# Week 3: Random Slopes – Guided Practical

# 1 Fitting and interpreting a random slopes model

Open the Stata dataset **hedon\_RandomIntercept.dta**.

In the Random Intercepts Practical, we investigated country differences in hedonism after controlling for individuals' age, income, education and gender. We assumed that the relationship between hedonism and each of these variables was the same for each country. But is that actually true? Are there for example country differences in the relationship between hedonism and income? To address this question, we will need to fit a model which allows the relationship between hedonism and income to differ across countries: **a random slopes model**.

a) Before running a random slopes model, study the variables of the dataset, centre the continuous variables around their mean and run a variance component model and a random intercept model including the variables age, income, years in education and gender.

To centre the variables, type:

*egen age2 = mean(age)*

*gen centage = age – age2*

*egen meanincome = mean(income)*

*gen centincome = income – meanincome*

*egen meanedu = mean(eduyrs)*

*gen centedu = eduyrs – meanedu*

To run a variance component model, type:

*mixed hed|| country:, ml variance*

*estimates store vc*

To run a **random intercepts** model, type:

*mixed hed centage centincome centedu female || country: , ml variance*

*estimates store ri*

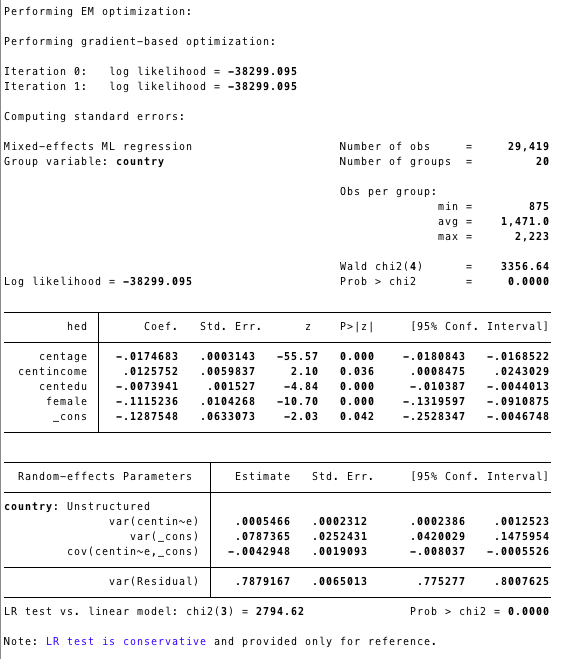
To run a **random slopes** model that allows the relationship between hedonism and income to differ across countries, type:

*mixed hed centage centincome centedu female||country:centincome, ml variance covariance(unstructured)*

**or**

*mixed hed centage centincome centedu female||country: income2, covariance(unstructured)*

*estimates store rs*

**

The option *covariance(unstructured)* provides the estimates for the covariance between intercepts and slopes. The random slopes model has two new parameters compared to random intercepts model: the covariance between intercepts and slopes and the variance of slopes. The covariance between intercepts and slopes is negative (when income is centred at its mean), equals to -0.004, and the pattern is of fanning in (countries with higher intercepts tend to have shallower slopes, i.e. more hedonistic countries at average income levels tend to have a smaller effect of income on hedonism).

b) Let’s compare our estimates for this model to those of the random interprets model.

*estimates table vc ri rs*

*lrtest ri rs*

The first thing to notice is that we have some extra terms in the random part. When we had just one parameter per level in the random part the table was quite easy to understand.

**Note:** We cannot directly compare the level 2 random parameters for our random slopes model with the level 2 random parameter for the random intercepts model. This is because the variance for the random slopes model depends on the value of income. We can compare the level 1 variance; we can see that it is roughly the same for the two models (it is very slightly smaller for the random slope model).

The coefficients of the variables in the fixed part have also changed only slightly.

The question that we asked at the beginning of the practical, which we were trying to investigate by fitting this model, was `Are there country differences in the relationship between hedonism and income?' We have not yet provided an answer. In fact, we will not really be able to do so until we perform a hypothesis test that will tell us whether or not there are country differences. In this section, we have simply made a start by fitting the appropriate model and trying to understand the estimates we get.

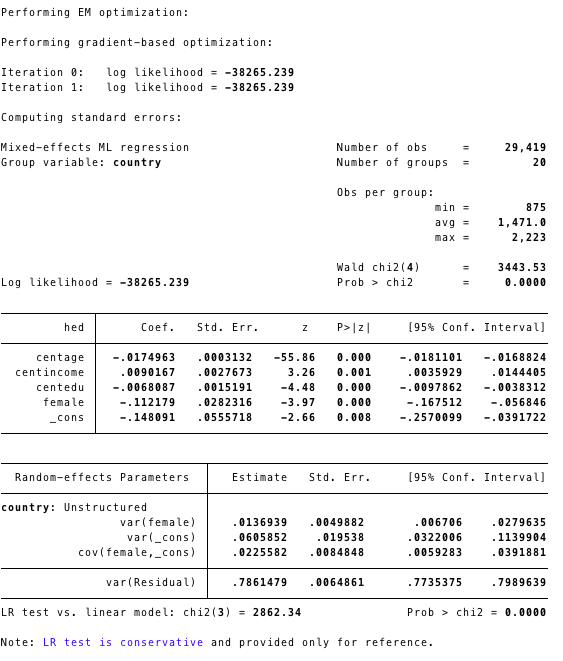
**Exercise** - What about gender: does the relationship between hedonism and gender differ across countries? Female is a categorical variable, but, we fit the random slope in exactly the same way. Run a random slopes model to investigate whether the relationship between hedonism and gender differs across countries. Interpret the covariance between intercepts and slopes. Has the level 1 variance changed much in this model compared to the random intercepts model?

**Answer** – Run the model:

*mixed hed centage centincome centedu female || country: female, covariance(unstructured)*

The covariance between the intercepts and slopes is positive, so countries for which men have comparatively high hedonism score also have a smaller difference between hedonism scores of men and women (recall that the coefficient of female is negative so that women have lower hedonism scores than men; therefore, a country with a positive value of covariance will have a smaller gap between men’s and women’s hedonism scores than average).

The level 1 variance is 0.786 in this model and it was 0.790 in the random intercepts model.



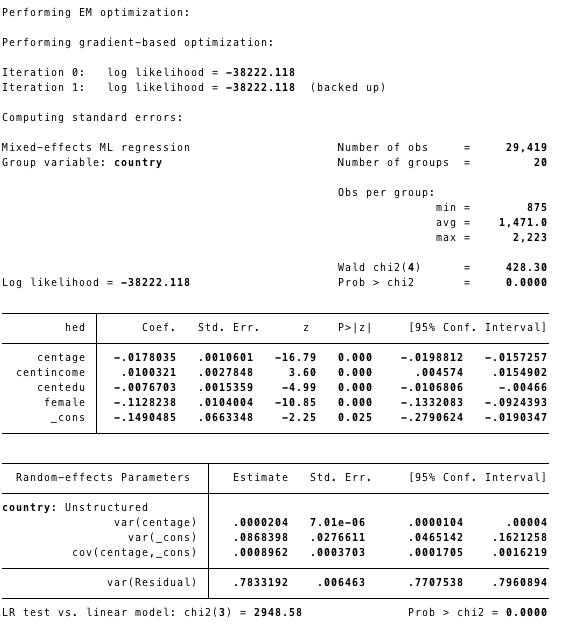
# 2 Hypothesis testing

We will now answer our question, ‘Are there country differences in the relationship between hedonism and income?’. We need to do a likelihood ratio test comparing the random slopes model to the random intercepts model. We want to compare our random slope model with a model which is exactly the same except that it does not allow the relationship between hedonism and income to differ across countries, i.e. it does not have random slopes. The deviance, i.e. the value of -2\*log(likelihood) for the random intercept model was 38321.519\*2= 76643.038, and the value for the random slopes model is 38299.095\*2=76598.19, so the test statistic is 76643.038-76598.19 = 44.848. We compare this against the distribution with two degrees of freedom (because we have two extra variance parameters in the random slopes model). The p-value is below 0.05. We can therefore reject the null hypothesis that the level-2 variance and the covariance between intercepts and slopes are 0. We conclude that there are indeed differences between countries in the relationship between hedonism and income.

All the fixed part coefficients are significant at the 5% level since they are more than twice than their standard errors.

c) Let us now investigate whether there are significant differences between countries in the relationship between hedonism and age. To do this we will fit a model with a random slope on age.

*mixed hed centage centincome centedu female || country: centage, ml variance covariance(unstructured)*

**

The estimate for the variance of the slopes is 0.000. We might think that if the variance of the slopes is 0 then the slopes are all the same and so there are no differences between countries in the relationship between hedonism and age. However, let's perform a likelihood ratio test. The -2\*log(likelihood) value for this model is 76444.236 and again we compare to the random intercepts model from before with value 76643.038, so the test statistic is 198.802, again with 2 degrees of freedom. The p-value is less than 0.000005; clearly, we should reject the null hypothesis. So, we conclude that there are differences between countries in the relationship between hedonism and age. But do the slopes have no variance? What is going on? The answer is that we must remember that the computer can only give us answers to so many decimal places. The variance of the slopes is 0.000 when rounded to 3 decimal places, but it is not actually zero, just very small.

Why might we expect the variance of the random slopes on age to be very small, just from looking at the values of age in our dataset? Remember that the variance depends on the scale of *x*. **Age** is measured in years, with a standard deviation of 18.15 for this dataset (*sum age*). The slope for age measures the average increase in hedonism for an extra year of age and the variance of the slopes measures the difference in the increase in hedonism for an extra year of age for country *j* compared to the average increase in hedonism for an extra year of age in the whole dataset. The standard deviation of **hed** is only 0.9854798. When we compare this to the standard deviation of **age** we can see that we will expect the slopes to be small. On top of that, recall that variance is the square of standard deviation, and that when you square a number smaller than 1, the answer has twice as many zeros before the first non-zero number after the decimal point.

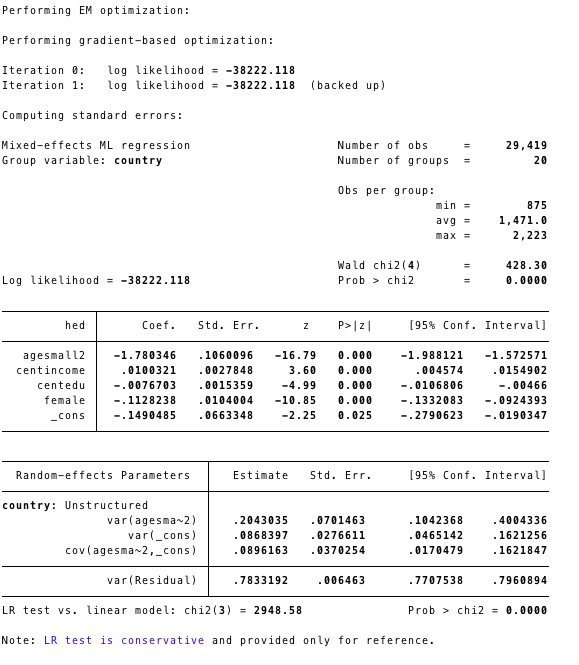
In order to get around this, we could create a new variable with values **age**/100 and put this in the model instead of **age**. The smaller values of this variable should make the slopes correspondingly bigger and the variance should become large enough not to appear as 0.000. Centre the new variable and run the random slopes model with the new variable:

*gen agesmall = age\*0.01*

*egen meanagesmall = mean(agesmall)*

*gen agesmall2 = agesmall – meanagesmall*

*mixed hed agesmall2 centincome centedu female || country: agesmall2, ml variance cov(uns)*



Notice that in the model with ‘agesmall’ instead of age, the value of -2\*log(likelihood) is exactly the same as the value for the random intercepts model with age, as are the level 1 and level 2 variances, and the coefficients of income, female and years of education and their standard errors, while the coefficient of age has simply been multiplied by 100 to give the coefficient of agesmall in the new model. None of this is surprising: these two models are equivalent since all we have done is to take a linear transformation of one of the variables. The main difference of this model compared to the previous one is that the variance of the slopes is non-zero, so our conclusion (from the result of the likelihood ratio test) that there are differences between countries in the relationship between hedonism and age no longer seems nonsensical!

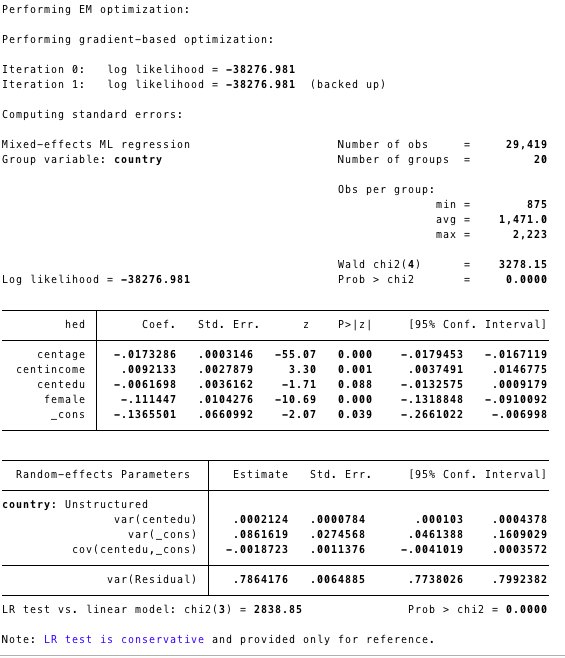
Of course, the disadvantage of this model is that the parameters relating to **agesmall** are not so readily interpretable. For this reason, we will return to using **age**, rather than **agesmall**, in future models. However, let’s interpret the coefficient for age and the variance of slopes of this model. Our estimate for age tells us that on average the hedonism score decreases by 1.781 for every 100 extra years, and our estimate of the variance tells us that the variance of the country differences from this average slope of a 1.781 decrease in hedonism for every 100 extra years of age is 0.204.

**Exercise** - Use the estimates from this model to calculate estimates for a 1 year increase in age.

**Answer** – We need to divide by 100 to find that on average the hedonism score decreases by 0.01780 for every extra year of age. Similarly, we need to divide by 10,000 (100\*100; recall that the variance is the square of the standard deviation and the standard deviation is on the same scale of the slopes) to get the variance of the slopes on the year instead of the hundred scale (it will be 2.04\*10-5).

d) Now let’s study whether there are significant differences between countries in the relationship between hedonism and education. Run the model:

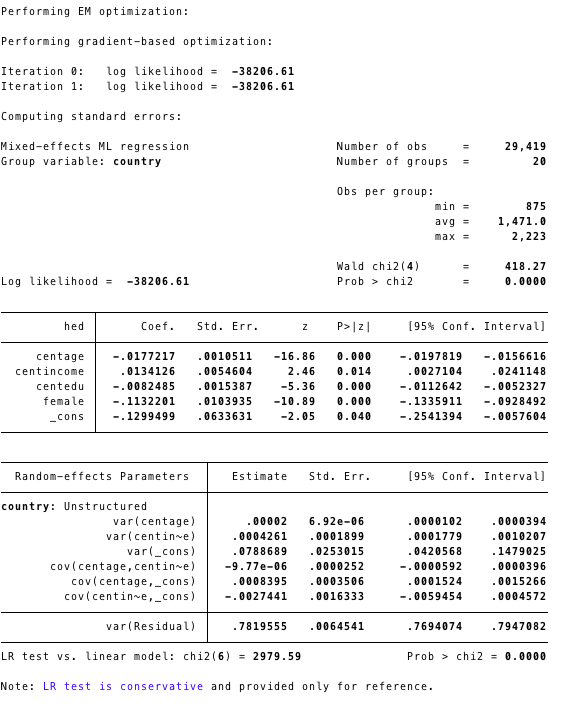
*mixed hed centage centincome centedu female || country: centedu, ml variance cov(uns)*

**

Again, we compare this model to the random intercepts model, so the test statistic is 76643.038 – 76553.962=89.076. A comparison with the x2 distribution with 2 degrees of freedom gives us a p-value well below 0.05, so we conclude that there are significant differences between countries in the relationship between hedonism and years spent in education (*centedu*).

Let's now investigate whether there are significant differences between countries both in the relationship between hedonism and income and in the relationship between hedonism and age. In other words, we'll try fitting a model with random slopes on both income and age, and test whether all the random parameters of this model are significant.

*mixed hed centage centincome centedu female || country: centage centincome, ml variance cov(uns)*



We have a choice when it comes to the likelihood ratio test. We can regard this model, which has a random slope on both income and age (RS2), as being an extension of the model which has a random slope only on income (RS1), or as being an extension of the model we fitted earlier in this section, which had a random slope only on age (note that we have already checked whether the random slope on just income and the random slope on just age are significant.) Which two models we compare using the likelihood ratio test depends on how we're working. If we were building up the model gradually, starting with a single-level model (SL), then going on to the random intercepts (RI) model and then RS1 and then progressing to RS2, we might only want to know whether RS2 was more appropriate than RS1, and therefore just compare these two models. If we were building up the model and had gone from RI to the model with random slopes only on age instead of to RS1, then of course we would want to know whether RS2 was more appropriate than the model with random slope only on age, and would compare those two models. If, on the other hand, we were exploring all possible models thoroughly with the aim of finding the one that fits best we would probably want to know both whether RS2 is better than RS1 and whether RS2 is better than the model with random slopes only on age. We would then perform two likelihood ratio tests.

Let's compare model RS2 with model RS1. The value of -2\*log(likelihood) is 76413.22 for RS2 and 76598.19 for RS1 so the test statistic is 76598.19-76413.22 = 184.97. This time we compare against the distribution with 3 degrees of freedom, because we have added in 3 extra parameters to RS2 compared to RS1: the variance of the slopes on age, the covariance between the intercepts and the slopes on age, and also the covariance between the slopes on age and the slopes on income. The p-value is less than 0.05 so we reject the null hypothesis and conclude that model RS2 is more appropriate than RS1: there are differences between countries in the relationship between age and hedonism on top of differences between countries in the relationship between income and hedonism.

Now perform a hypothesis test to compare model RS2 to the model with a random slope only on age, and interpret the result. To do this, we need to perform a likelihood ratio test. The value of -2\*log(likelihood) is 76413.22 for RS2 and 76444.236 for the model with a random slope on age only so the test statistic is 76444.236-76413.22 = 31.016. We compare this test statistic again the distribution with 3 degrees of freedom because we have added 3 extra parameters (the variance of the slopes on income, the covariance between the intercepts and the slopes on income, and also the covariance between the slopes on age and the slopes on income). This gives us a p-value below 0.05, so we conclude that there are differences between countries in the relationship between income and hedonism on top of the differences between countries in the relationship between age and hedonism.

Let's return to the three extra parameters we introduced in model RS2 compared to model RS1. The variance of the slopes on age and the covariance between intercepts and the slopes on age can be interpreted just as for the model with a random slope only on age. So, we can see that countries with greater slopes on age have greater intercepts: the pattern is of fanning out (again, we need to visualise the model by plotting predictions to be really confident in our interpretation but we'll leave that for now). The covariance between the slopes on age and the slopes on income is negative, so countries with a positive random slope on income tend to have a negative random slope on age and vice-versa. Recall that the average slope for income is 0.013, i.e. positive, and the average slope for age is -0.018 i.e. negative. So, a positive random slope on income adds a positive value to 0.013 resulting in a steeper positive slope, and a negative random slope on age adds a negative value to -0.018 resulting in a steeper negative slope. In other words, countries that have a larger (more steeply positive) effect of income on hedonism tend also to have a larger (more steeply negative) effect of age on hedonism, while countries that have a smaller (less steeply positive) effect of income on hedonism tend also to have a smaller (less steeply negative) effect of age on hedonism.

The covariance between the intercepts and the slopes on age is positive (0.0008), so countries with higher than average (less negative or positive) values of hedonism for men of average age, education and income tend to have a smaller (less steeply negative) effect of age on hedonism. The covariance between the intercepts and the slopes on income is negative (-0.0027), so countries with higher than average (less negative or positive) values of hedonism for men of average age, education and income tend to have a smaller (less steeply positive) effect of income on hedonism.

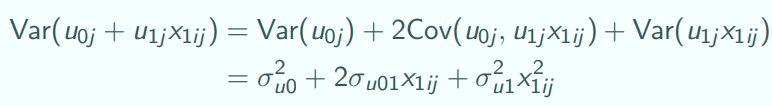
# 3 Calculating the level-2 variance and the VPC

Run the model RS1 (random slope only on income) again.

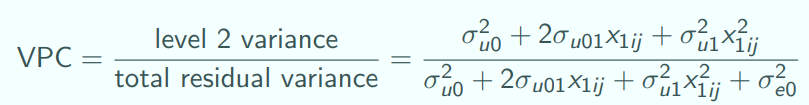
*mixed hed centage centincome centedu female || country: centincome, ml variance cov(uns)*

The level-1 variance = var(Residual) = σ2eo = 0.788.

The level-2 variance is more complicated to calculate. The formula for level-2 variance is:



And the formula for the variance partitioning coefficient (VPC) is:



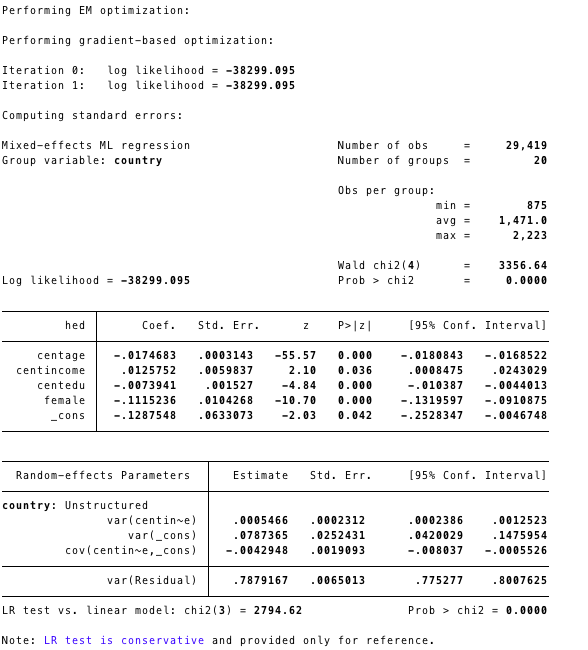
# 

# 4 – Calculating and plotting residuals and predictions

In this section, we will explore our estimated values, looking first at the residuals which are our estimates of how each country’s hedonism differs from the average hedonism (the intercept residuals) and of how each country’s relationship between hedonism and an explanatory variable differs from the average relationship (the slope residuals). We will then go on to use our estimates of the residuals and of the fixed part coefficients to draw graphs of our models.

e) Run again the RS1 model, with a random slope on income:

*mixed hed centage centincome centedu female || country: centincome, ml variance cov(uns)*



We can now calculate the residuals, in the same way as we did for the variance components and random intercepts models.

To calculate the level-1 residuals:

*predict lev1, resid*

To calculate the level-2 residuals:

*predict r1 lev2, reffects*

To summarise the level-2 residuals by country:

*gen b0=\_b[\_cons] + r1 + lev2*

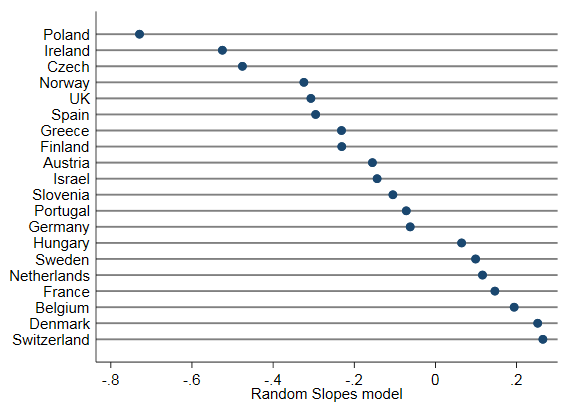
*bysort country: gen tolist = \_n==1*

*list country b0 if tolist*

**

To plot the level-2 residuals against countries, type (in one line):

*graph dot (mean) b0, over(country, gap(country) sort(b0)) cw linetype(line) ytitle(Random Slopes model)*

**

The most hedonistic country after controlling for income is Switzerland (the country with the largest positive residual), while the least hedonistic country appears to be Poland.

To check for model diagnostics:

*kdensity lev1, normal*

*pnorm lev1*

*qnorm lev1*

*kdensity lev2, normal*

*pnorm lev2*

*qnorm lev2*

**Exercise**: Which country is the most hedonistic after controlling for age, income, education and gender and allowing for the relationship between hedonism and age to differ across countries?

[Answer: Denmark]

*mixed hed centage centincome centedu female || country: centincome, ml variance cov(uns)*

Examine intercept and slope residuals for countries:

*estat recov, corr*

*predict u1 u02, reffects*

*egen pickone=tag(country)*

*scatter u1 u02 if pickone==1, yline(0) xline(0) ///*

*ytitle("Slope of income (u1j)") xtitle("Intercept (u0j)")*

*drop predhed*

*predict predhed, fitted*

Add a random coefficient for gender:

*mixed hed age2 income2 edu2 female || country: income2 female, ml variance cov(uns)*