

Practical 3: Multilevel Models for Binary Responses in MLwiN

Corrections made following Craig Duncan

1 The Bangladesh Demographic and Health Survey 2004 Dataset

You will be analysing data from the Bangladesh Demographic and Health Survey (BDHS), a nationally representative cross-sectional survey of women of reproductive age (13-49 years).

The response variable is a binary indicator of whether a woman received antenatal care from a medically-trained provider (a doctor, nurse or midwife) at least once before her most recent live birth.

We consider multilevel models to allow for and to explore between-community variance in antenatal care. The data have a two-level hierarchical structure with 5366 women at level 1, nested within 361 communities at level 2. In rural areas a community corresponds to a village, while an urban community is a neighbourhood based on census definitions.

We consider a range of predictors. At level 1, we consider variables such as a woman's age at the time of the birth and education. Level 2 variables include an indicator of whether the region of residence is classified as urban or rural. There are also community-level contextual measures for wealth and maternal education

The file **antenatal.wsz** contains the following variables:

Variable name	Description and codes
Comm	Community identifier
Womid	Woman identifier
Antemed	Received antenatal care at least once from a medically-trained provider, e.g. doctor, nurse or midwife (1=yes, 0=no); this needs to be 1 and 0 and not other coding such as 2 and 1
Bord	Birth order of child (ranges from 1 to 13)
Mage	Mother's age at the most recent child's birth (in years)
Urban	Type of region of residence at survey (1=urban, 0=rural)
Meduc	Mother's level of education at survey (1=none, 2=primary, 3=secondary or higher)
Islam	Mother's religion (1=Islam, 0=other)
Wealth	Household wealth index in quintiles (1=poorest to 5=richest)

Cons	A column of ones. This variable will be included as an explanatory variable in all models and its coefficient will be the intercept
ComWealth	A precision-weighted estimate of community wealth; see Appendix (Section 11) for how this has been calculated
ComEduc	A precision-weighted estimate of community maternal education

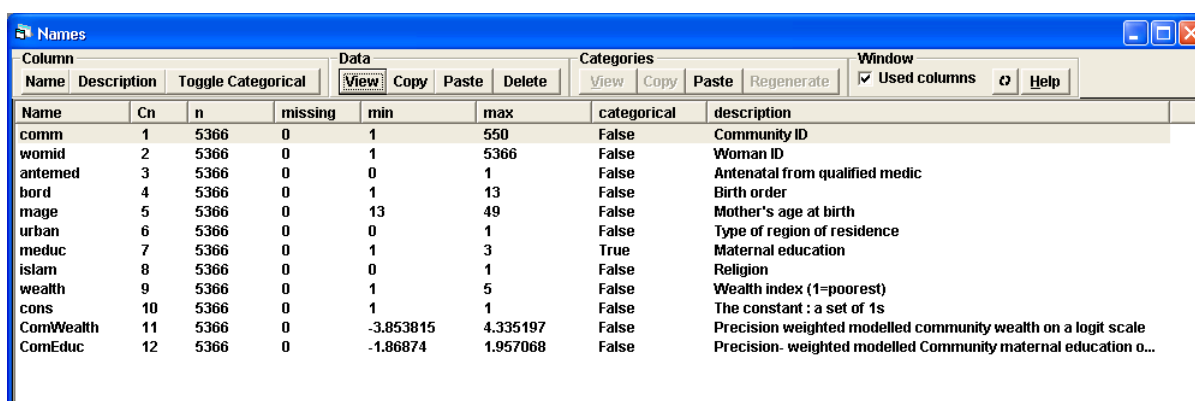
A word of warning!

This is a large and rather forbidding practical that may take you quite a while to complete in its entirety. This is because in addition to demonstrating how to specify and estimate a discrete outcome model through quasi-likelihood and MCMC estimation, we have also tried to provide a feel of a real project as you work through an analysis. To this end we have exemplified a number of procedures that you could undertake post-estimation to help interpret the estimates and we have done so for logits, odds and probabilities. We have also included some important ‘tricks of the trade’ that you have to use when you encounter problems. Thus we deal with what to do when confronted with the dreaded “**MCMC error 0135 prior variance matrix is not positive definite**”. We have also tried to give some explanation of what is going on behind the scenes as it were. You should try and complete up to and including Section 3, as this will give you experience of MCMC estimation. You can complete the remainder at your discretion.

- In MLwiN go to the **File** menu, then **Open worksheet** and select the file **antenatal.wsz**

The **Names** window will appear.

- Click the check box next to **Used columns** to view only those columns that contain data



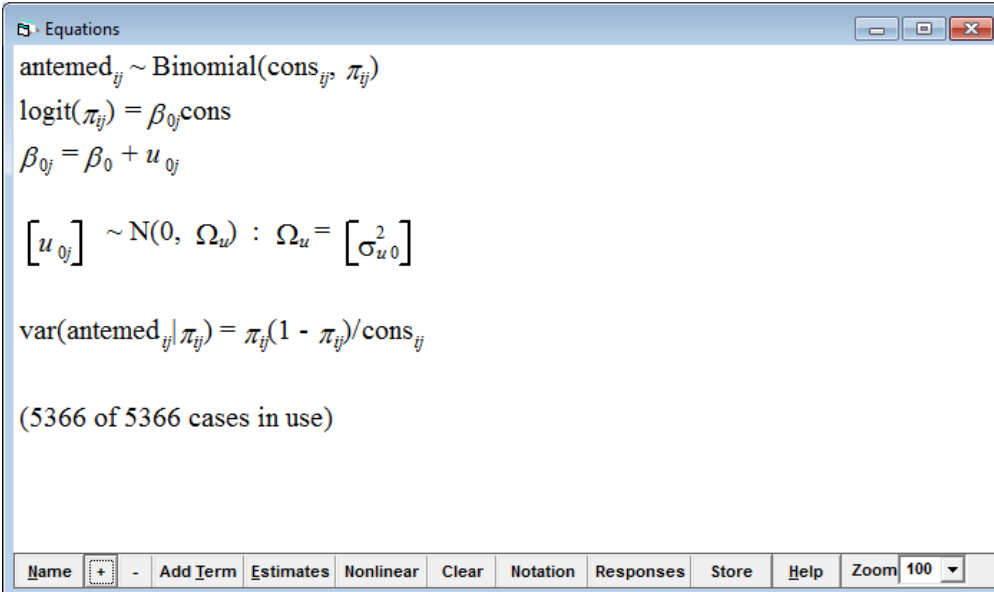
The screenshot shows the 'Names' window in MLwiN. It has a menu bar with 'Data', 'Categories', and 'Window'. Below the menu bar is a toolbar with buttons for 'View', 'Copy', 'Paste', 'Delete', 'Regenerate', and 'Help'. The main area is a table with columns: Name, Cn, n, missing, min, max, categorical, and description. The table lists 12 variables: comm, womid, antemed, bord, mage, urban, meduc, islam, wealth, cons, ComWealth, and ComEduc. The 'Used columns' checkbox is checked.

Name	Cn	n	missing	min	max	categorical	description
comm	1	5366	0	1	550	False	Community ID
womid	2	5366	0	1	5366	False	Woman ID
antemed	3	5366	0	0	1	False	Antenatal from qualified medic
bord	4	5366	0	1	13	False	Birth order
mage	5	5366	0	13	49	False	Mother's age at birth
urban	6	5366	0	0	1	False	Type of region of residence
meduc	7	5366	0	1	3	True	Maternal education
islam	8	5366	0	0	1	False	Religion
wealth	9	5366	0	1	5	False	Wealth index (1=poorest)
cons	10	5366	0	1	1	False	The constant : a set of 1s
ComWealth	11	5366	0	-3.853815	4.335197	False	Precision weighted modelled community wealth on a logit scale
ComEduc	12	5366	0	-1.86874	1.957068	False	Precision- weighted modelled Community maternal education o...

2 Specifying and Estimating a Two-Level Random Intercept Model

We will begin by fitting a null or empty two-level model, which is a model with only an intercept and community random effects. The data have already been sorted so that women are nested in communities. To specify a two-level logit model in MLwiN:

- From the **Model** menu, select **Equations**
- Click the red **y** in the **Equations** window
- From the drop-down list labelled **y**: select **antemed**
- From the drop-down list labelled **N levels**: select **2-ij**
- From the drop-down list labelled **level 2(j)**: select **comm**
- From the drop-down list labelled **level 1(i)**: select **womid**
- Click **done**
- Click on **N(XB, Ω)** and check **Binomial**. A list of link functions will appear. We will retain the default of **logit** so click **Done**
- We now need to specify the denominator which for a binary response is always equal to 1, so click on the red n_{ij} and select **cons** from the drop-down list. Check **Done**. If you look in the **Names** window you will see that a new variable called **denom** has been added to the worksheet. This has been created from **cons**; receiving care is therefore a 1 out of 1 trial; not receiving is 0 out of 1; there is also a new variable called **bcons** which is a placement for the level 1 binomial weight, which can be ignored; you may have to refresh the **Names** window to see the update; other new variables in the c1090's can be ignored
- Click on x_0 and select **cons** from the drop-down list. Check **j(comm)** to add a j subscript to the intercept β_0 , then click **Done**
- Click the **+** button twice to see the full model specification



The screenshot shows the 'Equations' window in MLwiN. The model is specified as follows:

$$\text{antemed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

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

$$\text{var}(\text{antemed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{cons}_{ij}$$

(5366 of 5366 cases in use)

The bottom of the window features a toolbar with buttons: Name, +, -, Add Term, Estimates, Nonlinear, Clear, Notation, Responses, Store, Help, Zoom, and a dropdown menu set to 100.

The intercept β_{0j} consists of two components: a fixed effect, β_0 , shared by all communities, and a random effect u_{0j} , specific to community j . The random effect is assumed to follow a Normal distribution with covariance matrix Ω_u which in this simple model contains just one element, the between-community variance, σ_u^2 .

We can now check the model-implied hierarchy:

- From the **Model** menu, select **Hierarchy Viewer**; close after viewing

Hierarchy viewer

Summary

level	range	total
comm(j)	1.. 361	361
womid(i)	1.. 25	5366

Options... Help

Details

L2 ID: 1, j = 1 of 361 N1 14	L2 ID: 2, j = 2 of 361 N1 19	L2 ID: 3, j = 3 of 361 N1 21	L2 ID: 4, j = 4 of 361 N1 18	L2 ID: 5, j = 5 of 361 N1 22
L2 ID: 6, j = 6 of 361 N1 20	L2 ID: 7, j = 7 of 361 N1 21	L2 ID: 8, j = 8 of 361 N1 19	L2 ID: 9, j = 9 of 361 N1 12	L2 ID: 10, j = 10 of 361 N1 23
L2 ID: 11, j = 11 of 361 N1 12	L2 ID: 12, j = 12 of 361 N1 19	L2 ID: 13, j = 13 of 361 N1 10	L2 ID: 14, j = 14 of 361 N1 22	L2 ID: 15, j = 15 of 361 N1 20
L2 ID: 16, j = 16 of 361 N1 12	L2 ID: 17, j = 17 of 361 N1 11	L2 ID: 18, j = 18 of 361 N1 19	L2 ID: 19, j = 19 of 361 N1 14	L2 ID: 20, j = 20 of 361 N1 12
L2 ID: 21, j = 21 of 361 N1 11	L2 ID: 22, j = 22 of 361 N1 7	L2 ID: 23, j = 23 of 361 N1 10	L2 ID: 24, j = 24 of 361 N1 9	L2 ID: 25, j = 25 of 361 N1 12
L2 ID: 26, j = 26 of 361 N1 10	L2 ID: 27, j = 27 of 361 N1 17	L2 ID: 28, j = 28 of 361 N1 16	L2 ID: 29, j = 29 of 361 N1 13	L2 ID: 30, j = 30 of 361 N1 20

Because the data have been correctly sorted so that women are nested in communities, we see that we indeed have 361 communities and 5366 women with a maximum of 25 women in any community, with 14 women in the first community, 19 in the second and so on.

Before fitting the model, we have to specify details about the estimation procedure to be used. There are several estimation procedures available for binary and other categorical response models. In MLwiN, there are two options: quasi-likelihood methods and Markov chain Monte Carlo (MCMC) methods. There are four varieties of quasi-likelihood methods in MLwiN: 1st and 2nd order marginal quasi-likelihood (MQL1 and MQL2) and 1st and 2nd order penalised or predictive quasi-likelihood (PQL1 and PQL2). We will begin with the default procedure MQL1 to obtain starting values for MCMC; this procedure has the greatest chance of converging but can give rather poor, often downwardly biased estimates, and that is why we will later use MCMC procedures.

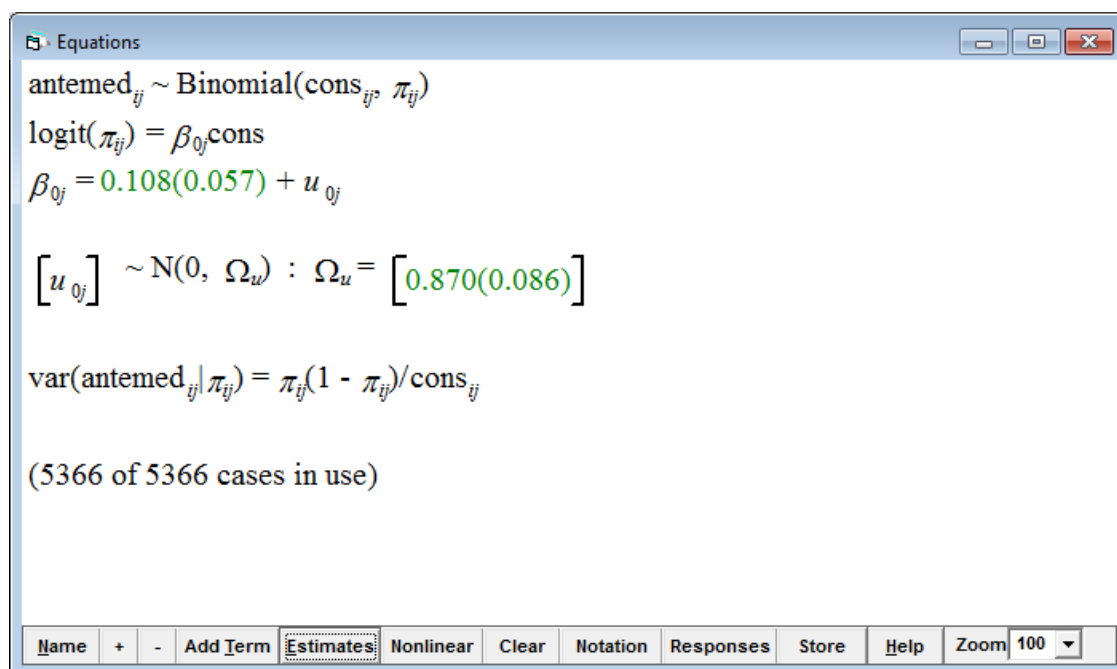
To specify the quasi-likelihood estimation procedure:

- Click on the **Nonlinear** button at the bottom of the **Equations** window
- In the **Nonlinear Estimation** window, click on **Use Defaults**, then **Done**

- Click once on **Estimates** so that the parameters to be estimated (β_0 and σ_u^2) appear in blue

Now to fit the model:

- Click **Start** which will estimate the model to convergence and the estimates should turn green; i.e. there has been little change since the previous iteration
- Click once on **Estimates** to see the estimated coefficients (and standard errors in brackets)



The screenshot shows the 'Equations' window in MLwiN. The equations displayed are:

$$\text{antedmed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons}$$

$$\beta_{0j} = 0.108(0.057) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.870(0.086)]$$

$$\text{var}(\text{antedmed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{cons}_{ij}$$

(5366 of 5366 cases in use)

The 'Estimates' button in the bottom toolbar is highlighted with a dashed border, indicating it is the active view.

These estimates provide the starting values for the MCMC procedure. To switch to MCMC:

- Click on **Estimation control** and then **MCMC**
- The user can set the **Burn-in Length**, **Monitoring Chain Length** and **Thinning**, which are explained briefly below
- We will accept the defaults for now, so click **Done**
- Click the **+** button to see the specifications of the prior distributions
- Now click **Start**

After a couple of minutes, depending on the speed of your computer, (it will go through an Adaptive phase, then a burn-in which is discarded, and then a Monitoring phase) you should obtain the following estimates:

Equations

$$\text{antedmed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons}$$

$$\beta_{0j} = 0.157(0.075) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [1.510(0.161)]$$

$$\text{var}(\text{antedmed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{cons}_{ij}$$

PRIOR SPECIFICATIONS

$$p(\beta_0) \propto 1$$

$$p(1/\sigma_u^2) \sim \text{Gamma}(0.001, 0.001)$$

Deviance(MCMC) = 6047.276(5366 of 5366 cases in use)

Name * - Add Term Estimates Nonlinear Clear Notation Responses Store Help Zoom 100

In Bayesian statistics, parameters are treated as random quantities with two probability distributions: a *prior* (representing information before the sample data are collected) and a *posterior* (combining the prior information with the sample data). We have assumed a ‘diffuse’ or ‘weakly-informative’ prior because we have no strong prior information about likely parameter values. Using an MCMC method (specifically, Metropolis Hastings), we have taken a large number of random draws from the joint posterior distribution of the parameters to obtain a *chain* of values for each parameter.

The **Monitoring chain length** is the number of random draws, which is set by default to 5000. We will often require more than 5000 draws, and we can use various diagnostics to judge how many draws are needed. Unlike frequentist approaches such as maximum or quasi-likelihood, where convergence is reached when the difference between parameter estimates from two successive iterations is within a pre-specified tolerance, MCMC convergence is more difficult to assess. In MCMC the values in a chain should be drawn from the *distribution* of the parameter, rather than approach a single point as in maximum likelihood. In practice, the analyst must judge whether the process has reached convergence. For this reason, the parameter estimates from MCMC will always appear in blue, regardless of how many draws have been taken. We are looking for convergence to an equilibrium distribution which is shown by a lack of overall trending in the parameter estimates.

The parameter estimates shown in the **Equations** window are the mean of the 5000 parameter values drawn from the posterior distribution, and the standard error (in brackets) is the standard deviation of the 5000 values. We can also construct an interval estimate for each parameter. For example, the 2.5 and 97.5 percentiles of the ordered values represent a 95% credible interval for the population parameter, which is analogous to the frequentist 95% confidence interval, but can be asymmetric. Values outside this credible interval would receive little support from the data and this model.

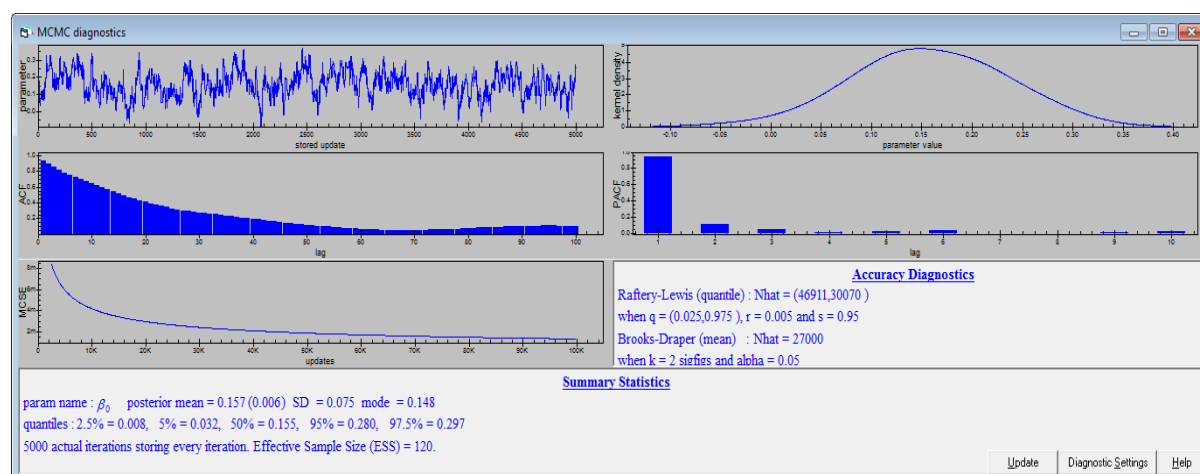
Depending on the starting values, initial draws may not be from the desired posterior distribution; the chains may take some time to ‘settle down’. The period before a chain has reached equilibrium (converged) is known as the **burn-in**, and this part of the chain is discarded before summary statistics are calculated. In MLwiN the default burn-in is 500 draws. This is usually sufficient as the quasi-likelihood estimates give good starting values. This will generally be the case for fixed estimates and for the random estimates with a hierarchical structure. You may have to increase it with more complex structures such as the cross-classified model as the standard IGLS procedures are not really designed for such methods.

Comparing the MCMC estimates with the MQL1 estimates, we see that the MCMC estimates are much larger (the between community variance is now 1.510 compared to 0.870) and we should use these MCMC estimates as we know that MQL-1 estimates can be downwardly biased.

3 MCMC Diagnostics

To assess convergence, we can view the **trajectories** of the model parameters. We will start with the intercept term (the coefficient associated with **cons**).

- From the **Model** menu, select **Trajectories**
- The coefficient of **cons** is β_0 (click **Estimates** to see the parameters in mathematical notation). Click on the graph for β_0 and respond ‘yes’ to the question ‘Calculate MCMC diagnostics?’



At the bottom of the **MCMC diagnostics** window, various summary statistics of the chain for β_0 are shown. These include the mean and standard deviation, also shown in the **Equations** window, and the lower and upper limit of the 95% credible interval, 0.008 to 0.297. We have little evidence that the estimate goes negative, so we can be pretty confident when we convert the logit estimate to a probability that more than 50 percent of the women in the study have received antenatal care (a logit of 0.0 equals a probability of 0.50)

A large number of **MCMC diagnostics** are produced. We will consider only a few here; readers are referred to Browne (2009) for full details.¹ The top-left graph shows the 'trace' of the 5000 values drawn from the monitoring phase of the posterior distribution of β_0 . If convergence has been reached, the trace should resemble 'white noise', in which case the chains are said to be 'mixing well'. In this case, there appears to be no overall trend. This suggests that convergence has indeed been achieved, and we do not need to increase the burn-in (from the default 500) and discard more of the earlier part of the chain.

The top-right graph shows a smoothed histogram of the parameter values. This should show an approximately Normal distribution for this fixed part parameter, which causes the 95% credible interval to be symmetric about the mean. Another implication of the Normal distribution is that the mean, median and mode of the chain of estimates will coincide. Variance parameters, in contrast, will often tend to have a skewed distribution as they cannot go negative during estimation, leading to an asymmetric credible interval.

The second row of graphs shows the autocorrelation function and the partial autocorrelation function. Ideally the 5000 parameter values should be independent draws from the posterior distribution, so that they rapidly explore the joint distribution of the parameters and not become stuck in one part of it and 'learn' slowly. But it is common for them to be correlated, especially the estimates of the intercept term. Here, we see high correlations even at lag 10, i.e. values that are 9 apart in the chain are correlated. The non-independence of successive chain values can also be seen in the trace (top-left plot) interpreted earlier. The high autocorrelation is also reflected in the **Effective Sample Size (ESS)** shown at the bottom of the **Diagnostics** window. Although 5000 draws have been taken, the ESS is only 120; we have only the equivalent of 120 independent draws for the values of the intercept. While 120 may be sufficient to say something with confidence about values of central tendency for a parameter, it is rather too low on which to base percentiles like the 2.5% and 97.5% credible intervals. The solution to this is either to run the chains for longer or we can use procedures to make the chains less correlated such as hierarchical centering (see Chapter 25, *MCMC Manual*,) which we will demonstrate later. For now we will simply increase the length of the chain.

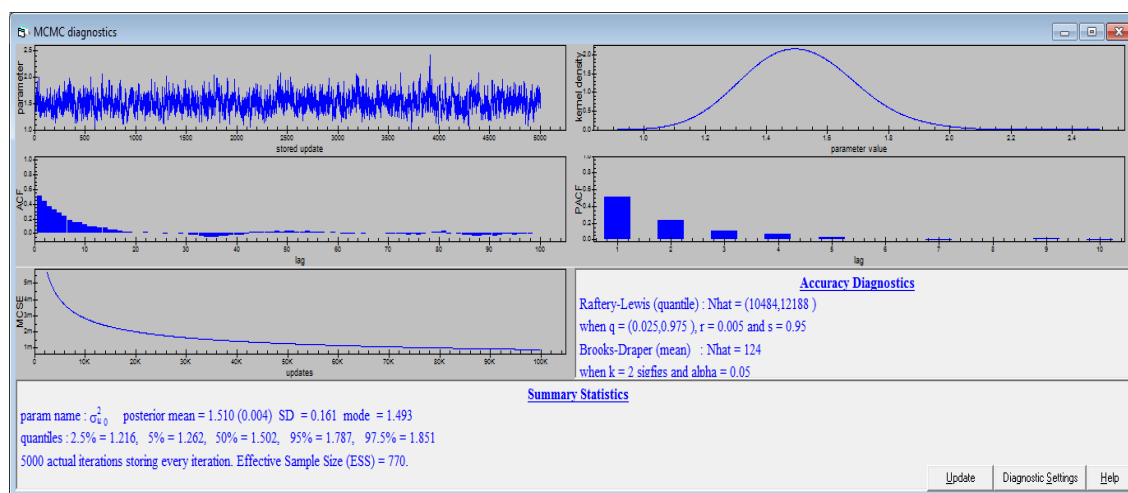
There are a number of procedures that will guide us in deciding for how long to run the chains. Under **Accuracy Diagnostics**, the **Raftery-Lewis** and **Brooks-Draper** diagnostics are shown. The Raftery-Lewis statistic (**Nhat**) is an estimate of the chain length required to estimate a particular quantile to a given accuracy. Here, we need a chain of length 46,890 to estimate the 2.5 percentile and length 30,056 to estimate the 97.5 percentile, i.e. almost 47,000 to estimate a 95% credible interval. The Brooks-Draper statistic is an estimate of the chain length required to estimate the mean of the parameter to k significant figures to a given accuracy. The default value of k is 2, and 27,134 is the estimated chain length to obtain an estimate of β_0 . The Raftery-Lewis and Brooks-Draper diagnostics should be used alongside the other plots to assess convergence. In this case, all diagnostics suggest that a longer chain is required. Diagnostics should be examined for all parameters in turn.

¹ Browne, W.J. (2012) *MCMC Estimation in MLwiN*, v2.25. Centre for Multilevel Modelling, University of Bristol, Chapter 4.

A quick way of doing this (see later) is to **Store** the model estimates (bottom button of the equations window) and this can be used to give the ESS of every parameter in the model. But it is better practice to look at the trajectories and inspect then to see that convergence has been achieved. After all, a lot of effort has been put into the data collection and similar attention is needed for model estimation.

We now turn to inspect the trajectories for the level-2 variance

- Close the MCMC diagnostics window for the intercept and click on the graph for σ_u^2 to view the diagnostics for the community-level variance parameter
- Respond 'yes' to the question 'Calculate MCMC diagnostics?'



The trajectory for σ_u^2 more closely resembles white noise and the effective sample size for this parameter is 770. The MCMC chains are exploring the marginal distribution of this parameter more quickly and accumulating evidence about its distribution. The 95% credible interval for σ_u^2 is (1.216, 1.851); the 2.5% point far exceeds its lowest possible value of zero, so there is strong evidence of differences between communities in antenatal care uptake. The distribution is slightly positively skewed (the mean of 1.51 is larger than the 50% quartile, which is the median of 1.502 which is larger than the mode of 1.493). The floor of zero is a long way from the centre of this distribution. Careful inspection of the trace plot will show some further evidence of skewness; there is the occasional spike in the trajectories, and the smoothed histogram extends out to in excess of 2.4. In many situations the posterior distribution of the variance will be more skewed than this, especially when the lower tail of the distribution approaches zero.

We can also assess the importance of community effects using the Deviance Information Criterion (**DIC**), a likelihood-based measure for comparing non-nested models and a generalisation of Akaike's Information Criterion (AIC). As with the AIC, the DIC combines goodness of fit with model complexity (the number of parameters), so that DIC values for different models can be compared directly. The model with the lowest DIC is deemed the 'best' model. Unlike the AIC, the complexity of the model - the 'degrees of freedom consumed in the fit' is estimated in the DIC case

and can be a non-integer value. Further details of the DIC can be found in Spiegelhalter *et al* (2002) and in the MCMC manual, Chapter 3.²

To calculate the DIC:

- From the **Model** menu, select **MCMC** then **DIC Diagnostic**
- The DIC value will appear in the **Output** window

Bayesian Deviance Information Criterion (DIC)			
Dbar	D(thetabar)	pD	DIC
6047.28	5756.90	290.38	6337.65

To assess whether a multilevel is needed, we will compare the DIC for the current model with the DIC for the single-level model (with the same set of predictors).

- Close all windows except for the **Equations** window
- click on **Estimation control**, then **IGLS/RIGLS** followed by **Done** - as the model has been changed we have to revert to IGLS/RIGLS to get starting values for the MCMC procedures; this is always the case as you need updated starting values for the MCMC chains
- Remove the random effect from the intercept by clicking on **cons** and un-checking **j(comm)**. Click **Done**
- Click **Start** to fit the model
- To re-estimate the model using MCMC, click **Estimation control**, then **MCMC** followed by **Done**
- Click to **Start** again
- When estimation stops, go to the **Model** menu, select **MCMC** then **DIC Diagnostic**
- The DIC value for the single-level model will appear in the **Output** window

The DIC values for the multilevel and single-level models are shown below. Including community random effects leads to a very large reduction in the DIC value of 1100 (from 7437 to 6337), suggesting that the random effects model is the better model. Any reduction the DIC is an improvement but with small differences in the DIC, say less than 2, you cannot really distinguish between models.

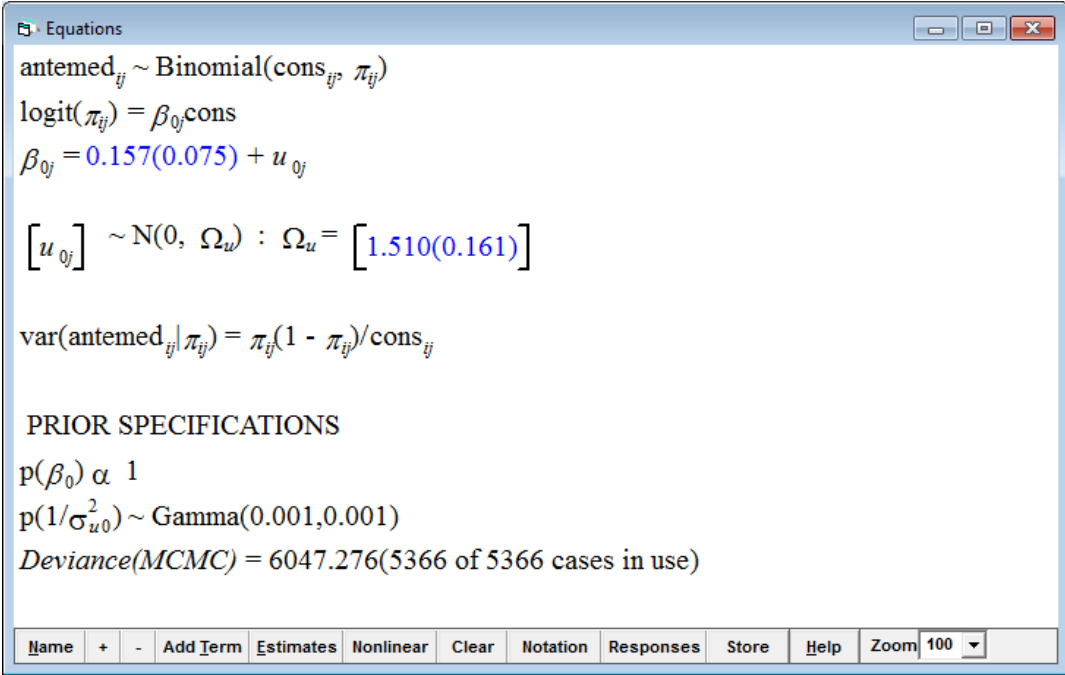
Bayesian Deviance Information Criterion (DIC)			
Dbar	D(thetabar)	pD	DIC
6047.28	5756.90	290.38	6337.65
7436.20	7435.20	0.99	7437.19

Note that the single level model has a **pD** (the estimated complexity) of close to 1 as only a single parameter is being estimated - the intercept. The level-1 variance is simply a function of that the mean. However, the introduction of the community level variance leads to a **pD** of 290.38. There are 361 communities, but when the

² Spiegelhalter, DJ.; Best, Nicola G.; Carlin, BP.; van der Linde, Angelika (2002) Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society, Series B* 64 (4): 583-639.

community differentials are conceived as random effects there are effectively only 290 or so parameters as these effects are estimated as coming from an overall distribution. The nominal 361 parameters are shrunk due to sharing a distribution and therefore do not contribute whole parameters to the parameter count. Despite this considerable extra complexity (moving from 1 to 290 terms), there is still a very large reduction in the penalized DIC. The level-2 variance term is needed to model antenatal care; communities are very different in their antenatal experience.

Now re-introduce the community random effects and refit the model using MCMC (starting with MQL1).



```

antemedij ~ Binomial(consij, πij)
logit(πij) = β0jcons
β0j = 0.157(0.075) + u0j

[u0j] ~ N(0, Ωu) : Ωu = [1.510(0.161)]

var(antemedij|πij) = πij(1 - πij)/consij

PRIOR SPECIFICATIONS
p(β0) ∝ 1
p(1/σu02) ~ Gamma(0.001,0.001)
Deviance(MCMC) = 6047.276(5366 of 5366 cases in use)

```

4 Interpretation of the Null Two-level Model

There are a number of ways of interpreting the size of the random effects in a logit model and will examine a variety of them here.

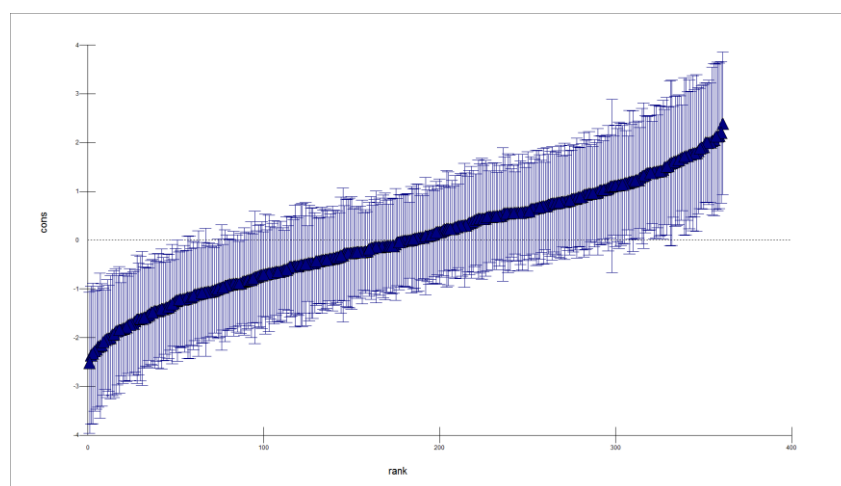
Visualised via a ‘Caterpillar plot’ of the logits

We begin by examining the community residuals on a logit scale. From the above MCMC estimates, we can say that the log-odds of receiving antenatal care from a medically-trained provider in an ‘average’ community (one with $u_{0j} = 0$) is estimated as $\hat{\beta}_0 = 0.157$. The intercept for community j is $0.157 + u_{0j}$, where the variance of u_{0j} is estimated as $\hat{\sigma}_u^2 = 1.510$.

We will now examine estimates of the community effects or residuals, \hat{u}_{0j} , obtained from the null model. To calculate the residuals and produce a ‘caterpillar plot’ with

the community effects shown in rank order together with 95% confidence intervals (these are not derived from the MCMC chains):³

- From the **Model** menu, select **Residuals**
- At the bottom of the **Residuals** window, change level from **1:womid** to **2:comm**
- In the box to the left of **SD(comparative) of residual to**, edit **1.0** to **1.96**
- Click **Calc**
- Click on the **Plots** tab and check **residual +/- 1.96 sd x rank** (the 3rd option)
- Click **Apply**



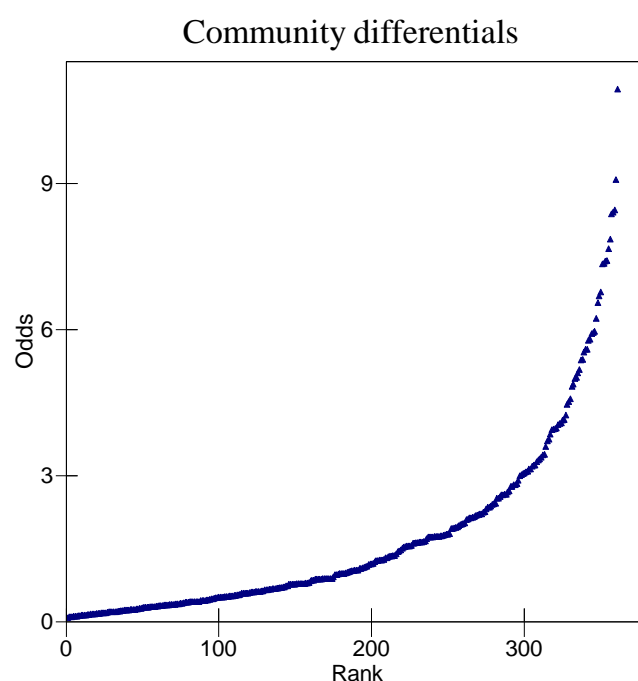
The plot (which automatically will have gone to graphic window **D10**) shows the estimated residuals for all 361 communities in the sample. For a substantial number of communities, the 95% confidence interval does not overlap the horizontal line at zero, indicating that uptake of antenatal care in these communities is significantly above average (above the zero line) or below average (below the zero line). For each community, the confidence intervals are quite wide. This is because the sample size within a community is relatively small (with a maximum of 25 women in a community; remember the results from the **Hierarchy Viewer**) leading to quite large standard errors for the estimated community residuals, \hat{u}_{0j} . The differences between communities are however large and we will now consider a number of other ways to get an appreciation of how large these differences are.

Visualised via a 'Caterpillar' plot of the odds

³ It is possible to store the residuals from the chains. These can require a lot of storage if you perform a long monitoring chain and there are many higher level residuals. You can store every 1 in 10 (say) of the chain by choosing **thinning** on the **Estimation Control** window. The summary statistics shown in the **Equations** window are unaffected by this thinning. However, the storage of residuals is controlled by this thinning and so are the stored estimates (in column **c1090**). For more detail on thinning see the *MCMC Manual* (page 25), and detail on stored estimates (see section 4.9) and stored residuals (see section 4.6). The point of storing the estimates and residuals is that you can calculate a function of them such as the rank for the community and 95% credible intervals.

The first way is to redo the caterpillar plot but this time, plot the odds and not the logits. The easiest ways of doing this is plot the logits without the confidence intervals, and then transform the logits to odds.

- From the **Model** menu, select **Residuals**
- At the bottom of the **Residuals** window, change **level** from **1:womid** to **2:comm**
- Click **Calc** - the logit residuals are stored in c300
- Click on the **Plots** tab and check **residual x rank** (the 2nd option)
- Click **Apply**
- In **Command interface** found under **Data manipulation**, type the command in the lower narrow box at the bottom - **Calc c300 = expo(c300)**, and then press **Enter** so as to exponentiate the logits and turn them into relative odds; the graph should automatically update to produce the following plot (with titles modified)



The logit values are differentials from a value of zero representing the average across all women and communities ($\hat{\beta}_0$). When exponentiated this average value becomes 1 which is the base for the relative odds. The best-served community therefore from inspecting the graph has an odds that is over 9 times better than the overall experience. At the other end of the scale, the poorest performing areas have very low odds. These community differentials are very large.

Interpreted as coverage intervals on the probability scale

A second way of getting some feel for the size of these community effects is to calculate coverage intervals and to do so on the probability scale with which people are more generally familiar. Coverage intervals (not to be confused with confidence

intervals and credible intervals) use the mean and variance at a particular level and the Normality assumption for the differentials, to estimate what is the 95 coverage of possible observations at that level. That is we can estimate what is the interval such that it covers 95% of the community distribution with 2.5% of places in the lowest and highest tails. MLwiN has a facility to do this called **Customised predictions** which simulates from a Normal distribution with a mean of 0.157 and a variance of 1.510 and then converts the resultant logits to the probability scale.

- From the **Model** menu, select **Customised Predictions**
- In the **setup tab** tick on **Medians, Means and Coverage**, leaving all the other choices at their default values, so that 95% coverage intervals will be calculated on the probability scale on the basis of simulations; you will also see the columns where the results will be stored.
- Click **Fill Grid** at the bottom of the **Customised Predictions** window
- Click on the **Predictions** tab. The predictions table will appear. Click **Predict** to calculate the predictions (this is the part done by simulation)
- The predicted probabilities are stored in columns named for example **mean.pred**, with lower and upper limits of a 95% confidence interval for the predictions stored in **mean.low.pred** and **mean.high.pred**. You may need to increase the width of some columns to see their full contents. These values can be copied and pasted to a word-processor.

The **Customised Prediction** after **Fill grid** has been pressed is given below. The underlying procedure needed to simulate the population-average values is as follows:

- i) Generate M values for random effect u from $N(0, \hat{\sigma}_u^2)$ and denote the generated values by $u^{(1)}, u^{(2)}, \dots, u^{(M)}$
- ii) For each simulated value ($m = 1, \dots, M$) compute the estimate probabilities

$$\pi^{(m)} = \frac{\exp(\beta_0 + u^{(m)})}{1 + \exp(\beta_0 + u^{(m)})}$$

- iii) Calculate the mean of the probabilities computed in ii):

$$\pi = \frac{1}{M} \sum_{m=1}^M \pi^{(m)}$$

Customised predictions

Setup

Change Range

cons

Summary: 1

predicted cases: 1

draws from cov(Beta): 2000

nested draws from cov(u): 1000

simulations: 1000

Confidence Interval: 95

☒ Probabilities ☐ logit

☐ Differences

Predictions to:

☒ Medians

Median: c14 Low: c15 Upper: c16

☒ Means

Mean: c17 Low: c18 Upper: c19

Coverage

☒ Level 2 (comm) coverage interval: 95 Low: c20 Upper: c21

Fill Grid Predict Plot Grid

Here are the results of the simulation (they could differ slightly from these as they are a simulation)

cons.pred	median.pred	median.low.pred	median.high.pred
1	0.53981429	0.50210524	0.57549739

mean.pred	mean.low.pred	mean.high.pred	cover_low(l2).pred	cover_high(l2).pred
0.52841687	0.50052226	0.55703753	0.10613664	0.93386734

The **cons.pred** gives the value of the predictor we are using and this value of 1 for the constant signifies the estimate of $\hat{\beta}_0$. In the median community, some 53.9 percent of women are predicted to have used antenatal care, and the 95% confidence intervals around this value are from 50.2 to 57.5. You may think that these are rather wide given a sample of over 5000 women, but you must keep in mind this are an inference about the typical community and not a typical women. These median values are the cluster-specific estimates.

The next three values pertain to the population average results: in the mean community, the best estimate of uptake is 52.8% with a 95% confidence interval of 50.1 to 55.7 percent. It is interesting that even with a large cluster variance of

1.510, the differences between the population average and cluster-specific estimates are not huge. The final two values in the table give the 95% coverage intervals which are very large: in the community in the lowest 2.5% of the distribution only some 10.6 percent of women have accessed care; while the upper 2.5% cut-off is 93 percent. The coverage therefore extends from nearly no one to nearly everybody. The geography of care is very marked. More details on how the **Customised Prediction** facility works are given in the *Manual Supplement*.⁴

Interpreted as Variance Partitioning Coefficient

A third way of getting some feel for the size of these community effects is to calculate a Variance Partitioning Coefficient that provides an estimate of how much of the total unexplained variance (around the overall mean) lies at the community level. We will start by obtaining the estimate of the level 2 variance and the 95% credible intervals of this estimate.

- In the **Equations** window, click on **Store** in the bottom tool bar, name this stored model **Null RI**, then **OK**; remember this is just storing the model estimates for this particular model, and is not saving a worksheet and model specifications
- From the **Model** menu, select **Manage stored models** and highlight **Null RI** and tick on **Include extended information**, and select **Compare**. This will bring up the following tables of results which can be copied to a word processor.

	Null RI	S.E.	Corr	Median	CI(2.5%)	CI(97.5%)	ESS	Bayesian-p
Response	Antemed							
Fixed Part								
Cons	0.157	0.075		0.155	0.008	0.297	120	0.021
Random Part								
Level: comm								
cons/cons	1.510	0.161	1.000	1.502	1.216	1.851	770	
DIC:	6337.655							
pD:	290.379							
Units: comm	361							
Units: womid	5366							

We have removed the Binomial weight estimate *bcons* and $-2 \times \log\text{-likelihood}$ as you would do when publishing the results.

We postpone a detailed discussion of the elements in this table until later when we have added in some explanatory variables. For now we just look at the community level random intercepts variance which is on the logit scale and is given by line starting with cons/cons. The variance is estimated to be 1.51 and the 95% credible intervals are 1.216 and 1.851. The VPC in the two-level Normal-theory random intercepts model is readily calculated as:

⁴ Rasbash, J., Charlton, C., Jones, K. and Pillinger, R. (2012) *Manual Supplement to MLwiN v2.26*. Centre for Multilevel Modelling, University of Bristol.

$$VPC = \frac{\text{Level 2 Variance}}{\text{Level 2 Variance} + \text{Level 1 Variance}}$$

But this not straightforward in the Bernoulli model as the level 1 variance is not estimated but constrained. Moreover, the level-1 variance depends on underlying probability so that it is not a single constant value. Two methods can be used to overcome this. The simple and most used method is to treat the level-1, between individual variations as having a variance of a standard logistic distribution which is an unchanging value, 3.29.⁵ Consequently, the VPC will be given by:

$$VPC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + 3.29}$$

Therefore in the **Command interface** (you will find it under **Data Manipulation**) we calculate the VPC and its credible intervals by 'plugging in' the estimate and 95% intervals obtained from the MCMC chains

```
calc b1 = 1.510/( 1.510 + 3.29)
0.31458
```

```
calc b1 = 1.216/( 1.216 + 3.29)
0.26986
```

```
calc b1 = 1.851/( 1.851 + 3.29)
0.36005
```

Some thirty percent of the overall variance lies at the community level, confirming again that community differences are very substantial in gaining access to antenatal care. While this simple method is commonly used in practice it is approximate and takes no account of the variance changing with the mean. An alternative method developed by Browne *et al* (2005) is a simulation procedure that overcomes both these problems and is implemented as an MLwiN macro which comes with the software.⁶

Interpreted as Median Odds Ratios

A fourth and final way of getting some feel for the size of these community effects is to use Larsen's Median Odds Ratio.⁷ The MOR transforms the variance on the logit scale to a much more interpretable odds scale than can be compared to the relative

⁵If you fitted a single level model with no community variance the median estimate on a probability would be 51.2 and you would get the narrower 95% confidence intervals 50.0 to 52.6. But that would ignore that you are dealing with a clustered sample.

⁶ Browne WJ, Subramanian S V, Jones K, Goldstein H. (2005) Variance partitioning in multilevel logistic models that exhibit over dispersion *Journal of Royal Statistical Society A*, 168(3) 599-613. This is implemented in a macro that comes with the MLwiN software: VPC.txt.

⁷ Larsen K, and Merlo J. (2005) Appropriate assessment of neighbourhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression. *Am J Epidemiol*, 161,81-8; Merlo J, Chaix B, Yang M, Lynch J, Rastam L. (2005) A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health*, 59(6):443-9.

odds ratio for terms in the fixed part of the model (see later). MOR can be conceptualised as the increased odds (on average, hence the median) that would result from moving from a lower to a higher 'risk' area if two areas were chosen at random from the distribution with the estimated level 2 variance. The formula is as follows:

$$MOR = \exp[\sqrt{2 * \sigma_{u0}^2 * \Phi^{-1}(0.75)}]$$

$$MOR = \exp[\sqrt{2 * \sigma_{u0}^2 * 0.6745}]$$

$$MOR \approx \exp[0.95 \sqrt{\sigma_{u0}^2}]$$

Where σ_{u0}^2 is the level 2 between community variance on the logit scale (this would be replaced with a variance function if random slopes are involved); and $\Phi^{-1}(0.75)$ is the 75th percentile of the cumulative distribution function of the Normal distribution with mean 0 and variance 1. The credible intervals for a MOR can be obtained by 'plugging in' the credible intervals obtained from an MCMC run of the level 2 variance. Using the **Command Interface**

```
calc b1 = expo( (2 * 1.510 * 0.6745)^0.5)
4.1671
```

```
calc b1 = expo( (2 * 1.216 * 0.6745)^0.5)
3.5994
```

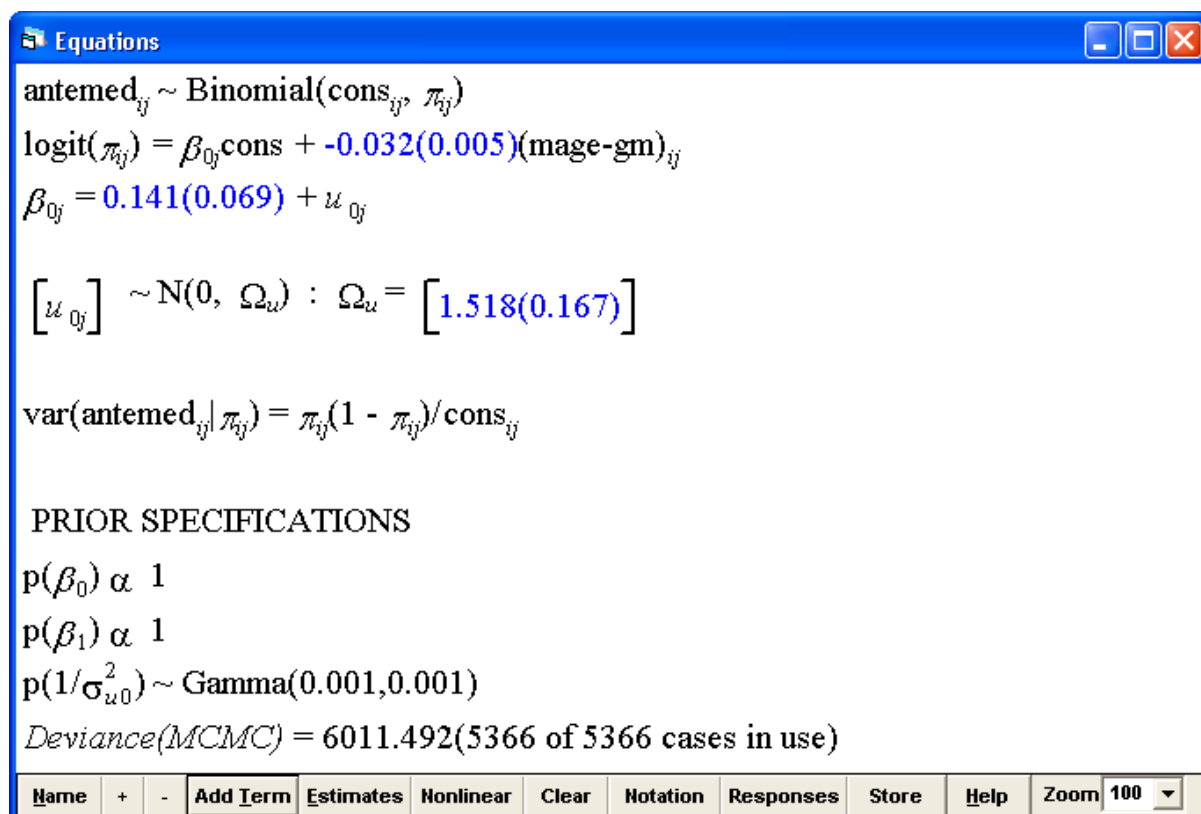
```
calc b1 = expo( (2 * 1.851 * 0.6745)^0.5)
4.8559
```

Again these are very substantial effects, and we will subsequently compare them with the effects of explanatory variables included in the model.

5 Adding Explanatory Variables

Next we include maternal age as an explanatory variable in the model, starting with a linear age effect.

- Before changing the model, we need to switch back to the IGLS estimation method. Click **Estimation control** and then **IGLS/RIGLS**, followed by **Done**
- In the **Equations** window, click **Add Term**
- From the variable drop-down list, select **mage**. Under **centring**, check **grand mean** to include mean-centred **mage**. Click **Done**
- Click **Start** to fit the model
- On convergence, click on **Estimation control** and then **MCMC**, followed by **Done**
- Finally click **Start** to initiate the chains



Note that the addition of age leads to little change in the estimate of the between-community variance, suggesting that the distribution of maternal age is similar across communities.

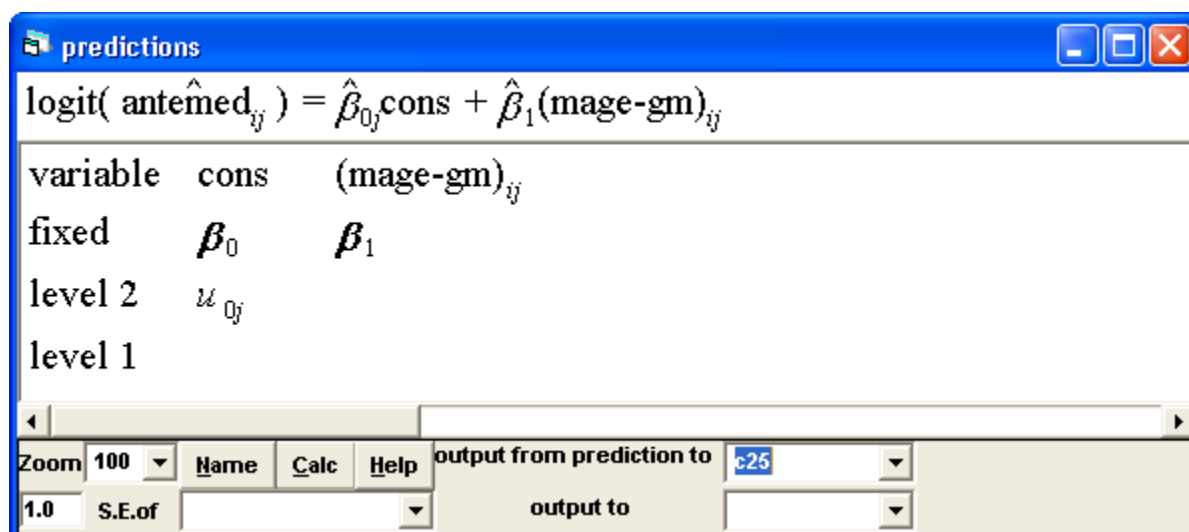
The equation of the average fitted regression line, expressing the relationship between the log-odds of receiving antenatal care and maternal age is:

$$\log\left(\frac{\hat{\pi}_{ij}}{1 - \hat{\pi}_{ij}}\right) = 0.141 - 0.032(\text{mage} - \text{gm})_{ij}$$

The fitted line for a given community will differ from the average line in its intercept, by an amount \hat{u}_{0j} for community j . A plot of the predicted community lines will therefore show a set of parallel lines on the logit scale. To produce this plot, we first need to calculate the predicted log-odds of antenatal care for each woman, based on her age at survey and community of residence.

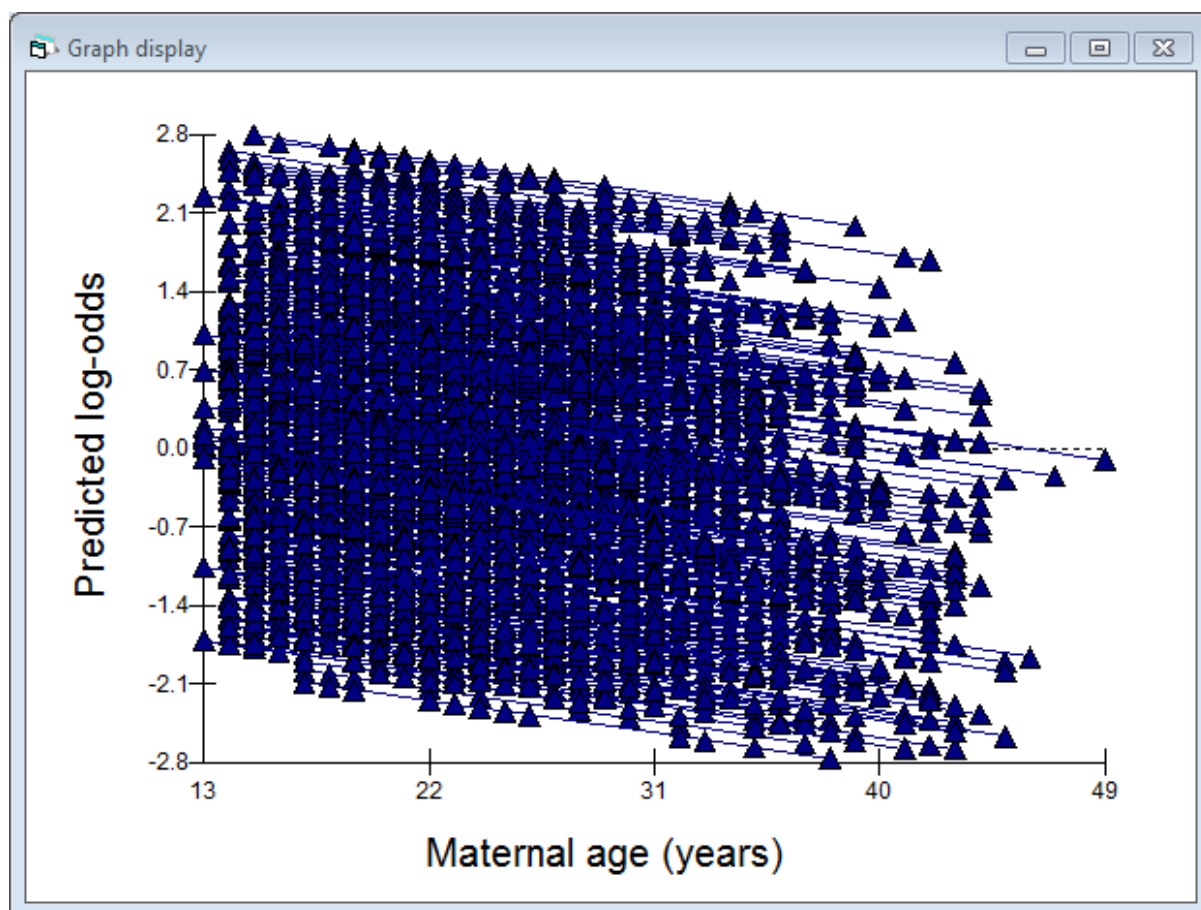
- From the **Model** menu, select **Predictions**; unlike the **Customised predictions** used earlier; this facility will make predictions for all level 1 units, that is every women in the worksheet
- Next to **fixed** click on β_0 and β_1
- Next to **level 2** click on u_{0j}
- Next to **output from prediction to**, select an empty column such as c25

Check that the **Predictions** window looks like the one below. In particular, note that the intercept β_0 has a j subscript which means that the prediction for a given individual will include the community-specific component \hat{u}_{0j} .

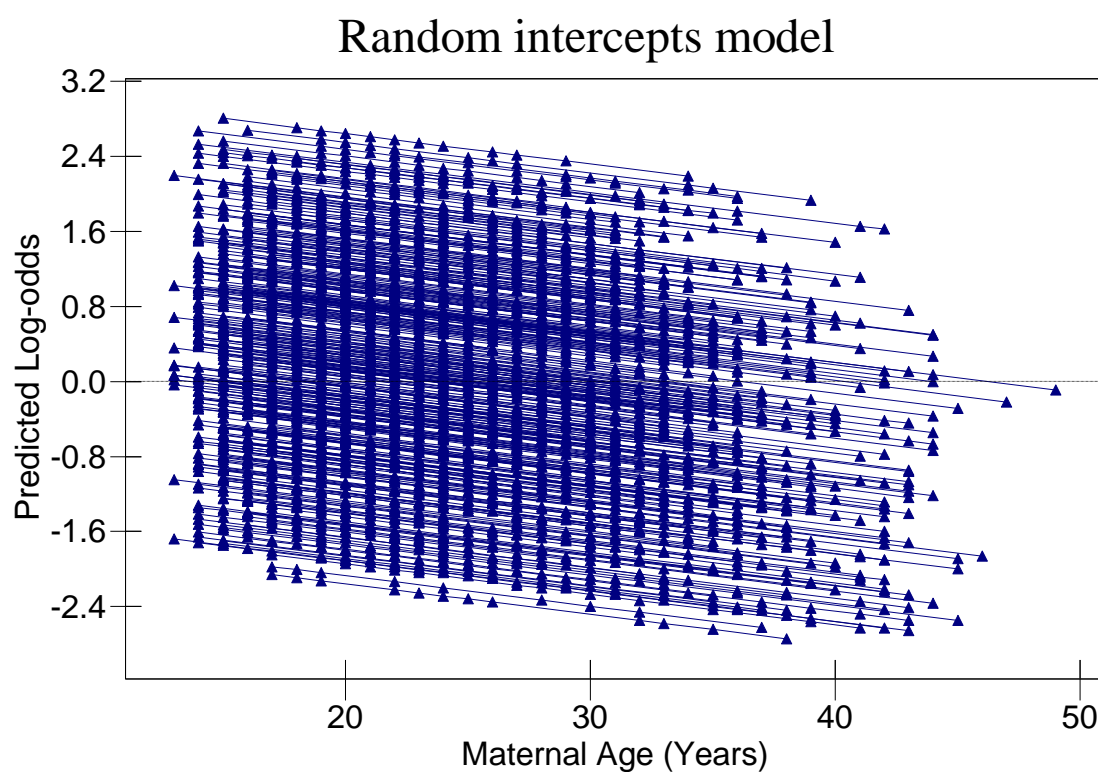


- Click **Calc**
- Go to the **Names** window and name c25 **pred_ri** (for 'prediction from random intercept' model)
- To plot the predicted community lines, go to the **Graphs** menu and select **Customised Graph(s)**
- From the drop-down list at the top left of the **Customised Graph** window, change **D10** to **D1** (or any empty **Display** with graph options unspecified)
- Next to **plot type**, select **line+point**
- Next to **y**, select **pred_ri**
- Next to **x**, select **mage**
- Next to **group**, select **comm**
- Click **Apply**
- As usual, titles may be added (as has been done below) by left - clicking anywhere on the plot and then on the **Titles** tab
- Right - clicking on the plot allows the export of high-quality graphics that can then be pasted to Word; choose the export dialog which brings up another window. The default is to send the image to the clipboard with a size of 152 by 101 millimetres in the form of an EMF. The latter stands for Enhanced Meta File so that the image is exported not as a pixelated bitmap but as a vector graphics which can be edited and changed.

Here is the bitmap version

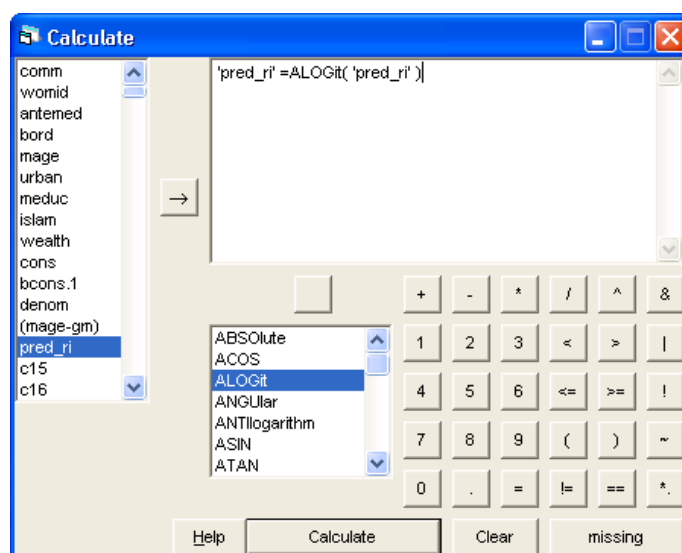


Followed by the vector graphics version



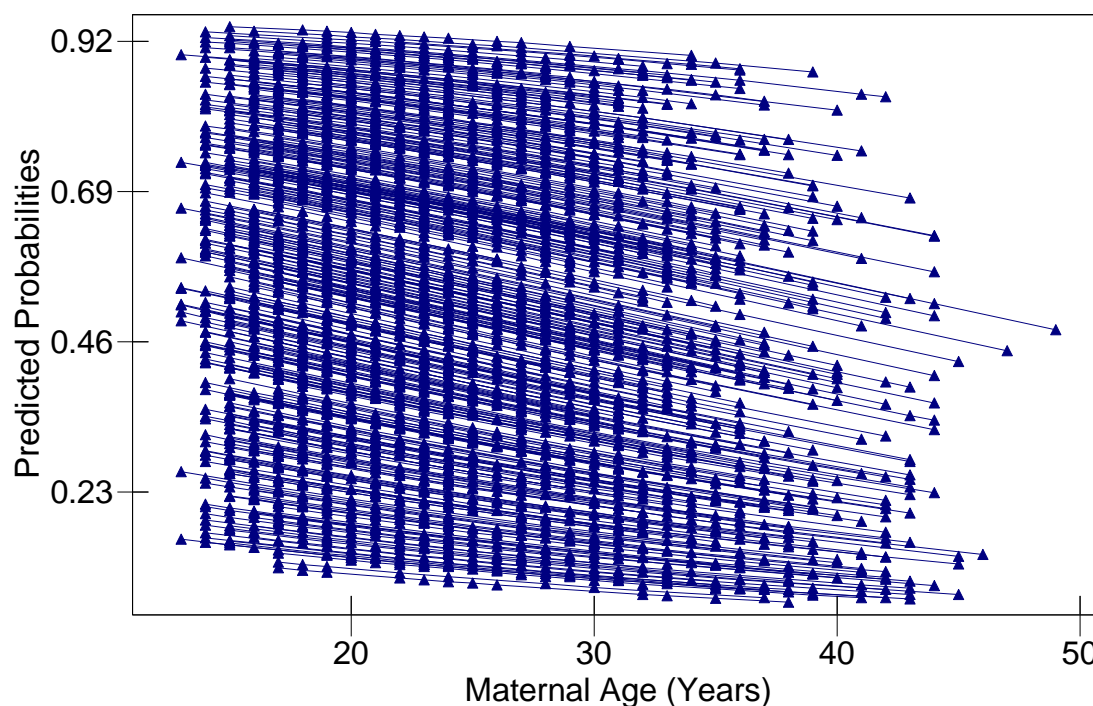
For a woman aged 22, the log-odds of receiving antenatal care range from about -2.2 to 2.5 depending on which community she lives in. This translates to a range in probabilities of $\exp(-2.2)/[1+\exp(-2.2)] = 0.10$ to $\exp(2.5)/[1+\exp(2.5)] = 0.92$, so there are strong community effects. To convert the predictions to probabilities

- From the **Data Manipulation** menu, select **Calculate**
- And specify the calculation (**ALOGit**) as follows so that the probabilities are stored back in **pred_ri** and overwrite the logits



The graph in **D1** will be updated automatically but the y-axis title will have to be changed to probabilities and not logits. The very large differences between communities are again very obvious.

Random intercepts model



Before adding further variables to the model, it is useful to store the model so that we can see how estimates change as we add in extra terms.

- In the equations window, click on **Store** in the bottom tool bar, called this stored model **Mage RI**, then **OK**;
- From the **Model** menu, select **Manage stored models** and highlight **Mage RI** and tick on **Include extended information**, and select **Compare** this will bring up the following table of results which can be copied to a word processor.

	Mage RI	S.E.	Corr	Median	CI(2.5%)	CI(97.5%)	ESS	Bayesian-p
Response	antemed							
Fixed Part								
Cons	0.141	0.069		0.140	0.005	0.276	134	0.022
(mage-gm)	-0.032	0.005		-0.032	-0.043	-0.022	1063	0.000
Random Part								
Level: comm								
cons/cons	1.518	0.167	1.000	1.509	1.212	1.866	809	
Level: womid								
bcons.1/bcons.1	1.000	0.000		1.000	1.000	1.000	0	
-2*loglikelihood								
DIC:	6302.464							
pD:	290.972							
Units: comm	361							
Units: womid	5366							

This gives a number of items that we have already discussed plus some new things. The bottom row give the structure of the model with 5366 women in 361 communities- these can change if there are missing values in predictors as any missing values for a variable result in that case being listwise deleted from the model. The DIC and pD are then given and these can be compared to the results from the previous null or empty two-level model. As we might expect the degrees of freedom have gone up by around one from 290 to 291, while the DIC has dropped quite considerably from 6337.65 to 6302.5. We really need to include maternal age in the model; and as the coefficient is negative, older women have on average received less antenatal care. The deviance (-2*loglikelihood) is not given as it is not reliable with these models. The level-1, women-level variance (bcons.1/bcons.1) is just a technical fix to fit the Bernoulli variance and does not need including when you publish papers. The level 2 between-community variance (cons/cons) is 1.518 and is based on an ESS of 809; the 95% credible intervals show that there is substantial community variance as the lower interval does not approach zero. The fixed part estimate for maternal age is negative and the 95% credible intervals do not straddle zero suggesting that there is a well-supported negative relation. The Bayesian p value is another by-product of the MCMC procedure. These are based on tail posterior probabilities and because the coefficient is estimated to be negative it gives the proportion of the estimates that are positive. Here the p value is 0.00; so a very small proportion of the estimates are positive. It is best to treat this as potentially useful informal diagnostic which suggests here that the overwhelming weight of evidence is that the relationship is negative. The MCMC chains for Age are relatively uncorrelated and the ESS is over 1000 so suggesting that we have run the

chains for long enough providing that they have converged. However, because the ESS for the intercept is only 134, then we really need to run the entire model for longer. In this model without random slopes the **Correlation** column does not provide any useful information.

We will now extend this model to include two further woman-level predictors: maternal education (**meduc**, with dummies for the 'primary' and 'secondary or higher' categories) and household wealth index (**wealth**, in quintiles and treated as a continuous variable, thereby treating it in a very parsimonious fashion by assuming that there is an underlying linear relationship on the logit scale).

- Switch back to the IGLS estimation method. Click **Estimation control** and then **IGLS/RIGLS**, followed by **Done**
- In the **Equations** window, click **Add Term**
- From the variable drop-down list, select **meduc**. This variable has already been declared as categorical (check the **Names** window to see). By default the first category (which has been named **meduc_1**) is taken as the reference. We will accept the default so click **Done**
- Click **Add Term** again and select **wealth**. Check **grand mean** to centre the variable and get a meaningful intercept, as there is no **wealth** value of 0; then click **Done**
- Click **Start** to fit the model using MQL1
- On convergence, click on **Estimation control** and then **MCMC**, followed by **Done**
- Finally click **Start**; after a while you should get the following

Equations

$$\text{antemed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons}_{ij} + -0.005(0.006)(\text{mage-gm})_{ij} + 0.546(0.087)\text{meduc_2}_{ij} + 1.311(0.104)\text{meduc_3}_{ij} + 0.400(0.030)(\text{wealth-gm})_{ij}$$

$$\beta_{0j} = -0.486(0.085) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.915(0.116)]$$

$$\text{var}(\text{antemed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{cons}_{ij}$$

PRIOR SPECIFICATIONS

$$p(\beta_0) \propto 1$$

$$p(\beta_1) \propto 1$$

$$p(\beta_2) \propto 1$$

$$p(\beta_3) \propto 1$$

$$p(\beta_4) \propto 1$$

$$p(1/\sigma_{u0}^2) \sim \text{Gamma}(0.001, 0.001)$$

Deviance(MCMC) = 5562.193(5366 of 5366 cases in use)

Name + - Add Term Estimates Nonlinear Clear Notation Responses Store Help Zoom 100

- In the equations window, click on **Store** in the bottom tool bar, calling this stored model **+Meduc&Wealth**, then **OK**;
- From the **Model** menu, select **Manage stored models** and highlight **+Med&Wlth** and tick on **Include extended information**, and select **Compare** this will bring up the following tables of results

	+Med&Wlth	S.E.	Corr	Median	2.5%	97.5%	ESS	Bayesian-p
Response	Antemed							
Fixed Part								
Cons	-0.486	0.085		-0.481	-0.656	-0.318	146	0.000
(mage-gm)	-0.005	0.006		-0.005	-0.016	0.007	677	0.201
meduc_2	0.546	0.087		0.544	0.379	0.724	229	0.000
meduc_3	1.311	0.104		1.313	1.114	1.506	225	0.000
(wealth-gm)	0.400	0.030		0.400	0.337	0.456	462	0.000
Random Part								
Level: comm								
cons/cons	0.915	0.116	1.00	0.908	0.709	1.159	489	
DIC:	5815.121							
pD:	252.928							
Units: comm	361							
Units: womid	5366							

This model should be run for longer, due to the relatively low ESS, and we leave that as an exercise for you. In the fixed part of the model, the **Mage** effect has attenuated quite considerably (it used to be -0.032, it is now -0.005) and the credible intervals now span zero and we get positive estimates 20 percent of the time as shown by the Bayesian p values. This suggests that the age effect was in some senses an artefact and that the effect was really due to level of education and wealth. The effects for **meduc** and **wealth** have 95% credible intervals that do not span zero and Bayesian p values of zero suggesting that these are substantive effects. Notice that the addition of **meduc** and **wealth** has also substantially reduced the between-community variance, suggesting that the distribution of one or both variables varies across communities. Clearly some communities have higher proportions of educated women and relatively wealthy households than others. In order to appreciate the relative size of the effects of the fixed part of the model we will use **Customised predictions** facility to predict probabilities for a specific set of predictor values that you can customize.

6 Predicted Probabilities via Customised Predictions

We can calculate ‘population-averaged’ probabilities that average over the values of u_{0j} drawn from a normal distribution with variance equal to the estimated level 2 variance, i.e. $N(0, \hat{\sigma}_u^2)$. The procedure for a model with one predictor x is as follows:

- iv) Generate M values for random effect u from $N(0, \hat{\sigma}_u^2)$ and denote the generated values by $u^{(1)}, u^{(2)}, \dots, u^{(M)}$

- v) For each simulated value ($m = 1, \dots, M$) compute, for a given value of x , the

$$\pi^{(m)} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x + u^{(m)})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x + u^{(m)})}$$

estimate probabilities

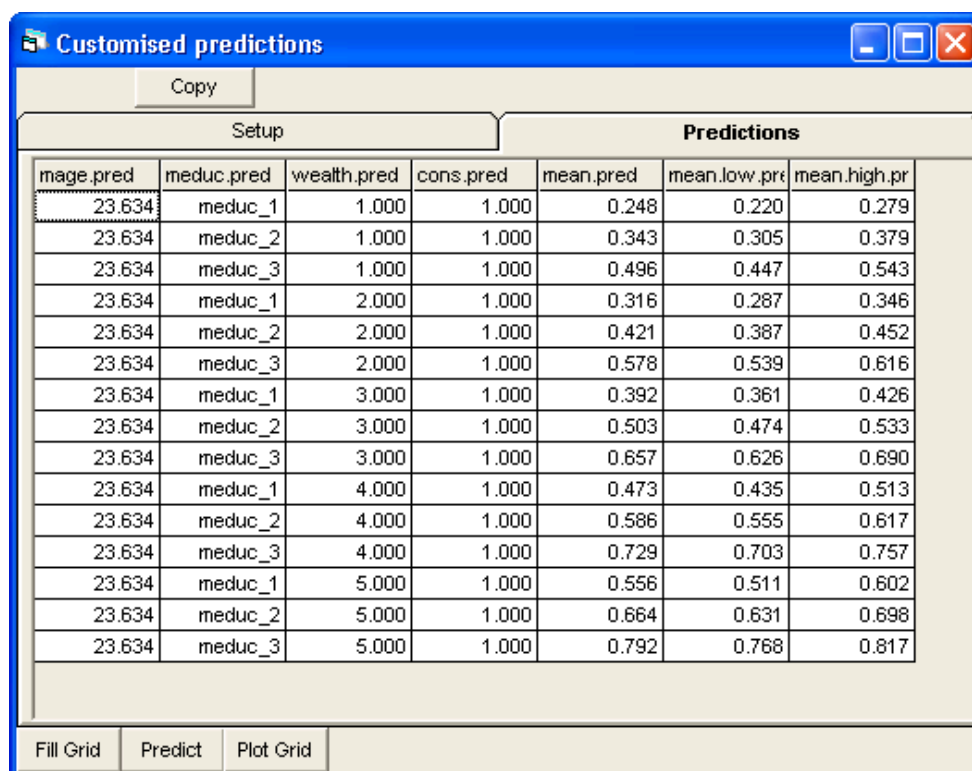
- vi) Calculate the mean of the probabilities computed in ii):

$$\pi = \frac{1}{M} \sum_{m=1}^M \pi^{(m)}$$

Steps i)-iii) can then be repeated for different values of x , so that the prediction is customised.

The above procedure is implemented in MLwiN using **Customised Predictions**.

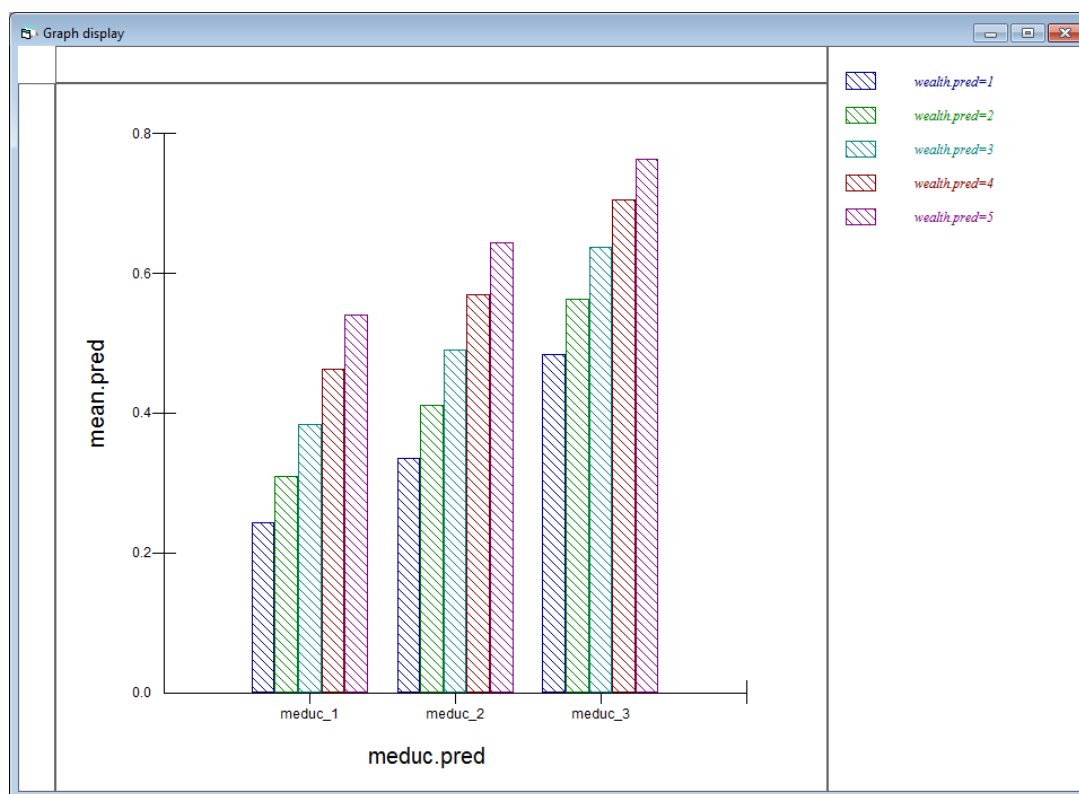
- From the **Model** menu, select **Customised Predictions**
- As the specification of the model has changed, click on **Clear**
- Click on **meduc**. Click on **Change Range** then check **Category**. Check each of **meduc_1**, **meduc_2** and **meduc_3** to obtain a prediction for each education category. Click **Done**
- Click on **wealth**, then **Change Range**. Click on the **Range** tab. Next to **Upper bound**, type **5**. Next to **Lower**, type **1**. Next to **Increment**, type **1**. Click **Done**
- Leave **mage** unchanged so that we will be making predictions for an average age female of **23.6**
- In the main **Customised Prediction** window, under **Predictions to**, check **Means** to get **population average** values [to get cluster specific check **Medians**]
- Click **Fill Grid** at the bottom of the **Customised Predictions** window
- Click on the **Predictions** tab. The predictions table will appear. Notice that there is a separate row for each combination of **meduc** and **wealth**, i.e. $3 \times 5 = 15$ entries. The other variable, **mage**, is set at its mean for each prediction.
- Click **Predict** to calculate the predictions (this is the bit done by simulation)
- The predicted probabilities are stored in a column named **mean.pred**, and lower and upper limits of a 95% confidence interval for the predictions are stored in **mean.low.pred** and **mean.high.pred**. You may need to increase the width of some columns to see their full contents



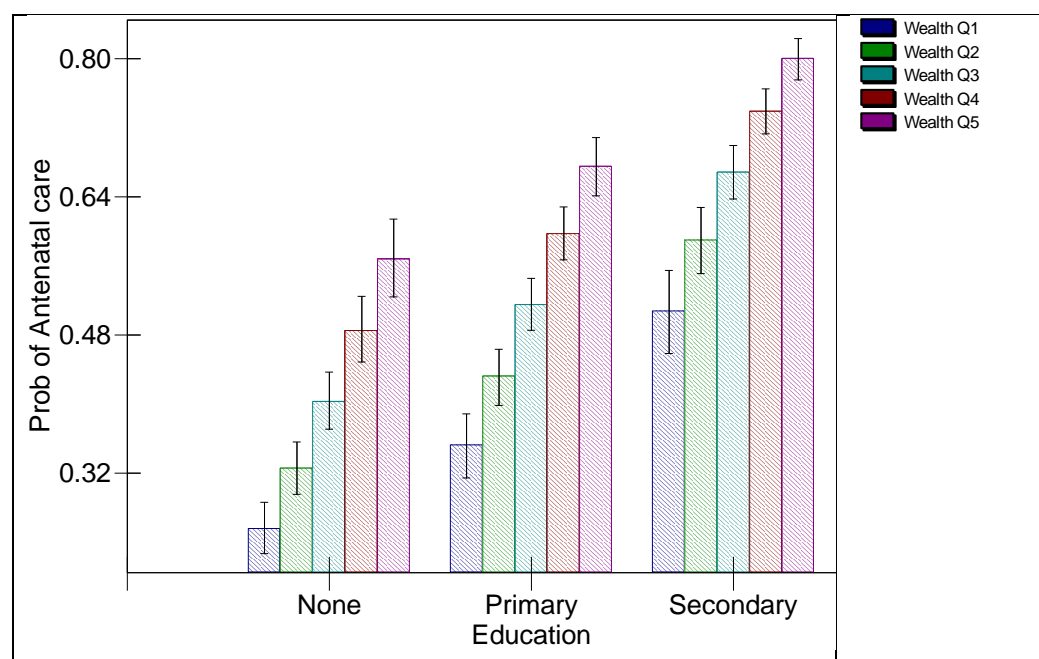
mage.pred	meduc.pred	wealth.pred	cons.pred	mean.pred	mean.low.pr	mean.high.pr
23.634	meduc_1	1.000	1.000	0.248	0.220	0.279
23.634	meduc_2	1.000	1.000	0.343	0.305	0.379
23.634	meduc_3	1.000	1.000	0.496	0.447	0.543
23.634	meduc_1	2.000	1.000	0.316	0.287	0.346
23.634	meduc_2	2.000	1.000	0.421	0.387	0.452
23.634	meduc_3	2.000	1.000	0.578	0.539	0.616
23.634	meduc_1	3.000	1.000	0.392	0.361	0.426
23.634	meduc_2	3.000	1.000	0.503	0.474	0.533
23.634	meduc_3	3.000	1.000	0.657	0.626	0.690
23.634	meduc_1	4.000	1.000	0.473	0.435	0.513
23.634	meduc_2	4.000	1.000	0.586	0.555	0.617
23.634	meduc_3	4.000	1.000	0.729	0.703	0.757
23.634	meduc_1	5.000	1.000	0.556	0.511	0.602
23.634	meduc_2	5.000	1.000	0.664	0.631	0.698
23.634	meduc_3	5.000	1.000	0.792	0.768	0.817

To plot the predicted probability of receiving antenatal care for each category of maternal education:

- Click on **Plot Grid**
- Next to **X:** check **meduc.pred**
- Next to **Y:** check **mean.pred**
- Next to **Grouped by:** check **wealth.pred**
- Click **Apply** (click **OK** if a message about the Graph display being already in use appears) or you could choose the graph Display **D2** so as not overwrite what is in **D1**
- The predictions are plotted as a grouped bar chart because MLwiN recognises **meduc** and **wealth** as categorical variables



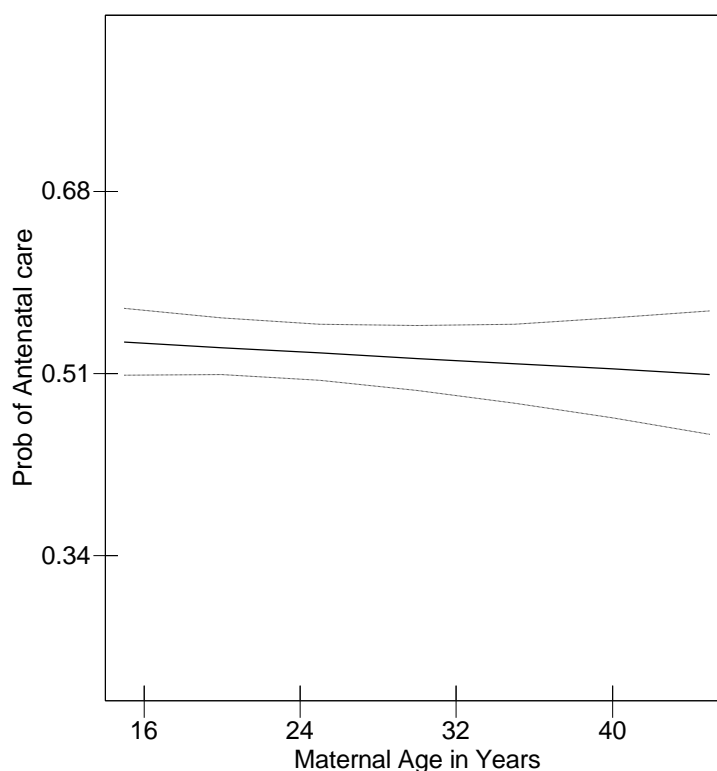
The bar chart shows the predicted probability of receiving antenatal care for each category of maternal education and wealth (for a woman of average age). The highest probability of antenatal care is for women with highest education level and in the wealthiest quintile. Choosing to put on 95% confidence levels (as error bars) and generally tidying up the plot gives the following graph. Note that the effect of education is predicted to be the same for each category of wealth because the model does not allow for an interaction between **meduc** and **wealth** (but see later). The variable **wealth.pred** has been changed to being categorical with the quartile labels.



If we repeat the **customised prediction** plot, but now **Change range of Mage** to range from 15 to 45 in steps of 5 and **Change range of meduc** and **wealth** back to their mean value, we get the following customised predictions

Customised predictions						
Copy						
Setup				Predictions		
mage.pred	meduc.pred	vwealth.pred	cons.pred	mean.pred	mean.low.pred	mean.high.pred
15.000	mean(meduc)	3.008	1.000	0.531	0.500	0.564
20.000	mean(meduc)	3.008	1.000	0.526	0.500	0.553
25.000	mean(meduc)	3.008	1.000	0.521	0.496	0.548
30.000	mean(meduc)	3.008	1.000	0.516	0.487	0.546
35.000	mean(meduc)	3.008	1.000	0.511	0.474	0.548
40.000	mean(meduc)	3.008	1.000	0.506	0.461	0.553
45.000	mean(meduc)	3.008	1.000	0.501	0.445	0.558

Plotting these results with same vertical scale as that for the above graph, we get the following graph with confidence lines as mother's age is a continuous variable



The effect of maternal age is therefore not large in comparisons to wealth and education status; the plot is for a typical person in terms of education and wealth.

7 Predicted Relative odds via Customised Predictions

The **Customised Predictions** has so far been used to predict probabilities; we will now use it to predict odds. This is a three-step procedure. We first choose the base category against which we want to compare the relative odds. We suggest that this is the category with the most negative logit estimates as people find it much easier to compare odds above 1 rather than below 1.⁸ A useful feature of the **Customised predictions** is that we can choose this base category after the model has been estimated. Second for the chosen values of the predictor variable we simulate the differential logits. Thirdly, we then exponentiate these logits to get the required odds.

The above procedure is implemented in MLwiN using **Customised Predictions**.

- From the **Model** menu, select **Customised Predictions**
- Click on **meduc**. Click on **Change Range** then check **Category**. Check each of **meduc_1**, **meduc_2** and **meduc_3** to obtain a prediction for each education category. Click **Done**
- Click on **wealth**, keep the value as the average , which is **3.0082**
- Click on **Age** and set to the average age female of **23.6**
- In the main **Customised Prediction** window, under **Predictions to**, check **Means** to get **population average** values [to get cluster specific, check **Medians**]; tick on **Differences** and for the **From Variable** choose **Meduc** from the dropdown list, choosing **Reference value** from the dropdown list to be to be the lowest logit, that is **meduc_1**
- Click **Fill Grid** at the bottom of the **Customised Predictions** window and then
- Click **Predict** to calculate the predictions (this is done by simulation)
- Click on the **Predictions** tab. The predictions table will appear. Notice that there is a separate row for each category of **meduc** and **wealth**. The other variables, **age**, and **wealth** are set to their respective means for each prediction.
- The predicted differential logits are stored in a column named **mean.pred**, and lower and upper limits of a 95% confidence interval for the predictions are stored in **mean.low.pred** and **mean.high.pred**.

Here are the predicted differential logits for the chosen values of the predictor variable

⁸ A 4 fold increase in odds is given by the relative odds of 4, a fourfold reduction in odds is represented by the relative odds of 0.25.

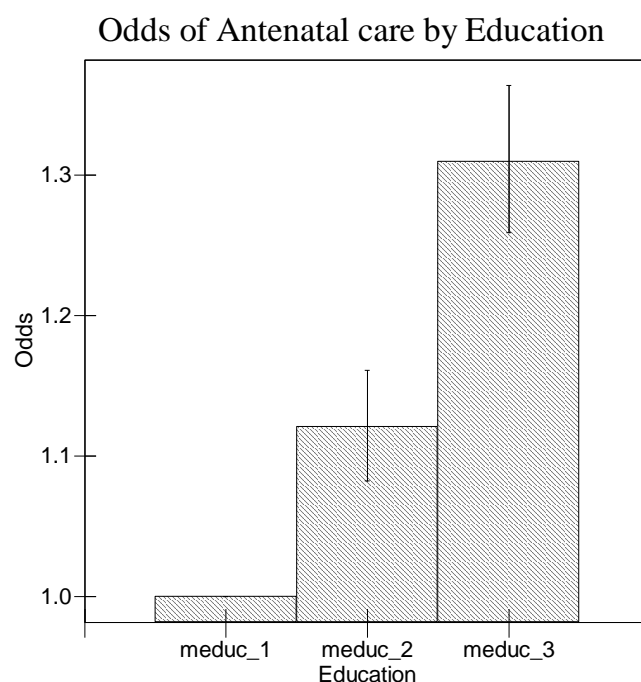
Setup				Predictions		
mage.pred	meduc.pred	wealth.pred	cons.pred	mean.pred	mean.low.pr	mean.high.pr
23.634	meduc_1	3.008	1.000	0.000	0.000	0.000
23.634	meduc_2	3.008	1.000	0.114	0.080	0.147
23.634	meduc_3	3.008	1.000	0.267	0.228	0.307

You can see that the logit for **Meduc_1** has been set to zero and all the other predicted values are logit differential from this. We now need to turn these logits into odds by exponentiation.

- In the **Names** window, find where the logits and 95% confidence intervals are stored, that is **mean.pred**, **mean.low.pred** and **mean.high.pred**, mine are c22, c23 and c29 respectively
- In the lower narrow box of the **Command window**, type
- **Expo c22 c23 c29 c22 c23 c29** so that the exponentiated logits and 95% confidence intervals are stored as odds back in the same columns; when asked to “clear prediction?”, respond **no** so that predictions are overwritten and not cleared. (Unfortunately, the values in the table will not be updated.)

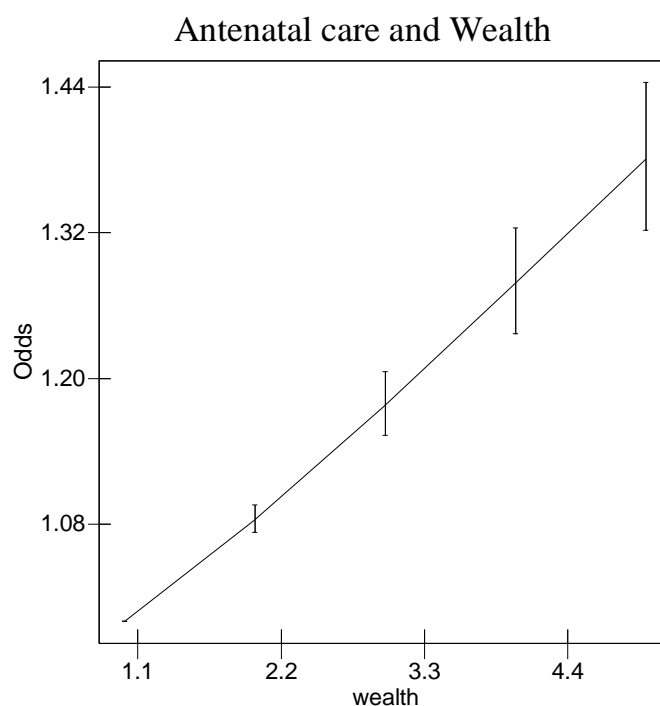
To plot the predicted odds of receiving antenatal care for each category of maternal education:

- Go back to the **Customised predictions** window
- Click on **Plot Grid**
- Next to **X:** check **meduc.pred**
- Next to **Y:** check **mean.pred**
- Tick on **Confidence interval as error bars** as **meduc** is a categorical variable
- Click **Apply** (click **OK** if a message about the Graph display being already in use appears or choose the graph Display **D3** so as not overwrite what is **D1** and **D2**)
- The predictions are plotted as a grouped bar chart because MLwiN recognises **meduc** as a categorical variable



The plot shows the relative odds for the base category of lowest education as set to 1 and no confidence intervals. The other two categories are significantly higher (the CI's do not cross the value 1) with antenatal care being some 10 percent higher those with primary education as compared to no education; and some thirty percent higher for secondary education compared to no education.

The process can be repeated for the quintiles of **Wealth** to get the following plot. Remember to set back the range for **meduc** to its average and change the range for **Wealth** to the values 1 to 5, with reference value set to 1, to calculate differences from the value 1, the lowest quintile.



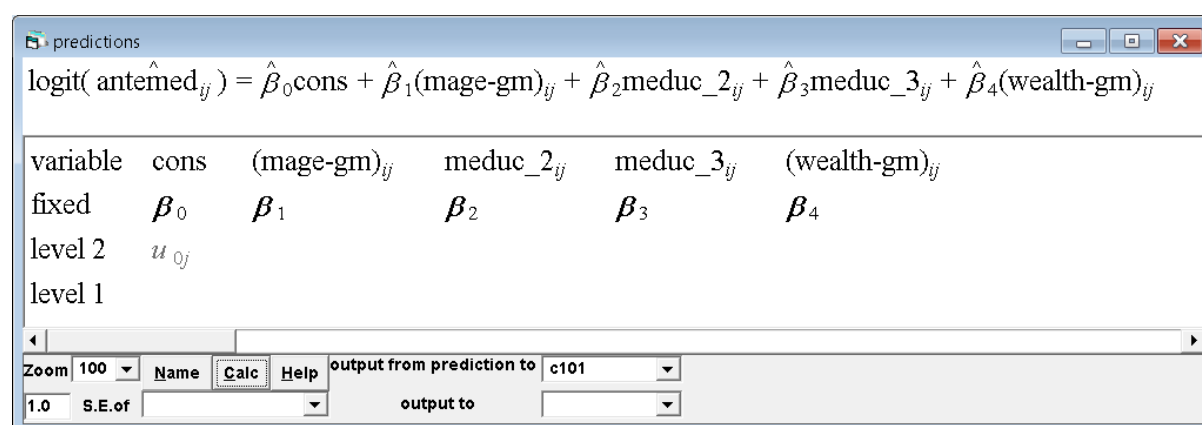
You can see the odds of receiving antenatal care are some 35 percent higher in the highest quintile of individual wealth compared to the lowest. Again these are significant effects but they much smaller than the Median Odds Ratio of over 400 percent considered earlier; the community differences are very much larger than the effects of education and wealth.

8 Estimating a pseudo R-squared

Snijders and Bosker (2012, equation 14.21) have proposed the following procedure for calculating a R^2 for the binary outcome model.⁹

1. Having estimated the model, use Predictions to calculate the predictions based only on the fixed part of the model (not the *Customised predictions* but for every level 1 unit in the dataset; and including no random effects))
2. Calculate the variance of these fixed-part predictions; this is the required 'explained' variance.
3. Calculate the ratio of this explained variance to the sum of the explained variance plus the unexplained variance from level 1 and the higher-level random parts.

All three elements have to be on the same scale, which here is the logit. The predictions based on all the estimated fixed coefficients are stored in c101 (notice that u_{0j} is greyed out):



In the command window; type

AVERage 'c101'

To get the following summary statistics.

⁹ Snijders, T. A. B. & Bosker, R. J. (2012) *Multilevel Analysis. Second edition*, London: Sage

	N	Missing	Mean	s.d.
c101	5366	0	0.13355	0.98139

Return to the command interface and calculate the variance as the square of the standard deviation and store in the box, b1

```
calc b1 = 0.98139 * 0.98139
```

Finally return to the command interface and calculate the R^2 with the level-2 estimated variance as 0.915, and 3.29 for the level-1 unexplained variance

```
calc b2 = (b1 / (b1 + 0.915 + 3.29) ) * 100
```

18.64

So some 19% of the variation in antenatal uptake can be accounted for. As Snijders and Bosker(2012,226) aver such estimates are typically “considerably lower” than for continuous outcomes. You also have to remember the level 1 variance of 3.29 cannot reduce even when important fixed-part variables are included in the model, and we turn to this issue in more detail in the next section.

Another easily calculated R^2 has been proposed by Tjur(2009) but it has no multilevel pedigree whatsoever; it simply works on the predictive power of the fixed part¹⁰ It consists of four stages.

1. Calculate the predicted fixed part values on the logit scale as above.
2. Use the alogit transformation to turn these into probabilities.
3. For each of the two categories of the dependent variable, calculate the mean of the predicted probabilities of an event.
4. Take the difference between the two means as the measure of predictive success

The intuition is clear - if the model makes good predictions, the cases with events should have high predicted values and the cases without events should have low predicted values. On the basis of the stored predictions on the logit scale in c101, use the Command interface to calculate the predicted probabilities and store in c102

```
calc c102= alogit(c101)
```

Again in the command window; calculate the mean for the two binary categories of 'antemed'

¹⁰ Tjur, T. (2009) “Coefficients of determination in logistic regression models—A new proposal: The coefficient of discrimination.” *The American Statistician* 63: 366-372; see also <http://www.statisticalhorizons.com/r2logistic#comments>

TABulate 'c102' 'antemed'

Which gives the summary statistics as follows:

Variable tabulated is c102

	0	1	TOTALS
N	2613	2753	5366
MEANS	0.429	0.618	0.526
SD'S	0.186	0.201	0.194

Finally calculate the absolute difference in the means

$$\text{Calc } b4 = (0.618 - 0.429) * 100$$

So that on this basis the fixed part estimates have a coefficient of determination of 18.900 which is very similar to the Snijder and Bosker procedure.

9 Returning to the size of the random effects

We have seen that the level 2 variance representing the unexplained between community variation was 1.511 when **mage** was included in the model, but reduced to 0.915 when **meduc** and **wealth** were additionally included. We can recalculate the variety of statistics discussed earlier to help interpret this level 2 variance. Thus, the VPC is now:

$$\text{calc } b1 = 0.915 / (0.915 + 3.29) \\ 0.21760$$

so that some 22 percent of the unexplained variation is at the community level.

And the MOR is:

$$\text{calc } b1 = \text{expo}(2 * 0.915 * 0.6745)^{0.5} \\ 3.0374$$

There is still a large threefold difference in the odds between randomly chosen high and low communities. This all suggests that there are still considerable differences between communities even when we have taken into account age, education and wealth.

We need, however, to reflect carefully when comparing the higher-level variances from different models as explanatory variables are included. Before we consider discrete outcome binomial models, it is useful to summarize what normally happens in Normal-theory models for the case of a two-level model.

- If a predictor variable is measured at level 2, it can only explain the unexplained variation at that higher level; thus the inclusion of important level 2 variables can reduce the level 2 variance, but not the level-1 variance.
- If a predictor variable is measured at level 1, it can account for the unexplained variation at level 1 and it can also reduce the higher-level variance if the predictor has an element that systematically varies at the higher-level. For this to happen, two conditions are needed, first if the predictor is treated as a response, there has to be a substantial higher-level variance; second the area mean (\bar{x}_{ij}) of the level 1 predictor has to be related to the original response variable.
- It is possible that the level 2 variance can increase with the introduction of a level 1 variable. In research on house prices, contextual neighbourhood effects may be masked so that expensive neighbourhood have small houses, and cheap neighbourhood have large houses, and when we take account of house size, the between district variance increases.

Given this pattern of changes, investigators often fit and report a sequence of models of growing complexity, paying due attention to the nature of the fixed parts estimates, but also inferring what are the effects of predictors by comparing the random parts of the model. These changes are usually given a substantive interpretation; thus in a value-added analysis, a researcher may report that the contextual effects have diminished, once intake ability has been taken into account.

Unfortunately, things are a lot trickier when modelling with a binomial variance. Using the simplifying assumption of the standard logistic distribution; the level 1 variance cannot change even when influential level 1 predictors are included in the model; it remains unchanged at 3.29. The level 2 variance is therefore being estimated relative to a numerically fixed benchmark; it is with model fit being re-scaled to this value. A thought experiment may help here. Imagine a null two level random intercepts model with no predictors, and which the level 2 variance is 0.35 (a VPC of 0.10). Now introduce an important level-1 predictor that has no strong level 2 component; it is 'pure' level 1 variable. As it is an important variable it should reduce the level 1 variance and leave the level 2 variance unchanged. But the level 2 variance is really scaled to the level 1 variance, the latter has gone down, but it cannot do so as it is fixed to 3.29. The consequence is that the level-2 variance will appear to go up to keep the relative scaling with the level 1 variance. Of the variance that remains a larger percentage must be at the higher level, as the level 1 value is fixed at 3.29. Sometimes this apparent increase is quite considerable if the level 1 predictor is an important one. In reality the matter is further complicated in that there may be an element of the level 1 variable that varies at the area level, and this might be reducing the level 2 variance but this is not showing as it is being swamped by the rise consequent on the explanatory power of the pure level-1 component of the variable. There is a final and important twist. As the level 2 variance increases the cluster-specific multilevel estimates can be expected to increase in absolute value; so that these constraints affect the fixed part estimates as well as the random part (see Snijders and Bosker, 2012, Chapter 17).¹¹

¹¹ Snijders, T. A. B. & Bosker, R. J. (2012) *Multilevel Analysis. Second edition*, London: Sage.

In short, in generalised linear models, changes in estimates are in part substantive and in part a technical consequence of scaling to the unchangeable level 1 variance. Adding important level 1 variables will generally increase the estimated level-2 unexplained variance. This will in turn lead to the estimates of already included predictors increasing in absolute size. The advice is to tread very carefully in comparing a sequence of binomial models. It will often be more helpful to include specifically the area means of the individual predictors (see later). It may also be helpful to run a further series of models in which the level 1 predictor is a response so that you ascertain the extent to which this variable varies at level 2. The conservative advice for binomial models is to compare estimates only within a model but not between models.

Thankfully in the present case the interpretation is relatively straightforward as the level-1 predictors of education and wealth are related to antenatal care and the level 2 variance has gone down considerably so that either education or wealth or both varies considerably between areas. (See the Appendix, Section 11, for confirmation of this.)

10 Two-level Random Slope Model: Allowing the Effect of Wealth to Vary Across Communities

The models fitted so far have allowed the probability of receiving antenatal care from a medically-trained provider to depend on the community of residence (as well as individual characteristics). This was achieved by allowing the model intercept to vary randomly across communities in a *random intercept* model. We have assumed, however, that the effects of individual characteristics such as age and education are the same in each community, i.e. the coefficients of all explanatory variables are fixed or unchanging across communities.

To obtain the DIC value for this random intercepts model:

- Go to the **Model** menu, select **MCMC** then **DIC Diagnostic**
- The DIC value will appear in the **Output** window

Bayesian Deviance Information Criterion (DIC)			
Dbar	D(thetabar)	pD	DIC
5562.19	5309.27	252.93	5815.12

We will now extend the random intercept model to allow both the intercept and the coefficient of one of the explanatory variables to vary randomly across communities.

- Open the **Equations** window to see the current random intercept model. The model contains three predictors: **mage**, **meduc** and **wealth**
- Switch back to the IGLS estimation method. Click **Estimation control** and then **IGLS/RIGLS**, followed by **Done**
- Click on **wealth-gm**, check **j(comm)** and click **Done**

- Click **Estimates** once to see the model with Greek letters for the parameters

Note that a new term u_{4j} has been added to the model, so that the coefficient of **wealth-gm** has become $\beta_{4j} = \beta_4 + u_{4j}$, and the community-level variance has been replaced by a matrix with two new parameters, σ_{u4}^2 and σ_{u04} . Note that the slope residual, and associated variance and covariance, have a subscript of '4' because **wealth** is the 4th explanatory variable in the model (not including the cons).

Equations

$$\text{antemed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_0 \text{cons} + \beta_1 (\text{mage-gm})_{ij} + \beta_2 \text{meduc_2}_{ij} + \beta_3 \text{meduc_3}_{ij} + \beta_{4j} (\text{wealth-gm})_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{4j} = \beta_4 + u_{4j}$$

$$\begin{bmatrix} u_{0j} \\ u_{4j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u04} \\ \sigma_{u04} & \sigma_{u4}^2 \end{bmatrix}$$

$$\text{var}(\text{antemed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{cons}_{ij}$$

Toolbar: Name, +, -, Add Term, **Estimates**, Nonlinear, Clear, Notation, Responses, Store, Help, Zoom 100

- Click on **Estimates** twice so that we can see the parameter estimates when the model has been fitted
- Click **Start** to fit the model using MQL1

Note that the two new parameters are estimated as zero. A zero variance estimate cannot be used as a starting value for MCMC, so we will refit the model using the better approximation of the PQL2 procedure.

- In the **Equations** window, click on **Nonlinear**
- Under **Linearization**, select **2nd order**
- Under **Estimation type**, select **PQL**
- Click **Done**, then **More** (to use the MQL1 estimates as starting values for PQL2 estimation)

The screenshot shows the 'Equations' window in MLwiN. The model is defined as follows:

$$\text{antemed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + -0.005(0.006)(\text{mage-gm})_{ij} + 0.540(0.085)\text{meduc_2}_{ij} + 1.296(0.097)\text{meduc_3}_{ij} + \beta_{4j}(\text{wealth-gm})_{ij}$$

$$\beta_{0j} = -0.489(0.079) + u_{0j}$$

$$\beta_{4j} = 0.407(0.030) + u_{4j}$$

$$\begin{bmatrix} u_{0j} \\ u_{4j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.839(0.095) & \\ -0.107(0.030) & 0.015(0.020) \end{bmatrix}$$

$$\text{var}(\text{antemed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{cons}_{ij}$$

The bottom of the window has a toolbar with buttons: Name, +, -, Add Term, Estimates, Nonlinear, Clear, Notation, Responses, Store, Help, Zoom, and a dropdown menu set to 100.

The estimates have now¹² converged to non-zero values, which can be used as starting values for MCMC.

- To switch to MCMC click on **Estimation control** and then **MCMC**, followed by **Done**
- Click **Start**
- **Close** the **equations** window to increase speed; there is some overhead being used in updating the estimates in the equations window

The screenshot shows the 'Equations' window after switching to MCMC. The model specification is identical to the previous one, but the estimates are updated:

$$\text{antemed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + -0.005(0.006)(\text{mage-gm})_{ij} + 0.553(0.083)\text{meduc_2}_{ij} + 1.315(0.100)\text{meduc_3}_{ij} + \beta_{4j}(\text{wealth-gm})_{ij}$$

$$\beta_{0j} = -0.505(0.087) + u_{0j}$$

$$\beta_{4j} = 0.409(0.033) + u_{4j}$$

$$\begin{bmatrix} u_{0j} \\ u_{4j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.896(0.109) & \\ -0.122(0.027) & 0.022(0.010) \end{bmatrix}$$

$$\text{var}(\text{antemed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{cons}_{ij}$$

Below the model specification, the 'PRIOR SPECIFICATIONS' section is visible:

PRIOR SPECIFICATIONS

$p(\beta_0) \propto 1$

$p(\beta_1) \propto 1$

$p(\beta_2) \propto 1$

$p(\beta_3) \propto 1$

$p(\beta_4) \propto 1$

$p(\Omega_u) \sim \text{inverse Wishart}_2[2 * S_u, 2], S_u = \begin{bmatrix} 0.839 & \\ -0.107 & 0.015 \end{bmatrix}$

The bottom of the window has the same toolbar as the previous screenshot.

¹² Another way of doing this would have been to edit the stored IGLS/RIGLS estimates which are in column c1096 and ensuring that all variances terms have a value that is not zero, for example, 0.001. In practice, this had to be done quite frequently to give the improved MCMC estimation procedures something to start working on. We will need to use this particular 'trick' later.

Notice that the PQL2 estimates are being used as a prior for the inverse Wishart distribution of the level 2 variance-covariance matrix. This is clearly an informative prior. However this is done in such a way that if you use the default, this prior is only weakly informative and is equivalent to only 2 degrees of freedom being imposed; the bulk on the information is coming from the data. For further discussion and a comparison with some alternatives, see Browne and Draper (2000, 2006).¹³

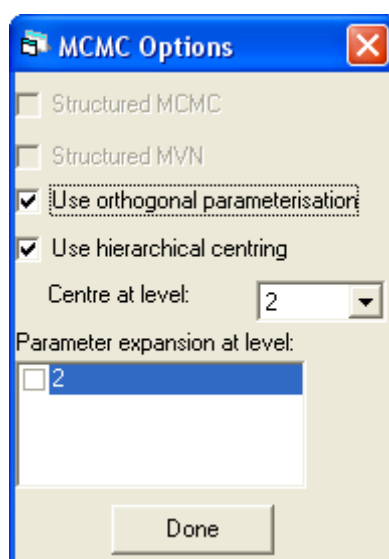
Once the model has finished running, re-open the **Equations** window, and **Store** this model as **Wealth_RS** to get the following results:

	Wealth_RS	S.E.	Corr	Median	CI(2.5%)	CI(97.5%)	ESS	Bayesian-p
Fixed Part								
Cons	-0.505	0.087		-0.510	-0.659	-0.322	36	0.000
(mage-gm)	-0.005	0.006		-0.005	-0.016	0.007	780	0.209
meduc_2	0.553	0.083		0.551	0.390	0.715	237	0.000
meduc_3	1.315	0.100		1.315	1.118	1.506	218	0.000
(wealth-gm)	0.409	0.033		0.410	0.345	0.471	90	0.000
Random Part								
Level: comm								
cons/cons	0.896	0.109	1.000	0.892	0.696	1.118	82	
(wealth-gm)/cons	-0.122	0.027	-0.895	-0.122	-0.178	-0.053	29	
(wealth-gm)/(wealth-gm)	0.022	0.010	1.000	0.019	0.012	0.050	18	
DIC:	5801.534							
pD:	252.112							

It is clear that the ESS for these estimates is un-acceptably low for several parameters. To have less correlated chains, that is to increase the ESS without increasing the default 5000 monitoring chains (and hence the run time), it is worth trying the following before **Starting** the chains

- **Model** on main menu
- Select **MCMC**, then select **MCMC options**
- Tick on **Use orthogonal parameterisation**
- Tick on **Use hierarchical centring**
- **Start**

¹³ Browne, WJ and Draper, D. (2000) Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models, *Computational Statistics*, 15:391-420; Browne, WJ and Draper, D. (2006). A comparison of Bayesian and likelihood- based methods for fitting multilevel models, *Bayesian Analysis*, 1:473-550.



These generally useful procedures are covered in detail in the *MCMC Manual*, Chapter 23 and 25. These procedures are not altering the specification of the model, but are designed to reduce the autocorrelation of the chains.

Store this model as **Wealth_RSHC** to get the following results

	WealthRSHC	S.E.	Corr	Median	2.5%	97.5%	ESS	BayesianP
Fixed Part								
Cons	-0.490	0.080		-0.492	-0.644	-0.331	415	0.000
(mage-gm)	-0.005	0.006		-0.005	-0.016	0.006	1211	0.173
meduc_2	0.541	0.085		0.540	0.375	0.710	892	0.000
meduc_3	1.301	0.100		1.301	1.108	1.493	895	0.000
(wealth-gm)	0.413	0.030		0.413	0.355	0.472	534	0.000
Random Part								
Level: comm								
cons/cons	0.864	0.111	1.000	0.856	0.660	1.098	260	
(wealth-gm) /cons	-0.122	0.032	-0.811	-0.121	-0.188	-0.064	13	
(wealth-gm) /(wealth-gm)	0.028	0.012	1.000	0.027	0.010	0.051	11	
DIC:	5809.188							
pD:	257.766							

This has generally improved the ESS for the fixed part, but the random part is still very low and we should increase the monitoring size to say 10,000, then 15,000 and so on until we are confident in the results. But we will simply press on here. A guide is an ESS of say at least 400 is needed to be able to say something useful about the posterior credible intervals.

We can first use the DIC to see whether the Random slopes model is an improvement over the Random intercepts model. We find that DIC has reduced from 5815.12 (RI) to 5809.18 (RS) which suggests that this more complex model is worth pursuing. (Notice that in the non-hierarchically centred model first estimated the DIC is 5801.534 compared to the current 5809.18 for the same RS model-; we really need to make sure that the chains have been run sufficiently long even for even comparing

an overall global statistic such as the DIC.) The mean estimate of each random parameter is larger than its standard deviation (e.g. 0.028 compared to 0.012 for the slope variance); the 95% credible intervals do not straddle zero; and the correlation between intercepts and slopes is a very sizeable -0.811. The value of correlation is not based on the MCMC chains it is simply based on the mean estimates of the covariance and variance as $-0.125 / (0.905^{0.5} * 0.020^{0.5})$.

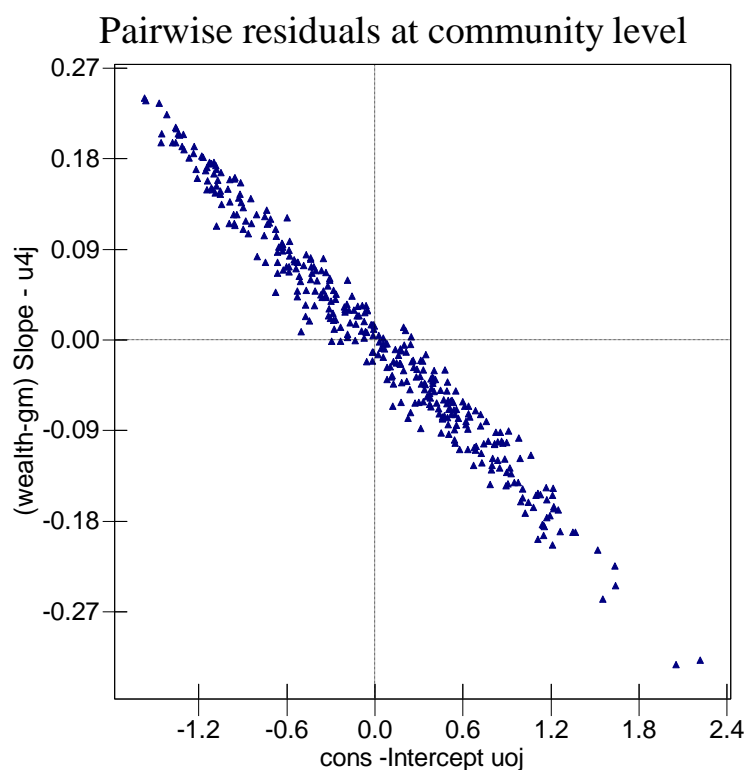
The effect of wealth on the log-odds of receiving antenatal care in community j is estimated as $0.4013 + \hat{u}_{4j}$, and the between-community variance in the effect of wealth is estimated as 0.028. Because **wealth** has been centred about its sample mean, the intercept variance $\hat{\sigma}_{u0}^2 = 0.864$ is interpreted as the between-community variance in the log-odds of antenatal care at the mean of the wealth index.

Examining the intercept and slope residuals for communities

The negative intercept-slope covariance estimate ($\hat{\sigma}_{u4} = -0.122$ and equivalently the correlation of -0.811) implies that communities with above-average antenatal care uptake (intercept residual $\hat{u}_{0j} > 0$) tend also to have below-average effects of wealth (slope residual $\hat{u}_{4j} < 0$). Put another way, there is less of an income gradient in use of antenatal care in communities with high uptake. To obtain a plot of the community intercepts versus the community slopes for wealth, \hat{u}_{0j} vs \hat{u}_{4j} :

- Make sure the **Residuals** window is not open from earlier
- From the **Model** menu, select **Residuals**
- At the bottom of the **Residuals** window, next to **level**, select **2:comm**
- Click **Calc** [Ignore numerical errors if reported and just move to the **Plots** tab of the residuals window]
- Click on the **Plots** tab and, under **pairwise**, check **residuals**
- Click **Apply**

You should obtain the following plot (titles have been added):



If we knew the geographical location of communities, it might be of interest to use this plot to identify communities that had low uptake and steep income gradients. Efforts to improve maternal health services might then be targeted towards such areas.

Community prediction lines

The equation for the fitted regression line for community j , for a woman of mean age and no education (the reference category of **meduc**) is:

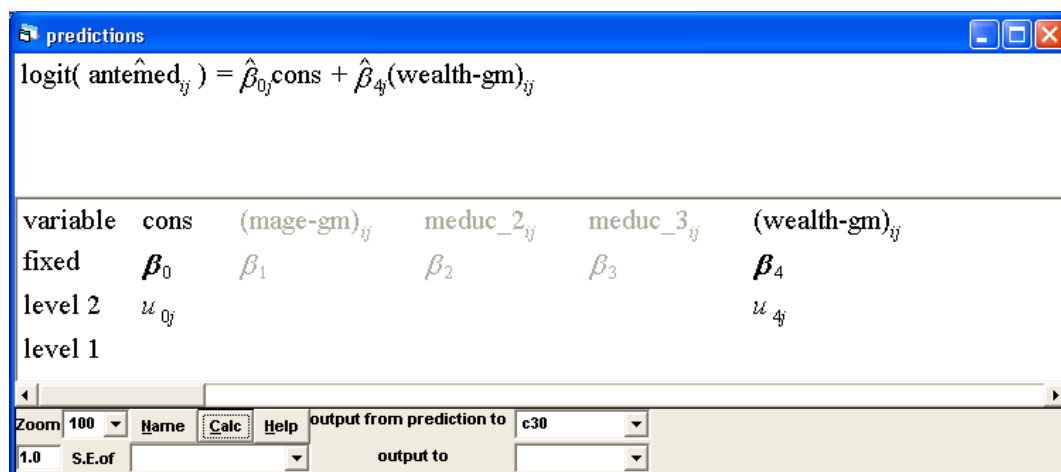
$$\log\left(\frac{\hat{\pi}_{ij}}{1 - \hat{\pi}_{ij}}\right) = (-0.490 + \hat{u}_{0j}) + (0.413 + \hat{u}_{4j}) \mathbf{wealth}_{ij}$$

where **wealth** has been centred about its sample mean. To obtain the fitted line for women with different ages or levels of education, only the intercept would change. For a woman with primary education, for example, the intercept would increase from -0.490 to $-0.490 + 0.541 = 0.06$.

To produce a plot of the predicted community lines, we first need to compute the predicted log-odds, $\text{logit}(\hat{\pi}_{ij})$, for each woman, based on their value of **wealth** and their community of residence.

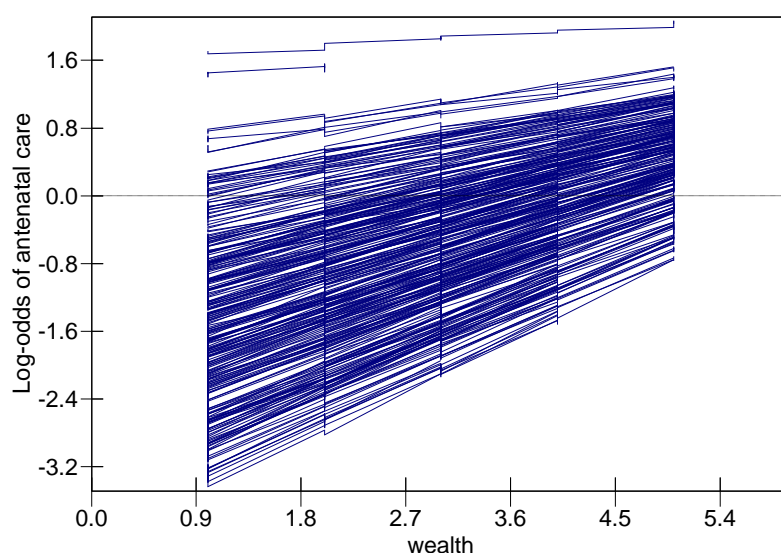
- Make sure the **Predictions** window is not open from earlier
- From the **Model** menu, select **Predictions**

- Next to **fixed** click on β_0 and β_4 to select them (make sure that β_1 is not selected)
- Next to **level 2** click on u_{0j} and u_{4j}
- Check that the prediction equation looks like the one shown below
- Next to **output from prediction to**, select **c30** (or any empty column)
- Click **Calc**
- Go to the **Names** window and name c30 **pred_rs** (for 'prediction from random slope model')



To plot the predicted community lines

- Go to the **Graphs** menu and select **Customised Graph(s)**
- Change to dataset **D4** (or any 'empty' display with no graph already specified)
- Next to **y**, select **pred_rs**
- Next to **x**, select **wealth**
- Next to **group**, select **comm**
- Change **plot type** to **line**
- Click **Apply**



Notice that some lines are shorter than others because not all communities contain women in the higher wealth quintiles. We can also see that the community lines are ‘fanning in’ as wealth increases. This is expected because of the negative correlation between the intercept and slope residuals. In short wealthier individuals tend to have a great chance of receiving antenatal care; the between-community differences are greatest for the least wealthy people.

Between-community variance as a function of wealth

From the plot of the predicted lines for each community, we can see that the lines are more spread out for the lower quintiles of the wealth index than at the higher quintiles. In other words, the variability in the log-odds of receiving antenatal care decreases as **wealth** increases. Fitting a random slope for **wealth** implies that the between-community variance is a function of **wealth**, rather than constant as in the random intercept model. The community-level variance function takes the following form:

$$\begin{aligned}\text{var}(u_{0j} + u_{4j}\mathbf{wealth}_{ij}) &= \text{var}(u_{0j}) + 2\text{cov}(u_{0j}, u_{4j})\mathbf{wealth}_{ij} + \text{var}(u_{4j})\mathbf{wealth}_{ij}^2 \\ &= \sigma_{u0}^2 + 2\sigma_{u04}\mathbf{wealth}_{ij} + \sigma_{u4}^2\mathbf{wealth}_{ij}^2\end{aligned}$$

which is estimated as (substituting estimates of σ_{u0}^2 , σ_{u04} and σ_{u4}^2):

$$0.860 - 0.122\mathbf{wealth}_{ij} + 0.028\mathbf{wealth}_{ij}^2.$$

To calculate the estimated community-level variance and its standard error:

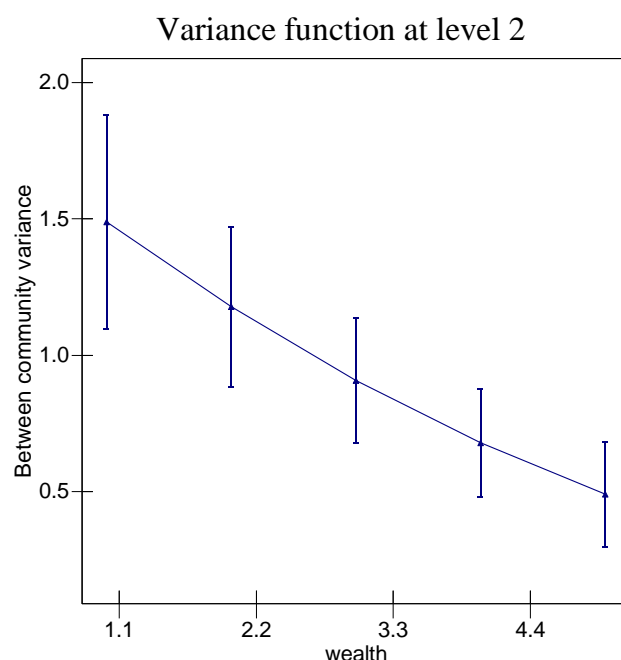
- From the **Model** menu, select **Variance function**
- At the bottom of the **Variance function** window, next to **level**, select **2:comm**

- Next to **variance output** to select **c31** (or any empty column) from the drop-down list
- In the text box immediately to the right, edit **1.0** to **1.96** and from the drop-down list next to **SE of variance output** to select **c32**
- Click **Calc**
- Go to the **Names** window and name c31 **l2var** and c32 **l2var_se** (you may need to click the refresh button under **Window** to see c31 and c32)

Column c31 now contains estimates of the between-community variance which will depend on wealth, and c32 contains 1.96 times the standard error of the variance estimates. The lower limit of a 95% confidence interval for the between-community variance (on the log-odds scale) is c31-c32 and the upper limit is c31+c32 that is c32 is an offset value.

We can now plot the between-community variance with 95% confidence intervals:

- Go to the **Graphs** menu and select **Customised Graph(s)**
- Change to dataset **D5** (or any 'empty' display with no graph already specified)
- Next to **y**, select **l2var**
- Next to **x**, select **wealth**
- Next to **plot type**, select **line+point**
- Click on the **error bars** tab
- Next to **y errors+**, select **l2var_se**
- Next to **y errors-**, select **l2var_se**
- Next to **plot as**, select **offsets** (not values)
- These settings for **y errors** and **plot as** will lead MLwiN to calculate and plot the lower and upper limits of the 95% confidence interval for the level 2 variance, using the values in **l2var_se**
- Click **Apply**



As expected from the ‘fanning in’ pattern of the community prediction lines, the between-community variance decreases as a function of individual wealth. Although the variance is a quadratic function of wealth, however, the plot shows mainly an underlying linear decrease. This is because the coefficient of the linear term in the variance function ($2\hat{\sigma}_{u04} = -0.44$) dominates over the coefficient of the quadratic term ($\hat{\sigma}_{u4}^2 = 0.028$), to the extent that $\hat{\sigma}_{u4}^2$ barely contributes to the total between-community variance. This is why we had difficulty in estimating this model in MQL-1: the variance term was being estimated as a negative value, and the default settings in IGLS resulted in both the variance and the covariance being set to zero. This is a misleading result as the covariance term (the linear part of the quadratic variance function) is revealed by MCMC estimation to be a substantively important term that is not zero.

11 Contextual effects and cross-level interactions

Also included in the worksheet are two community-level variables. **ComWealth** which has been derived from a multilevel model with the quintile of wealth as the response variable in a two level ordinal model. Similarly, **ComEduc** has been estimated from using **meduc** as an ordinal response variable in a two level model. The advantage of estimating these community estimates in a multilevel model is that they will be shrunken precision-weighted estimates that take account of measurement error from having an imbalanced sample with different number of respondents in different communities. An Appendix (section 11) details these procedures and how the contextual variables were derived.

ComWealth	A precision-weighted estimate of community wealth
ComEduc	A precision-weighted estimate of community maternal education

Returning to the Random slopes model, we will add in the conceptual variable **ComWealth** estimate and compare the DIC, and then add in a cross-level interaction between individual wealth and community wealth.

- Open the **Equations** window to see the current random slopes model. The model contains three individual level predictors: **mage**, **meduc** and **wealth**
- Switch back to the IGLS estimation method. Click **Estimation control** and then **IGLS/RIGLS**, followed by **Done**
- On the bottom tool bar, select **Add Term** choosing **ComWealth** from the drop down list - there is no need to centre as this variable will have a mean of around zero resulting from the way it has been derived, **Done**
- Click on **More** to get further quasi-likelihood iterations

This results in the following estimates

The screenshot shows the 'Equations' window in MLwiN. The model is defined as follows:

$$\text{antemed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + -0.006(0.006)(\text{mage-gm})_{ij} + 0.555(0.085)\text{meduc_2}_{ij} + 1.331(0.098)\text{meduc_3}_{ij} + \beta_{4j}(\text{wealth-gm})_{ij} + 0.400(0.041)\text{ComWealth}_j$$

$$\beta_{0j} = -0.475(0.076) + u_{0j}$$

$$\beta_{4j} = 0.280(0.031) + u_{4j}$$

$$\begin{bmatrix} u_{0j} \\ u_{4j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.628(0.077) & 0.000(0.000) \\ 0.000(0.000) & 0.000(0.000) \end{bmatrix}$$

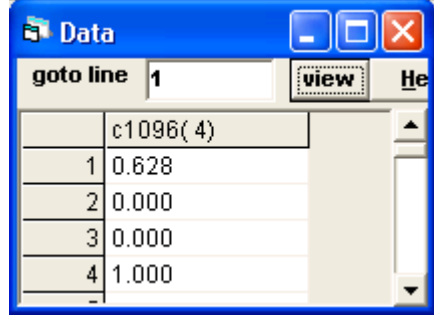
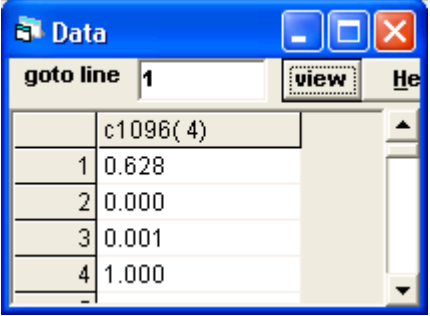
$$\text{var}(\text{antemed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{cons}_{ij}$$

The bottom toolbar shows buttons for Name, +, -, Add Term, Estimates, Nonlinear, Clear, Notation, Responses, Store, Help, and Zoom (set to 100).

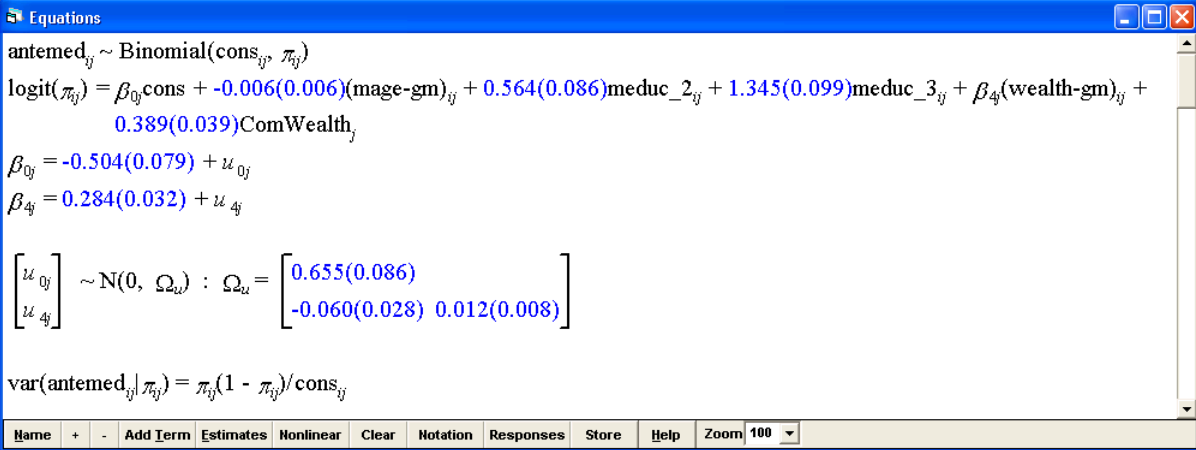
Notice that the estimate for **ComWealth** is large (0.400) in relation to its standard error (0.041) and that the software has correctly identified that this is a higher level variable with the subscript *j* (and not *ij*). Unfortunately, even the PQL2 quasi-likelihood procedures have estimated the covariance and variance associated with individual **wealth** to be zero.¹⁴ If you try to estimate this model with MCMC it will not be able to do so, and you will get the error: **MCMC error 0135 prior variance matrix is not positive definite**. To get around this, and make the prior variance positive definite, but without adding too much determining information to the prior we can edit column c1096 which contains the random part estimates and change the variance from 0.000 to 0.001 (leaving the covariance at 0.000)

¹⁴ If the variance term becomes zero during estimation, the associated covariance term is set to zero when using IGLS estimation. To see this tick on **Allow negative variances** at level 2 in **Estimation Control**, followed by **Start**; you will get an 'impossible' negative estimate of -0.002 but given the size of the standard error, this is a very uncertain estimate. Tick off **Allow negative variances** to get back to the default.

- In the **Names** window, highlight column c1096 and **View** it
- Replace the third value (the random slopes variance) of 0.000 with 0.001 by clicking into the field and typing the value.

Before	After	Meaning
		Level 2 variance Level 2 covariance Level 2 slope variance Level 1 constrained variance

The equations window should now update to the revised value and we can proceed to MCMC estimation. After **10,000** monitoring simulations the following values are obtained.



$\text{antemed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$
 $\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + -0.006(0.006)(\text{mage-gm})_{ij} + 0.564(0.086)\text{meduc_2}_{ij} + 1.345(0.099)\text{meduc_3}_{ij} + \beta_{4j}(\text{wealth-gm})_{ij} + 0.389(0.039)\text{ComWealth}_j$
 $\beta_{0j} = -0.504(0.079) + u_{0j}$
 $\beta_{4j} = 0.284(0.032) + u_{4j}$

$$\begin{bmatrix} u_{0j} \\ u_{4j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.655(0.086) \\ -0.060(0.028) & 0.012(0.008) \end{bmatrix}$$

 $\text{var}(\text{antemed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{cons}_{ij}$

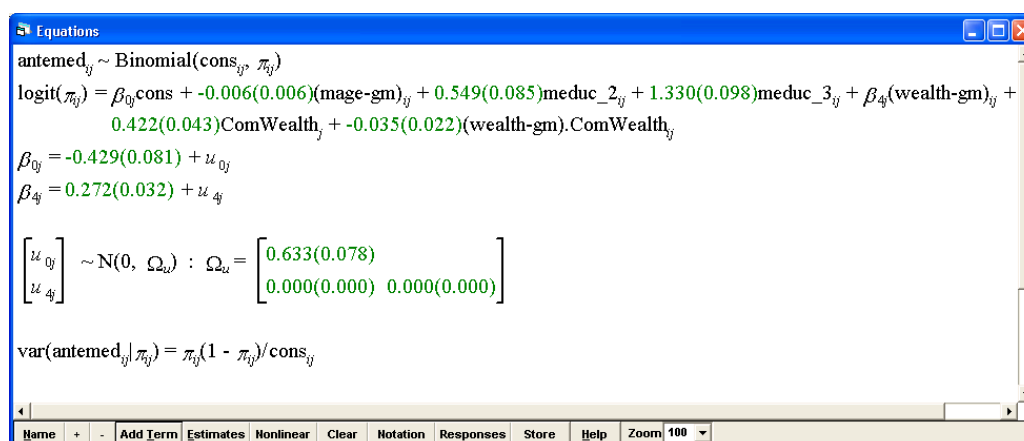
You will see in comparison to the previous model.

- The effect of individual wealth has reduced from 0.400 to 0.284 but it remains large in relation to its standard error (0.032); it is now the within- community individual effect and therefore has this more specific interpretation; previously it conflated the within and between effect of wealth.
- The effect of community wealth (0.389) is such that wealthier communities have higher rates of antenatal care; this between effect is large in comparison to its standard error (0.039).
- The unexplained variance function has changed from $0.864 - 0.122\text{wealth}_{ij} + 0.028\text{wealth}_{ij}^2$ to $0.655 - 0.060\text{wealth}_{ij} + 0.012\text{wealth}_{ij}^2$ which is a considerable reduction. You could plot this variance and do a before and after comparison.

- Examining the DIC this has reduced quite substantially from 5799.7 to 5762.1; community wealth is an important term in determining antenatal care.

The final model that we will consider is that with an additional cross-level interaction between individual and community wealth.

- Open the **Equations** window to see the current model. The model contains three individual level predictors: **mage**, **meduc** and **wealth** and **ComWealth**
- Switch back to the IGLS estimation method. Click **Estimation control** and then **IGLS/RIGLS**, followed by **Done**
- On the bottom tool bar, select **Add Term** choosing **Order 1** for an interaction between two variables; chose one of these to be individual **wealth** (it will automatically be centred around the grand mean as this was specified for the main effect); the other to be **ComWealth** from the drop down list (this will remain un-centred as in the main effect); **Done**
- Click on **More** to get further quasi-likelihood iterations



The screenshot shows the 'Equations' window in MLwiN. The model is defined as follows:

$$\text{antemed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + -0.006(0.006)(\text{mage-gm})_{ij} + 0.549(0.085)\text{meduc_2}_{ij} + 1.330(0.098)\text{meduc_3}_{ij} + \beta_{4j}(\text{wealth-gm})_{ij} + 0.422(0.043)\text{ComWealth}_{ij} + -0.035(0.022)(\text{wealth-gm}).\text{ComWealth}_{ij}$$

$$\beta_{0j} = -0.429(0.081) + u_{0j}$$

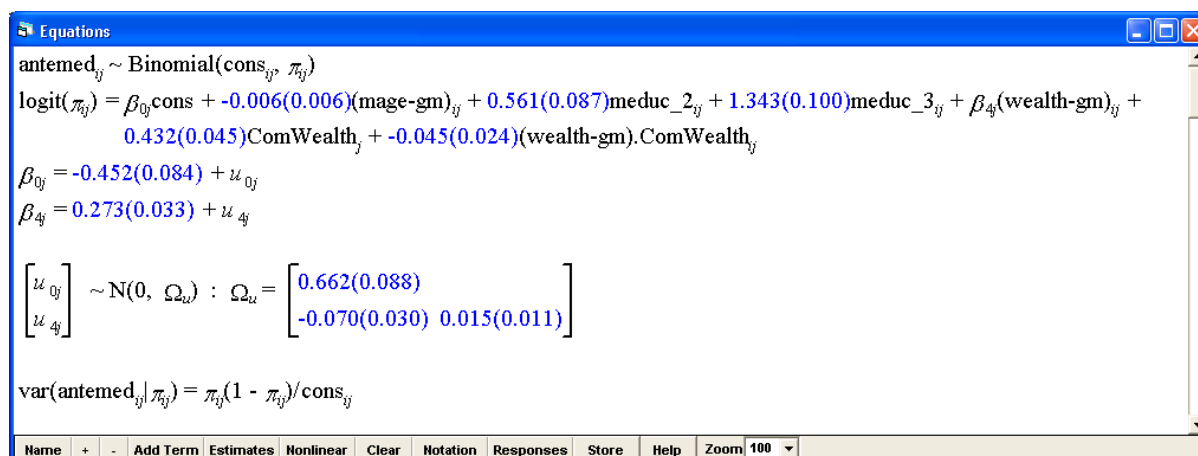
$$\beta_{4j} = 0.272(0.032) + u_{4j}$$

$$\begin{bmatrix} u_{0j} \\ u_{4j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.633(0.078) & 0.000(0.000) \\ 0.000(0.000) & 0.000(0.000) \end{bmatrix}$$

$$\text{var}(\text{antemed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{cons}_{ij}$$

The bottom toolbar shows the 'Add Term' button is selected.

Again we see estimates of zero for the level 2 slope variance. Using the procedure of putting a small value in for the level 2 variance in **c1096** prior to starting the MCMC estimation; the MCMC results are



The screenshot shows the 'Equations' window in MLwiN after MCMC estimation. The model is defined as follows:

$$\text{antemed}_{ij} \sim \text{Binomial}(\text{cons}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + -0.006(0.006)(\text{mage-gm})_{ij} + 0.561(0.087)\text{meduc_2}_{ij} + 1.343(0.100)\text{meduc_3}_{ij} + \beta_{4j}(\text{wealth-gm})_{ij} + 0.432(0.045)\text{ComWealth}_{ij} + -0.045(0.024)(\text{wealth-gm}).\text{ComWealth}_{ij}$$

$$\beta_{0j} = -0.452(0.084) + u_{0j}$$

$$\beta_{4j} = 0.273(0.033) + u_{4j}$$

$$\begin{bmatrix} u_{0j} \\ u_{4j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.662(0.088) & -0.070(0.030) \\ -0.070(0.030) & 0.015(0.011) \end{bmatrix}$$

$$\text{var}(\text{antemed}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{cons}_{ij}$$

The bottom toolbar shows the 'Add Term' button is selected.

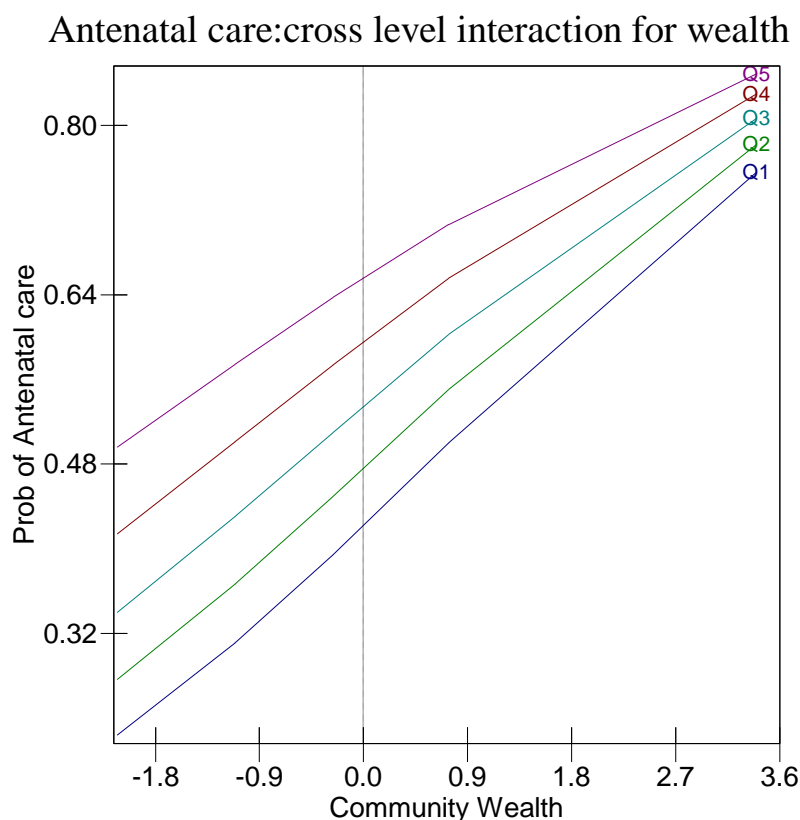
The estimate of -0.045 is not especially large in comparison to its standard error, the DIC has only reduced marginally from 5762.1 to 5758.2 while the Bayesian p value is 0.027. There is therefore some support for a cross-level interaction. A very effective way of conveying these results and evaluating the size of the effects is through the **Customised predictions**

- From the **Model** menu, select **Customised Predictions**
- Select **Clear** as the specification of the model has been changed
- Leave **Age** unchanged so that we will be making predictions for an average age female of **23.6** and also leave **Cons** at its value of 1
- Click on **meduc** and ensure that it does not have each category separately specified, but rather the average proportion, that is **meduc_2 (0.307)** and **meduc_3 (0.345)**
- Click on **wealth**, then **Change Range**. Click on the **Range** tab. Next to **Upper bound**, type 5. Next to **Lower**, type 1. Next to **Increment**, type 1. Click **Done**; this will give you a prediction for each and every quintile of individual wealth
- Click on **ComWealth**, this is currently set at its average value, select **Change Range**. Click on the **Percentile** tab. Type in 5 for the 5% percentile **Add**, then the 25, 50, 75 and 95 followed each time by **Add**, then **Done**; this will give the chosen percentiles for **ComWealth** which you will see in the **Summary** the actual values that will be used. There is no need to specify the cross-level interaction as the software 'knows' the specification of the model
- In the main **Customised Prediction** window, under **Predictions to**, check **Means** to get **population average** values [to get cluster specific check **Medians**]
- Click **Fill Grid** at the bottom of the **Customised Predictions** window and Click **Predict** to calculate the predictions by simulation
- Click on the **Predictions** tab. The predictions table will appear. Notice that there is a separate row for each combination of **ComWealth** and **wealth**, i.e. $5 \times 5 = 25$ entries.
- The predicted probabilities are stored in a column named **mean.pred**, and lower and upper limits of a 95% confidence interval for the predictions are stored in **mean.low.pred** and **mean.high.pred**.

Customised predictions							
Copy							
Setup				Predictions			
mage.pred	meduc.pred	wealth.pred	cons.pred	ComWealth.pred	mean.pred	mean.low.pr	mean.high.pr
23.634	lean(meduc)	1.000	1.000	-1.109	0.314	0.283	0.345
23.634	lean(meduc)	2.000	1.000	-1.109	0.370	0.343	0.398
23.634	lean(meduc)	3.000	1.000	-1.109	0.434	0.406	0.463
23.634	lean(meduc)	4.000	1.000	-1.109	0.504	0.468	0.540
23.634	lean(meduc)	5.000	1.000	-1.109	0.576	0.528	0.623
23.634	lean(meduc)	1.000	1.000	-2.126	0.227	0.188	0.270
23.634	lean(meduc)	2.000	1.000	-2.126	0.280	0.245	0.319
23.634	lean(meduc)	3.000	1.000	-2.126	0.344	0.307	0.382
23.634	lean(meduc)	4.000	1.000	-2.126	0.417	0.371	0.467
23.634	lean(meduc)	5.000	1.000	-2.126	0.498	0.433	0.564
23.634	lean(meduc)	1.000	1.000	-0.262	0.397	0.363	0.432
23.634	lean(meduc)	2.000	1.000	-0.262	0.452	0.424	0.482
23.634	lean(meduc)	3.000	1.000	-0.262	0.513	0.488	0.538
23.634	lean(meduc)	4.000	1.000	-0.262	0.576	0.548	0.604
23.634	lean(meduc)	5.000	1.000	-0.262	0.639	0.603	0.673
23.634	lean(meduc)	1.000	1.000	0.738	0.502	0.446	0.559
23.634	lean(meduc)	2.000	1.000	0.738	0.552	0.507	0.597
23.634	lean(meduc)	3.000	1.000	0.738	0.605	0.571	0.637
23.634	lean(meduc)	4.000	1.000	0.738	0.657	0.630	0.684
23.634	lean(meduc)	5.000	1.000	0.738	0.707	0.677	0.736
23.634	lean(meduc)	1.000	1.000	3.399	0.758	0.651	0.851
23.634	lean(meduc)	2.000	1.000	3.399	0.784	0.705	0.853
23.634	lean(meduc)	3.000	1.000	3.399	0.808	0.752	0.854
23.634	lean(meduc)	4.000	1.000	3.399	0.829	0.790	0.864
23.634	lean(meduc)	5.000	1.000	3.399	0.848	0.808	0.885

To produce a cross-level interaction plot

- Click on **Plot Grid**
- Next to X: check **ComWealth.pred**
- Next to Y: check **mean.pred**
- Next to **Grouped by**: check **wealth.pred**
- Click **Apply** (Confidence intervals are not used as it is difficult to see what is going on if these are included)



It is clear to see that the effects are large and that in particular community wealth is an important determinant of accessing care. The highest access is for the wealthiest quintile of wealth in the wealthiest communities. In the least wealthy communities, individual wealth makes a greater difference than in the wealthiest communities.

12 Appendix: deriving the precision-weighted community variables

Calculating Precision-weighted group level predictors

In the measurement of aggregate group level variables for higher-level units (communities in this case) based on level-1 variables (wealth and education in this case) there are two issues:

- the need to take account of the number of observations on which they are based and thereby minimize measurement error, if this is not addressed the

effect of contextual variables will be attenuated in the modelling of antenatal care;

- the need to calculate the means at one level net of the means at another higher level (this is not applicable here but this is often needed).

Both of these issues can be addressed by treating the level 1 predictor as a response and estimating a multilevel model. The higher level residuals are then the desired precision-weighted or shrunken group means (Jones *et al* 2014; Shin and Raudenbush, 2010, who consider more sophisticated methods for measurement error with missing values).¹⁵ These estimates are adaptive in that when they are not unreliable (because they are based on a large sample size in a community), there is no shrinkage. These multilevel estimates are thus corrected for reliability and minimise the effect of measurement error, and as Robinson, (1991) said, they are ‘a good thing’.¹⁶

It is important to acknowledge this method is based on the assumption that information is exchangeable between the different groups. Exchangeability refers to the idea that there is no systematic reason for distinguishing between the different groups, i.e. they are similar but not identical. Exchangeability can only be assessed on the basis of what we know about the dataset, i.e. are there no systematic and substantial differences between groups, which make them inappropriate for pooling of information. You have to make this judgement call. Here you might chose not to pool all the communities but only pool between urban areas and separately between rural areas.

In the present case both variables are measured on an ordinal scale. Chapter 11 of the *User manual* and Chapter 13 of the *MCMC manual* detail how to specify a multilevel ordinal model and how to estimate them in IGLS/quasi-likelihood and in MCMC respectively.

An ordinal model for wealth

Individual wealth was first ‘toggled’ in the **Names** window into a categorical variable and the labels were edited to Q1, Q2 to Q5. The model was then specified as a three level model with the ordered responses nested within women nested within community. When estimated by MQL-1, these estimates are

¹⁵ Shin, Yong-Yun and Raudenbush, Stephen W (2010). A latent cluster mean approach to the contextual effects model with missing data. *Journal of Educational and Behavioural Statistics*. Vol. 35, No. 1, 26-53; Jones, K Wright, C and Bell, A (2014) Do multilevel models ever give different results? Available from https://www.researchgate.net/publication/252146040_Do_multilevel_models_ever_give_different_results?ev=prf_pub

¹⁶ Robinson. G K (1991) That BLUP is a Good Thing: The Estimation of Random Effects, *Statistical Science* 6(1), 15-51.

Equations

$$\text{resp}_{ijk} \sim \text{Ordered Multinomial}(\text{cons}_{jk}, \pi_{ijk})$$

$$\gamma_{5jk} = \pi_{5jk}; \gamma_{4jk} = \pi_{5jk} + \pi_{4jk}; \gamma_{3jk} = \pi_{5jk} + \pi_{4jk} + \pi_{3jk}; \gamma_{2jk} = \pi_{5jk} + \pi_{4jk} + \pi_{3jk} + \pi_{2jk}; \gamma_{1jk} = 1$$

$$\text{logit}(\gamma_{2jk}) = 1.393(0.200)\text{cons.}(\geq Q2)_{ijk} + h_{jk}$$

$$\text{logit}(\gamma_{3jk}) = 0.491(0.199)\text{cons.}(\geq Q3)_{ijk} + h_{jk}$$

$$\text{logit}(\gamma_{4jk}) = -0.253(0.199)\text{cons.}(\geq Q4)_{ijk} + h_{jk}$$

$$\text{logit}(\gamma_{5jk}) = -1.122(0.199)\text{cons.}(\geq Q5)_{ijk} + h_{jk}$$

$$h_{jk} = v_{4k}\text{cons.}2345$$

$$\begin{bmatrix} v_{4k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 13.761(1.238) \end{bmatrix}$$

$$\text{cov}(y_{sjk}, y_{tjk}) = \gamma_{sjk}(1 - \gamma_{tjk})/\text{cons}_{jk} \quad s \leq t$$

(21464 of 21464 cases in use)

Name + - Add Term Estimates Nonlinear Clear Notation Responses Store Help Zoom 100

Running a MCMC analysis with hierarchical centering at level 3 and 50,000 monitoring chains results in the following estimates (note how the between community variance has been seriously over-estimated with the quasi-likelihood procedure): the between community differentials remain however substantial with a variance of 2.797

Equations

$$\text{resp}_{ijk} \sim \text{Ordered Multinomial}(\text{cons}_{jk}, \pi_{ijk})$$

$$\gamma_{5jk} = \pi_{5jk}; \gamma_{4jk} = \pi_{5jk} + \pi_{4jk}; \gamma_{3jk} = \pi_{5jk} + \pi_{4jk} + \pi_{3jk}; \gamma_{2jk} = \pi_{5jk} + \pi_{4jk} + \pi_{3jk} + \pi_{2jk}; \gamma_{1jk} = 1$$

$$\text{logit}(\gamma_{2jk}) = 1.942(0.097)\text{cons.}(\geq Q2)_{ijk} + h_{jk}$$

$$\text{logit}(\gamma_{3jk}) = 0.764(0.094)\text{cons.}(\geq Q3)_{ijk} + h_{jk}$$

$$\text{logit}(\gamma_{4jk}) = -0.294(0.093)\text{cons.}(\geq Q4)_{ijk} + h_{jk}$$

$$\text{logit}(\gamma_{5jk}) = -1.636(0.096)\text{cons.}(\geq Q5)_{ijk} + h_{jk}$$

$$h_{jk} = v_{4k}\text{cons.}2345$$

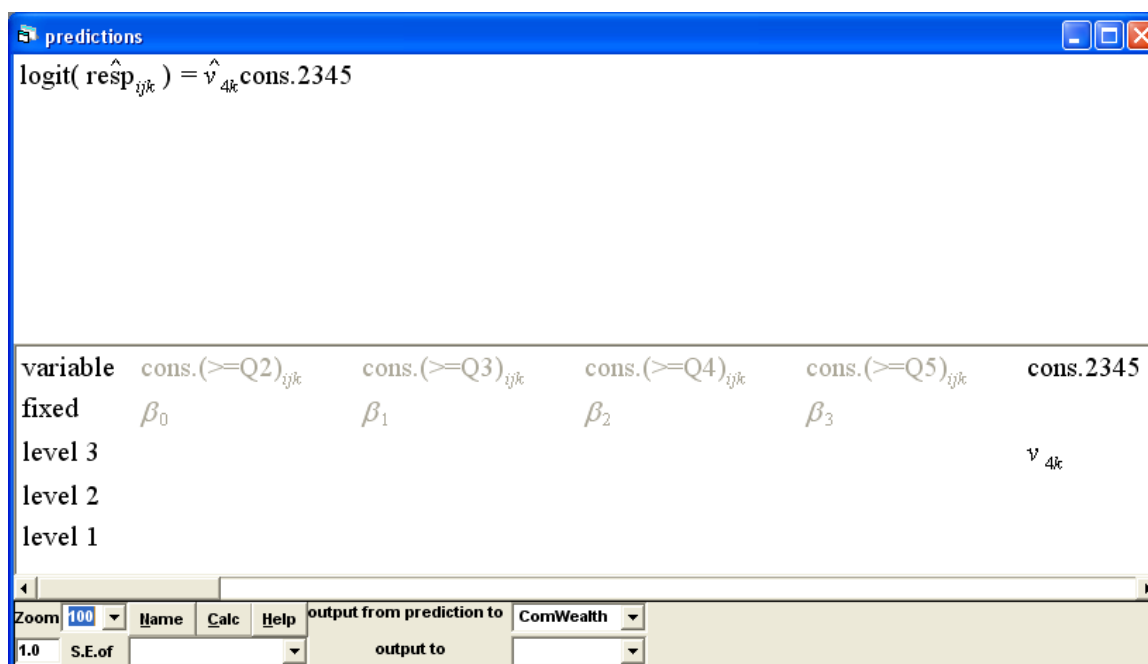
$$\begin{bmatrix} v_{4k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 2.797(0.257) \end{bmatrix}$$

$$\text{cov}(y_{sjk}, y_{tjk}) = \gamma_{sjk}(1 - \gamma_{tjk})/\text{cons}_{jk} \quad s \leq t$$

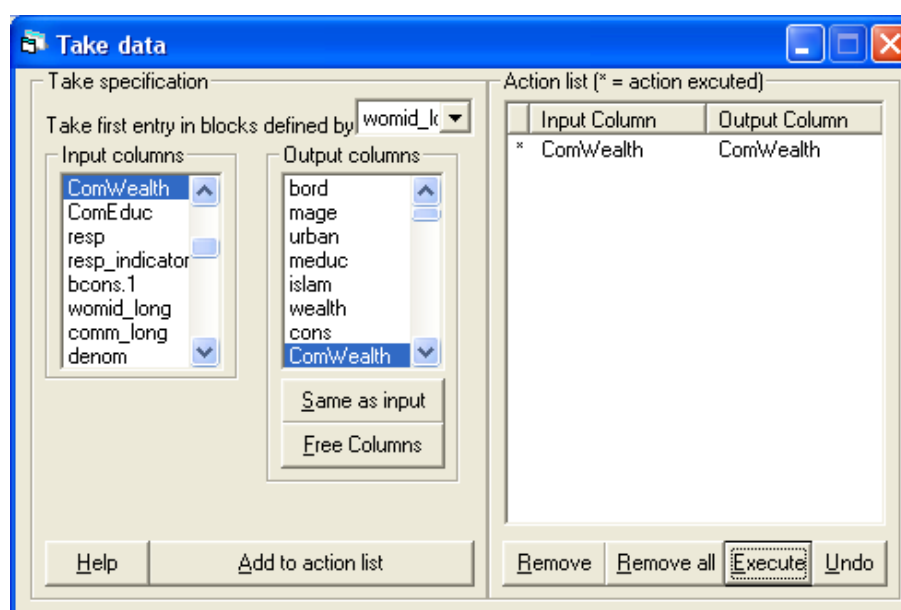
Deviance(MCMC) = 14334.218(21464 of 21464 cases in use)

Name + - Add Term Estimates Nonlinear Clear Notation Responses Store Help Zoom 100

The underlying latent community variable of wealth is v_4 ; high values of this mean a wealthy community. We can use the **prediction** window to calculate this for every community and to replicate it to every woman and to every response.



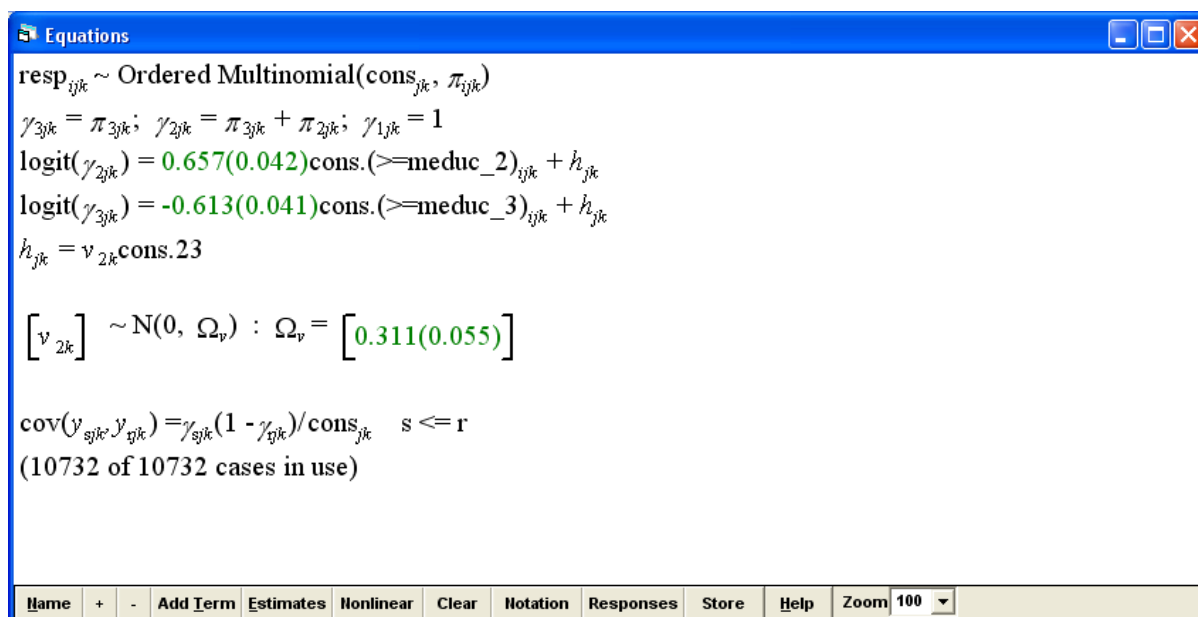
We can then **Take** or **unreplicate** to get a community wealth value for each women, with the blocks being defined by **womid_long** index which was created automatically as part of the setup of the multinomial model.



The underlying variable of community wealth **ComWealth** is the precision-weighted community mean; it is measured on the logit scale and centred on zero which is the average across all women and communities.

An ordinal model for maternal education

The same process is used for maternal education. These are the MQL-1 estimates



The screenshot shows the 'Equations' window in MLwiN. The equations displayed are:

$$\text{resp}_{ijk} \sim \text{Ordered Multinomial}(\text{cons}_{jk}, \pi_{ijk})$$

$$\gamma_{3jk} = \pi_{3jk}; \gamma_{2jk} = \pi_{3jk} + \pi_{2jk}; \gamma_{1jk} = 1$$

$$\text{logit}(\gamma_{2jk}) = 0.657(0.042)\text{cons.}(>=\text{meduc_2})_{ijk} + h_{jk}$$

$$\text{logit}(\gamma_{3jk}) = -0.613(0.041)\text{cons.}(>=\text{meduc_3})_{ijk} + h_{jk}$$

$$h_{jk} = v_{2k}\text{cons.23}$$

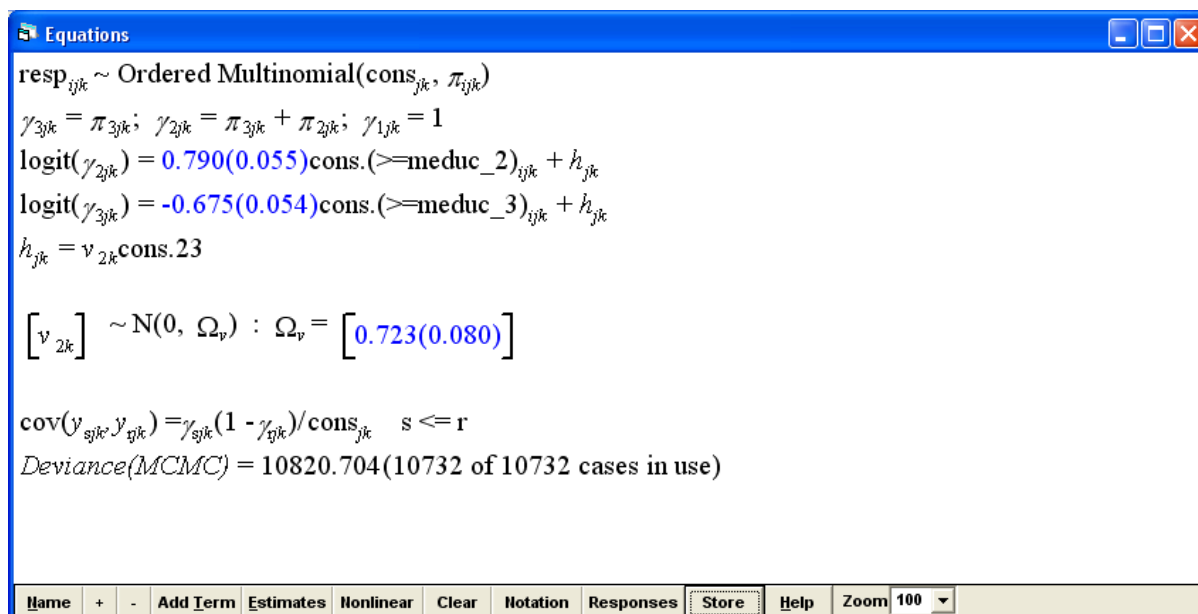
$$\begin{bmatrix} v_{2k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.311(0.055) \end{bmatrix}$$

$$\text{cov}(y_{sjk} y_{tjk}) = \gamma_{sjk}(1 - \gamma_{tjk}) / \text{cons}_{jk} \quad s \leq t$$

(10732 of 10732 cases in use)

The bottom toolbar includes buttons for Name, +, -, Add Term, Estimates, Nonlinear, Clear, Notation, Responses, Store, Help, and a Zoom dropdown set to 100.

And the MCMC estimates



The screenshot shows the 'Equations' window in MLwiN with the same model structure as the MQL-1 window, but with MCMC estimates:

$$\text{resp}_{ijk} \sim \text{Ordered Multinomial}(\text{cons}_{jk}, \pi_{ijk})$$

$$\gamma_{3jk} = \pi_{3jk}; \gamma_{2jk} = \pi_{3jk} + \pi_{2jk}; \gamma_{1jk} = 1$$

$$\text{logit}(\gamma_{2jk}) = 0.790(0.055)\text{cons.}(>=\text{meduc_2})_{ijk} + h_{jk}$$

$$\text{logit}(\gamma_{3jk}) = -0.675(0.054)\text{cons.}(>=\text{meduc_3})_{ijk} + h_{jk}$$

$$h_{jk} = v_{2k}\text{cons.23}$$

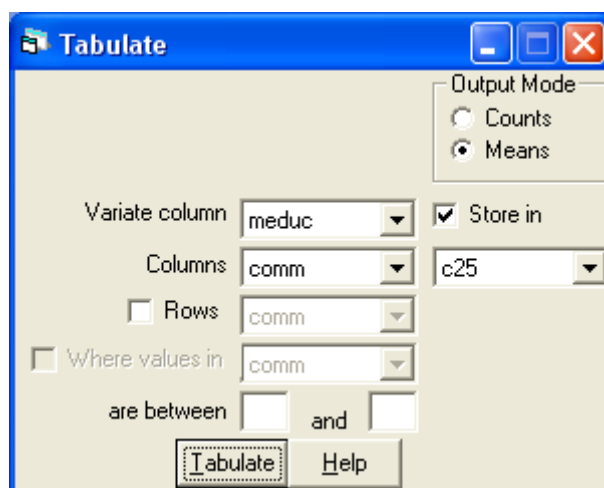
$$\begin{bmatrix} v_{2k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.723(0.080) \end{bmatrix}$$

$$\text{cov}(y_{sjk} y_{tjk}) = \gamma_{sjk}(1 - \gamma_{tjk}) / \text{cons}_{jk} \quad s \leq t$$

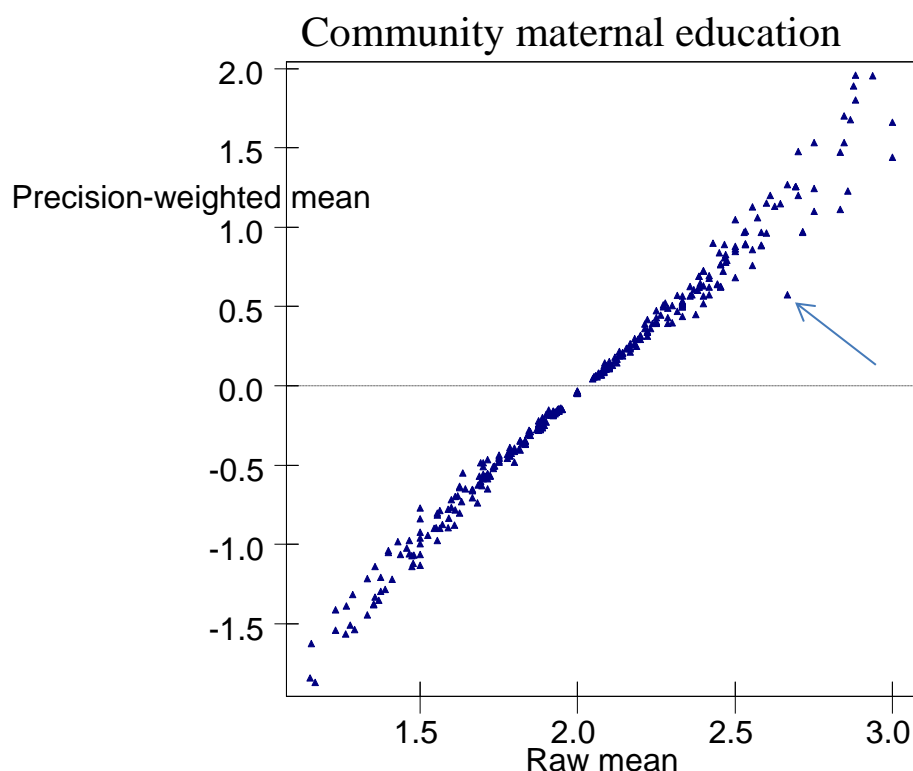
Deviance(MCMC) = 10820.704(10732 of 10732 cases in use)

The bottom toolbar is identical to the MQL-1 window, but the 'Store' button is highlighted.

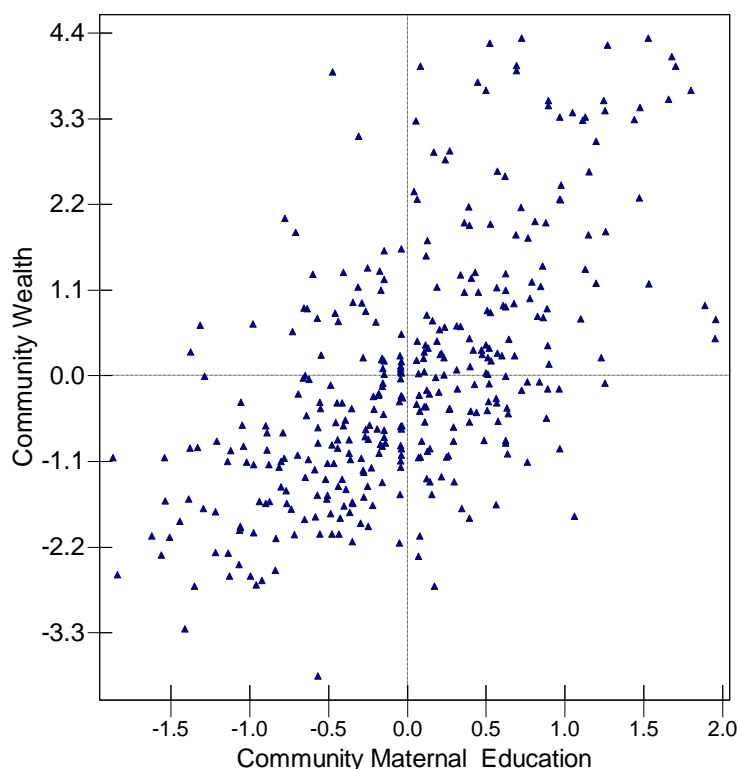
This time the quasi-likelihood have underestimated the between community variance. The variance of 0.723 suggests that there are community differences in education but they are not as large as those for wealth. **Prediction** followed by **Unreplicate** produces the required variable. This can be compared to the simple average raw mean for each community; the latter being produced as follows into **c25** via **Basic Statistics** and **Tabulate** which will be in the unreplicated form with only one value for each community.



The plot of the two estimates (the community unreplicated logit can be derived as the level 3 residual) shows that the raw mean values are pulled towards the “45 degree line” of no difference - on inspection you will find that the communities which are located furthest away from the line have a small sample size. Thus, the community being pointed to by the arrow, community 329, only has six respondents and is likely therefore to have substantial measurement error.



Comparing the variance of the two models as derived by MCMC, there are much greater between community differences in wealth than in education - the country is more segregated on wealth lines than on maternal education. A plot of the two variables reveals that there should not be issues of collinearity when both variables are included as contextual effects in a model to account for antenatal care.

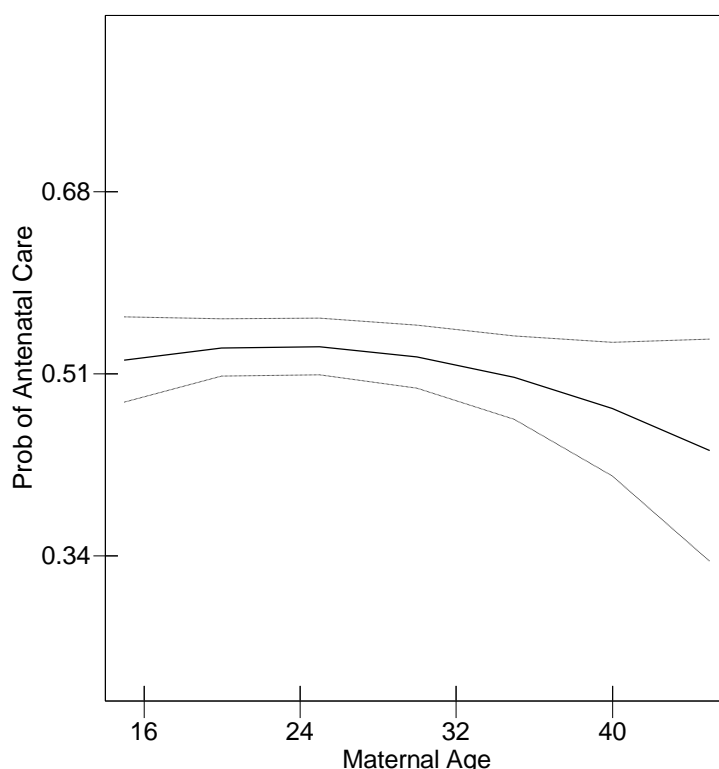


13 Extending the analysis

There are a number of ways the analysis can be extended and here we give some suggestions, and hints how they can be achieved in the MLwiN software.

Non-linear (quadratic) relationship with age

The relationship between the logit of antenatal care may non-linear and here is the result of 2nd order polynomial. These are the results for the random-intercepts model with only individual variables include. There is some evidence that older women have lower antenatal care when education and wealth are taken into account but the evidence is not strong. The 95% credible intervals of both the linear and quadratic estimates span zero, but the Bayesian p value for the quadratic part provides only some small evidence that the relation is negative.



	Quad Mage	S.E.	Median	CI(2.5%)	CI(97.5%)	ESS	Bayesian-p
Response	antemed						
Fixed Part							
cons	-0.455	0.084	-0.453	-0.624	-0.294	1369	0.000
meduc_2	0.555	0.085	0.557	0.385	0.716	1109	0.000
meduc_3	1.317	0.100	1.316	1.124	1.513	1002	0.000
(wealth-gm)	0.398	0.031	0.397	0.339	0.460	682	0.000
(mage-gm)^1	-0.001	0.007	-0.001	-0.013	0.013	1131	0.473
(mage-gm)^2	-0.001	0.001	-0.001	-0.002	0.000	1151	0.071
Random Part							
Between community	0.905	0.116	0.897	0.702	1.156	465	
DIC:	5816.250						
pD:	252.959						

Hints: in the **Equation** window, click on the **Mage** term, choose to **Modify** it and select a 2nd order polynomial that is quadratic in **Mage**. Don't forget to switch to IGLS, estimate and then switch to MCMC and then **Start**. You can use the customised graphics to make the plot; note that it 'understands' that **Mage** is now in a quadratic relationship. Orthogonal parameterisation and hierarchical centring at level 2 was used with the default monitoring chains to get these results with their good ESS.

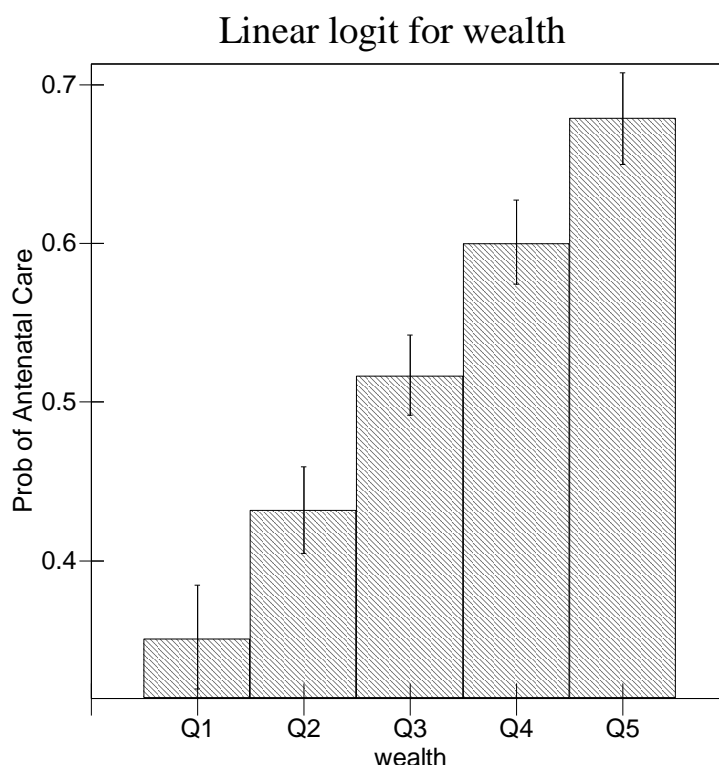
A second order orthogonal polynomial for wealth

We have so far treated wealth as a continuous variable when it is in fact ordinal. A variable measured in this way could be treated as categorical with 4 dummies being placed in the fixed part. A more parsimonious approach is to include this ordinal variable as an orthogonal polynomial. It is then possible to fit an underlying linear

trend, then as a quadratic and then as cubic to see if a parsimonious formulation is appropriate. Unlike the usual polynomial transformation and as its name suggests, this coding of an ordinal variable is achieved in such a way that each term is orthogonal from each other making for easier model estimation. The size of the effects are also directly comparable as each polynomial is automatically standardised. To undertake this modelling, begin by turning individual wealth into a categorical variable and then fit the variables as linear, quadratic, cubic, quartic and finally as a categorical variable (this has the same number of parameters as the quartic model). Work with a random-intercepts model with only age and maternal education included. The *Manual Supplement* fully explains the orthogonal polynomial technique in section 1.2. Each model has been run for 10,000 monitoring estimations and stored. The results given below show a very large reduction in the DIC when linear wealth is introduced but more complex models are not found to be an improvement. A customised predictions plot has also been plotted of the most parsimonious model, the underlying linear model of wealth.

	Age& Edu	S.E	Linear	S.E	Quad	S.E	Cubic	S.E	Quartic	S.E	Categ	S.E
Fixed Part												
cons	-0.73	0.09	-0.50	0.09	-0.48	0.08	-0.48	0.08	-0.49	0.08	-1.29	0.10
(mage-gm)	0.00	0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01
meduc_2	0.77	0.09	0.55	0.09	0.55	0.09	0.55	0.09	0.55	0.08	0.55	0.09
meduc_3	1.85	0.09	1.31	0.10	1.29	0.10	1.29	0.10	1.30	0.10	1.30	0.10
orthog_wealth^1			1.27	0.10	1.30	0.10	1.32	0.10	1.31	0.10		
orthog_wealth^2					0.16	0.08	0.18	0.08	0.17	0.08		
orthog_wealth^3							0.19	0.08	0.19	0.08		
orthog_wealth^4									-0.03	0.07		
Q2											0.47	0.11
Q3											0.69	0.11
Q4											1.06	0.12
Q5											1.78	0.14
Random Part												
Level: comm												
cons/con	1.31	0.15	0.91	0.11	0.88	0.11	0.87	0.11	0.86	0.11	0.87	0.11
DIC:	5916.56		5815.70		5818.00		5816.42		5820.02		5820.38	
pD:	279.08		252.68		251.13		250.14		251.75		252.46	

The downside of this approach is that compared to referenced code dummy variables the estimates do not give the mean differential, but a plot shows the nature of the relationship.



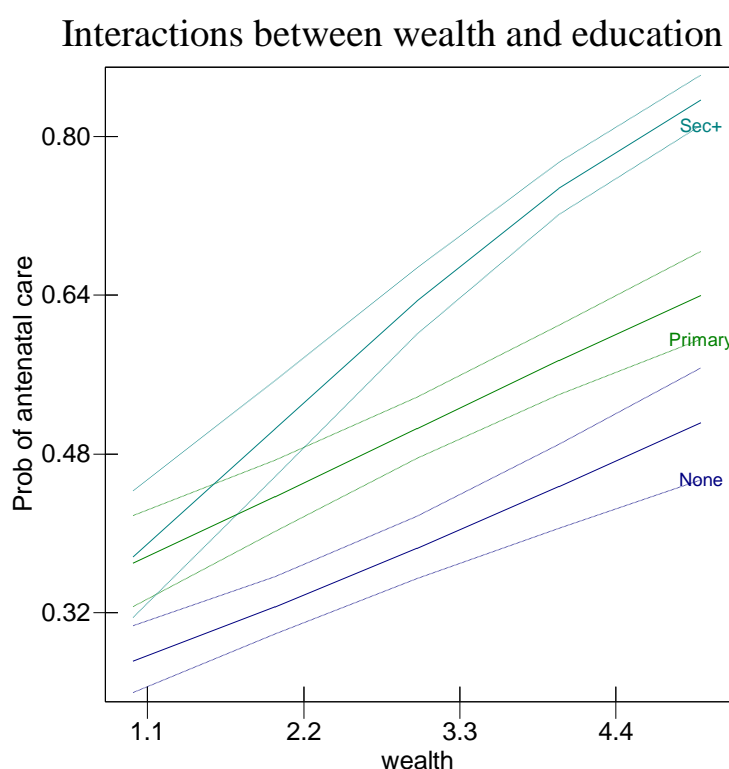
Hint: first remove linear wealth from the model and toggle to be a categorical variable. Edit the labels to simplify and identify quintiles as Q1,Q2... Q5. In the equations window use **add term** to include the categorical variable; tick on orthogonal polynomial and choose 1st order, estimate model and store results. Click on the wealth variable in the equations window and **modify** to 2nd order, estimate and store; repeat this for cubic and quartic - the latter is equivalent of a categorical specification (but the estimates are not so readily interpretable). Finally, click on wealth and tick off orthogonal polynomial and choose Q1 as the base category to produce the categorical model with dummies.

Interactions for education and wealth

This involves evaluating whether wealth has a differential effect for different degrees of education. These are the results for a model in which the interaction for education and wealth has been added to a random intercepts model already including the main effects for maternal age, education and wealth. The model has been run for 50,000 monitoring estimations. A customised predictions plot has also been plotted for the interaction.

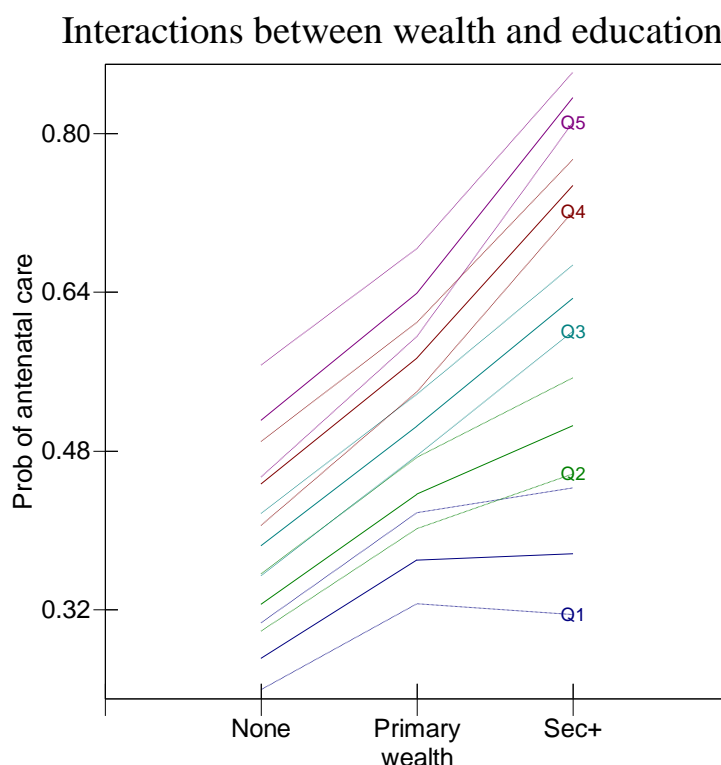
	Edu*Wealth	S.E.	Median	CI(2.5%)	CI(97.5%)	ESS	Bayesian-p
Fixed Part							
Cons	-0.544	0.081	-0.544	-0.698	-0.385	1349	0.000
(mage-gm)	-0.007	0.006	-0.007	-0.018	0.004	8357	0.119
meduc_2	0.585	0.087	0.586	0.411	0.756	2090	0.000
meduc_3	1.220	0.104	1.220	1.016	1.423	2098	0.000
(wealth-gm)	0.308	0.046	0.307	0.219	0.397	1876	0.000
meduc_2.(wealth-gm)	0.024	0.060	0.024	-0.095	0.142	2689	0.346
meduc_3.(wealth-gm)	0.326	0.067	0.326	0.195	0.458	3038	0.000
Random Part							
Level: comm							
cons/cons	0.887	0.111	0.881	0.687	1.124	4917	
DIC:	5796.031						
pD:	251.910						

The estimates are generally large in comparison to their standard error with the exception of **mage** with which we are familiar and the new term **meduc_2.(wealth-gm)** which is the differential interaction for primary education and wealth. A customised predictions plot will be very useful in interpreting this.



At low wealth there is much less difference between education groups, but for the most wealthy, education plays a much larger part. A combination of wealth and education is important in gaining and using antenatal care. This is a good example where a plot with confidence intervals is more informative than a table of estimates with standard errors.

Hint: in IGLS and in the **Equation** window, add 1st order interaction between maternal education and wealth using **Add term**. Estimate and then switch to MCMC, start the chain, wait, and store the results. In the customised predictions window, clear the previous settings, choose each category of maternal education and chose a range of values (1-5) for the quintiles of wealth. In the plot, wealth is the horizontal variable and the grouping variable is maternal education. We would argue that plotting wealth as the grouping variable and education as the horizontal axis (see below) is not as clear in conveying what is happening - it is worth experimenting with different visualizations when conducting you own analysis.



Cross-level interaction for individual education and community education

These are the results for a model in which Community education is included as a main effect and as a cross-level interaction with individual categorical maternal education. These have been added to a random slopes model for individual wealth that already includes a cross-level interaction for individual and community wealth. The model has been run for 50,000 monitoring estimations.

	EducCross	S.E.	Corr	Median	CI(2.5%)	CI(97.5%)	ESS	Bayesian-p
Fixed Part								
cons	-0.464	0.089		-0.464	-0.639	-0.291	7091	0.000
(mage-gm)	-0.006	0.006		-0.006	-0.017	0.005	10595	0.142
meduc_2	0.577	0.094		0.576	0.392	0.760	10896	0.000
meduc_3	1.351	0.105		1.351	1.145	1.559	12122	0.000
(wealth-gm)	0.274	0.035		0.274	0.207	0.342	9963	0.000
ComWealth	0.428	0.055		0.429	0.321	0.536	2364	0.000
(wealth-gm).ComWealth	-0.048	0.024		-0.048	-0.096	-0.001	3248	0.023
ComEduc	-0.017	0.133		-0.019	-0.275	0.246	3010	0.443
meduc_2.ComEduc	0.083	0.139		0.084	-0.191	0.355	9398	0.276
meduc_3.ComEduc	0.038	0.146		0.038	-0.253	0.320	6660	0.396
Random Part								
Level: comm								
cons/cons	0.656	0.088	1.000	0.652	0.497	0.842	2351	
(wealth-gm)/cons	-0.095	0.032	-0.70	-0.094	-0.164	-0.037	218	
(wealth-gm)/(wealth-gm)	0.029	0.014	1.000	0.027	0.008	0.061	162	
DIC:	5763.109							
pD:	243.144							

The main effect for individual education suggests that there is an individual effect but there is little support for the effect of community education and the cross-level interactions. The cross-level interaction plot shows this clearly. These results and the size of the effect for community wealth might lead to speculation that there is a supply-side issue with practitioners not receiving sufficient incentive to set up in the poorer areas or a diffusion issue where the provision has not yet spread to the poorer areas. Additionally it might be speculated that less educated women do not appreciate the benefits of antenatal care and the poorest households have low access to it.

Hint: in IGLS and in the **Equation** window, add the term for Community education and the 1st order interaction between individual maternal education and community education. Estimate and use the usual trick to make the level 2 variance-covariance positive definite. Switch to MCMC, start the chain, and store the results. In the customised predictions window, clear the previous settings, choose each category of maternal education and chose a range of values for community education keeping all the other predictors at their average value. In the graph, Community education is the horizontal variable and the grouping variable is maternal education.

