# Fixed and random effects models

## Transcript video 2

Full resource: https://www.ncrm.ac.uk/resources/online/all/?id=20842

Hello again. So, in this second presentation we are going to look at some worked examples of modelling fixed and random effects in this Stata, so we will look at a specific example and we will talk through the interpretation of fixed and random effects, and then we will look at some more advanced options.

So, to do this, we will need some data, and this slide shows a table of data that has been adapted from an example from Rabe-Hesketh and Skrondal 2008, where they modelled wages as the outcome variable. These data have an individual person identifier, so this is a variable NR. And we can see here that there are 545 unique cases on this variable. And this is 545 individuals observed at 4360 different occasions, so each individual in these data has been observed on 8 different occasions and we therefore have a balanced panel. The data are from the USA and they control for dummy categories, for Black ethnicities and also Hispanic ethnicities. In the USA, this is often termed as dummies for race. In the UK we’d probably describe these as dummy variables for ethnicity. The variable experience captures years of experience in the labour market, and this is the code EXPER from the table, so that’s a variable experience. There’s a variable that is a dummy variable for marriage, so capturing whether an individual is married or not in these data, and there is a variable for union, which codes whether an individual is a member of a trade union or not. As I mentioned, the outcome variable is wages, so it’s a log of wages in these data, and we have a variable that is EDUCT, which is years of education beyond the high-school level, so EDUCT is that variable. And there is an additional variable called YEART, which is the year from 1980. So, these are the data that we will be modelling in our examples.

Now, this slide indicates the code in Stata for fitting a fixed effect and a random effect model and then undertaking a Hausman test to compare the fixed and random effect results. The XTREG command in Stata can be used to fit both fixed and random effect models, and it is also common practice after doing this to compare the models using the Hausman effect. So Rabe-Hesketh and Skrondal 2008 provide a technical explanation of what the Hausman test is undertaking, but Allison 2009 provides a pithy definition of the Hausman test, explaining that Hausman test, the hypothesis that fixed effect coefficients are identical to the random effect coefficients. If they are identical then ordinarily we would prefer the random effects model because it also provides correct standard errors. If the coefficients are not identical then we may prefer the fixed effect model because, theoretically, the coefficients are considered to be unbiased.

So, this table provides modelling results from Stata comparing fixed and random effects models of our data. So, in comparing the fixed and random effects, it can be seen that the variable Black and Hispanic have not been estimated in the fixed effect model, they have dropped out of the model. So, they are time-constant and time-invariant, so they have been excluded by the fixed effect approach, which cannot estimate time-invariant variables. The variable education has also dropped out of the model as this is also time-invariant. YEART, which is year since 1980, has dropped out of the model because it is defined by an individual level constant related to the variable experience.

So, at this point it may seem that the random effects model is preferential because of the greater possibility to estimate substantively interesting associations. But if we were to compare the individual level covariates for union, married and experience between the fixed effect and random effect models, we notice that the random effects differ substantially from the fixed effect estimates. So, in a fixed effect model we can see that the parameter for union is 0.084, but for the random effect is it 0.111. For the married variable, the parameter is 0.061, but in a random effect the equivalent parameter is 0.076. In the fixed effect framework the parameter for experience is 0.06 and in the random effects is 0.033. So, if we accept that the fixed effect estimates are consistent and unbiased, then it appears that the random effects model estimates are likely to be biased by correlation with unobserved variables.

And indeed this is suggested by the results of the Hausman test. So, this slide shows the output for the Hausman test in Stata, and circled in red is the P value of significance. And we can see that the P value here is 0.0165, which is below 0.05, so it is showing that there is a statistically significant difference in the values for the parameters between the fixed effects and the random effects model.

Now, there are some alternatives, and a growing number of alternatives, to undertaking fixed and random effects analysis in this way. There’s a growing body of work demonstrating the possibility of estimating consistent fixed effect style estimates within a random effects framework. So, for example, Mundlak 1978 showed that the inclusion of cluster means for all within individual covariates can enable consistent estimation of within effects in a random effect framework. Allison 2009 put forward a hybrid model similar to that suggested by Mundlak using a group mean centring approach, while Bell et al similarly suggest an approach where XIT is divided into two parts, each with a separate effect. One part represents the average effect of XIT, the second part represents the average between effect of XIT.

In these cases an additional parameter is reported in the model output, and this represents the effect of the time-invariant variables between effects, so you get an additional term in your model that represents a between effect. And we can take a look at this. So, this table here shows a fixed effect and random effect output that we’ve all just seen, along with a Mundlak specification of the model. So, we can immediately see that there are two additional parameters, mn union and mn married for the cluster means, i.e. the individual means of within the individual covariates union and married. And it can be seen that the Mundlak estimates for union and married are identical to the fixed effect estimates, so this can be seen within the red boxes. So, for example, if we look at the fixed effect model, the parameter for union is 0.084, and in the Mundlak specification, parameter for union is also 0.084. In the fixed effect model the parameter for married is 0.061 and in the Mundlak specification the parameter for married is also 0.061.

However, the estimate for the variable experience or EXPER varies substantially between the Mundlak and the fixed effect model. Now, in this model experience is an age effect on ages, whilst years since 1980 is a period effect so it’s the YEART variable. It can be seen that summing these two estimates gives exactly the 0.06 reported in the fixed effect model as the estimate for EXPER or experience, so that’s given in the green box here. So, in the Mundlak specification of the model, the parameter estimated for experience is 0.028 and the parameter estimated for YEART is 0.032, and if we sum these two together you get exactly the 0.06 that is the estimate for experience or EXPER in the fixed effect model.

So, which approach should we choose when we’re undertaking fixed and random effect modelling? Well, classically, in econometrics the preference is for the fixed effects approach, and you can read this if you pick up the book Allison 2009. So, in there, Allison argues that individual level variation is likely to be correlated with the unobserved characteristics of individuals, and partialling out the contaminated variation produces approximately unbiased estimates, so therefore, in general terms, we prefer a fixed effect approach. But, by contrast, Bell and colleagues argue that a well-specified random effects models provides everything that a fixed effect model provides and more, making this the superior method. And that appears to be what we see in this specific example that we’ve shown today, so yeah, the Mundlak model seems to give us everything that the fixed effect model does, and more, and also has consistent estimates.

So, general recommendations, so as analysts we probably want to think about what we want to know substantively, and we probably want to ask whether we’re interested in the between effects or the within in between effect associations, or whether you’re interested in time invariant associations, and if you’re interested in these then probably a fixed effect approach isn’t going to get you very far. So, as analysts, it’s always worth undertaking a sensitivity analysis and trying out different specifications of models to try out a range of fixed and random effect approach, and also reporting on the alternative methods, perhaps in an appendix for a paper. And ultimately, as analysts, it’s up to us to provide robust and considered explanations for the approach that we’ve chosen.

So, here are the references that I’ve used in this presentation, but there’s a whole set of further references available in some of the other documentation for the resource, and I hope that you’ve found this resource useful for your work. Thank you.

National Centre for Research Methods (NCRM)
Social Sciences
Murray Building (Bldg 58)
University of Southampton
Southampton SO17 1BJ
United Kingdom

**Web** www.ncrm.ac.uk
**Email** info@ncrm.ac.uk
**Tel** +44 23 8059 4539
**Twitter** @NCRMUK