Chapter 11
Multilevel Linear Regression Using MLwiN: Mortality in England and Wales, 1979–1992

Abstract In this chapter, the reader becomes a user. This chapter contains the first of three tutorials that readers can work through, using the specialist multilevel modelling software MLwiN. We introduce MLwiN as a package; this software will also be used in the other two tutorials. This tutorial introduces practical linear multilevel analysis. It uses data on mortality in England and Wales over time. The dependent or outcome variable is the standard mortality ratio in a given year between 1972 and 1996, for districts which are nested within counties.

Because this is the first tutorial, we go into some detail regarding the use of MLwiN, and how to use it to manipulate and explore the data. The tutorial starts with the estimation of a single-level model, then moves on to a two-level and three-level model. We begin with a random intercept model and progress to a random slope model. Throughout the tutorial graphs are used to enable visualisation of the results of the analyses. At the end of this tutorial, we detail an alternative analysis of the data using a multilevel Poisson model.

Keywords Tutorial · Multilevel analysis · Linear regression · Poisson regression · Mortality

This chapter is based on training materials created by Leyland and McLeod (2000). The training materials in this chapter and the two chapters that follow are designed to be used either constituting part of a formal course or as a self-learning aid. They provide an introduction to the ideas behind multilevel modelling and a guide to analysis using the software package MLwiN. Further details on multilevel modelling and MLwiN are available from the Centre for Multilevel Modelling http://www.bristol.ac.uk/cmm/. The materials have been written for MLwiN v3.01. The teaching version of the software is available from https://www.bristol.ac.uk/cmm/software/mlwin/download/.

When working through the examples in this book, the user should periodically save the worksheet. Throughout these materials the instructions to the user appear in boxes. Selections to be made by the user appear in **bold type**, and variable names are given in CAPITALS. If you have to click on a term in an equation, this is presented in ***bold and italics***.

A. H. Leyland, P. P. Groenewegen, *Multilevel Modelling for Public Health and Health Services Research*, https://doi.org/10.1007/978-3-030-34801-4_11

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 1979–1992. These data can be found at the UK Data Service. 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 (districts in England and Wales) in each year. For further information on this and other available datasets, the user should visit the UK Data Service website https://discover.ukdataservice.ac.uk/.

Research Questions

In this tutorial, we will 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 and these store all the data and model settings from the last saved version. We will start by opening the file '*lmdp.wsz*'—an MLwiN worksheet that has already been prepared for analysis.

In MLwiN, go to the **File** menu
Select **Open worksheet**
Navigate to the folder containing the data file
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

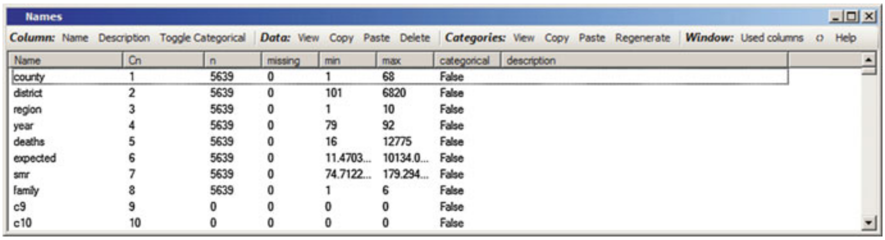
We can view a summary of this worksheet using the **Names** window. This will automatically appear when a worksheet is opened in MLwiN; at other times, the **Names** window can be called up as follows:

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

This shows a list of all the variables stored in the worksheet together with some summary information. The worksheet contains 8 variables; these are at the beginning of the worksheet in columns number 1–8. Each column contains 5639 data points and no missing values. Each data point (observation) corresponds to the annual number of deaths in a given district in England and Wales for 1 year in the period 1979–1992. 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 1 of 10 REGIONs. The data cover 14 YEARs from 1979 to 1992 inclusive. Note that there are only 5639 data points rather than the 5642 that might be expected (403 DISTRICTs with an observation for each of 14 years); 3 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 by applying the 1992 national age- and sex-specific mortality rates. 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

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

and reflects the excess deaths in an area, standardised for age and sex, over the national average mortality rate in 1992 (average = 100). The standardisation means that differences between areas in the age and sex structures of their populations are taken into account. The range from 75 to 179 implies a minimum mortality rate for one area in 1 year 25% below the 1992 average and a maximum 79% above the average. Finally, the variable FAMILY is a classification of districts into six groups devised by the UK's 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.



Name	Cn	n	missing	min	max	categorical	description
county	1	5639	0	1	68	False	
district	2	5639	0	101	6820	False	
region	3	5639	0	1	10	False	
year	4	5639	0	79	92	False	
deaths	5	5639	0	16	12775	False	
expected	6	5639	0	11.4703...	10134.0...	False	
smr	7	5639	0	74.7122...	179.294...	False	
family	8	5639	0	1	6	False	
c9	9	0	0	0	0	False	
c10	10	0	0	0	0	False	

Data Window

The data may be viewed and edited in a spreadsheet format.

Go to the **Data manipulation** menu
Select **View or edit data**
Alternatively, the **Data** window may be accessed from the **Names** window:
In the **Names** window, highlight columns 1–8 (use the shift or control keys to highlight multiple columns)
Click the **View** button in the **Data** section at the top of the **Names** window

The **view** button at the top of the **Data** window can be used to change or extend the selection of variables shown; simply select the desired variables from the drop-down list.

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	1.000	101.000	3.000	79.000	50.000	36.072	138.611	1.000
2	1.000	101.000	3.000	81.000	28.000	28.159	99.435	1.000
3	1.000	101.000	3.000	82.000	53.000	37.721	140.506	1.000
4	1.000	101.000	3.000	83.000	43.000	38.468	111.780	1.000
5	1.000	101.000	3.000	84.000	41.000	39.181	104.642	1.000
6	1.000	101.000	3.000	85.000	34.000	32.200	105.592	1.000
7	1.000	101.000	3.000	86.000	32.000	38.185	83.803	1.000
8	1.000	101.000	3.000	87.000	38.000	37.363	101.706	1.000
9	1.000	101.000	3.000	88.000	28.000	37.477	74.712	1.000
10	1.000	101.000	3.000	89.000	30.000	36.463	82.275	1.000
11	1.000	101.000	3.000	90.000	38.000	35.658	106.569	1.000
12	1.000	101.000	3.000	91.000	43.000	32.528	132.193	1.000
13	1.000	101.000	3.000	92.000	32.000	37.279	85.840	1.000
14	1.000	111.000	3.000	79.000	2160.000	1533.833	140.824	6.000
15	1.000	111.000	3.000	80.000	2094.000	1526.076	137.215	6.000
16	1.000	111.000	3.000	81.000	2081.000	1568.919	132.639	6.000
17	1.000	111.000	3.000	82.000	2041.000	1585.704	128.713	6.000
18	1.000	111.000	3.000	83.000	2011.000	1582.598	127.070	6.000
19	1.000	111.000	3.000	84.000	1854.000	1595.079	116.233	6.000
20	1.000	111.000	3.000	85.000	1997.000	1604.626	124.453	6.000
21	1.000	111.000	3.000	86.000	1971.000	1604.318	122.856	6.000
22	1.000	111.000	3.000	87.000	1868.000	1610.334	116.001	6.000
23	1.000	111.000	3.000	88.000	1864.000	1609.627	115.803	6.000
24	1.000	111.000	3.000	89.000	1760.000	1604.886	109.665	6.000

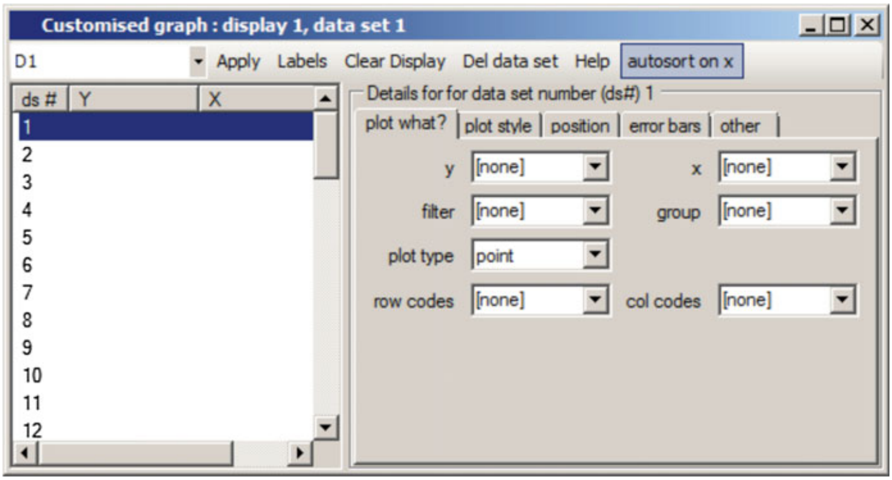
The first 13 observations are made on DISTRICT 101, COUNTY 1, REGION 3. The 13 observations on this DISTRICT can be seen to correspond to 13 YEARs of data; there is no observation for 1980. The estimated SMR in this district ranges from 75 in 1988 to 141 in 1982. The district classification (FAMILY)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

Graph Window

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

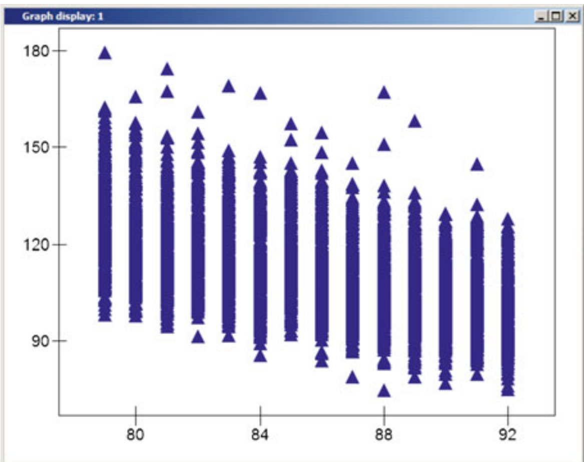
Go to the **Graphs** menu
Select **Customised Graphs**

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 (horizontal and vertical) axes showing lines, points or bars, and each display can show an array of up to 5 × 5 graphs. The **position** tab towards the top right of the customised graph window is used to specify the layout—the position of the graphs in the display. Finally, each graph can plot one or more *datasets*, each one consisting of a set of x and y coordinates selected from the worksheet columns. Different datasets may be specified by clicking on different rows in the table under the **ds#** heading shown at the left-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

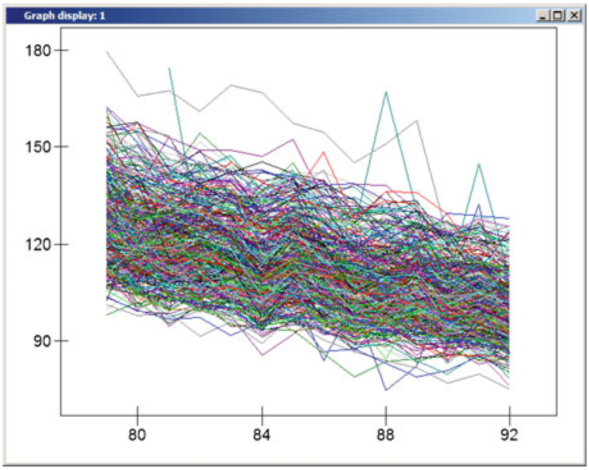
Select the **y** variable to be SMR from the drop-down list
Select the **x** variable to be YEAR from the drop-down list
Click the **Apply** button



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 overall SMR—for the whole of England and Wales—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 **rotate**
Click the **Apply** button



It is possible to identify points on the graph: point and click anywhere on the graph and the **Graph options** 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 minimise 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**.

This section has covered data exploration using:
Names window—data summaries

(continued)

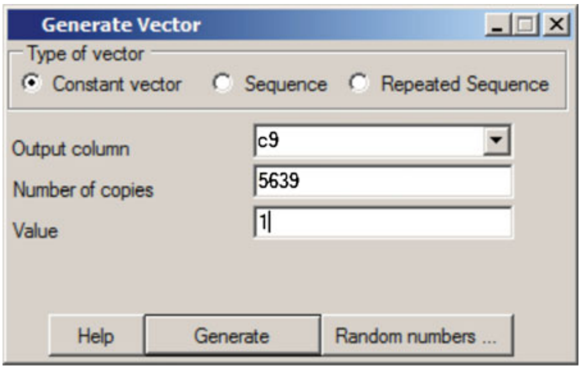
Data window—spreadsheet
Graph—scatter plot
Graph—line graph

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 using MLwiN, we need to create a column of 1's that spans the entire data set. This variable will also be used to model the variance at each level in a multilevel model. We use the **Generate vector** window to create a column containing 5639 occurrences of the value 1.

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

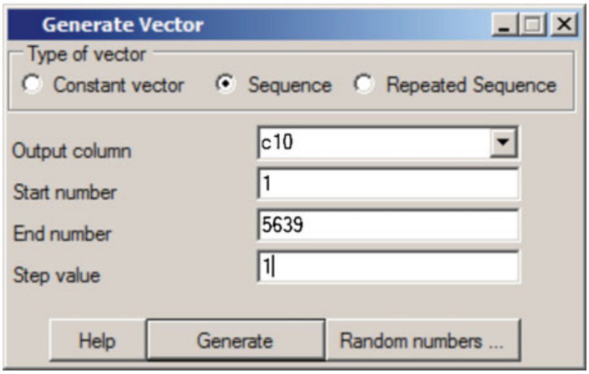


Returning to the **Names** window, column C9 now contains 5639 data points each with the value of 1. We can give this new variable a name:

Click on C9 in the **Names** window
Click on the **Name** button in the Column section at the top of the window
Type CONS and press <return>

We can also use the **Generate vector** window to create a unique identifier for every data point or observation.

In the **Generate vector** window:
Select **Type of vector** to be **Sequence**
Select C10 to be the **Output column**
Enter 1 beside **Start number**
Enter 5639 (the number of data points) beside **End number**
Enter 1 beside **Step value**
Click the **Generate** button



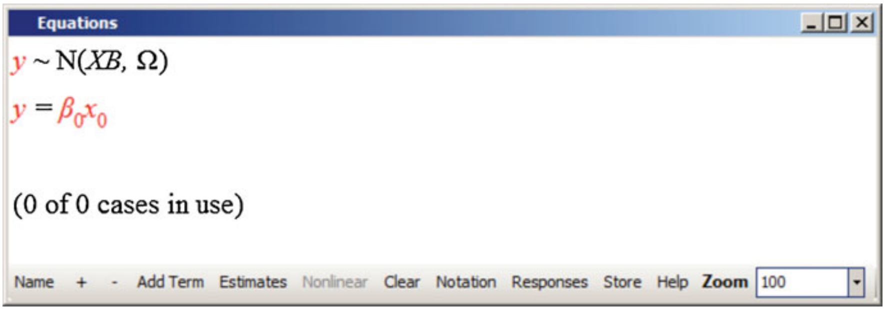
In the **Names** window, column C10 should now contain 5639 data points with a minimum of 1 and a maximum of 5639. We will name this variable:

Click on C10 in the **Names** window
Click on **Name** at the top of the window
Type ID and press <return>

Equations Window

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

Go to **Models** menu
Select **Equations**

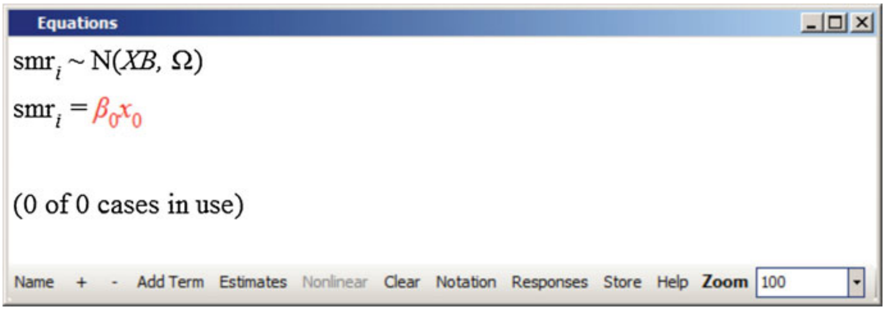


The terms in red are those which must be defined before a model can be fitted to the data. We begin by specifying our outcome:

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. We will start by fitting a single-level (Ordinary Least Squares—OLS) model. The level 1 units, our observations, are identified by the variable ID.

Select 1 – i for **N levels**
Select ID for **level 1(i)**
Click on **Done**



The red response variable ***y*** has been replaced by the term $\boldsymbol{smr_i}$, the black colour indicating that this term has been defined; moreover, the addition of a subscript

i indicates that this is a single-level model. In a similar manner, we can define CONS—the column of 1's that we just created—to be an independent variable.

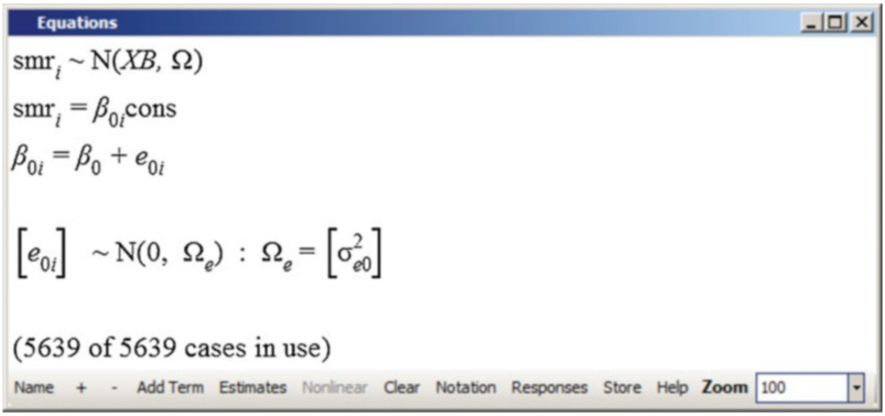
Click on the $\boldsymbol{\beta}_0\boldsymbol{x}_0$ term
Select CONS from the drop-down list

The check boxes indicate in what part of the model each variable is to be included; 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 option in this window relates to the random part of the model. We allow for random error at level 1 by setting the CONStant term to be random at this level.

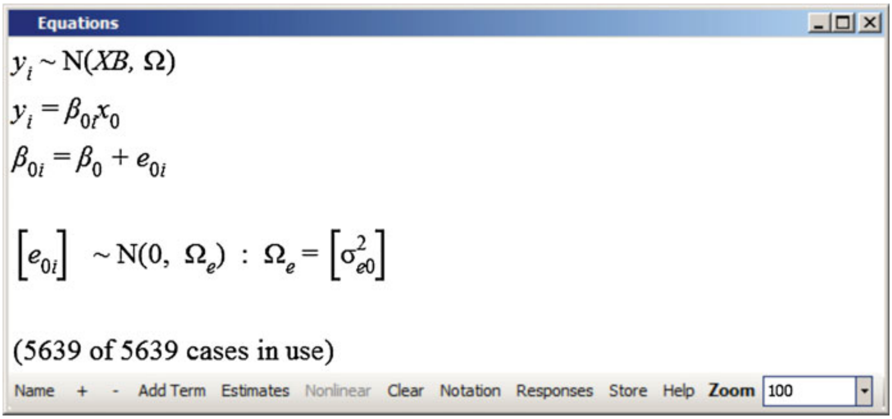
Click on the check box by **i(ID)**
Click on **Done**

Note that the term $\boldsymbol{\beta}_0$ changes to $\boldsymbol{\beta}_{0i}$, denoting the fact that it is random at level 1 (between occasions).

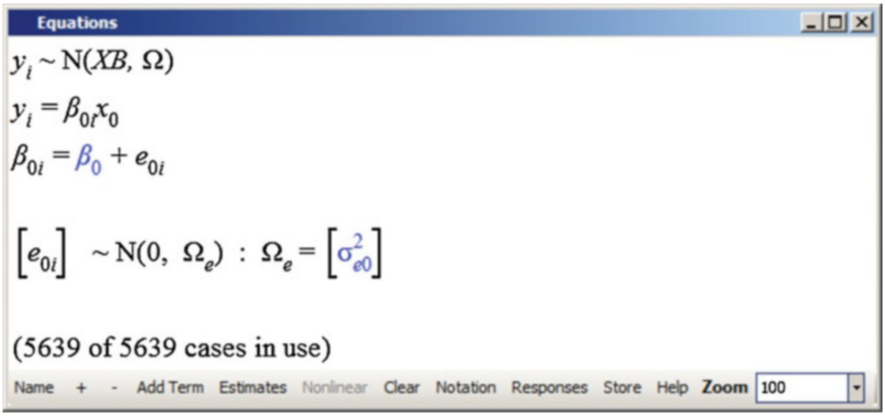
To expand this model to see the distributional assumptions and error structure, click twice on the '+' at the bottom of the **Equations** window.



This shows our assumption of a normal distribution for the residuals $e_{0i} \sim N\left(0, \sigma_{e0}^2\right)$. At any time we can toggle between the representation of the model that includes the names of all of the variables and a purely algebraic representation; simply click on the **Name** button at the bottom of the **Equation** window.



Note that the names of the specified dependent and independent variables, SMR and CONS, have been replaced by $\boldsymbol{y}$ and $\boldsymbol{x}$. If you click on the **Estimates** button, you can see that two terms in the model are blue: the grand intercept $\boldsymbol{\beta}_0$ and the level 1 variance $\boldsymbol{\sigma}_{e0}^2$. The fact that they are blue 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 0.000 because no model has yet been estimated). Since our variable CONS is just a column of 1's, the above equation is just fitting the SMR of the *i*th observation using a mean $\boldsymbol{\beta}_0$ and a residual or error term $\boldsymbol{e}_{0i}$. We are going to begin by assuming that these $\boldsymbol{e}_{0i}$ are independent and identically distributed. This means that if the SMR in a particular DISTRICT is higher than the mean one YEAR, we believe that the SMR in another YEAR is just as likely to be below the mean as above. In other words, we are fitting a model that assumes that there will not be certain DISTRICTs with persistently high SMRs and others where mortality is consistently below the mean.

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
Select YEAR from the drop-down list under **variable** in the **Specify term** window
Click **Done**

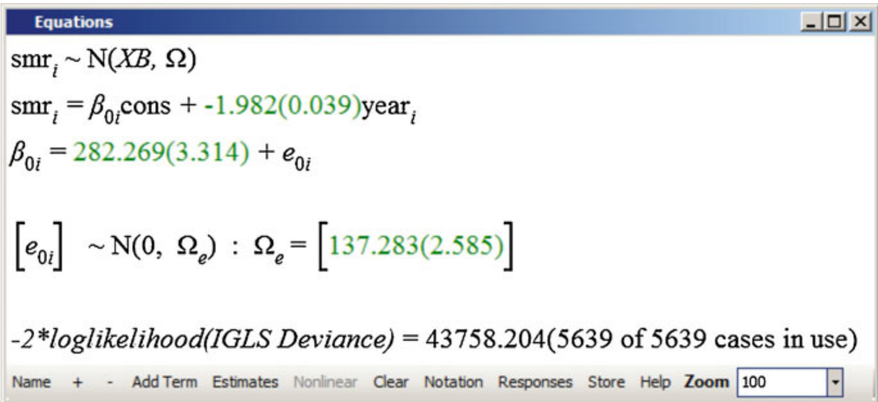
The third term to be estimated, $\boldsymbol{\beta}_1$, is the regression coefficient (slope) associated with YEAR and this will estimate the trend in SMRs during the study period.

Fitting the Model

The model is now ready to be estimated.

Click the **Start** button on the tool bar at the top left-hand corner of the MLwiN screen

After two iterations (the iteration number is given at the bottom of the MLwiN screen), the model converges; the blue estimates in the equation window turn green, indicating that they have converged.

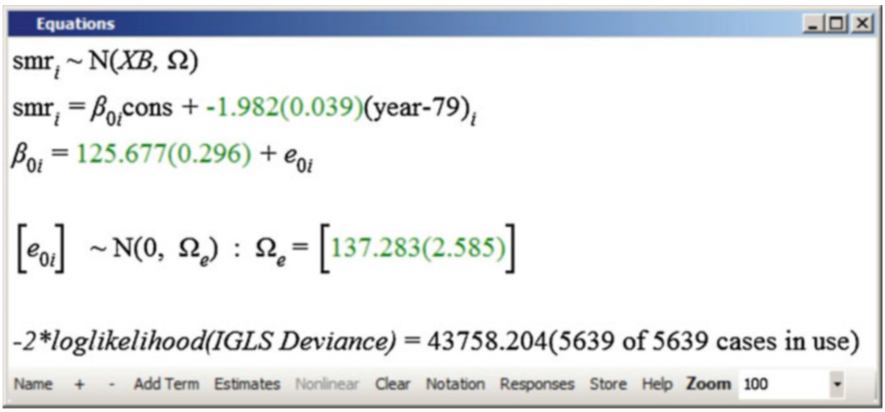


The parameter estimates are shown with their estimated standard errors in brackets. Our intercept is about 282, and the SMR has been decreasing at 1.982 per year. This decrease is highly significant in comparison with its standard error; a 95% confidence interval around this decrease would range from $1.982 \pm 1.96 * 0.039 = (1.906, 2.058)$. The variance of all of the observations around this fitted trend is 137. This means that the standard deviation is $\sqrt{137.283} = 11.717$ and so 95% of observations lie within ± 22.965 of the mean in any given year. 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∗*loglikelihood* is provided as an aid to model comparison and 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). We can use the **Equations** window to change this to a new variable which takes the value 0 in 1979 and 13 in 1982.

In the **Equation** window, click on the term ***year$_i$*** in the equation
Select **Modify term** in the X variable window
In the **centring** section of the **Specify term** window, check **around value** and type **79** in the corresponding box
Click **Done**

You will notice in the **Equations** window that the term ***year$_i$*** has been replaced by (***year*** − 79)$_i$. As the model has changed the estimates have changed from green to blue, indicating that we need to re-estimate this model. Note also that the new variable appears in column 11 in the **Names** window. Rather than click on the **Start** button again to estimate this model, click on the **More** button in the top left-hand corner of the MLwiN screen to continue estimation from the current values.



The estimated slope has not changed and nor has the variance. There is, however, a big change in the intercept. In 1979, the average SMR was therefore about 126.

We can store the results of successive models to allow easy comparison.

Click on the **Store** button at the bottom of the **Equations** window
Enter a suitable name in the box in the **Model name** window, e.g. Trend 1-level
Click **OK**

This section has covered data manipulation using:
Generate vector window—creating a constant
Generate vector window—creating a sequence
Name window—naming variables
This section has also covered model set-up using:
Equations window—defining the response (dependent variable)
Equations window—adding an intercept (CONS)
Equations window—adding an explanatory (independent) variable
Equations window—modelling random error at level 1
Estimating a model—the **Start** and **More** buttons
Equations window—modifying a term in the regression model
Centring the data to assist model interpretation
Equations window—storing results

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 including the level of DISTRICT in the model, enabling us to partition the variance into that which is attributable to random variation between DISTRICTs and that which arises due to fluctuations between observations (YEARs) within DISTRICTs.

A 2-Level Variance Components Model

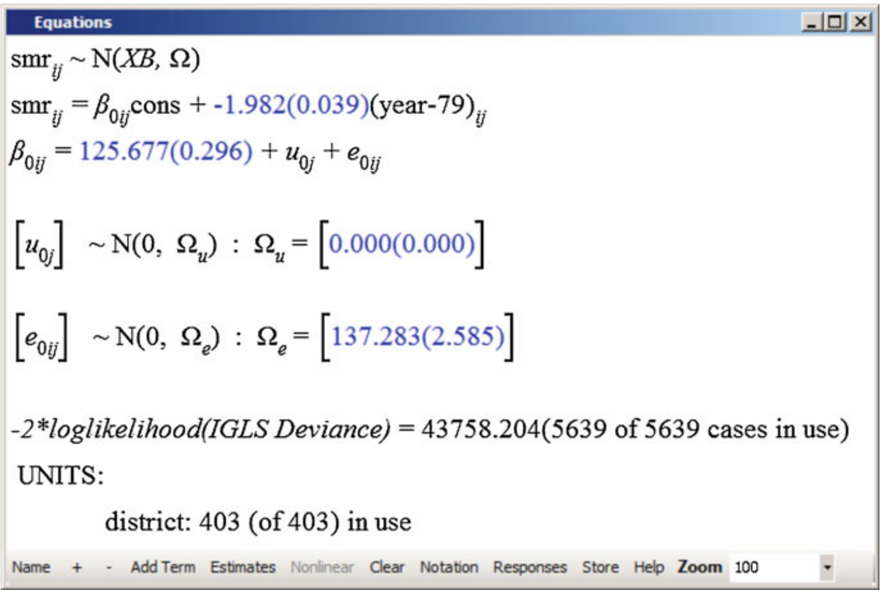
In the **Equations** window, we want to specify that our model has two levels, identified by DISTRICT (at level 2) and ID (at level 1).

Click on either ***smr**$_i$* term
Change **N levels** to ***2 – ij***
Select DISTRICT from the drop-down list by **level 2(j)**
Click **Done**

We now need to fit a random intercept across DISTRICTs. We do this by allowing the coefficient of the CONStant to vary randomly across DISTRICTs as well as at level 1.

Click on $\boldsymbol{\beta_{0i}}$
Check the box by **j(DISTRICT)**
Click **Done**

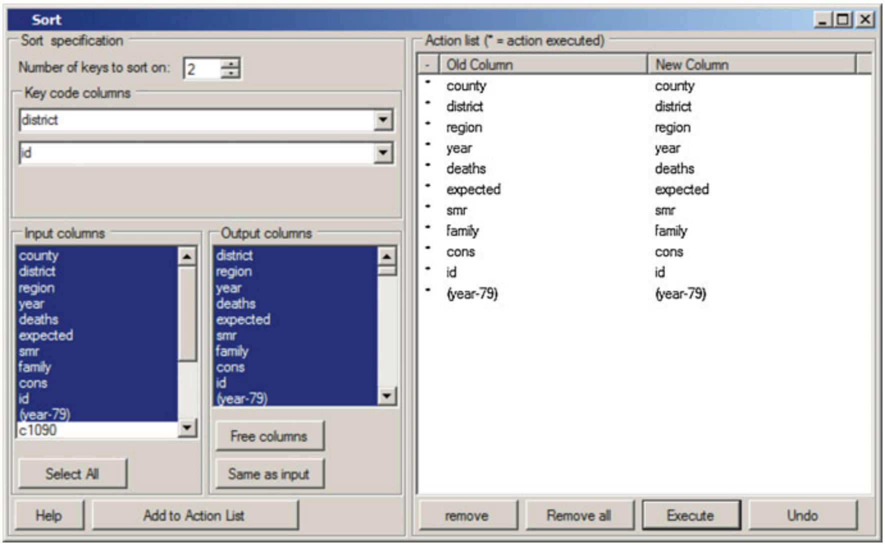
The intercept term now has an additional subscript (j), indicating that it varies across DISTRICTs as well as across YEARs. The intercept now has three parts: the overall fixed part intercept for 1979, the error term $\boldsymbol{e_{0ij}}$ and a term $\boldsymbol{u_{0j}}$ which is specific to DISTRICT j. The $\boldsymbol{u_{0j}}$ are random effects at level 2 and are assumed to be normally distributed. The intercept for the jth district in 1979 will be given by $\boldsymbol{\beta_0} + \boldsymbol{u_{0j}}$. The parameter estimates have again changed from green to blue, indicating that the model has changed and must be estimated again.



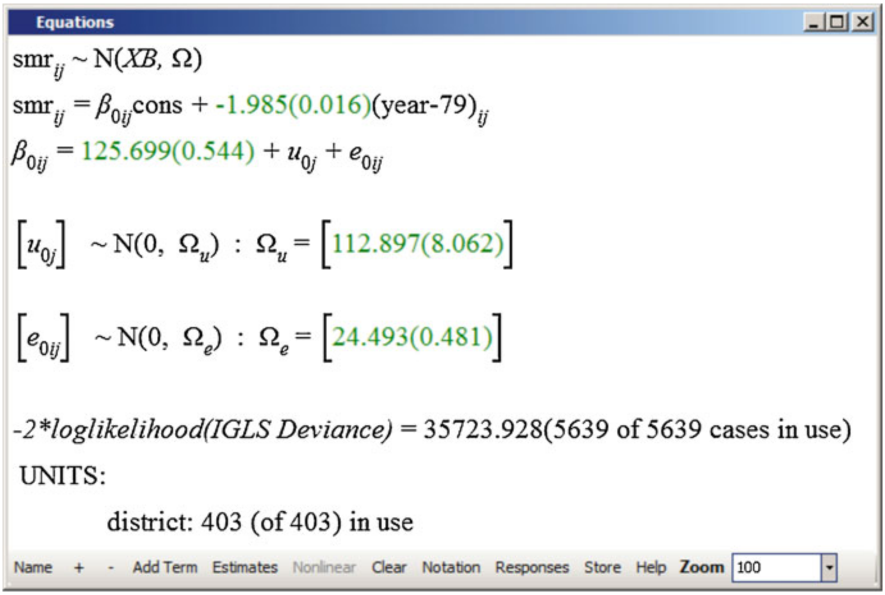
Sorting the Data

Before fitting a multilevel model, the data need to be sorted within their hierarchy (in this example by DISTRICTs and then by ID within DISTRICT). If your data are not sorted, then MLwiN will produce estimates but these will not be correct. *Failure to sort your data when using MLwiN is a common reason for getting 'strange' results!*

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 ID as the second
Select all named variables, from COUNTY to (YEAR-79), 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**



This model may now be fitted by clicking **More**.



If we store these results, we can then make a comparison with the OLS estimates previously obtained.

Click on the **Store** button at the bottom of the **Equations** window
Enter a suitable name in the box in the **Model name** window, e.g. Trend 2-level
Click **OK**
Go to the **Model** menu
Select **Compare stored models**

Results Table

Copy

	Trend 1-level	S.E.	Trend 2-level	S.E.
Response	smr		smr	
Fixed Part				
cons	125.677	0.296	125.699	0.544
(year-79)	-1.982	0.039	-1.985	0.016
Random Part				
Level: cons				
Level: id				
Var(cons)	137.283	2.585	24.493	0.481
Level: district				
Var(cons)			112.897	8.062
Units: cons	1			
Units: id	5639		5639	
Units: district			403	
Estimation:	IGLS		IGLS	
-2*loglikelihood:	43758.204		35723.928	

There is little change in the estimates of the intercept and slope between the two models. However, in the random part most of the variation has moved up to level 2, indicating that there is substantial variation between DISTRICTs rather than year-on-year variation within DISTRICTs. The total variance in our second model is obtained by summing the variances between and within DISTRICTs ($\boldsymbol{\sigma}^2_{u0} + \boldsymbol{\sigma}^2_{e0}$); the label 'CONS/CONS' in the first column indicates that these terms are the variances of the intercepts (remembering that we created and used the variable CONS to model the variance at each level). The total variance in our 'Trend 2-level' model is 137.390, very close to the estimate of 137.283 obtained for $\boldsymbol{\sigma}^2_{e0}$ in our 'Trend 1-level' model. The proportion of the total variance which arises due to differences between DISTRICTs is 112.897/(112.897 + 24.493) or 82.2%. This figure is known as the *intra-unit* or *intraclass 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.

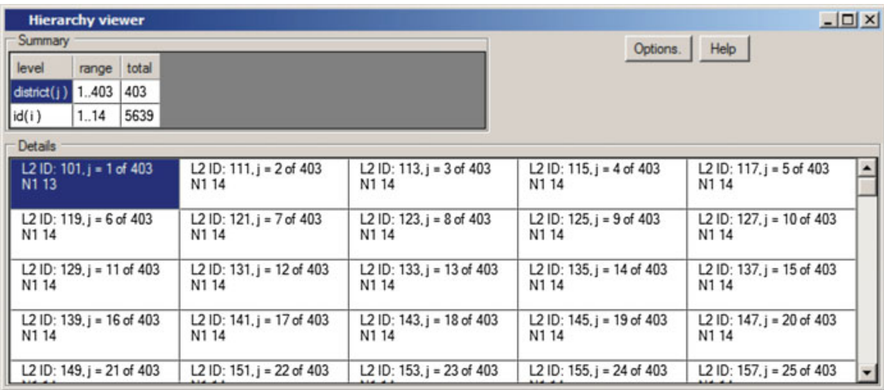
Note that the addition of a single variance term has produced a substantial reduction in the value of $-2{*}\log(\text{likelihood})$ from 43758 to 35724. This reduction has been brought about by the addition of a single parameter—the variance σ_{u0}^2—to our single-level model. Changes in the value of −2∗log(likelihood) for nested models (that is, for models that differ only by having terms added) are assessed using a chi-squared distribution with the number of degrees of freedom equivalent to the number of additional parameters. Since the reduction in −2∗log(likelihood) is of the order of 8000, we can dispense with the formal hypothesis testing (which will be covered later in this chapter) and conclude that the full model—that including the level of DISTRICT—is a significant improvement on the single-level model.

Predictions and confidence envelopes
At this stage, you may wish to look at the section Predictions and confidence envelopes at the end of this chapter. This compares the precision of estimates from the 2-level model with those from a single-level model. The work is in a section at the end of this chapter because it covers predictions, a subject which is given more attention later in this tutorial. You may read through this section or work through the example, in which case you will be prompted to save your worksheet at the current point. Alternatively, you can save the worksheet now as, for example, lmdpapp1.wsz and return to this section later.

The Hierarchy Viewer

We can view the data structure using the **Hierarchy viewer**. This will tell us how many lower-level units are in each high level unit.

Go to **Model** menu
Select **Hierarchy Viewer**

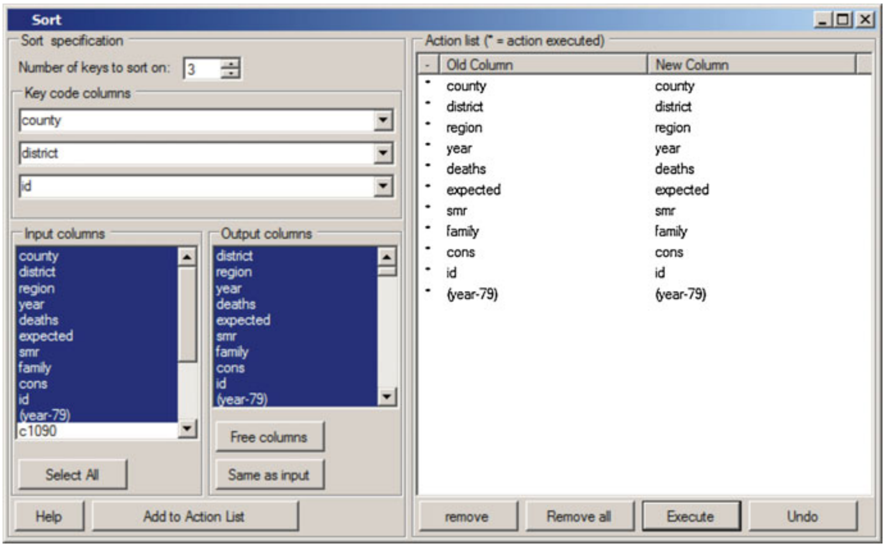


The box in the top left corner provides a summary of the data hierarchy: there are 403 DISTRICTs at level 2 and up to 14 observations at level 1 (defined by ID) within each DISTRICT, with a total of 5639 observations. Every level 2 unit (DISTRICT) has a box in the grid in the main part of the **Hierarchy viewer** screen; the first level 2 unit has identifying code 101 and has 13 level 1 units (observations). The second DISTRICT has identifying code 111 and so on. (These identifying codes are those found in the column DISTRICT.) The **Hierarchy viewer** is a useful tool to check that your data structure is correctly specified; failure to sort the data, for example, may lead to a data structure containing too many high level units.

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 sort the data again according to this new hierarchy of COUNTY then DISTRICT then ID.

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 ID 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**

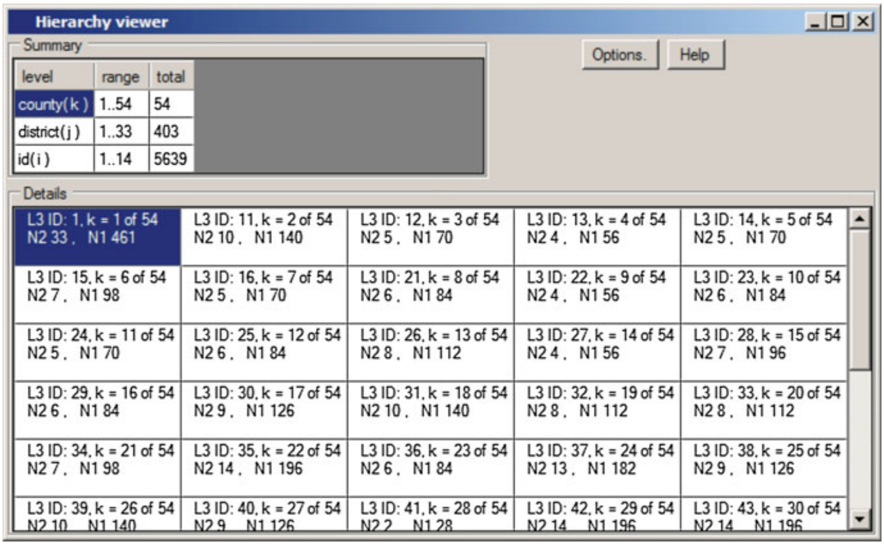


Now return to the **Equations** window. We are going to add COUNTY in at level 3 and make the coefficient of CONS, the intercept, random across COUNTY (as well as DISTRICT and ID).

> Click on either smr_{ij} term
> 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**

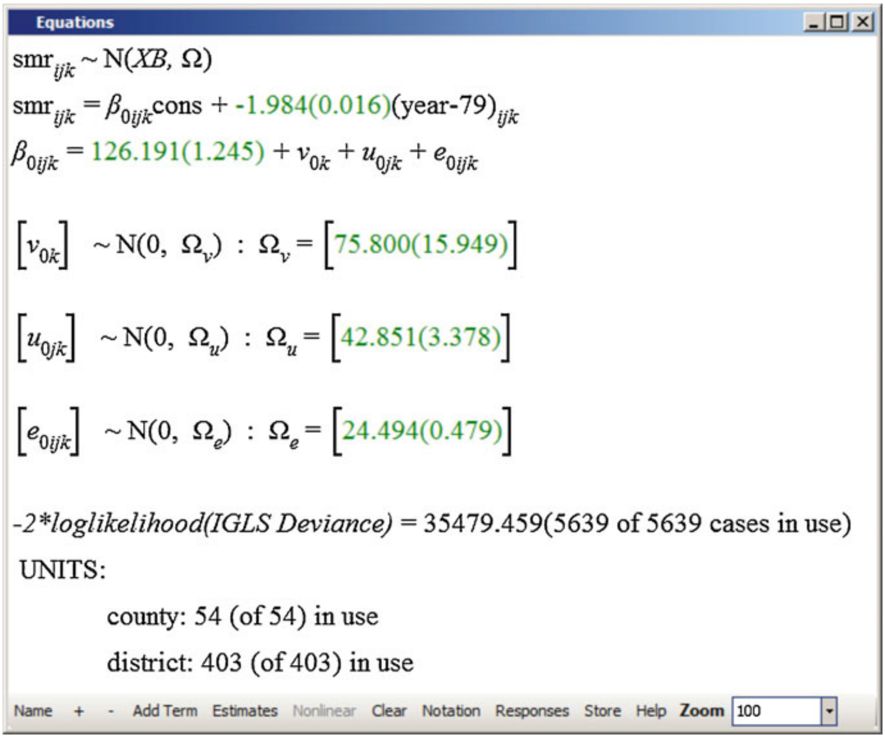
The intercept term now has an additional subscript to indicate that it varies across COUNTY as well as across DISTRICTs and ID. The terms v_{0k} are level 3 random effects and are again assumed to arise from a normal distribution. We can check the data structure using the **Hierarchy viewer**:

> Go to **Model** menu
> Select **Hierarchy Viewer**



We still see 5639 observations and 403 DISTRICTs, but these are nested in 54 COUNTYs. The first COUNTY has 33 DISTRICTs and a total of 461 observations at level 1.

Click **More** to estimate the new model



Click on the **Store** button at the bottom of the **Equations** window
Enter a suitable name in the box in the **Model name** window, e.g. Trend 3-level
Click **OK**
Go to the **Model** menu
Select **Compare stored models**

Results Table

Copy

	Trend 1-level	S.E.	Trend 2-level	S.E.	Trend 3-level	S.E.
Response	smr		smr		smr	
Fixed Part						
cons	125.677	0.296	125.699	0.544	126.191	1.245
(year-79)	-1.982	0.039	-1.985	0.016	-1.984	0.016
Random Part						
Level: cons						
Level: id						
Var(cons)	137.283	2.585	24.493	0.481	24.494	0.479
Level: district						
Var(cons)			112.897	8.062	42.851	3.378
Level: county						
Var(cons)					75.800	15.949
Units: cons	1					
Units: id	5639		5639		5639	
Units: district			403		403	
Units: county					54	
Estimation:	IGLS		IGLS		IGLS	
-2*loglikelihood:	43758.204		35723.928		35479.459	

The fixed part is more or less 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% (75.800/[75.800 + 42.851 + 24.494]) of the total variation can be seen to be between COUNTYs with 30% between DISTRICTs and just 17% due to year-on-year fluctuations.

This section has covered multilevel model set-up using:
Equations window—adding additional levels
Equations window—random intercepts (CONS) at different levels
Sort window—sorting the data by the hierarchy
Hierarchy viewer window—viewing the data structure
Results table window—comparing a series of models

Interpreting the Model

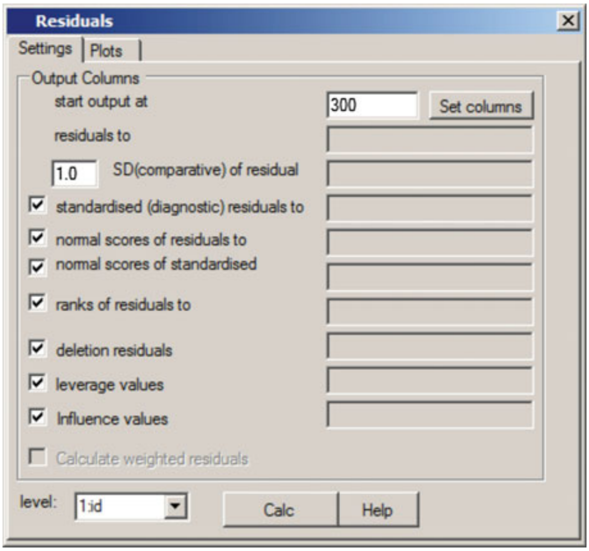
Residuals

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

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

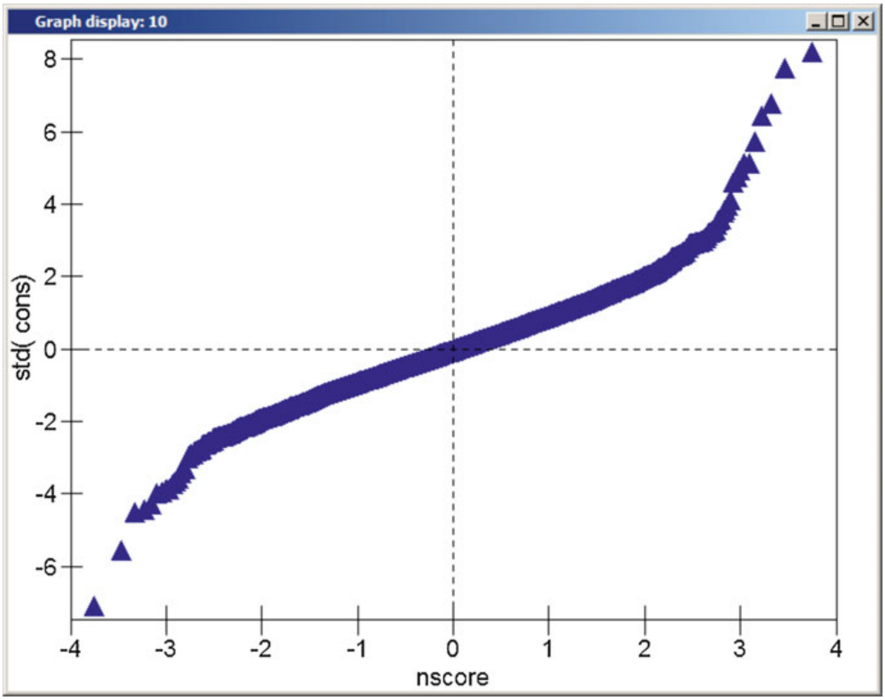
The terms inside the first set of brackets comprise the fixed part of the model, i.e. the fitted values for all data points. The terms inside the second set of brackets comprise the random part of the model and describe the departures from the fitted values at each level of the hierarchy. Thus, the difference between the observed and fitted values is comprised of residuals at three levels—the $\boldsymbol{v_{0k}}$, $\boldsymbol{u_{0jk}}$ and $\boldsymbol{e_{0ijk}}$ in the regression equation. (Remember that $\boldsymbol{x_0}$ 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 using similar residual diagnostics as those that would be appropriate if using OLS. First we will consider the residuals at level 1.

Go to the **Models** menu
Select **Residuals**



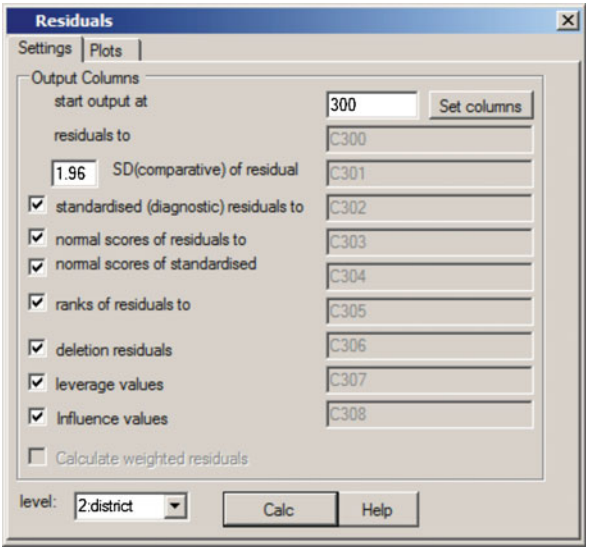
There is a variety of options which allow a range of standard diagnostic checks to be carried out—for example, to check the normality of the data or to look for outliers. By default all nine functions are calculated and the results are stored in columns c300–c308; this can be changed by entering a different number in the box by **start output at**. The drop-down box in the bottom left corner specifies the level at which the residuals are calculated; the default is level 1. We will calculate the residuals at level 1—the e_{0ijk}—and plot the standardised residuals against their normal scores.

In the **Residuals** window, 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**

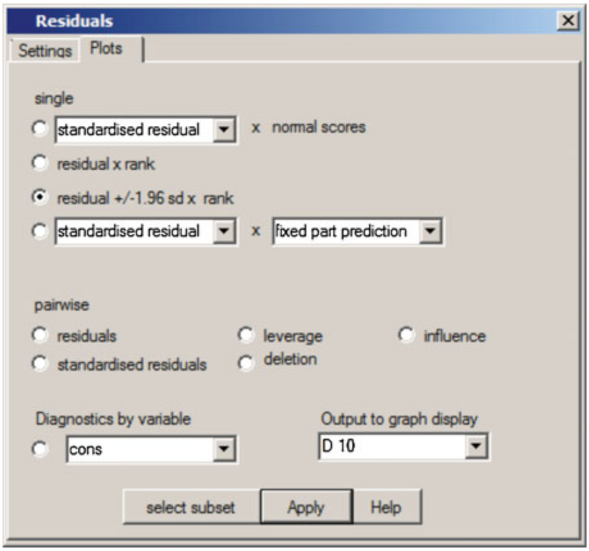


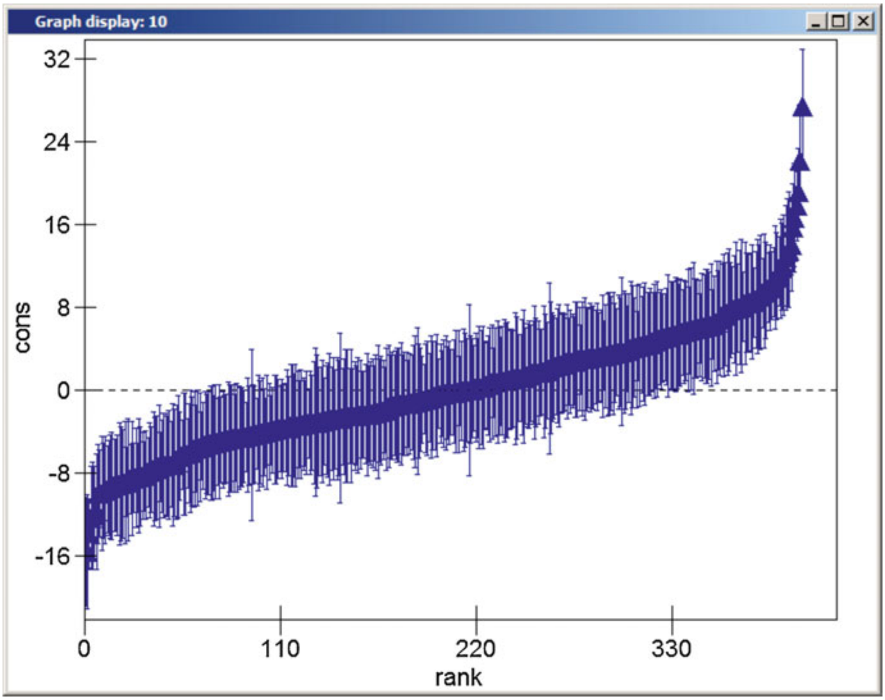
The points in the resulting graph should lie on a straight line; the fact that they do not suggests that there is some departure from normality. For the moment we will ignore this and look at the residuals at level 2 (DISTRICT), calculating these and 1.96 times their standard deviation (so that we can examine 95% confidence intervals).

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**

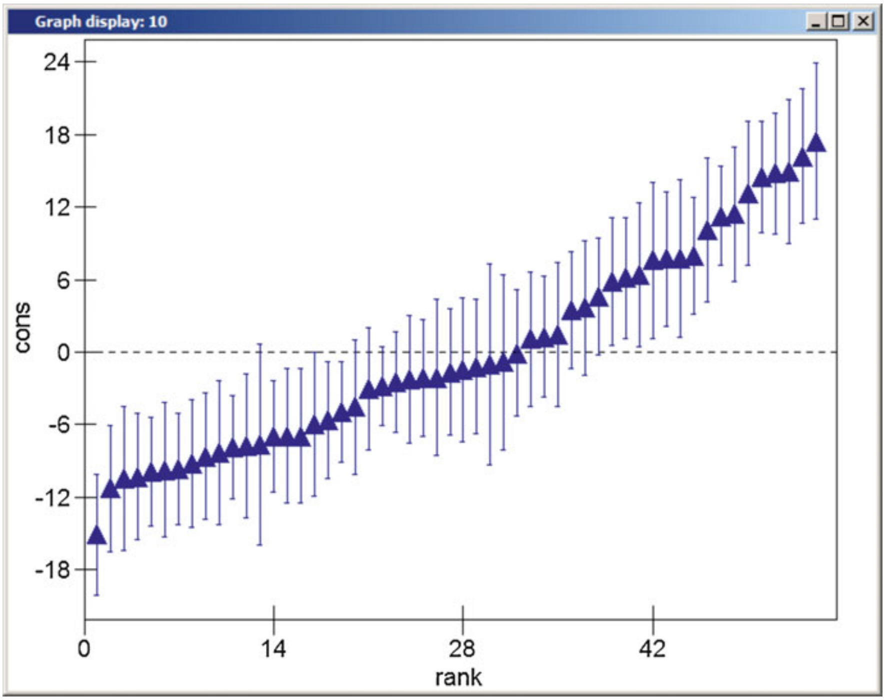

Residuals
Settings
Plots
single
standardised residual
x normal scores
residual x rank
residual +/-1.96 sd x rank
standardised residual
x
fixed part prediction
pairwise
residuals
leverage
influence
standardised residuals
deletion
Diagnostics by variable
cons
Output to graph display
D 10
select subset
Apply
Help


Graph display: 10
32
24
16
8
0
-8
-16
cons
0
110
220
330
rank

This plot shows the residuals or random effects for each of the 403 DISTRICTs, 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 27 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. (We could add the residual for each DISTRICT to that of the appropriate COUNTY and plot these composite residuals $v_{0k} + u_{0jk}$.) 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 to have SMRs which are significantly different from 100; a COUNTY with a positive residual may contain DISTRICTs with negative residuals because the components of the composite random part—$\boldsymbol{u}_{0jk}$ and $\boldsymbol{v}_{0k}$—are assumed to be independent.

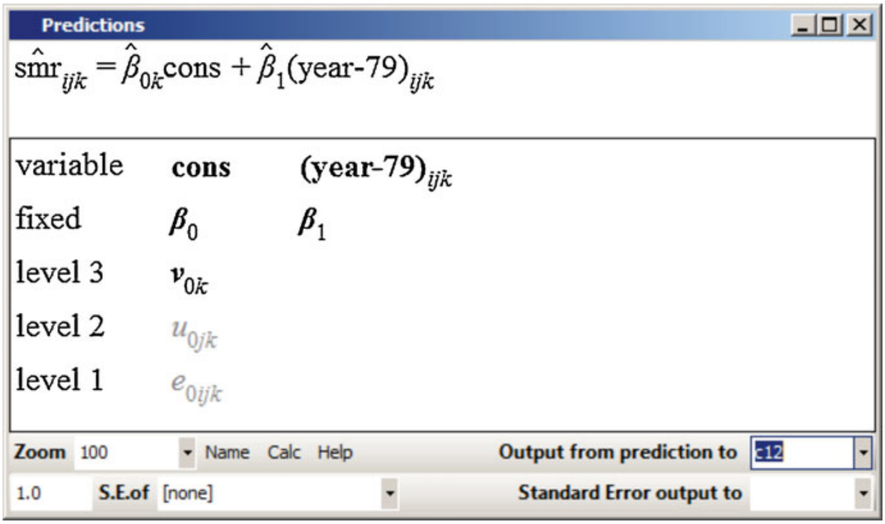
Predictions Window

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 COUNTY level (the $\boldsymbol{v}_{0k}$).

Go to the **Model** menu
Select **Predictions**

The elements of the model are arranged in two columns in the bottom half of the **Predictions** window, one for each explanatory variable. Initially, all the terms are in grey indicating that none has been selected and that they 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 (***cons*** or $(\boldsymbol{year} - 79)_{ijk}$) selects all the terms in that column (turning them black), whilst clicking on individual terms (such as $\boldsymbol{\beta}_0$ or $\boldsymbol{v}_{0k}$) toggles that term in or out of the prediction equation. To make predictions for the 54 COUNTYs at level 3, we need to include the fixed part and the level 3 residuals.

Click on ***cons*** and $(\boldsymbol{year} - 79)_{ijk}$
Click on $\boldsymbol{u}_{0jk}$ and $\boldsymbol{e}_{0ijk}$ to remove these terms from the prediction
In the drop-down list by **output from prediction to** select C12
Click on **Calc**

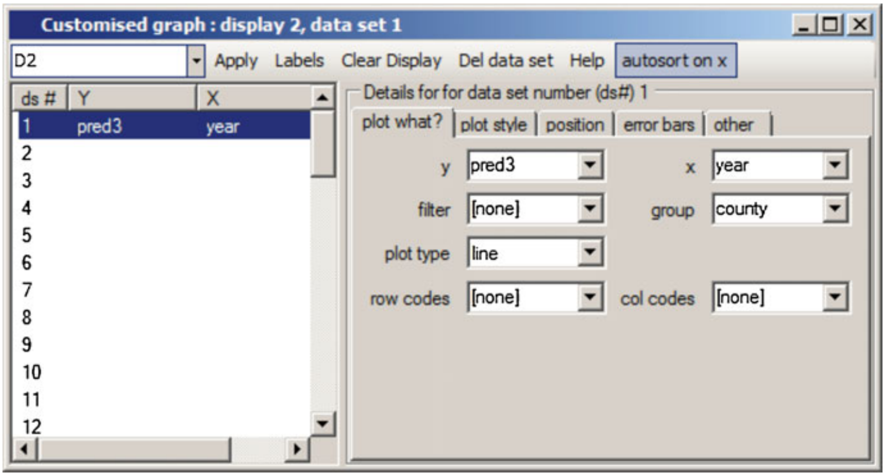


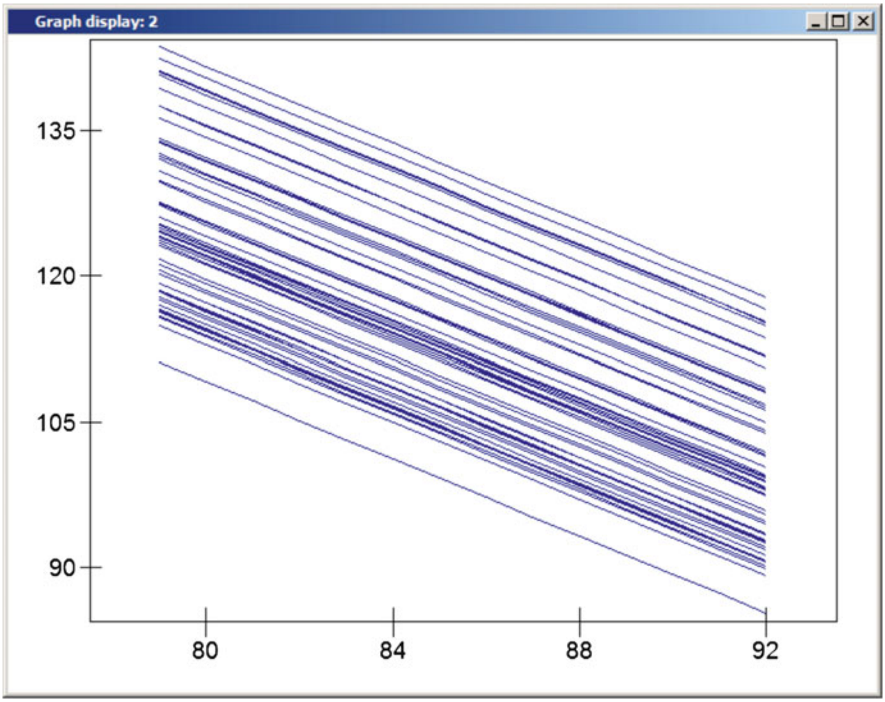
The results from this prediction are now in C12. (You may need to click on the **Refresh** button in the **Window** section at the top right-hand corner of the **Names** window to see the values that have been put in this column.) The COUNTY level predictions range from 85.3 to 143.6. Use the **Names** window to name this variable PRED3 to indicate that it is a prediction including the level 3 (COUNTY) random effects. Then plot the predicted values for each COUNTY against YEAR.

In the **Names** window, click on C12
Click on **Name** in the Column section at the top of the **Names** window
Type PRED3 and press <return>
Go to the **Graph** menu
Select **Customised Graph**

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

Select **D2** from the drop-down box in the top left-hand corner
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


Customised graph : display 2, data set 1
D2
Apply
Labels
Clear Display
Del data set
Help
autosort on x
ds #
Y
X
1 pred3 year
2
3
4
5
6
7
8
9
10
11
12
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
row codes [none]
col codes [none]


Graph display: 2
135
120
105
90
80
84
88
92

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

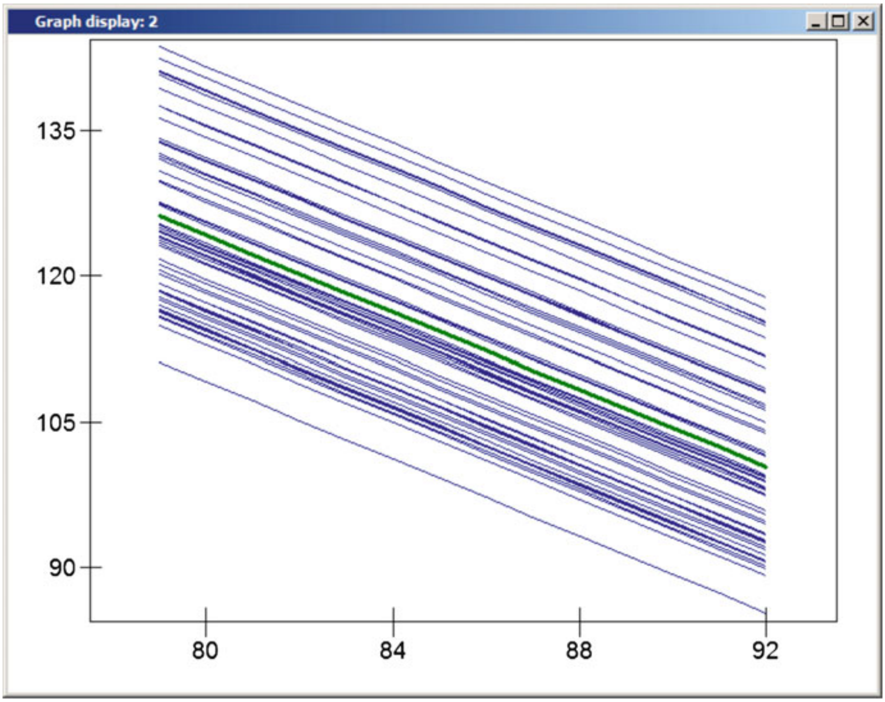
$$\hat{y}_{ijk} = \hat{\beta}_0 x_0 + \hat{\beta}_1 x_{1ijk}$$

This means that we only wish to include the fixed part of the model—all of the residual terms in the equation window should be grey.

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 C13
Click on **Calc**

In the **Names** window, change the name of C13 to PREDFP to indicate that it is a prediction from the fixed part only. We will plot the predicted values from the fixed part as dataset number 2 in display 2, plotting the mean over the prediction for each COUNTY.

Open 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 **green**
Change the **line thickness** to **3**
Click the **Apply** button

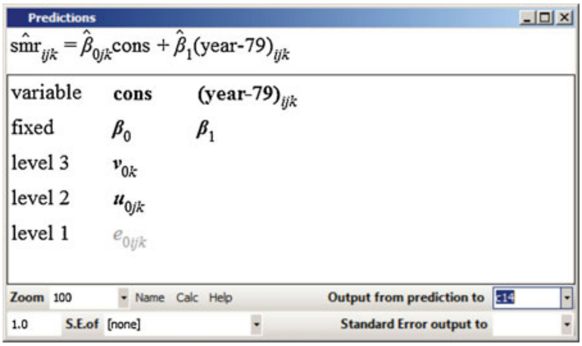


The national mean SMR is highlighted in green with the predicted mean for each COUNTY shown around it. The lines are all parallel since the effect of each COUNTY, $\boldsymbol{v}_{0k}$, is assumed to be the same throughout the study period. This residual is the horizontal distance between the national intercept and the COUNTY-specific intercept; a positive value of $\boldsymbol{v}_{0k}$ 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 by including all terms apart from the level 1 residuals $\boldsymbol{e}_{0ijk}$:

$$\widehat{\boldsymbol{y}}_{ijk} = \widehat{\boldsymbol{\beta}}_{0jk}\boldsymbol{x}_0 + \widehat{\boldsymbol{\beta}}_1\boldsymbol{x}_{1ijk}$$

Return to the **Predictions** window
Click on $\boldsymbol{v}_{0k}$ and $\boldsymbol{u}_{0jk}$ to add them to the prediction equation
In the drop-down list by **output from prediction to** select C14
Click on **Calc**



In the **Names** window, change the name C14 to PRED2 to indicate that these predicted values include the level 2 (DISTRICT) random effects.

We can look at these three sets of predictions using the **View or edit data** window.

Go to the **Data Manipulation** menu
Select **View or edit data**
Click on **view** to see a choice of variables
Select COUNTY, DISTRICT, YEAR, SMR, PRED3, PREDFP and PRED2 (multiple columns can be selected using the Control key)
Click on **OK**

Data

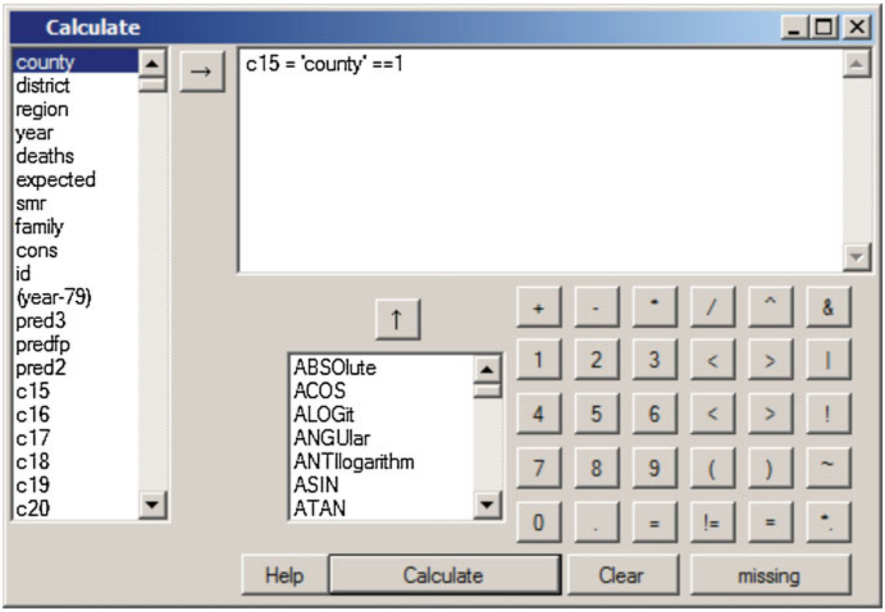
goto line 1 view Show value labels Font Help

	county(5639)	district(5639)	year(5639)	smr(5639)	pred3(5639)	predfp(5639)	pred2(5639)
1	1.000	101.000	79.000	138.611	123.371	126.191	119.130
2	1.000	101.000	81.000	99.435	119.402	122.222	115.161
3	1.000	101.000	82.000	140.506	117.418	120.238	113.177
4	1.000	101.000	83.000	111.780	115.434	118.253	111.192
5	1.000	101.000	84.000	104.642	113.449	116.269	109.208
6	1.000	101.000	85.000	105.592	111.465	114.284	107.223
7	1.000	101.000	86.000	83.803	109.480	112.300	105.239
8	1.000	101.000	87.000	101.706	107.496	110.315	103.254
9	1.000	101.000	88.000	74.712	105.511	108.331	101.270
10	1.000	101.000	89.000	82.275	103.527	106.347	99.285
11	1.000	101.000	90.000	106.569	101.542	104.362	97.301
12	1.000	101.000	91.000	132.193	99.558	102.378	95.316
13	1.000	101.000	92.000	85.840	97.573	100.393	93.332
14	1.000	111.000	79.000	140.824	123.371	126.191	133.459
15	1.000	111.000	80.000	137.215	121.387	124.207	131.474
16	1.000	111.000	81.000	132.639	119.402	122.222	129.490
17	1.000	111.000	82.000	128.713	117.418	120.238	127.506
18	1.000	111.000	83.000	127.070	115.434	118.253	125.521
19	1.000	111.000	84.000	116.233	113.449	116.269	123.537
20	1.000	111.000	85.000	124.453	111.465	114.284	121.552

The variable PREDFP contains just the values from the fixed part of the model—the intercept and slope. These values change across YEAR—the slope—but are constant (in the same YEAR) across DISTRICTs and COUNTYs. PRED3 contains the predicted mean for each COUNTY and although they vary from one YEAR to another they are the same for all DISTRICTs in the same COUNTY. PRED2 contains predictions for each DISTRICT within each COUNTY. The slope is constant across time and does not vary between DISTRICTs or COUNTYs; for any of our three predictions, the difference between the predictions in neighbouring years is 1.984 (the coefficient of YEAR in our current **Equations** window).

To illustrate the different prediction lines in a single chart, select a single COUNTY, e.g. COUNTY number 1. (You can use the **Hierarchy viewer** to see which COUNTY codes exist; for example, there is no COUNTY with code between 2 and 10 inclusive.) To create an indicator for COUNTY number 1, for example, we use the logical function == (two equals signs) meaning 'is equal to' in the **Calculate** window:

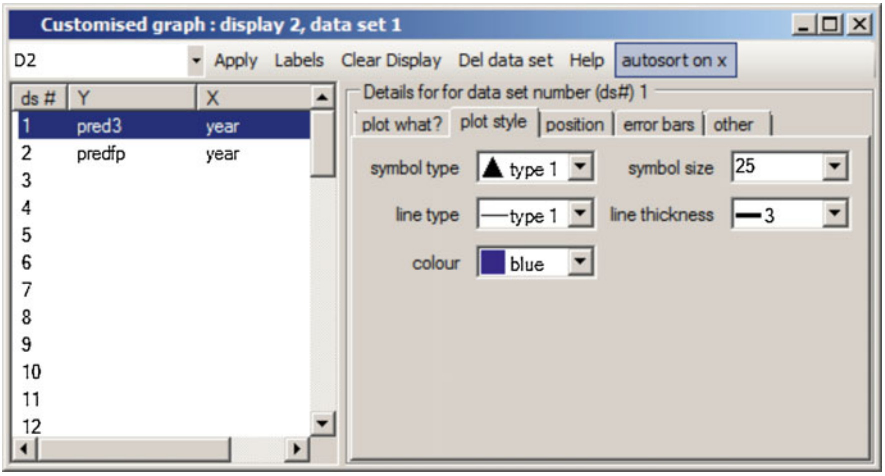
Go to **Data manipulation** menu
Select **Calculate**
Select the empty column C15 from the list of variables and press the right arrow button near the top of the **Calculate** window
Click on the **=** button on the window's keypad
Select COUNTY from the list of variables and press the right arrow button
Use the window's keypad to enter **==1**
Press **Calculate**



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 C15 to COUNTY1. Then, in the **Customised Graph** window we can filter out all COUNTYs apart from the one that we have chosen.

Return to the **Customised Graph** window
Ensure **D2** is selected
Highlight data set number **1** under **ds #**
Select the **filter** to be COUNTY1 under the **plot what?** tab
Click on the **plot style** tab
Change the **line thickness** to **3**
Click the **Apply** button

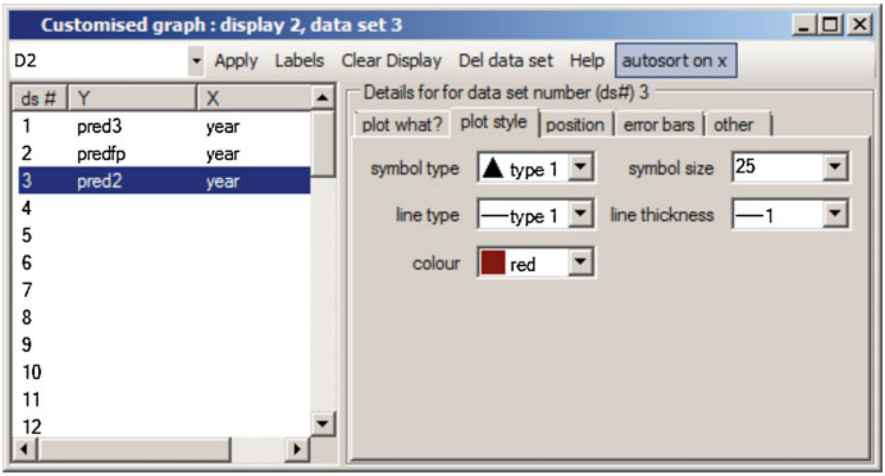


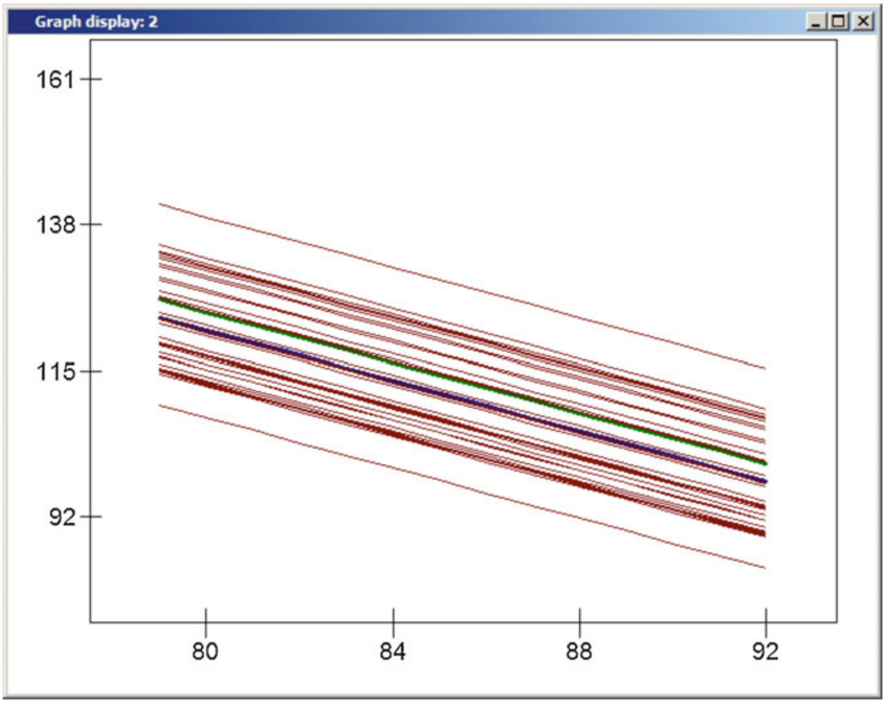
The resulting graph now has just two lines—one for the national mean and one for the selected COUNTY. To plot the predicted lines for the DISTRICTS in COUNTY number 1, we again need to use the filter; we can plot the DISTRICT predictions in red.

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

(continued)

Select **group** to be DISTRICT
Select the **plot type** to be **line**
Click the **plot style** tab
Change the **colour** to **red**
Click the **Apply** button





In addition to the national mean (green) and COUNTY mean (blue), the graph now displays the DISTRICT predictions for the selected COUNTY. The vertical distance between the green and blue lines is the level 3 (COUNTY) residual $\boldsymbol{v}_{01}$ (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 $\boldsymbol{u}_{0j1}$. The vertical distance between each DISTRICT mean and the national mean is then the composite residual $\boldsymbol{v}_{01} + \boldsymbol{u}_{0j1}$. 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 $\boldsymbol{v}_{01} + \boldsymbol{u}_{0j1}$ is greater than zero.

This section has covered model diagnostics and interpretation using:
Residuals window—checking normality at level 1
Residuals window—higher-level residuals with confidence intervals
Predictions window—predictions from the fixed part of the model
Predictions window—predictions including residuals
Graph—plotting predicted values
Graph—overlaying (multiple) graphs
Calculate window—creating a new variable

Model Building

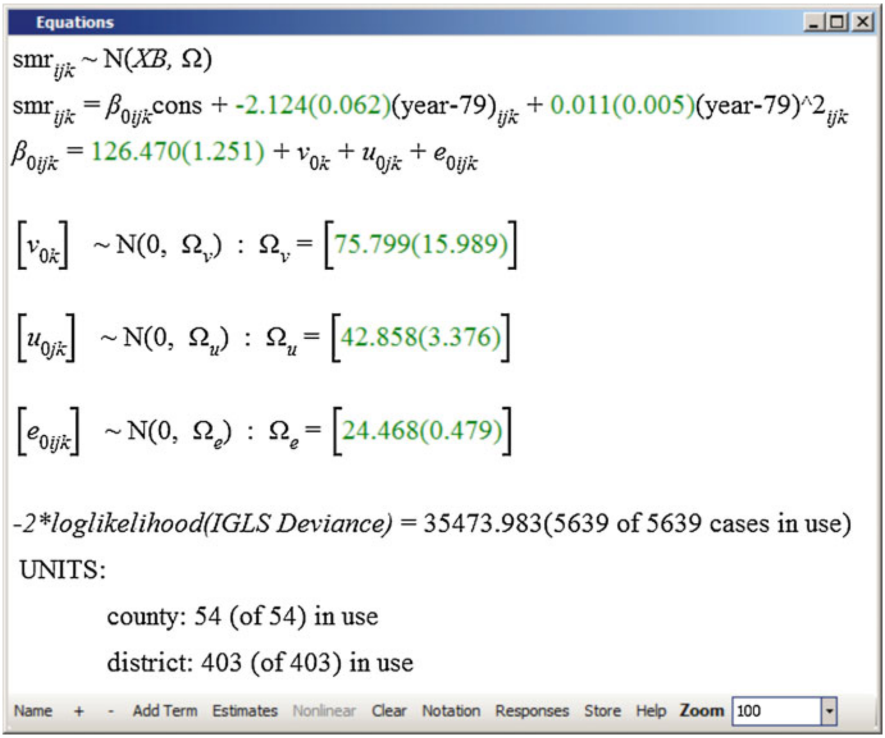
Adding More Fixed Effects

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, we add a quadratic term in year since the assumption of a linear trend may be too simplistic. We can use the ^ (to the power of) function in the **Calculate** window to raise our trend variable (YEAR-79) to the power of 2.

Go to **Data manipulation** menu
Select **Calculate**
Select the empty column C16 from the list of variables and press the right arrow button
Click on the **=** button on the window's keypad
Select (YEAR-79) 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 C16 to (YEAR-79)^2. We can add this term in the **Equations** window and re-estimate the model:

Return to the **Equations** window
Click on **Add Term**
Select (YEAR-79)^2 from the drop-down list under **variable** in the **Specify term** window
Click **Done**
Click on the **More** button to re-estimate the model



Click on the **Store** button at the bottom of the **Equations** window
Enter a suitable name in the box in the **Model name** window, e.g. M4
Click **OK**
Go to the **Model** menu
Select **Compare stored models**

Results Table

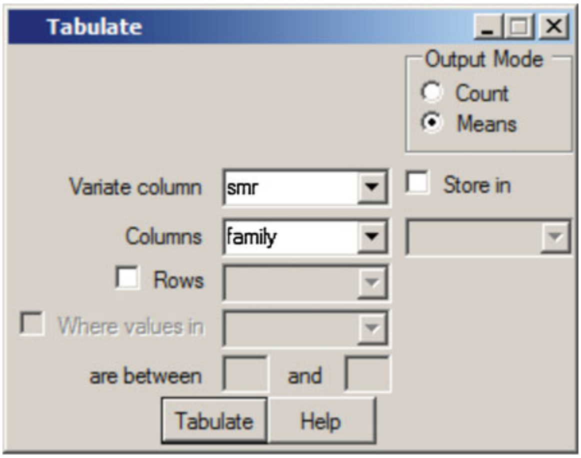
Copy

	Trend 1-level	S.E.	Trend 2-level	S.E.	Trend 3-level	S.E.	M4	S.E.
Response	smr		smr		smr		smr	
Fixed Part								
cons	125.677	0.296	125.699	0.544	126.191	1.245	126.470	1.251
(year-79)	-1.982	0.039	-1.985	0.016	-1.984	0.016	-2.124	0.062
(year-79)^2							0.011	0.005
Random Part								
Level: cons								
Level: id								
Var(cons)	137.283	2.585	24.493	0.481	24.494	0.479	24.468	0.479
Level: district								
Var(cons)			112.897	8.062	42.851	3.378	42.858	3.376
Level: county								
Var(cons)					75.800	15.949	75.799	15.989
Units: cons	1							
Units: id	5639		5639		5639		5639	
Units: district			403		403		403	
Units: county					54		54	
Estimation:	IGLS		IGLS		IGLS		IGLS	
-2*loglikelihood:	43758.204		35723.928		35479.459		35473.983	

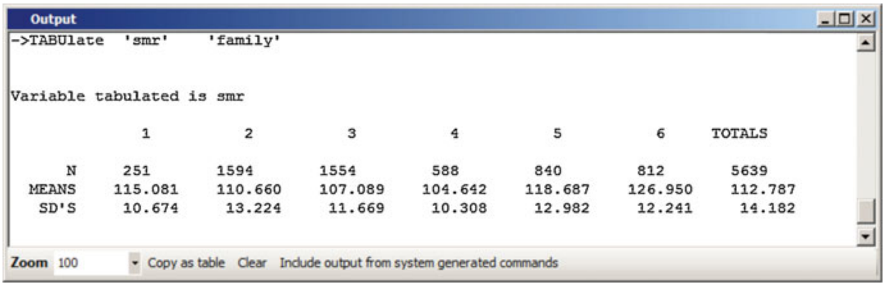
The reduction in −2∗log(likelihood) is 5.476 from 1 degree of freedom—comfortably greater than the critical value of 3.84—so this term has significantly improved the fit of the model. The addition of this term has, however, done nothing to reduce the variance at any of the three levels in the model.

The next covariate we can consider adding to the fixed part of the model is the variable FAMILY, a classification of the DISTRICTs into different types. Before adding this to the model, we can see how mean SMRs differ across the categories of family. We do this using the **Tabulate** window:

Go to **Basic statistics** menu
Select **Tabulate**
In the **Output mode** section at the top right of the **Tabulate** window, select **Means**
From the drop-down list next to **Variate column,** select SMR
From the drop-down list next to **Columns,** select FAMILY
Click **Tabulate**



The output window opens containing the table of means and SDs:



	1	2	3	4	5	6	TOTALS
N	251	1594	1554	588	840	812	5639
MEANS	115.081	110.660	107.089	104.642	118.687	126.950	112.787
SD'S	10.674	13.224	11.669	10.308	12.982	12.241	14.182

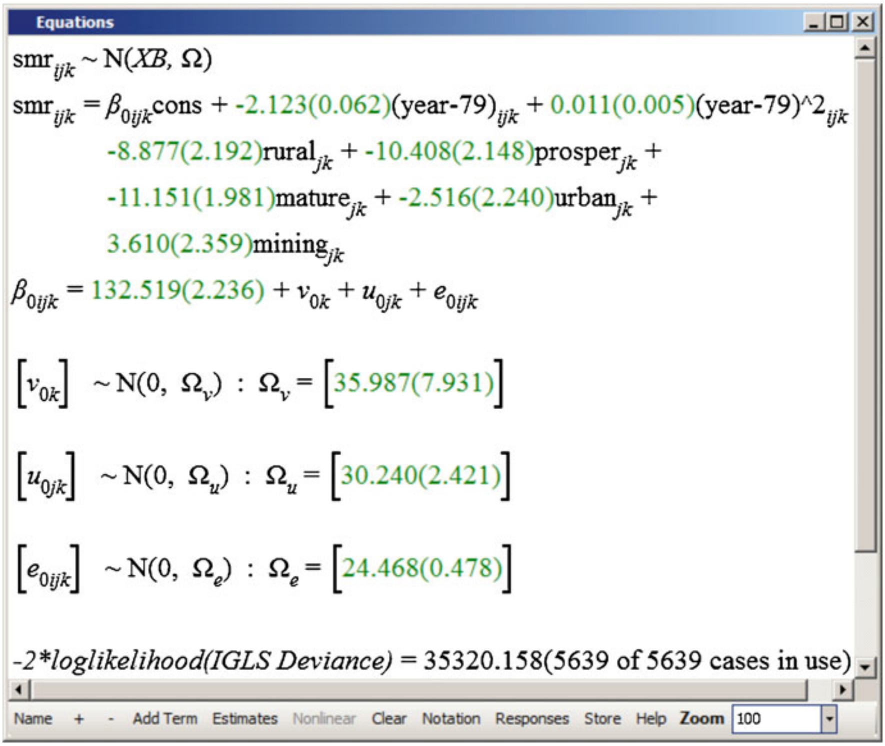
This shows lower mean SMRs in categories 3 and 4 (prospering and maturer areas) and higher SMRs in mining areas (category 6). To add a categorical variable such as this to our model, we first need to specify that it is categorical; we do this using the **Names** window.

In the **Names** window, click on the variable FAMILY
Click on the **Toggle Categorical** button in the **Column** section at the top of the **Names** window
Click on the **View** button in the **Categories** section
With family_1 highlighted, click **Edit** and type LONDON
Highlight family_2, click **Edit** and type RURAL
Highlight family_3, click **Edit** and type PROSPER
Highlight family_4, click **Edit** and type MATURE
Highlight family_5, click **Edit** and type URBAN
Highlight family_6, click **Edit** and type MINING
Click on **OK**

We can now add the variable FAMILY to the model. As with any categorical variable, we fit one fewer dummy variable than the number of categories; for this reason, we need a reference category against which all comparisons will be made. We will use LONDON as the reference category.

In the **Equations** window, click on the **Add term** button
Select FAMILY from the drop-down list under **variable** in the **Specify term** window
Check that LONDON is selected as the **Reference category**
Click **Done**
Click on the **More** button to re-estimate the model

We have created five dummy variables named RURAL, PROSPER, etc., which take the value 1 for a DISTRICT if it is of that type, 0 otherwise. These variables can be seen in the **Names** window in columns 17–21.



In the **Equations** window, note that the dummy variables representing the categories of FAMILY have subscripts **jk** as opposed to the variables (YEAR-79) and (YEAR-79)^2 which have subscripts **ijk**. This is because the FAMILY variable is measured at the DISTRICT level—it remains constant for each DISTRICT from one year to another.

Click on the **Store** button at the bottom of the **Equations** window
Enter a suitable name in the box in the **Model name** window, e.g. M5
Click **OK**
Go to the **Model** menu
Select **Compare stored models**

Results Table

Copy

	Trend 1-level	S.E.	Trend 2-level	S.E.	Trend 3-level	S.E.	M4	S.E.	M5	S.E.
Fixed Part										
cons	125.677	0.296	125.699	0.544	126.191	1.245	126.470	1.251	132.519	2.236
(year-79)	-1.982	0.039	-1.985	0.016	-1.984	0.016	-2.124	0.062	-2.123	0.062
(year-79)^2							0.011	0.005	0.011	0.005
rural									-8.877	2.192
prosper									-10.408	2.148
mature									-11.151	1.981
urban									-2.516	2.240
mining									3.610	2.359
Random Part										
Level: cons										
Level: id										
Var(cons)	137.283	2.585	24.493	0.481	24.494	0.479	24.468	0.479	24.468	0.478
Level: district										
Var(cons)			112.897	8.062	42.851	3.378	42.858	3.376	30.240	2.421
Level: county										
Var(cons)					75.800	15.949	75.799	15.989	35.987	7.931
Units: cons	1									
Units: id	5639		5639		5639		5639		5639	
Units: district			403		403		403		403	
Units: county					54		54		54	
Estimation:	IGLS		IGLS		IGLS		IGLS		IGLS	
-2*loglikelihood:	43758.204		35723.928		35479.459		35473.983		35320.158	

The intercept or coefficient of the CONS term has changed as this is now the estimated mean in 1979 for areas in Inner London (the reference category). There has been a significant reduction in −2∗log(likelihood) 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.)

We can calculate the explained variance ($\boldsymbol{R}^2 = 1 - \widehat{\boldsymbol{\sigma}}^2 / \boldsymbol{s}_y^2$) at any time by making a comparison of the variance in our current model, $\widehat{\boldsymbol{\sigma}}^2$, and the variance in the original data, $\boldsymbol{s}_y^2$. (See, for example, Gelman and Hill 2007.) From M5 in the table above, we have $\widehat{\boldsymbol{\sigma}}^2 = 90.695$. To obtain $\boldsymbol{s}_y^2$ we could refit the first model—labelled 'Trend 1-level' above—excluding the trend variable (YEAR-79) from the fixed part. Alternatively, we can use the **Averages and Correlation** window to obtain the SD of the dependent variable SMR.

Go to the **Basic Statistics** menu
Select **Averages and Correlations**
Ensure that **Averages** is selected in the **Operation** section
Select SMR from the drop-down list
Click **Calculate**

In the output window, we can see that the variable SMR has a mean of 112.79 and a standard deviation of 14.182, giving a variance of 201.129. So the $\boldsymbol{R}^2$ for M5 is 0.550.

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.

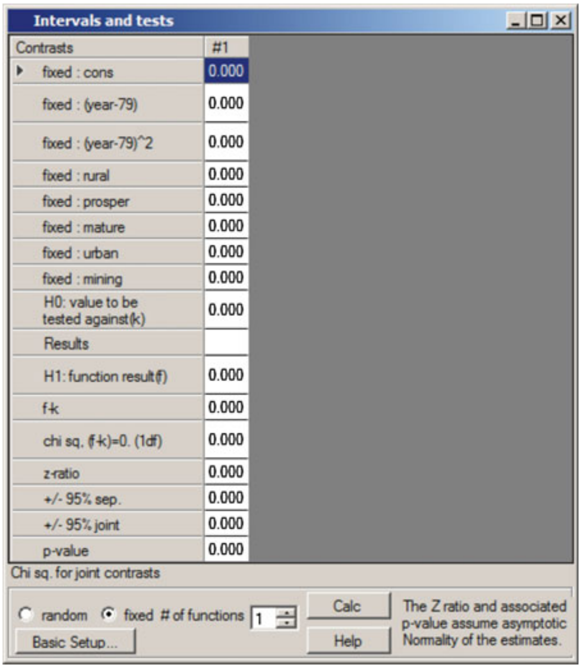
$$\textit{Hypothesis}\,1: \quad \boldsymbol{\beta}_6 = 0$$

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.

$$\textit{Hypothesis } 2: \quad \boldsymbol{\beta}_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5 \textit{ or equivalently}$$
$$(\boldsymbol{\beta}_3 - \boldsymbol{\beta}_4 = 0) \textit{ and } (\boldsymbol{\beta}_3 - \boldsymbol{\beta}_5 = 0) \textit{ implying } (\boldsymbol{\beta}_4 - \boldsymbol{\beta}_5 = 0)$$

The **Intervals and tests** window gives us a choice of testing contrasts among the fixed or random parameters; in this case, we want to test the fixed parameters.

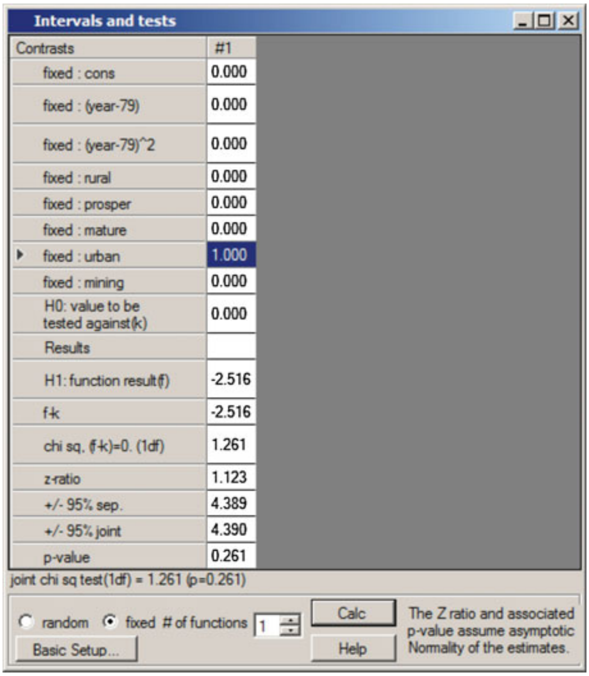
Go to the **Model** menu
Select **Intervals and tests**
Select **fixed** at the bottom of the window



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 whilst two functions are required for hypothesis 2. The boxes beside each fixed parameter are used to enter the function of the parameters to be tested, whilst 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**

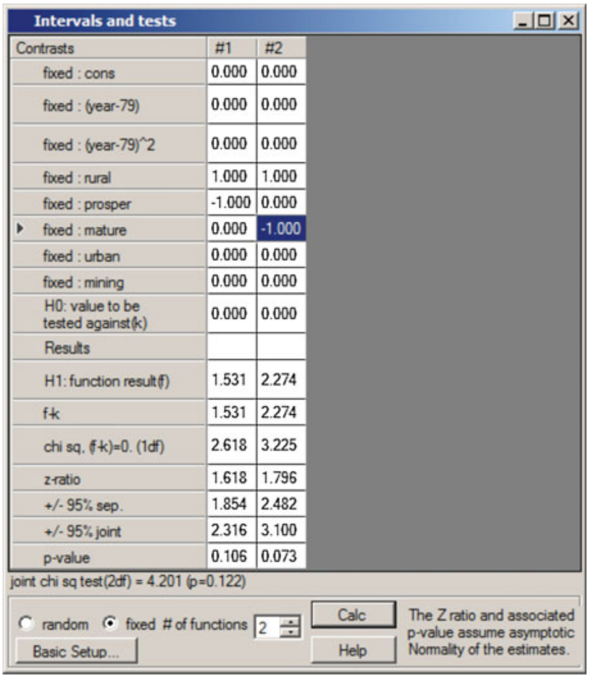
Note that the function **f** is a single multiple of the URBAN parameter and so equals $\boldsymbol{\beta}_6$, and because $k = 0$, $(f - k)$ also equals the parameter $\boldsymbol{\beta}_6$. The test statistic, based on Wald's Test, appears at the bottom of the window, **joint chi sq test(1df) = 1.261**, and this may be compared to a chi-squared distribution to either accept or reject the hypothesis that $\boldsymbol{\beta}_6 = 0$. In this instance we can see that the p-value of 0.261 is greater than the conventional threshold of 0.05 and, as such, 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), we need to set up the two tests corresponding to RURAL − PROSPER = 0 and RURAL – MATURE = 0. (The third test, corresponding to PROSPER – MATURE = 0, is implied by the other two tests.)

Ensure **fixed** is selected at the bottom of the **Intervals and tests** window
Change the **# of functions** to **2**
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: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 $(1 \times \boldsymbol{\beta}_3) - (1 \times \boldsymbol{\beta}_4) = 0$ (i.e. $\boldsymbol{\beta}_3 = \boldsymbol{\beta}_4$).



This time we are jointly testing two functions and therefore base the test on two degrees of freedom. The resulting p-value of 0.122 indicates that we cannot reject the hypothesis that the mean SMRs of categories RURAL, PROSPER and MATURE are the same.

In practice at this stage we might want to collapse the variable FAMILY into just three categories: a baseline category comprising Inner LONDON and URBAN areas 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 now with all six categories.

This section has covered model building using:
Equations window—adding an explanatory (independent) variable
Tabulate window—tabulating variable means across categories
Names window—declaring categories for a variable
Averages and correlation window—obtaining the mean and standard deviation of a variable
Intervals and tests window—testing hypotheses involving single and multiple parameters

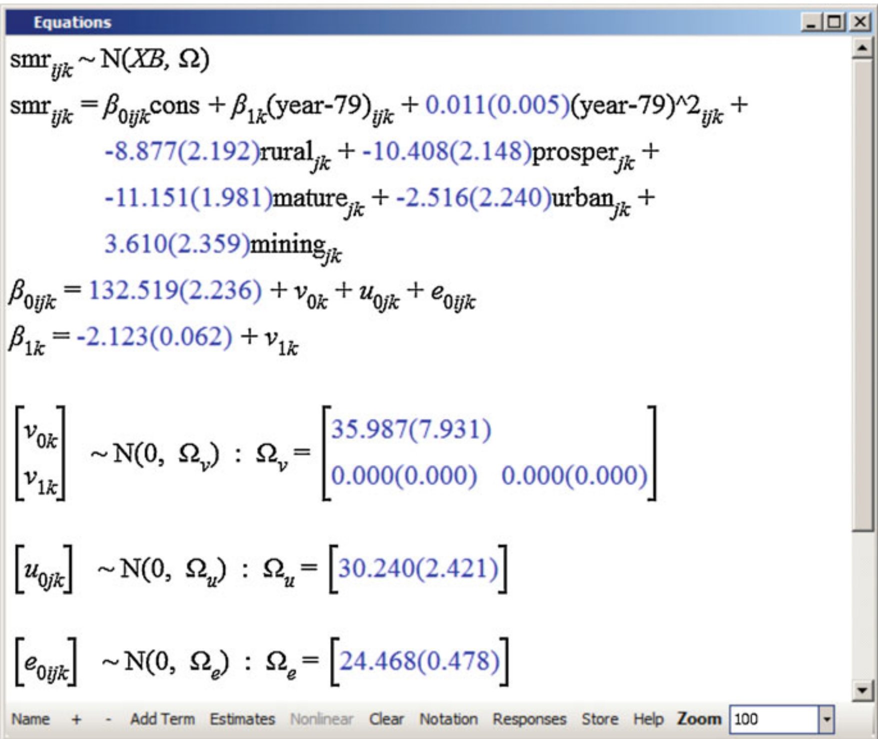
Random Coefficients

We now consider another important class of multilevel model: random coefficients (also known as random slopes). 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.

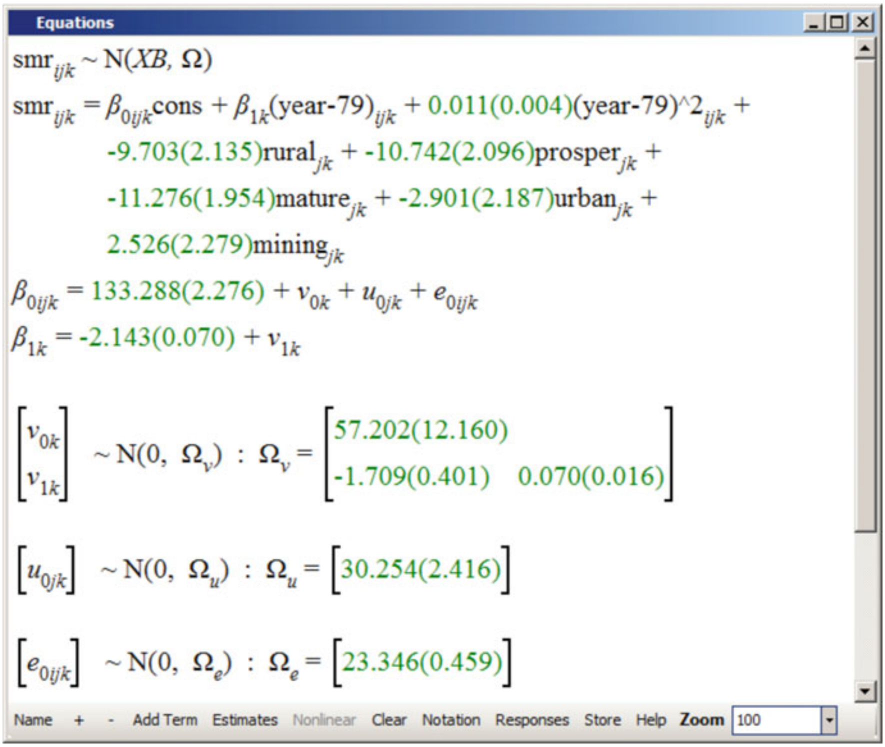
Random Slopes

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). This is achieved in the **Equations** window by specifying that we want the coefficient of (YEAR-79) to vary randomly across COUNTYs.

Return to the **Equations** window
Click on ***(year − 79)***$_{ijk}$ and check the box by **k(COUNTY)**
Then click **Done**



The coefficient of $(\boldsymbol{year} - \boldsymbol{79})_{\boldsymbol{ijk}}$ has changed from $\boldsymbol{\beta}_1$ to $\boldsymbol{\beta}_{1\boldsymbol{k}}$ indicating that this parameter now varies randomly across COUNTYs. The estimate of $\boldsymbol{\beta}_{1\boldsymbol{k}}$ is now given as a mean $\boldsymbol{\beta}_1$, common to all COUNTYs, plus a level 3 residual $\boldsymbol{v}_{1\boldsymbol{k}}$, unique to the kth COUNTY. The level 3 residuals $\boldsymbol{v}_{0\boldsymbol{k}}$ and $\boldsymbol{v}_{1\boldsymbol{k}}$ now have a joint multivariate normal distribution with variances $\boldsymbol{\sigma}^2_{\boldsymbol{v}0}$ and $\boldsymbol{\sigma}^2_{\boldsymbol{v}1}$ respectively and covariance $\boldsymbol{\sigma}_{\boldsymbol{v}01}$. Click on **More** to estimate this model.



Click on the **Store** button at the bottom of the **Equations** window
Enter a suitable name in the box in the **Model name** window, e.g. M6
Click **OK**
Go to the **Model** menu
Select **Compare stored models**

Results Table

Copy

	Trend 1-level	S.E.	Trend 2-level	S.E.	Trend 3-level	S.E.	M4	S.E.	M5	S.E.	M6	S.E.
cons	125.677	0.296	125.699	0.544	126.191	1.245	126.470	1.251	132.519	2.236	133.288	2.276
(year-79)	-1.982	0.039	-1.985	0.016	-1.984	0.016	-2.124	0.062	-2.123	0.062	-2.143	0.070
(year-79)^2							0.011	0.005	0.011	0.005	0.011	0.004
rural									-8.877	2.192	-9.703	2.135
prosper									-10.408	2.148	-10.742	2.096
mature									-11.151	1.981	-11.276	1.954
urban									-2.516	2.240	-2.901	2.187
mining									3.610	2.359	2.526	2.279
Random Part												
Level: cons												
Level: id												
Var(cons)	137.283	2.585	24.493	0.481	24.494	0.479	24.468	0.479	24.468	0.478	23.346	0.459
Level: district												
Var(cons)			112.897	8.062	42.851	3.378	42.858	3.376	30.240	2.421	30.254	2.416
Level: county												
Var(cons)					75.800	15.949	75.799	15.989	35.987	7.931	57.202	12.160
Covar((year-79)/cons)											-1.709	0.401
Var((year-79))											0.070	0.016
Units: cons	1											
Units: id	5639		5639		5639		5639		5639		5639	
Units: district			403		403		403		403		403	
Units: county					54		54		54		54	
Estimation:	IGLS		IGLS		IGLS		IGLS		IGLS		IGLS	
-2*loglikelihood:	43758.204		35723.928		35479.459		35473.983		35320.158		35138.778	

Note that should you want you can select a subset of stored models to compare by going to the **Manage stored models** window in the **Model** menu.

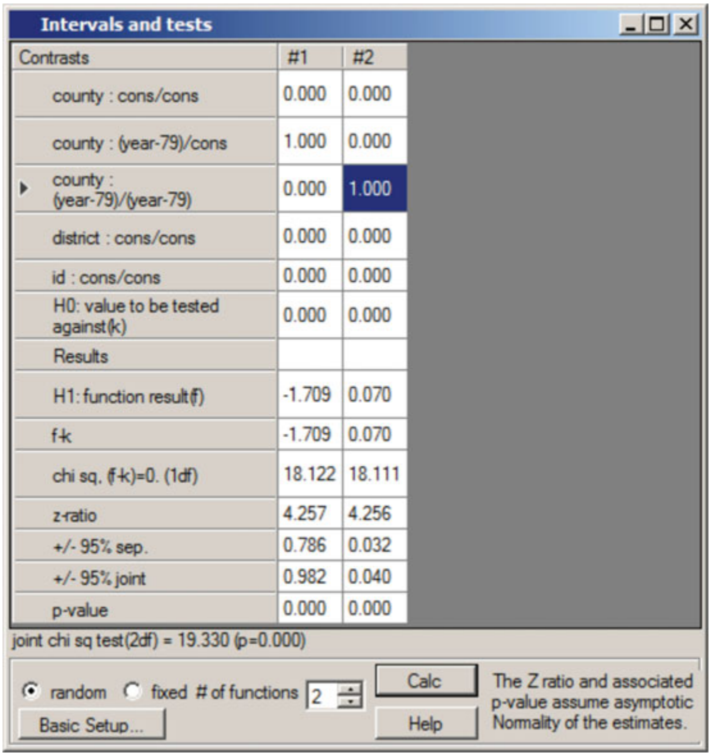
There is little change in the fixed part of the model, nor in the level 1 or level 2 variances. There has, however, been a large reduction in the value of $-2*\log$ (likelihood). Therefore, the addition of random slopes has improved the overall fit of the model. (If the covariance between the intercept and slope at the COUNTY level does not show up in the results table then go to **Manage stored models** in the **Model** menu, ensure that the box by covariance in the **Metric** section is checked, and click on the **Compare** button.) The three random terms at level 3 now refer to the variance of the intercept (CONS) for COUNTYs—$\boldsymbol{\sigma}_{\nu 0}^2$, the variance of the slope (YEAR79) for COUNTYs—$\boldsymbol{\sigma}_{\nu 1}^2$, and the covariance between the two, $\boldsymbol{\sigma}_{\nu 01}$. 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. This time we are testing contrasts on two random parameters.

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

$$\textit{Hypothesis } 3: \quad \boldsymbol{\sigma}_{v1}^2 = \boldsymbol{\sigma}_{v01} = 0 \;\; or$$
$$\boldsymbol{\sigma}_{v1}^2 = 0 \;\; and \;\; \boldsymbol{\sigma}_{v01} = 0$$

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



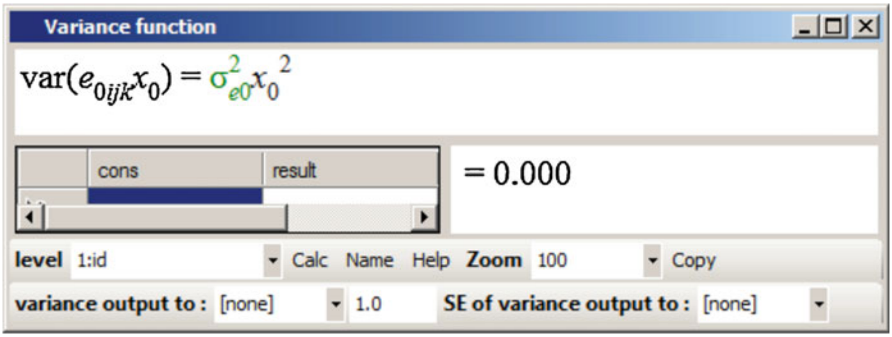
Contrasts	#1	#2
county : cons/cons	0.000	0.000
county : (year-79)/cons	1.000	0.000
county : (year-79)/(year-79)	0.000	1.000
district : cons/cons	0.000	0.000
id : cons/cons	0.000	0.000
H0: value to be tested against(k)	0.000	0.000
Results		
H1: function result(f)	-1.709	0.070
f-k	-1.709	0.070
chi sq. (f-k)=0. (1df)	18.122	18.111
z-ratio	4.257	4.256
+/- 95% sep.	0.786	0.032
+/- 95% joint	0.982	0.040
p-value	0.000	0.000

The value of 19.330 is highly significant when compared with a chi-squared distribution with two degrees of freedom ($p < 0.001$); we therefore reject the hypothesis that the two random terms are not significantly different from 0. In general, when testing the significance of random parameters (variances and covariances), using either the likelihood ratio test (comparing values of $-2*\log(\text{likelihood}))$ or the Wald test (using the **Intervals and tests** window), we need to halve the p-value. This is essentially because variances are non-negative and the alternative hypothesis is therefore one-sided. For a more detailed explanation of this issue, the reader is referred to Snijders and Bosker (2012).

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

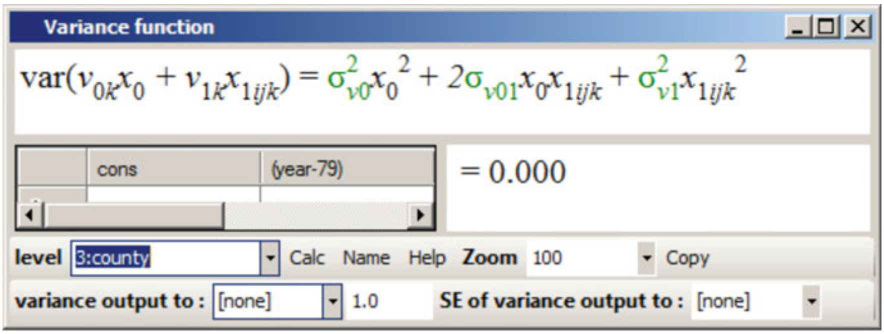
Variance Function Window

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



The purpose of this window is to display and calculate the variance function at any level of the current model. The variance function for level 1 is shown by default; this only involves one term because the current model assumes that the level 1 variance is constant for all observations. (Remember that x_0 is our CONStant and takes the value 1 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**



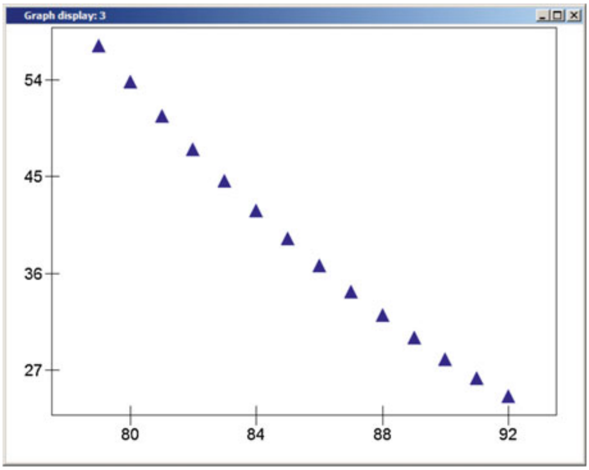
The current model has two terms random at level 3, the intercept and the slope, so the level 3 variance is a function of two random variables. The function shown is the variance of the sum of the two random terms $v_{0k}x_0$ and $v_{1k}x_{1ijk}$. Since x_0 is just the CONStant term, taking the value 1, the level 3 variance is a quadratic in x_{1ijk} (YEAR-79). We can use the **Variance function** window to calculate this function and use the **Graph** window to plot it. This will tell us how the variance between COUNTYs has been changing over time.

Note that the columns in the table in the **Variance function** window named **cons**, **(year-79)**, **result** and **result se** allow us to estimate the variance function at specific values of (YEAR-79). 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 C22
Click **calc**

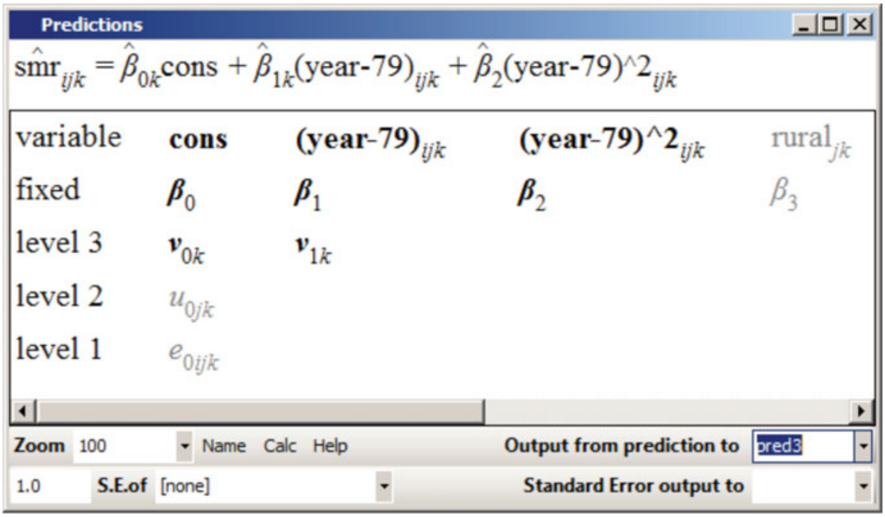
In the **Names** window name C22 VARF3. To plot the level-3 variance across the observed values of YEAR79:

From the **Graph** menu, select the **Customised Graph** window
Select a new display **D3**
Highlight **ds # 1**
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 **cons**, ***(year − 79)***$_{ijk}$ and ***(year − 79)^2***$_{ijk}$ to ensure that they are included
Click on $\boldsymbol{u}_{0jk}$ and $\boldsymbol{e}_{0ijk}$ to remove them from the prediction but ensure that $\boldsymbol{v}_{0k}$ and $\boldsymbol{v}_{1k}$ are included
Select PRED3 for **output from prediction to**
Click **Calc**



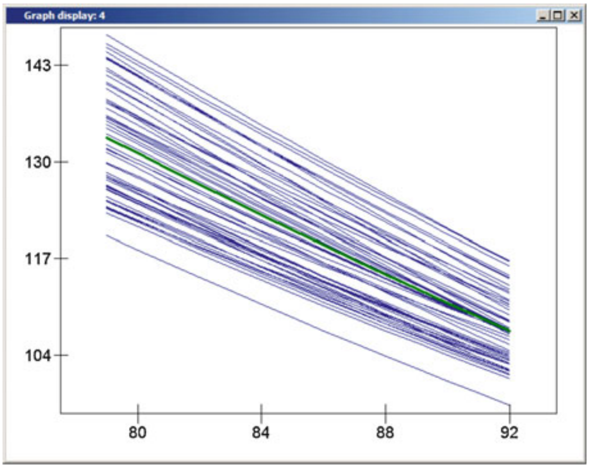
Next re-calculate the predicted values using the fixed part of the model only:

Click on $\boldsymbol{v}_{0k}$ and $\boldsymbol{v}_{1k}$ to remove these terms from the prediction
Select PREDFP for **output from prediction to**
Click **Calc**

We have ignored the categories of the FAMILY variable indicating the type of each DISTRICT. This is because we are only interested at the moment in seeing how mortality has changed over time in each COUNTY, and not how mortality varies according to FAMILY. (The inclusion of the categories of FAMILY would give us up to six lines for each COUNTY, corresponding to the different DISTRICT types within each COUNTY.) We can plot these new level 3 predictions using the

Customised graph window, overlaying the national mean in green on top of the COUNTY-specific slopes.

Return to the **Customised Graph** window
Select a new display **D4**
Highlight **ds # 1**
Select **y** to be PRED3
Select **x** to be YEAR
Select COUNTY as the **group**
Change **plot type** to **line**
Click **Apply**
Highlight **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 **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 towards the mean line over time demonstrates the decrease in level 3 variation over time.

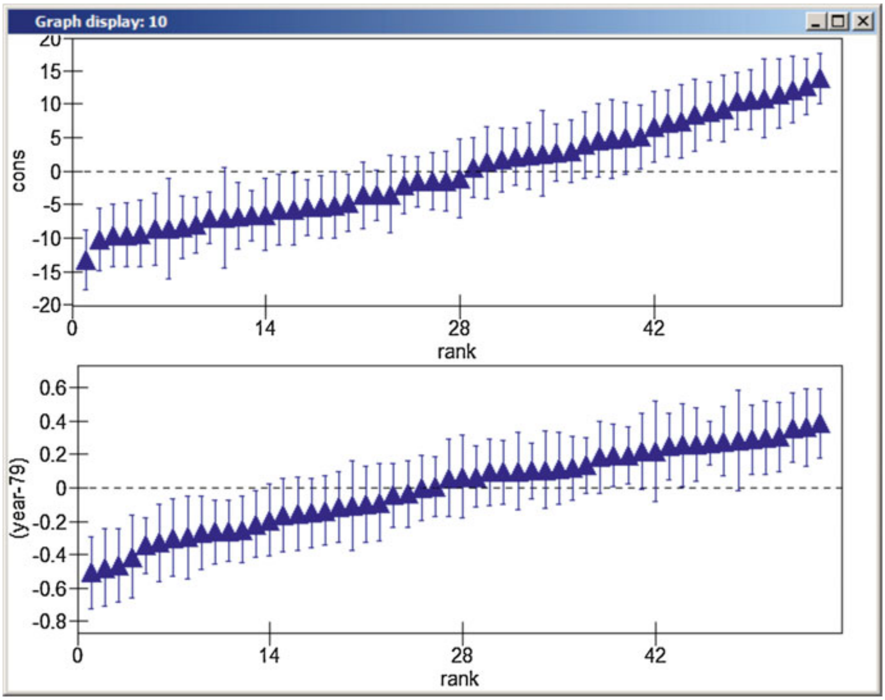
Higher-Level Residuals

There are now two sets of residuals at the COUNTY level; we can look at these using the **Residuals** window.

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

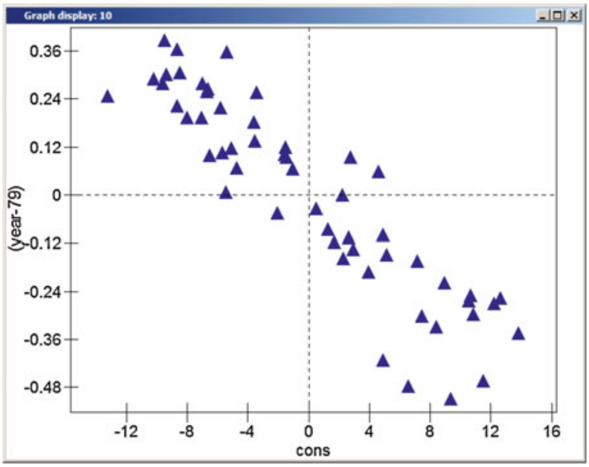
Each of the output items now requires two columns: the first column relates to the intercept CONS and the second to the slope YEAR79. For example, C300 will store the residual for CONS and C301 the residual for YEAR79. We can plot both sets of residuals, together with 95% confidence intervals:

> 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**



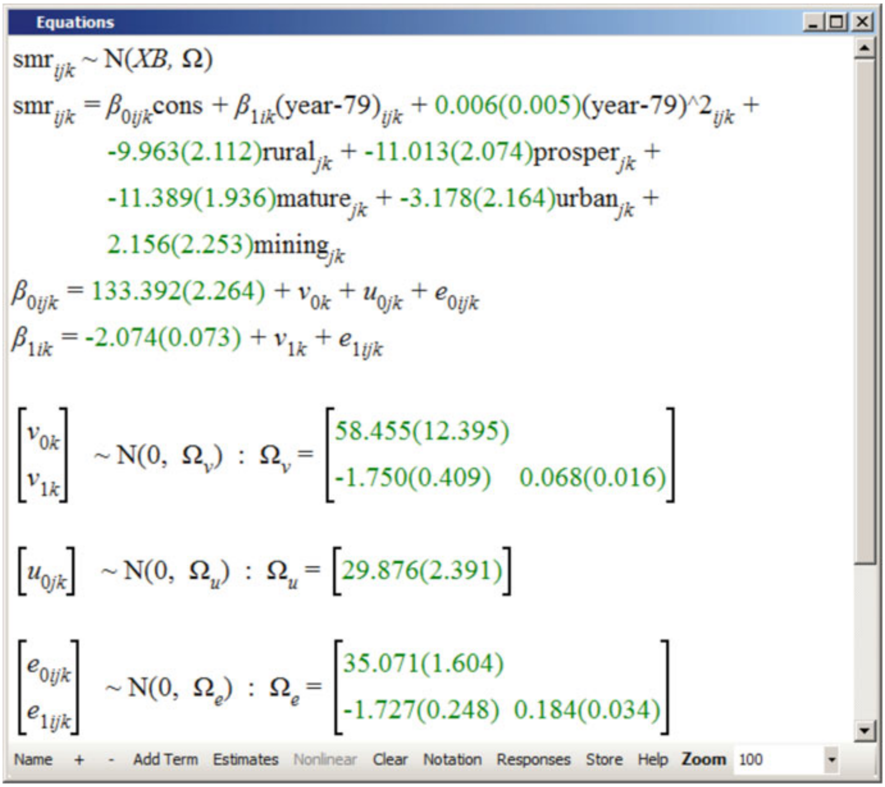
This shows the strong negative correlation between the two sets of residuals. Those in the top left quadrant refer to those COUNTYs with negative intercept (CONS) residuals and positive slope (YEAR-79) residuals. This suggests that those COUNTYs which had lower than average SMRs in 1979 experienced a 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 (YEAR-79) 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. We do this by making the coefficient of the variable YEAR79 random across observations (ID—our level 1 identifier).

Return to the **Equations** window
Click on $(\boldsymbol{year - 79})_{ijk}$
Check the box at **i(id)**
Click **Done**

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



Click on the **Store** button at the bottom of the **Equations** window
Enter a suitable name in the box in the **Model name** window, e.g. M7
Click **OK**
Go to the **Model** menu
Select **Compare stored models**

Results Table

Copy

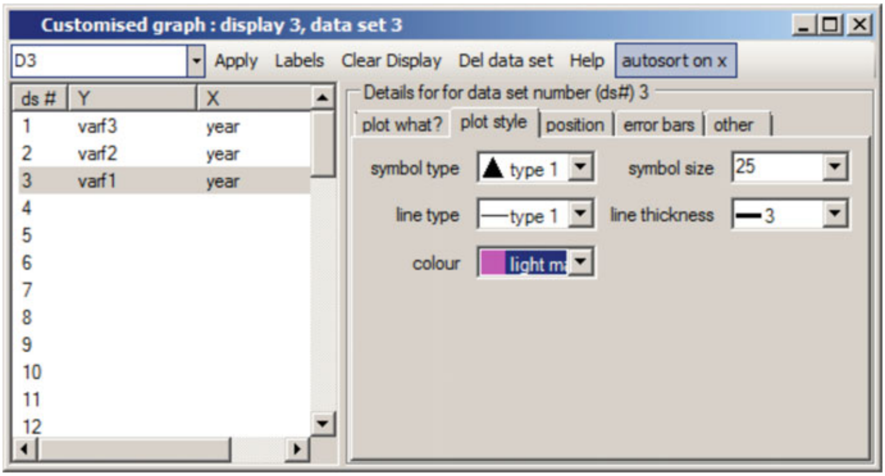
	M6	S.E.	M7	S.E.
Level: county				
Var(cons)	57.202	12.160	58.455	12.395
Covar((year-79)/cons)	-1.709	0.401	-1.750	0.409
Var((year-79))	0.070	0.016	0.068	0.016
Level: district				
Var(cons)	30.254	2.416	29.876	2.391
Level: id				
Var(cons)	23.346	0.459	35.071	1.604
Covar((year-79)/cons)			-1.727	0.248
Var((year-79))			0.184	0.034
Units: county	54		54	
Units: district	403		403	
Units: id	5639		5639	
Estimation:	IGLS		IGLS	
-2*loglikelihood:	35138.778		35032.992	

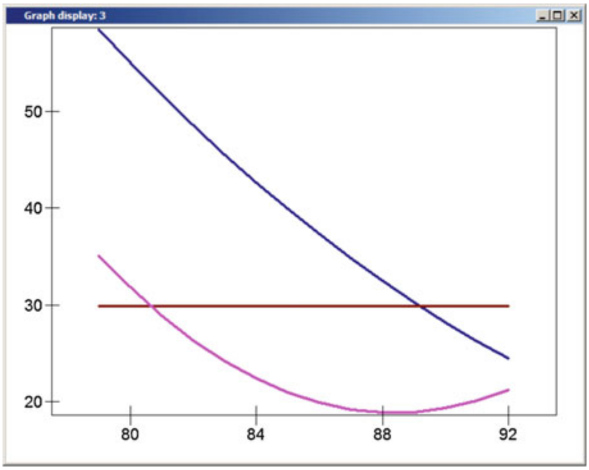
There is evidence of heterogeneity with a substantial reduction in −2∗log(likelihood). This means that the degree of scatter of individual observations about the predicted DISTRICT (level 2) means is not constant over time; it appears to have been decreasing. We can use the **Variance function** window 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:ID** is selected to be the **level**
In the drop-down menu by **variance output to**, select C23
Click **Calc**
Select **2:DISTRICT** to be the **level**
In the drop-down menu by **variance output to**, select C24
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 C23 VARF1 and C24 VARF2. We can plot all of these variance functions on the same scale across the observed values of YEAR; this will show us how the variance at COUNTY, DISTRICT and YEAR level have been changing over time.

Go to the **Customised Graph** window
Select display **D3**
Highlight **ds#1**
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**
Under the **plot what?** tab, 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 **red** and the **line thickness** to be **3**
Click **Apply**
Under the **plot what?** tab, 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 **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 may have increased slightly since then. (This may also be an 'edge effect'.) 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 changing 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 the yearly observations within DISTRICTs 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.

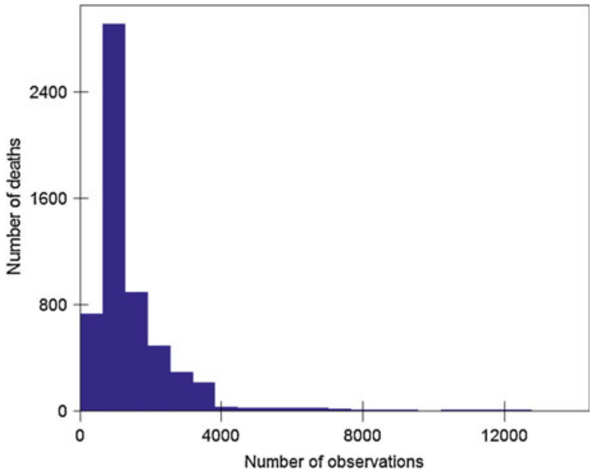
The example used has been illustrative of the methods employed when fitting a multilevel model; it is not, however, the way in which we would normally model such data. The following section goes on to consider a generalised linear model for these data; however, before proceeding to the more complex modelling it is important to have a good understanding of the basics covered up to this point.

This section has covered random coefficients using:
Equations window—making a variable random at different levels
Intervals and tests window—testing hypotheses about random parameters, e.g. the significance of a random slope
Variance function window—calculating a non-constant variance
Graph window—plotting random slopes
Predictions window—predictions including random intercepts and random slopes
Residuals window—plotting intercept and slope residuals with confidence intervals
Residuals window—pairwise comparisons of intercept and slope residuals
Equations window—modelling heterogeneity at level 1

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 are discrete rather than continuous—you cannot 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 appropriate.



Instead we can fit a generalised linear model and approximate a Poisson distribution for the data. This is the basis of the analysis conducted by Leyland (2004) on data including these.

Setting Up a Generalised Linear Model in MLwiN

First open the original worksheet lmdp.ws again.

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

We use the **Generate vector** window to create a constant and a unique identifier for every data point or observation.

Select the **Generate vector** window from the **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
Select **Type of vector** to be **Sequence**

(continued)

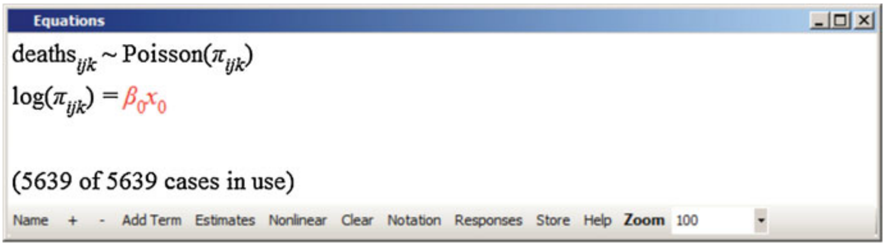
Select C10 to be the **Output column**
Enter 1 beside **Start number**
Enter 5639 (the number of data points) beside **End number**
Enter 1 beside **Step value**
Click the **Generate** button

In the **Names** window name C9 CONS and C10 ID. Then go to the **Equations** window. We will set DEATHS to be the response variable in a 3-level model: ID (observations) in DISTRICTs in COUNTYs.

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 ID 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, we will specify a **Poisson** distribution with a **log** link.

Click on the **N** that defines the normal distribution
Check the box marked **Poisson**
Click on **Done**



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 (the relationship between the response variable, DEATHS, and any of our explanatory variables) is taken to be the natural logarithm.

Now return to the **Equations** window and add CONS to the fixed part of the model only.

Click on $\boldsymbol{\beta}_0 \boldsymbol{x}_0$
Select CONS from the drop-down list
Click on **Done**

This time you may notice that there was no possibility to make the CONStant random at the lowest level (ID). This is because we have already defined the error structure at the lowest level when we specified that the data had a Poisson distribution. MLwiN automatically generated a new variable—which it called BCONS.1—which it will use in the estimation. This new variable can be seen in the **Names** window.

We are using the CONStant in the fixed part of the model to estimate the intercept or mean. We are going to fit a single-level Poisson model to start with, ignoring any variation between DISTRICTs and COUNTYs. As in the introductory tutorial, we will fit a quadratic in YEAR centred around 1979.

Go to **Data manipulation** menu
Select **Calculate**
Select the empty column C12 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 or by pressing the **Clear** button in the **Calculate** window
Next, select the empty column C13 and press the right arrow button
Click the '=' button on the keypad
Select C12 from the list of variables and press the right arrow button
Use the window's keypad to enter **^2**
Press **Calculate**

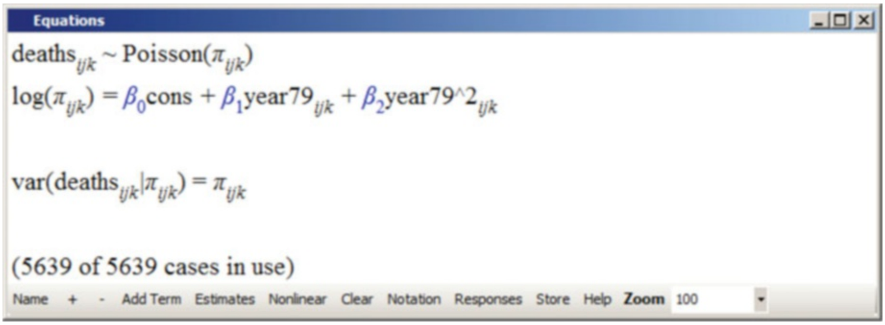
Use the **Names** window to name C12 and C13 YEAR79 and YEAR79^2, respectively. Next, return to the **Equations** window to add both terms to the fixed part of the model only.

Click on the **Add term** button
Select the **variable** YEAR79 from the drop-down list
Click on **Done**

(continued)

Click on the **Add term** button
Select the **variable** YEAR79^2 from the drop-down list
Click on **Done**

The **Equations** window should now look like this (remember you can use the **Name**, **Estimates** and **+** buttons to display more information about the current model in the **Equations** window):



The response, DEATHs, is assumed to follow a Poisson distribution with parameter π_{ijk}. 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_0, does not vary across COUNTYs or DISTRICTs since we have included no random effects at these levels. This model will provide estimates of how the average number of deaths has changed over time (the fixed part) allowing just for random fluctuations from one year to the next (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 10,000 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

$$\log\left({}^{\pi_{ijk}}\!/_{E_{ijk}}\right) = \beta_0 x_0 + \beta_1 x_{1ijk} + \beta_2 x_{2ijk}$$

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

$$\log\left(\pi_{ijk}\right) = \log\left(E_{ijk}\right) + \beta_0 x_0 + \beta_1 x_{1ijk} + \beta_2 x_{2ijk}$$

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 YEAR79^2. MLwiN provides a facility to do this; **the variable to be offset must be named OFFS**. We can create a variable containing the logarithm of the expected count using the LOGE function in the **Calculate** window.

Go to **Data manipulation** menu
Select **Calculate**
Select the empty column C14 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 C14 OFFS. This variable is now included in all subsequent Poisson models unless it is renamed.

Non-linear Estimation

As mentioned above, generalised linear models are approximated in MLwiN 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 non-linear 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** give 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 **first order** or **second order** approximation to the Taylor series.
- **Estimation type** gives us the option of using marginal quasi-likelihood (**MQL**) 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 **second 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 **first order, MQL** when exploring the data and to use **second order, PQL** to test the model and obtain final estimates.

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

This section has covered setting up a GLM using:
Equations window—changing the distributional assumptions
Calculate window—using arithmetical functions
Adding an OFFSet to a Poisson model
Equations window—non-linear estimation options

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.

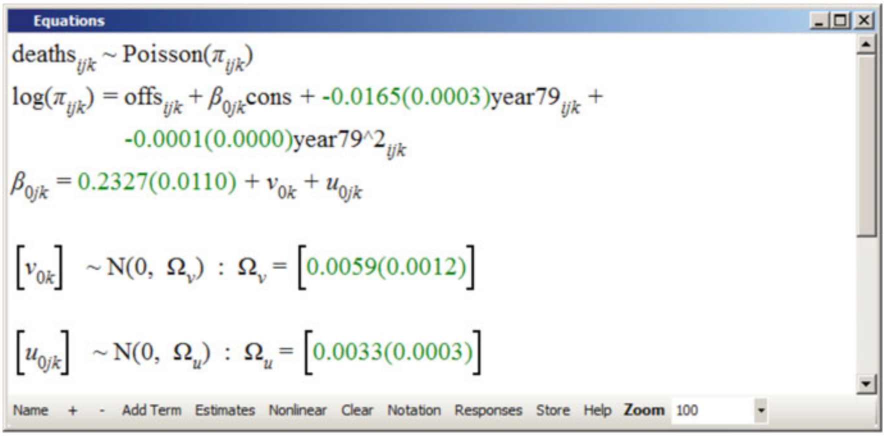
Go to the **Options** menu
Select **Numbers**
Increase the **# digits after decimal point** to 4
Click **Apply** and then **Done**

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 predicted number of deaths in 1979 (when both YEAR79 and YEAR79^2 equal 0) has been fitted as 1.272 (=$e^{0.2403}$) times the expected number of deaths. Since the expected number of deaths varies from DISTRICT to DISTRICT, so will the predicted number of deaths. Note that MLwiN does not give values of −2∗*loglikelihood* for generalised linear models.

We can now consider the effects of COUNTY and DISTRICT by letting the intercept or mean CONS vary at random across these two levels.

Return to the **Equations** window
Click on CONS
Click on the check box by **j(DISTRICT)**
Click on the check box by **k(COUNTY)**
Click on **Done**
Click on the **More** button to re-estimate the model

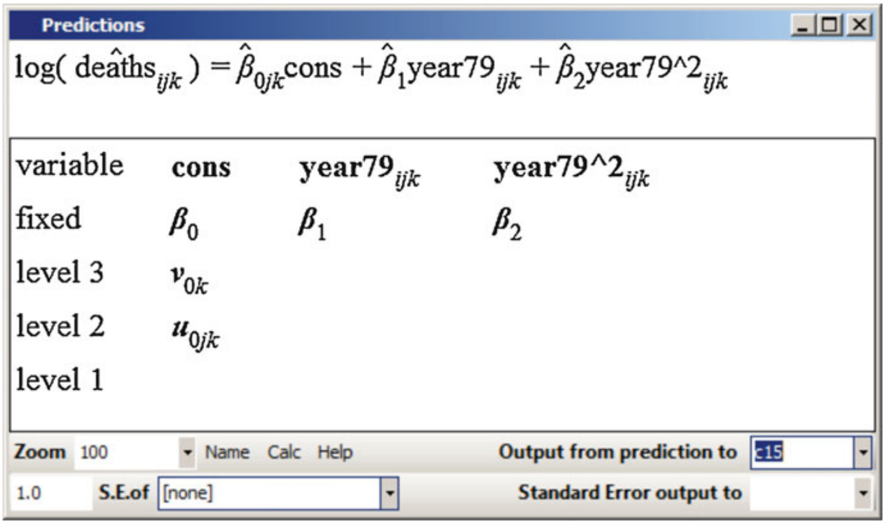


The parameter estimates in the fixed part are little changed; what is clear, however, is that there is variation over and above the Poisson variation in the counts that we might expect from one year to the next. Of the higher-level variation, about

64% (0.0059/[0.0059 + 0.0033]) is at the COUNTY (as opposed to the DISTRICT) level; this figure is very similar to that obtained from the 3-level variance components model of the SMR.

We can see what is going on more clearly using the **Graph** window. First of all we will get **Predictions** by DISTRICT and output these to C15.

> Go to **Model** menu
> Select **Predictions**
> Click on ***cons***, ***year79***$_{ijk}$ and ***year79^2***$_{ijk}$ to include all terms in the **Predictions** equation
> Select C15 for **output from prediction to**
> Click **Calc**



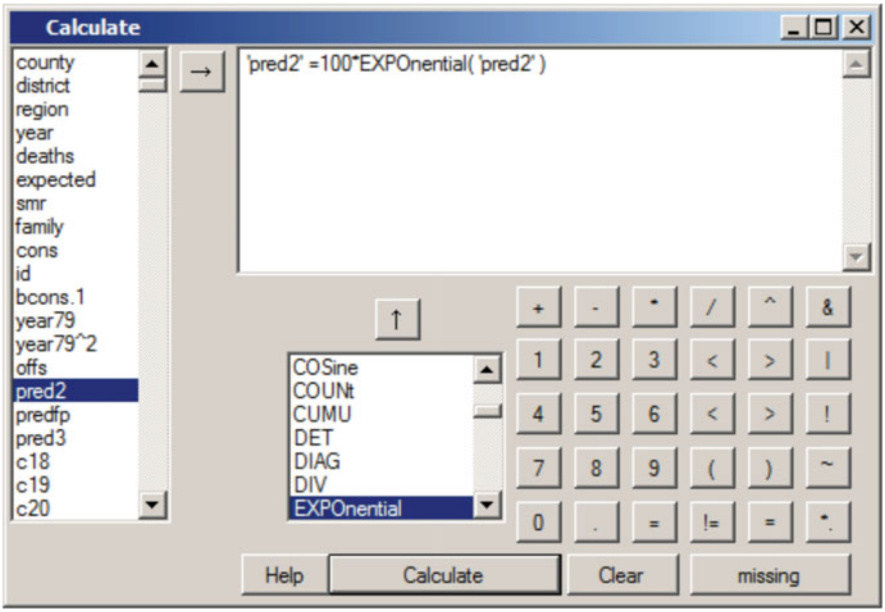
In the **Names** window, change the name of C15 to PRED2. In a similar manner we can put the predicted values for the fixed part in c16 and the level 3 predictions in c17.

> Return to the **Predictions** window
> Click on v_{0k} and u_{0jk} to remove them from the **Predictions** equation
> Select C16 for **output from prediction to**
> Click **Calc**
> Click on v_{0k} to include it in the **Predictions** equation
> Select C17 for **output from prediction to**
> Click **Calc**

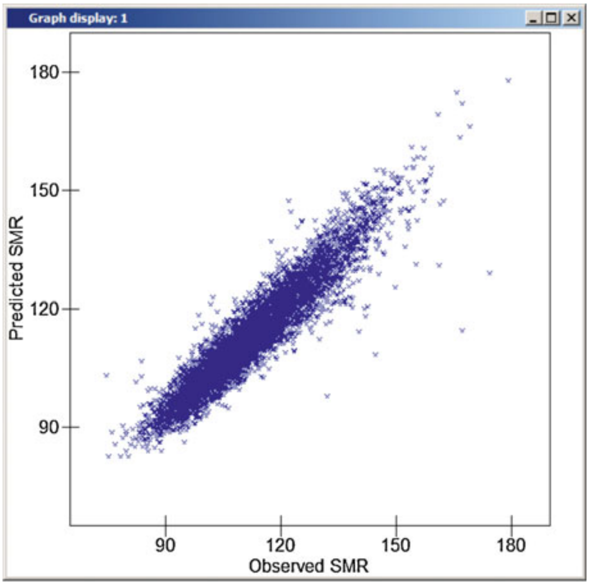
Name the variables C16 and C17 PREDFP and PRED3, respectively. You will note from the summary statistics in the **Names** window that these prediction equations are on the log scale; they also do not include our OFFSet term. As such, we really have the predicted values $\log\left(\widehat{y}/_{E_{ijk}}\right)$.

We can very easily convert these to predicted SMRs by taking the EXPOnents 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 EXPOnential 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.



The variability in the predicted SMRs (range 82–178) is slightly less than in the observed SMRs (range 75–179). However, some of the points on this graph are a long way from the diagonal (if a point lies on the diagonal, then the observed and predicted SMRs are equal). We can identify some of the points that lie further from the diagonal by clicking on those points in the graph. Some of those in the lower right-hand quadrant—where observed SMRs are considerably larger than the predicted—can be identified as belonging to district 2835 in county 28. It may be worth examining some of these points in more detail using the **View or edit data** window, clicking the **view** button to select the required variables and resizing the window if necessary:

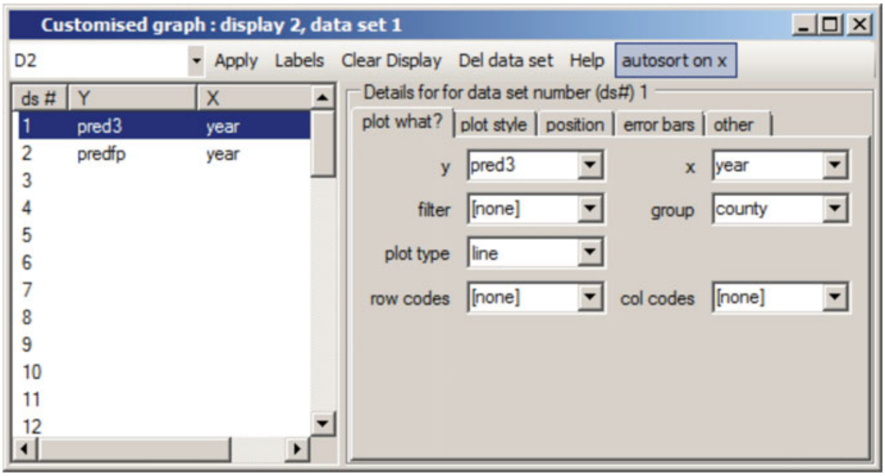
Data

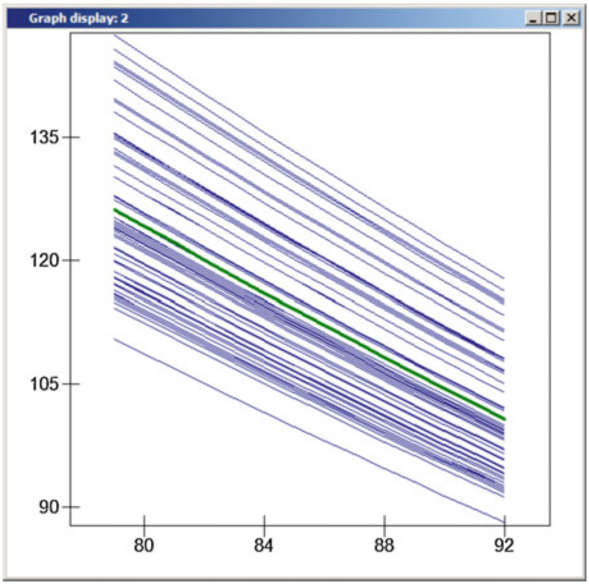
goto line 1 view Show value labels Font Help

	county(5639)	district(5639)	year(5639)	deaths(5639)	expected(5639)	smr(5639)	pred2(5639)	predfp(5639)	pred3(5639)
1586	28.0000	2830.0000	83.0000	1034.0000	899.8230	114.9115	115.2132	118.0290	113.7186
1587	28.0000	2830.0000	84.0000	1042.0000	924.8376	112.6684	113.2637	116.0318	111.7943
1588	28.0000	2830.0000	85.0000	1103.0000	950.5851	116.0338	111.3322	114.0531	109.8879
1589	28.0000	2830.0000	86.0000	1062.0000	970.8724	109.3862	109.4191	112.0932	107.9996
1590	28.0000	2830.0000	87.0000	1036.0000	1005.3714	103.0465	107.5245	110.1523	106.1296
1591	28.0000	2830.0000	88.0000	1084.0000	1034.8282	104.7517	105.6485	108.2305	104.2780
1592	28.0000	2830.0000	89.0000	1075.0000	1060.7599	101.3424	103.7915	106.3281	102.4450
1593	28.0000	2830.0000	90.0000	1056.0000	1091.5106	96.7467	101.9534	104.4451	100.6308
1594	28.0000	2830.0000	91.0000	1140.0000	1112.5481	102.4675	100.1346	102.5818	98.8355
1595	28.0000	2830.0000	92.0000	1080.0000	1123.5365	96.1250	98.3350	100.7383	97.0593
1596	28.0000	2835.0000	81.0000	20.0000	11.4703	174.3632	128.8103	122.0782	117.6199
1597	28.0000	2835.0000	82.0000	18.0000	14.7798	121.7880	126.6645	120.0445	115.6605
1598	28.0000	2835.0000	83.0000	17.0000	13.4120	126.7520	124.5378	118.0290	113.7186
1599	28.0000	2835.0000	84.0000	22.0000	15.8785	138.5518	122.4305	116.0318	111.7943
1600	28.0000	2835.0000	85.0000	17.0000	11.9065	142.7790	120.3427	114.0531	109.8879
1601	28.0000	2835.0000	86.0000	16.0000	16.2276	98.5976	118.2747	112.0932	107.9996
1602	28.0000	2835.0000	87.0000	25.0000	20.0699	124.5646	116.2268	110.1523	106.1296
1603	28.0000	2835.0000	88.0000	27.0000	16.1551	167.1298	114.1990	108.2305	104.2780
1604	28.0000	2835.0000	89.0000	19.0000	14.5950	130.1811	112.1917	106.3281	102.4450
1605	28.0000	2835.0000	90.0000	20.0000	18.5083	108.0595	110.2049	104.4451	100.6308
1606	28.0000	2835.0000	91.0000	29.0000	20.0285	144.7934	108.2388	102.5818	98.8355
1607	28.0000	2835.0000	92.0000	20.0000	18.1693	110.0757	106.2936	100.7383	97.0593
1608	29.0000	2905.0000	79.0000	1231.0000	902.8111	136.3519	136.5076	126.1988	135.1608

For most years in district 2835 there were more deaths than expected; however, the expected number of deaths (and therefore the population) was rather small (range 16–29). The observed SMRs for this district show considerable disparity, ranging from 174 in 1981 to 99 in 1986. The values in PRED3 suggest that county 28 as a whole has an SMR which is slightly below average, the value of 97 in 1992 being lower than the fitted average in PREDFP of 101. PRED2 contains the predicted SMRs based on the fixed part of the model—containing just an intercept and a linear and quadratic term in YEAR—and the residuals at levels 2 (DISTRICT) and 3 (COUNTY). Since both sets of residuals have been shrunk towards zero, the predicted SMRs are also known as shrunken estimates. (This name may seem confusing, since the estimates for individual years are not always closer to the average in PREDFP. For example, in 1982 the observed SMR of 122 is nearer to the average of 120 than the shrunken estimate of 127. This is because the shrunken estimate for any one year is derived from data for all years in that district—and, indeed, for all districts and all counties—and in this sense is thought to be a closer approximation to a 'true', underlying relative risk of mortality.) Note that the values of the predicted SMR are much closer to the observed values for the previous DISTRICT, 2830, reflecting the larger number of expected deaths and the consequent increase in confidence that the observed rate is close to the 'true' mortality rate.

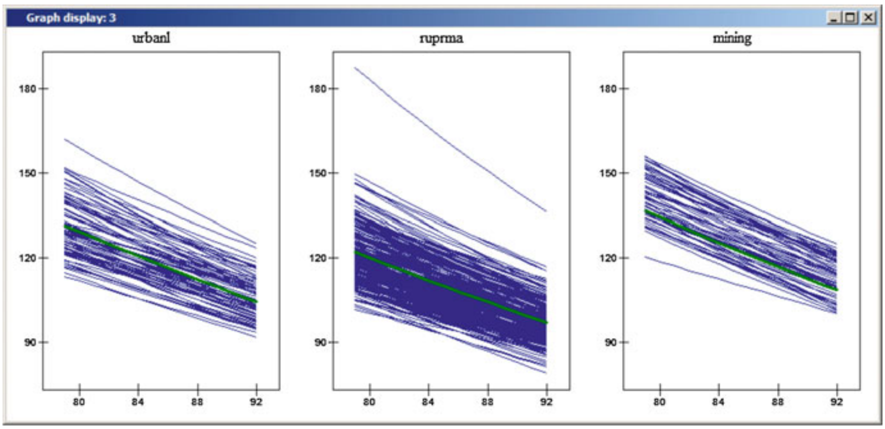
We can also plot the predicted values at national and COUNTY level by YEAR:





This graph illustrates the convergence of SMRs that we noted in the previous analysis even though we have not included a random slope; this is to be expected since the assumption of Poisson variation means that we can expect the variance 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.

> This section has GLM interpretation using:
> **Equations** window—interpreting parameter estimates
> **Calculate** window—converting predicted values back to SMRs

Predictions and Confidence Envelopes

Compare the parameter estimates obtained from the basic 1-level and 2-level models with YEAR79 as the sole covariate. Note particularly the standard errors in the fixed part of the two models; whilst the standard error associated with the intercept (CONStant) has increased with the addition of another level, as we might expect, that associated with the slope (YEAR79) has actually decreased from 0.0387 to 0.0164. One of the reasons for fitting a multilevel model is that single-level models tend to underestimate the standard errors in the fixed part, so what is the cause of this counter-intuitive result?

Results Table

Copy

	Trend 1-level	S.E.	Trend 2-level	S.E.
Response	smr		smr	
Fixed Part				
cons	125.6769	0.2962	125.6986	0.5439
(year-79)	-1.9822	0.0387	-1.9845	0.0164
Random Part				
Level: cons				
Level: id				
Var(cons)	137.2831	2.5854	24.4935	0.4807
Level: district				
Var(cons)			112.8966	8.0625

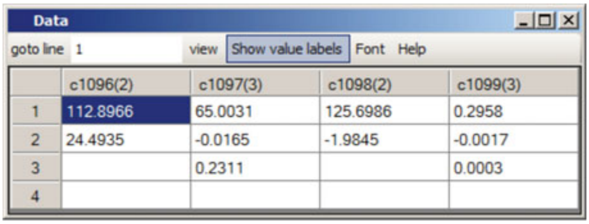
To understand this apparent anomaly, it is necessary to consider the confidence that we have in any predicted value $\widehat{y}_{ij} = \widehat{\beta}_0 + \widehat{\beta}_1 x_{1ij}$. The variance of our predicted value $\widehat{y}_{ij}$ is given by

$$\mathrm{var}\left(\widehat{y}_{ij}\right) = \mathrm{var}\left(\widehat{\beta}_0\right) + x_{1ij}^2 \mathrm{var}\left(\widehat{\beta}_1\right) + 2x_{1ij}\mathrm{cov}\left(\widehat{\beta}_0, \widehat{\beta}_1\right)$$

From the current (2-level) model, we have estimates of $\mathrm{var}\left(\widehat{\beta}_0\right)$ and $\mathrm{var}\left(\widehat{\beta}_1\right)$ as $(0.5439)^2$ and $(0.0164)^2$, respectively. However, there is no estimate of the covariance in the **Equations** window. This parameter is stored by MLwiN, but we will have to find it.

Columns C1096–C1099 are used by MLwiN to store, respectively, the random parameter estimates, their estimated covariance matrix, the fixed parameters and their estimated covariance matrix. Both covariance matrices are stored in lower diagonal form. Take a look at these four columns in the **Data** window.

Go to **Data Manipulation** menu
Select **View or edit data**
Click on the **view** button
Scroll down and highlight C1096, C1097, C1098 and C1099
Click the **OK** button and resize the window if necessary



	c1096(2)	c1097(3)	c1098(2)	c1099(3)
1	112.8966	65.0031	125.6986	0.2958
2	24.4935	-0.0165	-1.9845	-0.0017
3		0.2311		0.0003
4				

Looking at columns C1098 and C1099 we find the estimated distribution of the fixed parameters to be

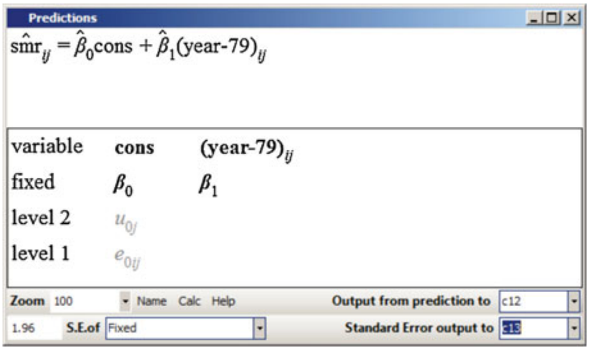
$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} \sim N\left(\begin{bmatrix} 125.6982 \\ -1.9845 \end{bmatrix}, \begin{bmatrix} 0.2958 & \\ -0.0017 & 0.0003 \end{bmatrix} \right)$$

Our estimates of the two parameters are therefore not independent; we find a negative correlation of about -0.2 between the intercept and the slope. We can use the **Predictions** window to plot the predicted line and a 95% confidence envelope. First save the data so that we can return to the current model when we have finished our exploration.

Go to **File** menu
Select **Save worksheet as**. . .
Type **lmdpapp1.ws** as the new filename
Click the **Save** button

The Predictions window is described in more detail when it is used later in this tutorial. At this stage we will do little more than detail the commands. To start with we will obtain the predicted values of the SMR for each COUNTY, DISTRICT and YEAR based on the fixed part of the model alone, together with 1.96 times the standard error of these estimates. (A 95% confidence interval can be constructed as the estimate ± 1.96 standard errors.) At the moment, the fixed part of the model just contains the intercept and the time trend YEAR79.

Go to **Model** menu
Select **Predictions**
Click on β_0 and β_1
In the drop-down list by **Output from prediction to** select C12
Edit the multiplier of **S.E.** to 1.96
In the drop-down list by **S.E. of** select **Fixed**
In the drop-down list by **Standard Error output to** select C13
Click on **Calc**

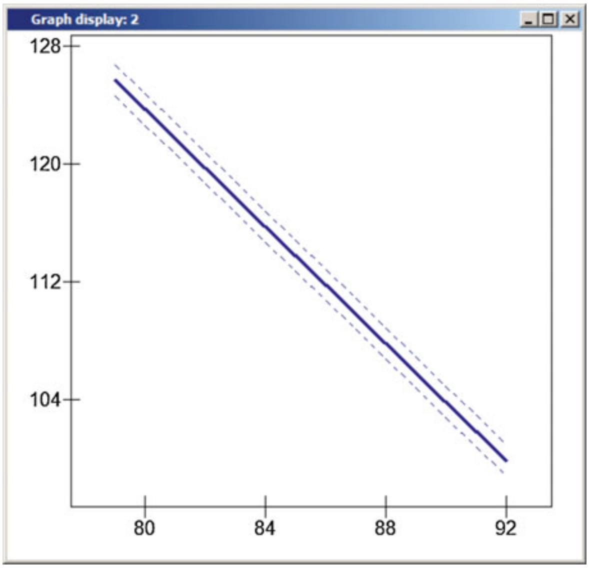


C12 now contains our predicted regression line and C13 contains 1.96 times the standard error of the fixed parameters. We can plot these using the **Customised Graph** window, plotting the predicted values against YEAR as a line graph.

Go to the **Graphs** menu
Select **Customised Graph(s)**
Select the second dataset, **D2**, from the pull-down list at the top left of the window
From the drop-down list by **y** select C12
From the drop-down list by **x** select YEAR
Change the **plot type** to **line**
Click the **Apply** button

This produces a line graph of the predicted mean SMR. We can add confidence intervals around this line using the **y errors** feature on the **error bars** tab:

In the **Customised Graph** window, click on the **error bars** tab
Select C13 to be the **y errors +** and the **y errors –**
Change the **y error type** to **lines**
Click on **Apply**



This is the predicted regression line together with 95% confidence intervals. We will now compare this with the single-level model. We start by removing CONS from the random part of the model at level 2 and then we re-estimate the model.

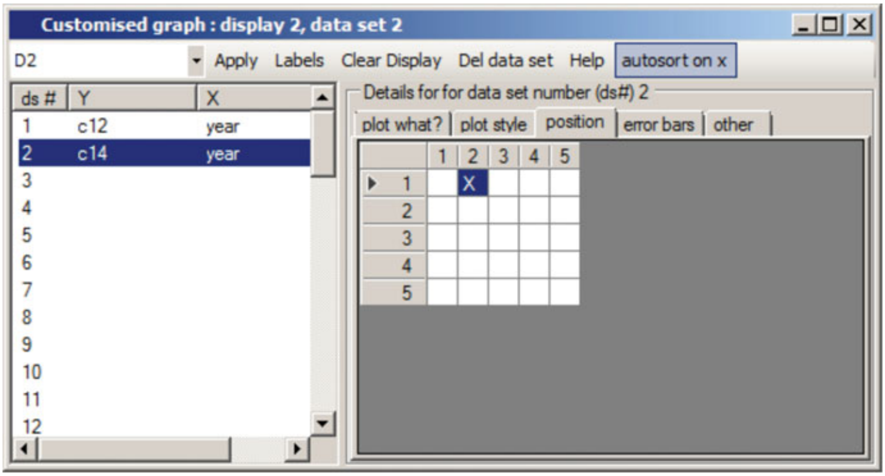
In the **Equations** window, click on u_{0j}
Remove the tick by **j(district)**
Click on **Done**
Click on **More**

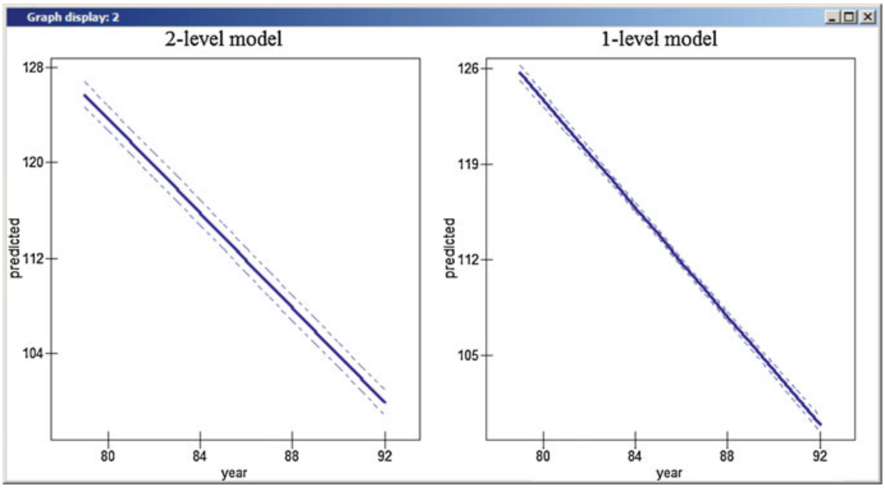
This returns us to the single-level model that we had fitted previously. We can now use the **Predictions** window again to obtain the predicted values of the SMR based on the fixed part of the model, together with appropriate multiples of the standard errors of these estimates.

In the **Predictions** window, ensure that both fixed terms are included but not the random terms
In the drop-down list by **Output from prediction to** select C14
Edit the multiplier of **S.E.** to 1.96
In the drop-down list by **S.E. of** select **Fixed**
In the drop-down list by **Standard Error output to** select C15
Click on **Calc**

We can plot the estimates from this single-level model alongside those from the 2-level model using the **position** feature of the **Customised Graph** window.

In the **Customised Graph** window, click on row **2** under **ds #**
From the drop-down list by **y** select C14
From the drop-down list by **x** select YEAR
Change the **plot type** to **line**
Click on the **error bars** tab
Select C15 to be the **y errors +** and the **y errors –**
Change the **y error type** to **lines**
Click on the **position** tab
Click in the box on row **1**, column **2**
Click on **Apply**





(Note: the above graphs have added titles.) We can see that the confidence envelope around the predicted mean is much tighter under the 1-level model than the 2-level model. We can confirm this by looking at the variables C12–C15 in the **Names** window:

Names

Column: Name Description Toggle Categorical *Data:* View Copy Paste Delete

Name	Cn	n	missing	min	max	categorical
c12	12	5639	0	99.8995...	125.698...	False
c13	13	5639	0	1.04552...	1.06599...	False
c14	14	5639	0	99.9086...	125.676...	False
c15	15	5639	0	0.30813...	0.58057...	False

C13, 1.96 times the standard error under the 2-level model, varies between 1.046 and 1.066; C15, 1.96 times the standard error under the single-level model, takes values ranging between 0.308 and 0.581. The single-level estimates show signs of ‘misestimated precision’—ignoring the data hierarchy leads to a confidence envelope that is too tight.

Retrieve the saved worksheet before returning to the section on the hierarchy viewer:

Go to **File** menu
Select **Open** worksheet
Choose **lmdpapp1.ws** as the filename
Click the **Open** button

References

Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge

Leyland AH (2004) Increasing inequalities in premature mortality in Great Britain. J Epidemiol Commun Health 58:296–302

Leyland AH, McLeod A (2000) Mortality in England and Wales, 1979-1992. An introduction to multilevel modelling using MLwiN. MRC Social and Public Health Sciences Unit occasional paper no. 1. MRC Social and Public Health Sciences Unit, Glasgow

Snijders TAB, Bosker RJ (2012) Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage, Los Angeles