component with significant autocorrelation indicating a poor model. We found that larger variances of the survey error improved the estimated components and some model diagnostics.

Further work is required on the design based variance estimates as we have assumed these to be constant over time, which is not the case. The wave specific estimates and estimates of unemployment used in our models are not those of published unemployment as the weighting differs and we may expect that the estimates have a greater variance than those published due to simplified calibration weighting.

Below we fit models assuming $var(e_t) = 0.023$ and $var(e_t) = 0.062$ and with and without survey error autocorrelation, giving four permutations of each model.

5.1 Univariate Models

Table 3 shows the estimated parameters for the variances of the slope and irregular component using model A with different fixed inputs for the autoregressive parameter and survey error. Note that the estimated variances for the level and seasonal components were generally found not to be significantly different from zero and so have been assumed to be zero. The values in parenthesis are the standard errors of the estimates computed using the Hessian of the negative loglikelihood described by Petris (2010: 146).

Model A1 includes a term for the survey error autocorrelation (ϕ = 0.46) and published variance for the survey error. For this model the variance of the irregular component is large and the monthly decomposition of the irregular component has significant autocorrelation, as does the quarterly version of the irregular component. One possible cause of this could be an underestimate of the design based variance for the data used in the model.

In model A2 where there is no attempt to model the survey error autocorrelation, again the irregular component shows significant autocorrelation, as would be expected. Only in model A3 where the survey error in increased, while accounting for survey error autocorrelation does the model improve in terms of autocorrelation in the irregular. However, the variance of the irregular component is then found not to be significantly different from zero.

Table 3: Hyperparameters (with standard errors) under four scenarios of model A

	A1	*A2*	*A3*	*A4*
Hyper-parameter	ϕ = 0.46 $var(e_t) = 0.023$	ϕ = 0 $var(e_t) = 0.023$	ϕ = 0.46 $var(e_t) = 0.062$	ϕ = 0 $var(e_t) = 0.062$
$\hat{\sigma}_R^2$	0.00024 (0.00012)	0.00015 (0.00007)	0.00051 (0.00018)	0.00016 (0.00008)
$\hat{\sigma}_I^2$	0.58954 (0.05243)	0.90593 (0.06472)	0.01434 (0.01455)	0.74645 (0.06265)

Table 4 shows the estimated parameters for the variances of the slope and irregular components using model B with different fixed inputs for the autoregressive parameter and survey error. Under model B there is no attempt to estimate a monthly irregular component, rather the observation error, is the irregular component which is therefore a rolling quarterly error term.

Model B1, that includes a term for the survey error autocorrelation and variance based on published estimates, results in a model with approximately normally distributed errors and no evidence of autocorrelation in the residuals. Models B2 to B4 all show evidence of autocorrelation in the residuals.

Table 4: Hyperparameters (with standard errors) under four scenarios of model B

	B1	*B2*	*B3*	*B4*
Hyper-parameter	$\phi = 0.46$ $var(e_t) = 0.023$	$\phi = 0$ $var(e_t) = 0.023$	$\phi = 0.46$ $var(e_t) = 0.062$	$\phi = 0$ $var(e_t) = 0.062$
$\hat{\sigma}_R^2$	0.00097 (0.00023)	0.00102 (0.00022)	0.00051 (0.00017)	0.00076 (0.00019)
$\hat{\sigma}_v^2$	0.00449 (0.00054)	0.00421 (0.00049)	0.00288 (0.00047)	0.00341 (0.00047)

Both models A and B provide a monthly trend estimate, and both enable an adjustment to the rolling quarterly data for survey error autocorrelation. Model A3 in theory allows estimation of a monthly population value, from which it would also be possible to derive a monthly seasonally adjusted estimate, without requiring monthly input data. However, the estimation of the irregular component is poor in this model. Model B1 has reasonable model diagnostics and provides a monthly trend estimate, but does not provide a monthly population estimate or monthly seasonally adjusted estimate. The resulting trend and seasonally adjusted estimates for some of these models are presented in section 6.

5.2 Multivariate Models

As noted in section 2, design based variance estimates of the wave specific estimates of unemployment have not been estimated and used in the following models. This will be the subject of further work. As was found for the univariate model, an increased variance for the survey error improved estimation of the population irregular component, and may be justified by the simplified calibration used to obtain the wave specific estimate of unemployment.

Model C3 exhibits a small amount of autocorrelation in the monthly Irregular component but is much better than the other models where there is very clear evidence of autocorrelation. Model C3 results in a similar estimate of the variance for the slope and irregular as the univariate model A3. The main difference between them is that the variance of the irregular component is found to be significantly different from zero for model C3. This slight improvement may be due to dealing with the wave specific correlation structure.

Table 5: Hyperparameters (with standard errors) under four scenarios of model C

	C1	*C2*	*C3*	*C4*
Variable	$\Phi_1 = 0$	$\phi_i = 0$, for i=1,..,5	$\Phi_1 = 0$	$\phi_i = 0$, for i=1,..,5
	$\Phi_2 = 0.65$	$var(e_t) = 0.023$	$\Phi_2 = 0.65$	$var(e_t) = 0.062$
	$\Phi_3 = \Phi_4 = \Phi_5 = 0.53$		$\Phi_3 = \Phi_4 = \Phi_5 = 0.53$	
	$var(e_t) = 0.023$		$var(e_t) = 0.062$	
$\hat{\sigma}_R^2$	0.00041	0.00034	0.00057	0.00034
	(0.00015)	(0.00015)	(0.00017)	(0.00014)
$\hat{\sigma}_I^2$	0.24883	0.44091	0.01803	0.34788
	(0.02768)	(0.03737)	(0.00896)	(0.03739)

All of the scenarios tested for model D resulted in autocorrelation in the residuals for one or more of the wave specific estimates. Improvements in model diagnostics may be possible if design based variances are estimated for the wave specific estimates or with the inclusion of other variables in the state vector to account for difference between wave specific estimates as is done by Brakel and Krieg (2009). This will be the subject of future work.

Table 6: Hyperparameters (with standard errors) under four scenarios of model D

	D1	*D2*	*D3*	*D4*
Variable	$\Phi_1 = 0$	$\phi_i = 0$, for i=1,..,5	$\Phi_1 = 0$	$\phi_i = 0$, for i=1,..,5
	$\Phi_2 = 0.65$	$var(e_t) = 0.023$	$\Phi_2 = 0.65$	$var(e_t) = 0.062$
	$\Phi_3 = \Phi_4 = \Phi_5 = 0.53$		$\Phi_3 = \Phi_4 = \Phi_5 = 0.53$	
	$var(e_t) = 0.023$		$var(e_t) = 0.062$	
$\hat{\sigma}_R^2$	0.00292	0.00210	0.00224	0.00147
	(0.00054)	(0.00040)	(0.00048)	(0.00030)
$\hat{\sigma}_v^2$	0.01674	0.01979	0.00666	0.00759
	(0.00084)	(0.00101)	(0.00042)	(0.00049)

6. Discussion

Figure 4 shows monthly trend and seasonally adjusted estimates of the unemployment rate and monthly and annual changes from a selection of models. Trend and seasonally adjusted estimates from directly seasonally adjusting the rolling quarterly unemployment rate using X-13ARIMA-SEATS are shown for comparison (denoted X-11 as they are adjusted using the X-11 algorithm). Models that do not adjust for survey error autocorrelation show very similar estimates to the X-11 results (not shown in figure 4).

As can be seen estimation of the population trend and seasonally adjusted estimates show different interpretations of levels and movements in the series compared to the case where survey error autocorrelation is not adjusted for in the X-11 algorithm. The series shown in figure 4 are for estimates up to January 2013, as the last rolling quarterly estimate is for the period December to February 2013. However, the models also provide a monthly estimate for February 2013, which would allow more timely estimates of unemployment and allow presentation of month on month changes at the more recent time point than currently published. We do not comment further on the results as further work is required on the design based variances, but note that the approach is promising.

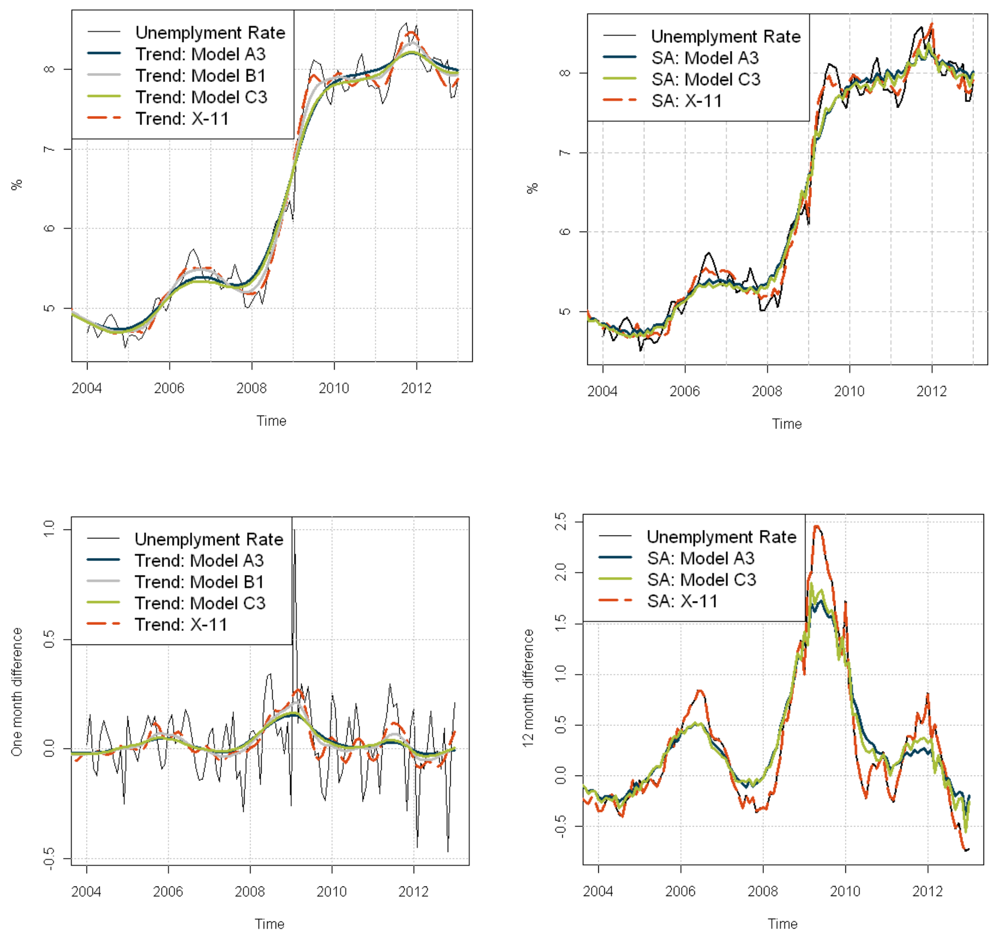

Figure 4: Trend and seasonally adjusted estimates for unemployment rate

It is clear that further work is required to refine the models and also reduce some of the assumptions required by providing wave specific monthly estimates with associated design based variances. This will require a reformulation of the model, and additional unobserved components should be tested for, such as trend differences between waves as estimated in Brakel and Krieg (2009). Further work is also required to analyse the performance of the model in terms of revisions, identification of turning points and using the results to provide timely estimates of change, as in Harvey and Chung (2000), who find a bivariate model using ILO unemployment and the Claimant Count provide improved estimates of monthly change.

Acknowledgements

This work builds on previous work by Jan Angenendt, Julian Chow, and Gary Brown with support and advice from Danny Pfeffermann, Moshe Feder, Jan van den Brakel, John Aston, Paul Smith, Denise Silva, Daniel Ayoubkhani and Neil Parkin. Thanks to Katie Stuart, Philip Ball and Greg Dixon for their work on producing wave specific estimates of unemployment. Any errors are those of the authors.

References

Brakel, J.A van den and Krieg, S. (2009). 'Structural time series modelling of the monthly unemployment rate in a rotating panel', *Discussion paper 09031*, CBS, Statisitcs Netherlands http://www.cbs.nl/NR/rdonlyres/DF041949-9046-45ED-A673-72787ECCE778/0/200931x10pub.pdf

Chandler, M., Palmer, N. andYasir, S. (2011). 'Labour Force Survey Single Month Estimates', Office for National Statistics, http://www.ons.gov.uk/ons/rel/lmac/labour-force-survey-single-month-estimates/labour-force-survey-single-month-estimates/new-component.pdf

Harvey, A. and Chung, C-H (2000). 'Estimating the underlying change in unemployment in the UK' *Journal of the Royal Statistical Society Series A*, Vol. 163 Part 3 pp303-339

ONS (2011). 'Labour Force Survey User Guide Volume 1 – LFS Background and Methodology 2011' http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/labour-market-statistics/index.html

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

Pfeffermann, D., Feder, M. And Signorelli, D (1998). 'Estimation of autocorrelations of survey errors with application to trend estimation in small samples' *Journal of Business and Economic Statistics*, Vol. 16 pp339-348

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Steel, D. (1997). 'Estimates of Unemployment and Employment' *Journal of the Royal Statistical Society Series A*, Vol. 160 Part 1 pp5-46

USCB (2011), *X-12-ARIMA Reference Manual version 0.3*, Time Series Research Staff, U.S. Census Bureau, http://www.census.gov/srd/www/x12a/ .

Wei, W.W.S. (2006). *Time Series Analysis: Univariate and Multivariate Methods*, second edition, Boston: Addison-Wesley Publishing Company,Inc.