

NEW EXPERIMENTAL STATISTICS
ESTIMATING UNEMPLOYMENT AT PARLIAMENTARY CONSTITUENCY LEVEL

Bob Watson, Denise Silva and Philip Clarke
Office for National Statistics

Key Points

- In 2003 ONS produced model-based estimates of unemployment for Local Authorities (LAs), borrowing strength from the Annual Population Survey and the count of claimants of Jobseeker's Allowance (see pp 37-43, Labour Market Trends, January 2003).
- This methodology has been extended to produce estimates of unemployment for Parliamentary Constituencies (PCs).
- The methodology needed refining because some of the parameters used in the original model were not available for PCs, and to build in consistency between modelled estimates where PCs and LAs cover the same geographic areas.
- The new model-based estimates for PCs will be published as experimental statistics at the end of July 2009, alongside the 'Local area labour markets: statistical indicators' publication.
- In publishing the new experimental statistics, ONS welcomes feedback from potential users.

Background

The demand for small area labour market statistics has increased over the past few years, particularly in the context of the measurement of social exclusion, social wellbeing and, more recently, effects of the downturn in the economy. The Annual Population Survey (APS) is a key source of information on the labour market, but it is primarily designed to produce Government Office Region estimates, and the sample is not large enough to provide reliable labour force estimates for all local areas. Therefore, direct APS estimates of unemployment consistent with International Labour Organisation (ILO) definitions are of limited use for parliamentary constituencies (PCs). The use of statistical modelling techniques has enabled the quality of small area statistics to be enhanced. This article presents an overview of the methodology for producing PC model-based unemployment estimates and some analysis.

In an earlier article ONS reported on progress made on the development of a model-based approach to estimating levels and rates of unemployment for small areas (see pp 37-43, Labour Market Trends, January 2003). The article described the model-based approach and presented results for unitary authorities and local authority districts (LAs) in Great Britain for the years 1995/96 to 1999/2000. The model was developed to improve the APS estimates of unemployment for small areas by using supplementary information from the claimant count – the number of people claiming Jobseeker's Allowance, a socio-economic area indicator and a random area effect. These were originally published as experimental statistics (for a definition of experimental statistics see www.statistics.gov.uk/press_release/experimental.asp.) and became National Statistics in July 2006.

Following this, ONS has extended this model to produce estimates of unemployment based on ILO definitions for PCs. This model is based on the one currently used to produce unemployment estimates for LAs. It takes, as an input, data from APS and uses the claimant count as a covariate. The estimates are annual estimates, i.e. they use an average of the previous twelve month's claimant

count totals and twelve months of survey data. The aim is to publish updated unemployment and unemployment rate estimates for PCs every quarter.

The results have been compared with the direct survey estimates of unemployment for PCs, which have shown that the model estimates are more reliable. Similarly to the model-based estimates for LAs, the model-based estimates for PCs are constrained to regional totals. Also, where a PC is the same geographic area as a LA, the model-based estimate for the PC is constrained to be the same as the model-based estimate for the LA. There are 35 PCs which are identical areas to LAs. There are also a further 48 LAs which are the same areas as a number of PCs. For most of these areas there are two or three PCs within the LA, but some have more than this. Birmingham, for example, has 11 PCs within the LA. As a further constraint, where the LA is the same as a number of PCs, the total unemployment has been constrained to the LA model based estimate.

Issues

There were two main issues that needed to be addressed before the model for unemployment at PC level could be finalised:

1. Constraining to local authority estimates where parliamentary constituencies and local authorities coincide

An issue arises of alternative estimates for the same areas from the LA and the PC models where areas coincide exactly. There are 35 such instances. This is addressed within the estimation procedure of the PC model by constraining to the LA model-based estimate. Subsequently calibration to direct survey estimates is carried out at regional level. ONS conducted a quality check on the pre-constrained estimates for these PCs to assess the compatibility of the separate models and the extent of the adjustments caused by subsequent constraining. This showed close agreement between the estimates from the two models - median absolute relative differences equal or under 3.0% and maximum absolute relative difference of around 11.5% in unemployment level. In addition, all PC model estimates fall within the 95% confidence interval of the LA model estimates.

2. Cases where several parliamentary constituencies make up one local authority

While the estimation procedure constrains to LA model-based estimates for single matching areas, there are cases where several PCs exactly make up one LA. These are predominantly among London boroughs and metropolitan districts which for the most part are made up of two or three PCs (although a few are made up of many more, the maximum being Birmingham which has eleven). Altogether there are 48 such cases affecting 150 PCs. ONS conducted a quality investigation of the aggregated PC unemployment level estimates against the LA model-based estimates for five annual time periods. In summary the results showed close agreement with the LA estimates. All aggregated estimates fall within the 95% confidence intervals of the LA model-based estimate and have median absolute relative differences equal or under 3.5%. However, occasionally, outlying LAs gave rise to a maximum absolute relative difference of about 15%.

The development of this project exposed the duality between producing estimates based on statistical procedures that are optimised for a given geography and the need for calibrating/constraining the estimates for presentational purposes. Consistency is an important feature of any statistical system and the creation of a small area estimation framework has to address

the issue of consistency between model-based estimates produced for different geographies as well as the consistency of model-based estimates and direct surveys estimates published for higher/broader geographies.

Further Development of the Model

The original model to estimate LAs was taken as the basis for the PC model. The LA model is supplied with data from the APS for 406 LAs (City of London and Isles of Scilly are excluded) by male/female and three age groups giving counts of those unemployed in the sample, total sample size and direct survey estimates of total unemployed, total economically active and total population. Also supplied is the total count of those claiming Jobseeker's Allowance by these categories and the Government Office Region (GOR) and socio-economic classification. The latter is based on the published socio-economic classification of LAs based on the 2001 Census.

The model constructed to estimate to PCs is supplied with the same data for the 628 PCs with the exception of the socio-economic classification variable (cluster term). No such classification has been published for PCs and, after investigation, it was decided that the inclusion of a cluster term could be omitted. As for the LA estimation, the PC model combines fixed-effects for age, sex and GOR with a random area effect (see p43 Labour Market Trends, February 2006).

The model based approach relies on determining a strong relationship between unemployment and auxiliary information. **The underlying idea of the method is to exploit similarities in order to borrow strength over areas.** The main source of this auxiliary information is the number of beneficiaries of Jobseeker's Allowance, the claimant count.

The small area estimation model at PC level is an area level logistic mixed model that relates the probability of being unemployed for an individual of a particular sex and age group within a PC with the corresponding claimant count information, incorporating additional explanatory variables accounting for age, sex and regional differentials. The PC random effects capture unexplained sources of variation and area heterogeneity that may not be explained by the auxiliary data. The model formulation is given by:

$$\text{logit}(p_{di}) = \ln \left\{ \frac{p_{di}}{1 - p_{di}} \right\} = \mathbf{X}_{di}^T \boldsymbol{\beta} + u_d \quad \text{such that} \quad p_{di} = \frac{\exp(\mathbf{X}_{di}^T \boldsymbol{\beta} + u_d)}{1 + \exp(\mathbf{X}_{di}^T \boldsymbol{\beta} + u_d)}$$

where p_{di} is the probability that an individual in (age-sex) group i from PC d is unemployed, $\boldsymbol{\beta}$ is the vector of fixed effect coefficients and u_d is the random area effect. The model predictors \mathbf{X}_{di} are:

- indicators of age-sex groups (male/female for age groups: 16 to 24; 25 to 49; 50 and over) and of the 11 GORs in GB;
- the logit of the claimant count proportion in each age-sex group within the PC;
- the logit of the claimant count in the PC.

The model-based estimate of the unemployment level in each PC is produced by combining the APS sample count of number of unemployed with a predicted value for the non-sample count obtained from the modelling procedure. The unemployment rate is simply the ratio of the number of people who are unemployed in the PC to the economically active population in the PC. An

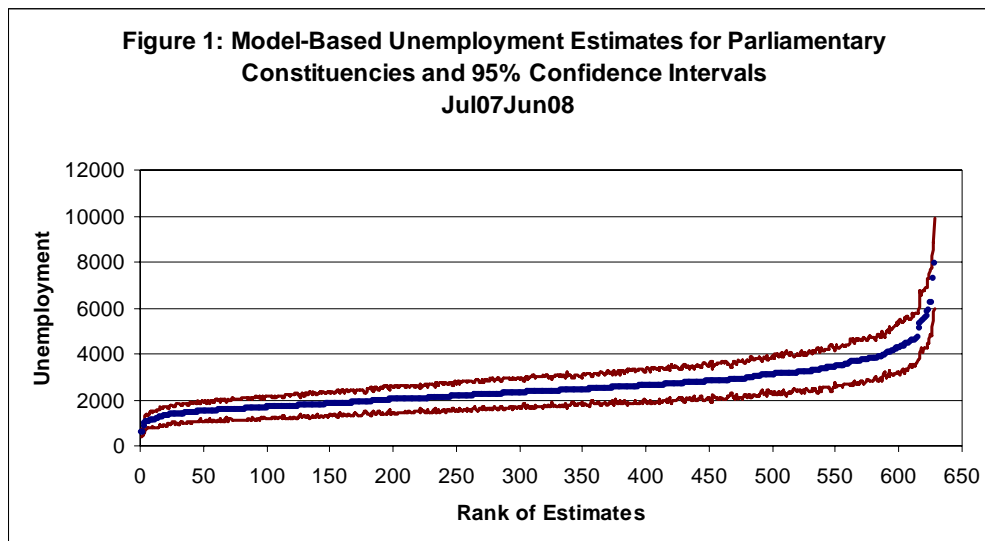
estimate for the number of economically active is obtained as the sum of the model-based estimate of unemployment and the direct survey estimate of employment.

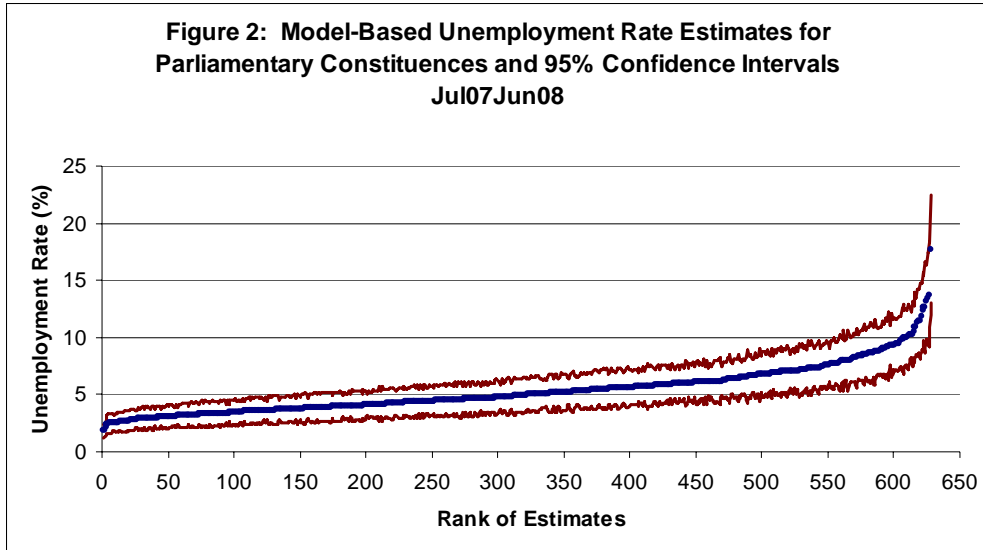
As stated previously, investigation of how closely the model-based estimates for these PCs align with the model-based estimates for the LAs was carried out. Even though the results were good it was felt the loss of consistency did not make the use of the unconstrained model-based PC estimates feasible and, as a result, a further study of the estimates has been conducted. The estimation system was re-written to give an option to constrain the PC estimates to sum to LA estimates where two or more PCs are jointly coterminous with a LA while maintaining the overall calibration of estimates by GOR and by age-sex group nationally.

Following this work it was clear that constraining to single PCs matching a LA resulted in a very small and consistent change in the remaining PCs in the same regions tested, to the order of +0.1%. Thus there is no cause for concern in this single PC constraining scenario. Further constraining to LA estimates where LAs match multiple PCs also showed an even distribution of change among the non-matching PCs, in this case of around -0.4% in Scotland, -0.5% in Greater London and -3% in West Midlands. The impact of calibration on the remaining non-matched PCs was evaluated since although the non-matching PCs are not directly involved in the calibration equations for constraining to LA values, the additional constraints also have an effect on those PCs which do not form part of the LA. This is because an overall calibration to the direct estimates of total unemployed at each GOR level and by sex-age group nationally is also implemented (in the same manner that is used to produce model-based estimates for LAs). The results show that little statistical damage is made by implementing the constraining for multiple PCs coterminous with LAs

Results

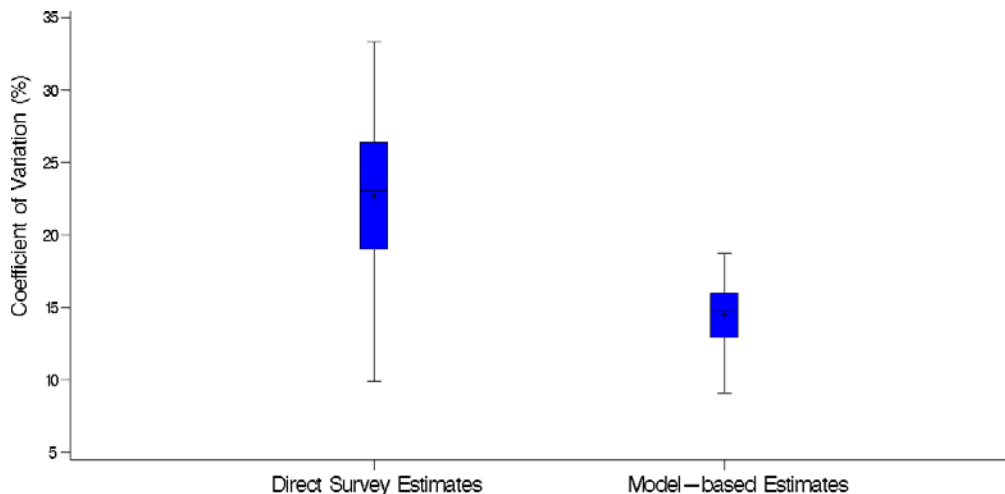
Following the studies ONS took the decision to publish experimental unemployment figures for PCs that are constrained to ensure consistency everywhere. Figures 1 and 2 present an example of unemployment and unemployment rate estimates, with corresponding 95% confidence intervals, for the period July 2007-June 2008.

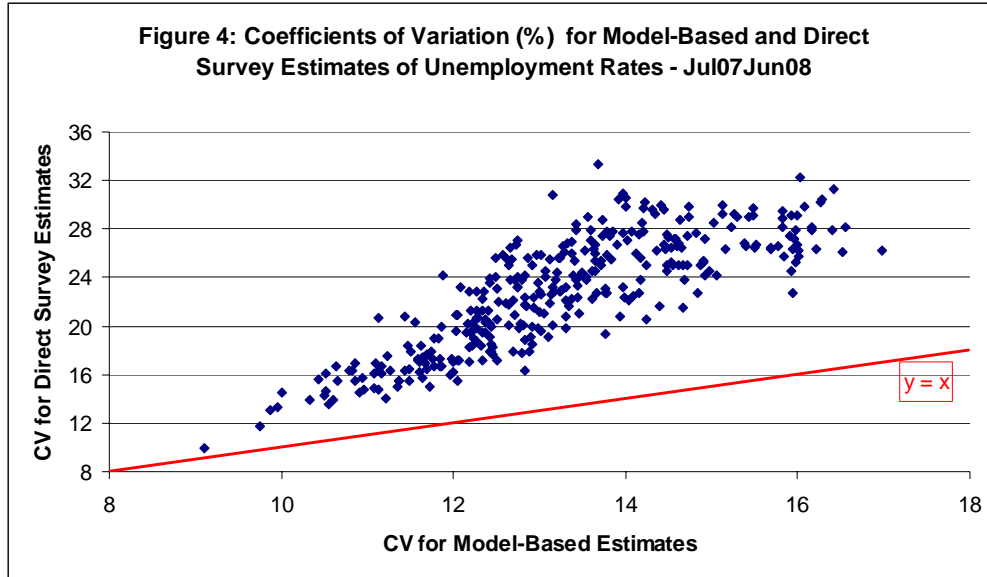




The model-based estimates are more reliable than the direct APS estimates. Figure 3 and 4 compare the coefficients of variation (CVs), defined as the standard error divided by the corresponding estimate, for the model-based and direct survey estimates of unemployment rates. The box-plots show that whereas more than half of the CVs of the APS estimates are higher than 20%, all model-based estimates present a CV lower than this threshold. In fact, almost 40% of the APS unemployment rate estimates for PCs published via Nomis for the period July2007-June2008 are marked as unreliable. In addition, Figure 4 provides evidence that for every PC the precision of the model based figures is better than for the survey direct results since the small area estimation technique produces estimates with lower CVs.

Figure 3: Distribution of the Precision of Parliamentary Constituency Unemployment Rate Estimates – Jul07Jun08

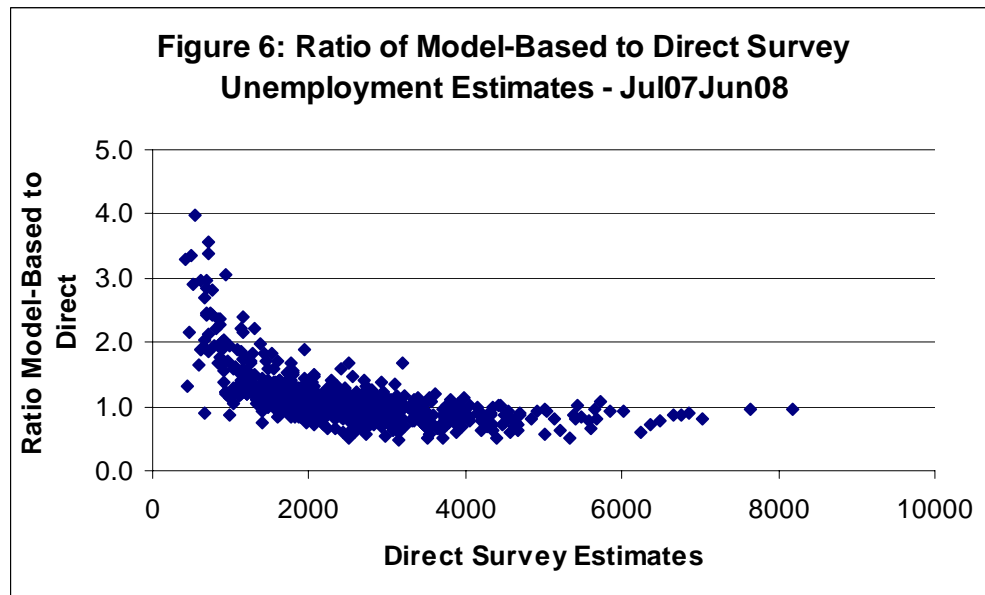
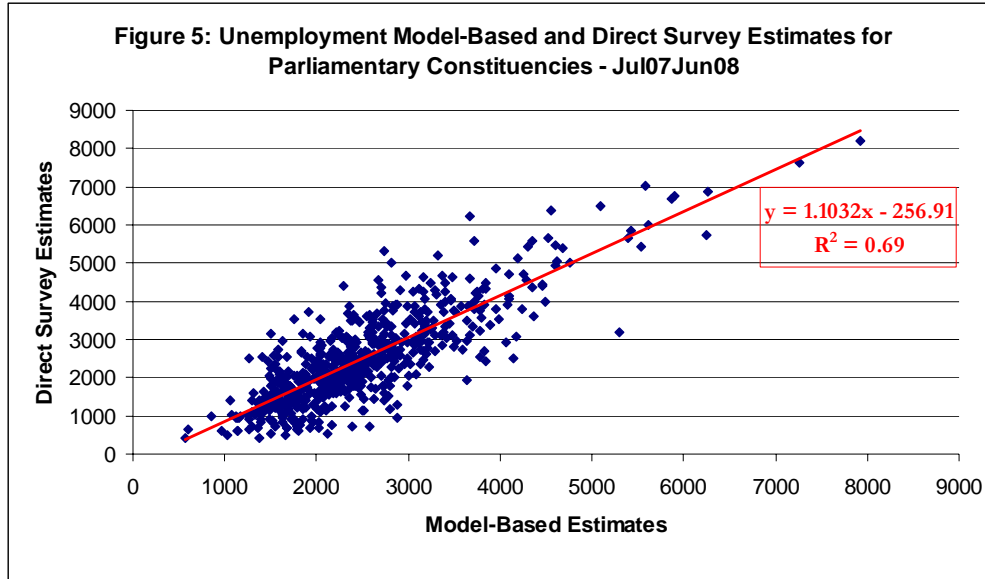




The small area estimation technique was used to overcome the problem of APS small sample sizes within PCs. However, although more precise, the resulting model based estimates are biased¹ (in a statistical sense). The aim of the estimation procedure is to balance the trade-off between variance and bias, producing estimates with good precision and with as little bias as possible. Although the direct estimates are very variable, they are nevertheless unbiased thus a plot of direct estimates (on the y axis) and model-based estimates (on the x axis) should display direct estimates randomly scattered about the model estimates and will give a regression line close to the $y = x$ line if the model based estimates are also unbiased.

Figure 5 presents a plot of the unemployment estimates to investigate if the model-based estimates are approximately unbiased. The results show that although the regression line slope is 1.10 with a negative intercept, the direct and modelled estimates do appear to track each other. Figure 6 shows that at low unemployment levels, the modelled estimates tend to be higher than the direct estimates. Outside these low levels however, the ratios model-to-direct appear to be scattered around 1 whereas, for higher levels of unemployment, model-based estimates are in general lower than the corresponding survey figures. This may be partly due to particular age-sex groups in some PCs having no-one in the sample being recorded as unemployed. Also, the model-based estimates are produced using a model fitted on data from all GB and the estimation procedure tends to shrink the estimates towards the overall mean. In addition, the final estimates are calibrated to national level direct estimates. These factors combined explain the observed patterns in Figures 5 and 6.

¹ Bias refers to a systematic error contributing to the difference between the sample estimate and true population value (the smaller the better)



ONS developed a modelling methodology to produce reliable unemployment and unemployment rate estimates for PCs and also the corresponding quality measures. The model-based estimates are of a different nature from the direct APS estimates. The former relies on the model that relates the APS estimates to the auxiliary data. The results show that the model is plausible and provides more reliable small area estimates.

Next Steps

The experimental model-based estimates will be published on the NS web site in July with the 'Local area labour markets: statistical indicators' publication. This area of the website already includes data for PCs, and in future it is intended to include model-based estimates of unemployment for PCs, rather than estimates of unemployment from the APS (the publication already includes the model-

based estimates of unemployment for LAs). The model-based estimates of unemployment for PCs will be published along with confidence intervals. Model-based unemployment and unemployment rate estimates will be produced quarterly based on annual data. The modelling process produces standard errors for the estimates, from which the confidence intervals will be calculated. The direct survey estimates will continue to be made available via Nomis.

The House of Commons Library will provide a focal point for use of these estimates. ONS has regular contact with the Library, and has already discussed making a presentation on the estimates at the House of Commons. The audience for this will be Library statisticians and Member's researchers.

In publishing the experimental data ONS welcomes feedback from potential users. The statistics are not yet National Statistics and do not replace previously published estimates.

References

Hastings, D., Maine, N., Brown, G. and Cruddas, M. Development of improved estimation methods for local area unemployment levels and rates. Labour Market Trends. January 2003. pp 37-43.

Further information

For further information and comments, contact:

Bob Watson
Room 1.024,
Office for National Statistics,
Government Buildings,
Cardiff Road,
Newport
Gwent NP10 8XG

Email: bob.watson@ons.gov.uk
Tel: 01633 455070

For further information about the methodology, contact:

Denise Silva
Room 4200N,
Office for National Statistics,
Segensworth Road,
Titchfield,
Fareham,
Hants PO15 5RR

Email: denise.silva@ons.gov.uk
Tel: 01329 444982