Model-Based Estimates of ILO Unemployment for LAD/UAs in Great Britain

Guide for Users

July 2006

1. Summary

The Office for National Statistics (ONS) has jointly developed with Professor Ray Chambers of the University of Southampton, a new modelling methodology to produce estimates of unemployment level and rate on the International Labour Organisation (ILO) definition for local authority districts and unitary authorities (LAD/UAs). The new methodology has undergone academic review and the estimates made under this methodology have undergone validation.

This new methodology builds on previous modelling work reported in [1] where estimates were produced and published as experimental statistics. The new methodology is an improvement for the following reasons: -

- the estimates produced are always more precise than the direct estimates produced from the Labour Force Survey. Under the previous methodology there were a few (typically 4-6) LAD/UAs where the model-based estimates were less precise than the corresponding direct estimates;

- the estimates produced are, in effect, a weighted combination of the direct estimate and the previous model-based estimate. The weighting is variable and dependent upon survey sample size. The estimates become closer to the direct estimate as the survey sample size, and hence confidence in the direct estimate, increases. Conversely, for LAD/UAs where the survey sample size is small the new model-based estimates will be close to the previous model-based estimates.

The new estimates are now accredited as National Statistics and this User Guide accompanies the first issue of contemporary estimates for the year January to December 2005. It describes the development of the methodology and gives the user guidance on the use and limitation of the estimates. Using historic data covering the periods 1996/97 to 2003/04 as well as the 2005 data, model-based estimates and

confidence intervals of the unemployment rate and level for 406[1] LAD/UAs in England, Scotland and Wales have been produced and verification and validation tests are described.

Thematic maps displaying the estimates of unemployment rate at the LAD/UA level for Great Britain are presented in Section 7. A Technical Report which describes the methodology and conduct of diagnostics in detail is forthcoming and will also be released on the website.

[1] There are 408 LAD/UAs in Great Britain. For this project, and in common with the Labour Force Survey, two local authorities: The Isles of Scilly and The City of London are not included for reasons of disclosure.

2. Background

2.1 Introduction

The UK Labour Force Survey is the key source of national information on the labour market; however it is not able to deliver direct estimates of unemployment with adequate precision for every local authority district or unitary authority (LAD/UA) in Great Britain because the sample size in many areas is insufficient. The Office for National Statistics (ONS) has explored the use of model-based estimation techniques to improve the precision of estimates of unemployment for LAD/UAs.

The Labour Force Survey is a continuous, large-scale survey, with a sample of around 60,000 households in each three-month period. These include around 150,000 people, of whom over 110,000 are aged 16 or over, in each three-month period. Since 2000 the Labour Force Survey sample has included enhancements (the boost [2]) to the underlying sample to ensure at least a minimum number of economically active adults are included in the sample for each area (these are not necessarily LADs as local education authorities and London boroughs were the target for the English boost). A boost was first applied to England (2001/02). In the following year a boost was applied to Wales and finally in Scotland from 2003. With these boosts the survey is known as the Annual Population Survey (APS). These data are used to measure unemployment according to the International Labour Organisation (ILO) definition on an annual basis. Historically the sample size within an individual LAD/UA was often too small to provide reliable estimates resulting in only about a quarter of the annual estimates (22% in 1999/00) of unemployment qualifying for publication under the publication rules existing at that time. More recently the rules have been relaxed allowing more of the estimates to be published but still many Local Authorities have no published estimate of unemployment and for those that have, many of the estimates lack precision. More information about the Labour Force Survey may be found on the National Statistics web site, in particular see [2] and [3].

Against this background, a project was established with the aim of producing sets of LAD/UA level estimates of unemployment levels and rates using a model-based methodology jointly developed by the ONS and Professor Ray Chambers of the University of Southampton.

2.2 Model-based approach

The model-based approach is based on determining a strong relationship between ILO unemployment (as measured by the Labour Force Survey) and auxiliary or covariate information (usually from Census or administrative sources). This relationship is then used to provide more reliable estimates of ILO unemployment for LAD/UAs. The main source of this additional information is the register of the number of recipients of job-seekers allowance (the 'claimant count').

The new methodology is known as the **random effects** model and differs from the previous **fixed effects** methodology (estimates using this methodology have been published as experimental statistics) in that it allows this relationship to be different for different LAD/UAs recognising that there may be between area differences that are not explained by the auxiliary data. The main benefits of the using the new methodology are:

- It produces model-based estimates that are generally more precise than the previous ones estimated from the fixed effects model. They are always more precise than the direct estimates produced under the Labour Force Survey. Under the previous fixed effects model-based methodology there were a few (typically 4-6) LAD/UAs where the model-based estimates were less precise than the corresponding direct estimates.

- It produces estimates that are, in effect, a weighted combination of the direct Labour Force Survey estimate and the fixed effects estimate. The weighting is variable and dependent upon survey sample size with the estimate becoming closer to the direct estimate as the survey sample size, and hence confidence in the direct estimate, increases. Conversely, for LAD/UAs where the survey sample size is small the new model-based estimates will be very close to the previous fixed effects estimates.

To ensure that the model-based estimates are consistent with the Labour Force Survey published estimates at high geographical levels, the model-based estimates are constrained to the direct LFS estimates of unemployment for Government Office Regions (GORs) in England and the estimate for the countries of Wales and Scotland.

The estimates are *annual estimates*, i.e. they use an average of the previous twelve months claimant count totals and twelve months of survey data. The aim is to eventually publish updated estimates every quarter, however they will be still be based on the previous twelve months of data and hence be quite highly correlated with the estimates from the previous quarter because ¾ of the data is common to both sets of estimates. **It will therefore not be valid to compare estimates made for periods under 12 months apart.** In the case of periods a year apart a comparison *can* be made since the data is different, however due to the wave nature of the LFS there is still a 50% overlap in respondents hence the data are not independent. This means an approximation is made in estimating the precision of the difference. This is discussed further in section 4.3.

It is important to recognise that this model-based approach gives estimates that are of a different nature from the standard estimates from the Labour Force Survey. This is because they are dependent upon correctly specifying the relationship between unemployment and the covariate information. A brief explanation of the methodology is provided in the following section.

3. Guide to the Methodology

This section provides a description of the new methodology for producing model-based estimates of ILO unemployment levels and rates at the LAD/UA level. A full description of the methodology can be found in the Technical Report.

3.1 How do model-based estimates differ from standard survey estimates?

The standard methodology for making estimates of ILO unemployment for LAD/UAs from the Labour Force Survey produces unbiased estimates, which have a high level of sampling variability. To provide more accurate estimates a model-based approach has been adopted. This methodology is dependent upon the correct specification of the model, the quality and relevance of the input data sources and the fit of the model.

The premise behind the model-based methodology is that a strong relationship can be found between the variable of interest, in this case ILO unemployment as measured by the Labour Force Survey, and other auxiliary non-sample information (mainly provided from Census and administrative sources). Strength can then be borrowed from this relationship to provide more reliable estimates of ILO unemployment for LAD/UAs. The main source of auxiliary information here is administrative data on the number of recipients of unemployment benefit (the 'claimant count').

During its research ONS investigated a number of different relationships and different sources of information. It is satisfied that while there are some limitations with the methodology (see section 4 on Guidance on use and limitations of the estimates) the models are well specified and the modelling assumptions hold, so that the model-based estimates for ILO unemployment for LAD/UAs in Great Britain improve on those directly estimated from the Labour Force Survey.

3.2 The model for ILO unemployment -The area random effects model

For each year a model is defined that relates the Labour Force Survey estimates of the proportion of the population aged 16 and over in each age/sex class (male/female for age groups: 16 to 24; 25 to 49; 50 and over) within each LAD/UA to the following predictors:

- the logit[2] of the claimant count proportion in each age/sex class within each LAD/UA;
- the logit of the claimant count proportion in the LAD/UA;
- the age/sex group;
- the 12 government office regions; and
- the seven supergroups under the National Statistics 2001 Area Classifications for Local Authorities of Great Britain [4], referred to as *cluster* in this document

The outcome variable in the model is the logit of the Labour Force Survey direct estimate of the proportion of the population aged 16 and over in each age/sex group that are ILO unemployed.

Explicitly, the model for the unemployed proportion p_{ij} in age-sex class i, LAD j is:

$$
\begin{aligned}
logit(p_{ij}) = \beta_0 & \\
& + \beta_{1-5}[dummy\ age/sex\ variables] \\
& + \beta_{6-16}[dummy\ GOR\ variables] \\
& + \beta_{17-22}[dummy\ Cluster\ variables] \\
& + \beta_{23}[logit(claimant\ count\ proportion_{ij})] \\
& + \beta_{24-28}[logit(claimant\ count\ proportion_{ij})].[dummy\ age/sex\ variables] \\
& + \beta_{29}[logit(claimant\ count\ proportion_{j})] \\
& + U_j\ (area\ random\ effect\ term_j)
\end{aligned}
$$

This uses the same predictors as in the previous (fixed effects) model [1] but includes an extra term (U_j) which models between-area variation - a random effect term. It is the inclusion of this term, to model between area variation not explained by the auxiliary data, that differentiates the two methodologies.The inclusion of the area random effects term in the model gives the estimates the properties described in section 2.2. In effect the estimate is now a weighted average of the direct estimates and the fixed effects estimates where more weight is given to the direct estimate as the sample size increases. This ensures convergence to the direct estimate with

2 It is usual to model a function of the variable of interest when that variable is a proportion p. The logistic transformation is defined as follows: -

$$logit(p) = \log_e\left(\frac{p}{1-p}\right)$$

increasing sample size and avoids the situation of having to choose which estimate to use when the direct estimate is more precise than the model estimate.

3.3 Description of the data

The Labour Force Survey data:

For each period an annual sample is used. This consists of independent data from the preceding four quarters (e.g. for an estimate covering the 12 months March – February, data from the preceding quarters:- March to May, June to August, September to November and December to February are used[3]). The response variable is then the logit of the proportion of responses giving ILO unemployment as their economic status for each age/ sex group for each LAD.

Claimant count data:

This is the average of the monthly count of unemployment benefit claims for the same annual period. The data enters the model at two levels: the first, at the age/sex by LAD/UA level to reflect the observation that the relationship between ILO unemployment and the claimant count varies by age, sex and area; the second, at the LAD/UA level as one would expect the overall employment conditions within a LAD/UA to impact on the employment prospects of any one group within the LAD/UA, and the overall claimant count should be related to these conditions.

Indicator variables:

The remaining variables are included to account for regional and socio-economic factors.

3.4 Deriving the estimates

To derive the required estimates of unemployment for LAD/UAs, ONS:

- uses the model to estimate the proportion, *p*, of ILO unemployed for each age/sex group within each LAD/UA;

[3] These are the previously used seasonal quarters. Calendar quarters are now used : January – March, April – June, July – September and October – December.

- multiplies each of these proportions by the Labour Force Survey direct estimate of the resident population – this gives the model-based estimates of the ILO unemployment levels for each age/sex group;
- ensures these estimates are consistent with the published Labour Force Survey estimates of ILO unemployment for national age/sex groups, and for regional and socio-economic classes by using a standard method of calibration;
- adds the calibrated estimates for the age/sex groups within each LAD/UA to give the model-based estimates of ILO unemployment levels for LAD/UAs; and
- calculates the model-based estimates of ILO unemployment rates[4] for LAD/UAs by dividing the model-based estimates of ILO unemployment levels by an estimate of the economically active population. (This estimate is the sum of the Labour Force Survey direct estimate of the employed population and the model-based estimate of the ILO unemployment level.)

3.5 Quality measures for model-based estimates of ILO unemployment

The standard error is the key indicator of the quality of the model-based estimates. It represents 'uncertainty' in the estimation arising from the following sources:

- from using the model to make estimates of the proportion of ILO unemployed for classes within LAD/UAs;
- from sampling error due to using Labour Force Survey data to make estimates of the populations for classes within LAD/UAs; and
- a between area effect to allow for any between LAD/UA effects the model has missed.

There is an additional source of variability in the standard errors of the model-based estimates of ILO unemployment rates:

- that takes into account that the economically active population is estimated by adding the Labour Force Survey estimates of the employed population to the model-based estimates of the ILO unemployed level.

Each of the estimates is accompanied by a confidence interval which has been determined from the standard error.

[4] The unemployment rate differs from the value p in that it is the defined as the proportion of the number of *economically active* people who are unemployed rather than the proportion of the *working age population* who are unemployed, some of which are economically inactive for some reason.

For example, the model-based estimate of the ILO unemployment rate (UR) for Salisbury in 2001/02 is 2.7% with a standard error (SE) of 0.3%. The upper and lower 95% confidence limits are:

$$UR_{upper} = UR + 1.96\,SE$$
$$UR_{lower} = UR - 1.96\,SE$$

These gives [2.1%, 3.3%] as the 95% confidence interval around the estimate of 2.7%.

The interpretation of the 95% confidence is that, assuming the model holds, on average we would expect it to contain the true value 95% of the time.

Confidence intervals for the model-based estimates of the level of ILO unemployment are formed in the same way.

4. Guidance on use and limitations of the estimates

The model based estimates may be found here. Before using the estimates we recommend you read the following guidance notes.

The main limitation of the estimates for unemployment, either those estimated directly from the Labour Force Survey or model-based, is that they are *subject to variability*. ONS has produced confidence intervals associated with the model-based estimates of both rates and levels for each LAD/UA in order to make the accuracy of the estimates clear. This has been described in section 3.5.

Four further limitations of the estimates must be considered:

- the consistency and accuracy of unemployment estimates for other, often larger geographical areas;
- the conclusions that may be drawn from the estimates on the overall distribution of unemployment and the ranking of specific areas;
- consistency with different time periods; and
- consistency with other labour market variables.

These are considered in the following sections.

4.1 Consistency and accuracy of estimates for other geographical areas

The model-based methodology produces LAD/UA level estimates of annual unemployment rates and levels. These LAD/UA level estimates can be aggregated to provide unemployment estimates for larger geographical areas, however, this method is approximate and hence it is not possible to assess the precision of the aggregated estimates .

The model-based methodology has been developed to ensure that the LAD/UA estimates are constrained to direct survey estimates from the Labour Force Survey for GORs in England and the estimate for the countries of Scotland and Wales. For example, the model-based estimates for the LAD/UA levels in Wales when added together correspond to the Labour Force Survey estimate of annual level of unemployment for Wales. However, the model-based estimates may not be consistent with Labour Force Survey estimates of unemployment for other geographical levels.

4.2 Distribution and ranking of unemployment levels and rates

In common with any ranking based on estimates, great care must be exercised in interpreting the ranking of the LAD/UAs. One needs to take into account the variability of the estimates when using these figures. For example, the confidence interval around the highest ranked LAD/UA suggests that the estimate lies among the group of LAD/UAs with the highest unemployment levels (or rates) rather than being the LAD/UA with the highest unemployment level (or rate). Estimates for two particular LAD/UAs can only be described as significantly different if the confidence intervals for those estimates do not overlap.

Although these model-based estimates can be used to rank LAD/UAs by unemployment rate or level, they cannot be used to make any inferences on the distribution of unemployment across the LAD/UAs. The estimation procedure will tend to shrink estimates towards the average level of unemployment for the whole population, so model-based estimates at each end of the scale tend to be over or under-estimated. Nevertheless estimates can be used to make certain inferences, e.g. the rate of unemployment for LAD/UA A is greater than that for LAD/UA B (if the appropriate confidence intervals do not overlap). However, making assertions such as x% of LAD/UAs have a rate of unemployment greater than y is not valid.

4.3 Consistency with different time periods

The model has been developed to provide the best estimates of unemployment in a survey year and so some caution has to be exercised when interpreting differences in the estimates over time as a measure of change.

It has been explained in 2.2 that comparisons must not be made for estimates for two periods less than a year apart due to data in common. In the case of two annual periods which do not overlap, *year a* and *year b*, a confidence interval for the difference between a model-based estimate for a particular LAD/UA may be constructed using the following as the standard error[5]: -

$$SE_{diff} = \sqrt{SE^2_{year\,a} + SE^2_{year\,b}}$$

The resulting 95% confidence interval for the difference Δ between the two model-based estimates is therefore:

$$(\Delta - 1.96.SE_{diff}\,,\Delta + 1.96.SE_{diff})$$

This result is an approximation when the years are directly successive as the survey data is then not independent because of respondent overlap (see 2.2) and will therefore exhibit some correlation. However this correlation is almost certainly positive, so it follows that the standard error will be over estimated and thus the confidence interval will be conservative (larger than necessary). As the distance between time periods increases, the respondent overlap decreases until with time periods two years apart it disappears and then the standard error is no longer an approximation.

[5] The expression uses the standard errors for each of the individual years. While these are not displayed, they can be calculated as : (confidence interval maximum value – estimate)/1.96.

4.4 Consistency with other labour market variables

When estimates of ILO unemployment, employment and working age inactivity[6] are taken directly from the Labour Force Survey these estimates added together will equal the total working age population. However, model-based estimates of ILO unemployment are not equal to the Labour Force Survey direct estimates of ILO unemployment, and so the model-based estimates of ILO unemployment plus the available estimates of employment and working age inactivity (from the Labour Force Survey) will not add to the working age population (see section 6 Future plans).

4.5 Example of data use

Given that the model-based estimates are subject to limitations an example of appropriate and inappropriate use of the estimates has been produced.

4.5.1 LAD/UA comparisons

When comparing two model-based estimates, one LAD/UA may only be said to have a statistically significant lower or higher unemployment level or rate if the confidence intervals (section 3.5) for the two LAD/UAs do not overlap. For example, using Table 4.1 it may be said that LAD/UA C has a significantly lower model-based unemployment level than LAD/UA A since the 95% confidence intervals do not overlap. However, it would be wrong to say that LAD/UA B has a significantly lower model-based unemployment level estimate than LAD/UA A, since the confidence intervals overlap.

	95% confidence intervals for the unemployment estimate		
	Estimate	**Lower Confidence Limit**	**Upper Confidence Limit**
LAD/UA A	1660	1260	2060
LAD/UA B	1110	910	1310
LAD/UA C	1080	960	1200

Table 4.1: Model-based estimates of unemployment level and associated confidence intervals for three LAD/UAs.

[6] The working age economically inactive population is the remainder of the working age population who are neither employed or ILO unemployed. Thus the three 'states' of employment add to the working age population.

5. Further analysis and validation of the estimates.

Model-based estimates and their standard errors using the random effects methodology have been produced for LAD/UAs in England, Scotland and Wales for the years March – February 1996/97 through to 2003/04 in addition to the current estimation covering the period January – December 2005. Previous model-based estimates using the previous fixed effect methodology have been published as experimental statistics and published for the years March – February 1995/96 to 2000/01. As well as the methodology used, the current estimates differ in two other substantial ways:

- they use rebased data, that is the population estimates and Labour Force Survey direct estimates used in the model have been adjusted following the 2001 Census [5];

- the National Statistics 2001 Area Classification for Local Authorities [4] has been used as a covariate data source in the model in place of the previous 1999 revision. (The 2001 classification system was developed as a result of the 2001 Census).

Thus, comparison between the new and previously published model-based estimates should not be made.

The success of the model may be judged by looking at the 'publication rate' and the reliability of the model-based estimates in comparison with the direct Labour Force Survey estimates. The publication rate is defined as the number of LAD/UAs where the model based standard error is no greater than 20% of the value of the estimate. This is referred to as the publication rate as it was previously used by the ONS as a criterion by which to judge whether a (direct) estimate of unemployment could be published. The gain of the model is defined as the ratio of the direct estimate standard error to the model-based standard error.

Ideally the model-based estimates will be more reliable than the direct Labour Force Survey estimates and will thus have values of gain greater than one. This will avoid

any issues as to which estimate to use when it comes to publication. Table 5.1 shows the number of LAD/UAs where estimates are publishable each year using the Labour Force Survey estimates, the previous fixed effects model-based estimates and the new random effects model-based estimates for the years 96/97 through 02/03.

Table 5.1 also shows the number of LADs where the model-based estimates are less reliable than the Labour Force Survey direct estimates.

	Model-based estimates are less reliable than the direct Labour Force Survey estimates		**Publication level (from a total of 406 LADs)**		
Year : March to February	**Previous model – fixed effects**	**New model – random effects**	**Previous model – fixed effects**	**New model – random effects**	**Labour Force Survey direct**
96/97	4	0	406	406	23
97/98	2	0	406	406	82
98/99	5	0	406	406	72
99/00	4	0	405	406	90
00/01	1	0	406	406	115
01/02	0	0	404	404	138
02/03	0	0	406	406	138

Table 5.1. Gain and publication rates for the model-based and Labour Force Survey direct estimates of unemployment levels.

It is clear from table 5.1 that both methodologies perform well. Nearly every estimate is publishable under the 20% rule (the exceptions in the years 99/00 and 01/02 only just exceed this figure). The model-based estimates are always more reliable than the Labour Force Survey direct estimates under the new methodology, indicating an improvement over the original methodology where for a small number of LADs the Labour Force Survey direct estimates are more reliable.

5.1 Effect of rebasing and reclassification

The set of model-based estimates to which this document refers were calculated using the estimates of LAD populations rebased to estimates given by the 2001 Census. Also the 2001 ONS area classifications were used in place of the 1999 ONS area classifications used in the previous model.

Historic data for 1996/97 to 2003/04 was used in order to gauge the impact of these two changes. The model was firstly fitted on the data based on the earlier population estimates and using the old classifications. Then the model was then fitted to the rebased data still using the old classifications. Finally the model was fitted to the rebased data now using the 2001 area classifications.

Rebasing resulted in a decrease in model-based estimates of levels by an average of 2.8% and rates by an average of 2.9%. The additional effect of reclassification on the rebased levels and rates was much smaller, showing an average increase of 0.18% and 0.19% respectively.

Although the overall effect of reclassification on the estimates is small, in some individual cases the effect has been quite large. For example, in 2001/02 the model-based estimates for rates & levels for Barking & Dagenham are lower under the new classifications (‘*Cities & Services’*) than they would have been under the old classifications (‘*Mining & Manufacturing & Industry’*). Conversely, the estimates for rates and levels for Darlington in the same year are higher. Darlington changes from ‘*Coast and services’* to ‘*Mining & Manufacturing*’ under reclassification.

5.2 Validation of the model and estimates

A number of diagnostic checks have been used to assess the appropriateness of the models developed for producing estimates of unemployment. These involve using direct survey estimates which are unbiased to check for bias in the model-based ones, splitting the data into two sets and cross validating a model based on one half of the data by applying it to the other half and thirdly checking model stability by applying a model based on a previous year’s data to the current year. Also an extensive

simulation exercise was carried out in order to check that the size of the model-based confidence intervals were realistic and not too small. These are described in detail in the Technical Report. The analysis shows that in general the models are well specified and the assumptions are sound. This provides confidence in the accuracy of the estimates and the associated confidence intervals. In addition the methodology used to produce the model-based estimates has undergone an academic review.

As well as validating the methodology for making the estimates, it is necessary to validate the estimates themselves. This is of crucial importance in establishing the plausibility of the model-based estimates. The validation of the model-based estimates was two-fold: Firstly a comparison of the model–based estimates with an external data source (the Census) has been completed; at the same time the estimates and supporting documentation were sent to members of the Central and Local Government Information Partnership Labour Market Statistics subgroup (CLIP/LMS) for their comments as to the plausibility and usefulness of the estimates; These are described below.

5.2.1 Census comparison

The 2001 Census provided a snapshot of the entire UK population on a particular date (29th April 2001). A range of questions was asked including questions on employment. Although the Census differs from the Labour Force Survey in several ways[7], it does provide an independent data source that can be used for comparison with the model-based estimates.

The 2001/02 model-based estimates were compared against the Census estimates at LAD/UA level. The comparison was carried out for levels and rates of unemployment for all LAD/UAs in Great Britain and repeated by GOR; by high unemployment areas; by low unemployment areas and by socio-economic groups (cluster).
Given the differences between the data sources the aim of the validation was to see whether there is broad agreement or any areas of large disagreement.

[7] The Census is a self completion questionnaire whilst the Labour Force Survey is a mixture of face to face interviewing and telephone interviewing. There are also other slight differences such as the way full time students are presented.

5.2.1.1 Comparison with the Census – GB Levels

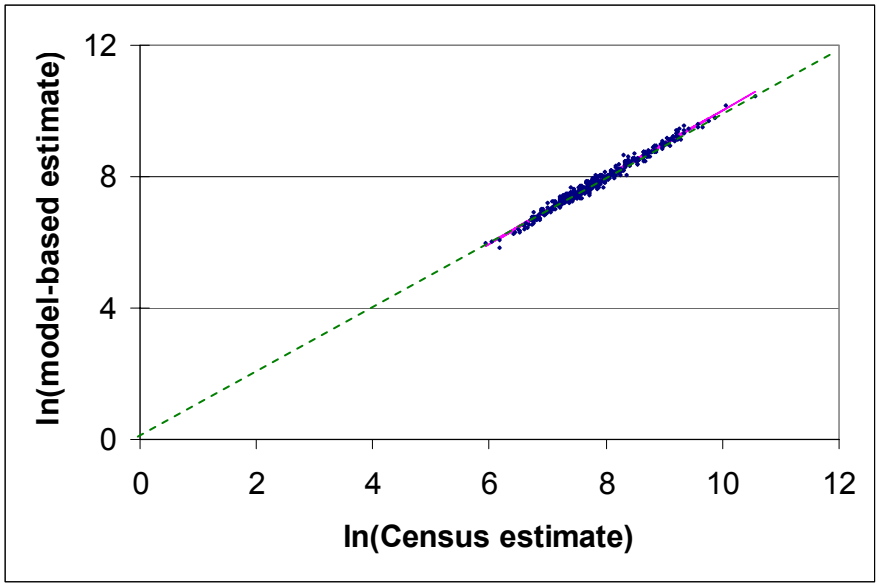


Figure 5.1. Model based-estimates for 2001/02 level plotted against Census estimates of unemployment level.

Figure 5.1 displays a comparison of the model-based estimates of unemployment levels with the Census estimates on the natural logarithm scale. All 406 LAD/UA's are included. The line of best fit is in magenta and the line Y=X is shown in green – clearly there is good agreement. Analysis shows that these two lines are not statistically different at a 95% level.

A log scale has been used in this analysis because it prevents the few larger LAD/UAs having undue influence on the fitting of the regression line to the data.

The Census comparison was carried out in order to look for any systematic errors in the model-based estimates. The correlation between the model-based estimates and Census is strong and there are no patterns in the plots of the model-based estimates against Census that would suggest systematic errors. We would not expect the Census and LFS estimates to be exactly the same due to differences in definitions, time periods and collection methods; the comparisons shown here give confidence in the estimates.

5.2.2 User validation

Labour Market Division of ONS carried out a user validation exercise of the estimates produced by the random-effects model. Users consulted included the members of the CLIP/LMS sub-group, others who had taken part in the user validation of the original (fixed-effects) model, and others suggested by members of the CLIP group.

For the user validation, estimates for seven years (1996/97 to 2002/03) and methodological documentation were sent to respondents, and they were asked to complete a questionnaire on the plausibility of the estimates for the respondents' areas.

Six responses were received, though these were not all from single local authorities. For example, one was a joint response from a unitary authority and a county council. Also, a regional body and a devolved administration responded, covering many local authorities.

A few specific concerns were raised. In one authority which produces its own estimates of unemployment, census population estimates were revised after the rebasing had taken place. This meant that the LFS data which weight to the population would give unemployment proportions inconsistent with the revised populations. As these proportions are input to the model this will affect the authority's model-based estimate. In the case of another authority there was an issue concerning the use by the claimant count of frozen 1991 wards as the building block for higher geographies. This is not the fault of the model and will be addressed when the claimant count uses the output area as the geographic building block.

In summary, the respondents agreed that the model-based estimates were a positive step, and better than what was currently available. The specific concerns which were raised have been satisfactorily addressed, and they were not due to deficiencies in the modelling methodology.

5.3 Conclusions

The diagnostic analysis has confirmed that the models are well specified, stable and the assumptions are sound. The additional verification exercises taken together have ensured that the models developed are robust, make the best use of the available data and that the model-based estimates are plausible and informative to users. This gives confidence in the accuracy of the estimates and the associated confidence intervals.

In addition to this in-house analysis, the methodology has undergone an academic review and been presented at conferences [6].

6. Future plans

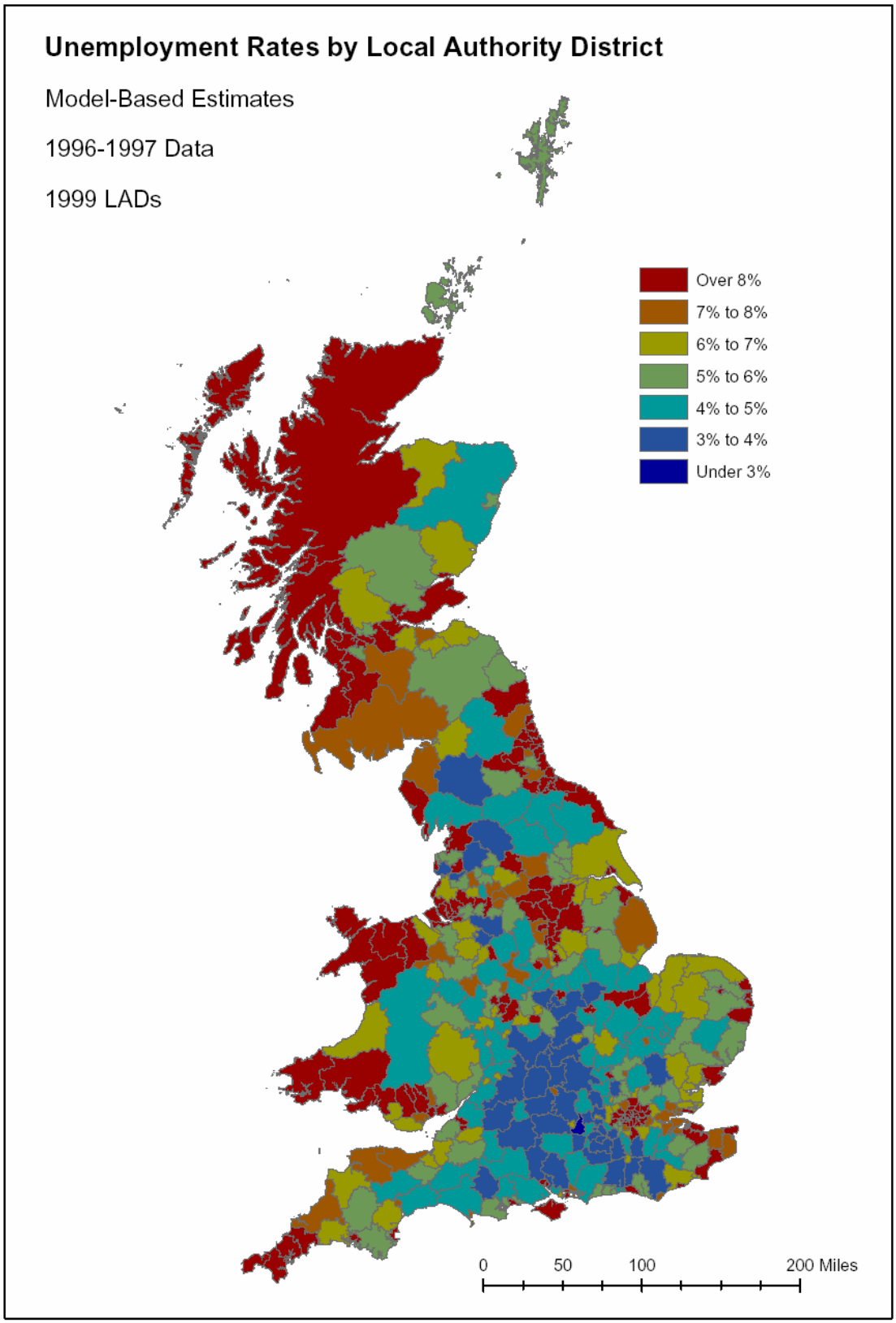
These new estimates are an improvement on previous work and ONS are currently planning further methodological developments.

Under the Guidance and Limitations section (4.4), it was stated that the model-based unemployment estimates would not to be consistent with other labour market variables estimated directly from the LFS such as economic inactivity at LAD/UA level. The ideal solution is to calculate consistent model-based estimates for ILO unemployment, employment and inactivity simultaneously. However, this is not an easy task and is beyond the scope of the initial remit to produce estimates of ILO unemployment at LAD/UA level. ONS is carrying out research to develop methodology for such simultaneous estimation.

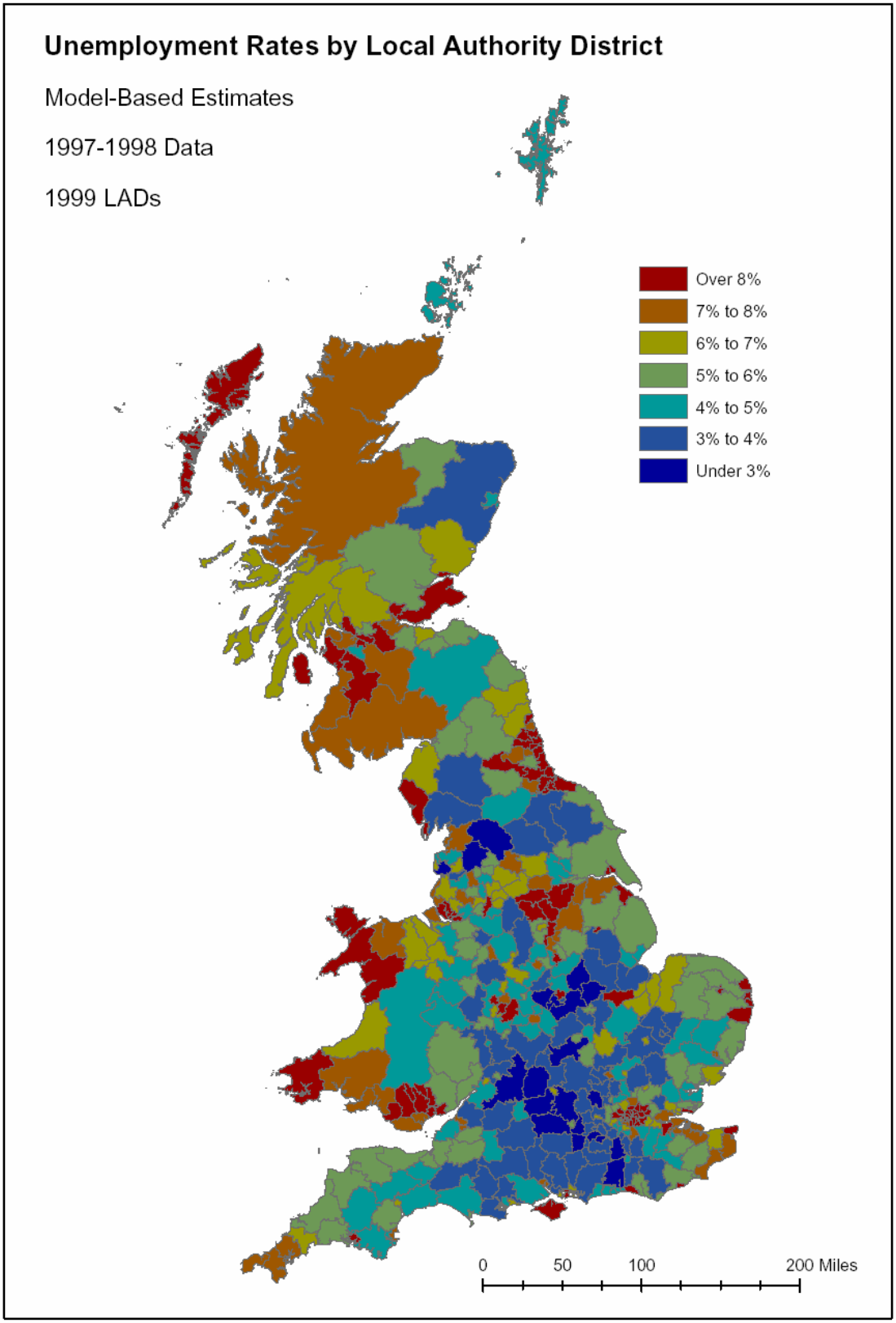
7. Maps

It is informative to display the model-based LAD/UA level estimates of unemployment rate on thematic maps. Displaying levels is less so because of widely differing populations between some adjacent LAD/UAs. Maps 7.1 through Map 7.9 shows the model-based unemployment rates for the years March to February 1996/97 through 2003/04 and January to December 2005.

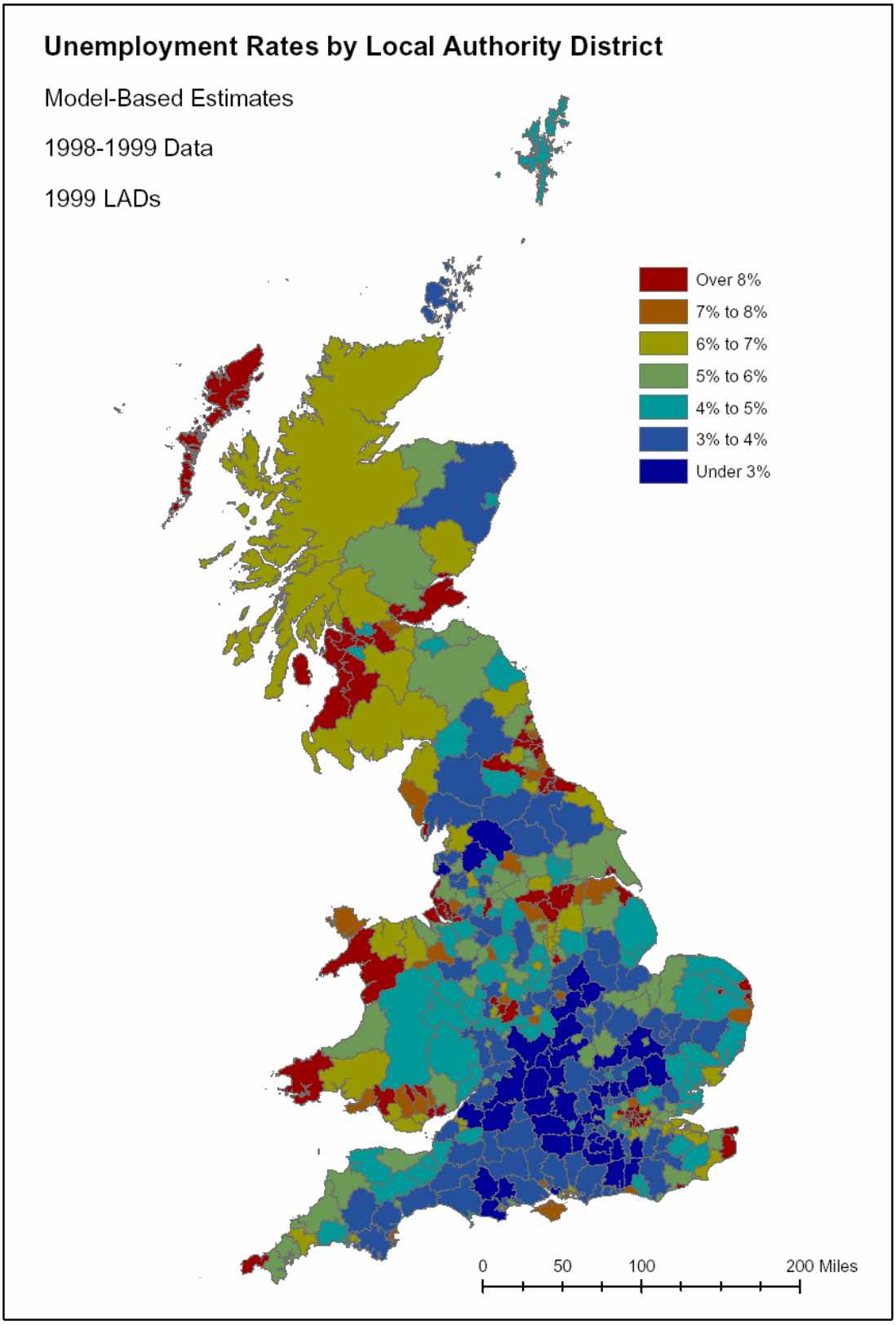
The colours and interval ranges in the map have been chosen to aid interpretation. There are more intervals covering the areas of lower unemployment because these contain the majority of LAD/UAs. The colours darken at either extreme, so the dark blue represents the LAD/UAs with the lowest unemployment rates, whilst the dark red representing the areas of highest unemployment.



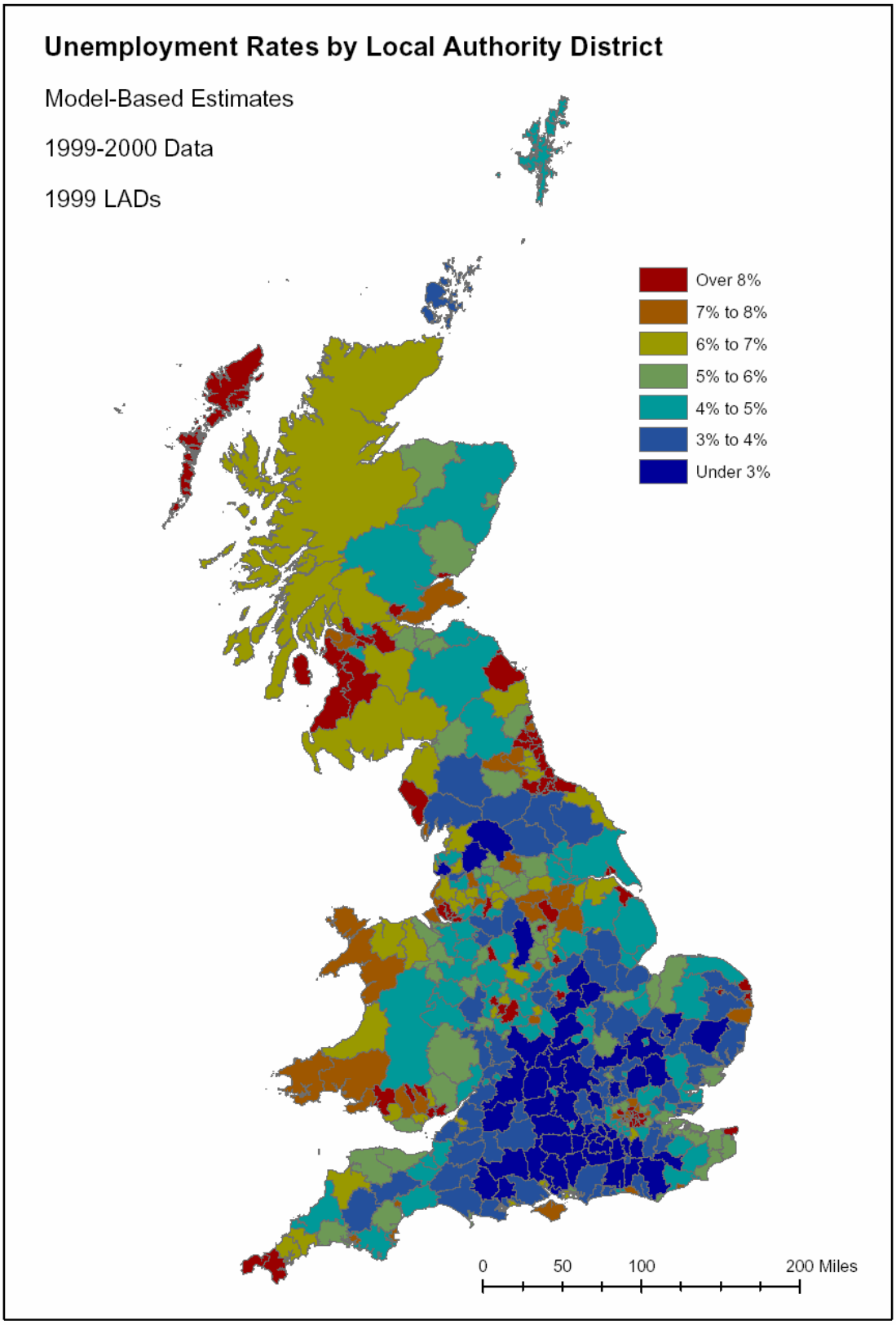
Map 7.1. Model-based unemployment rates for the year 1996-1997



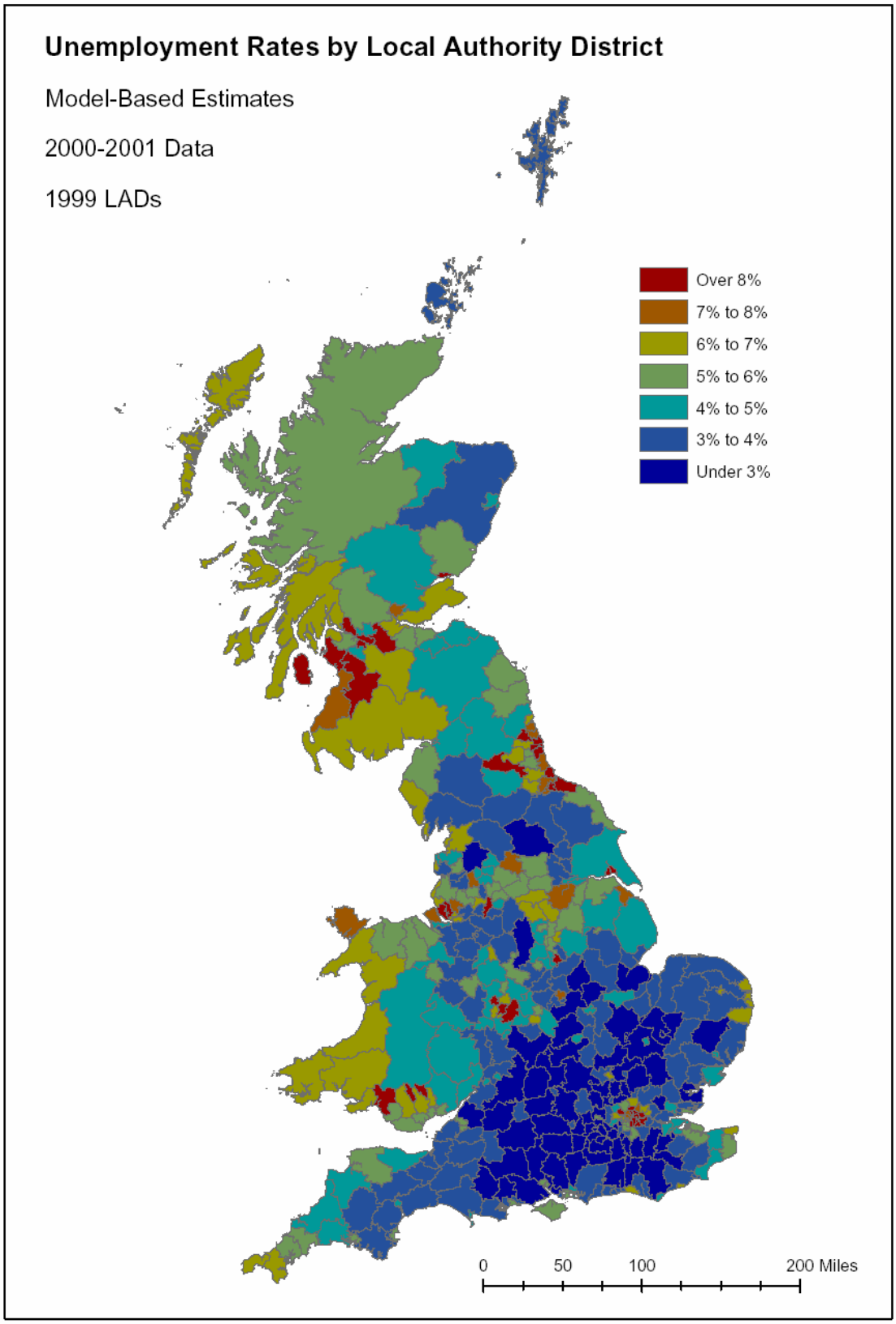
Map 7.2. Model-based unemployment rates for the year 1997-1998



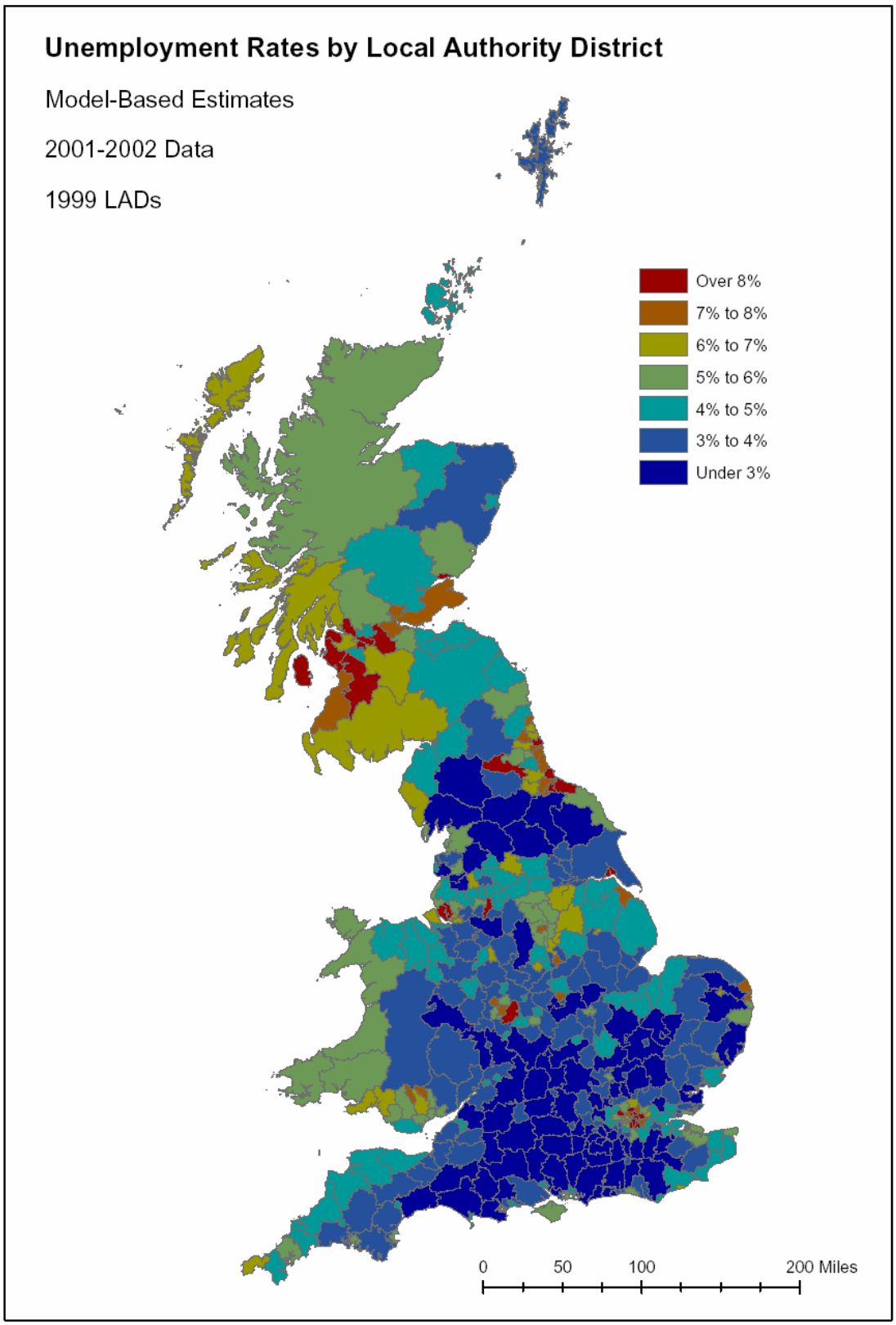
Map 7.3. Model-based unemployment rates for the year 1998-1999



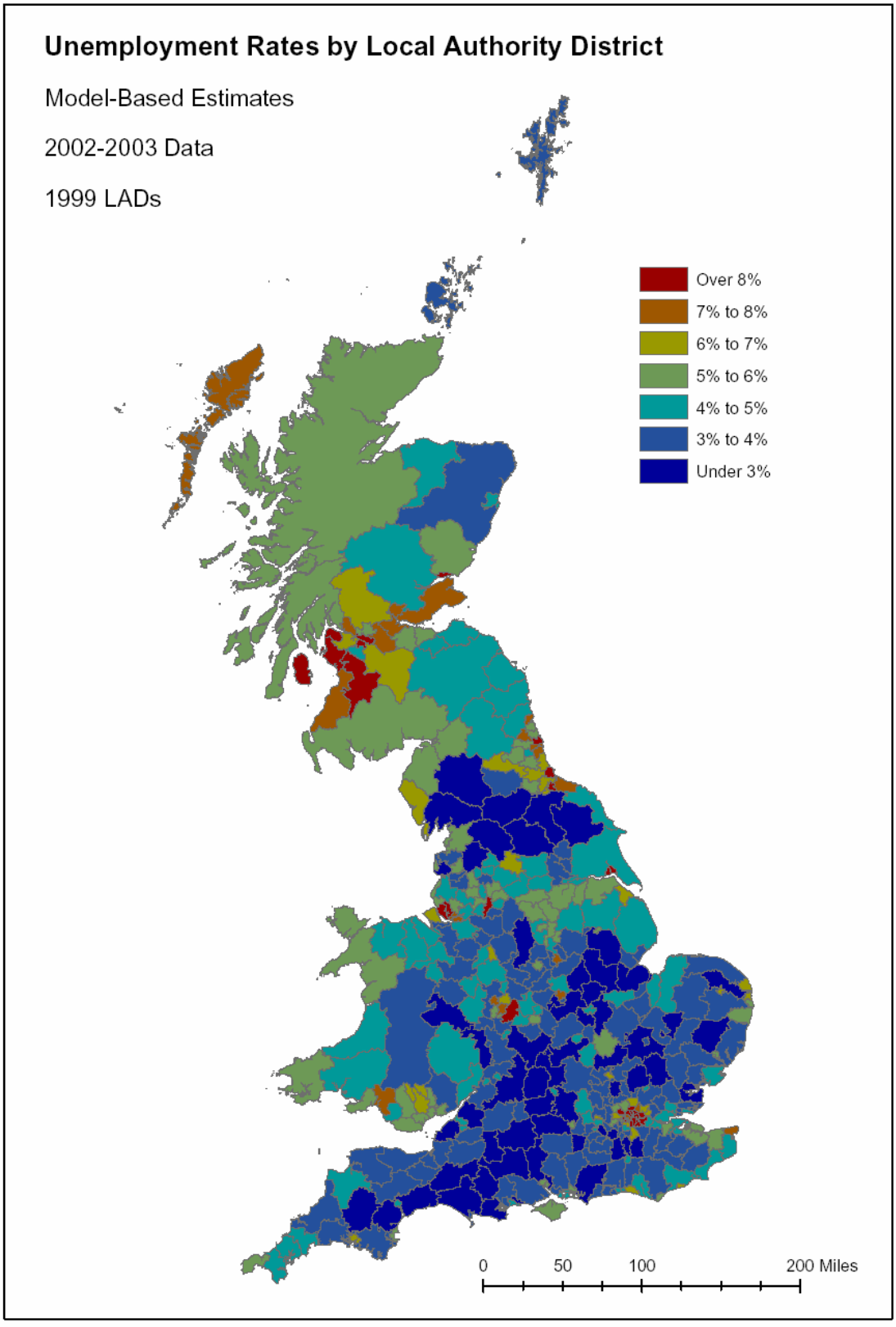
Map 7.4. Model-based unemployment rates for the year 1999-2000



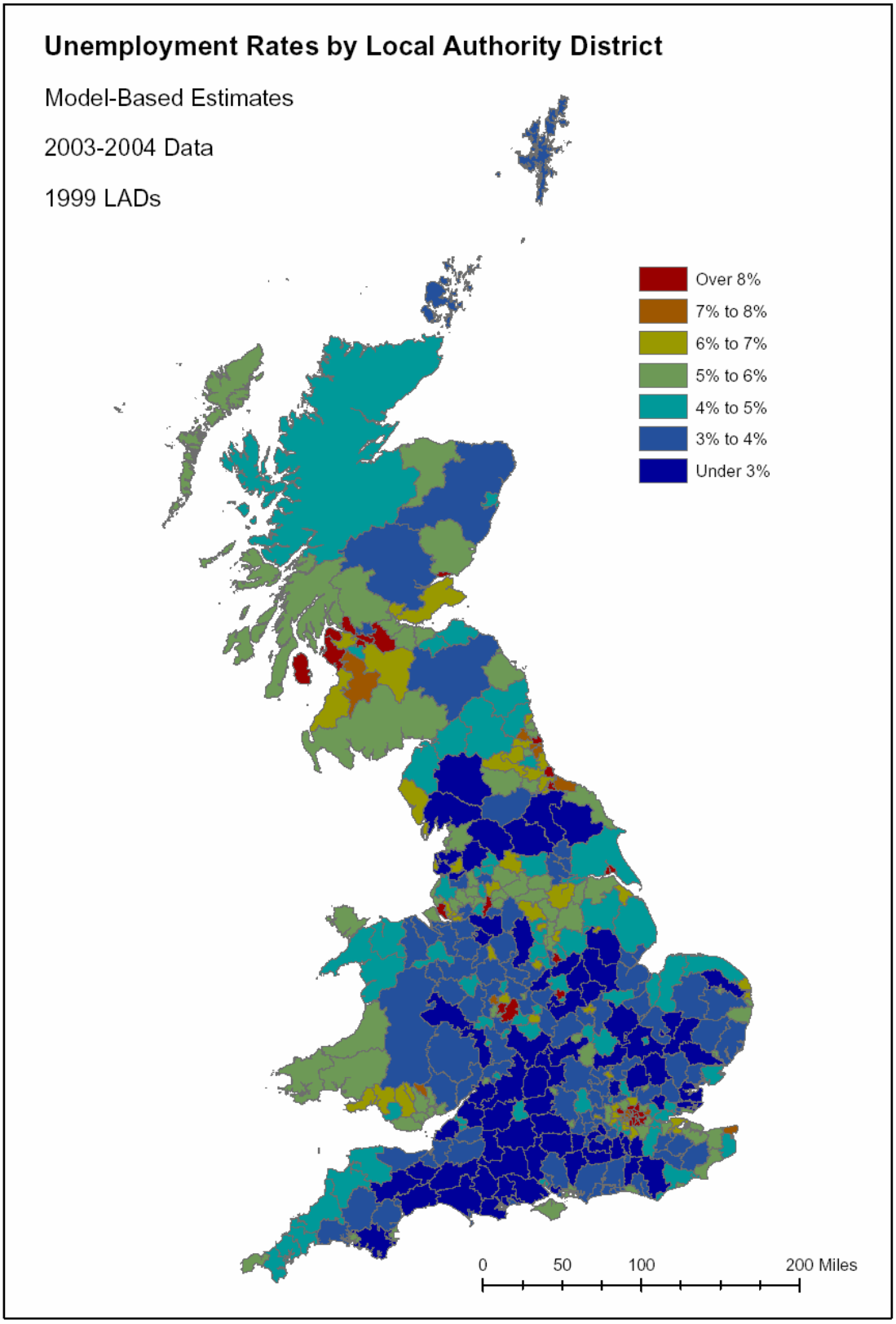
Map 7.5. Model-based unemployment rates for the year 2000-2001



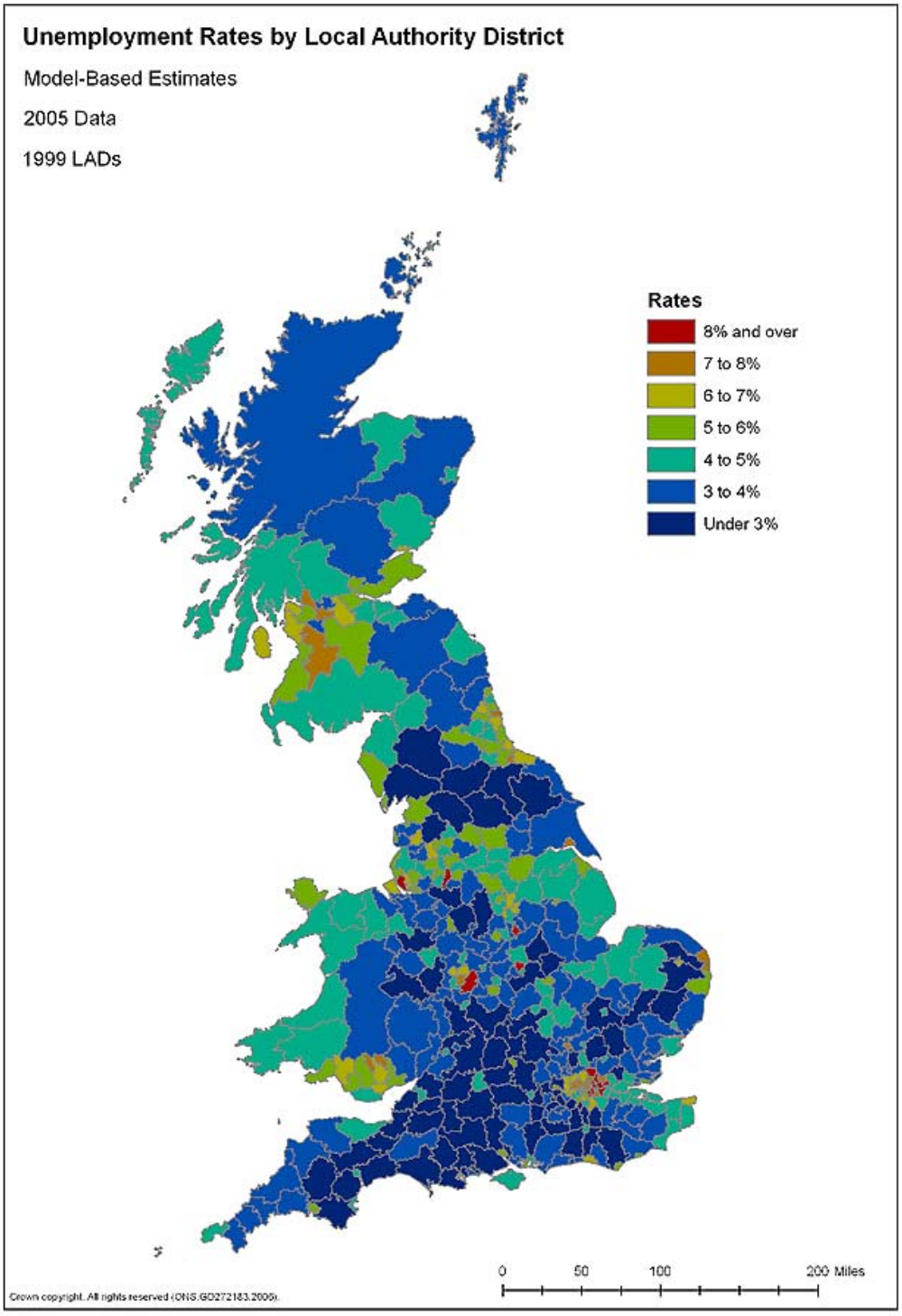
Map 7.6. Model-based unemployment rates for the year 2001-2002



Map 7.7 Model-based unemployment rates for the year 2002-2003



Map 7.8 Model-based unemployment rates for the year 2003-2004



Map 7.9 Model-based unemployment rates for the year 2005

References and links

[1] Development of improved estimation methods for local area unemployment levels and rates.
Authors: Gary Brown, Marie Cruddas, David Hastings. Labour Market Trends, vol. 111, no 1
http://www.statistics.gov.uk/cci/article.asp?id=372

[2] Summary publication accompanying the publication of the 2003 estimates November 2004
http://www.statistics.gov.uk/downloads/theme_labour/ALALFS/AnnexA.pdf

[3] Labour Force Survey User Guide – Volume 6: Local Area Data. November 2004. http://www.statistics.gov.uk/downloads/theme_labour/vol6_2003.pdf

[4] National Statistics 2001 Area Classification for Local Authorities.
http://www.statistics.gov.uk/about/methodology_by_theme/area_classification

[5] Keeping Labour Force Survey in Line with population. Labour Force Survey revision plans.
http://www.statistics.gov.uk/cci/nugget.asp?id=342

[6] ASC 2005 conference. Maximising data value.
http://www.asc.org.uk/Events/Sep05/Slides/Curtis.pps