Scientific Team Project, Multi-level modeling

Adel Memariani

Abdullah Al Zubaer

Karthik Kadajji

May 6, 2020

3 Synthetic Data

In this section we are going to discuss our work related to synthetic data that we have created. We have performed exploratory analysis and fitted two level MLM on the synthetic data. The rest of this section is structured in the following order. In [sec:Synthetic Data]Section 3.1 we are going to present description of our dataset. In [sec:Exploratory Analyses2]Section 3.2. we are going to do exploratory analysis on our dataset, with results and discussion. In [sec:Fitting Two Level MLM on Synthetic data]Section 3.3 we are going to fit two level MLM on our synthetic data, test the assumptions of MLM of our final model and finally present the result with discussions.

3.1 Description of Synthetic Data

ID	Semester	ECTS_Cleared	Gender	Marital_Status
1	1	21	1	0
1	2	22	1	0
1	3	24	1	0
2	1	27	1	0
2	2	33	1	0
2	3	30	1	0
3	1	21	1	0
3	2	30	1	0
3	3	39	1	0
4	1	21	1	0
4	2	33	1	0
4	3	33	1	0
5	1	16	1	0
5	2	17	1	0
5	3	18	1	0
6	1	27	0	0
6	2	28	0	0
6	3	30	0	0

Table 1: Snippet of our synthetic data(showing first 6 individual)

Table 1 shows our synthetic data that we have created. The dataset is in person-period format which means that each students will have multiple records, one for each occasion. Here occasion represents number of semesters. There are three waves of longitudinal data for 60 individuals. In contrast, there is another way to store our longitudinal data which is in person-level format. For analysis of longitudinal data, person-period format is preferred over person-level format[2]. We have created the synthetic data following some assumptions. The assumptions are listed below

1. Male married individual clears more ECTS compared to male unmarried individual. Male married individual decreases the number of ECTS cleared decreases as number of semester compared to male unmarried individual; which is the slope.

2. Female married individual clears less ECTS in their first semester compared to female unmarried individual. Female married and unmarried individual has increasing number of ECTS cleared as number of semester increases; which is the slope.

Total number of students in our dataset is 60, there are 57 fictive students and 3 real students which represents each of us. ID is a unique identifier for each individuals and is unique for all occasion of an individual's measurements. We have two time-invariant predictors (which means the value of these predictors is constant for each individuals), Marital_Status, representing if the student is married or not. It is represented in a binary format, where married student = 1, and unmarried student = 0. The second time-invariant predictor we have is Gender, which is also in binary format, where Gender = 1 represents male, and Gender = 0 represents female. Our final predictor is a time-variant predictor (which means the value of this predictor varies depending on the time of measurement), which is semester (it is our time indicator) and represent at which point of time we have the individual's record. Its numerical value ranges from 1 to 3. There are three waves of longitudinal data for 60 individuals Our outcome variable is ECTS_cleared which means, ECTS achieved by a student in each semester during the occasion of measurement. While preparing the dataset we had different number of semester that we were currently enrolled in. To have the equal number of waves among us we chose to have semester ranging from 1 to 3, which is the minimum number of semester that we have among us.

3.2 Exploratory Analyses

We have carried out exploratory analyses on our synthetic data. This analysis will help us to observe the patterns of each individuals' in our dataset, and will confirm the predictors that we are using if they have affects on the growth trajectories of the individuals or not, across groups. Which will subsequently guide us in choosing predictors that are important, and leave out the predictors that does not have affects on the growth trajectory of the individuals and wll not be part of our analysis.

This analysis is important because we are going to fit two level MLM on our synthetic data, which will be discussed in [sec:Fitting Two Level MLM on Synthetic data]Section 3.3, and we have to know if the predictors that we are using has affects on the individuals or not.

3.2.1 Intradividual Change

After creating our synthetic data in person period format we have explored how each individuals changes with respect to semesters. Figure 2 shows us how individuals outcome changes over each semester. This is also known as *Intraindividual-Changes*, which means we can track the changes of each individuals separately with respect to semester. We have 60 students in our dataset and due to space constrain we have shown only 16 individuals. This study helps us to show how individual changes over time.

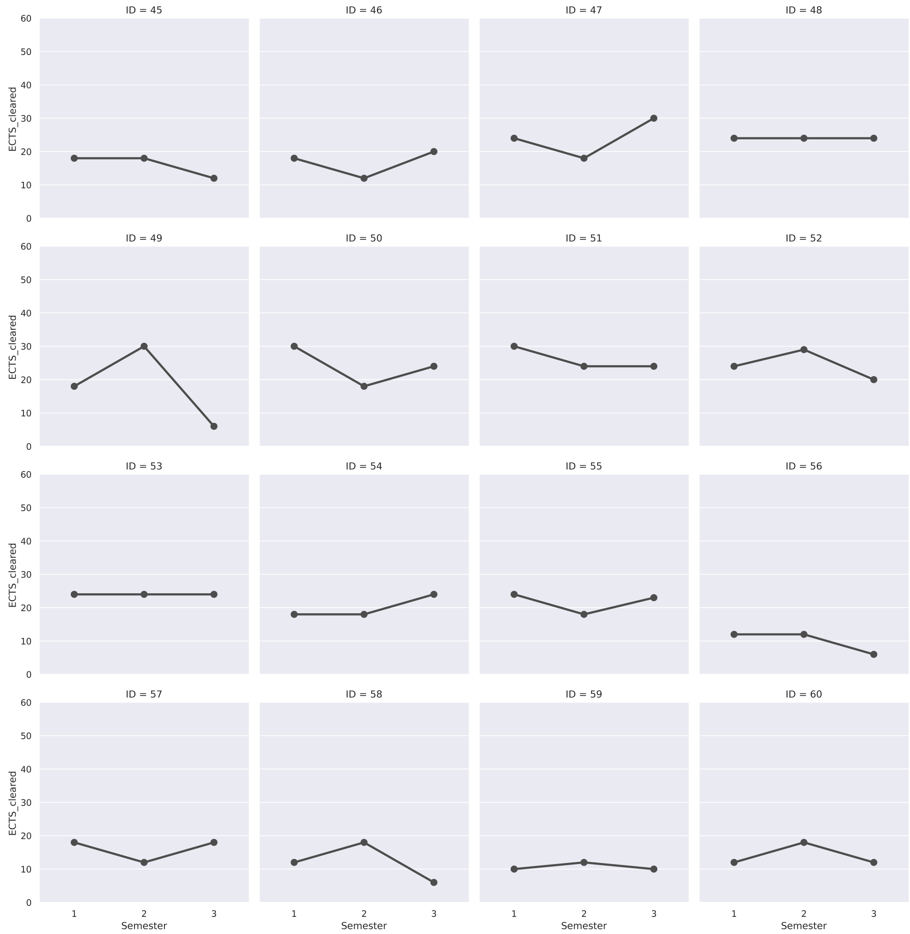

Figure 2: Snippet of empirical growth plot of 16 individuals.

We can observe from the Figure 2 that each individuals clears ECTS differently from others and also ECTS cleared is not constant for a particular individual, it changes as the number of semester increases.

3.2.2 OLS Trajectories For All The Individuals

We have used ordinary least squares (OLS) regression with common functional form of linear change model for fitting to all the individuals in our synthetic data. We have performed this study to have an over all idea how each individual's ECTS_Cleared changes with respect to semester.

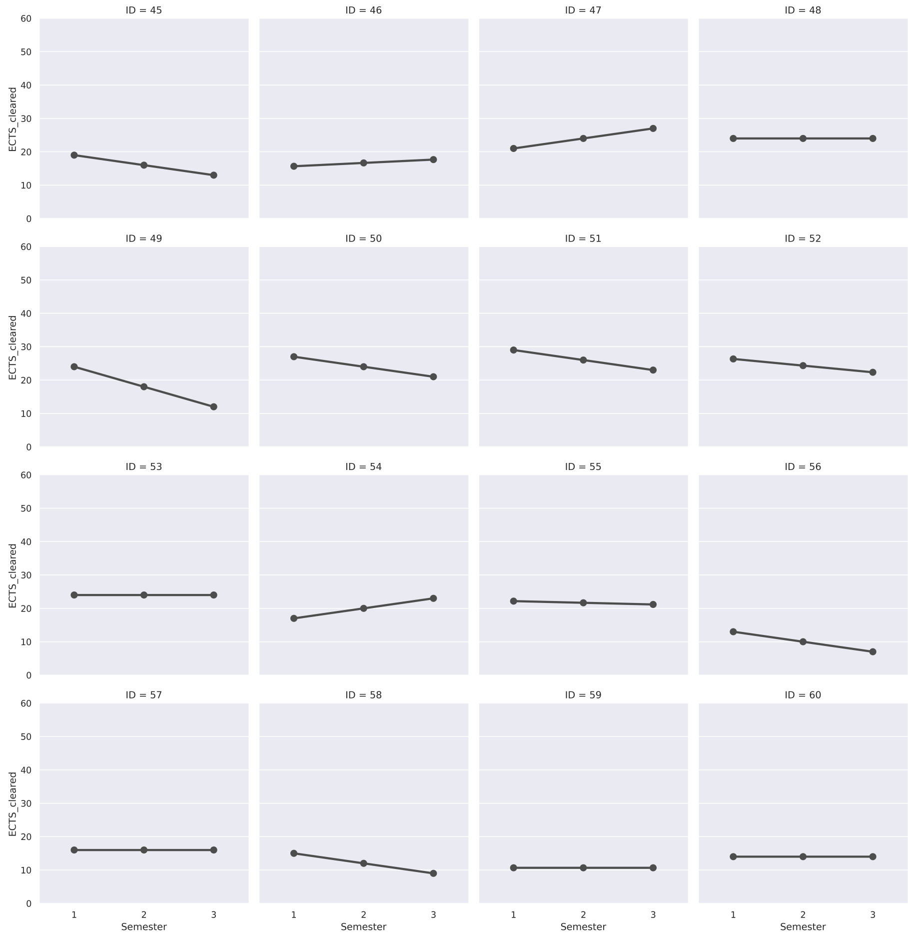

Figure 3: Snippet of fitted OLS trajectories of 16 individuals.

We have fitted OLS regression with common functional form of linear change to all the individuals and have showed 16 individuals on Figure 2 due to space constrain. We can observe from Figure 2 that individual 46, 47 and 54 shows steady growth, individual 50 shows steady decline and individual 59 shows flat growth, with respect to ECTS_Cleared.

3.2.3 Population OLS Trajectories

We have fitted OLS trajectories for all the individuals and studied how each individual's ECTS_Cleared changes when the number of semester increases. In Figure 4 we can observe that the number of ECTS cleared for each individual increases and decreases as the number of semester increases. This study was done for having an over all view of the characteristics of our synthetic data.

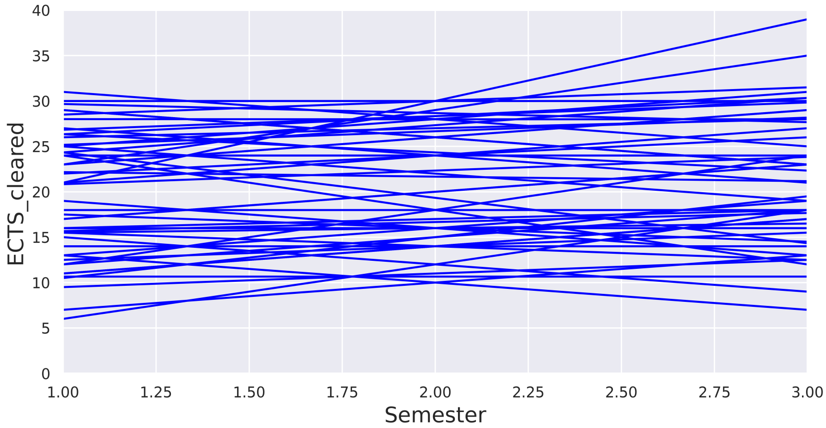

Figure 4: Fitted OLS trajectories for all the individuals

3.2.4 Interindividual Change

We have two time-invariant predictor in our dataset, marital status and gender. In this subsection we are going to study the effect of these predictor in our dataset which means, how individuals changes their trajectories when they are married or unmarried and gender is male or female. This study was carried out to test the assumptions that we have mentioned on [sec:Synthetic Data]Section 3.1

We have two research questions that we want to study

1. How does every individual changes with respect to our predictors in groups

2. Does the change in trajectory for every individual varies with respect to our predictors in groups

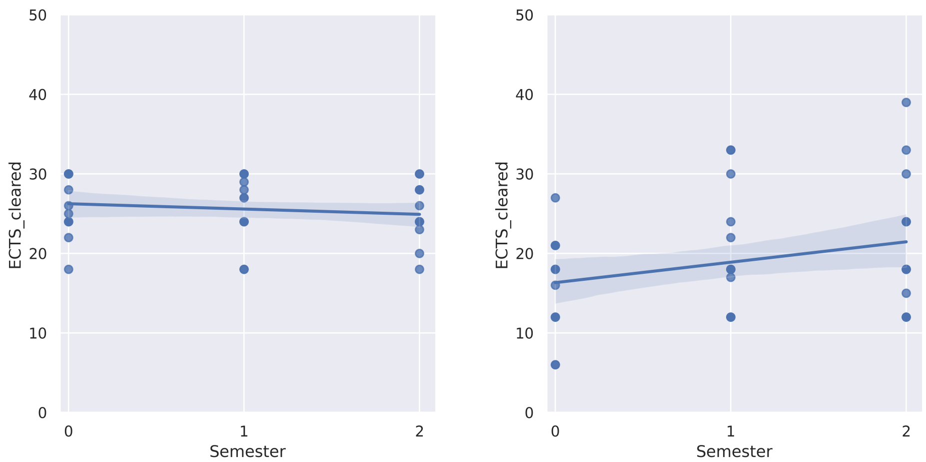

Figure 5: Fitted average OLS trajectory for male married and male unmarried individuals(left figure for male married, right figure for male unmarried)

The group has been created by considering Gender and Martial Status. We have separated all the individuals based on Gender and Marital Status, and then analysed the change in group with respect to these predictors. For analysis we have fitted average OLS change trajectory for each group.

From Figure 5 we can conclude the following. On average male married individuals clears more ECTS compared to males unmarried on their first semester but declines in further semester. On average male unmarried clears less ECTS compared to male married on their first semester but increases in further semesters.

Next we have studied how females who are married and unmarried changes their trajectories and compared them.

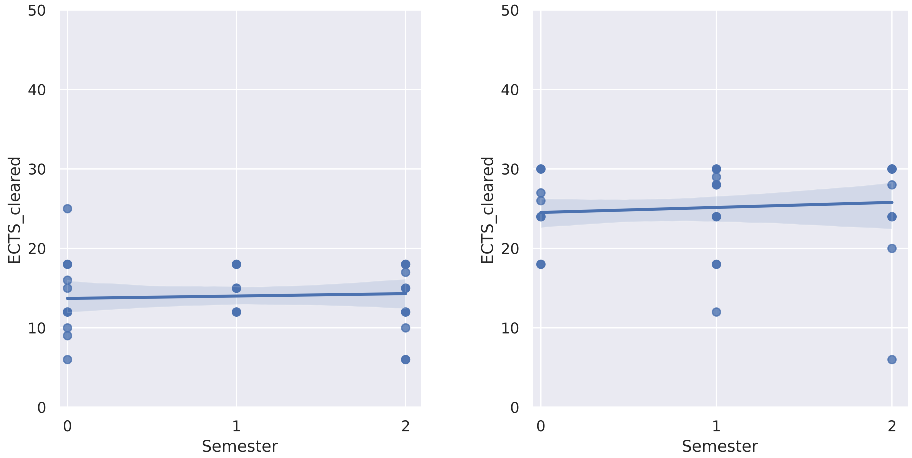

Figure 6: Fitted average OLS trajectory for female married and female unmarried individuals((left figure for female married, right figure for female unmarried)

From Figure 6 we can conclude the following. On average female married individuals clears less ECTS compared to females unmarried on their first semester. On average female unmarried clears more ECTS compared to male married on their first semester. Female married has slightly less increasing slope compared to female unmarried individuals.

3.2.5 Conclusion

From our exploratory analysis of our synthetic data we can conclude that the assumption, mentioned on [sec:Synthetic Data]Section 3.1, that we have considered while creating the dataset is reflected on our experiment. Which also reflects that the predictors that we have introduced does affects number of ECTS cleared when the number of semester the individual increases.

3.3 Fitting Two Level MLM on Synthetic data

As we have already discussed before in the introduction, mutlilevel model for change helps us to study within-person and between-person change. In this section we are going to discuss about two level MLM, but one characteristic of MLM is that it can be extended to more than two level, as long as we have sufficient data. Regarding our study we have two level MLM, level 1 and level 2. Level 1 describes the change of each individual with respect to time variant predictor(s) i.e. within person change which is known as *Intraindividual Change*. We can have more than one time variant predictor in level 1, we have used only one time variant predictor which is semester each individual is currently in. Level 2 describes the changes between individuals , which is known *Interindividual Change* . Level 2 outcome variables are the individual growth parameter we have in level 1 which is intercept and slope. We bring predictor to level 2 sub model depending on the hypothesis we are currently studying. We have random effects in level 1 and level 2 which allows each person to have trajectories around the population average value, which are the fixed effects. One of the assumptions of MLM is, the residuals at each level are normally distributed. One of the main reason to observe heterogeneity to occur in residuals is, if we have missed predictor(s) that was necessary for capturing the characteristics of our dataset.In rest of the section we will be discussing in detail of the models we have fitted to our dataset and in [sec:Models]Section 3.3.1 we have presented our results.

We have fitted 9 two level MLM on our synthetic data. We have followed the instruction of the book[2] to fit the models which is discussed in the seminar paper of *Adel Memariani*. Our outcome variable is *ECTS Cleared*, predictors for level 1 and level 2 are *Marital Status* and *Gender*. And with nesting of semesters nested in students. Our research questions on our synthetic data is:

- Does individual trajectories of clearing ECTS differs accordingly to their marital status and gender?

Due to our research interest, we have chosen predictor marital status and gender as our predictor in question, and we are going to explore the effect of having these two predictor in our model. We have first fitted unconditional means model and then unconditional growth model. These two models are our base models and we have compared the rest of the models with it. This comparison helps us to find out, if the new models that we are building for the dataset fits better or not, and if it is beneficial to explore more to find the "best" i.e. model which describes the characteristics of our dataset best.

3.3.1 Models

The unconditional means model is the simplest model that we can fit to the dataset. The main characteristic of this model is the fact that there is no predictor(s) involved in any of the levels.

Model A

$$Y_{ij} = \pi_{0i} + \epsilon_{ij} \tag{1}$$
$$\pi_{0i} = \gamma_{00} + \zeta_{0i} \tag{2}$$
$$Y_{ij} = \gamma_{00} + (\epsilon_{ij} + \zeta_{0i}) \tag{3}$$

In the above Model A, equation 1 is the level 1, equation 2 is level 2 and equation 3 is the composite model representation of both the level. In the equation above, 'i' represents each individual, ranging from 1 to 60, 'j' represent at which semester the individual is, it ranges from 1 to 3, Y_{ij} is our outcome variable which is ECTS cleared. π_{0i} represents each individuals rate of change, ϵ_{ij} represents the portion of the outcome variable that is unexplained at semester j for individuals i, which is known as the random effects. γ_{00} represents the population average for initial status. ζ_{0i} is the level 2 residuals (random effects) which represents the part of the initial status unexplained by level 2. The assumption for the level 1 and level 2 residuals is that they are independently and identically distributed with homoscedastic variance.

Our next model is unconditional growth model.

Model B

$$Y_{ij} = \pi_{0i} + \pi_{1i} Semester_{ij} + \epsilon_{ij} \tag{4}$$
$$\pi_{0i} = \gamma_{00} + \zeta_{0i} \tag{5}$$
$$\pi_{1i} = \gamma_{10} + \zeta_{1i} \tag{6}$$

$$Y_{ij} = \gamma_{00} + \gamma_{10} Semester_{ij} + (\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} Semester_{ij}) \tag{7}$$

The main difference between Model A and Model B , is the presence of time variant predictor in level 1. We have two outcome variable in level 2, equation 5 and 6. These are the individual growth parameters from level 1. $Semester_{ij}$ is our time variant predictor, and π_{1i} is the newly defined slope parameter for each individual i. γ_{00} and γ_{00} are the level 2 intercepts which represent the average initial status and rate of change respectively, which are the fixed effects. ζ_{0i} and ζ_{0i} are the level 2 residuals and we assume that they are independently drawn from a bivariate normal distribution with mean 0, variances σ_0^2 σ_1^2 and covariance σ_{01}.

After fitting the above two models we have included our time invariant predictors at level 2 with different combinations. On our next model, Model C, we include gender on both the outcome variable of level 2.

Model C

$$Y_{ij} = \pi_{0i} + \pi_{1i} Semester_{ij} + \epsilon_{ij} \tag{8}$$
$$\pi_{0i} = \gamma_{00} + \gamma_{01} Gender_i + \zeta_{0i} \tag{9}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11} Gender_i + \zeta_{1i} \tag{10}$$

$$Y_{ij} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{10} Semester_{ij} + \gamma_{11} Gender_i * Semester_{ij} \tag{11}$$
$$+(\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} Semester_{ij}) \tag{12}$$

Our aim of building this model is to study if gender has effects on the individuals initial status which is ECTS cleared in the first semester and to study if gender has effects on the rate of change of the individuals, which means how Gender can effect the increase or decrease of clearing ECTS. We have two new fixed effects due to inclusion of gender as our predictor which are γ_{01} and γ_{11}, both of these fixed effects captures the effect of our predictor, gender, on the initial status and rate of change. Equation 11 is the composite model and here we can observe the interaction between level 2 predictor and level 1 predictor, which is captured by γ_{11}.

Our next model is Model D

Model D

$$Y_{ij} = \pi_{0i} + \pi_{1i} Semester_{ij} + \epsilon_{ij} \tag{13}$$
$$\pi_{0i} = \gamma_{00} + \gamma_{02} MaritalStatus_i + \zeta_{0i} \tag{14}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{12} MaritalStatus_i + \zeta_{1i} \tag{15}$$

$$\begin{split} Y_{ij} = \gamma_{00} + \gamma_{02} MaritalStatus_i + \gamma_{10} Semester_{ij} \\ + \gamma_{12} MaritalStatus_i * Semester_{ij} + (\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} Semester_{ij}) \end{split} \tag{16}$$

In this model we have removed gender as our level 2 predictor and replaced it with marital status. And the reason is, we wanted to study if marital status has effects on the initial status and rate of change of the individuals. Rest of the parameter were discussed previously.

Our next model is Model E

Model E

$$Y_{ij} = \pi_{0i} + \pi_{1i} Semester_{ij} + \epsilon_{ij} \tag{17}$$
$$\pi_{0i} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{02} MaritalStatus_i + \zeta_{0i} \tag{18}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11} Gender_i + \gamma_{12} MaritalStatus_i + \zeta_{1i} \tag{19}$$

$$\begin{split} Y_{ij} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{02} MaritalStatus_i + \gamma_{10} Semester_{ij} + \\ \gamma_{11} Gender_i * Semester_{ij} + \gamma_{12} MaritalStatus_i * Semester_{ij} \\ + (\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} Semester_{ij}) \end{split} \tag{20}$$

In this model, Model E, we have introduced both the predictors in level 2, keeping in mind that both gender and marital status might effect on the initial status and rate of change for each individual.

Model F

$$Y_{ij} = \pi_{0i} + \pi_{1i} Semester_{ij} + \epsilon_{ij} \tag{21}$$
$$\pi_{0i} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{02} MaritalStatus_i + \zeta_{0i} \tag{22}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{12} MaritalStatus_i + \zeta_{1i} \tag{23}$$

$$\begin{split} Y_{ij} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{02} MaritalStatus_i + \gamma_{10} Semester_{ij} \\ + \gamma_{12} MaritalStatus_i * Semester_{ij} \\ + (\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} Semester_{ij}) \end{split} \tag{24}$$

In this model, Model F, we can observe the omission of gender on the outcome of the slope in level 2 compared to Model E. With similar reasoning that gender might not have any effects on the rate of change of each individual

Model G

$$Y_{ij} = \pi_{0i} + \pi_{1i} Semester_{ij} + \epsilon_{ij} \quad (25)$$
$$\pi_{0i} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{02} MaritalStatus_i + \zeta_{0i} \quad (26)$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11} Gender_i + \zeta_{1i} \quad (27)$$

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{02} MaritalStatus_i \\ + \gamma_{10} Semester_{ij} + \gamma_{11} Gender_i * Semester_{ij} \\ + (\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} Semester_{ij}) \end{aligned} \quad (28)$$

In this model, Model G, we have replaced marital status with gender as a predictor for the outcome of the slope in level 2. Following similar reasoning that Marital Status might not have any effect on the rate of change of each individual.

Model H

$$Y_{ij} = \pi_{0i} + \pi_{1i} Semester_{ij} + \epsilon_{ij} \quad (29)$$
$$\pi_{0i} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{02} MaritalStatus_i + \gamma_{03} MaritalStatus_i * Gender_i + \zeta_{0i} \quad (30)$$
$$\pi_{1i} = \gamma_{10} + \zeta_{1i} \quad (31)$$

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{02} MaritalStatus_i \\ + \gamma_{03} Gender_i * MaritalStatus_i + \gamma_{10} Semester_{ij} \\ + (\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} Semester_{ij}) \end{aligned} \quad (32)$$

In this model, Model H, we have studied the interaction between marital status and gender on level 2 for the outcome of intercept and have removed all predictor for the rate of change outcome variable. We have done this interaction between our two predictor because we want to study if the intercept depends on the interaction of these two predictors.

Model I (Final Model)

$$Y_{ij} = \pi_{0i} + \pi_{1i} Semester_{ij} + \epsilon_{ij} \quad (33)$$
$$\pi_{0i} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{03} MaritalStatus_i * Gender_i + \zeta_{0i} \quad (34)$$
$$\pi_{1i} = \gamma_{10} + \gamma_{12} MaritalStatus_i + \zeta_{1i} \quad (35)$$

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{01} Gender_i + \gamma_{03} MaritalStatus_i * Gender_i + \gamma_{10} Semester_{ij} \\ + \gamma_{12} MaritalStatus_i * Semester_{ij} \\ + (\epsilon_{ij} + \zeta_{0i} + \zeta_{1i} Semester_{ij}) \end{aligned} \quad (36)$$

In this model, Model I, as we can see from the equation 32, 33 and 34 that we have removed and added predictors on level 2. For the outcome of intercept we have removed the effect of marital status and for the outcome of slope we have added marital status as a predictor. We have labelled this model as our "final" model and the result is discussed in the following section.

3.3.2 Results and Discussion

In this section we are going to discuss the result of fitting 2 level MLM on our synthetic data. We have done our analysis using the package "lme4" [3], and full maximum likelihood estimation (FML) method. There are two kind maximum likelihood estimation, full maximum likelihood and restricted maximum likelihood(RML). The difference between two type is when we use FML it describes the fit of the full model including fixed effects and random effects, whereas RML describes the fit of only random effects. Since we will be comparing models that varies both in the fixed and the variance component we have used FML. For comparing the models we have used Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) as our goodness-of-fit measure. We can use AIC and BIC score to compare all the models, and the model that has the lowest score fits better to our dataset. We have done T-tests using Satterthwaite's method on the fixed effects.

		Parameter	Model A	Model B	Model C	Model D
Fixed Effects						
Initial Status,						
π_{0i}	Intercept	γ_{00}	20.91***	20.19***	19.11***	20.42***
	Gender	γ_{01}			2.17	
	Marital_Status	γ_{02}				-.45
	Marital_Status*Gender	γ_{03}				
Rate of change,						
π_{1i}	Intercept	γ_{00}		0.71	0.47	1.60**
	Gender	γ_{11}			0.48	
	Marital_Status	γ_{12}				-1.78*

Table 2: Results of fitting 2 level MLM for change to our synthetic data. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. T-tests use Satterthwaite's method

		Parameter	Model E	Model F	Model G	Model H	Model I
Fixed Effects							
Initial Status,							
π_{0i}	Intercept	γ_{00}	19.34***	19.06***	20.33***	24.47***	23.90***
	Gender	γ_{01}	2.17	2.71	2.17	-6.99 ***	-6.95***
	Marital_Status	γ_{02}	-.45	-.45	-2.44	-11.03***	
	Marital_Status*Gender	γ_{03}				18.92 ***	18.87 ***
Rate of change,							
π_{1i}	Intercept	γ_{00}	1.36*	1.6**	0.47	.708	1.6**
	Gender	γ_{11}	0.48		0.48		
	Marital_Status	γ_{12}	-1.78*	-1.78*			-1.78*

Table 3: Results of fitting 2 level MLM for change to our synthetic data. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. T-tests use Satterthwaite's method(Continued..)

		Parameter	Model A	Model B	Model C	Model D
Variance Component						
Level 1	Within person	σ_ϵ^2	15.34	13.93	13.93	13.93
Level 2	In initial status	σ_0^2	33.80	34.18	32.00	34.13
	In rate of change	σ_1^2		0.91	0.85	0.12
	Covariance	σ_{01}		-0.07	-0.13	-0.31
Goodness-of-fit	AIC		1130.1	1131.8	1132.7	1127.5
	BIC		1139.7	1151.0	1158.2	1153.0

Table 4: Results of fitting 2 level MLM for change to our synthetic data. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. T-tests use Satterthwaite's method(Continued..)

		Parameter	Model E	Model F	Model G	Model H	Model I
Variance Component							
Level 1	Within person	σ_ϵ^2	13.93	13.93	13.93	13.91	13.46
Level 2	In initial status	σ_0^2	35.97	33.04	33.94	6.64	6.99
	In rate of change	σ_1^2	0.06	0.12	0.85	0.931	0.58
	Covariance	σ_{01}	-0.64	-0.48	-0.33	0.81	1.00
Goodness-of-fit	AIC		1128.0	1126.5	1132.5	1081.1	1076.8
	BIC		1159.9	1155.2	1161.2	1109.8	1108.7

Table 5: Results of fitting 2 level MLM for change to our synthetic data. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. T-tests use Satterthwaite's method(Continued..)

Table 2, 3, 4 and 5 shows the result of all the models after fitting 2 level MLM. We have already discussed our reasoning for fitting Model A to Model I in section **??**. In the following we are going to discuss our final model and the reason why we think that Model I is the "best" model compared to other models.

Model I captures the characteristics of our dataset "best" compared to other models. The fixed effects associated with all the predictor we have used are statistically significant. Which implies that the effects of the predictor are being captured by Model I. The goodness-of-fit measure, AIC and BIC score (as discussed in the beginning of this subsection) is the lowest for Model I, indicating that Model I fits better to our dataset compared to other models. Also from the results we can observe that the covariance parameter for Model I is 1. Which indicates that there is very high correlation between the intercept and slope. Which in our case means the following. The number of ECTS cleared in the last semester (third semester in our dataset) is correlated with number of ECTS cleared in the first semester. An individual who will clears more ECTS in the first semester will clear more ECTS in their last semester. And an individual who will clear less ECTS in their first semester will clear less ECTS in their last semester.

For further testing of Model I, to confirm that Model I is the best fit model to our dataset we have done pairwise model comparison using ANOVA. We have compared all the models, Model A to Model H, with Model I. The table 6 below shows the results.

	Df	AIC	BIC	Chisq	Df	Pr(>Chisq)
Models						
Model A	3	1130.1	1139.7			
Model I	10	1076.8	1108.7	67.321	7	5.131e-12 ***
Model B	6	1131.8	1151.0		4	
Model I	10	1076.8	1108.7	63.055		6.606e-13 ***
Model C	8	1132.7	1158.2			
Model I	10	1076.8	1108.7	59.869	2	9.99e-14 ***
Model D	8	1127.5	1153.0			
Model I	10	1076.8	1108.7	54.711	2	1.317e-12 ***
Model E	10	1128.0	1160.0			<2.2e-16 ***
Model I	10	1076.8	1108.7	51.234	0	
Model F	9	1126.5	1155.2			
Model I	10	1076.8	1108.7	51.731	1	6.364e-13 ***
Model G	9	1132.5	1161.2			
Model I	10	1076.8	1108.7	57.669	1	3.102e-14 ***
Model H	9	1081.	1109.8			
Model I	10	1076.8	1108.7	6.31	1	0.01201 *

Table 6: Pairwise model comparison using ANOVA (Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1)

A “good model” will fit our data better compared to other model. We can compare two models using ANOVA and results of ANOVA guides us if the new model that we have build fits better to our dataset compared to the other model. We have compared Model A with Model I, Model B with Model I and so on.

From the result above we can clearly see that the p-value of Model I is always significantly better than the other models. Even though the degree of freedom of Model I is higher than all the models, which means it is more complex than others, but still Model I is significantly better. And we an conclude from this that Model I fits better to our dataset compared to all other models.

3.3.3 Testing The Assumptions And Results

When we fit a MLM we include some assumptions and if the assumptions are not holding then the results will not be interpretable. For this reason we have tested the three assumptions, which are - linearity assumption, normality assumption and homoscedasticity assumption. Below we have discussed the three assumptions and later in this subsection we have presented the result of our test.

- Linearity assumption: The change trajectories of each individual is linear and the relationship between individual growth parameter and the predictors are linear. To test this assumption we have to examine outcome variable against the predictors for each level. Since we have 2 level MLM, for level 1 we have to examine outcome against the predictor and for level 2, we have to examine the outcomes against all the predictors. For level 1 we have assumed it to be linear. And, since we have binary value predictor in level 2, we can assume that linearity assumption is met for level 2 without doing further test since, a linear model is acceptable for binary predictors.

- Normality assumption: The residuals in all the levels are normally distributed. To test this assumption we have to study the behavior of the residuals using exploratory analysis. For all the residuals, plot a normal probability plot against the residuals with respect to their normal score. If the distribution of the residuals are normal then the points will make a line.

For testing if normality assumption is fulfilled for our dataset we have done the following study. We have extracted the residuals from level 1 and level 2 and plotted them against their normal score.

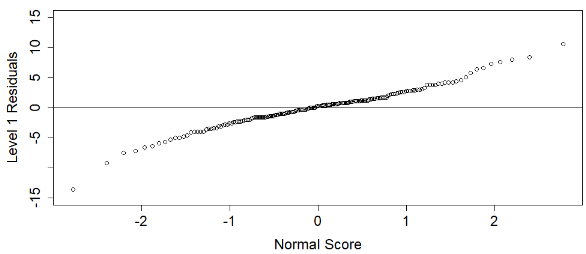

Figure 7: Examining normality assumptions in MLM for change. Normal probability plot for the raw residuals at level 1

Residuals of level 1 exhibiting long tails. It is showing reasonable linear pattern in the center of the data however the lower tail shows deviation from straight line. We can conclude the residuals are normally distributed in the center of the data but not perfectly normally distributed due to lower and upper tail showing deviation from straight line.

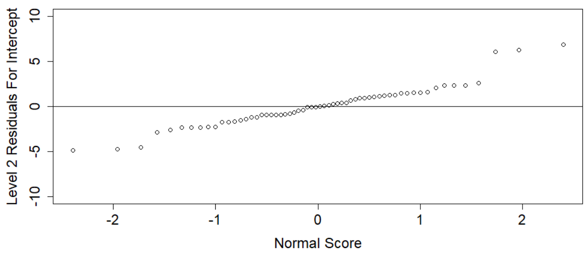

Figure 8: Examining normality assumptions in MLM for change. Normal probability plot for the raw residuals at level 2

Residuals of level 2 for intercept exhibiting long tails. It is showing reasonable linear pattern in the center of the data however the lower and upper tail shows significant deviation from straight line. We can conclude the residuals are normally distributed in the center of the data but not perfectly normally distributed due to lower and upper tail(it deviates more than lower tale) showing deviation from straight line.

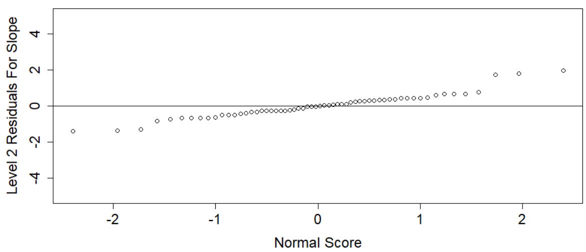

Figure 9: Examining normality assumptions in MLM for change. Normal probability plot for the raw residuals at level 2

Residuals of level 2 for slope exhibiting long tails. It is showing reasonable linear pattern in the center of the data however the lower and upper tail shows significant deviation from straight line and lower tale showing slight deviation from straight line. We can conclude the residuals are normally distributed in the center of the data but not perfectly normally distributed due to lower and upper tail(it deviates more than lower tale)showing deviation from straight line.

We can conclude from our study of normality assumptions that the residuals are normally distributed

- Homoscedasticity assumption: Variance of the residuals is equal at each level for every predictor. Since we have 2 level MLM to test this assumption we have to plot raw residuals against the predictors: level 1 residuals with respect to level 1 predictor, level 2 residuals with respect to level 2 predictor(s). If the assumptions is fulfilled then we will observe residual variability will be close for all the predictors value.

For testing if homoscedasticity assumption is fulfilled for our dataset we have done the following study. We have extracted the residuals from level 1 and level 2 (for both intercept and slope) and plotted them against the predictors. Level 1 residuals against level 1 predictor which is semester, level 2 residuals of intercept against gender and marital status and level residuals of slope against gender and marital status.

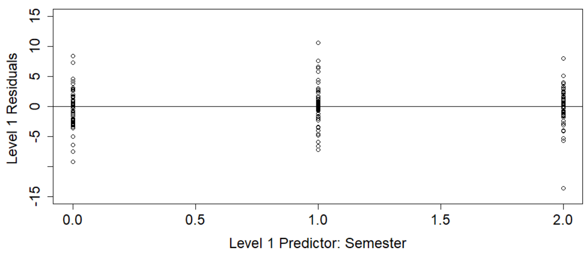

Figure 10: Examining homoscedasticity assumptions in MLM for change. Level 1 residuals vs level 1 predictor Semester.

The level 1 residuals shows approximately equal range and variability in all 3 semesters.

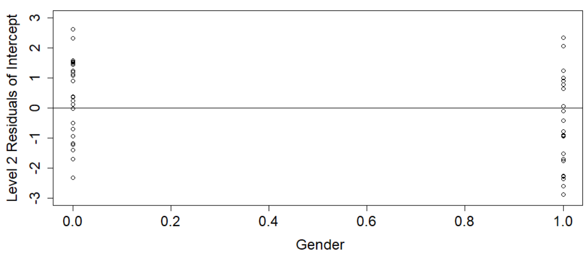

Figure 11: Examining homoscedasticity assumptions in MLM for change. Level 2 residuals of intercept vs level 2 predictor Gender.

The level 2 residuals of intercept shows approximately equal range and variability for male and female individuals.

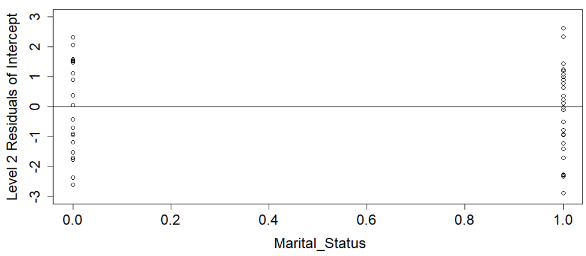

Figure 12: Examining homoscedasticity assumptions in MLM for change. Level 2 residuals of intercept vs level 2 predictor Marital Status.

The level 2 residuals of intercept shows approximately equal range and variability for married and unmarried individuals.

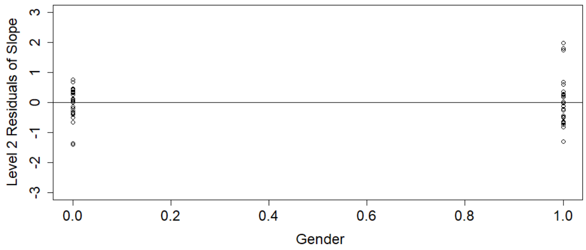

Figure 13: Examining homoscedasticity assumptions in MLM for change. Level 2 residuals of slope vs level 2 predictor Gender.

The level 2 residuals of slope shows unequal range and unequal variability for male and female individuals.

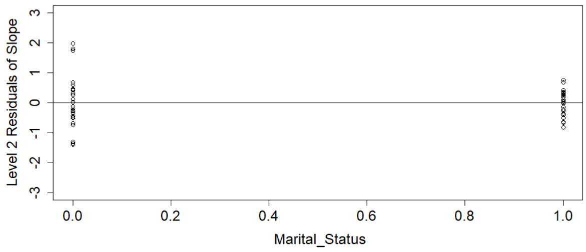

Figure 14: Examining homoscedasticity assumptions in MLM for change. Level 2 residuals of slope vs level 2 predictor Marital Status.

The level 2 residuals of slope shows unequal range and and unequal variability for married and unmarried individuals.

We conclude from our study of homoscedasticity assumption is partially fulfilled.

References

[1] Thomas Probst, Rüdiger C. Pryss, Berthold Langguth, Josef P. Rauschecker, Johannes Schobel, Manfred Reichert, Myra Spiliopoulou, Winfried Schlee, and Johannes Zimmermann. Does tinnitus depend on time-of-day? An ecological momentary assessment study with the "TrackYourTinnitus" application. *Frontiers in Aging Neuroscience*, 9(AUG):1–9, 2017.

[2] Judith D Singer, John B Willett, John B Willett, et al. *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press, 2003.

[3] Douglas Bates, Deepayan Sarkar, Maintainer Douglas Bates, and L Matrix. The lme4 package. *R package version*, 2(1):74, 2007.

Appendices

A R codes to fit a multilevel model with lme4 package

Before using lme4 package, we need to install it at first. The following command will install the lme4 package in R-studio:

```
install.packages("lme4")
```

In order to use the functionalities in the lme4 package we need to load it into the memory with the below command:

```
library(lme4)
```

The R command to show a summary of the fitted model (random effects, fixed effects and some information about the good-ness of fit i.e. AIC, BIC):

```
summary(Model-Name)
```