**Technical appendix**

Jon Minton

6/1/2016

**Introduction**

The purpose of this appendix is to describe the modelling strategy in more detail than the space constraints in the main document will allow. In essence, the modelling strategy involves three stages:

1. Model Comparison: Comparison of model specifications for statistical fit, to select a ‘best model’
2. Model Simulation: Projection of estimated mortality rates for different ages and sexes for the years 2010-2015.
3. Death count estimation: Construction of period lifetables based on projected values from stage 2, and comparison with actual death counts for each age, sex and year.

**Overall modelling strategy**

The overall aim of the modelling was to produce plausible estimates of the expected number of deaths, at each age in single years, and for both sexes, over the period 2010-2015 inclusive, given that age-sex specific mortality rates have tended to change at most ages over the period 1961-2015. The overall modelling strategy was to fit the same model specification to each of 180 distinct age and sex specific log mortality rates. A number of different model specifications were explored, and for each model family the age and sex specific log mortality rates were regressed against a number of terms that included time (years since the start of the time series) as a predictor variable. In the simpler model specifications, only a linear effect on time (‘year’) was included, and in more complicated model specifications various polynomials of time (year2, year3 and year4) were also included. In addition, some model specifications included dummy variables indicating either the presence of a Labour government (LAB), or a recession (REC). Model specifications differed also in whether they included interaction terms between LAB, REC, and the polynomial terms for time.

Where a model specification was used that included an interaction term between LAB and year and not higher polynomials on year, this can be interpreted as a change in the gradient of change in log mortality rates over time during years in which the Labour government was in power compared with other years. A similar interpretation can be provided to models including interactions between REC and year.

The R programming language was used to produce the analyses, and the function lm used to produce the models.

**Model Comparison**

A series of 24 model specifications were compared. In each of these model specifications the log mortality rates for each of 180 separate age/sex combinations were regressed against a different series of possible explanatory variables. The data used were from the ONS. For years 1961 to 2014 the data included and provided population and death counts for males and females separately for each age in single years from birth to over 89 years of age.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Number** | **Model Specification**  (R code using lm function) | **Number of terms** | **Comments** |
| 1 | lmr ~ year \* (lab + recession) | 6 | Linear trend over time, intercept and trend varying with government and economic recession |
| 2 | lmr ~ year \* (lab + recession) + I(year^2) | 7 | As (1), but with non-interacting 2nd order polynomial with year |
| 3 | lmr ~ year \* lab | 4 | Linear trend over time, intercept and trend varying with government |
| 4 | lmr ~ year \* recession | 4 | Linear trend over time, intercept and trend varying with economic recession |
| 5 | lmr ~ year + lab | 3 | Linear trend over time, intercept varying with government |
| 6 | lmr ~ year + recession | 3 | Linear trend over time, intercept varying with economic recession |
| 7 | lmr ~ year + I(year^2) | 3 | Nonlinear trend over time: first two polynomials. |
| 8 | lmr ~ year + I(year^2) + I(year^3) | 4 | Nonlinear trend over time: first three polynomials. |
| 9 | lmr ~ recession \* (year + I(year^2)) | 6 | Nonlinear trend over time, first two polynomials. Intercept and interaction terms with economic recession. |
| 10 | lmr ~ lab \* (year + I(year^2)) | 6 | Nonlinear trend over time, first two polynomials. Intercept and interaction terms with government. |
| 11 | lmr ~ lab + (year + I(year^2)) | 4 | Nonlinear trend over time, first two polynomials. Separate intercept with government. |
| 12 | lmr ~ lab \* (year + I(year^2) + I(year^3)) | 8 | Nonlinear trend over time: first three polynomials. Intercept and interaction terms with government. |
| 13 | lmr ~ lab + (year + I(year^2) + I(year^3)) | 5 | Nonlinear trend over time: first three polynomials. Intercept with government. |
| 14 | lmr ~ recession \* (year + I(year^2) + I(year^3)) | 8 | Nonlinear trend over time: first three polynomials.  Intercept and interaction terms with recession. |
| 15 | lmr ~ recession + (year + I(year^2) + I(year^3)) | 5 | Nonlinear trend over time: first three polynomials.  Intercept with recession. |
| 16 | lmr ~ (year + I(year^2) + I(year^3) + I(year^4)) | 5 | Nonlinear trend over time: first four polynomials. |
| 17 | lmr ~ recession \* (year + I(year^2) + I(year^3) + I(year^4)) | 10 | Nonlinear trend over time: first four polynomials.  Intercept and interaction with recession. |
| 18 | lmr ~ recession + (year + I(year^2) + I(year^3) + I(year^4)) | 6 | Nonlinear trend over time: first four polynomials.  Intercept with recession. |
| 19 | lmr ~ lab \* (year + I(year^2) + I(year^3) + I(year^4)) | 10 | Nonlinear trend over time: first four polynomials.  Intercept and interaction with government. |
| 20 | lmr ~ lab + (year + I(year^2) + I(year^3) + I(year^4)) | 6 | Nonlinear trend over time: first four polynomials.  Intercept with government. |
| 21 | lmr ~ lab \* year + I(year^2) + I(year^3) + I(year^4) | 7 | Nonlinear trend over time: first four polynomials.  Intercept with government and interaction with linear year term only. |
| 22 | lmr ~ lab + year + I(year^2) + I(year^3) + I(year^4) | 6 | Nonlinear trend over time: first four polynomials.  Intercept for government type. |
| 23 | lmr ~ lab \* recession + year + I(year^2) + I(year^3) + I(year^4) | 8 | Nonlinear trend over time: first four polynomials.  Intercepts for government and recession, and interaction between them but not time. |
| 24 | lmr ~ lab \* recession \* year + I(year^2) + I(year^3) + I(year^4) | 11 | Nonlinear trend over time: first four polynomials.  Intercept and linear trend term for government type, recession, and interaction between government and recession. |

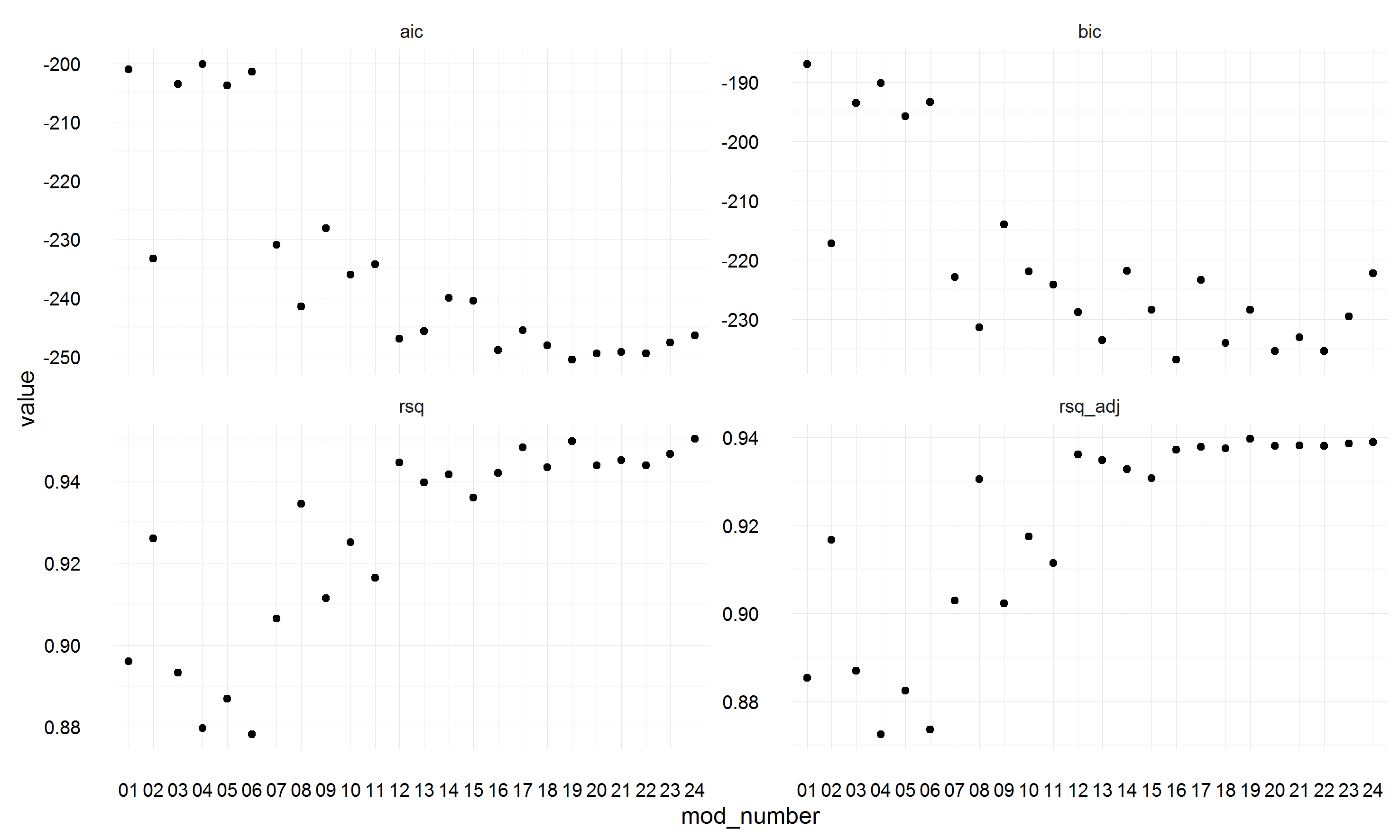
The variables lab and recession are both dummy terms indicating years in which there was a Labour government, and years in which there was an economic recession respectively. Therefore, the lab variable had a value of 1 for the years 1964, 1974-1978, and 1997-2010 inclusive, and a value of 0 for other years; and the recession variable had a value 1 for the years 1961, 1973-1975, 1990-1991, and 2008-2009 inclusive, and a value of 0 for other years.

For the years 1961 to 2014 inclusive, population counts and death counts were available disaggregated by ages in single years from 0 to 103 years, and for the year 2015 the population counts and death counts were available disaggregated by ages in single years from 0 to 89 years of age. For comparability with the 2015 data release, therefore, all analyses were performed only on ages 0 to 89 years.

**Comparison of model families**

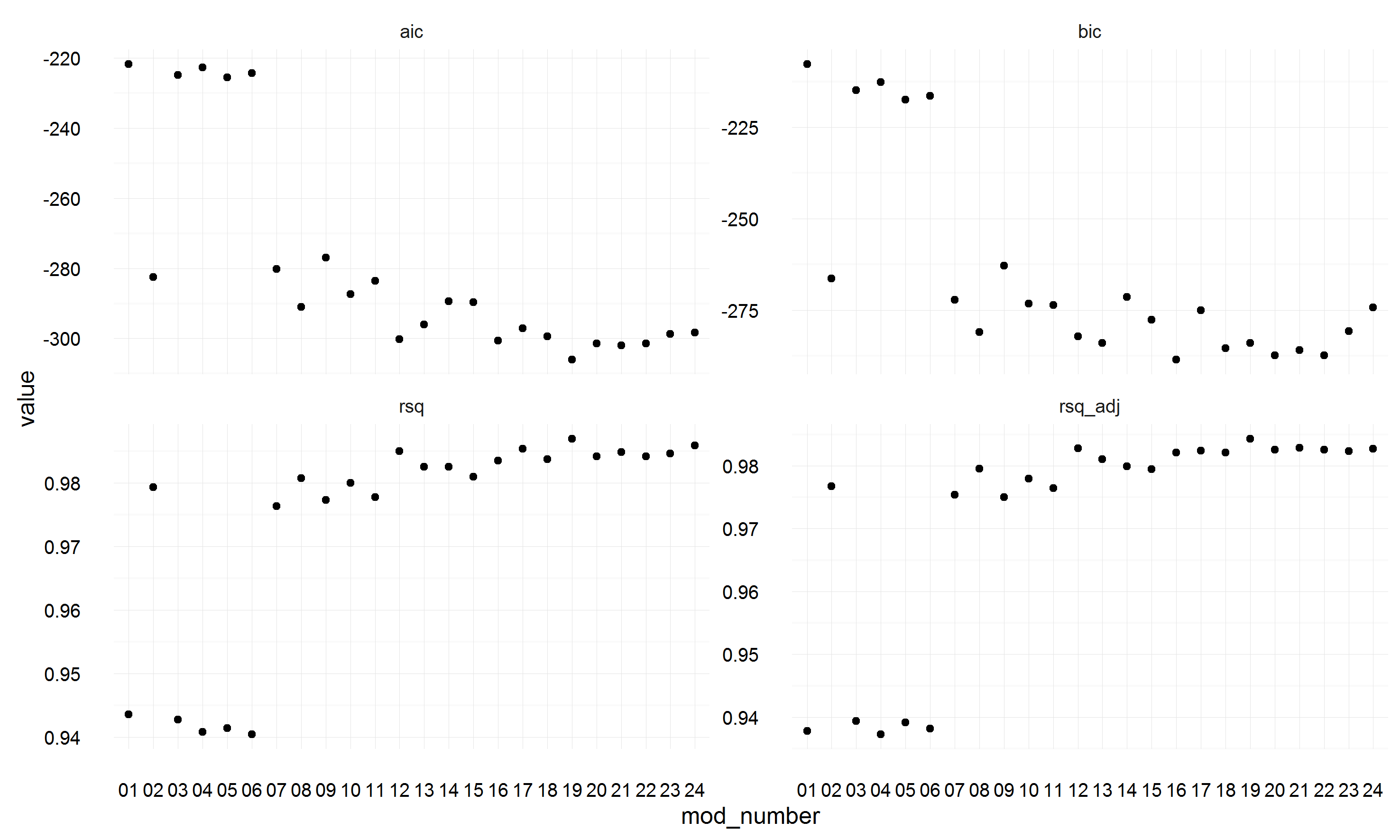
Each model specification creates 180 separate regression models, one for each sex, and each age from 0 to 89 years. Not all models are nested (i.e. satisfy the condition that one model, the restricted model, can be expressed as another model, the unrestricted model, with certain terms set to zero), and so the two standard metrics for comparing penalised model fit, the Akaike Information Criterion (AIC) and the related Bayesian Information Criterion (BIC), were used to compare between model families, along with R-squared and Adjusted R-squared. As each model family produced 180 separate regression models, the mean AIC, mean BIC, R-squared and Adjusted R-Squared values from all models in the same family were compared in the first instance.

The figure below shows the mean AIC and BIC (lower is better), R-squared adjusted R-squared values (higher is better) for each of the 24 model families.



It is clear from this figure that different measures give different indications about which model family is ‘best’, with AIC indicating that model 19 has the highest penalised model fit, whereas BIC suggests model 16 has the best fit. Model 19 also appears to have almost the highest R-squared, and the highest adjusted R-squared scores.

Given the equivocal results above, a further analysis was performed using only the models at age 0, and 50-89 years, as deaths at these ages contribute disproportionately to deaths overall in any given year. The results of this subgroup analysis are shown below:



Within these age groups, model 19 had the lowest AIC and highest R-squared and adjusted R-squared values, as well as amongst the lowest BIC value. For this reason we decided to use model 19 in all subsequent analyses.

**Model simulation approach**

After the best performing model family was identified, the next stage was to produce counterfactual

### **Model**

For each sex, and for each age in single years, a, from birth to 95 years old, a separate linear regression model was fit with the following specification:

|  |  |
| --- | --- |
|  | (1) |

Where is the mortality rate (death count divided by population count) in year t, at age a, and for sex s; t is year; L is a dummy variable indicating the years, 1997 to 2010, in which New Labour were in government; R is a dummy variable indicating 2008 and 2009, the years in which the UK economy entered a recession as a result of the GFC, and is an error term. The R term is included to capture any additional short-term changes in mortality rates to be captured in a separate term rather than influence the coefficients including New Labour years, and . The use of interaction terms Lt and Rt allowed for the gradients of change in log mortality rates over time to be different over the New Labour and GFC recession periods.

The above model specification was fit to ONS data for each year from 1990 to 2010 inclusive. Redefining , projected log mortality rates were calculated for years 2011 to 2015 inclusive by setting t to these year values and L to 1, i.e.

|  |  |
| --- | --- |
|  | (2) |

Predicted numbers of deaths at each age, for each sex, and in each year from 2011 to 2015 were therefore calculated by multiplying the relevant age-year-sex specific population counts by the requisite projected mortality rates, i.e.

|  |  |
| --- | --- |
| or equivalently | (3) |

Where is the projected mortality rate rather than log rate.

The age-sex specific differences in deaths are therefore , and the total difference in deaths by age A, shown in figures xxx, is .

### **Summary of regression coefficients for different ages in standard model**

The following figure shows the regression coefficients for each of the age-specific models produced. The bands show 1.96 standard deviations above and below the central parameter estimates indicated by the line. GFC: Global Financial Crisis; NL: New Labour.