

Mortality in England and Wales, 1979-1992
An Introduction to Multilevel Modelling using MLwiN

Alastair H Leyland and Alice McLeod

MRC Social and Public Health Sciences Unit
University of Glasgow
4 Lilybank Gardens
Glasgow G12 8RZ

MRC Social and Public Health Sciences Unit occasional paper no. 1, 2000.

The Social and Public Health Sciences Unit is a Research Unit supported by the Medical Research Council and the Chief Scientist Office of the Scottish Executive Health Department at the University of Glasgow. The production of these training materials has been supported by the Economic and Social Research Council as part of the Analysis of Large and Complex Datasets programme.

Mortality in England and Wales, 1979-1992

An Introduction to Multilevel Modelling using MLwiN

Alastair H Leyland and Alice McLeod

INTRODUCTION TO THE DATASET

The data are taken from the local mortality datapack and detail deaths from all causes in England and Wales in the period 1979 to 1992. The data are stored at the Data Archives at the University of Essex, maintained by the Economic and Social Research Council. The raw data comprise two files: one containing information on deaths over this time period and the other detailing the populations of the relevant areas (counties in England and Wales) in each year. For further information on this and other available datasets the user should visit the Data Archives website: <http://dawww.esx.ac.uk/>

RESEARCH QUESTIONS

In the following tutorial we will attempt to answer the following research questions:

1. What is happening to mortality rates over time?
2. How much variation in mortality rates is there between districts of England and Wales?
3. Is this variation just between districts, or are there also differences between the mortality rates of counties?
4. Does mortality vary according to the type of area?
5. What is happening to the variation in mortality rates over time?

INTRODUCTION TO MLwiN

Opening a worksheet

MLwiN files are known as worksheets which include all the data and model settings from the last saved version.

Go to the **File** menu
Select **Open worksheet**
Open the worksheet called **lmdp.ws**

The name of the current file appears in the bar at the top of the MLwiN window.

Names window

To view a summary of this worksheet

Go to the **Data manipulation** menu
Select **Names**

This brings up a list of all the variables stored in the worksheet together with some summary information. The worksheet contains 8 variables, each with 5639 data points and no missing values. Each data point corresponds to the annual number of deaths for all districts in England and Wales in the period 1979-92. COUNTY, DISTRICT and REGION are area identifiers; there are 403 county DISTRICTs (coded from 101 to 6820) which are nested within 54 COUNTYS (coded from 1 to 68), and these in turn lie within one of 10 REGIONS. The data cover 14 YEARS from 1979 to 1992. Note that there are only 5639 data points rather than the 5642 that might be expected (403 county districts with an observation for each of 14 years); three data points have been removed because extreme outlying values made them implausible. The next two columns show the number of DEATHS observed in each district at each time point – ranging from 16 to 12,775 – and the number that would be EXPECTED. The expected number of deaths has been calculated on the basis of the age and sex structure of that area's population in each year had the 1992 national age- and sex-specific mortality rates applied throughout. This worksheet has been constructed using the two raw data files contained in the local mortality datapack – the number of deaths and the populations. The OBSERVED and EXPECTED deaths are combined to form the standardised mortality ratio (SMR) for each year in each district. This is calculated as

$$smr = \frac{\text{observed deaths}}{\text{expected deaths}} \times 100$$

and reflects the excess deaths in an area, standardised for age and sex, over the national average mortality rate in 1992 (average = 100). The range from 75 to 179 implies a minimum mortality rate for one area in one year 25% below the 1992 average and a maximum 79% above the average. Finally, the variable FAMILY is a classification of districts into 6 groups devised by the Office for National Statistics:- 1 – Inner London, 2 – Rural areas, 3 – Prospering areas, 4 – Maturer areas, 5 – Urban centres, 6 – Mining and industrial areas. All of the remaining columns are empty; the default name for such columns is 'C' followed by the column number.

The screenshot shows a window titled "Names" with a search bar containing "1 county" and buttons for "Refresh", "Categories", and "Help". Below the search bar is a table with the following data:

	Name	n	missing	min	max
1	county	5639	0	1	68
2	district	5639	0	101	6820
3	region	5639	0	1	10
4	year	5639	0	79	92
5	deaths	5639	0	16	12775
6	expected	5639	0	11.47031	10134.06
7	smr	5639	0	74.71229	179.295
8	family	5639	0	1	6
9	c9	0	0	0	0
10	c10	0	0	0	0

Data window

The data may be viewed and edited in a spreadsheet format

Go to the **Data manipulation** menu

Select **View or edit data**

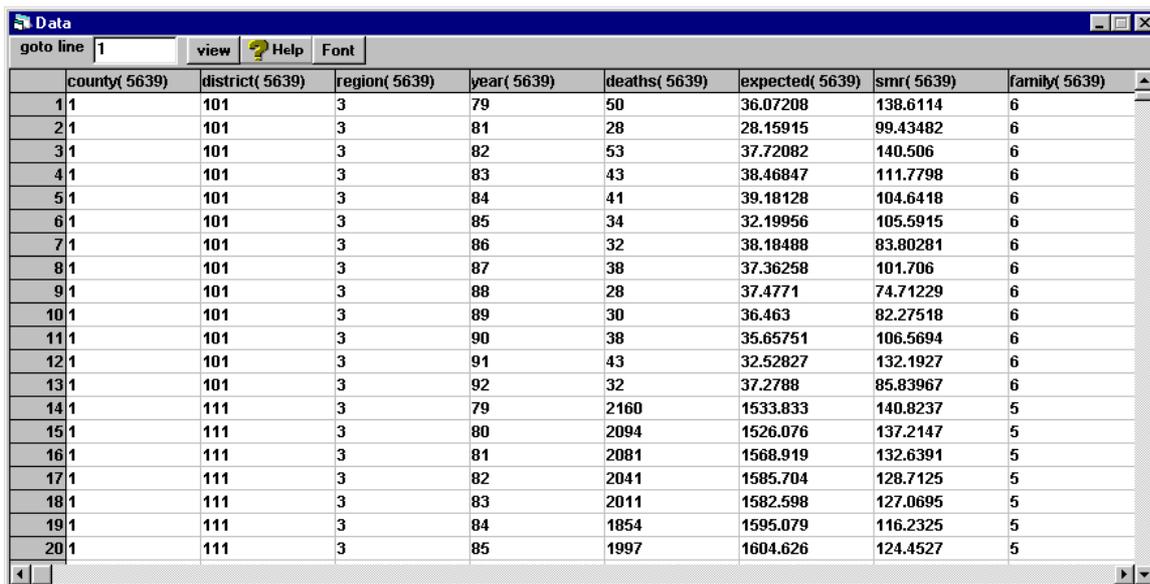
By default only the first 3 columns are shown. The **view** button within this window can be used to change or extend the selection of variables shown.

Click the **view** button

To select all variables, click on COUNTY, hold down the left mouse button and drag to FAMILY

Click the **OK** button

All windows can be re-sized by clicking on the borders and dragging; also the scroll bars at the bottom and on the right hand side can be used to view more of the selected data.



	county(5639)	district(5639)	region(5639)	year(5639)	deaths(5639)	expected(5639)	smr(5639)	family(5639)
1	1	101	3	79	50	36.07208	138.6114	6
2	1	101	3	81	28	28.15915	99.43482	6
3	1	101	3	82	53	37.72082	140.506	6
4	1	101	3	83	43	38.46847	111.7798	6
5	1	101	3	84	41	39.18128	104.6418	6
6	1	101	3	85	34	32.19956	105.5915	6
7	1	101	3	86	32	38.18488	83.80281	6
8	1	101	3	87	38	37.36258	101.706	6
9	1	101	3	88	28	37.4771	74.71229	6
10	1	101	3	89	30	36.463	82.27518	6
11	1	101	3	90	38	35.65751	106.5694	6
12	1	101	3	91	43	32.52827	132.1927	6
13	1	101	3	92	32	37.2788	85.83967	6
14	1	111	3	79	2160	1533.833	140.8237	5
15	1	111	3	80	2094	1526.076	137.2147	5
16	1	111	3	81	2081	1568.919	132.6391	5
17	1	111	3	82	2041	1585.704	128.7125	5
18	1	111	3	83	2011	1582.598	127.0695	5
19	1	111	3	84	1854	1595.079	116.2325	5
20	1	111	3	85	1997	1604.626	124.4527	5

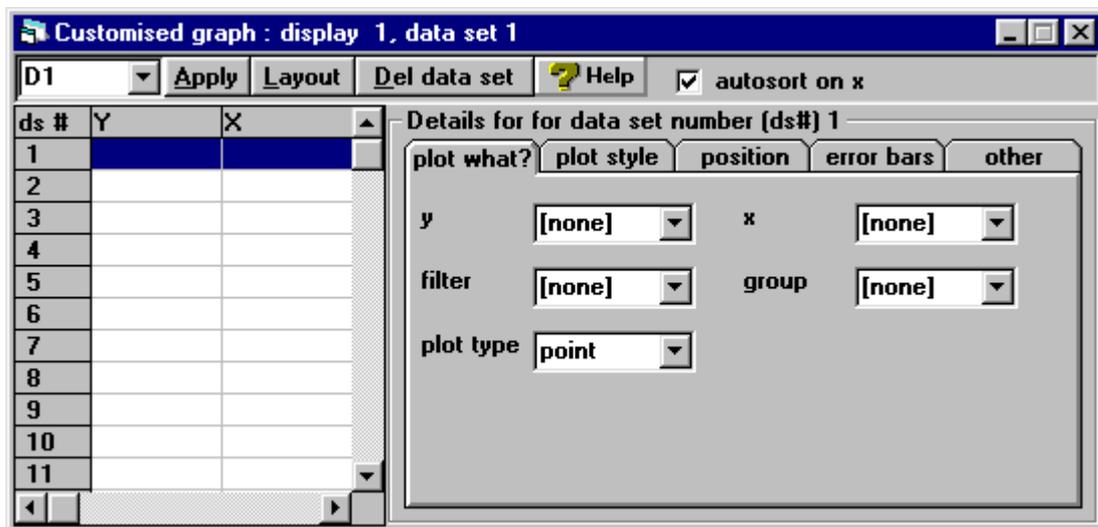
The first 13 observations are made on DISTRICT 101, COUNTY 1, REGION 3. The 13 observations on this DISTRICT can be seen to correspond to 13 YEARS of data; there is no observation for 1980. The estimated SMR in this district varies from 75 in 1988 to 141 in 1982. The district classification was group 1 – Inner London.

Graph window

Before starting to model we may wish to examine the data in a graph.

Go to the **Graphs** menu
Select **Customised Graphs**

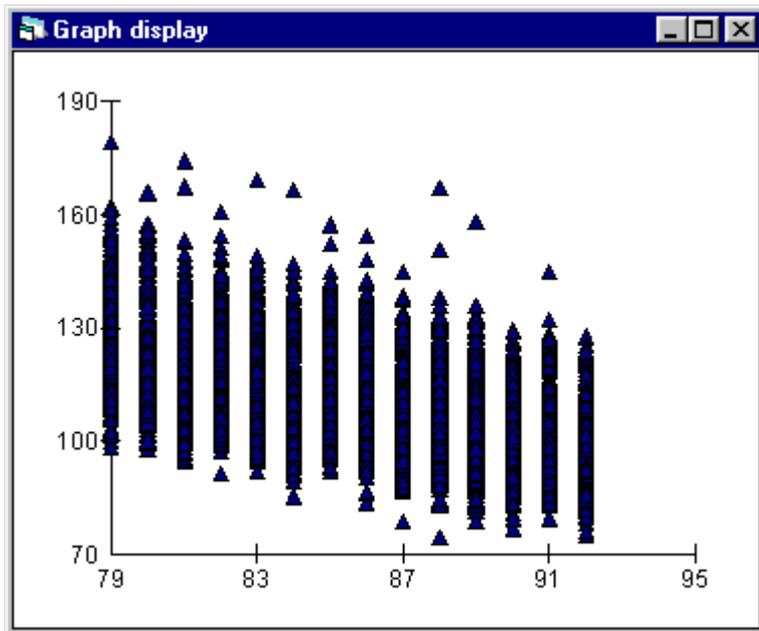
The graphical output in MLwiN is separated into three components. A *display* is what can be displayed on the computer screen at any one time and up to ten different displays may be specified. The pull-down menu at the top-left hand corner of the customised graph window corresponds to the display function – this currently shows **D1** denoting display 1. Each display can contain a number of *graphs*. A graph is a frame with x and y axes showing lines, points or bars, and each display can show an array of up to 5x5 graphs. The **Layout** button at the top of the customised graph window is used to specify the layout of the display. Finally, each graph can plot one or more *datasets*, each one consisting of a set of x and y coordinates held in the worksheet columns. Different datasets may be specified by clicking on different rows under the **ds#** heading shown at the right hand side of the customised graph display.



To obtain a scatter plot of SMRs by year, ensure that the **plot what?** tab is selected and

Select the **y** variable to be SMR
Select the **x** variable to be YEAR

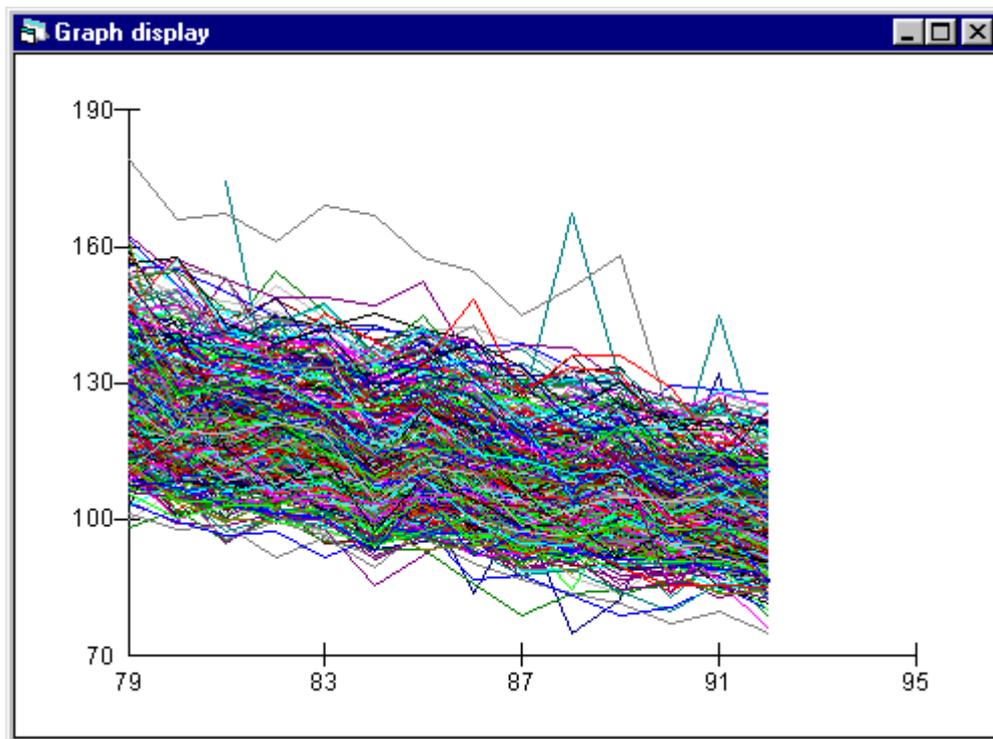
Click the **Apply** button



It is clear that there have been considerable reductions in SMR over these 14 years; nearly every district had an SMR greater than 100 in 1979. (The fact that standardisation was to 1992 means that the average SMR was 100 for that year.)

To change this graph to a line plot with a line for each district

In the **Customised Graph** window, select **group** to be DISTRICT
Change **plot type** to **line**
Select the **plot style** tab
Change **colour** to **16 rotate**
Click the **Apply** button



It is possible to identify points on the graph: point and click anywhere on the graph and the **Identify point** window will appear with details of the closest data point. Also included in the **Graph options** window are facilities for adding titles to the graph and axes, and for making other changes to the display including the scales.

Closing windows

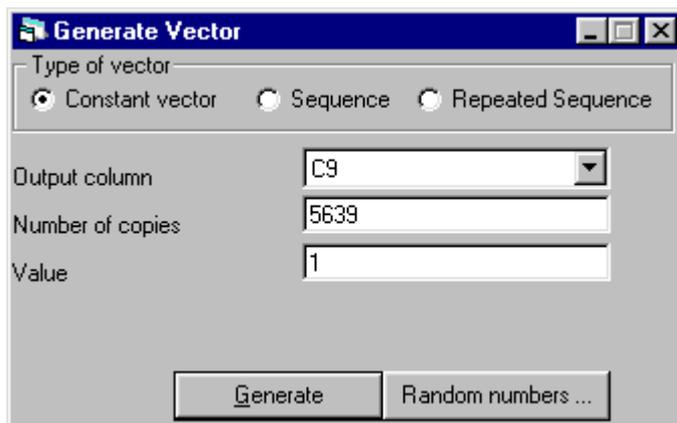
At any time you may wish to close or minimize windows to prevent your screen from becoming too cluttered. You may do this, as with any other Windows package, by clicking on the X or _ buttons respectively in the top right corner of each window. Alternatively you may go to the **Window** menu and select **close all windows**.

MODEL SPECIFICATION

Creating new variables

A number of functions are available in MLwiN that allow the creation of new variables or amendments to existing variables. In order to include a constant or intercept term in a model, we need to create a column of 1's that spans the entire data set.

Go to **Data manipulation** menu
Select **Generate vector**
Select **Type of vector** to be **Constant vector**
Select C9 to be the **Output column**
Enter 5639 (the number of data points) beside **Number of copies**
Enter 1 beside **Value**
Click the **Generate** button



Returning to the **Names** window, column C9 now contains 5639 data points each with the value of 1.
To name this new variable

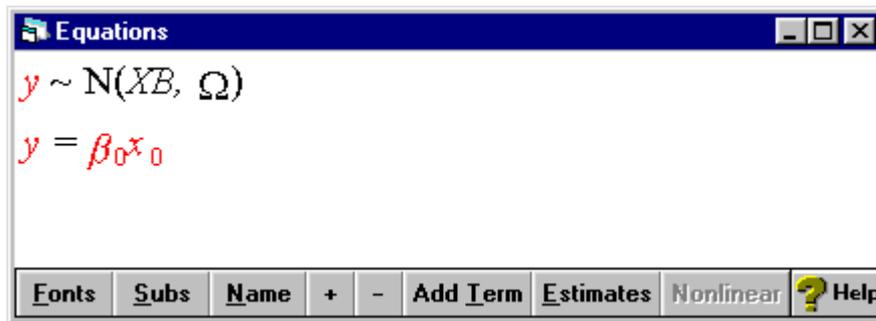
Click on C9 in the **Names** window
Highlight C9 in the box at the top of the window
Type CONS and press <return>

Equations window

Specifying models in MLwiN is done mainly via the **Equations** window.

Go to **Models** menu

Select **Equations**



The items in red must be defined before a model can be fitted to the data.

Click on either of the **y** terms

Select SMR as the dependent variable

The structure of the hierarchical model is also specified at this stage; first by stating the number of levels the model will have and then by specifying what the levels of the hierarchy are using the appropriate identifier variables. For a 2-level model of YEARS nested within DISTRICTs

Select 2 – ij for **N levels**

Select DISTRICT for **level 2(j)**

Select YEAR for **level 1(i)**

Click on **Done**

The red response variable **y** has changed colour indicating that this term has been defined; moreover, the addition of the subscripts **i** and **j** indicates that this is a 2-level model. The ordering tells us that YEAR **i** is nested within DISTRICT **j**. In a similar manner we can define CONS to be an independent variable.

Click on the $\beta_0 x_0$ term

Select CONS from the drop-down list

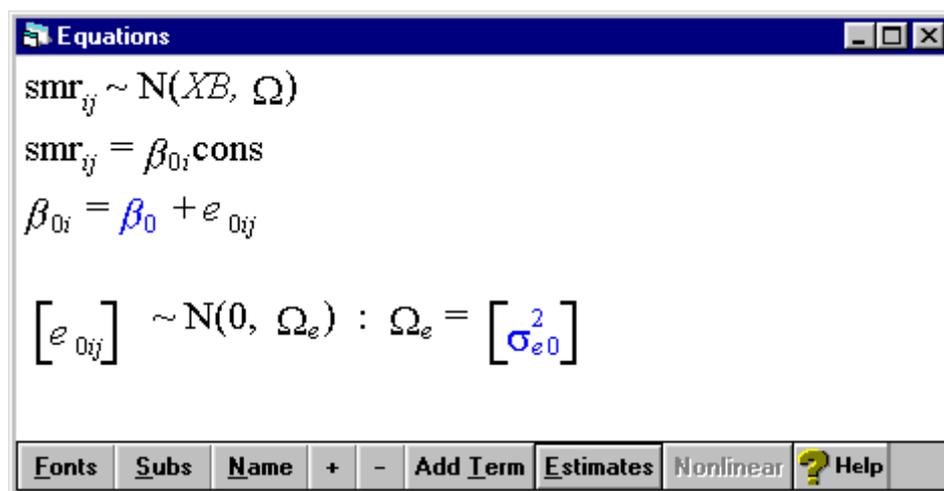
The check boxes indicate what part of the model each variable is in; by default, CONS has been added to the fixed part of the model and its coefficient will provide an estimate of the intercept. The other options in this window relate to the random part of the model. We will start by fitting a model with the intercept random at the level of YEAR (level 1).

Click on the check box by **i(YEAR)**
 Click on **Done**

Note that the term β_0 changes to β_{0i} , denoting the fact that it is random at level 1.

At any time we can toggle between a purely algebraic representation and one that gives a little more detail of the models that are being fitted.

In the **Equations** window, click on the **Name** button
 Click the **Estimates** button



Note that y and x have been replaced by the names of the specified dependent and independent variables, SMR and CONS. Clicking the **Estimates** button expands the model to include the distributional assumptions that are applied to the random terms, in this case to the error term e_{0ij} . Two terms in the model are blue: the grand intercept β_0 and the level 1 variance σ_{e0}^2 . This indicates that these terms are to be estimated; when the model converges, the blue will change to green. Clicking on the **Estimates** button again will replace these two terms with their current estimates (both the default value of 1.00 because no model has yet been estimated).

In addition to the mean we will add year as an independent variable in the fixed part of the model in order to answer our first research question.

Click on the **Add Term** button

Click on the red x_1 which appears and select YEAR from the drop-down list

Click **Done**

The third term to be estimated, β_1 , is the coefficient associated with year and this will estimate the trend in SMRs during the study period.

Sorting the data

Before fitting any model, the data need to be sorted within their hierarchy (in this example by YEAR within DISTRICTs).

Go to **Data manipulation** menu

Select **Sort**

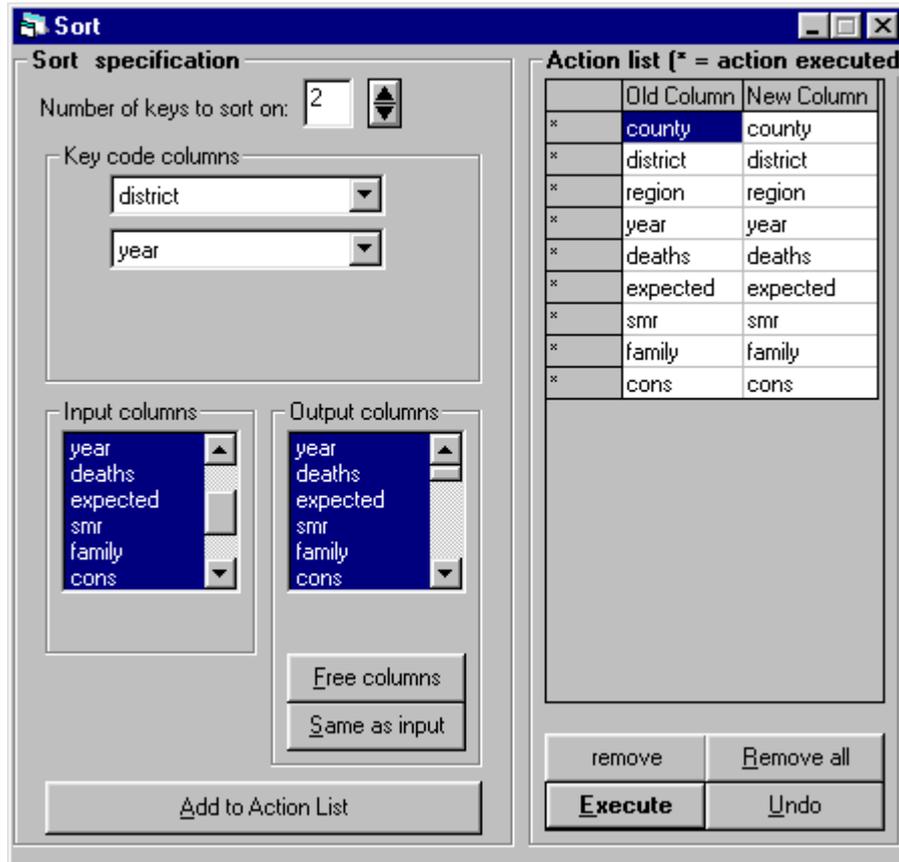
Increase the **Number of keys to sort on** to 2

Select DISTRICT as the first **Key code column** and YEAR as the second

Select all named variables under the heading **Input columns**

Press **Same as input** button to overwrite current columns with sorted data

Press **Add to action list** and then **Execute**



Fitting the model

The model is now ready to be estimated.

Click the **Start** button on the tool bar at the top of the MLwiN screen

After one iteration the model converges and the blue estimates in the equation window turn green.

$smr_{ij} \sim N(XB, \Omega)$
 $smr_{ij} = \beta_{0i} \text{cons} + -1.990(0.039) \text{year}_{ij}$
 $\beta_{0i} = 282.955(3.319) + e_{0ij}$
 $[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [137.284(2.585)]$
 $-2 * \log \text{likelihood(IGLS)} = 43758.250(5639 \text{ of } 5639 \text{ cases in use})$

Fonts Subs Name + - Add Term Estimates Nonlinear ? Help Clear

The estimated standard errors of each of the parameter estimates are shown in brackets. Our intercept is about 283, and the SMR has been decreasing at 1.990 per year. This decrease is highly significant in comparison with its standard error. The variance of all of the observations around this fitted trend is 137. The current model has a single term to describe the variation around the mean and is therefore just an ordinary least squares (OLS) regression model, but it is a starting point for our multilevel analysis. The value $-2 * \log(\text{likelihood})$ is provided as an aid to model selection.

Before continuing, consider the interpretation of the intercept term. This is the predicted value of the SMR in all districts when the variable YEAR takes the value 0 – in other words, in 1900. Since the data do not cover this period it is not sensible to make any inference about the SMR at this time, and we can change the origin to something more meaningful. Explanatory variables are frequently centred around an average value; in this case, however, we will set the origin at the first year for which we have data (1979).

Go to **Data manipulation** menu

Select **Calculate**

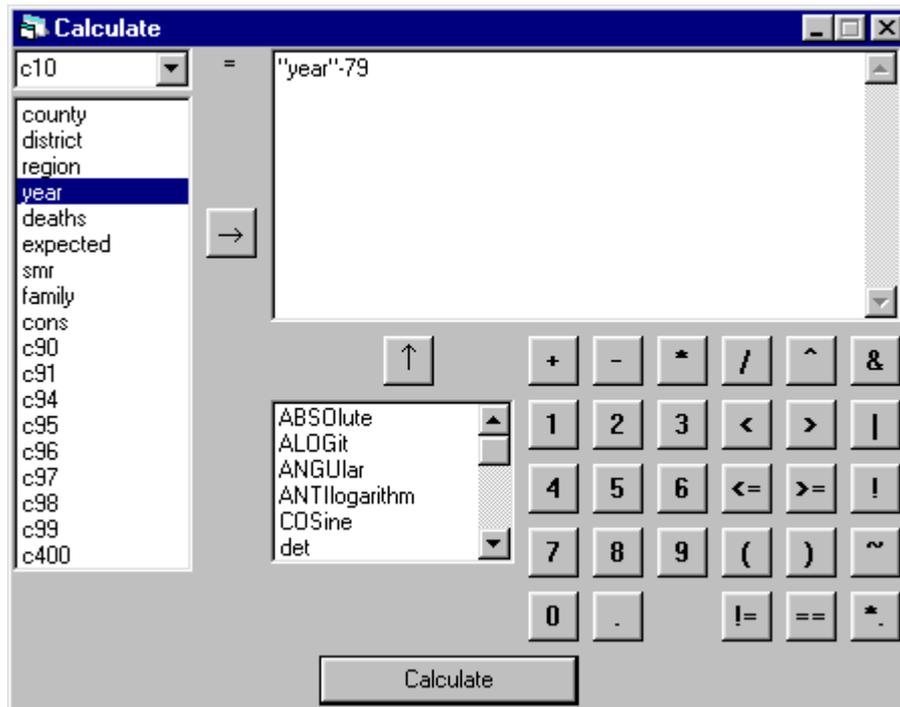
Select the empty column C10 from the list of variables and press the right arrow button

Click on the = button on the window's keypad

Select YEAR from the list of variables and press the right arrow button

Use the window's keypad to enter **-79**

Press **Calculate**



This has created a new year variable with an origin of 1979; name this new variable YEAR79 using the **Names** window.

To change the trend variable in the current model return to the **Equations** window

Click on **year_{ij}**
Select YEAR79 from the drop-down list
Click **Done**

The parameter estimates turn blue indicating that the model has changed and that it must be re-estimated. Rather than click on the **Start** button again, click on the **More** button to continue estimation from the current values.

Equations

$$\text{smr}_{ij} \sim N(XB, \Omega)$$

$$\text{smr}_{ij} = \beta_{0i} \text{cons} + -1.982(0.039) \text{year}79_{ij}$$

$$\beta_{0i} = 125.677(0.296) + e_{0ij}$$

$$[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [137.283(2.585)]$$

-2*loglikelihood(IGLS) = 43758.210(5639 of 5639 cases in use)

Fonts Subs Name + - Add Term Estimates Nonlinear ? Help Clear

The estimated slope has changed very slightly, but the big change is in the intercept. In 1979 the average SMR was therefore about 126.

VARIANCE COMPONENTS

All of the variance in the current model is at the lowest level of observation; this is just an ordinary least squares (OLS) regression equation. This model may be expanded by partitioning the variance into that which is attributable to random variation between DISTRICTs and that which arises due to fluctuations between YEARS within DISTRICTs.

A 2-level variance components model

In the **Equations** window

Click on β_{0i}
Check the box by **j(DISTRICT)**
Click **Done**

The intercept term now has an additional subscript, indicating that it varies across DISTRICTs as well as across YEARS. The intercept is partitioned into three parts: the overall fixed part intercept for 1979, the error term e_{0ij} and a term u_{0j} which is specific to DISTRICT j . The u_{0j} are random effects at level-2 and are assumed to arise from a normal distribution. The intercept for the j^{th} district in 1979 will be given by $\beta_0 + u_{0j}$. This model may be fitted by clicking **More**.

Equations

$$\text{smr}_{ij} \sim N(XB, \Omega)$$

$$\text{smr}_{ij} = \beta_{0ij}\text{cons} + -1.985(0.016)\text{year}79_{ij}$$

$$\beta_{0ij} = 125.698(0.544) + u_{0j} + e_{0ij}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 112.897(8.062) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 24.494(0.481) \end{bmatrix}$$

$-2*\text{loglikelihood(IGLS)} = 35723.940(5639 \text{ of } 5639 \text{ cases in use})$

Fonts Subs Name + - Add Term Estimates Nonlinear ? Help Clear

There is little change in the estimates of the intercept and slope in the fixed part of the model. However, most of the variation is now at level 2 – between DISTRICTs rather than YEARS. The total variation ($\sigma_{u0}^2 + \sigma_{e0}^2$) is 137.391, very close to the estimate of σ_{e0}^2 obtained in the single level model. The proportion of the total variance which arises due to differences between DISTRICTs is approximately 113/137 or 82.2%. This figure is known as the *intra-unit* or *intra-class correlation*, and indicates that the correlation between two observations made in different YEARS on the same DISTRICT is 0.822. The level 1 variance may be interpreted as the variation between years within districts. So, in answer to the second research question, it would appear that the majority of the variation in mortality is due to between-district differences rather than year-on-year fluctuations. Note that the addition of a single variance term has produced a substantial reduction in the value of $-2*\log(\text{likelihood})$.

Adding a further level

We can add COUNTY as a third level to the model and examine the relative importance of these large areas compared to the smaller DISTRICTs. First resort the data according to this new hierarchy.

Go to **Data manipulation** menu

Select **Sort**

Increase the **Number of keys to sort on** to 3

Select COUNTY as the first **Key code column**, DISTRICT as the second and YEAR as the third

Select all named variables under the heading **Input columns**

Press **Same as input** button to overwrite current columns with sorted data

Press **Add to action list** and then **Execute**

Now return to the **Equations** window

Click on y_{ij} or smr_{ij}

Change **N levels** to **3 – ijk**

Select COUNTY from the drop-down list by **level 3(k)**

Click **Done**

Click on β_{0ij}

Check the box by **k(COUNTY)**

Click **Done**

Click **More** to estimate the new model

The intercept term now has an additional subscript to indicate that it varies across COUNTY as well as across DISTRICTs and YEARS. The terms v_{0k} are level-3 random effects and are again assumed to arise from a normal distribution.

Equations

$$\text{smr}_{ijk} \sim N(XB, \Omega)$$

$$\text{smr}_{ijk} = \beta_{0ijk} \text{cons} + -1.984(0.016) \text{year79}_{ijk}$$

$$\beta_{0ijk} = 126.190(1.245) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 75.800(15.949) \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 42.851(3.378) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 24.494(0.479) \end{bmatrix}$$

$-2 * \log \text{likelihood(IGLS)} = 35479.460(5639 \text{ of } 5639 \text{ cases in use})$

Fonts Subs Name + - Add Term Estimates Nonlinear Help Clear

The fixed part is unchanged as is the level 1 (between years within DISTRICTs) variance. However, the higher level variance has been partitioned further into that attributable to COUNTYs and that due to differences between DISTRICTs within COUNTYs. About 53% of the total variation can be seen to be between COUNTYs with 30% between DISTRICTs and just 17% between YEARS.

INTERPRETING THE MODEL

Residuals

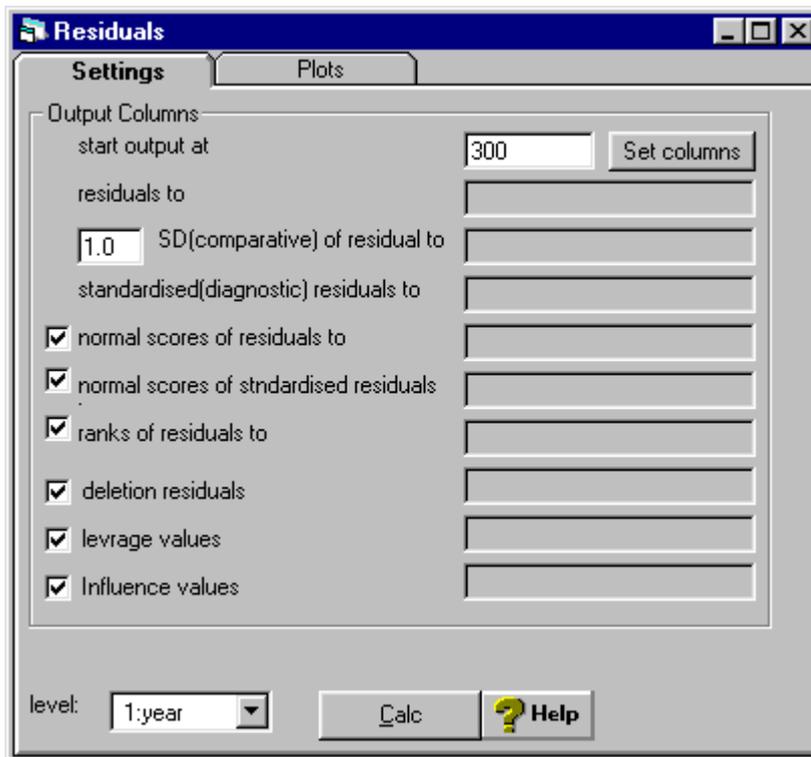
In an ordinary least squares (OLS) regression equation, the residual or error term is the difference between the observed and fitted values. In the above model, the equation may be written as

$$y_{ijk} = (\beta_0 x_0 + \beta_1 x_{1ijk}) + (v_{0k} x_0 + u_{0jk} x_0 + e_{0ijk} x_0)$$

The terms inside the first set of brackets comprise the fixed part of the model, i.e. the fitted values for all data points. The terms inside the second set of brackets comprise the random part of the model and describe the variation around the fitted values at each level of the hierarchy.. Thus, the difference between the observed and fitted values is comprised of residuals at three levels – the v_{0k} , u_{0jk} and e_{0ijk} in the regression equation. (Remember that x_0 is the variable CONS i.e. it takes the value 1 for every observation.) Each set of residuals is assumed to follow a Normal distribution and this assumption may be checked in the same way as residual diagnostics in OLS. First consider the residuals at level 1.

Go to the **Models** menu

Select **Residuals**



There are a variety of options which allow a range of standard diagnostic checks to be carried out – for example, to check the normality of the data or to look for outliers. By default all 9 functions are calculated and the results are stored in columns c300-c308. The box in the bottom left corner specifies the level at which the residuals are calculated; the default is level 1.

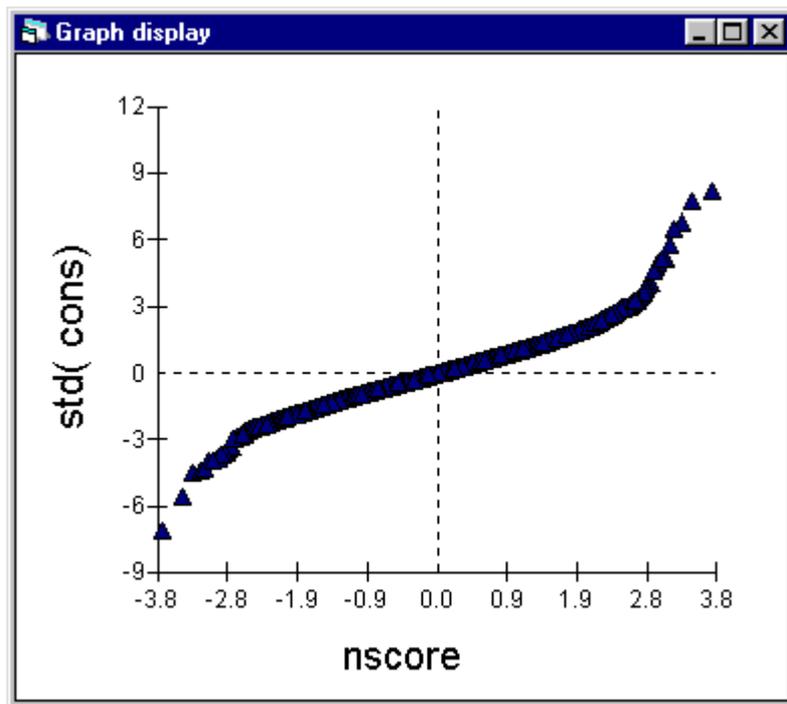
Click on the **Set columns** button

Click **Calc**

Select the **Plots** tab at the top of the **Residuals** window

Select the first option **standardised residual x normal scores**

Click **Apply**.



The points in the resulting graph should lie on a straight line; the fact that they don't suggests that there may be some departure from Normality. For the moment we will ignore this and look at the residuals at level 2 (DISTRICT).

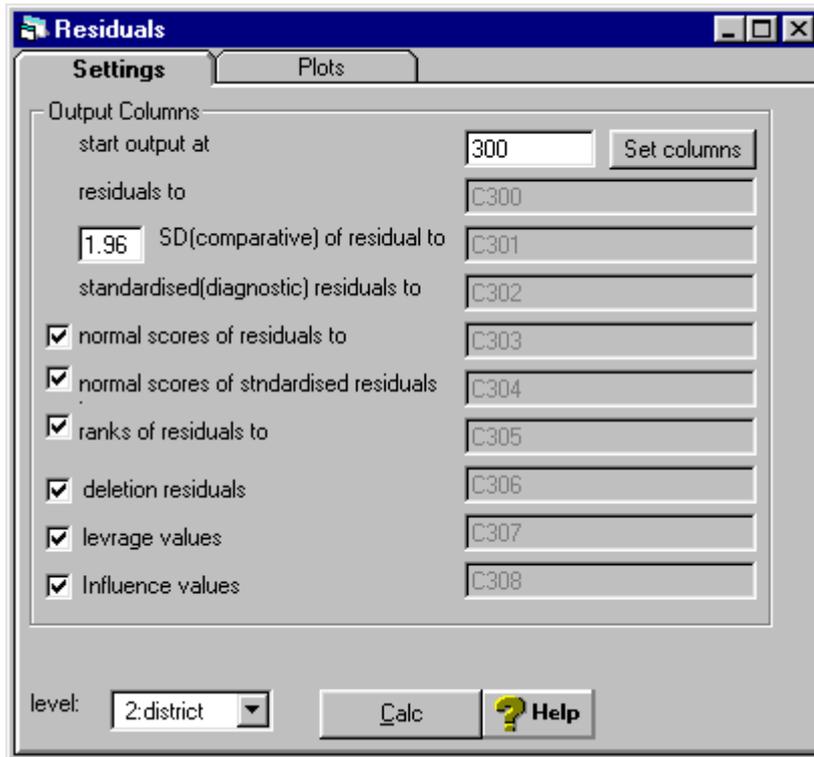
Click on the **Settings** tab in the **Residuals** window

Select **2: DISTRICT** to be the **level** at which the residuals are calculated

Change the multiplier in the box by **SD(comparative) of residual** to 1.96

Click on **Set columns**

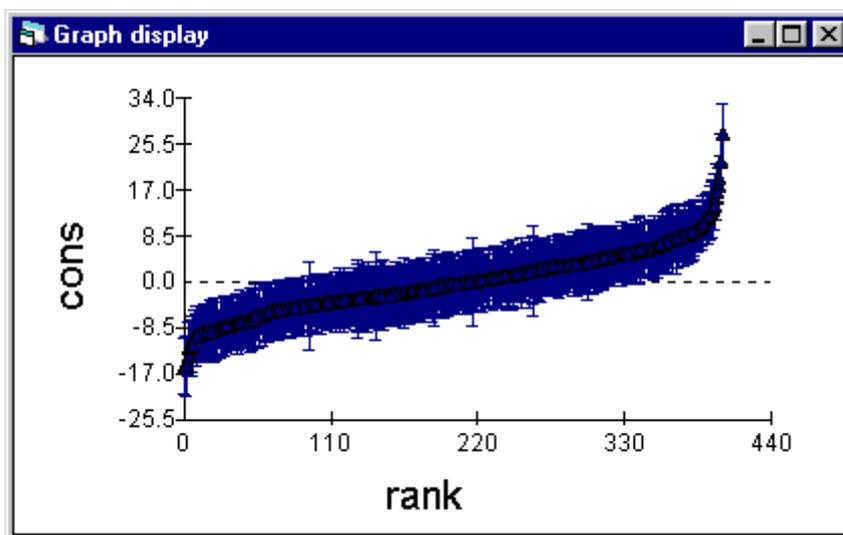
Click **Calc**



Select the **Plots** tab

Choose a plot of **residual ± 1.96 sd x rank**

Click on **Apply**

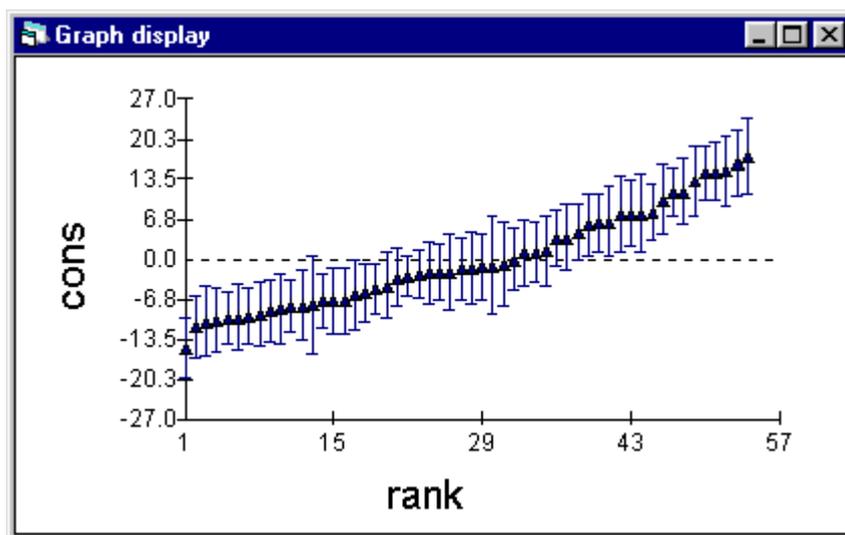


This plot shows the residuals or random effects for each DISTRICT, ordered from those with the smallest residuals on the left to those DISTRICTs with the largest residuals on the right. The range of values is from a reduction in the SMR of 16 points to an increase of 28 points. Since there is another level above DISTRICT, that of COUNTY, the residuals do not represent differences from the national

average but from the COUNTY average. The residuals are accompanied by error bars of half-width 1.96 S.D.; a DISTRICT whose error bar does not cross the horizontal line through zero has an SMR which is significantly different from the COUNTY average.

Finally, consider the residuals at level 3 (COUNTY).

Click on the **Settings** tab in the **Residuals** window
Select **3:COUNTY** to be the **level** at which the residuals are calculated
Ensure the multiplier in the box by **SD(comparative) of residual** is set to 1.96
Click on **Set columns**
Click **Calc**
Select the **Plots** tab
Choose a plot of **residual +/-1.96 sd x rank**
Click on **Apply**



The range of values of the COUNTY residuals is from a reduction in SMR of 15 points to an increase of 17 points. Although this is not as great as the range that is apparent among the DISTRICTs, bear in mind that there are considerably fewer COUNTYs than DISTRICTs (54 as opposed to 403). Thirty-three of the COUNTYs have residuals which are significantly different from zero. Note that not all DISTRICTs within these COUNTYs need have SMRs which are significantly different from 100; a REGION with a positive residual may contain DISTRICTs with negative residuals because the components of the composite random part – u_{0jk} and v_{0k} – are assumed to be independent.

Predictions window

A number of different predictions may be made from a multilevel model depending on whether one includes fixed effects only or a combination of fixed and random effects. For example, prediction lines for COUNTYs are derived from the fixed part of the model together with the residuals from the appropriate level – the v_{0k} in this case.

Go to the **Model** menu

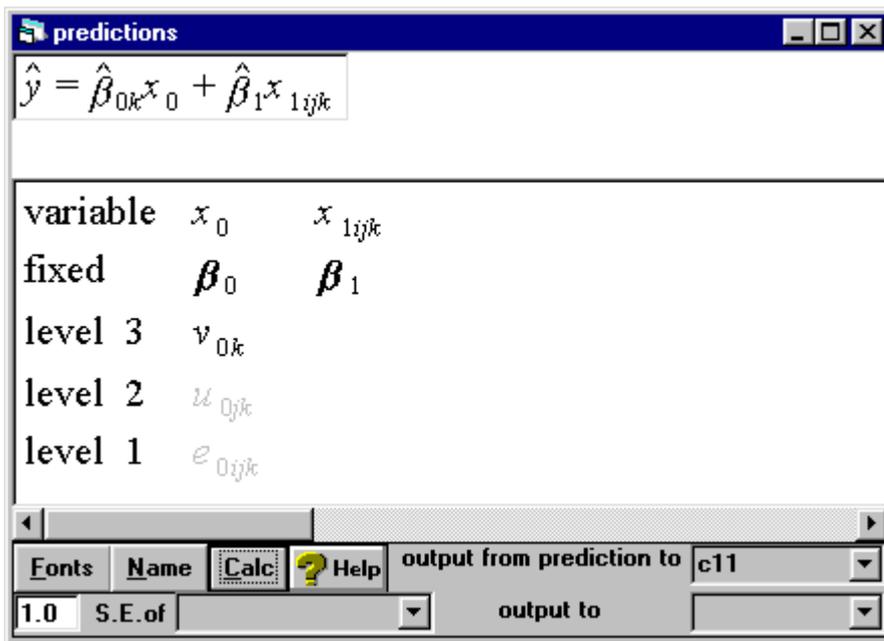
Select **Predictions**

The elements of the model are arranged in two columns, one for each explanatory variable. Initially, all the terms are in grey indicating that they have not been selected and are not included in the prediction equation at the top of the **Predictions** window. The prediction equation is built by selecting the appropriate terms; clicking on the variable name at the head of the column (x_0 or x_{1ijk}) selects all the terms in that column, whilst clicking on individual terms (such as β_0 or v_{0k}) toggles them in or out of the prediction equation.

Click on x_0 , x_{1ijk} , then u_{0jk} and e_{0ijk} to remove them from the prediction

In the drop-down list by **output from prediction** to select C11

Click on **Calc**



The results from this prediction are now in C11; using the **Names** window, name this variable PRED3 to indicate that it is a prediction including the level 3 (COUNTY) random effects. (Note: if the **Names** window was already open then column C11 may still appear to be empty; click on the **Refresh** button at the top of the **Names** window to overcome this problem.) Now plot the predicted values for each COUNTY against YEAR

Go to the **Graph** menu
 Select **Customised Graph**

Note that details of earlier graphs are still held. D1 contains plots of the crude data while D10 contains the plot of residuals carried out in the previous section. To create a new graph

Select **D2** from the list
 Select the y variable to be PRED3
 Select the x variable to be YEAR
 Select **group** to be COUNTY
 Select **plot type** to be **line**
 Click the **Apply** button

This produces a plot of 54 parallel lines, one for each COUNTY. We will superimpose on this graph the prediction of the fixed part of the model, the mean line given by

$$\hat{y}_{ijk} = \hat{\beta}_0 x_0 + \hat{\beta}_1 x_{1ijk}$$

Return to the **Predictions** window

Click on **v_{0k}** to remove it from the prediction equation

In the drop-down list by **output from prediction to** select C12

Click on **Calc**

In the **Names** window, change the name of C12 to PREDFP to indicate that it is a prediction from the fixed part only.

Return to the **Customised Graph** window

Ensure **D2** is selected

Under **ds #** (dataset number) click on number **2**

Select the **y** variable to be PREDFP

Select the **x** variable to be YEAR

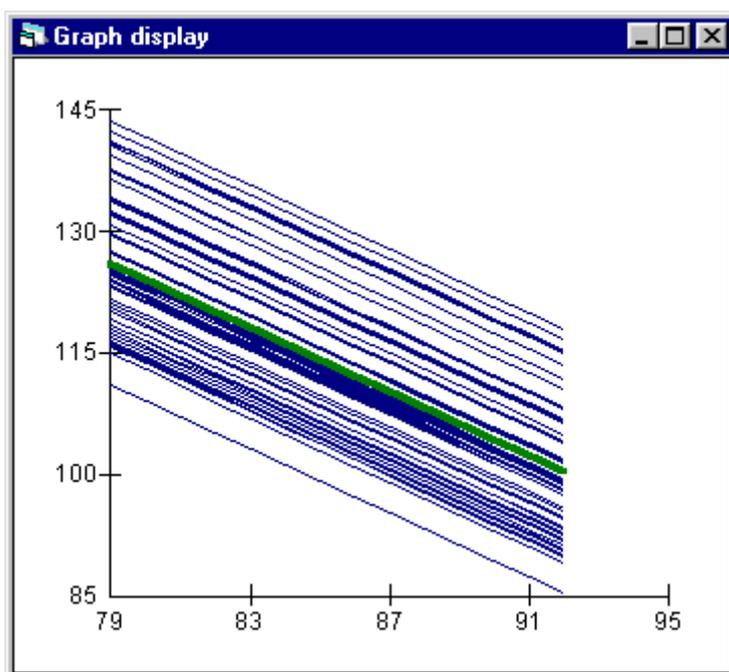
Select **plot type** to be **line**

Click the **plot style** tab

Change the **colour** to **2 green**

Change the **line thickness** to **3**

Click the **Apply** button



The national mean SMR is highlighted in green with the predicted mean for each COUNTY shown around it. The lines are all parallel since the effect of each COUNTY, v_{0k} , is assumed to be the same throughout the study period. This residual is the horizontal distance between the national mean and the COUNTY mean; a positive value of v_{0k} indicates the COUNTY mean SMR is greater than the national mean.

Now look at the predicted means for DISTRICTs within a specific COUNTY. First we need to generate the predicted values for each DISTRICT:

$$\hat{y}_{ijk} = \hat{\beta}_{0jk}x_0 + \hat{\beta}_1x_{1ijk}$$

Return to the **Predictions** window
Click on v_{0k} and u_{0jk} to add them to the prediction equation
In the drop-down list by **output from prediction to** select C13
Click on **Calc**

In the **Names** window change the name C13 to PRED2 to indicate that these predicted values include the level 2 (DISTRICT) random effects.

To illustrate the different prediction lines in a single chart, select a single COUNTY, e.g. COUNTY number 1. To create an indicator for COUNTY number 1:

Go to **Data manipulation** menu
Select **Calculate**
Select the empty column C14 from the list of variables and press the right arrow button
Click on the = button on the window's keypad
Select COUNTY from the list of variables and press the right arrow button
Use the window's keypad to enter ==1
Press **Calculate**

Note the logical command “==” (two equals signs) means “is equal to”. This will create a dummy variable with the value 1 if the data are from COUNTY number 1, 0 otherwise.

Go to the **Names** window and change the name of C14 to COUNTY1.

Return to the **Customised Graph** window

Ensure **D2** is selected

Highlight data set number **1** under **ds #**

Select the **filter** to be COUNTY1

Click on the **plot style** tab

Change the **line thickness** to **3**

Click the **Apply** button

The resulting graph now has just two lines – one for the national mean and one for the selected COUNTY. To plot the predicted lines for the DISTRICTS in COUNTY number 1

Return to the **Customised Graph** window

Select **ds # 2**

Select the **filter** to be COUNTY1

Click the **Apply** button

Under **ds #** click on number **3**

Select the **y** variable to be PRED2

Select the **x** variable to be YEAR

Select the **filter** to be COUNTY1

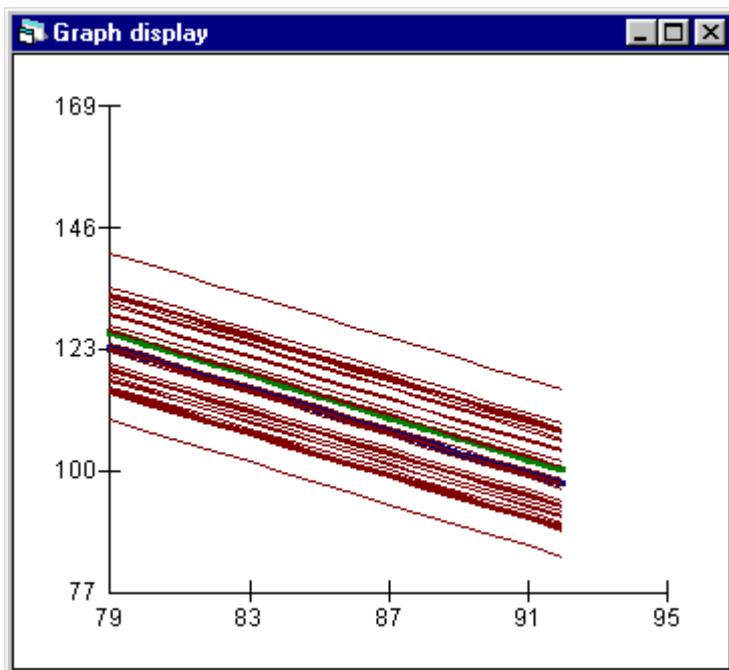
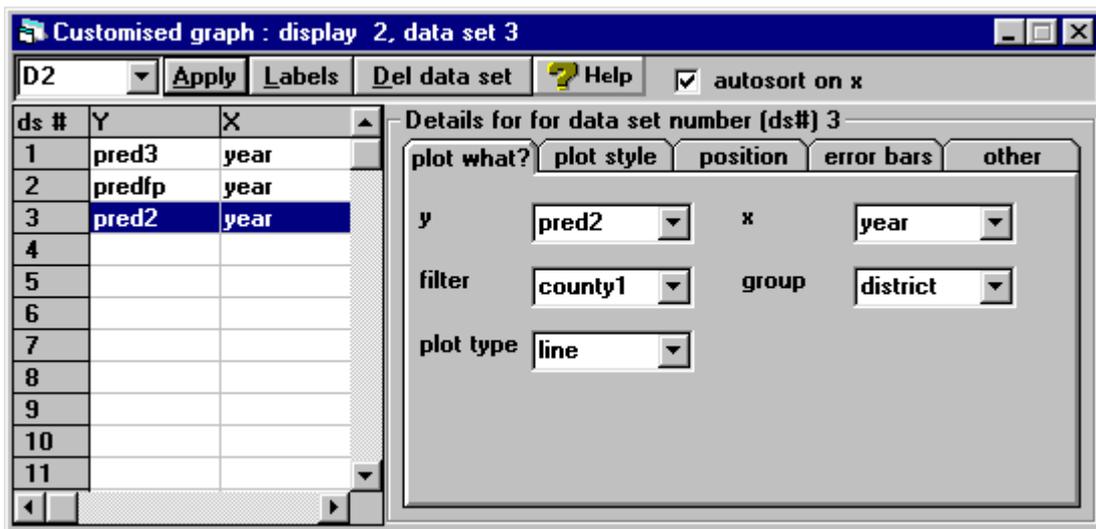
Select **group** to be DISTRICT

Select the **plot type** to be **line**

Click the **plot style** tab

Change the **colour** to **4 red**

Click the **Apply** button



In addition to the national mean (green) and COUNTY mean (blue) the graph now displays the DISTRICT means for the selected COUNTY. The vertical distance between the green and blue lines is the level 3 (COUNTY) residual v_{01} (the subscript k is replaced by the number of the COUNTY). The fact that the COUNTY mean is below the national mean indicates that this residual is negative. The vertical distance between each DISTRICT mean and the COUNTY mean is the level 2 (DISTRICT) residual u_{0j1} . The vertical distance between each DISTRICT mean and the national mean is then the composite residual $v_{01} + u_{0j1}$. You may note that, despite the average for this COUNTY being below the national average, some of the DISTRICT means still lie above the national average (the green line) because the composite residual $v_{01} + u_{0j1}$ is greater than zero.

MODEL BUILDING

Adding more fixed effects

The models fitted so far include only an intercept term (CONS) and a trend coefficient (YEAR) in the fixed part. Now consider the addition of further variables. Firstly, add a quadratic term in year since the assumption of a linear trend may be too simplistic.

Go to **Data manipulation** menu

Select **Calculate**

Select the empty column C15 from the list of variables and press the right arrow button

Click on the = button on the window's keypad

Select YEAR79 from the list of variables and press the right arrow button

Use the window's keypad to enter 2

Press **Calculate**

In the **Names** window change the name of C15 to YEAR79 2 .

Return to the **Equations window**

Click on **Add Term**

Click on x_2 and select YEAR79 2 , then click on **Done**

Click on the **More** button to re-estimate the model

Equations

$$\text{smr}_{ijk} \sim N(XB, \Omega)$$

$$\text{smr}_{ijk} = \beta_{0ijk} \text{cons} + -2.124(0.062) \text{year79}_{ijk} + 0.011(0.005) \text{year79}^2_{ijk}$$

$$\beta_{0ijk} = 126.470(1.251) + v_{0k} + u_{0ijk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 75.802(15.988) \end{bmatrix}$$

$$\begin{bmatrix} u_{0ijk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 42.858(3.376) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 24.468(0.479) \end{bmatrix}$$

$-2 * \text{loglikelihood(IGLS)} = 35473.980(5639 \text{ of } 5639 \text{ cases in use})$

Fonts Subs Name + - Add Term Estimates Nonlinear Help Clear

The reduction in (-2*loglikelihood) is 5.48 from 1 degree of freedom so this term has significantly improved the fit of the model. The addition of this term has, however, done nothing to reduce the variance at any of the three levels in the model.

The next covariate to consider adding to the fixed part of the model is the variable FAMILY, a classification of the DISTRICTs into different types. It is a categorical variable with 6 categories and in order to model the effect of this covariate on the average SMR, it is necessary to create six dummy variables, one for each DISTRICT classification. The easiest way of entering categorical variables such as this is to use the **Main Effects and Interactions** window.

Under the **Model** menu select **Main Effects and Interactions**

Under **Categorical variable**, click on [none] and select FAMILY from the pull-down list

Click **OK** on the dialog box to open the **Set category names** window

Click on **0** under **name** and type LONDON

Click on **1** under **name** and type RURAL

Click on **2** under **name** and type PROSPER

Click on **3** under **name** and type MATURE

Click on **4** under **name** and type URBAN

Click on **5** under **name** and type MINING

Click on **Apply** and then close the **Set category names** window by clicking on **Quit**
 Under **Categorical variable**, click on [none] and select FAMILY from the pull-down list
 Under **View**, click on **Main effects**
 Under **Main effects included**, click on each variable apart from LONDON
 Click **Build**
 Click on the **More** button to re-estimate the model

We have created six dummy variables named LONDON, RURAL etc., and each observation takes the value 1 if the DISTRICT is of that type, 0 otherwise. (The **Set category names** window may also be opened from the **Names** window by highlighting the relevant variable – in this case FAMILY – and clicking on the **Categories** button.) As with OLS regression, when a covariate has n categories, only $n-1$ dummies are fitted in the model, the remaining category being used as a baseline against which comparisons are drawn. In this example, category 0, Inner London, will be used as the baseline.

Equations

$$\text{smr}_{ijk} \sim N(XB, \Omega)$$

$$\text{smr}_{ijk} = \beta_{0ijk}\text{cons} + -2.123(0.062)\text{year79}_{ijk} + 0.011(0.005)\text{year79}^2_{ijk} +$$

$$-8.876(2.192)\text{rural}_{ijk} + -10.407(2.148)\text{prosper}_{ijk} +$$

$$-11.150(1.981)\text{mature}_{ijk} + -2.514(2.240)\text{urban}_{ijk} + 3.611(2.359)\text{mining}_{ijk}$$

$$\beta_{0ijk} = 132.518(2.236) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 35.987(7.931) \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 30.240(2.421) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 24.468(0.478) \end{bmatrix}$$

$-2*\text{loglikelihood(IGLS)} = 35320.150(5639 \text{ of } 5639 \text{ cases in use})$

Fonts Subs Name + - Add Ierm Estimates Nonlinear ? Help Clear

The intercept or CONS term has changed as this is now the estimated mean in 1979 for areas in Inner London. There has been a significant reduction in $-2*\text{loglikelihood}$ with the loss of just 5 degrees of freedom. The total variance has been reduced by 36.6% from 143 to 91; whilst the year-on-year (level 1) variation has changed little, the between DISTRICT (level 2) variance has been reduced by 29%

and the between COUNTY (level 3) variance by 53%. The addition of a level 2 variable has then had the greatest effect on the apparent variation between level 3 units, indicating that to a large extent there is homogeneity of the type of DISTRICT found within each COUNTY. (This is not surprising; as an example, consider the fact that all of the DISTRICTs classified as being Inner London must lie within the same COUNTY i.e. London.).

Intervals and tests window

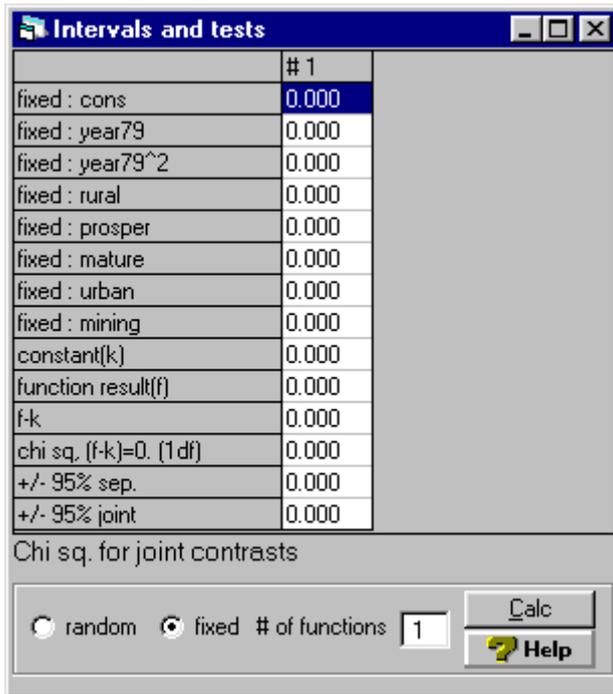
So far the change in likelihood has been used to assess improvement in the fit of the model to the data. It is also possible to carry out hypotheses tests for either fixed or random parameters using the **Intervals and tests** window. To illustrate how tests are formulated, consider the following two hypotheses. Firstly, if we are interested in testing whether SMRs in urban DISTRICTs are the same as those in Inner London then, since Inner London is the baseline category, this is equivalent to testing whether the coefficient for URBAN is significantly different from 0, i.e.

Hypothesis 1: $\beta_6=0$

We are not limited to single parameter tests but can also formulate significance tests involving a function of two or more parameters, as well as joint significant tests involving two or more functions of the model parameters. For example, consider a test of the hypothesis that SMRs in rural, prospering and mature DISTRICTs are the same, i.e.

Hypothesis 2: $\beta_3=\beta_4=\beta_5$ or
 $(\beta_3-\beta_4=0)$ and $(\beta_4-\beta_5=0)$ and $(\beta_3-\beta_5=0)$

Go to **Model** menu
Select **Intervals and tests**
Select **fixed** at the bottom of the window



The **# of functions** relates to the number of functions or contrasts of the parameter estimates being tested under a single hypothesis; for hypothesis 1 only one function is necessary while three functions are required for hypothesis 2. The boxes beside each fixed parameter are used to enter the function of the parameters to be tested, while the constant (k) contains the value to which the function is compared which, in both of the following cases, is the default value zero. So for hypothesis 1:

Select the box beside **fixed : urban**

Type 1

Press **Calc**

Note that the function **f** is a single multiple of the URBAN parameter and so equals β_6 , and because $k=0$, $(f-k)$ also equals the parameter β_6 . The test statistic, based on Wald's Test, appears in the bottom half of the window, **joint chi sq test(1df)=1.260**, and this may be compared to a chi-squared distribution to either accept or reject the hypothesis that $\beta_6=0$. We can obtain the p-value for the chi-squared or other distributions using the **Tail Areas** window:

Go to the **Basic statistics** menu

Select **Tail areas**

Select **Chi squared** under **Operation**

Enter the test statistic value **1.260** in the box beside **Value**

Enter the value **1** in the box beside **Degrees of freedom**

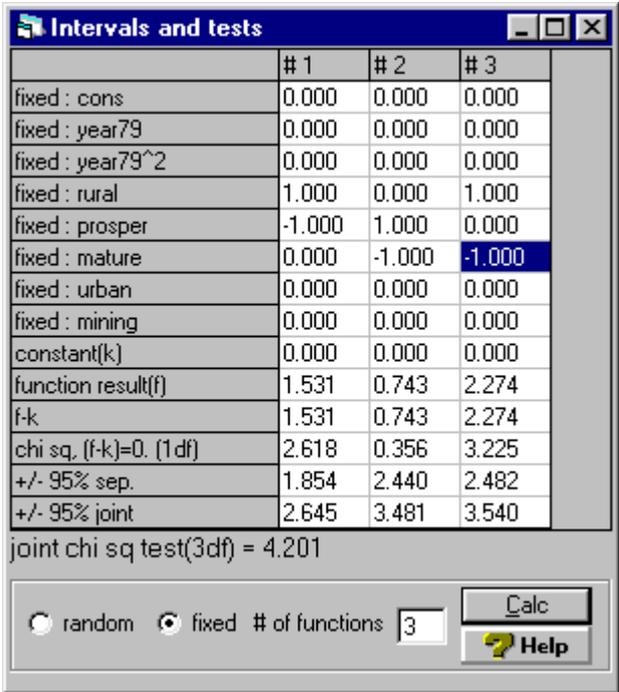
Press **Calculate**

The probability value of 0.26165 appears in the output window which should open automatically. In this case, $1.260 < \chi^2(1, 0.95) = 3.84$, so we do not reject the hypothesis that the mean SMR is the same in Inner London and urban DISTRICTs.

Now to formulate a test for *Hypothesis 2* (if the **Intervals and tests** window is still open, close it down and open it again to erase details of the previous test).

Ensure **fixed** is selected at the bottom of the **Intervals and tests** window
Change the **#of functions** to **3**
In the first column, enter a **1** beside **fixed:rural** and a **-1** beside **fixed:prosper**
In the second column, enter a **1** beside **fixed:prosper** and a **-1** beside **fixed:mature**
In the third column, enter a **1** beside **fixed:rural** and a **-1** beside **fixed:mature**
Press **Calc**

Each column specifies a function of the parameters which is compared to **constant (k)** equal to zero; for example, in column 1, the function is $(1 \times \beta_4) - (1 \times \beta_5) = 0$ (i.e. $\beta_4 = \beta_5$).



We return to the **Tail areas** window

Go to the **Basic statistics** menu

Select **Tail areas**

Select **Chi squared** under **Operation**

Enter the test statistic value **4.201** in the box beside **Value**

Enter the value **3** in the box beside **Degrees of freedom**

Press **Calculate**

This time we are jointly testing 3 functions and therefore require 3 degrees of freedom. The resulting p-value of 0.24056 indicates that we cannot reject the hypothesis that the mean SMRs of RURAL, PROSPER and MATURE are the same.

In practice at this stage we might want to collapse the variable FAMILY into just 3 categories: a baseline category comprising Inner London and Urban areas, achieved by deleting the variable URBAN from the current model, and a combination of Rural, Prospering and Maturer areas, which would involve creating a new variable using the **Calculate** window and replacing the variables RURAL, PROSPER and MATURE in the model with this new variable. However, we will continue for the moment with all six categories.

RANDOM COEFFICIENTS

We now consider another important class of multilevel model; random coefficients. In variance components models only the intercept is considered random; however, in the following model we will also allow the slope to vary across higher levels.

Random slopes

The following section considers the possibility that the rate at which the SMRs have been decreasing may vary from one COUNTY to another. The models fitted so far have contained random intercepts for both COUNTY and DISTRICT; however, the following model will also consider random slopes across the level 3 units (COUNTYs).

Return to the **Equations** window
 Click on **year79_{ijk}** and check the box by **k(COUNTY)**
 Then click **Done**

Equations

$$\text{smr}_{ijk} \sim N(XB, \Omega)$$

$$\text{smr}_{ijk} = \beta_{0ijk}\text{cons} + \beta_{1k}\text{year79}_{ijk} + 0.011(0.005)\text{year79}^2_{ijk} +$$

$$-8.876(2.192)\text{rural}_{ijk} + -10.407(2.148)\text{prosper}_{ijk} +$$

$$-11.150(1.981)\text{mature}_{ijk} + -2.514(2.240)\text{urban}_{ijk} + 3.611(2.359)\text{mining}_{ijk}$$

$$\beta_{0ijk} = 132.518(2.236) + v_{0k} + u_{0ijk} + e_{0ijk}$$

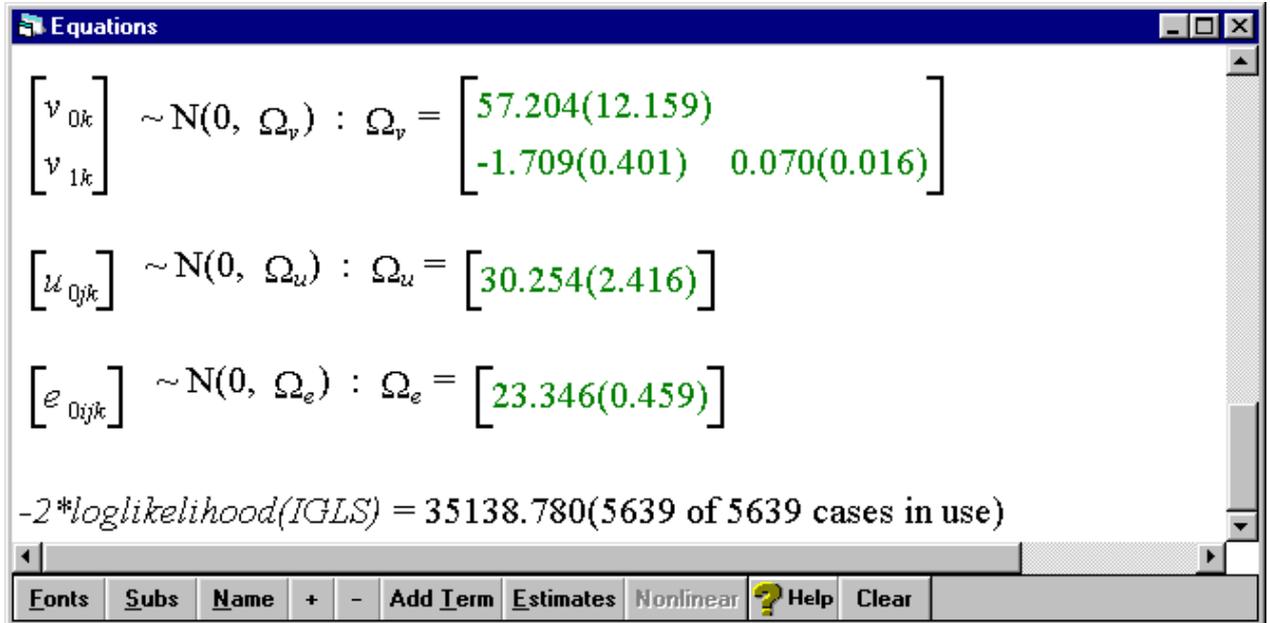
$$\beta_{1k} = -2.123(0.062) + v_{1k}$$

$$\begin{bmatrix} v_{0k} \\ v_{1k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 35.987(7.931) & \\ 0.000(0.000) & 0.000(0.000) \end{bmatrix}$$

Fonts Subs Name + - Add Term Estimates Nonlinear ? Help Clear

The coefficient of **year79_{ijk}** has changed from β_1 to β_{1k} indicating that this parameter now varies randomly across COUNTYs. The estimate of β_{1k} is now given as a mean β_1 , common to all

COUNTYs, plus a level 3 residual v_{1k} , unique to the k^{th} COUNTY. The level 3 residuals v_{0k} and v_{1k} now have a joint multivariate Normal distribution with variances $\sigma_{v_0}^2$ and $\sigma_{v_1}^2$ respectively and covariance $\sigma_{v_{10}}$. Click on **More** to estimate this model.



There is little change in the fixed part of the model, nor in the level 1 or level 2 variances. There has, however, been a large reduction in the value of $-2*\log(\text{likelihood})$. Therefore, the addition of random slopes has improved the overall fit of the model. The three random terms at level 3 now refer to the variance of the intercept (CONS) for COUNTYs – $\sigma_{v_0}^2$, the variance of the slope (YEAR79) for COUNTYs – $\sigma_{v_1}^2$, and the covariance between the two, $\sigma_{v_{10}}$. Whilst the two additional random terms appear large compared to their standard error, it is possible to test this formally using the **Intervals and tests** window.

Go to **Model** menu
 Select **Intervals and tests**
 Select **random** at the bottom of the window
 In the box beside **# of functions** type 2

There are two functions to test; our hypothesis is

Hypothesis 3: $\sigma_{v_1}^2 = \sigma_{v_{10}} = 0$

or

$\sigma_{v_1}^2 = 0$ and $\sigma_{v_{10}} = 0$

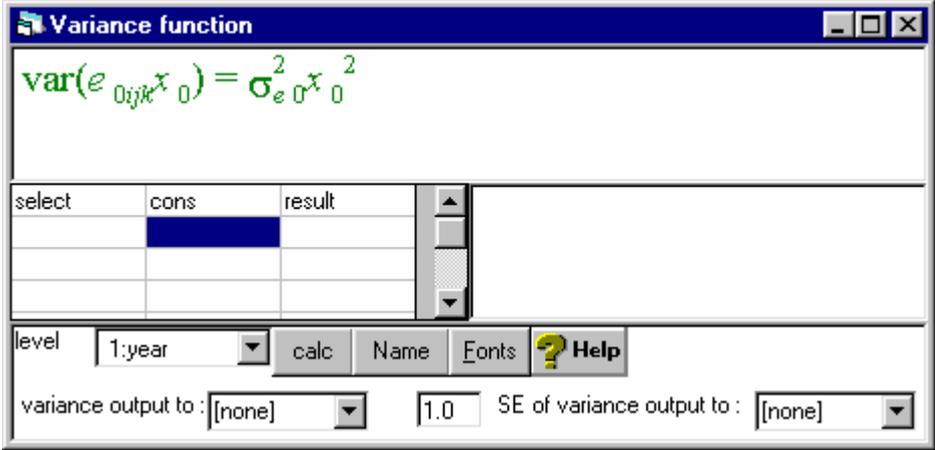
In the first column, enter a **1** beside **county:year79/cons**
In the second column, enter a **1** beside **county:year79/year79**
Press **Calc**

The value of 19.332 is extremely significant when compared with a chi-squared distribution with 2 degrees of freedom; we therefore reject the hypothesis that the two random terms are not significantly different from 0.

The level 3 variance is now more complex and more difficult to interpret; however, the **Variance function** window can be used as an aid.

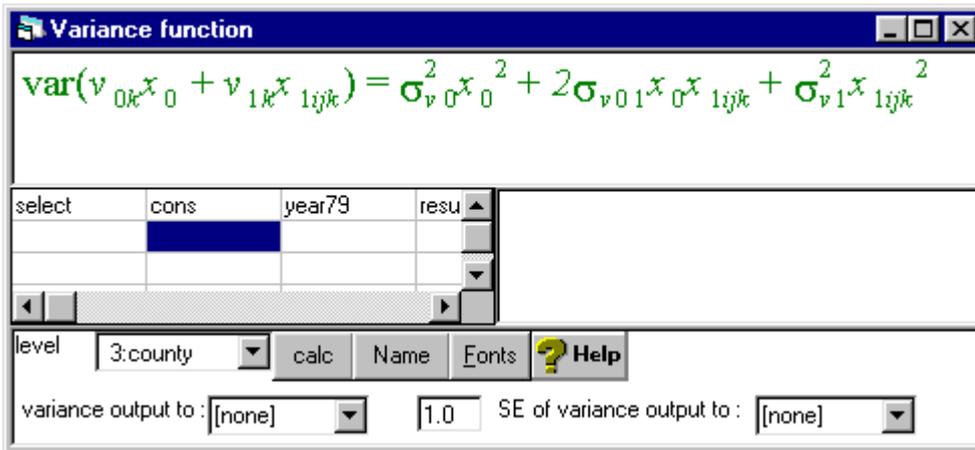
Variance function window

Go to **Model** menu
Select **Variance function**



The purpose of this window is to display and calculate the variance function at each level of the current model. The variance function for level 1 is currently shown; this only involves one term because the current model assumes that the level 1 variance is constant for all observations. To view the level 3 variance function:

In the drop-down list by **level** in the bottom left-hand corner, select **3:COUNTY**



The current model has two terms random at level 3, the intercept and the slope, so the level 3 variance is a function of two random variables. The function shown is the variance of the sum of the two random terms $v_{0k}x_0$ and $v_{1k}x_{1ijk}$. Since x_0 is just the CONSTANT term, taking the value 1, the level 3 variance is a quadratic in x_{1ijk} (YEAR79). We can use the **Variance function** window to calculate this function and use the **Graph** window to plot it.

Note that the columns in the table in the **Variance function** window named **select**, **cons**, **year79** and **result** allow us to estimate the variance function at specific values of YEAR79. However, rather than enter the values from 0 to 13 it is simpler to estimate the function for all data points.

In the drop-down menu by **variance output to**, select C21

Click calc

In the **Names window** name C23 VARF3. To plot the level-3 variance across the observed values of YEAR79:

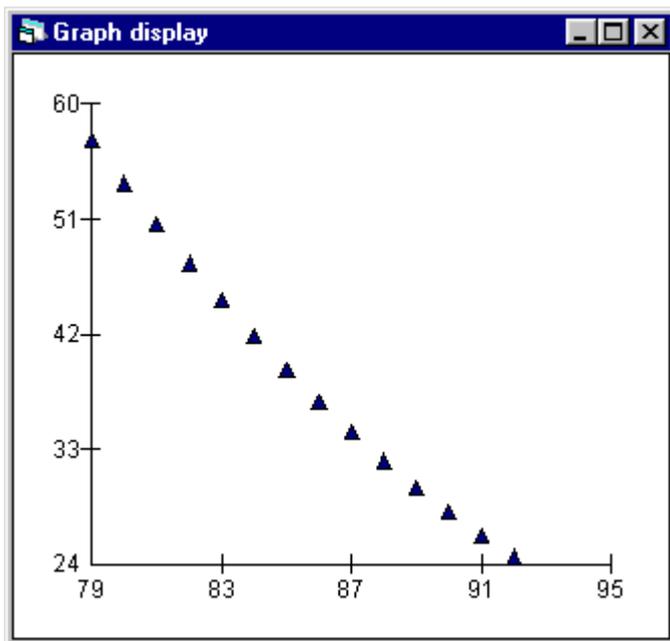
Go to the **Customised Graph** window

Select a new display **D3**

Select **y** to be VARF3

Select **x** to be YEAR

Click Apply



The level 3 (between COUNTY) variance has steadily decreased from a high of 57.2 in 1979 to a low of 24.6 in 1992. It therefore appears that absolute differentials between COUNTYs have been decreasing over time. Another way of examining this change is by looking at the prediction graphs. First calculate the predicted values using the random intercepts and slopes at COUNTY level:

Choose the **Predictions** window from the **Model** menu
 Click on x_0 , x_{1ijk} and x_{2ijk} to ensure that they are included
 Click on u_{0jk} and e_{0ijk} to remove them from the prediction
 Select PRED3 for **output from prediction to**
 Click **Calc**

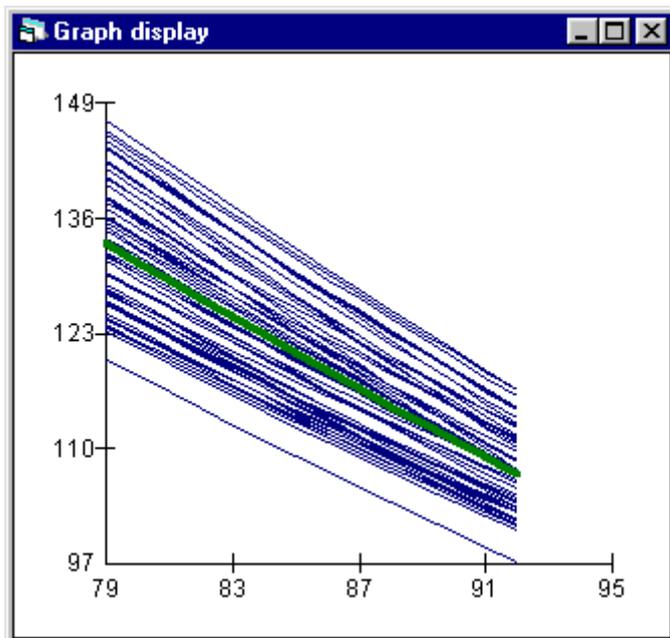
Next re-calculate the predicted values using the fixed part of the model only:

Click on v_{0k} and v_{1k} to remove these terms from the prediction
 Select PREDFP for **output from prediction to**
 Click **Calc**

To plot these new predictions:

Return to the **Customised Graph** window
 Select a new display **D4**
 Under **ds # 1**

Select **y** to be PRED3
Select **x** to be YEAR
Select COUNTY as the **group**
Change **plot type** to **line**
Click **Apply**
Choose **ds # 2**
Select **y** to be PREDFP
Select **x** to be YEAR
Change **plot type** to **line**
Under the **plot style** tab set **colour** to **2 green**
Set **line thickness** to **3**
Click **Apply**



The plot shows the individual predicted trends for each COUNTY plotted around the mean trend line shown in green. The fact that the COUNTY lines are converging toward the mean line over time demonstrates the decrease in level-3 variation over time.

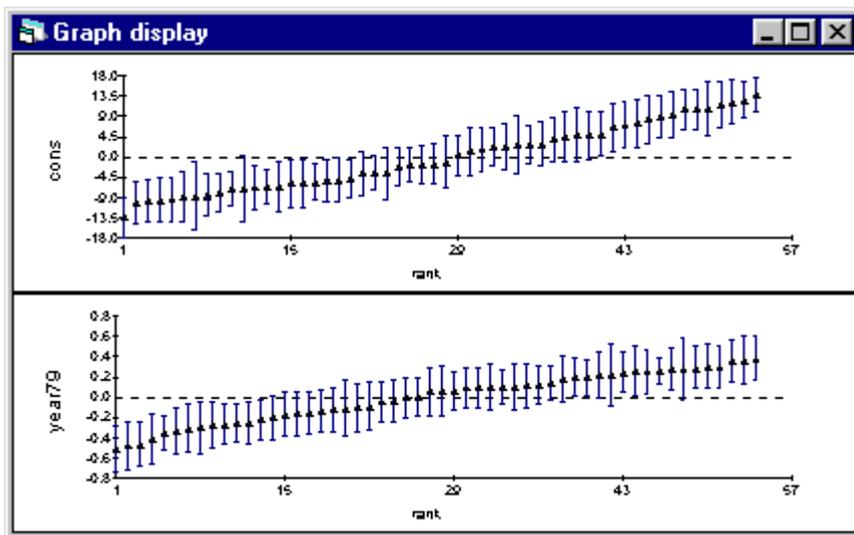
Higher-level residuals

There are now two sets of residuals at COUNTY level.

Under the **Model** menu, open the **Residuals** window
Click on the **Settings** tab
Select the **level** to be **3:COUNTY**
Choose a multiplier of 1.96 for the **SD (comparative) of residual to**
Click on **Set columns**

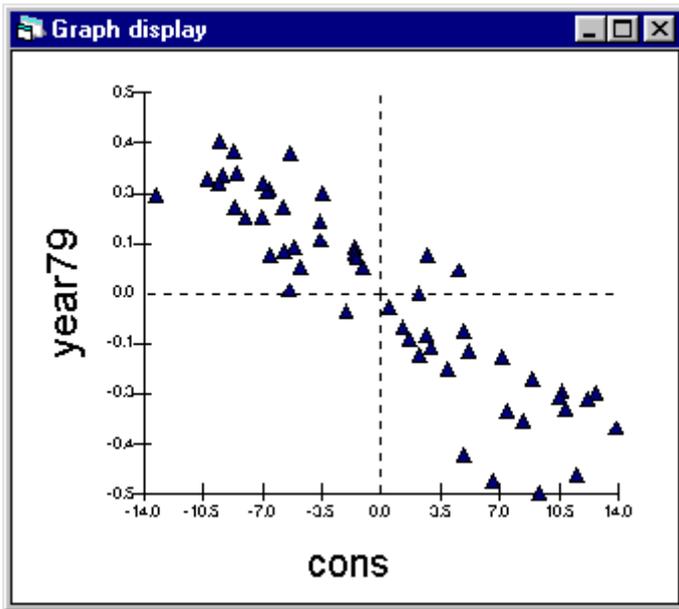
Each item in the list now has two columns assigned to it: the first column for the intercept residual, the second for the slope residual.

Click **Calc**
Select the **Plots** tab
Select **residual +/- 1.96 sd x rank**
Click **Apply**



These plots can be used to examine how many COUNTYs have slopes which differ from the average as well as how many have intercepts which differ from the average. Note that a COUNTY's rank for the intercept residual will not necessarily be the same as its rank for the slope residual. To see how the intercept and slope residuals are correlated between COUNTYs:

Return to the **Plots** tab in the **Residuals** window
Under the **pairwise** heading, select a **residuals** plot
Click **Apply**



This shows the strong negative correlation between the two sets of residuals. Those in the top left quadrant refer to those COUNTYs with negative intercept (CONS) residuals and positive slope (YEAR79) residuals. This suggests that those COUNTYs which had the lower than average SMRs in 1979 experienced a more gradual decrease in SMR over the 14 YEARS. Similarly, the COUNTYs featured in the bottom right quadrant are those which had above average SMRs in 1979 (positive CONS residual) but which experienced mortality decreasing at a faster than average rate (negative YEAR79 residual).

Complex level 1 variation

The multilevel framework allows variables to be random at any level so, for example, we may wish to extend the previous model such that trends in SMR not only vary across COUNTYs but also vary across DISTRICTs at level-2. However, random variables at level 1 have a slightly different interpretation; this concerns the effects of heterogeneity (i.e. non-constant variance). In this example, we may consider whether the variation between observations is constant throughout the 14 years or whether it changes.

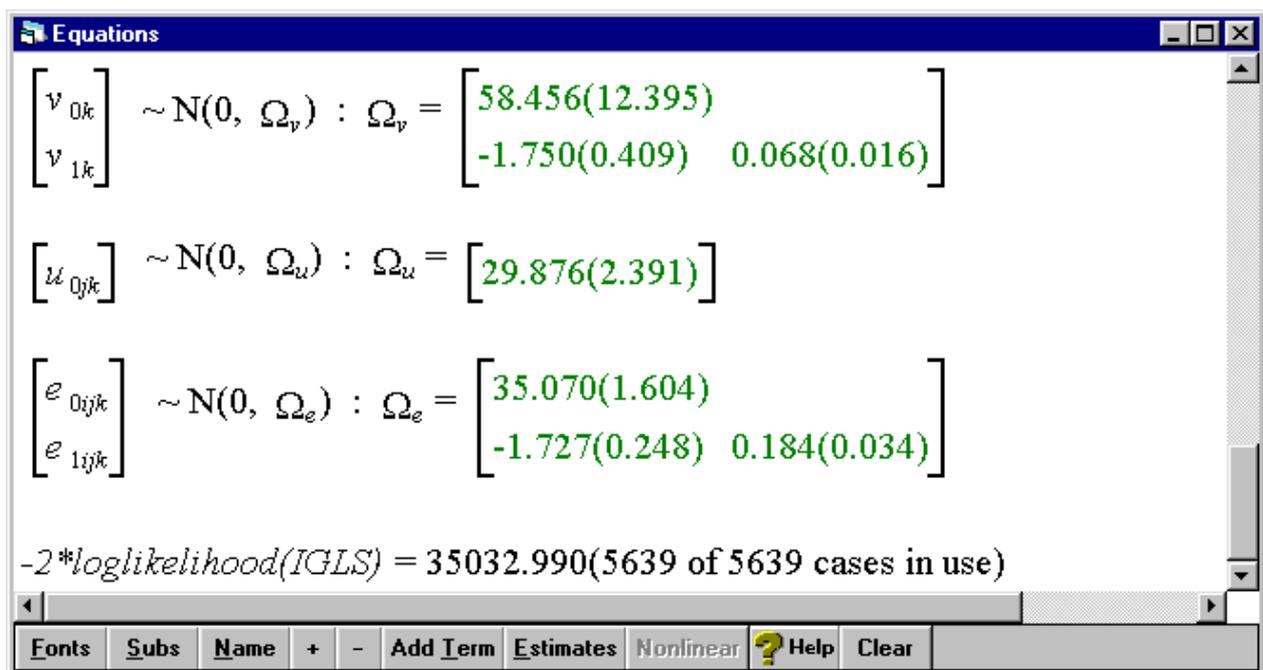
Return to the equations window

click on **year79_{ijk}**

check the box at **i(year)**

click done

Now estimate this model by clicking on the **More** button.



There is evidence of heterogeneity with a substantial reduction in $-2*\loglikelihood$. This means that the degree of scatter of individual observations about the predicted DISTRICT (level 2) means is not constant over time; it appears to have been decreasing. We can use the **Variance function** window to estimate the variance at each level, creating two new variables VARF2 and VARF1 and plotting these three variables against YEAR in the **Graph** window.

Open the **Variance function** window under the **Model** menu

Ensure that **1:YEAR** is selected to be the **level**

In the drop-down menu by **variance output to**, select C22

Click calc

Select **2:DISTRICT** to be the **level**

In the drop-down menu by **variance output to**, select C23

Click calc

Select **3:COUNTY** to be the **level**

In the drop-down menu by **variance output to**, select VARF3

Click calc

In the **Names window** name C22 VARF1 and C23 VARF2. To plot all of these variance functions across the observed values of YEAR79:

Go to the **Customised Graph** window

Select display **D3**

Select **y** to be VARF3

Select **x** to be YEAR

Select the **plot type** to be **line**

Under the **plot style** tab, select the **line thickness** to be **3**

Click **Apply**

Select **ds#2** with VARF2 as the **y** variable, YEAR as the **x** variable, and the **plot type** to be **line**

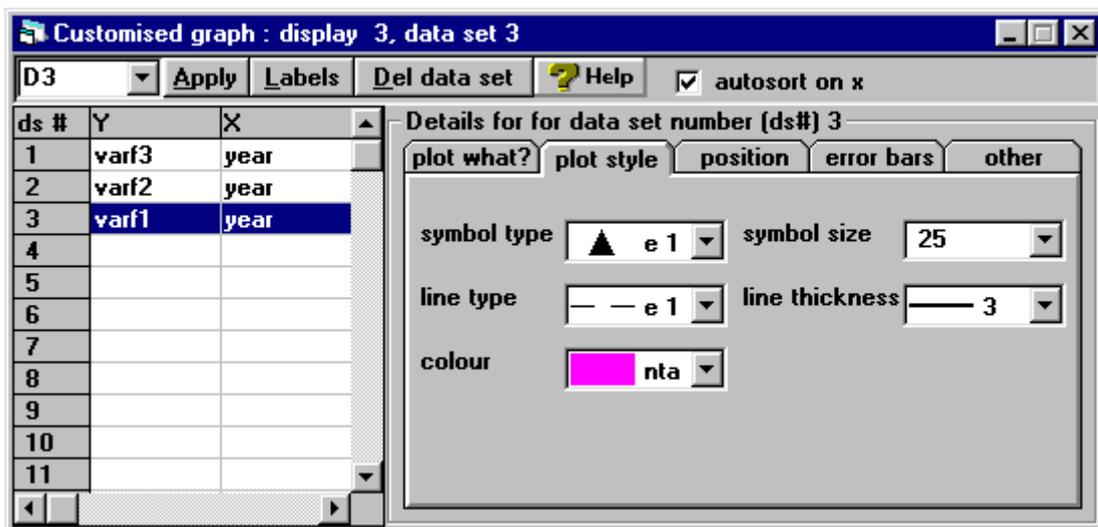
Under the **plot style** tab, select the **colour** to be **4 red** and the **line thickness** to be **3**

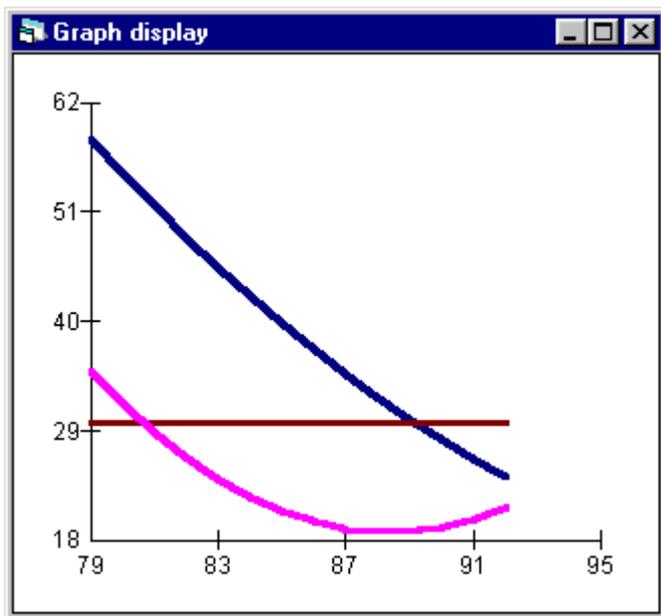
Click **Apply**

Select **ds#3** with VARF1 as the **y** variable, YEAR as the **x** variable, and the **plot type** to be **line**

Under the **plot style** tab, select the **colour** to be **13 Light magenta** and the **line thickness** to be **3**

Click **Apply**





We have not fitted any random effects at level 2, so the variation between DISTRICTs within COUNTYs is assumed to be constant. The variation between COUNTYs decreased steadily between 1979 and 1992; however, the level 1 variance decreased from 1979 to 1988 but it appears to have increased slightly since then. The total variation has decreased from 123 in 1979 to just 76 in 1992. In a similar manner it is possible to explore the extent to which the level 2 variation (between DISTRICTs) has also been varying over time.

By this stage the user has become familiar with the basics of model fitting for continuous (normally distributed) responses. The fixed part of the model can be built up as with an ordinary least squares (OLS) regression model, including any combination of continuous and categorical variables and interactions between them. The significance and effect of variables can be examined through changes in the likelihood or through comparisons of the parameter estimates with their estimated standard errors.

The difference between such models and OLS regression is the ability to separate the variance into the different levels in the model – COUNTY, DISTRICT and YEAR in this example – and then to model this variance by considering other variables to be random at any of the levels. At higher levels this has the interpretation of fitting random slopes; at the lowest level this is modelling heterogeneity (non-constant variance) within the data. We are again able to test for the significance of any of these random terms.

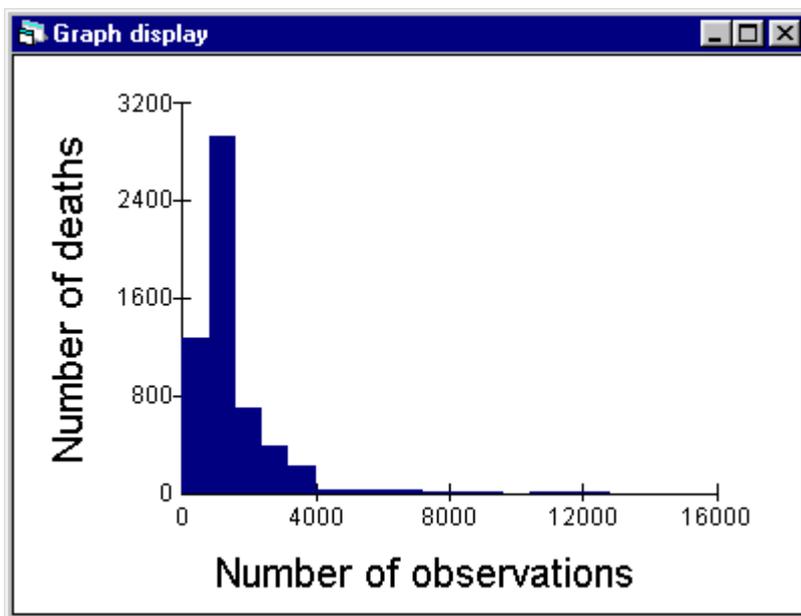
The example used has been illustrative of the methods employed when fitting a multilevel model; it is not, however, the way in which we would normally model such data. The next section goes on to

consider a generalised linear model for these data; however, before proceeding to the more complex modelling it is important to have a good understanding of the basics covered up to this point.

A POISSON MODEL – INTRODUCTION

The model that we have fitted assumes that the standardised mortality ratio follows a normal distribution. We found that the variance decreased over the period 1979-1992; over this time the standardised mortality ratio also fell. This suggests that there may be a link between the variance in a particular year and the average mortality rate in that year. We have also attached equal importance to every area and in every year; this is probably not sensible since the size of areas in terms of their populations and the number of deaths observed varies considerably both across areas and over time. One possibility would be to weight each observation according to the population of the district in that year; this requires weighting at each level of analysis and would ensure that areas from which we have the most information – the largest areas in terms of their populations – are afforded the most weight. In this section we adopt an alternative approach.

The local mortality datapack is based on **counts** of deaths. Instead of modelling a transformation of this response – the SMR – we can consider modelling the actual counts of deaths. Such data tend to be discrete rather than continuous – you can't observe fractions of deaths – and they also tend to be extremely skewed (see histogram below). Therefore, the assumption of a normal distribution is usually not sensible.



Instead we can fit a generalised linear model and approximate a Poisson distribution for the data.

SETTING UP A GENERALISED LINEAR MODEL IN MLWIN

First open the original worksheet LMDP.ws again.

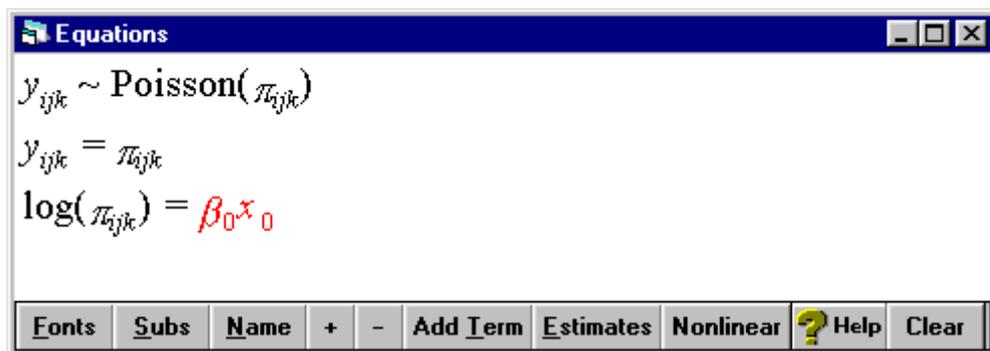
Go to the **File** menu
Select **Open worksheet**
Open the worksheet called **lmdp.ws**

Then go to the **Equations** window.

Click on either of the **y** terms
Select DEATHS as the dependent variable
Select 3 – ijk for **N levels**
Select COUNTY for **level 3(k)**
Select DISTRICT for **level 2(j)**
Select YEAR for **level 1(i)**
Click on **Done**

So far we have simply repeated the steps for the 3-level model in the introductory tutorial with the response variable being DEATHS rather than SMR. We now have to amend the default distribution for the response. In the **Equations** window

Click on the **N** that defines the Normal distribution
Select **Poisson** from the list
Change the **logit** link function to **log**



These steps have specified the response to be a Poisson random variable, which defines the lowest level variance function, and the linearising function of the response to be the natural logarithm. As in the normal multilevel model, we need to define a CONStant again.

Go to **Data manipulation** menu
Select **Generate vector**
Select **Type of vector** to be **Constant vector**
Select C9 to be the **Output column**
Enter 5639 (the number of data points) beside **Number of copies**
Enter 1 beside **Value**
Click the **Generate** button

The linearising function used when estimating generalised linear models means that the level one variance is on a different scale to the variance at higher levels. This means that we require two copies of the CONS column; one to model the level 1 variation and one to model the higher level variation.

Go to **Data manipulation** menu
Select **Calculate**
Select C10 and press the right arrow button
Click the '=' button on the keypad
Select C9 and press the right arrow button
Click the **Calculate** button

Open the **Names** window and name C9 and C10 CONS and PCONS respectively. Now return to the **Equations** window.

Click on β_0x_0
Select CONS from the drop-down list
Click on the check box by **j(DISTRICT)**
Click on the check box by **k(COUNTY)**
Click on **Done**

With these commands, we have added the constant to the fixed part of the model to estimate the intercept and have also allowed the intercept to vary across DISTRICTs and COUNTYs. Still in the **Equations** window:

Click on the **Add term** button

Click on the new explanatory variable and choose PCONS from the drop-down list

Remove this from the fixed part by clicking on the check-box by **Fixed parameter**

Click on the check box by **i(YEAR)**

Click on **Done**

These commands have added PCONS to the model in order to model the level-1 Poisson variation only. As in the introductory tutorial, we will fit a quadratic in YEAR.

Go to **Data manipulation** menu

Select **Calculate**

Select the empty column C11 and press the right arrow button

Click the '=' button on the keypad

Select YEAR from the list of variables and press the right arrow button

Use the window's keypad to enter **-79**

Press **Calculate**

Clear this calculation using the backspace or delete buttons on your keyboard

Next, select the empty column C12 and press the right arrow button

Click the '=' button on the keypad

Select C11 from the list of variables and press the right arrow button

Use the window's keypad to enter **^2**

Press **Calculate**

Use the **Names** window to name C11 and C12 YEAR79 and YEAR79^2 respectively. Next, use the **Add Terms** button to add both terms to the fixed part of the model only. The **Equations** window should now look like this (remember you can use the **Name** and + buttons to display more information about the current model in the **Equations** window):

Equations

$$\left. \begin{aligned} \text{deaths}_{ijk} &\sim \text{Poisson}(\pi_{ijk}) \\ \text{deaths}_{ijk} &= \pi_{ijk} + e_{1ijk} \text{pcons}^* \end{aligned} \right\}$$

$$\log(\pi_{ijk}) = \beta_{0jk} \text{cons} + \beta_2 \text{year}79_{ijk} + \beta_3 \text{year}79^2_{ijk}$$

$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} \sigma_{v0}^2 \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 \end{bmatrix}$$

$$\text{pcons}^* = \text{pcons} \pi_{ijk}^{0.5}$$

$$\begin{bmatrix} e_{1ijk} \end{bmatrix} \sim (0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma_{e1}^2 \end{bmatrix}$$

Fonts Subs Name + - Add Term Estimates Nonlinear Help Clear

The response, DEATHS, is set equal to the predicted number of deaths, π_{ijk} , to be estimated, plus an error term e_{1ijk} (multiplied by a factor PCONS* which may be ignored temporarily). The e_{1ijk} 's are the level 1 residuals and the variance of these residuals is constrained by the assumption of a Poisson distribution. The predicted number of deaths is then estimated by taking the log of π_{ijk} (i.e. linearising the response) and setting this equal to the linear predictor on the right hand side. This linear predictor is estimated as a quadratic function of time and the intercept in the predictor, β_{0jk} , varies across both COUNTYs and DISTRICTs via the random effects v_{0k} and u_{0jk} , respectively. As in the first part of this tutorial, these random effects are assumed to be normally distributed with zero means. The current model will provide estimates of how the average number of deaths has changed over time (the fixed part) and how the average number of deaths varies between districts and counties (the random part).

The offset

The model described above will fit the observed number of DEATHS in an area using just a mean and a linear and quadratic term in YEAR. However, unlike the SMR this response variable has not been scaled. That is, the SMR of an average DISTRICT in 1992 should be 100; the number of DEATHS in

that DISTRICT may be 10 or 10000 depending on the size of the population. All that an SMR of 100 tells us is that the observed number of DEATHS is the same as the EXPECTED number; we are now trying to fit that observed number and so need to account for the EXPECTED number in our model. We will do this by including it as an *offset* term. We can think of this as modelling the log of the ratio of the predicted deaths π_{ijk} to the EXPECTED deaths E_{ijk} as

$$\log\left(\frac{\pi_{ijk}}{E_{ijk}}\right) = \beta_{0,jk}x_0 + \beta_2x_{2ijk} + \beta_3x_{3ijk}$$

In terms of the predicted number of deaths this can be rewritten as

$$\log(\pi_{ijk}) = \log(E_{ijk}) + \beta_{0,jk}x_0 + \beta_2x_{2ijk} + \beta_3x_{3ijk}$$

In other words, the logarithm of the EXPECTED number of deaths in each area, based on population size and age-sex composition, is entered into the regression equation but its coefficient is fixed at 1 rather than being estimated freely, as is the case with the covariate coefficients for CONS, YEAR79 and YEAR79². MLwiN provides a facility to do this; the variable to be offset must be named OFFS.

Go to **Data manipulation** menu

Select **Calculate**

Select the empty column C13 and press the right arrow button

Click the '=' button on the keypad

Select LOGE from the list of functions and press the up arrow button

Click the '(' button on the keypad

Select EXPECTED from the list of variables and press the right arrow button

Click the ')' button on the keypad

Click the **Calculate** button

In the **Names** window, name C13 OFFS. This variable is now included in all subsequent Poisson models unless it is renamed or removed using the OFFSET command.

Nonlinear estimation

As mentioned above, generalised linear models are approximated in MLwiN by using a linearising function based upon an expansion of the Taylor series. Specialist knowledge of this approximation is not necessary, however, users should be aware of the following options which are available when using nonlinear estimation.

Click the **Nonlinear** button at the bottom of the **Equations** window

A window appears and provides details of the options for three settings:

- **Distributional assumptions** gives us the options of **Poisson** or **extra Poisson** variation at level 1. A **Poisson** distribution has an equal mean and variance such that $E(y_{ijk}) = Var(y_{ijk}) = \pi_{ijk}$. However, it may be that such a distribution does not fit the data well; the most common situation is one in which the tail of the observed distribution is too heavy. We can sometimes obtain a better approximation to the data by allowing **extra Poisson** variation; the mean remains unchanged but we fit the variance as $Var(y_{ijk}) = \pi_{ijk} \sigma_e^2$. **Poisson** (distributional) variation can then be seen to be a special case of this in which $\sigma_e^2 = 1$.
- **Linearisation** gives us the choice of using a **1st order** or **2nd order** approximation to the Taylor series.

- **Estimation type** gives us the option of using marginal quasi-likelihood (**ML**) or penalised quasi-likelihood (**PQL**).

The latter two options affect the way in which coefficients are estimated. Bias in parameter estimates tends to be lower when using **2nd order** approximations and **PQL** estimation; however, there is an associated cost in as much as estimation may take longer. The **PQL** estimation procedure is also somewhat less robust and you may experience problems with convergence. A guideline is often to use **1st order, ML** when exploring the data and to use **2nd order, PQL** to test the model and obtain final estimates.

We will begin by using the default settings, assuming **Poisson** variation and a **1st order, ML** estimation procedure. These options may be set by clicking the **Use Defaults** button in the **Nonlinear Estimation** window and then clicking **Done**.

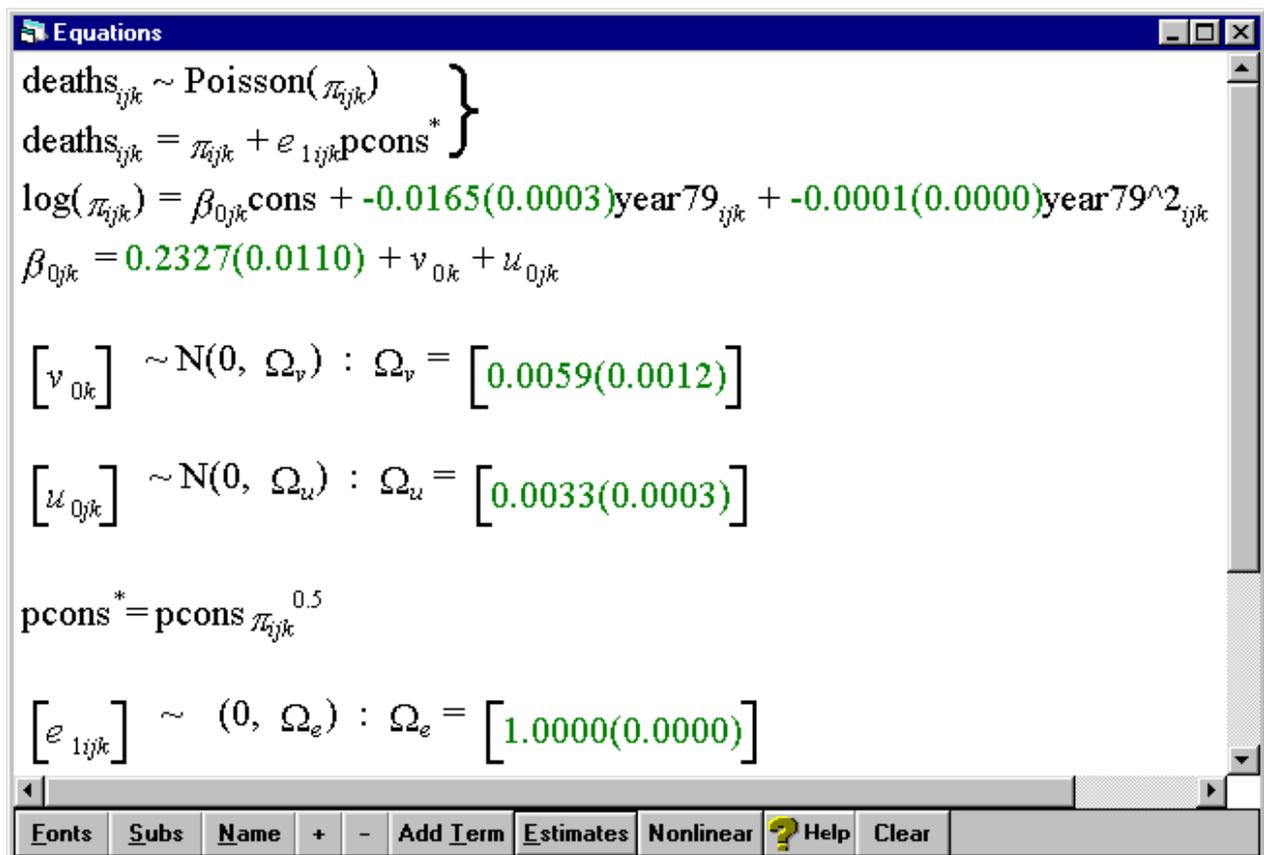
MODEL INTERPRETATION

Press the **Start** button to estimate the model

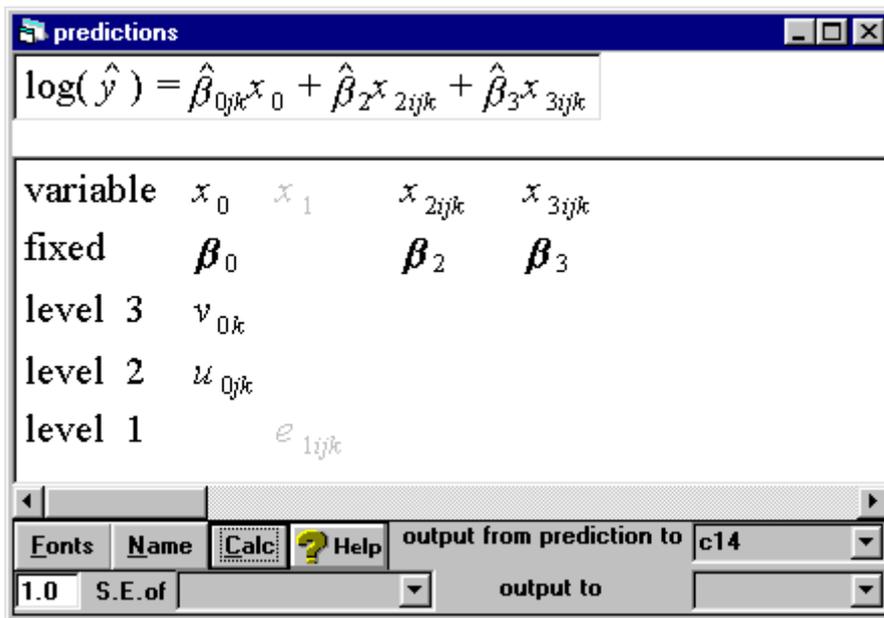
To view the estimates, it will be helpful to change the precision of the display.

Go to the **Options** menu
 Select **Numbers**
 increase the # **digits after decimal point** to 4
 Click **Apply** and then **Done**

By clicking on the **Estimates** button in the **Equations** window, the following should appear:



The parameter estimates are now on the log scale and should be treated as such with the OFFSET term included; for example, the average number of deaths in 1979 has been fitted as 1.262 ($e^{0.2327}$) times the expected number. We can see what is going on more clearly using the **Graph** window. First of all we will get Predictions by DISTRICT, output these to c14 and name this column PRED2.



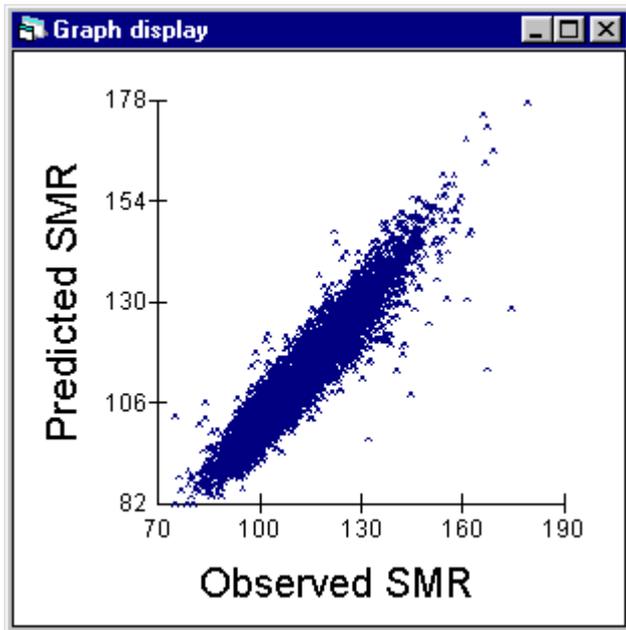
In a similar manner put the predicted values for the fixed part in c15 and name this column PREDFP, and the level 3 predictions in c16 and name this column PRED3. You will note from the summary statistics in the **Names** window that these prediction equations are on the log scale; they also do not include our OFFSet term. As such, we really have the predicted values

$$\log\left(\frac{\hat{y}}{E_{ijk}}\right)$$

We can very easily convert these to predicted SMRs by taking the EXPOnts in the **Calculate** window:

- Go to **Data manipulation** menu
- Select **Calculate**
- Select the PRED2 and press the right arrow button
- Type =100* using the keypad
- Select the function EXPOntial from the list and press the up arrow button
- Click the '(' button on the keypad
- Select PRED2 from the list of variables and press the right arrow button
- Click the ')' button on the keypad
- Click the **Calculate** button

Repeat this process for the variables PREDFP and PRED3. We can now plot the predicted SMR against the observed values; PRED2 includes DISTRICT and COUNTY effects but assumes that the year-on-year fluctuations are part of a Poisson process.



We can also plot the predicted values at national and COUNTY level by YEAR:

ds #	Y	X
1	pred3	year
2	predfp	year
3		
4		
5		
6		
7		
8		
9		
10		
11		

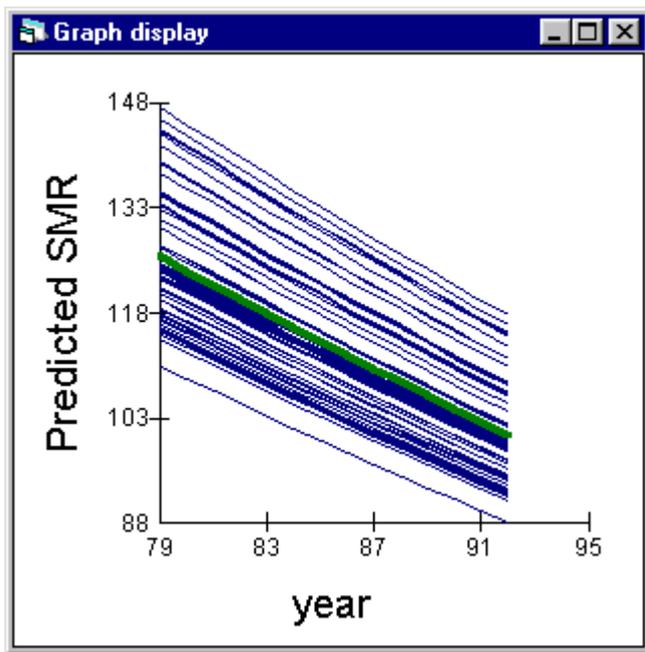
Details for for data set number (ds#) 1

plot what? plot style position error bars other

y: pred3 x: year

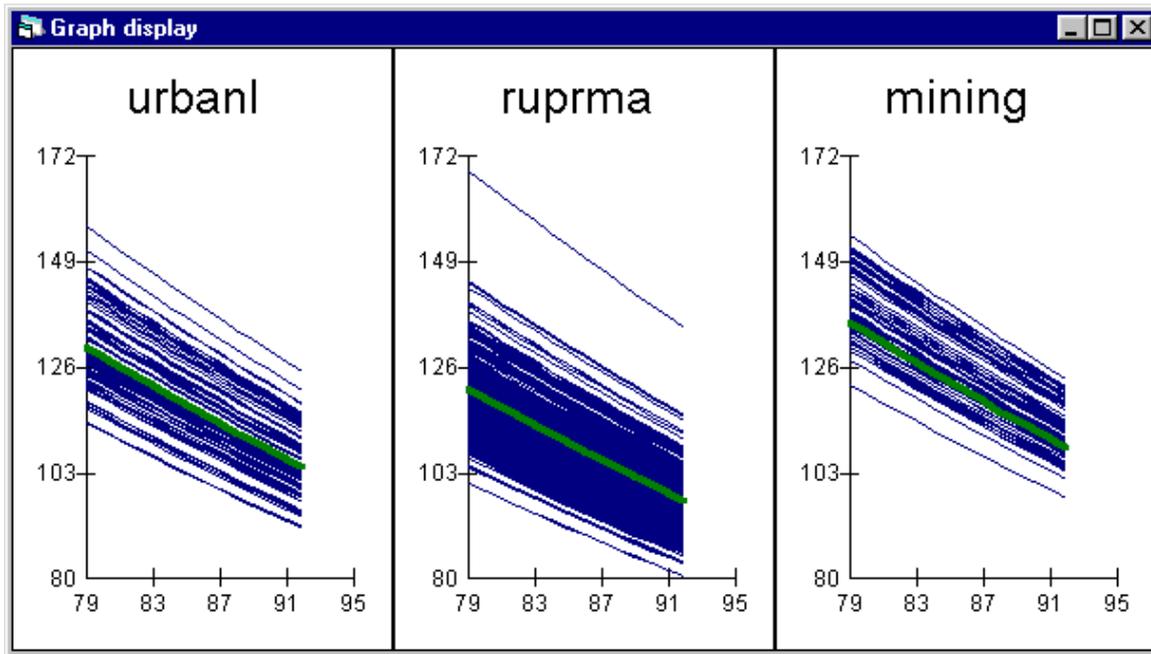
filter: [none] group: county

plot type: line



This graph illustrates the convergence of SMRs that we noted in the previous analysis; this is to be expected since the assumption of Poisson variation means that we can expect the variation to decrease as the number of DEATHS decreases.

You can continue to build up the model as before, entering random effects where appropriate. The plots of predicted SMRs can be broken down into the three area groupings – urban areas and inner London (URBANL), rural, prospering and maturer areas (RUPRMA) and MINING using the layout option of the Graph window. These might look as follows.



These graphs indicate that there are clear differences between the three types of area in terms of their mean SMR, with MINING areas tending to have the highest SMRs. One of the RURAL districts – DISTRICT 4820 – appears to be outlying with the highest predicted SMR over the period.

The user should now be familiar with the ideas behind modelling mortality data using Poisson regression models. There are many unanswered questions within this data set which stem from the above introductory analysis; for example, are the trends over time the same for the three area groupings or should different slopes be fitted for each? And are the slopes random across DISTRICTs and COUNTYs? The above plot also suggests that the variance of the intercept for DISTRICTs might be different for the three area types, possibly being higher for the RURAL, PROSPERing and MATURER areas and lower for the MINING areas. The user is encouraged to explore the data further with these questions in mind.