xtsum2, a new tool to analyse variance in Stata

March 23, 2021

Contents

1 Abstract

The decomposition of variance is one of the fundamentals in modern econometrics, whatever the aim of the analysis. The statistical software, Stata, provide a lot of commands to make that, as `mixed` `xtreg` `anova` `xtsum` `loneway`. These programs compute the components of variability as partial standard deviations, partial variances, or partial sums of squares. In this article, we introduce a new command, Xtsum2, that in addition to those magnitudes calculate the percentage shares of the variability components related to total variability. Moreover, in the output are collected most of the information can help us to understand how our response variable is nested within such a categorical variable. We present the program, showing how it works in different contexts, using different datasets from hierarchical to panel as well. Thereafter we compare it with the command already available in Stata, to show the wherewithals of this new command, which avoids unnecessary and annoying steps.

2 Introduction

In data analysis, one of the biggest problems is the lack of independence among the observations of the dataset of our interest, which we could figure out when the observations are gathered without guaranteeing in such a way that they are completely random.

A particular case of lack of independence is when we omit to consider some clustering process, the observations in the dataset of our interest, could be similar to each other belonging to the same group. If we don't consider this possibility, it could drive us to estimate models with less efficiency, not valid or even biased, according to the preeminence of the clustering in the variables of our interest and especially when concerning the aim of the analysis.

For instance, if we are going to analyse salary, in such a sample of workers belonging to different firms, probably no concerning the clustering process will lead us to lack of important information. The firm which the workers are nested could have different company policies concerning the salary, for example, a firm would prefer to give their employees a minor salary opting to rather more investment in corporate welfare, or another one would prefer to pay more dividend to the shareholders instead of a higher salary. There could be other million reasons about the choice of different corporate policy, that get as results differents level of salary.

In this example the clustering process is evident and unless the aim of the analysis is among firms very similar to each other, the categorical variable that identifies the belonging to differents firm is better to be included. In other situations, could be no longer so clear if we have to concern a particular clustering process, as in the example that follows.

Suppose that we are now interested in student's knowledge and we know that the students are registered sorting by the last name, in the register. A possibility is to think that the observations are nested in letters classes, in other words, we could think that the knowledge of a student which the last name starting with A is different from a student that the last name starting with B, and a student that starting with C, and so on until Z. In the latter example, with high probability, there's no necessity to add the letter-nesting information to the analysis of knowledge, but we can't reject first, despite we lead to think that it doesn't work. Indeed some studies figure out evidence that the last name could get on differents features (one of them is Claudia Olivetti, 2013), so why the letter-nesting should not affect the student's knowledge? Maybe the pupils which have a surname starting with A or another of the first letter, are questioned first than the others in the lower half of the alphabet and could lead them to more study.

It is undeniable that the two examples are poles apart, but between them, there is a range of shades of grey. It is important to use a method that allows us to understand how and how much a clustering process affects the observation which we are interested in.

This is a huge reason for which we are interested in the variance components but there are others, the aim of a lot of analysis is proper how a variable is distributed among such nesting. For instance, we could want to analyse the differences in the GDP per capita among different countries. In another case, we are interested in the effects of different medicines tested on a sample of patients, which are divided into groups according to the medicine administered. In those situations, we are even more interested in the components of variance, and in all situations in which the objective of the study lies in the way that the observations are nested.

As well known variance gets us a lot of information concerning the distribution of a certain variable, but we can think about that as a total from which we want to split out its components. That decomposition, as we will show, leads to close off the components attributable to the clustering process and the parts that don't.

Since that, we can estimate the components of variance that is very useful but remain an absolute measure it is unbounded from its total. Often is even more useful a relative measure to these magnitudes as a ratio between a component and its total or the related percentage share.

A measure of that is the intraclass correlation coefficient, which measures which amount of observations nested in the same class are correlated. As we will show, it is equal to the ratio between the between classes variation on the total of variance.

There is a lot of commands, as we will show, available in Stata to estimate the components of variance, and in some of them the intraclass correlation. The innovation of `xtsum2` is that calculates the components of variance in terms of standard deviations and of percentage shares too.

In section 3 we show how theoretically the variance is decomposed, following Fisher's decomposition. We demonstrate that the total variance could be decomposed in its components and how to obtain the percentage shares of them.

In section 4 we present the variance component models, which actually regress the response variable against a categorical variable that identifies the clustering. Those models are, fixed effect model, random effect model, and two kinds of mixed model, the first using full maximum likelihood approach and the second using the restricted maximum likelihood one. We highlight the strength and the weakness of each method according to the different aims of analysis.

In section 5 we focus on the command presented in this paper, we show the syntax and how to use it, highlighting the options which allow different methods of estimations. Finally, we explain the notation used by `xtsum2` that is the same used in all the paper and the help file.

The latter, but not the least, is section 6, in which we present some examples, using different datasets. We apply the command to these differents situations focusing on the choice of the best method to estimate the components of variance, in each one case.

In the first example we make a comparison among the various methods of estimation, showing the best one in that case to estimate the components of variance. In the next examples, for no turn out to be redundant, we avoid the comparison to focus on the choice of the best method and comment it.

Finally, we leave to the reader an appendix section, in which could find all the Stata statements used in this paper, the statements proper of the command, and statements for the help file.

3 The decomposition of variance

The variance of response variables, whose observations are nested within one or more categorical variables, can be decomposed. It is very important to analyse in which dimensions have concentrated the variability of our variables of interest.

We present, from a mathematical standpoint, how the variance is decomposed, we follow the method of Ronald Fisher, the first who developes the analysis of variance that we all know now (Searle 2016) Before starting is necessary clarification about the notation used from now on, since we'll generalize from hierarchical dataset to longitudinal as well.

As said before, the data that we analyse are nested within a categorical variable, we could think about the latter as a distinction between two levels. The observations occupy the first level and the classes to which the observations belong occupy the second level, for instance, the workers assemble the first level and the firms where they work are units of the second level, the schools occupy the second level, and the students the first one. The above could be applied to panel data as well, but with further clarifications, panel data can be seen as a hierarchical one where the higher level is the individual level and the lower one is made up by the temporal observations of each individual. We refer to the single unit of the first level as j and the ones of the second level as i, each class is composed by J_i, according to the size of the cluster i and the number of classes is given by N.

There are two ways to decompose variance, the first one by splitting the share of variance into between and within classes, and the second one divides the variation into three parts, between first level, between second level, and within first-second level.

Let's start to analyse the total sum of squares, a measure of variability, that is the sum of square deviations from the overall mean for every i and every t (Bontempi M. E., Golinelli R., 2014)

$$TSS = \sum_{i=1}^{N} \sum_{j=1}^{J_i} (x_{ij} - x_{..})^2 \tag{1}$$

Then we can get its two transformations that make ahead for the two relative ways for decompose variance, explained before, let's start to see the first one.

By adding and subtracting the individual mean across the time into the square, we can see it as two separate components, the deviation of the observation of the individual i at time t from the average of that individual i over the time and the deviation of the average of individual i over the time from the overall mean.

$$= \sum_{i=1}^{N} \sum_{j=1}^{J_i} (x_{ij} - x_{..} + x_{i.} - x_{i.})^2$$

by solving the square and separating the terms of the summation we get

$$= \sum_{i=1}^{N} \sum_{j=1}^{J_i} [(x_{ij} - x_{i.} + x_{i.} - x_{..})^2 + 2(x_{ij} - x_{i.})(x_{i.} - x_{..})]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{J_i} (x_{ij} - x_{i.})^2 + \sum_{j=1}^{N} \sum_{i=1}^{J_i} (x_{i.} - x_{..})^2$$

where the duble product go to 0 because of the property of the mean in which the summation of the deviations from the mean is equal to 0.

$$\sum_{i=1}^{N} \sum_{j=1}^{J_i} (x_{ij} - x_{..})^2 = \sum_{i=1}^{N} \sum_{j=1}^{J_i} (x_{ij} - x_{i.})^2 + \sum_{i=1}^{N} \sum_{i=1}^{J_i} (x_{i.} - x_{..})^2$$

In the second term, the summation by Ti simplifies because there are no values that change over time. These two components are the within individuals sum of squares and the between individuals sum of squares

$$WSS^N = \sum_{i=1}^{N} \sum_{j=1}^{J_i} (x_{ij} - x_{i.})^2 \tag{2}$$

$$BSS^N = J_i \sum_{n=1}^{N} (x_{i.} - x_{..})^2 \tag{3}$$

with similar calculation we can obtain the within temporal periods sum of squares and the between temporal periods sum of squares, replacing the individual means with the average of the specific periods

$$WSS^j = \sum_{j=1}^{J} \sum_{i=1}^{N_j} (x_{ij} - x_{.j})^2 \tag{4}$$

$$BSS^j = N_j \sum_{j=1}^{J} (x_{.j} - x_{..})^2 \tag{5}$$

Let's see the second way to decompose variance, starting from equation (1) we add and subtract the average of individual i over periods, the average of period t over individuals and the overall mean.

$$\sum_{i=1}^{N} \sum_{j=1}^{J_i} (x_{ij} - x_{..})^2 = \sum_{i=1}^{N} \sum_{j=1}^{J_i} (x_{ij} - x_{..} + x_{i.} - x_{i.} - x_{.j} + x_{.j} + x_{..} - x_{..})^2$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{J_i} [(x_{ij} - x_{i.} - x_{.j} + x_{..}) + (x_{i.} - x_{..}) + (x_{.j} - x_{..})]^2$$

For the same reason see before, making explicit the squares, the double products go to 0 and only the squares remain.

$$
\begin{aligned}
&= \sum_{i=1}^{N}\sum_{j=1}^{J_i}(x_{ij} - x_{i.} - x_{.j} + x_{..})^2 + \sum_{i=1}^{N}\sum_{j=1}^{J_i}((x_{i.} - x_{..})^2 + \sum_{i=1}^{N}\sum_{j=1}^{J_i}(x_{.j} + x_{..})^2 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J_i}(x_{ij} - x_{i.} - x_{.j} + x_{..})^2 + \sum_{i=1}^{N}\sum_{j=1}^{J_i}((x_{i.} - x_{..})^2 + \sum_{i=1}^{N} J_i(x_{.j} + x_{..})^2
\end{aligned} \tag{6}
$$

Now we can isolate three different parts, the second term is the between classes deviation like in equation (3), the third term is the between first level deviation like in equation (5) and the first one component is the within individuals-temporal periods sum of squares, named residual sum of squares as well, that represents both the heterogeneity, individual and temporal.

$$
RSS = \sum_{j=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{i.} - x_{.j} + x_{..}\right)^2 \tag{7}
$$

Because is the within individual deviation less the between temporal deviation, so intuitively the individual variability depurated from the part of that attributable to time (like economic cycle). At the same time, rearranging the terms it is the within temporal deviation less the between individual one, that is the internal temporal variability depurate from the part of it attributable to individuals (the permanent individual characteristics).

This term is very important because captures both the temporal and individual unobserved heterogeneity, which features those characteristics proper to individual and every period that cannot be explained by the temporal and individual means of our variable of interest.

The sum of squares is a measure of variability but is not related to its specific degrees of freedom and is not express in the same unit of measure of the data, because is a square operator.

The standard deviation instead has both the characteristics so we can get more information about our data. We obtain this magnitude by dividing the sum of squares by their degrees of freedom and then taking the square root of it.

Before we see the formulas for the total sum of squares and its components, now from these we can calculate the relative standard deviation relating to the specific degrees of freedom.

$$\sigma^{r} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{..}\right)^2}{NJ_i - 1}} \tag{8}$$

$$\sigma^{wi} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{iJ} - x_{i.}\right)^2}{N\left(j_i - 1\right)}} \tag{9}$$

$$\sigma^{bi} = \sqrt{J_i\frac{\sum_{i=1}^{N}\left(x_{i.} - x_{..}\right)^2}{J_i\left(N - 1\right)}} = \sqrt{\frac{\sum_{i=1}^{N}\left(x_{i.} - x_{..}\right)^2}{\left(N - 1\right)}} \tag{10}$$

$$\sigma^{wj} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{.j}\right)^2}{J_i\left(N - 1\right)}} \tag{11}$$

$$\sigma^{bj} = \sqrt{N\frac{\sum_{i=1}^{T_i}\left(x_{i.} - x_{..}\right)^2}{N\left(J_i - 1\right)}} = \sqrt{\frac{\sum_{i=1}^{J_i}\left(x_{i.} - x_{..}\right)^2}{\left(J_i - 1\right)}} \tag{12}$$

Now we can have much more information about our data with these measures of variability more interpretable than the sum of squares.

Furthermore, a tool for variance analysis is to know which of the components weigh more on the total variance. To do that we have to start again from the total sum of squares, that as said before we can see that in two way, the first split out into two components, the within classes one plus the between classes one and the second split out in three, the between first level, the between second level and the residual.

What we have to do is to calculate the percentage shares between the partial sums of squares and the total sum of squares.

$$PS^{wi} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{i.}\right)^2}{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{..}\right)^2}100 \tag{13}$$

$$PS^{bi} = \frac{J_i\sum_{i=1}^{N}\left(x_{i.} - x_{..}\right)^2}{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{..}\right)^2}100 \tag{14}$$

$$PS^{wj} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{.j}\right)^2}{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{..}\right)^2}100 \tag{15}$$

$$PS_{bj} = \frac{N\sum_{i=1}^{J_i}\left(x_{i.} - x_{..}\right)^2}{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{..}\right)^2}100 \tag{16}$$

$$PS_{r} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{i.} - x_{.j} + x_{..}\right)^2}{\sum_{i=1}^{N}\sum_{j=1}^{J_i}\left(x_{ij} - x_{..}\right)^2}100 \tag{17}$$

4 Variance-components models

After understand the magnitudes that we are looking for, we introduce in this section how the variance components are going to be estimate, such the models that allow us to do that. Specifcly we show the manner to analyse variance using the statistical software Stata.

These models are named variance-components models, because work with nested data, no covariates or explanatory variables are included and the aim of them is proper the variance components (Brown 2017).

In this discussion we don't care about the relationship between the dependent variable and some explanatory, that would be put off at the time of the choice of the right model for analyse the data of interest.

We are interested in a certain response variable, whose observations could be or must be analysed as the units of clusters. A model like that, without considering the clustering process, is simply the response explained by its overall mean and the deviation from the latter. However whether the residuals are correlated within classes, there is an obvious problem of lack of independence between the observations and will lead our estimation to be biased (J.J.Hox 2010).

What we have to do is to split the deviation from the overall mean into two uncorrelated terms, the first is a class-specific parameter, equal to all units of the same class and the second is an idiosyncratic component and changes among cluster and units of a cluster.

In that way, we can estimate the components of our interest, the variance of the first component is interpretable as the between classes variance and the variance of the idiosyncratic component as the within classes variance. (Rabe-Hesketh A.Skrondal, 2012)

There are a lot of methods to decompose variance, each one with its strengths and weakness, we present them showing when a model is more suitable than another according to the data that we are interested in.

4.1 Multilevel model

The multilevel regression model finds application only to the data that by their nature could be clustered, so which have a certain hierarchical structure and the observations could be nested within specific classes.

We can extend this model to panel data too, if we think about them as hierarchical, with the second level represented by individuals and the one represented by the observations at different times per individual.

To perform this model, we need to implement more than one regression, one per dimension. In this context, we focus only on the dataset with two levels because the program that we discuss works only with this kind. So we present the `mixed` model with two regressions, as well as the command's action field.

The first regression is implemented at the second level dimension, in which the response variable is the random intercepts, which will be used in the second regression. The intercepts are explained by a coefficient that doesn't change among the levels and an error term that varies among the i-classes.

$$\beta_{0i} = \gamma_{00} + u_{0i} \tag{18}$$

The second regression, indeed, takes the random coefficient $\beta_{0i}$from equation (18) to explain the variable that we want to analyse.

$$y_{ij} = \beta_{0i} + e_{ij} \tag{19}$$

As we can see at the observation level there is an error term that now, varies among both levels. Now with simple steps, we can replace the expression of the coefficient given by the equation (18) into the equation of interest (19), to obtain:

$$y_{ij} = \gamma_{00} + u_{0i} + e_{ij} \tag{20}$$

In this last form, we can notice that the variable of our interest is explained by a constant equal for each i and each j, a component of error equal for each j that belongs to the same cluster, and another component which changes both for j and i.

All of these coefficients estimated in the two different equations are implemented with the maximum likelihood method, which finds out the coefficients that have generated the data with the highest probability. This estimator maximizes the likelihood function.

There are two approaches to do that, the first, called full maximum likelihood (FML), and the second restricted maximum likelihood (RML), each of them with its strength and weakness.

We can think of the full model as an estimation of the mean and the variance components, using the maximum likelihood method, in which the variance estimator will be adjusted dividing the sum of squares by $n - s$, where s is the number of degrees of freedom, that otherwise will be biased.

Another way to arrive at the same result, keeping the bias adjustment, is to use the maximum likelihood to estimate the deviations of the observations from the mean. These residuals have a singular multivariate normal distribution, and the estimation of variance components is equivalent to the bias adjustment of the full maximum likelihood model (J.de Leeuw E.Meijer 2007). For this reason, this variant of the full model is also called residual maximum likelihood

In practice the FML maximizes the likelihood function letting both the coefficients and the components of variance vary, the RML does this but proceeds in two steps and includes in the function no coefficients, which will be estimated at a later time.

The responses are transformed in such a way that their distribution no longer depends on the regression parameter β_{0i}. The likelihood then depends on the parameters of the random part of the model (S.Rabe-Hesketh, A.Skrondal 2012) The full maximum likelihood is certainly easier to calculate, but treat the coefficient as fixed, so loses degrees of freedom, the restricted one estimates the coefficients only after removing fixed effects. Based on this the full model may be biased especially in the small samples (J.Hox 2010)

The two methods however are used both, because treat coefficients as fixed have the other side of the coin, which allows to comparison with other models. The restricted model, since doesn't include the coefficients into the likelihood function, is not to be able to do the test on the fixed parts of two models compared.

After have been estimated the variance components it is appreciated to get a measure of the significance of them. Mixed models estimate the standard errors for whole the variance components, in the three-way decompositions method too, we might be tent to use them to involve a t-test, making a ratio between each value and its standard error. Unfortunately, that is not appropriate because the space of variance components goes from zero to $+\infty$, instead, the t-statistic is thought to test values that may go from $-\infty$ to $+\infty$, which in that case have sense to test if the parameter is equal to zero, but perhaps we apply to the variance components we are testing the absence of between variance instead of its significance.

Another way to reach our aim is the log likelihood ratio test, which, as the name suggests, is allowed only in the context of maximum likelihood estimations. This test is especially suitable with our objective because occurs properly on parameter those are restricted to be non-negative. The idea behind it is to compare between two models, the first including the factor variable that identifies the class level and a second, restricted, without that. (R.G.Gutierrez, S.Carter, D.M. Drukker 2001)

Performing the procedure in Stata we can use the command `mixed` to estimate this model, using no option the program implements the maximum likelihood approach by default, or you can add the option `mle` and will bring you to the same result. Instead, if we want to compute the model with the restricted maximum likelihood, we just add the option `reml`. The syntax of these commands is a little bit complicated, please refer to [ME]mixed and Marchenko (2006) for details, that lack in this paper.

4.2 Fixed effect model

The fixed effect model, is based on the idea that there is a different intercept for each individual at the first level. The regression that does this is the Least Squares dummy variable one (LSDV), it simply adds at the pooled OLS model, N dummies variables, one per group of the second level.

There is an obvious problem to estimate this model, there are too many regressors, according to the size of N, and losing that size of degrees of freedom (J.Wooldridge 2017). A solution to bypass the problem is the within transformation, which allows us to keep the information of the clustering without losing degrees of freedom.

This transformation start from two different regressions to arrive at one, by difference between the two. The first one is the LSDV, represented in equation (21), so a simple response variable explained by the effect of each class at the second level, u_i, so insert the dummies, and an error term eij, which varies at the observation level

$$y_{ij} = u_i + e_{ij} \tag{21}$$

The error terms have the following characteristics, every e_{ij} has the same variance σ_e^2, are independently and identically distributed and that pairs of different e_{ij} have zero covariance. (Searle, 2016) The equation below, (22) is the between regression equation, in which we force the data to be cross-section suppressing the first level into its means. The response variable is the partial mean of each class at the second level and is explained by the mean of the partial means, so the overall mean and the means for each class of the error terms.

$$\overline{y_i} = \overline{u_i} + \overline{e_i} \tag{22}$$

Now subtracting the equation (22) from the equation (21) we get the fixed effect (or within) transformation

$$y_{ij} - \overline{y_i} = e_{ij} - \overline{e_i} \tag{23}$$

This last equation gets as response variable the deviation of the observation from the partial mean of the class to which it belongs. The right term of this equation is the deviation of the error term from the mean residual of the class to which the error belongs.

After gets this transformation we apply the OLS estimator to find out the parameter of the model and as we can see no dummies variables are inserted, however, we get the same estimate of the coefficient of LSDV keeping the degrees of freedom. In this way, we capture all the information that we need about our variable of interest related to the classes. We can estimate the fixed effects as in LSDV and then compute the variance of those as the between classes variance.

As we see, the model estimates the coefficient as fixed, theoretically the datasets suitable to this kind of estimations are the ones in which the observations are nested in specific and finite classes. In other words, when the classes are proper those that we want to analyse, they are not a sample of the population of the classes. For instance, if we are interested in the distribution of salary among age-groups, we don't choose them randomly in a sample of age-groups but we choose the ones that focus the aim of our analysis. Instead, if we want to study the salary among firms, these are taken from the population of firms randomly and in this case, the fixed effects are no longer the best way to estimate variance components. (Searle, 1971)

The fixed effect model of course doesn't allow the log likelihood ratio test such in the multilevel model, because it is not estimated using the maximum likelihood method but the ordinary latest square one.

In this case, the more appropriate test, being the coefficients treat as fixed, is the F-statistic. It verifies if the coefficients of the dummies, those identify the differents classes, are jointly equal to zero. The idea behind this test is similar to the loglikelihood ratio one, it compares the pooled ordinary latest square model, as the restricted one, with the latest square dummies variables model. In a panel data context, this test could be made on the coefficients of time dummies variable as well. Stata do all of that in a unique command, `xtreg` using the option `fe`, so we can implement this model easily.

4.3 Random effect model

The random effect model, like the previous one, it's made up of one regression.

$$y_{ij} = \beta_0 + v_{ij} \tag{24}$$

where

$$v_{ij} = u_i + e_{ij} \tag{25}$$

The error term is composed, the first part u_i varies only among the second level while the second one, e_{ij} varies for the lower one. We have to attribute some probability properties to these part of error terms, all e_{ij} have the same variance σ_u^2, the first moment is equal to zero and the covariance between e_{ij} too. Regarding the class-component error terms u_i we make similar considerations, they are independently and identically distributed, haveing zero mean, and then that all have the same variance σ_u^2 (Searle, 2016)

The random effect estimator uses a GLS transformation, in which the observations are weighted for the degree of intra-class correlation. To do that we have to transform the equation (24) using the θ parameter, which is equal to zero if there isn't intra-class correlation and equal to 1 if there is a total correlation among classes (Verbeek M. 2017).

We subtract to each term of the equation (24) the product between θ and the class mean of that term since the mean of β_0 is zero the equation become

$$y_{ij} - \theta\overline{y_i} = \beta_0\left(1-\theta\right) + \left(v_{ij} - \theta\overline{v_i}\right) \tag{26}$$

The θ parameter will be estimate using the formula

$$\hat{\theta} = 1 - \sqrt{\frac{\widehat{\sigma_e^2}}{\widehat{\sigma_e^2} + J_i\widehat{\sigma_u^2}}} \tag{27}$$

We face suddenly a problem, how gets consistent estimates of the variance at the first level σ_u^2 and at the second one σ_e^2 We need to implement an auxiliary regression, using pooled OLS model. From this model, we save the estimates of the residuals and using for estimate the first level variance (Wooldridge J.M.2020)

$$\widehat{\sigma_u^2} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{s=j+1}^{J}\widehat{v_{ij}}\widehat{v_{is}}}{[NJ(J-1)/2]} \tag{28}$$

and then we find by difference the last component

$$\sigma_e^2 = \sigma_v^2 - \sigma_u^2 \tag{29}$$

Now we have all the elements for a consistent estimate of θ and so we could make up the transformation as in equation (26).

We could estimate the random coefficients in two different ways from equation (26), the first one is the ordinary latest squares that minimize the sum of squares of the residual. The second one is the maximum likelihood method that indeed maximizes the likelihood function. It reaches the coefficients that with the highest probability generate the data that we analyse.

In contrast to the case in which are suitable use fixed effect models, the random effects one works well in the dataset where the observations are nested within some categorical variables in which the classes are taken randomly from a population of classes. Resume the previous instance, the case in which we want to analyse salary among firms is a perfect example of analysis that is appropriate to estimate variance components as random coefficients (Brown W.J. 2017).

Sometimes is no longer easy to decide what kind of effect to choose, as in the previously examples that it is evident. In panel data, for instance, we could think to fit temporal periods as fixed because they are well known to be not random, but in some cases, we have to do because of the aim of the analysis. If we are interested in the level of fine dust in the air and we have collected observations in different cities over differents years, is more appropriate to treat the temporal factor as random and the cities as fixed, because the purpose of this analysis is not in the specifics years but the differents observations gathered per city, in different occasions. keeping in mind the covariance between successive years.

There are also cases in which the years instead need to be estimated as fixed, but always when the objective of the analysis is in some specifics periods.(Searle,2016)

The random intercept model estimated with OLS method doesn't allow neither the log likelihood ratio test nor the F-statistic one. It's necessary to figure out another test specifically for this kind of model, fortunately, we can use the Breush-Pagan test LM test. This belongs to the Lagrange multiplier based on the type of test, which tests the effect on the first-order conditions for a maximum of the likelihood of imposing the hypothesis (T.Breush,A.Pagan 1980). This particular case of Breush-Pagan test allows us to compare the model using the categorical variable and the model without it, analysing the effect on the first-order condition between them.

In Stata, all the steps to arrive at these two results are done automatically using the `xtreg` command (see [xt]xtreg). The option `re` allows the ordinary latest squares methods and the option `mle` the maximum likelihood one.

5 The xtsum2 command

`xtsum2` decomposes the variance of a single variable or a set of variables, identify as `varlist` using both two ways shown above, reporting the standard deviations and the percentage shares of the specific component of variance.

All illustrated in a very intuitive table, in which the top part reports the values of the second way of decomposition of variance, so the components between individuals, between temporal periods, residual and the bottom part reports the within individuals component.

The commands need to know how the dataset is structured, so we have to specify before running which are the factor-variables that identify the classes. The Stata command which makes that is `xtset` (see [xt]xtset), it is necessary that it is the last command executed before `xtsum2`, specifying the two dimensions that we want to consider for decompose variance, otherwise the program return error.

The factor-variables that identify the dimensions of the dataset, which are inserted in the previews command `xtset`, must be numeric and its label no longer than 32 characters. Moreover, that variable has to be enumerated in progressive order from 1 to j_i for each category of it, if not, we suggest generating a new variable according to these features. We give a plain application of this problem in example 6.6. Furthermore, the factor variable that identifies observation level must be more numerous than the one that identifies the classes in which the observations are nested.

`Xtsum2` don't work with time-series operators, we can't put the lags of any variable into `varlist`. We need to generate the lags with the commands available in Stata (see [U]tsvarlist) before running this program. In example 6.4, to give a clear application of this problem, we'll show the necessary statement to use the time series operators into `xtsum2`.

The command shows two other sections that are useful when we analyse unbalanced data. The second on the left reports the number of observations, the total for both two dimensions and then for each one individually, shows the total, the minor, the grater, and the medium.

The last section shows differents statistics according to the model option chosen, for fixed effect model in this section there is the F-statistics, on both classes and observations level. The random effect model gives the Braush-Pagan test and the multilevel models give the likelihood ratio test. All of them are better explained in section 4.

In addition to those tests for all the model options, make a statistic for the balancedness degree.

The null hypothesis in the up part of this section is that the individual dummies are all equal to zero and in the down part is that the temporal dummies are all equal to zero.

5.1 Sintax

`xtsum2` *`varname`* `[if] [in] [,options]`

Options

`Common`: allow to do the analysis only on the observations common to each variable insert in `varlist`.

`fe`: The program estimates the variance components via fixed effect model. This option is the default.

`re`: The program estimates the variance components via random effect model.

`mle`: The program estimates the variance components via mixed model using full maximum likelihood approach.

`rml` The program estimates the variance components via mixed model using restricted maximum likelihood approach.

`ur` The program adds the unit-roots test in the test section.

5.2 Notation

An explanation with the notation is necessary, to understand well what the program results mean because in literature there's no longer a unique standard to discern the structure of the data and it returns in the output and in the stored results of the program. The same notation will be used while showing the examples.

The command thinks the dimensions of the dataset that it analyses in this way, at the second level we find the cluster dimension, which could be the classes in which the pupil are nested, the firm which the workers lend work, or the countries where the individuals live and so on. The single class is identified with the letter i, for a total of N classes.

The first level is represented by the individuals that belong to the class i, and they are identified as j, for a total of J_i individual for the class i. That can be the Ji pupils that belong to the class i, the J_i workers of the firm i, or the J_i people that live in the country i.

This notation could be extended to panel data as well, in which the individual i, which can be like for hierarchical the countries or workers or individuals, are observed for J_i times, that could be days, months, years, and so on.

6 Exemples

Now we present some examples using different datasets, implementing the procedure for estimates standard deviation components and the percentage shares, using both `xtsum2` and the existing commands in Stata and compare these showing that the new command is simpler and avoid intricate calculation.

6.1 Interest rate and inflaction rate in Euro-area

The dataset consists of the surveys of nominal interest rate and inflation rate, for 11 countries in the euro area from 1980 to 1998, so 19 periods and a total of 209 observations. This dataset has been used, to explain the nominal interest rate using the inflation rate, so the first one is the dependent variable on which we analyse variance components. The dataset is a macro type one because is composed of time series for few individuals, the countries. It is strongly balanced because all the time series are complete, no observation is missing. In this case, we have to handle panel data, so investigate where the variation is concentrated, countries, years, or both. It will be very helpful to choose the right model to explain the interest rate and to capture a consistent and unbiased effect of the inflation rate on the dependent variable and moreover. We start with the new command to get a clear panoramic vision on which dimension of the panel has concentrated the variance.

The classes in the dataset are finite and not random, the countries are the ones that use Euro as currency, so are proper the classes that we should analyse. As said in the preview sections, in these situations is better to use a fixed effect model to estimate standard deviations and percentage shares, instead of random effect one. We exclude the mixed model too because for panel data the more suitable models are the first two. The syntax, in this case, is the simplest, because the fixed effect model is the default option, perhaps we specify `fe`, or not, the result doesn't change.

```
. xtset codice anno
. xtsum2 tint
209
-------------------------------- index i: codice   index t: anno           ----------------------

Variable                       |   Mean   Std.Dev.  % SS |  Observations  |  Test
-------------------------------+-------------------------+----------------+-----------------------
tint     overall x..           | .0945924                | NT          209| codice effects:
            overall xit-x..    |          .03436         | N            11|  F(10, 180)    =  41.89
        (1)  between xi.-x..   |          .02169   36.41 | Nmin         11|      Prob > F =   0.00
        (2a) between x.t-x..   |          .02439   47.95 | N-bar        11|
        (2b) resid xit-xi.-x.t+x..|       .01461   15.64 | Nmax         11| anno effects:
        ------ (2a) + (2b): ------|       ----------------| T            19|  F(18, 180)    =  30.65
        (2)  within  xit-xi.   |          .02808   63.59 | Tmin         19|      Prob > F =   0.00
         (% (2a) between x.t-x..) |               (75.40)| T-bar        19|
                               |                         | Tmax         19| balancedness  = 100.00
--------------------------------------------------------------------------------------------------
```

such we can see the output report both the way of decomposition of variance, in the higher part we have between countries(1), between years(2a) and residual components(2b) and in the bottom part the within one(2).

If we look at the column of the percentage, we have suddenly a lot of information about nominal interest rate, the variation from one year to another is higher than the variation from one country to another and the variation within too. This last one is the variation within a country, so over the years, so we understand that the time dimension is very important in this dataset, omit it in a hypothetical model could cause less efficiency or worse biased estimation. The residual percentage share means that the dependent variable is not fully explained by the deviation of the country means and year means from the overall mean, remain 15,64% of unobserved heterogeneity for both intra-individual and intra-temporal. The intra-individual-temporal variation represents the variation at the country level, so across time, depurate from the variation of the characteristics due to the time course, like the economic cycle. At the same time represent the differences between every country in a specific year from the mean of the specific year, also that depurate from characteristic attributable at the heterogeneity of the countries.

After that, we continue giving a panoramic of the methods for decomposing variance existing in Stata and showing how intricate is the statement to arrive at the percentage share see above.

There is more than one method for estimate components of variance in Stata, the most common are `xtsum`, `anova`, `loneway`, `xtreg`, `mixed` (for see in details anova and mixed see Marchenko, 2006).

The command `xtsum` returns the variance components, in form of standard errors, it uses the first way of decomposition of variance producing the between individual standard deviation and the within individual standard deviation.

```
. xtsum tint

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
tint     overall |  .0945924    .0343588   .0400857   .1947852 |     N =     209
         between |              .0216913   .0687109   .1339803 |     n =      11
         within  |              .0273994   .0006978   .1673036 |     T =      19
```

The command `xtsum` needs that the data are set in their two dimensions by using `xtset`, in the output above, since the dataset is a panel one, at the first level there are the observations of the interest rate at differents year per country and at the second one, the countries. therefore we can calculate the between temporal and within temporal standard deviation too by setting the dataset exchanging the individual dimension with the temporal dimension and then executing `xtsum`. We can find similar estimations by fit the multilevel model or the random effect model using the variance component model for both. Because the random part of these models calculates just these components, in form of standard deviations. In the statement of `mixed`, we have to specify that we want standard deviations because it calculates variance by default.

```
. mixed tint || codice: , stddev
Mixed-effects ML regression                     Number of obs     =        209
Group variable: codice                          Number of groups  =         11
                                                Obs per group:
                                                              min =         19
                                                              avg =       19.0
                                                              max =         19
                                                Wald chi2(0)      =          .
Log likelihood =  437.28532                     Prob > chi2       =          .
------------------------------------------------------------------------------
        tint |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0945924   .0062358    15.17   0.000     .0823704    .1068144
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
codice: Identity             |
                   sd(_cons) |   .0196528   .0046415      .0123706    .0312217
-----------------------------+------------------------------------------------
                sd(Residual) |   .0280828   .0014112      .0254487    .0309895
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 57.65         Prob >= chibar2 = 0.0000


. xtreg tint, re
Random-effects GLS regression                   Number of obs     =        209
Group variable: codice                          Number of groups  =         11
R-sq:                                           Obs per group:
     within  = 0.0000                                         min =         19
     between = 0.0000                                         avg =       19.0
     overall = 0.0000                                         max =         19
                                                Wald chi2(0)      =          .
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =          .
------------------------------------------------------------------------------
        tint |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0945924   .0065402    14.46   0.000     .0817738    .1074109
-------------+----------------------------------------------------------------
     sigma_u |  .02071247
     sigma_e |   .0280828
         rho |  .35232331   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The standard deviation within individuals is equal to 0.0280828 in the whole methods saw up until now. This component concerns the observations level and the differents method estimation of it differ in the way of calculation of how the observations are nested so are different in the component between. The standard deviation between individuals could be slightly different from each other, because of the different calculation methods. The command `mixed` estimates a multilevel model, using the maximum likelihood approach and `xtreg` estimates a random effect one that uses the weighted latest squares methods. Despite this, all of them are consistent and unbiased. The random effect models, as seen in section 4, after produce the θ-transformation could be estimated using the maximum likelihood approach as for mixed model, rather than ordinary latest squares.

```
. xtreg tint, re mle
Random-effects ML regression                    Number of obs     =        209
Group variable: codice                          Number of groups  =         11
Random effects u_i ~ Gaussian                   Obs per group:
                                                              min =         19
                                                              avg =       19.0
                                                              max =         19
                                                Wald chi2(0)      =       0.00
Log likelihood  =  437.28532                    Prob > chi2       =          .
------------------------------------------------------------------------------
        tint |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0945924   .0062358    15.17   0.000     .0823704    .1068144
-------------+----------------------------------------------------------------
    /sigma_u |   .0196528   .0046415                      .0123707    .0312217
    /sigma_e |   .0280828   .0014111                      .0254489    .0309892
         rho |   .3287432   .1070614                      .1529894    .5544545
------------------------------------------------------------------------------
LR test of sigma_u=0: chibar2(01) = 57.65              Prob >= chibar2 = 0.000
```

The results, of course, are identical to the ones finds out using `mixed`, because using both the method of maximum likelihood approach.

Comparing the latter result of the between countries standard deviation, with the one estimated with the multilevel model, we can see that are completely identical. It may not surprise us because both the commands use the maximum likelihood to fit these standard deviation components.

As seen in section 4, another approach to estimate the random coefficients of mixed models is the restricted maximum likelihood rather than full maximum likelihood. To do this in Stata it is necessary to add to the syntax used before for, the specific option, `reml`.

In this way, we can see if the restrictions on the parameters using this second method induce differences in the estimation of variance components.

```
. mixed tint || codice: , reml stddev
Mixed-effects REML regression                   Number of obs     =        209
Group variable: codice                          Number of groups  =         11
                                                Obs per group:
                                                              min =         19
                                                              avg =       19.0
                                                              max =         19
                                                Wald chi2(0)      =          .
Log restricted-likelihood =  433.15026          Prob > chi2       =          .
------------------------------------------------------------------------------
        tint |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0945924   .0065402    14.46   0.000     .0817738    .1074109
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
codice: Identity             |
                   sd(_cons) |   .0207125   .0050806      .0128068    .0334983
-----------------------------+------------------------------------------------
                sd(Residual) |   .0280828   .0014112      .0254487    .0309895
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 59.63         Prob >= chibar2 = 0.0000
```

We know that the restricted maximum likelihood is better than the full one, because, as seen in section 4, the latter loses degrees of freedom because treat the coefficient as fixed, losing efficiency. Moreover, there is the possibility that in the small samples they might be even biased, our sample is neither so small but nor so big, N is equal to 209. Therefore is better trust in the result of RML. In this case, the result of the between components using the FML might be overestimated. As we can see the within component doesn't change once again, remains the same for the whole methods using so far.

Another thing to notice is that the between standard deviation, calculated using the residual maximum likelihood is precisely equal to the one calculated with the first random effect model. This because the random effect model as the RML doesn't estimate the coefficient as fixed and both apply to the data a quasi-demand transformation, fitting not the simple observations but the deviation of the observation from its partial mean. The two transformations, however, are not equal because in the random effect model the partial mean is multiplied for the θ parameter that measures the intraclass correlation. The multilevel model with the residual maximum likelihood Another model to estimate within and between standard deviations is the fixed effect model, which is the same method use into the new command `xtsum2`, infact such we can see below the results are identicals.

```
. xtreg tint, fe
Fixed-effects (within) regression               Number of obs      =       209
Group variable: codice                          Number of groups   =        11
R-sq:                                           Obs per group:
     within  =        .                                        min =        19
     between =        .                                        avg =      19.0
     overall =        .                                        max =        19
                                                F(0,198)           =      0.00
corr(u_i, Xb)  =        .                       Prob > F           =         .
------------------------------------------------------------------------------
        tint |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0945924   .0019425    48.70   0.000     .0907617    .0984231
-------------+----------------------------------------------------------------
     sigma_u |  .02169134
     sigma_e |   .0280828
         rho |  .37367367   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(10, 198) = 11.34                    Prob > F = 0.0000
```

The latter one command to estimate these components of variance is loneway.

```
. loneway tint codice
                      One-way Analysis of Variance for tint:
                                                Number of obs =          209
                                                    R-squared =       0.3641
    Source                  SS         df      MS            F     Prob > F
-------------------------------------------------------------------------
Between codice          .08939767     10    .00893977     11.34     0.0000
Within codice           .15615142    198    .00078864
-------------------------------------------------------------------------
Total                   .24554909    208    .00118052
         Intraclass       Asy.
         correlation      S.E.       [95% Conf. Interval]
         ------------------------------------------------
            0.35232     0.11472       0.12748     0.57716
         Estimated SD of codice effect           .0207125
         Estimated SD within codice              .0280828
         Est. reliability of a codice mean        0.91178
              (evaluated at n=19.00)
```

Now we can find out for all the procedures above, the relative percentage share. It is necessary to do some calculations using `display`, it works like a calculator, we state it to do the ratio between the variance components (in the first row of the output below the between one and in the second the within one) and the sum of them.

xtsum

```
. xtset codice anno
. xtsum tint
. display r(sd_b)^2/(r(sd_w)^2+r(sd_b)^2)
.38527502
. display r(sd_w)^2/(r(sd_w)^2+r(sd_b)^2)
.61472498
```

mixed (FML)

```
. mixed tint || codice:
. display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
> + exp(2*[lnsig_e]_cons))
.32874348
. display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
> + exp(2*[lnsig_e]_cons) )
.67125652
```

random effect model

```
. xtreg tint, re
. display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
.35232331
. display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
.64767669
```

random effect model using mle approach

```
. xtreg tint, re mle
. display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
.32874318
. display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
.67125682
```

mixed (RML)

```
. mixed tint || codice:, reml
. display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)  +
exp(2*[lnsig_e]_cons))
.35232362
. display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)  +
exp(2*[lnsig_e]_cons))
.64767638
```

fixed effect model

```
. xtreg tint, fe
. display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
.37367367
. display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
.62632633
```

loneway

```
. loneway tint codice
. display r(sd_b)^2/(r(sd_b)^2+r(sd_w)^2)
.35232331
. display r(sd_w)^2/(r(sd_b)^2+r(sd_w)^2)
.64767669
```

We have to make a clarification on these percentage shares estimated. As shown in section 3 the equations (14) and (15) are different from the estimations done before. They are obtained making the ratio between the sum of squares, not the variances, because based on identity (3) we calculate a composition ratio between the total sum of squares and the specific partial sum of squares. Whereby the commands used before are just an approximation of the real percentage share however they give us a good direction of the between and within percentage shares. If we want to be accurate to arrive exactly at the same formulas, more countless steps are required. We need to estimate the partial sums of squares from the partial standard deviations manually, because all the models used above except for Anova, don't compute these magnitudes or don't save that in the stored results, in any way, they are not available and we have to do that ourselves. Below we present the statements for doing that using all the methods available in Stata.

xtsum

```
. xtset codice anno
. xtsum tint
. local wi_ss=r(sd_w)^2*(r(N)-r(n))
. local bi_ss=r(sd_b)^2*(r(N)-r(Tbar))
. display `wi_ss'/(`wi_ss'+`bi_ss')
.62444552
. display `bi_ss'/(`wi_ss'+`bi_ss')
.37555448
```

mixed (FML)

```
. xtdescribe
. local N=r(sum_w)
. local J=r(mean)
. mixed tint || codice: , stddev
. local wi_ss=(exp([lnsig_e]_cons))^2*(e(N)-`n')
. local bi_ss=(exp([lns1_1_1]_cons))^2*(e(N)-`t')
. display `wi_ss'/(`wi_ss'+`bi_ss')
.68029255
. display `bi_ss'/(`wi_ss'+`bi_ss')
.31970745
```

random effect model

```
. xtreg tint, re
. local wi_ss=e(sigma_e)^2*(e(N)-e(N_g))
. local bi_ss=e(sigma_u)^2*(e(N)-e(Tbar))
. display `wi_ss'/(`wi_ss'+`bi_ss')
.6570297
. display `bi_ss'/(`wi_ss'+`bi_ss')
.3429703
```

random effect model using mle approach

```
. xtreg tint, re mle
. local wi_ss=e(sigma_e)^2*(e(N)-e(N_g))
. local bi_ss=e(sigma_u)^2*(e(N)-e(g_avg))
. display `wi_ss'/(`wi_ss'+`bi_ss')
.68029284
. display `bi_ss'/(`wi_ss'+`bi_ss')
.31970716
```

mixed (RML)

```
. xtdescribe
. local N=r(sum_w)
. local J=r(mean)
. mixed tint || codice:, reml
. local wi_ss=(exp([lnsig_e]_cons))^2*(e(N)-`n')
. local bi_ss=(exp([lns1_1_1]_cons))^2*(e(N)-`t')
. display `wi_ss'/(`wi_ss'+`bi_ss')
.65702939
. display `bi_ss'/(`wi_ss'+`bi_ss')
.34297061
```

fixed effect model

```
. xtreg tint, fe
. local wi_ss=e(sigma_e)^2*(e(N)-e(g_avg))
. local bi_ss=e(sigma_u)^2*(e(N)-e(N_g))
. display `wi_ss'/(`wi_ss'+`bi_ss')
.61662458
. display `bi_ss'/(`wi_ss'+`bi_ss')
.38337542
```

loneway

```
. xtdescribe
. local N=r(sum_w)
. local J=r(mean)
. loneway tint codice
. local wi_ss=r(sd_w)^2*(r(N)-`n')
. local bi_ss=r(sd_b)^2*(r(N)-`t')
. display `wi_ss'/(`wi_ss'+`bi_ss')
.6570297
. display `bi_ss'/(`wi_ss'+`bi_ss')
.3429703
```

This information is very useful for analyse variance and then for choosing the best model to explain the data but don't care about the decomposition among the first level, that would be useful when we handle a panel dataset, like in this case. Anyway could be useful in some particular kind of hierarchical dataset, in which the observations of the first level are not randomly nested within the unit of the second level. We can calculate the between years and within years standard deviation, as said before, by setting the dataset exchanging the dimension, and then run one of the commands shown above, but the intra-individual temporal component would remain unknown.

There is another way to find the between temporal component, that allows us to find the residual component too. We need to use `anova` or a `mixed` version in which we specify both the dimensions, the countries, and the years, where the outcome becomes the second way of decomposition of variance, so it estimates the component we lack, the residual standard deviation, in addition to the between temporal.

We start estimating mixed models, first using the maximum likelihood approach and then the restricted maximum likelihood one.

```
. mixed tint ||_all:R.codice || anno:, stddev
Mixed-effects ML regression                     Number of obs     =        209
-------------------------------------------------------------
                |     No. of       Observations per Group
 Group Variable |     Groups    Minimum    Average    Maximum
----------------+--------------------------------------------
           _all |          1        209      209.0        209
           anno |         19         11       11.0         11
-------------------------------------------------------------
                                                Wald chi2(0)      =          .
Log likelihood =  535.60322                     Prob > chi2       =          .
------------------------------------------------------------------------------
        tint |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0945924   .0083649    11.31   0.000     .0781975    .1109872
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
                sd(R.codice) |   .0208225   .0045925      .0135143    .0320827
-----------------------------+------------------------------------------------
anno: Identity               |
                   sd(_cons) |   .0236886   .0040102      .0169997    .0330092
-----------------------------+------------------------------------------------
                sd(Residual) |   .0146092     .00077      .0131753    .0161991
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 254.28                Prob > chi2 = 0.0000
```

```
. mixed tint ||_all:R.codice || anno:, reml stddev
Mixed-effects REML regression                    Number of obs     =       209
-------------------------------------------------------------
                |     No. of       Observations per Group
 Group Variable |     Groups    Minimum    Average    Maximum
----------------+--------------------------------------------
           _all |          1        209      209.0        209
           anno |         19         11       11.0         11
-------------------------------------------------------------
                                                 Wald chi2(0)      =         .
Log restricted-likelihood =   531.7491           Prob > chi2       =         .
------------------------------------------------------------------------------
        tint |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0945924   .0085469    11.07   0.000     .0778407     .111344
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
                sd(R.codice) |   .0214309   .0049094      .0136788    .0335763
-----------------------------+------------------------------------------------
anno: Identity               |
                   sd(_cons) |    .023984   .0041324      .0171105    .0336188
-----------------------------+------------------------------------------------
                sd(Residual) |   .0146086   .0007699      .0131749    .0161983
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 256.83                Prob > chi2 = 0.0000
```

Such see above, in this output, we get three random-effect parameters, the first one is the between individual standard deviation, the second one is the between temporal standard deviation, which still had not been found in the method used above. Actually, that can be estimated using `xtsum`, `mixed` or `xtreg`, however, we have to set the data, before using, so that the years result like individuals and country look like the temporal periods. In this way, we obtain the standard deviation within temporal.

The latter is the residual standard deviation, which is very important because captures both the unobserved temporal and individual heterogeneity.

As for the component between countries and within countries saw before, we can calculate the percentage shares for the between temporal component and for the residual one as well. We have to run the statements below, which computes first the partial sum of squares and base on it the percentage shares.

mixed (FML)

```
. display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)  +
exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
.35887527
. display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)  +
exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
.46446797
. display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)  + exp(2*[lnsig_e]_cons)
+ exp(2*[lns2_1_1]_cons))
.17665676
```

mixed (RML)

```
. display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)  +
exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
.36803645
. display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)  +
exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
.46095127
. display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)  + exp(2*[lnsig_e]_cons)
+ exp(2*[lns2_1_1]_cons))
.17101228
```

Finally the statements for the percentage share using the ratio between the sums of square instead of the variances

mixed (FML)

```
. xtdescribe
. local N=r(sum_w)
. local J=r(mean)
. mixed tint || codice: , stddev
. local wi_ss=(exp([lnsig_e]_cons))^2*(e(N)-`n')
. local bi_ss=(exp([lns1_1_1]_cons))^2*(e(N)-`t')
. display `wi_ss'/(`wi_ss'+`bi_ss')
.68029255
. display `bi_ss'/(`wi_ss'+`bi_ss')
.31970745
```

mixed (RML)

```
. local bi_ss_mixed3_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`n')
. local bj_ss_mixed3_reml=(exp([lns2_1_1]_cons))^2*(e(N)-`t')
. local r_ss_mixed3_reml=(exp([lnsig_e]_cons))^2*(e(N)-`t')
. display`r_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'+
`bj_ss_mixed3_reml')
.16840266
. display `bi_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'+
`bj_ss_mixed3_reml')
.37768009
. display `bj_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'+
`bj_ss_mixed3_reml')
.45391725
```

Now we see the latter `anova`, this command estimates the components of variance see above for the second version of `mixed` but in the basic form of partial sums of squares and means of square

```
. anova tint codice anno
                       Number of obs =       209    R-squared     =  0.8436
                       Root MSE      =   .014609    Adj R-squared =  0.8192
                Source | Partial SS        df        MS          F     Prob>F
            -----------+----------------------------------------------------
                 Model |  .20713514        28   .00739768      34.66  0.0000
                codice |  .08939767        10   .00893977      41.89  0.0000
                  anno |  .11773747        18   .00654097      30.65  0.0000
              Residual |  .03841395       180   .00021341
            -----------+----------------------------------------------------
                 Total |  .24554909       208   .00118052
```

We have to calculate the percentage shares and the standard deviations, that these last ones don't be calculated by `anova`. Starting from the partial sum of squares we can find out all these components.

We can obtain the percentage directly from the sum of squares, that as seen before is the more correct way from an algebric point of view

```
. xtdescribe
. local N=r(sum_w)
. local J=r(mean)
. display ((e(ss_2)/e(df_2))/`n')^0.5
.02438511
. display ((e(ss_1)/e(df_1))/`t')^0.5
.02169134
. display (e(rss)/e(df_r))^0.5
.01460859
. display  e(ss_1)/(e(ss_1)+e(ss_2)+e(rss))*100
36.407249
. display e(ss_2)/(e(ss_1)+e(ss_2)+e(rss))*100
47.948648
. display e(rss)/(e(ss_1)+e(ss_2)+e(rss))*100
15.644104
```

We saw all the methods available in Stata to decompose variance comparing with `xtsum2`. To get a global vision about every value that has been calculated before, we leave to the reader into the first subsection of the appendix, all the Stata statements using in this paper for each example.

6.2 Wagepan

This dataset is a panel one, from the National Longitudinal Survey (Youth Sample), in which are observed the wage of workers, in form of a logarithm, and other individual characteristics of them. The classes, in this case, are represented by the employees that are observed for several years. The dataset is strongly balanced, no observation is missing, in the sample, there are 545 workers per 8 temporal periods, the years from 1980 to 1987. The class level is hugely bigger than the temporal one, for this reason, this kind is usually named, in panel data anlysis, micro type. As opposite a dataset in which are gathered low observations for much temporal period is usually named macro.

This dataset has been used in Vella, Verbek (1998), the aim of the analysis is to estimate the union premium for young workers in the period considered.

The workers are gathered randomly from the population of workers, and we are not interested in this specific sample but we should know something about the population from which come the sample. In this situation, it is better to use the random effect model to estimate the components of variance. We start again with the `xtsum2` using the specif option for this kind of variance component model, `re`.

```
. xtset nr year
       panel variable:  nr (strongly balanced)
        time variable:  year, 1980 to 1987
                delta:  1 unit

. xtsum2 lwage, re
--------------------------------- Level 2 (i): nr   Level 1 (j): year       -----------------------
                                  Hypothetical full sample, Ni x Nj: 4360
                                  Ni: 545  |   Nj: 8
---------------------------------------------------------------------------------------------------
Variable                          |   Mean   Std.Dev.  % SS |  Observations  |  Test
----------------------------------+-------------------------+----------------+-----------------------
lwage       overall xij-x..       |  1.649    .5326         | NiNj       4360| nr effects:
       (1) between xi.-x..        |           .3907   53.74 | Ni          545|  Chi2(1)       =3389.35
       (2a) between x.j-x..       |           .1546    7.37 | Ni-bar   545.00|   Prob > Chi2 =    0.00
       (2b) res xij-xi.-x.j+x..   |           .3554   38.89 | Ni-max      545| year effects:
       ------ (2a) + (2b): ------|         ---------------| Nj            8|  Chi2(1)       =6403.78
       (2)  within xij-xi.        |           .3872   46.26 | Nj-bar        8|   Prob > Chi2 =    0.00
        (% (2a) between x.j-x..) |                  (15.93)| Nj-max     8.00| balancedness  =   100.0
---------------------------------------------------------------------------------------------------
```

Such we can see, the between individual variation is slightly bigger than within individual one, in percentage terms, so the differences the salary from one worker and another is slightly bigger than the differences among each worker across the time. The between temporal variation is very low, only 7.37% on the total, meaning that the differences between one year and another are very low, almost negligible. If we make box plots among the years we can see better that.

```
graph box lwage, over(year) yline(1.649147)
```

This graph shows also that there was a rise up, in the period considered, of the wage earned by the workers, it could explain the differences between a year and another.

The temporal dimension, however, is not completely negligible because the within individual variation is a variation among time in substance and is 46,26% of the total. Furthermore, the F statistic on the time dummies rejects the null hypothesis that there are no differences between a period and another. The residual percentage share is 38.89%, it represents both heterogeneities individual and temporal.

It captures by one side the differences within worker depurate from the one attributable at the period, like the economic cycle, so remain only the characteristics proper of the workers. On the other side, it captures the differences among workers in each specific period depurate from the characteristic proper of the worker.

To conclude, the individual dimension keeps more information than the temporal one so if we are going to estimate a model we have to give more importance to the country dimension and less to the temporal one, but is not careful ignore it.

We don't discuss the comparison between `xtsum2` and the other commands available in Stata, for this example and those that will follow, it could be redundant. In any way, in the appendix, there is all the information as well as example 1. In that, we find all the statements to arrive at the same result using all the commands available in Stata to estimate the components of variance of our interest.

6.3 Abdata

This dataset is a panel one, represent a survey on companies, used to estimate a labor demand model by Blundell, Bond (1998).

In this panel the temporal dimension is represented by the year from 1976 to 1984, for a total of 9 temporal periods, instead, the individual dimension is represented by the companies in the UK, for a total of 140. Furthermore, the companies are cluster within the industrial sector, so there are three levels, the higher is that one, then there are the companies one, and the lower one is the temporal observation of each company.

The dataset is a micro type, such as the dataset in example 2, which contains much more companies than years.

The variable on which we will do the decomposition of variance is the logarithm of the number of employees in a firm that through a dynamic model is explained by its lags and the characteristics of the firm.

This example is different from the other ones because of its unbalancedness, not every company has been registered for all the years in the period considered, we'll see after run xtsum2 in its third section, the missing observations, and their implications.

In this case as the example 6.2, the classes, which here are represented by firms, are gathered from a population of firm, which we are interested in, so the right option of `xtsum2` it's `re`, as follow,

```
. xtset id year
       panel variable:  id (unbalanced)
        time variable:  year, 1976 to 1984
                delta:  1 unit

. xtsum30calcoli6 n, re
-------------------------------- Level 2 (i): id   Level 1 (j): year       ----------------------
                                 Hypothetical full sample, Ni x Nj: 1260
                                 Ni: 140   Nj: 9
-----------------------------------------------------------------------------------------------
Variable                        |   Mean   Std.Dev.  % SS |  Observations   |  Test
--------------------------------+-------------------------+-----------------+-----------------------
n           overall xij-x..     |  1.056    1.342         | NiNj        1031| id effects:
       (I) between xi.-x..      |           1.294   92.49 | Ni           140|  Chi2(1)       =3031.22
       (IIa) between x.j-x..    |           .3508    6.09 | Ni-bar    114.56|   Prob > Chi2 =    0.00
       (IIb) res xij-xi.-x.j+x..|           .1731    1.43 | Ni-max       140| year effects:
       ----- (IIa) + (IIb): -----|          ----------------| Nj             9|  Chi2(8)       = 415.66
       (II)  within xij-xi.     |           .3954    7.51 | Nj-bar         7|   Prob > Chi2 =    0.00
        (% (IIa) between x.j-x..)|                 (80.99)| Nj-max      9.00|| balancedness  =  81.83
-----------------------------------------------------------------------------------------------
```

The between companies variation is the higher one on the total 92.49%, it seems that almost all the variance is concentrated in the differences between the mean of each company and the overall mean.

The principal reason why is that in the sample there are firms very different from each other. If we look at the number of employees there are companies with more than 100 compared with others that have less than 1, such mean of the year. If we take a look at the within component, we can see that is very low, it means that the percentage increase of the employees in the same firm between one year and another is very small.

In the last section, we can see a measurement of the unbalancedness, there are 81.83% of the observations on the total that we would like to consider. To see well where the information is missing we have to run the Stata command `xtdescribe`

```
. xtdescribe
      id:  1, 2, ..., 140                                    n =        140
    year:  1976, 1977, ..., 1984                             T =          9
           Delta(year) = 1 unit
           Span(year)  = 9 periods
           (id*year uniquely identifies each observation)
Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         7       7       7         7         8       9       9
     Freq.  Percent    Cum. |  Pattern
 ---------------------------+-----------
       62     44.29   44.29 |  1111111..
       39     27.86   72.14 |  .1111111.
       19     13.57   85.71 |  .11111111
       14     10.00   95.71 |  111111111
        4      2.86   98.57 |  11111111.
        2      1.43  100.00 |  ..1111111
 ---------------------------+-----------
      140    100.00         |  XXXXXXXXX
```

Such we can see, the observation is missing in the beginning or/and in the ending of the period, no one in the middle or any jump. Nevertheless, only 10% of the company are observed for whole 9 periods. More companies have been interviewed for only 7 periods than the others interviewed in 8 or 9 periods, instead, the median period is 7, so at least half of the firms have been interviewed for 7 periods.

6.4 TCSC

In this example, we use TSCS data from Milner and Kubota (2005). The dataset is composed of observations of 101 countries over the time, from 1980 to 1999, so the first level is represented by the countries one and the second one by the years

Milner and Kubota in their paper, showing how more democratic countries are more open on trade liberalism. The dependent variable, of course, is the country's statutory rate and the explanatory ones are the policy score, which measures the degree of democracy of the state, the gdp per capita, and the natural logarithm of the population.

The dataset is unbalancedness but different from example 6.3 because the variables were not detected for all the countries and all the years alike. There may be a variable observed for the country i and the year t and another one that instead for the same i and same t is missing.

As in example 6.1 the sample is composed of some countries, in this case, more, observed for several years. The reasoning for the choice of the model is similar to the first dataset, we are interested proper in the classes that there are in the dataset, the countries Therefore in this case is better to use a fixed effect model, specifying the option `re` in the syntax of `xtsum2`.

In this case is useful that the analysis of variance is doing on a restricted sample, that remains only the i and t in which all the variables are present. When we estimate a model, whatever it is, we use only the observations common to each variable, the independent one, and the set of explanatory that we decide to insert. If we make decisions on a wrong analysis of variance it could take us to a wrong model to explain our data, even worse if the object of the analysis about the behavior of a variable related to a categorical variable. What we should do is to restrict the model in that way. `xtsum2` simplifies this problem with the option `common`, it rejects the observations that are not observed for all the variables.

This option is very useful, otherwise, we have to manually create a subset in which we keep only the observations observed for all the variables.

An example for doing that is to create a dummy variable that gets this feature, so we can state to get value 1 when all the variables are different from missing value and then replace the missing values with 0.

After creates the dummy variable we have to use one of the methods available in Stata, adding an `if` expression that rejects the observations in which dummy is equal to zero. Below the log for doing that using `xtsum` as a method, of course, we can implement the same procedure using the others, we can find those statements in the appendix.

```
. gen dummy=1 if polity!=. & gdppc!=. & tariff!=. & lnpop!=.
(4,596 missing values generated)
. replace dummy=0 if dummy==.
(4,596 real changes made)
. xtset country date
       panel variable:  country (strongly balanced)
        time variable:  date, 1970 to 1999
                delta:  1 unit
. xtsum tariff polity lnpop gdppc if dummy==1
Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
tariff   overall |  21.68411   15.04545         .1      102.2 |     N =     774
         between |             11.17842       .375   63.45625 |     n =     101
         within  |             8.599189  -16.74089   63.25911 | T-bar = 7.66337
                 |                                            |
polity   overall |   1.18863   7.148984        -10         10 |     N =     774
         between |             6.454548        -10         10 |     n =     101
         within  |              3.59562   -13.3447   12.18863 | T-bar = 7.66337
                 |                                            |
lnpop    overall |  16.47991    1.53594    12.7189   20.93034 |     N =     774
         between |             1.519106   13.09863   20.85007 |     n =     101
         within  |             .1133827   15.99519    16.9258 | T-bar = 7.66337
                 |                                            |
gdppc    overall |  2642.515   3694.001     107.06   37841.16 |     N =     774
         between |             3983.491     107.06   27032.56 |     n =     101
         within  |             930.4761  -6104.327   13451.12 | T-bar = 7.66337
```

Now we present the two outputs of `xtsum2`, the first with and the second without using the `common` option, comparison the results and understanding if a change has occurred.

```
.xtset country date
.xtsum2 tariff policy lnpop gdppc
Variable                         |   Mean   Std.Dev.  % SS |  Observations  |  Test
---------------------------------+-------------------------+----------------+-----------------------
tariff   overall x..             | 20.53826                | NT          907| country effects:
            overall xit-x..      |          15.06          | N           125|  F(124, 753)  =  17.60
       (1)  between xi.-x..      |           12.5   68.39 | Nmin          1|      Prob > F =   0.00
       (2a) between x.t-x..      |          4.841   10.00 | N-bar        30|
       (2b) resid xit-xi.-x.t+x..|           7.68   21.61 | Nmax         72| date effects:
       ------ (2a) + (2b): ------|        ----------------| T            30|  F(29, 753)   =  12.01
       (2)  within  xit-xi.      |          9.115   31.61 | Tmin          1|      Prob > F =   0.00
        (% (2a) between x.t-x..) |                 (31.63)| T-bar       7.3|
                                 |                         | Tmax         30| balancedness  =   3.12
------------------------+-----------------------------------------------------------------------------
polity   overall x..             |-2.215342                | NT         3246| country effects:
            overall xit-x..      |          6.927          | N           133|  F(132, 3085) =  59.27
       (1)  between xi.-x..      |          5.655   66.18 | Nmin        102|      Prob > F =   0.00
       (2a) between x.t-x..      |          2.087    8.77 | N-bar       112|
       (2b) resid xit-xi.-x.t+x..|          3.556   25.05 | Nmax        125| date effects:
       ------ (2a) + (2b): ------|        ----------------| T            29|  F(28, 3085)  =  38.55
       (2)  within  xit-xi.      |          4.113   33.82 | Tmin          5|      Prob > F =   0.00
        (% (2a) between x.t-x..) |                 (25.92)| T-bar        24|
                                 |                         | Tmax         29| balancedness  =  11.18


------------------------------------------------------------------------------------------------------
lnpop    overall x..             | 15.10395                | NT         4709| country effects:
            overall xit-x..      |          1.997          | N           164|  F(163, 4517) =9741.65
       (1)  between xi.-x..      |          1.993   98.93 | Nmin        161|      Prob > F =   0.00
       (2a) between x.t-x..      |          .1806    0.79 | N-bar       162|
       (2b) resid xit-xi.-x.t+x..|          .1083    0.28 | Nmax        164| date effects:
       ------ (2a) + (2b): ------|        ----------------| T            29|  F(28, 4517)  = 451.71
       (2)  within  xit-xi.      |          .2104    1.07 | Tmin          9|      Prob > F =   0.00
        (% (2a) between x.t-x..) |                 (73.68)| T-bar        29|
                                 |                         | Tmax         29| balancedness  =  16.22
------------------------------------------------------------------------------------------------------
gdppc    overall x..             | 2887.762                | NT         3547| country effects:
            overall xit-x..      |           4657          | N           152|  F(151, 3367) = 212.68
       (1)  between xi.-x..      |           4438   90.26 | Nmin         90|      Prob > F =   0.00
       (2a) between x.t-x..      |          245.1    0.27 | N-bar       122|
       (2b) resid xit-xi.-x.t+x..|           1471    9.47 | Nmax        152| date effects:
       ------ (2a) + (2b): ------|        ----------------| T            29|  F(28, 3367)  =   3.40
       (2)  within  xit-xi.      |           1485    9.74 | Tmin          4|      Prob > F =   0.00
        (% (2a) between x.t-x..) |                  (2.75) | T-bar        23|
                                 |                         | Tmax         29| balancedness  =  12.21
------------------------------------------------------------------------------------------------------
```

```
.xtset country date
.xtsum2 tariff policy lnpop gdppc, common
-------------------------------- index i: country   index t: date          ----------------------|
Variable                        |  Mean   Std.Dev.  % SS |  Observations  |  Test
--------------------------------+------------------------+----------------+----------------------
tariff   overall x..            | 21.68411               | NT         774| country effects:
            overall xit-x..     |          15.05         | N          101|  F(100, 654)  =  19.42
       (1)  between xi.-x..     |           12.4   67.33 | Nmin        12|      Prob > F =   0.00
       (2a) between x.t-x..     |          5.194   11.34 | N-bar       39|
       (2b) resid xit-xi.-x.t+x..|         7.555   21.33 | Nmax        62| date effects:
       ------ (2a) + (2b): ------|        ----------------| T           20|  F(19, 654)    =  18.29
       (2)  within  xit-xi.     |          9.216   32.67 | Tmin         1|      Prob > F =   0.00
        (% (2a) between x.t-x..) |                (34.70)| T-bar      7.7|
                                |                        | Tmax        19| balancedness  =   2.67
-------------------------------------------------------------------------------------------------
polity   overall x..            |  1.18863               | NT         774| country effects:
            overall xit-x..     |          7.149         | N          101|  F(100, 654)  =  23.08
       (1)  between xi.-x..     |          6.206   74.70 | Nmin        12|      Prob > F =   0.00
       (2a) between x.t-x..     |          1.966    7.20 | N-bar       39|
       (2b) resid xit-xi.-x.t+x..|         3.307   18.10 | Nmax        62| date effects:
       ------ (2a) + (2b): ------|        ----------------| T           20|  F(19, 654)    =  13.69
       (2)  within  xit-xi.     |          3.854   25.30 | Tmin         1|      Prob > F =   0.00
        (% (2a) between x.t-x..) |                (28.45)| T-bar      7.7|
                                |                        | Tmax        19| balancedness  =   2.67

-------------------------------------------------------------------------------------------------
lnpop    overall x..            | 16.47991               | NT         774| country effects:
            overall xit-x..     |          1.536         | N          101|  F(100, 654)  =8108.35
       (1)  between xi.-x..     |          1.538   99.46 | Nmin        12|      Prob > F =   0.00
       (2a) between x.t-x..     |          .1076    0.47 | N-bar       39|
       (2b) resid xit-xi.-x.t+x..|        .04676    0.08 | Nmax        62| date effects:
       ------ (2a) + (2b): ------|        ----------------| T           20|  F(19, 654)    = 204.82
       (2)  within  xit-xi.     |          .1215    0.54 | Tmin         1|      Prob > F =   0.00
        (% (2a) between x.t-x..) |                (85.61)| T-bar      7.7|
                                |                        | Tmax        19| balancedness  =   2.67
-------------------------------------------------------------------------------------------------
gdppc    overall x..            | 2642.515               | NT         774| country effects:
            overall xit-x..     |           3694         | N          101|  F(100, 654)  = 106.58
       (1)  between xi.-x..     |           3590   93.66 | Nmin        12|      Prob > F =   0.00
       (2a) between x.t-x..     |          322.5    0.73 | N-bar       39|
       (2b) resid xit-xi.-x.t+x..|           952    5.62 | Nmax        62| date effects:
       ------ (2a) + (2b): ------|        ----------------| T           20|  F(19, 654)    =   4.44
       (2)  within  xit-xi.     |          997.2    6.34 | Tmin         1|      Prob > F =   0.00
        (% (2a) between x.t-x..) |                (11.43)| T-bar      7.7|
                                |                        | Tmax        19| balancedness  =   2.67
-------------------------------------------------------------------------------------------------
```

As we can see all the variables lose a lot of observations but not in the same way. The independent variable loses only 133 observations, not so much, indeed the two decompositions of variance are very similar. All the other variables instead, lose more than 3000 observations and this might cause differences between the decompositions with common option and without.

For the democratic degree of the countries the two analyses of variance changes, adding the option, the variability between country rises and all of within country, between years and residual decrease. The gross domestic product per capita changes to a little bit but is not so meaningful because already without the sectioned sample the variability is very concentrated among countries, adding that it adding it is even more concentrated. The letter one variables, the natural logarithm of the populations indeed remains almost such without the restricted sample.

In many cases, using panel data, we may find that these by their nature can't be handle as they are, but we have to analyse them in terms of first or second differences. It could be like that, because often the temporal dimension present unit roots, which lead the estimator of our interest to be biased The new command `xtsum2` allows doing this, but we have to modify manually the variables. We necessary state to Stata to compute the first or second differences and then implement the new program using them as `varlist`.

```
. gen d_tariff= d.tariff
(4,822 missing values generated)
. gen d_polity=d.polity
(2,302 missing values generated)
. gen d_lnpop=d.lnpop
(827 missing values generated)
. gen d_gdppc=d.gdppc
(1,977 missing values generated)
. xtsum2 d_tariff d_polity d_lnpop d_gdppc, common
Variable                        |   Mean   Std.Dev.  % SS |  Observations  |  Test
--------------------------------+-----------------------+----------------+----------------------
d_tariff overall x..            |-1.302259              | NT          487| country effects:
         overall xit-x..        |          4.389        | N            79|  F(78, 390)   =   1.14
         -between xi.-x..       |          1.913   18.80 | Nmin          5|      Prob > F =   0.22
         -within  xit-xi.       |          4.316   81.20 | N-bar        26|
           of which: ---------- |        ----------------| Nmax         47| date effects:
           between x.t-x..      |          .6875    2.33 | T            19|  F(18, 390)   =   0.64
           (in % of within)     |                (2.87)  | Tmin          1|      Prob > F =   0.87
         -resid xit-xi.-x.t+x.. |          4.351   78.87 | T-bar       6.2|
                                |                       | Tmax         17| bal: in/out   = 100.00
                                |                       |                | bal: missing  =   9.07
--------------------------------+-----------------------+----------------+----------------------
d_polity overall x..            | .2361396              | NT          487| country effects:
         overall xit-x..        |          1.803        | N            79|  F(78, 390)   =   1.02
         -between xi.-x..       |          .7387   16.61 | Nmin          5|      Prob > F =   0.44
         -within  xit-xi.       |          1.797   83.39 | N-bar        26|
           of which: ---------- |        ----------------| Nmax         47| date effects:
           between x.t-x..      |          .4098    4.90 | T            19|  F(18, 390)   =   1.35
           (in % of within)     |                (5.88)  | Tmin          1|      Prob > F =   0.15
         -resid xit-xi.-x.t+x.. |          1.783   78.49 | T-bar       6.2|
                                |                       | Tmax         17| bal: in/out   = 100.00
                                |                       |                | bal: missing  =   9.07
```

```
--------------------------------+------------------------+----------------+----------------------
d_lnpop  overall x..            | .0212265               | NT         487| country effects:
         overall xit-x..        |         .0105          | N           79|  F(78, 390)    =  25.56
         -between xi.-x..       |       .009664   83.76  | Nmin         5|       Prob > F =   0.00
         -within  xit-xi.       |        .00462   16.24  | N-bar       26|
           of which: ---------- |        ----------------| Nmax        47| date effects:
           between x.t-x..      |       .001423    1.74  | T           19|  F(18, 390)    =   2.61
            (in % of within)    |                (10.73) | Tmin         1|       Prob > F =   0.00
         -resid xit-xi.-x.t+x.. |       .004465   14.50  | T-bar      6.2|
                                |                        | Tmax        17| bal: in/out    = 100.00
                                |                        |               | bal: missing   =   9.07
--------------------------------+------------------------+----------------+----------------------
d_gdppc  overall x..            | 61.15557               | NT         487| country effects:
         overall xit-x..        |         224.8          | N           79|  F(78, 390)    =   8.79
         -between xi.-x..       |         178.7   62.50  | Nmin         5|       Prob > F =   0.00
         -within  xit-xi.       |         150.2   37.50  | N-bar       26|
           of which: ---------- |        ----------------| Nmax        47| date effects:
           between x.t-x..      |         34.73    2.27  | T           19|  F(18, 390)    =   1.39
            (in % of within)    |                 (6.04) | Tmin         1|       Prob > F =   0.13
```

As we can see, the program works as well as using level variables, with some differences in the results.

Of course, the number of observations is less, because the differences are calculated only when there are two observations with periods adjacent, perhaps there are jumps in the variables we can't calculate them.

We can note that in the variables which measure the trade openness, the between countries and between temporal percentage share decrease much compared with the level, making them within countries and the residual one growth. These results are very different from each other, so we have to be careful to do this in dataset unbalanced because remove information could lead us to distort the nature of the data.

6.5 Data from J.Abrevaya (2006)

This last dataset is a hierarchical one, no temporal dimension is considered. The sample has two-level, the second one is represented by the mothers, and the first one by the children of each mom.

Some of the variables included are related to the second level, these are features of the mothers like age, level of education, married or not, and so on. The variables at the first level, of course, are related to every child, the most important for the analysis that we presented in this example is the number of cigarettes smoked by the mothers while pregnant and the weight at birth of both of that child. This because the relationship between these two variables is the aim of the study example that we rely on (Abrevaya, J. 2006). Finally, the dataset consists also of other characteristics proper of the child, like male or female, week of gestation, and so on.

The paper quotes above analyses the relationship between the weight at birth and the number of cigarettes smoked by the mother while the state of pregnancy, this to research one of the possible damage caused by smoke, the decrease of the child's weight at birth. We proceed to analyse the variance of the response variable to observe how its variability is decomposed according to the different mums to know if and how much the mothers clustering explain the weight at born of their children.

The sample is obviously unbalanced, otherwise, every mum would have the same number of children which is unthinkable. Nevertheless, there isn't a lot of variability among them, the mothers in this dataset have two or three sons.

Since the dataset is a hierarchical one, the more suitable variance component model is the mixed model using restricted maximum likelihood approach. The full maximum likelihood, treat the coefficient as fixed, causing you to lose degrees of freedom, and since the classes are very much we might lose a lot of degrees of freedom. In this case we use the option `mle` in the syntax of `xtsum2`, as follow,

```
. xtset momid idx
       panel variable:  momid (unbalanced)
        time variable:  idx, 1 to 3
                delta:  1 unit

. xtsum2 birwt, rml
---------------------------------- Level 2 (i): momid   Level 1 (j): idx      -----------------------
                                   Hypothetical full sample, Ni x Nj: 8604
                                   Ni: 3978   Nj: 3
-----------------------------------------------------------------------------------------------------
Variable                          |   Mean   Std.Dev.  % SS |  Observations  |  Test
----------------------------------+-------------------------+----------------+-----------------------
birwt       overall xij-x..       |   3470    527.1         | NiNj       8604| lrtest:
       (1) between xi.-x..        |          369.04    58.83 | Ni         3978|
       (2a) between x.j-x..       |           43.92     |.56 | Ni-bar  2868.00|  LR chi2(1)    = 681.28
       (2b) res xij-xi.-x.j+x..   |          375.67    40.61 | Ni-max     3978|   Prob > chi2 =    0.00
       ------ (2a) + (2b): ------|          ----------------| Nj            3|
       (2)  within xij-xi.        |           377.7    41.17 | Nj-bar        2| balancedness   =  72.10
        (% (2a) between x.j-x..) |                   (2.03) | Nj-max     3.00|
-----------------------------------------------------------------------------------------------------
```

As we can see from the column of the percentage share, the between mums variability is the greater one, equal to 58.83%, so we find that the bigger differences in the weight of the infants are due to the differences between a mother than another rather than to differences within the mothers, in other words among the children of the same mum.

The within component is in proportion smaller, equal to 40.17% but is not completely negligible, for example, explain the data with a between regression, like the one shown in (22) in section 4 is not so smart. We would lose a lot of information, even worse usings fixed effect estimator which completely overlook the between variability component or pooled OLS model that overlook both, The best we can do is to choose a model that considers both the variability better still if keep in mind the proportion of the two components of variability. For example, estimating some kind of random effect model or multilevel one.

The between first level variability is almost equal to zero, that is logical if we think of the equation (5) and how it is estimated. We have three categories of the variable that identify the children, 1, 2, and 3, as firstborn, second born, and third born. This sum of squares be calculated in this way, first, the program computes the three partial means one for all the babies that are identified as firstborn and the other two for the second and third-born. After which calculates the sum of the deviations of these partial means from the overall mean. If we found another result bigger than zero, would mean that the weight at birth is related to the birth order, therefore the fact that the mother has already had babies before, affects the new baby weight of birth. All of that would not be supported by scientific evidence, that bound this magnitude at genetic factors or environmental ones. we find proper the result that we expected, further confirmation is given by the F test, which can't refuse the null hypothesis that the coefficient of the dummies that identify, first, second, and third born, are jointly equal to zero. The residual percentage, of course, tends to the within one, because if we have a look at the equation (7), this component is calculated as the difference between the within component and the between j one.

The percentage of unbalancedness, in this case, have no much statistical sense, because it means that there are the 72.1% of the babies related to an abstract total, where the mums of the dataset have all three children, in the examples see before where the dataset was panel had much more meaning. Nevertheless, this measure of balancedness, is not always useful in hierarchical dataset, for example, if in a hierarchical one, the categorical variable that identifies the unit of the first level, don't be assigned as random but have some meaning, in that case, this percentage will keep useful information. To get some more information about this categorical variable we can look at the third section of the program, "observations", at the T-bar item, equal to 2.2, which mean that the mums have in mean 2.2 children. This magnitude gives to us information more clearly than the percentage of unbalancedness in a case like this.

6.6 PISA 2003

In this last example, we use a subsample from PISA 2003, a dataset built by OECD, that collects information about students and schools to which the students belong.

To sampling the dataset we based on the paper of T.Mustafa (2009), which investigates inequalities in educational attainments, in an international comparison.

We don't make use of all the countries of PISA dataset, would be too many, are insert only are five, Finland, Italy, Belgium, Germany, and Korea, the others are no object of this analysis. It be necessary to use the `if` option of `xtsum2` to perform the five different decomposition of variance. we keep the categorical variable existing in PISA dataset, which identifies the countries, and we use that to implement the `if` option.

Assessed students in the OECD dataset, are aged between 15 years and 3 months and 16 years and two months, regardless of the grade in which they are enrolled, to compare students with the same school experience.

It is necessary to generate a new variable that identifies the students because the one we find in PISA 2003 is not suitable for the program. The problem is that this factor variable starts from 1 to NJ_i, `xtsum2` doesn't work with this kind of numeration, it is necessary to create a new variable that starts from 1 to J_i in according to every school size. A manner to implement it in Stata is the command `generate` with the prefix `bysort` as in the statement below

```
bysort schoolid:  g stucod=_n
```

For the details about that statement, that will not be deepened in this discussion, see [D]generate and [D]bysort.

In this example, we use the mark in mathematics as the response variable, like in the paper quoted above and we're going to see how the variance is decomposed. At the first level, we find the students, and the second are the schools to which they belong.

We make five subsamples, one for each country, executing five analyses of variance to understand if there are any differences among them. The distribution of variability of the marks in mathematics may be different between countries that have different school organizations, in a country the major components could be between schools, and in another one could be the variability within the schools.

Since the data collected from OECD consist of different educational systems, the marks in mathematics are measured with different range. To understand how are that range in the various countries, we can use the `summarize` command as follow.

Belgium

```
. summm math
```

Variable	Obs	Mean	Std. Dev.	Min	Max
math	7,279	65.32889	17.42846	0	100

Finland

```
. summ math
```

Variable	Obs	Mean	Std. Dev.	Min	Max
math	5,597	7.552796	1.416086	4	10

Germany

```
. summ math
```

Variable	Obs	Mean	Std. Dev.	Min	Max
math	4,347	3.040488	1.022399	1	6

Italy

```
. summ math
```

Variable	Obs	Mean	Std. Dev.	Min	Max
math	11,375	6.103736	1.48095	1	10

Korea

```
. summ math
```

Variable	Obs	Mean	Std. Dev.	Min	Max
math	5,324	3.06837	1.451781	1	5

Looking at the two latter columns, we can get the range, in Italy, the mark starts from a minimum of 1 and a maximum of 6, for Germany and Finland the mark is in tenths, the Korea range starts from 1 to 5, finally, Belgium register the mark in hundredths.

Now that we know the differences among the countries we can start the comparison in the analysis of variance. As in the example 6.5, we use the multilevel model using the restricted maximum likelihood approach and what follows is the log of xtsum2 per country with the option `rml`,

Belgium

```
. xtset schoolid stucod
       panel variable:  schoolid (unbalanced)
        time variable:  stucod, 1 to 155
                delta:  1 unit

. xtsum2 math, rml
--------------------------------- Level 2 (i): schoolid   Level 1 (j): stucod----------------------
                                  Hypothetical full sample, Ni x Nj: 7279
                                  Ni: 271    Nj: 155
---------------------------------------------------------------------------------------------------
Variable                         |   Mean    Std.Dev.  % SS |  Observations   |  Test
---------------------------------+--------------------------+-----------------+--------------------
math        overall xij-x..      |  65.33     17.43         | NiNj        7279|  lrtest:
       (1) between xi.-x..       |             3.95    5.25 | Ni           271|
       (2a) between x.j-x..      |             .748     .19 | Ni-bar     46.96|  LR chi2(2)    = 162.86
       (2b) res xij-xi.-x.j+x..  |            16.95   94.55 | Ni-max       232|   Prob > chi2 =   0.00
       ------ (2a) + (2b): ------|          ----------------| Nj           155|
       (2)  within xij-xi.       |            16.95   91.10 | Nj-bar        27| balancedness   =  17.33
        (% (2a) between x.j-x..) |                    (.19) | Nj-max    149.00|
---------------------------------------------------------------------------------------------------
```

Germany

```
. xtset schoolid stucod
       panel variable:  schoolid (unbalanced)
        time variable:  stucod, 1 to 35
                delta:  1 unit

. xtsum2 math, rml
--------------------------------- Level 2 (i): schoolid   Level 1 (j): stucod----------------------
                                  Hypothetical full sample, Ni x Nj: 5597
                                  Ni: 197    Nj: 35
---------------------------------------------------------------------------------------------------
Variable                         |   Mean    Std.Dev.  % SS |  Observations   |  Test
---------------------------------+--------------------------+-----------------+--------------------
math        overall xij-x..      |  7.553     1.416         | NiNj        5597|  lrtest:
       (1) between xi.-x..       |            0.300    4.52 | Ni           197|
       (2a) between x.j-x..      |            0.027    0.37 | Ni-bar    159.91|  LR chi2(2)    = 162.86
       (2b) res xij-xi.-x.j+x..  |            1.384   95.44 | Ni-max       194|   Prob > chi2 =   0.00
       ------ (2a) + (2b): ------|          ----------------| Nj            35|
       (2)  within xij-xi.       |            1.406   95.64 | Nj-bar        28| balancedness   =  81.17
        (% (2a) between x.j-x..) |                   (0.00) | Nj-max     35.00|
---------------------------------------------------------------------------------------------------
```

Finland

```
. xtset schoolid stucod
       panel variable:  schoolid (unbalanced)
        time variable:  stucod, 1 to 25
                delta:  1 unit

. xtsum2 math, rml
--------------------------------- Level 2 (i): schoolid   Level 1 (j): stucod-----------------------
                                  Hypothetical full sample, Ni x Nj: 4347
                                  Ni: 207   Nj: 25
----------------------------------------------------------------------------------------------------
Variable                         |  Mean   Std.Dev.  % SS |  Observations   |  Test
---------------------------------+------------------------+-----------------+-----------------------
math       overall xij-x..       |  3.04    1.022         | NiNj       4347|  lrtest:
       (1) between xi.-x..       |          0.200    3.87 | Ni          207|
       (2a) between x.j-x..      |           0.00    0.39 | Ni-bar   173.88|  LR chi2(2)    = 162.86
       (2b) res xij-xi.-x.j+x..  |          1.003   96.13 | Ni-max      201|   Prob > chi2 =   0.00
       ------ (2a) + (2b): ------|        ----------------| Nj           25|
       (2)  within xij-xi.       |          1.025   96.32 | Nj-bar       21| balancedness  =  80.50
        (% (2a) between x.j-x..) |                 (0.00) | Nj-max    25.00|
----------------------------------------------------------------------------------------------------
```

Italy

```
. xtset schoolid stucod
       panel variable:  schoolid (unbalanced)
        time variable:  stucod, 1 to 35
                delta:  1 unit

. xtsum2 math, rml
--------------------------------- Level 2 (i): schoolid   Level 1 (j): stucod-----------------------
                                  Hypothetical full sample, Ni x Nj: 11375
                                  Ni: 404   Nj: 35
----------------------------------------------------------------------------------------------------
Variable                         |  Mean   Std.Dev.  % SS |  Observations   |  Test
---------------------------------+------------------------+-----------------+-----------------------
math       overall xij-x..       | 6.104    1.481         | NiNj      11375|  lrtest:
       (1) between xi.-x..       |          .4810   10.55 | Ni          404|
       (2a) between x.j-x..      |          .0418    0.08 | Ni-bar   325.00|  LR chi2(2)    = 660.54
       (2b) res xij-xi.-x.j+x..  |          1.402   89.37 | Ni-max      401|   Prob > chi2 =   0.00
       ------ (2a) + (2b): ------|        ----------------| Nj           35|
       (2)  within xij-xi.       |          1.401   86.61 | Nj-bar       28| balancedness  =  80.45
        (% (2a) between x.j-x..) |                 (0.09) | Nj-max    35.00|
----------------------------------------------------------------------------------------------------
```

Korea

```
. xtset schoolid stucod
       panel variable:  schoolid (unbalanced)
        time variable:  stucod, 1 to 40
                delta:  1 unit

. xtsum2 math, rml
--------------------------------- Level 2 (i): schoolid   Level 1 (j): stucod-----------------------
                                  Hypothetical full sample, Ni x Nj: 5324
                                  Ni: 149   Nj: 40
----------------------------------------------------------------------------------------------------
Variable                         |  Mean   Std.Dev.  % SS |  Observations   |  Test
---------------------------------+------------------------+-----------------+-----------------------
math       overall xij-x..       | 3.068    1.452         | NiNj       5324|  lrtest:
       (1) between xi.-x..       |          .6804   22.12 | Ni          149|
       (2a) between x.j-x..      |          .0671    0.22 | Ni-bar   133.10|  LR chi2(2)    = 980.24
       (2b) res xij-xi.-x.j+x..  |           1.28   77.66 | Ni-max      146|   Prob > chi2 =   0.00
       ------ (2a) + (2b): ------|        ----------------| Nj           40|
       (2)  within xij-xi.       |          1.277   77.88 | Nj-bar       36| balancedness  =  89.63
        (% (2a) between x.j-x..) |                 (0.28) | Nj-max    40.00|
----------------------------------------------------------------------------------------------------
```

In this case, in which we have to compare magnitudes registered with a different unit of measure, the percentage share is essential.

The first thing that we notice, is the within schools percentage share, clearly is the greater one in all country. We understand that the differences in the mark are due more to the differences between students rather than the fact that a student belongs to a school instead of another. So the variability is bounded much at the individual characteristics of the students like attitude to mathematics or the number of hours of study per day, and so on. Since the within schools percentage share is the greater one, let's see how much is big from a country to another. The countries, in which this magnitude is bigger than the others, are Belgium for first with 94.55% follow by Germany with 91.61% and Finland with 92.17%, so very small differences among these. In these countries, the share of variability due to the school is less than 10%, which means that there are very small differences between a school and another. The students in Germany, Finland, and Belgium will not find significant differences in the ways of teaching maths from a school to another. In Italy the differences between schools are a little bit bigger with a between schools percentage share equal to 86.61%, so the choice of the school from the students is a little bit more substantial to learn mathematics. The country with more disparities in our subsample is Korea. The latter has a value of 75.78% which means that only three-quarters of the variability in the mark is due to the student's characteristics and a quarter of the differences over the mark is due to the school where the students belong. The differences among the percentage could be caused by the different school systems, for example, the amount of public expenditure give to instruction that in the case is small could produce more disparities between public and private schools.

Before continuing the comparison, it is necessary to explain the meaning of the between students variability.

If we think about how the program calculates the magnitude we understand immediately that in this case and As in the example 6.5 it doesn't have any statistical meaning.

If we go back to the equation (5), we see that in this case, the sum of square calculated is between the student that in the varied school occupied a certain position. The student in the dataset has a code number that identifies them within the schools, starting from 1 to J_i as the number of students in each school.

What `xtsum2` doing is to calculate the partial mean of the students that have code number equal to 1, the partial mean of those who have code number equal to 2, and so on until the maximum J_i. Thereafter computes the deviations of these partial means from the overall mean.

The student code however is assigned randomly, so don't have any sense use this as a categorical variable to decompose the variability over that. Therefore the between second level is a component of variance that we can't explain as the residual one, that as we know is equal to the within component, have no sense, in that case, split out into these two components.

7 Conclusion

The command `xtsum2`, as we saw, is suitable in many situations in which the variance needs to be decomposed. The program works well both with hierarchical datasets and with panel datasets. As we saw in examples 6.5 and 6.6, specifying the maximum likelihood option or the restricted maximum likelihood one, `xtsum2` computes different variance component models, to guarantee the best model according to hierarchical datasets. Also on panel datasets, could make an analysis of variance using the new command, choosing the best variance component model also in this situation. The strength of this program is proper the fact that it allows many options for all the situations, it could calculate four different models, the fixed effect model that is more suitable in panel dataset in which the class analysed are proper the ones that we should make inference on. The random effect model also is more suitable for the panel dataset but is better to use when the classes are randomly extracted from a population of classes in which we are interested. The multilevel models are more suitable for hierarchical datasets, as said before, but the safer one is that use restricted maximum likelihood approach. This because doesn't treat the coefficient as fixed, such as the full maximum likelihood approach, but estimates in two steps removing the restrictions in the second one, giving as result degrees of freedom kept. This is a problem especially when the class level is very big because we have to estimate a lot of fixed effect, but this model is used as well, perhaps this problem could be ignored. So the program can make this model too, to the situations that required this specific model, for instance, the full maximum likelihood one is better to make some inference with, proper for the fact that treat the coefficient as fixed, allowing more tests.

There is another option very useful, `common`, that as shown in example 6.4, when the variables that we should analyse are not gathered in the same way and so we have to consider only the observations common to all of them. The program allows that without sectioning the dataset, avoiding a lot of annoying steps.

Another goodness of the new command is the output, very clear and understandable, it is split out into three sections that keep a lot of information. For first the standard deviations and percentage share section, that show us all the variance components according to the different models chosen. This program avoids a lot of steps to calculate the percentage share that as we see in the various examples is very useful when we are interested in the estimation of variance component.

The second report all the informations about how the observations are nested, such as the number of classes, the numerousness of classes the total of observations, and the medium value of the first two for handle unbalanced dataset as well without use always other commands to know them.

The last one, but not the least, reports useful statistics, the index of balancedness common to all the options that concern the model choice. That makes a ratio between the observations that we get and a theoretical sample perhaps the real sample is balancedness. The other statistics are different according to the specific model, and give us a measure of the significance of the variance components, using the F-statistic for the fixed effect model, the Breush-Pagan test for random effect model, and the likelihood ratio test for the multilevel model. In this manner, we have a measure of that for all the models that we need to use.

In conclusion, the program adds to Stata a useful tool, because gives us a lot of information without using many commands already available in Stata, avoiding waste of time.

8 Appendix

8.1 Examples statement

Interest rate and inflaction rate in Euro-area

```
use "R_euroarea.dta", clear

xtset codice anno
xtsum2 tint

xtsum tint
display r(sd_b)^2/(r(sd_w)^2+r(sd_b)^2)
display r(sd_w)^2/(r(sd_w)^2+r(sd_b)^2)
local bi_ss_xtsum=r(sd_b)^2*(r(N)-r(Tbar))
local wi_ss_xtsum=r(sd_w)^2*(r(N)-r(n))
display `bi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')
display `wi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')

xtdescribe
local N=r(sum_w)
local J=r(mean)

mixed tint || codice:  , stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')
display `wi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')

mixed tint || codice:, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')
display `wi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')

mixed tint ||_all:R.codice || anno:, stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + xp(2*[lns2_1_1]_cons))
```

```
local r_ss_mixed3_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bi_ss_mixed3_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bj_ss_mixed3_mle=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bi_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bj_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')

mixed tint ||_all:R.codice || anno:, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + exp(2*[lns2_1_1]_cons))
local r_ss_mixed3_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bi_ss_mixed3_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bj_ss_mixed3_reml=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bi_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bj_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')

xtreg tint, re
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re=e(sigma_u)^2*(e(N)-e(Tbar))
display `wi_ss_re'/(`wi_ss_re'+`bi_ss_re')
display `bi_ss_re'/(`wi_ss_re'+`bi_ss_re')

xtreg tint, re mle
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re_mle=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re_mle=e(sigma_u)^2*(e(N)-e(g_avg))
display `wi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')
display `bi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')

xtreg tint, fe
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local bi_ss_fe=e(sigma_u)^2*(e(N)-e(N_g))
local wi_ss_fe=e(sigma_e)^2*(e(N)-e(g_avg))
display `bi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')
display `wi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')
```

```
anova tint codice anno
display ((e(ss_2)/e(df_2))/(e(df_1)+1))^0.5
display ((e(ss_1)/e(df_1))/(e(df_2)+1))^0.5
display (e(rss)/e(df_r))^0.5
display e(ss_1)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(ss_2)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(rss)/(e(ss_1)+e(ss_2)+e(rss))*100

loneway tint codice
display r(sd_w)^2/(r(sd_b)^2+r(sd_w)^2)
display r(sd_b)^2/(r(sd_b)^2+r(sd_w)^2)
local bi_ss_loneway=r(sd_b)^2*(r(N)-`J')
local wi_ss_loneway=r(sd_w) ^2*(r(N)-`n')
display `bi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
display `wi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
```

wagepan

```
use "wagepan.dta", clear

graph box lwage, over(year) yline(1.649147)

xtset nr year
xtsum2 lwage

xtsum lwage
display r(sd_b)^2/(r(sd_w)^2+r(sd_b)^2)
display r(sd_w)^2/(r(sd_w)^2+r(sd_b)^2)
local bi_ss_xtsum=r(sd_b)^2*(r(N)-r(Tbar))
local wi_ss_xtsum=r(sd_w)^2*(r(N)-r(n))
display `bi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')
display `wi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')

xtdescribe
local N=r(sum_w)
local J=r(mean)

mixed lwage || nr:  , stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')
display `wi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')

mixed lwage || nr:, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')
display `wi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')

mixed lwage ||_all:R.nr || year:, stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + xp(2*[lns2_1_1]_cons))
local r_ss_mixed3_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bi_ss_mixed3_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
```

```
local bj_ss_mixed3_mle=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bi_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bj_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')

mixed lwage ||_all:R.nr || year:, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + exp(2*[lns2_1_1]_cons))
local r_ss_mixed3_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bi_ss_mixed3_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bj_ss_mixed3_reml=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bi_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bj_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')

xtreg lwage, re
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re=e(sigma_u)^2*(e(N)-e(Tbar))
display `wi_ss_re'/(`wi_ss_re'+`bi_ss_re')
display `bi_ss_re'/(`wi_ss_re'+`bi_ss_re')

xtreg lwage, re mle
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re_mle=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re_mle=e(sigma_u)^2*(e(N)-e(g_avg))
display `wi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')
display `bi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')

xtreg lwage, fe
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local bi_ss_fe=e(sigma_u)^2*(e(N)-e(N_g))
local wi_ss_fe=e(sigma_e)^2*(e(N)-e(g_avg))
display `bi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')
display `wi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')

anova lwage nr year
```

```
display ((e(ss_2)/e(df_2))/(e(df_1)+1))^0.5
display ((e(ss_1)/e(df_1))/(e(df_2)+1))^0.5
display (e(rss)/e(df_r))^0.5
display e(ss_1)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(ss_2)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(rss)/(e(ss_1)+e(ss_2)+e(rss))*100

loneway lwage nr
display r(sd_w)^2/(r(sd_b)^2+r(sd_w)^2)
display r(sd_b)^2/(r(sd_b)^2+r(sd_w)^2)
local bi_ss_loneway=r(sd_b)^2*(r(N)-`J')
local wi_ss_loneway=r(sd_w) ^2*(r(N)-`n')
display `bi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
display `wi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
```

abdata

```
use "abdata.dta", clear

xtset id year
xtsum2 n

xtsum n
display r(sd_b)^2/(r(sd_w)^2+r(sd_b)^2)
display r(sd_w)^2/(r(sd_w)^2+r(sd_b)^2)
local bi_ss_xtsum=r(sd_b)^2*(r(N)-r(Tbar))
local wi_ss_xtsum=r(sd_w)^2*(r(N)-r(n))
display `bi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')
display `wi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')

xtdescribe
local N=r(sum_w)
local J=r(mean)

mixed n || id:  , stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')
display `wi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')

mixed n || id:, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')
display `wi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')

mixed n ||_all:R.id || year:, stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + xp(2*[lns2_1_1]_cons))
local r_ss_mixed3_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bi_ss_mixed3_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bj_ss_mixed3_mle=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
```

```
+`bj_ss_mixed3_mle')
display `bi_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bj_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')

mixed n ||_all:R.id || year:, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + exp(2*[lns2_1_1]_cons))
local r_ss_mixed3_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bi_ss_mixed3_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bj_ss_mixed3_reml=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bi_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bj_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')

xtreg n, re
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re=e(sigma_u)^2*(e(N)-e(Tbar))
display `wi_ss_re'/(`wi_ss_re'+`bi_ss_re')
display `bi_ss_re'/(`wi_ss_re'+`bi_ss_re')

xtreg n, re mle
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re_mle=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re_mle=e(sigma_u)^2*(e(N)-e(g_avg))
display `wi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')
display `bi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')

xtreg n, fe
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local bi_ss_fe=e(sigma_u)^2*(e(N)-e(N_g))
local wi_ss_fe=e(sigma_e)^2*(e(N)-e(g_avg))
display `bi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')
display `wi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')

anova n id year
display ((e(ss_2)/e(df_2))/(e(df_1)+1))^0.5
display ((e(ss_1)/e(df_1))/(e(df_2)+1))^0.5
```

```
display (e(rss)/e(df_r))^0.5
display e(ss_1)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(ss_2)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(rss)/(e(ss_1)+e(ss_2)+e(rss))*100

loneway n id
display r(sd_w)^2/(r(sd_b)^2+r(sd_w)^2)
display r(sd_b)^2/(r(sd_b)^2+r(sd_w)^2)
local bi_ss_loneway=r(sd_b)^2*(r(N)-`J')
local wi_ss_loneway=r(sd_w) ^2*(r(N)-`n')
display `bi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
display `wi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
```

TSCS

```
use "TCSC.dta", clear

gen dummy=1 if polity!=.  & gdppc!=.  & tariff!=.  & lnpop!=.
replace dummy=0 if dummy==.
xtsum polity lnpop gdppc tariff if dummy==1
display r(sd_b)^2/(r(sd_w)^2+r(sd_b)^2)
display r(sd_w)^2/(r(sd_w)^2+r(sd_b)^2)
local bi_ss_xtsum=r(sd_b)^2*(r(N)-r(Tbar))
local wi_ss_xtsum=r(sd_w)^2*(r(N)-r(n))
display `bi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')
display `wi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')

xtdescribe
local N=r(sum_w)
local J=r(mean)

xtset country date
xtsum2 polity lnpop gdppc tariff, common

gen d_tariff= d.tariff
gen d_polity=d.polity
gen d_lnpop=d.lnpop
gen d_gdppc=d.gdppc
xtset country date
xtsum2 d_tariff d_polity d_lnpop d_gdppc, common

xtdescribe if dummy==1
local n=r(sum_w)
local J=r(mean)

mixed tariff || country:  if dummy==1, stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')
display `wi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')

mixed tariff || country:if dummy==1, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')
```

```
display `wi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')

mixed tariff ||_all:R.country || date:if dummy==1, stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + xp(2*[lns2_1_1]_cons))
local r_ss_mixed3_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bi_ss_mixed3_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bj_ss_mixed3_mle=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bi_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bj_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')

mixed tariff ||_all:R.country || date:if dummy==1, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + exp(2*[lns2_1_1]_cons))
local r_ss_mixed3_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bi_ss_mixed3_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bj_ss_mixed3_reml=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bi_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bj_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')

xtreg tariff if dummy==1, re
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re=e(sigma_u)^2*(e(N)-e(Tbar))
display `wi_ss_re'/(`wi_ss_re'+`bi_ss_re')
display `bi_ss_re'/(`wi_ss_re'+`bi_ss_re')

xtreg tariff if dummy==1, re mle
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re_mle=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re_mle=e(sigma_u)^2*(e(N)-e(g_avg))
display `wi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')
```

```
display `bi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')

xtreg tariff if dummy==1, fe
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local bi_ss_fe=e(sigma_u)^2*(e(N)-e(N_g))
local wi_ss_fe=e(sigma_e)^2*(e(N)-e(g_avg))
display `bi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')
display `wi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')

anova tariff country date if dummy==1
display ((e(ss_2)/e(df_2))/(e(df_1)+1))^0.5
display ((e(ss_1)/e(df_1))/(e(df_2)+1))^0.5
display (e(rss)/e(df_r))^0.5
display e(ss_1)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(ss_2)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(rss)/(e(ss_1)+e(ss_2)+e(rss))*100

loneway tariff country if dummy==1
display r(sd_w)^2/(r(sd_b)^2+r(sd_w)^2)
display r(sd_b)^2/(r(sd_b)^2+r(sd_w)^2)
local bi_ss_loneway=r(sd_b)^2*(r(N)-`J')
local wi_ss_loneway=r(sd_w) ^2*(r(N)-`n')
display `bi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
display `wi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
```

Data from J.Abrevaya (2006)

```
use "smoking.dta", clear

xtset momid idx
xtsum2 birwt

xtsum birwt
display r(sd_b)^2/(r(sd_w)^2+r(sd_b)^2)
display r(sd_w)^2/(r(sd_w)^2+r(sd_b)^2)
local bi_ss_xtsum=r(sd_b)^2*(r(N)-r(Tbar))
local wi_ss_xtsum=r(sd_w)^2*(r(N)-r(n))
display `bi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')
display `wi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')

xtdescribe
local n=r(sum_w)
local J=r(mean)

mixed birwt || momid:  , stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')
display `wi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')

mixed birwt || momid:, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')
display `wi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')

xtreg birwt, re
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re=e(sigma_u)^2*(e(N)-e(Tbar))
display `wi_ss_re'/(`wi_ss_re'+`bi_ss_re')
display `bi_ss_re'/(`wi_ss_re'+`bi_ss_re')

xtreg birwt, re mle
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
```

```
local wi_ss_re_mle=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re_mle=e(sigma_u)^2*(e(N)-e(g_avg))
display `wi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')
display `bi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')

xtreg birwt, fe
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local bi_ss_fe=e(sigma_u)^2*(e(N)-e(N_g))
local wi_ss_fe=e(sigma_e)^2*(e(N)-e(g_avg))
display `bi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')
display `wi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')

anova birwt idx momid
display ((e(ss_2)/e(df_2))/(e(df_1)+1))^0.5
display ((e(ss_1)/e(df_1))/(e(df_2)+1))^0.5
display (e(rss)/e(df_r))^0.5
display e(ss_1)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(ss_2)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(rss)/(e(ss_1)+e(ss_2)+e(rss))*100

loneway birwt momid
display r(sd_w)^2/(r(sd_b)^2+r(sd_w)^2)
display r(sd_b)^2/(r(sd_b)^2+r(sd_w)^2)
local bi_ss_loneway=r(sd_b)^2*(r(N)-`J')
local wi_ss_loneway=r(sd_w) ^2*(r(N)-`n')
display `bi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
display `wi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100

xtset idx momid xtdescribe
local n=r(sum_w)
local J=r(mean)

mixed birwt ||_all:R.idx || momid:, stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + xp(2*[lns2_1_1]_cons))
local r_ss_mixed3_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bj_ss_mixed3_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bi_ss_mixed3_mle=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bi_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bj_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
```

```
mixed birwt ||_all:R.momid || idx:, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + exp(2*[lns2_1_1]_cons))
local r_ss_mixed3_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bj_ss_mixed3_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bi_ss_mixed3_reml=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bi_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bj_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
```

PISA 2003

```
xtset schoolid stucod
xtsum math
display r(sd_b)^2/(r(sd_w)^2+r(sd_b)^2)
display r(sd_w)^2/(r(sd_w)^2+r(sd_b)^2)
local bi_ss_xtsum=r(sd_b)^2*(r(N)-r(Tbar))
local wi_ss_xtsum=r(sd_w)^2*(r(N)-r(n))
display `bi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')
display `wi_ss_xtsum'/(`wi_ss_xtsum'+`bi_ss_xtsum')

xtset schoolid stucod
xtsum2 math

xtdescribe
local n=r(sum_w)
local J=r(mean)

mixed math || schoolid:, stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')
display `wi_ss_mixed2_mle'/(`wi_ss_mixed2_mle'+`bi_ss_mixed2_mle')

mixed math || schoolid:, reml stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons))
local bi_ss_mixed2_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local wi_ss_mixed2_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n')
display `bi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')
display `wi_ss_mixed2_reml'/(`wi_ss_mixed2_reml'+`bi_ss_mixed2_reml')

xtreg math , re
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re=e(sigma_e)^2*(e(N)-e(N_g))
local bi_ss_re=e(sigma_u)^2*(e(N)-e(Tbar))
display `wi_ss_re'/(`wi_ss_re'+`bi_ss_re')
display `bi_ss_re'/(`wi_ss_re'+`bi_ss_re')

xtreg math , re mle
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
local wi_ss_re_mle=e(sigma_e)^2*(e(N)-e(N_g))
```

```
local bi_ss_re_mle=e(sigma_u)^2*(e(N)-e(g_avg))
display `wi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')
display `bi_ss_re_mle'/(`wi_ss_re_mle'+`bi_ss_re_mle')

xtreg math , fe
display e(sigma_u)^2/(e(sigma_u)^2+e(sigma_e)^2)
display e(sigma_e)^2/(e(sigma_u)^2+e(sigma_e)^2)
local bi_ss_fe=e(sigma_u)^2*(e(N)-e(N_g))
local wi_ss_fe=e(sigma_e)^2*(e(N)-e(g_avg))
display `bi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')
display `wi_ss_fe'/(`wi_ss_fe'+`bi_ss_fe')

anova math schoolid stucod
display ((e(ss_2)/e(df_2))/(e(df_1)+1))^0.5
display ((e(ss_1)/e(df_1))/(e(df_2)+1))^0.5
display (e(rss)/e(df_r))^0.5
display e(ss_1)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(ss_2)/(e(ss_1)+e(ss_2)+e(rss))*100
display e(rss)/(e(ss_1)+e(ss_2)+e(rss))*100

loneway math schoolid
display r(sd_w)^2/(r(sd_b)^2+r(sd_w)^2)
display r(sd_b)^2/(r(sd_b)^2+r(sd_w)^2)
local bi_ss_loneway=r(sd_b)^2*(r(N)-`J')
local wi_ss_loneway=r(sd_w)^2*(r(N)-`n')
display `bi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100
display `wi_ss_loneway'/(`wi_ss_loneway'+`bi_ss_loneway')*100

xtset stucod schoolid xtdescribe
local n=r(sum_w)
local J=r(mean)

mixed math ||_all:R.stucod || schoolid:, stddev
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + exp(2*[lns2_1_1]_cons))
local r_ss_mixed3_mle=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bj_ss_mixed3_mle=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bi_ss_mixed3_mle=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bi_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')
display `bj_ss_mixed3_mle'/(`r_ss_mixed3_mle'+`bi_ss_mixed3_mle'
+`bj_ss_mixed3_mle')

mixed math ||_all:R.schoolid || stucod:, reml stddev
```

```
display exp(2*[lns1_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lns2_1_1]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) +exp(2*[lns2_1_1]_cons))
display exp(2*[lnsig_e]_cons)/(exp(2*[lns1_1_1]_cons)
+ exp(2*[lnsig_e]_cons) + exp(2*[lns2_1_1]_cons))
local r_ss_mixed3_reml=(exp([lnsig_e]_cons))^2*(e(N)-`n'-`J'-2)
local bi_ss_mixed3_reml=(exp([lns1_1_1]_cons))^2*(e(N)-`J')
local bj_ss_mixed3_reml=(exp([lns2_1_1]_cons))^2*(e(N)-`n')
display `r_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bi_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
display `bj_ss_mixed3_reml'/(`r_ss_mixed3_reml'+`bi_ss_mixed3_reml'
+`bj_ss_mixed3_reml')
```

9 Bibliography

- Abrevaya, J. 2006. "Estimating the effect of smoking on birth outcomes using a matched"
- Blundell, R. and S. Bond (1998), "Initial conditions and moment restrictions in dynamic panel data models", Journal of Econometrics
- Brown W.J. 2017, "MCMC estimation in MLwiN"
- Bontempi M.E., Golinelli R. (2014), "Panel data econometrics: theory and applications in STATA"
- De Leeuw J. and Meijer E. (2007), "Hand book of multilevel analysis"
- Gibbons S. Overman H. G. And Pelkonen P (2012) "The decomposition of variance into individual and group components with an application to area disparities"(dove lo metto)
- Gutierrez R.G., Carter S., Drukker D.M. pg 15 "On boundary-value likelihood-ratio tests"
- Hox J.J. (2010) "multilevel analysis, techniques and applications"
- Marchenko Y. (2006) "Estimating variance components in stata"
- Milner and Kubota (2005) "Why the move to free trade? Democracy and trade policy in the developing countries
- Mustafa M. (2009) "The Anatomy of Inequalities in Attainments: An International Investigation on Stratification and Choice"
- Olivetti C. (2013) "in the name of the son (and the daughter): international mobility in the united states, 1850-1830"
- Rabe-Hesketh S. and Skrondal A. (2012) "multilevel and longitudinal modelling using Stata"
- Searle S.R. (1971), "Topics in variance components estimation"
- Searle S.R., Casella G. and McCulloch (2016), "variance components"
- Shunk R. (2013) "Within and between estimates in random-effects models: Advantages and drowbacks of correlated random effects and hybrid models"
- Vella F. And Verbeek M. (1998) "Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young man"
- Verbeek M. (2017) "A guide to modern econometrics" - 5. ed Hoboken : Wiley Custom
- Wooldridge J.M. (2020) "introductory econometrics, a modern approach"
- Wooldridge J.M 2020 "econometric analysis of cross section and panel data"