IMPLEMENTATION OF DIFFERENT ALGORITHMS IN LINEAR MIXED MODELS: CASE STUDIES WITH TIMSS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BURCU KOCA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

SEPTEMBER 2021

Approval of the thesis:

IMPLEMENTATION OF DIFFERENT ALGORITHMS IN LINEAR MIXED MODELS: CASE STUDIES WITH TIMSS

submitted by **BURCU KOCA** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Statistics, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Özlem İlk Dağ
Head of the Department, **Statistics** _____

Assist. Prof. Dr. Fulya Gökalp Yavuz
Supervisor, **Statistics, METU** _____

Examining Committee Members:

Prof. Dr. Olçay Arslan
Statistics, Ankara University _____

Assist. Prof. Dr. Fulya Gökalp Yavuz
Statistics, METU _____

Prof. Dr. Özlem İlk Dağ
Statistics, METU _____

Date: 06.09.2021

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Burcu Koca

Signature :

ABSTRACT

IMPLEMENTATION OF DIFFERENT ALGORITHMS IN LINEAR MIXED MODELS: CASE STUDIES WITH TIMSS

Koca, Burcu
Master of Science, Statistics
Supervisor: Assist. Prof. Dr. Fulya Gökalp Yavuz

September 2021, 82 pages

Mixed models are frequently used in longitudinal data types with time repetition over the same subject and clustered data types formed by observations gathered around certain groups. The modeling technique which models the dependency structure between repetitions and observations in the same cluster is required to use algorithms for parameter estimations. The same model can be solved with various algorithms arising from setup, inference and approach differences. In this study, several algorithms used for LMM, their development process and depending on what differences and similarities they can be resolved are explained with a data set related to an area that can contribute to society. In this sense, one of the sciences in which mixed models find application is education. The Trends in International Mathematics and Science Study (TIMSS) collects the most comprehensive and reliable information in the field of education internationally, and it is carried out every four years in 70 countries. With this data, several tests are prepared and applied to measure the success of students in science and mathematics in different countries, and demographic information about school, teacher, family and student is systematically collected with questionnaires that measure students' perspectives on lessons or parents' perspectives on schools. These results, beyond being a guide for

policy makers, can also guide the steps that countries will take in these areas. In this study where a multi-layered approach is preferred, the variables that are effective in students' mathematics achievement are determined as the student's gender, birth status in Turkey, emotional thinking, mathematical tendency, socioeconomic status, and family's thoughts about school. In addition, while many parameters give the same value in the algorithm comparison results; the fast algorithm is faster than the ecme algorithm. In terms of model setup, while lme and lmer functions are easy to implement and similar to each other; there are some differences in ecmeml, fastml and fastmcmc algorithms. The analysis is implemented solely with Turkey and then with England and South Africa for comparisons. While the same variables are statistically significant for all countries, LMM proves the superiority of England over others in math score when all situations are constant.

Keywords: EM Algorithm, Linear Mixed Model, Math Achievement, TIMSS

ÖZ

DOĞRUSAL KARMA MODELLERDE FARKLI ALGORİTMALARIN UYGULANMASI: TIMSS İLE ÖRNEK OLAYLAR

Koca, Burcu
Yüksek Lisans, İstatistik
Tez Yöneticisi: Dr. Öğretim Üyesi Fulya Gökalp Yavuz

Eylül 2021, 82 sayfa

Karma modeller zaman tekrarı bulunan boylamsal veri tiplerinde ve belirli gruplar etrafında toplanmış gözlemlerin oluşturduğu kümelenmiş veri tiplerinde sıklıkla kullanılmaktadır. Tekrarlar ve aynı küme içindeki gözlemler arasındaki bağımlılık yapısını modelleyebilen bu modelleme tekniği; parametre çıkarımları için algoritmalardan faydalanmayı gerekli kılmaktadır. Aynı model, kurulum, çıkarım ve yaklaşım farklılıklarından doğan çeşitli algoritmalar ile çözümlenebilmektedir. Bu çalışmada, LMM için kullanılan farklı algoritmaların neler olduğu ve nasıl bir gelişim süreci ile hangi farklılık ve benzerliklere göre çözümlenebildikleri anlatılmaktadır. Sözü edilen algoritmalar karşılaştırılırken toplumsal katkı sağlayabilecek bir alana ilişkin veri seti tercih edilmiştir. Bu anlamda, karma modellerin uygulama alanı bulacağı bilimlerden birisi eğitimdir. Eğitim alanında en kapsamlı ve güvenilir bilgiyi toplayan uluslararası bir çalışma olan Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS), her dört yılda bir 70 ülkede gerçekleştirilmektedir. Bu veri ile farklı ülkelerdeki öğrencilerin, bilim ve matematik alanında başarılarının ölçüldüğü testler hazırlanıp uygulanmakta ve bununla birlikte öğrencilerin derslere bakış açısını veya velilerin okullara bakış açısını ölçen anketler ile okul, öğretmen, aile ve öğrenciye ilişkin demografik bilgiler sistematik olarak

toplanmaktadır. Bu sonuçlar politika yapıcılara yol gösterici olma niteliği taşımanın ötesinde ülkelerin bu alanlarda atacakları adımlara da rehberlik edebilmektedir. Çok katmanlı bir yaklaşımın tercih edildiği bu çalışmada, öğrencilerin matematik başarısında etkin olan değişkenler öğrencinin cinsiyeti, Türkiye'de doğma durumu, duygusal düşüncesi, matematiksel eğilimi, sosyoekonomik statüsü ve ailenin okul hakkındaki düşünceleri olarak belirlenmiştir. Ayrıca algoritma karşılaştırma sonuçlarında bir çok parametre aynı değeri verirken; fast algoritması ecme algoritmasından daha hızlı çalışmaktadır. Model kurulumu açısından, lme ve lmer fonksiyonları kolay uygulanabilir olup birbirine benzerken; ecmeml, fastml ve fastmcmc algoritmalarında bir takım farklılıklar oluşmuştur. Analizler öncelikle Türkiye için gerçekleştirilmiş, sonrasında İngiltere ve Güney Afrika'yı kapsayan ülke karşılaştırmaları ile devam ettirilmiştir. Türkiye için başarıyı etkileyen faktörler ile üç ülkenin de birlikte incelendiği analizlerde etkin çıkan faktörler önemli ölçüde benzerken; diğer tüm faktörler sabit olduğunda, LMM İngiltere'nin diğer iki ülkeye göre başarısını desteklemektedir.

Anahtar Kelimeler: EM Algoritması, Doğrusal Karma Model, Matematik Başarısı, TIMSS

TO MY FAMILY…

ACKNOWLEDGMENTS

Firstly, I would like to express my deepest gratitude to my supervisor Assist. Prof. Fulya Gokalp Yavuz for their guidance, advice, criticism, encouragements and insight throughout the research. I hope we have the opportunity to work together for many more years.

I would also like to express my deepest gratitude to my thesis committee members Prof. Olcay Arslan and Prof. Ozlem Ilk Dag for their valuable guidance and contributions.

I would also like to offer my special thanks to my friend Emir Guler for his unwavering support and encouragement. I also thank to my friend Irem Akkaya and my colleagues Sevilay Dogan, Serenay Cakar and Hakkı Erduran for their valuable support.

Finally, I would like to thank my family members, Mehmet Uner Koca, Hasibe Koca and Onur Koca, for always encouraging me. I am lucky to have you.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

LMM: Linear Mixed Model

MLE: Maximum Likelihood Estimation

RMLE: Restricted Maximum Likelihood Estimation

EM: Expectation-Maximization

ECM: Expectation-Conditional Maximization

ECME: Expectation-Conditional Maximization Either

NR: Newton-Raphson

FS: Fisher Scoring

MH: Metropolis-Hasting

MCMC: Markov Chain Monte Carlo

TIMSS: Trends in International Mathematics and Science Study

CHAPTER 1

INTRODUCTION

Linear mixed models (LMM) expose the relationships and the characteristics between a response variable and explanatory variables, including the dependency structure in repeated, hierarchical or clustered data sets. Parameter estimations in LMM need some iterative algorithms such as Expectation-Maximization (EM) algorithm or Newton-Raphson algorithm, since parameter estimations have no closed form. Moreover, each parameter estimation includes some other unknown parameters. When the algorithm reaches convergency, parameter estimations are completed. This study shows implementation of appropriate algorithms for LMM.

Data structures in which dependencies exist between repeated measurements or between the subjects within a cluster are encountered in many areas. The most beneficial to society of these areas is education, which is now almost a necessity to cover science, technology, engineering and mathematics (STEM), and its continuous development is very crucial for national development (Langdon et al. 2011; Liou and Bulut, 2020). In the field of education, the Trends in International Mathematics and Science Study (TIMSS) is the most comprehensive international study that has been carried out for a long time and in a large-scale, allowing country comparisons, too. As it is known, like TIMSS, the Programme of International Student Achievement (PISA) is also a very comprehensive test that provides international comparison and also directs education policies (OECD, 2010). However, PISA is a test based on measuring how well learned skills can be applied to real life problems, while TIMSS measures how well conceptual skills are learned at school (Gronmo and Olsen, 2006). Since the latter has a more direct approach to measuring the achievement, TIMSS is chosen to identify the factors affecting success in this area.

Several studies related to the TIMSS concentrate on the measurement theory or validation of the test scores (e.g., Haladyna and Downing, 2004; Liou and Bulut, 2020). Since TIMSS results are important for national and international studies, academic researches, the development and regulation of education policies (Schmidt and Burroughs, 2016; Carnoy, Khavenson and Ivanova, 2015), it is essential to understand the result of this large-scale study and to understand the factors affecting the final scores.

LMM finds a place in several educational studies (for examples, see Mullis & Martin et al., 2012; Wiberg, 2019). Since the students are nested into the schools in TIMSS, the data structure is suitable to be analyzed with the mixed model.

The aim of this study is to examine different algorithm types in LMM that works for the nested structures and to test the performance of these analyzes on different models with a strong data set. And, by doing so, we try to understand what affects the achievement of students. For this aim, TIMSS 2019 database developed by the International Association for the Evaluation of Educational Achievement (IEA) from Turkey, England and South Africa national samples are used to understand the cognitive mathematics performances of students with demographic and affective (attitude, perception, self-confidence, motivation, etc.) latent traits. 4028, 3396, and 11891 number of students participated from Turkey, England and South Africa, respectively. The committees have some restrictions (TIMSS Technical Report, 2019) while working with students. For example, for fourth grade students, the mean age of students should be at least 9.5 at the time of test. However, some countries have different curriculum for fourth grade students. In this case, these countries participate with fifth or sixth grade students for fourth grade's test. In order not to make wrong inference, these countries are kept separate from other countries in the success comparisons. Turkey, England and South Africa are the examples of such countries.

While looking at the causes of success with a nested modeling technique, different types of algorithms of LMM are used to compare the results. Since the first

inferences proposed for LMM (Eisenhart,1947; Hendersen, 1950), several approaches have been developed for the implementation of it (e.g. Laird and Ware, 1982; Laird et al., 1987; Lindstrom and Bates, 1988). Although these different approaches have been used separately in many studies, they have not been examined together on a comprehensive and reliable data set like TIMSS. Since the inferences of fixed effects are more robust in these methods, it is anticipated that the main difference will arise from random effects. Although large differences are not expected, the results of the algorithms will be presented comprehensively to explain which algorithm will be more effective in which situations or the differences in the use and modeling of algorithms.

We use multiple countries with different achievement scores for jurisdiction. Besides, each country is from a different type of development scale. While England is a fully developed country, Turkey and South Africa are developing countries. Because each Turkish school is represented by one class in the TIMSS 2019 sample design, achievement gaps between schools could not be differentiated from achievement differences between classes from the same schools. As a result, in this study, schools and classes are referred to as the school/class level or unit.

Since this study uses the data from different cultures and family backgrounds, the understanding of a question may be different for each subject. We will assume that there is a measurement invariance which indicates that each individual from different countries has the same perception and interpretation of the items (Byrne and Watkins, 2003). In fact, the main goal is not to compare these cultures with each other. The main aim is to understand the factors that will affect this success by including different levels of success and to contribute a more generalized interpretation. By adding the country variable to the models as a covariate, the variability between these countries will be included in the model calculation.

CHAPTER 2

LITERATURE REVIEW

The first studies about linear mixed models (LMM) are implemented on animal breeding by Eisenhart (1947) and Hendersen (1950). LMM is applied over longitudinal and clustered data types. Clustered data includes a number of different data groups (Galbraith et al., 2010). Each group consists of multiple observations in nested or hierchical layers. Longitudinal data consists of repeated measurements of each unit and many longitudinal researches examine the changes in these characteristics over time or repeats (Laird and Ware, 1982). However, the measurements may have a significant variation due to the uncontrolled circumstances. Therefore, it is not suitable to perform general multivariate model analysis. Also, the general multivariate analysis may not be preferable or even inferences are not realizable for the unbalanced data sets (Laird and Ware, 1982). Laird and Ware (1982) indicate that two-stage models are more convenient for the longitudinal data analysis because it is not mandatory to provide balance in the data set. Additionally, explicit modelling and analysis of between- and within-individual variation are permitted in two-stage models. Laird and Ware (1982) provide with two-stage models known as LMMs with the EM Algorithm for parameter estimations.

Similarly, application of the EM Algorithm for repeated measures is discussed by Laird et al. (1987). They report that it is possible to use the EM Algorithm utilized with random-effect models for multivariate normal models with an arbitrary covariance structure and missing data. According to them, general formulation usage provides less messy applications (e.g., in terms of number of iterations) because the closed form solutions may exist for a much broader class of models.

Along with the EM Algorithm, parameter estimations may be implemented with other algorithms such as the Newton-Raphson Algorithm. For instance, Lindstrom and Bates (1988) implemented the Newton-Raphson and the EM Algorithm to LMM for repeated measures design. They provide with the formulation of derivatives for both maximum likelihood estimation (MLE) and restricted maximum likelihood estimation (RMLE). Then, they compare the two algorithms which are the Newton-Raphson and the EM Algorithms in the sense of convergency speed and overall performance. Lindstrom and Bates (1988) discuss that the EM algorithm is preferred over the NR algorithm because of the fast computation property. Also, EM guarantees to have parameter estimations in the parameter space and increases the likelihood at each step of the algorithm. However, they argue that NR is more amenable to handle the most of the extensions of the mixed models and the number of NR algorithm iterations is quite small than the number of EM iterations. Lindstrom and Bates (1988) finalize the study with discussing the extensions of the mixed effects model to integrate non-independent conditional error structure and the nested-type designs.

Expectation-Conditional Maximization Algorithm (ECM) which is a generalized EM is proposed by Meng and Rubin (1993). They attribute the popularity of the EM to its advantages of computations and convergency. However, in certain circumstances like computationally complicated MLE cases, these advantages cannot be applicable. Therefore, the estimation of complete data maximum likelihood is simpler with the conditional estimation of parameters. According to the proposed method (ECM), it is more beneficial to replace a complicated M-step of the EM algorithm with several CM-steps. They also support the ECM algorithm by showing all the properties that both the algorithms EM and ECM satisfy.

Another generalized EM algorithm is Expectation-Conditional Maximization Either (ECME) algorithm. Liu and Rubin (1994) present ECME by contributing to EM and ECM algorithms. According to the ECME, some or all steps of the ECM algorithm are replaced with the expectation of complete-data likelihood. Liu and Rubin (1994) report that ECME has two major advantages. The first one is having higher

maximization rate for some steps of the ECME. The second one is that ECME is faster than EM and ECM algorithms.

Jennrich and Schluchter published an article about unbalanced repeated measures models in 1982. The article mainly concentrates the point about the analysis of unbalanced or incomplete repeated measures data. Jennrich and Schluchter(1982) discuss the maximum likelihood analysis by using some iterative procedures. These procedures are Newton-Raphson algorithm, Fisher Scoring algorithm and Hybrid EM Scoring algorithm which is a generalized EM algorithm. Also, the article gives the examples of some certain covariance matrices structures to provide various ideas.

One of the recent studies about linear mixed-effects models and the EM algorithm is conducted by Gokalp Yavuz and Arslan (2018). Unlike other studies, Laplace distribution which is a member of the exponential power family is imposed into LMM to robustify the model. In this study, Gokalp Yavuz and Arslan (2018) use EM-type algorithm for parameter estimations with a robust distribution. Before this study, Pinheiro et al. (2001) also conduct a study related to LMM by imposing t-distribution into LMM and use an EM-type algorithm for the implementations. In both studies, the parameter estimations and predictions overcome the classical LMM approach.

Although there are some studies that apply LMM on the TIMSS data set, a study develops in line with this study has not been found in the literature. In the study of Ramirez (2006) the mathematical achievement of Chilean students is examined with a hierarchical modeling method and they compare three countries using TIMSS 1998/99 data. The main finding of this study is that the educational status of the family is very effective in the success of the students and students' mathematics achievement is higher, especially in advantageous regions, in schools that can determine their own curriculum. In the study of Liou and Bulut (2020), the science performance of eight-grade students is examined with TIMSS 2011 data according to the question types and domains. In addition to item difficulty analysis, they use cumulative link mixed modeling approach to find item format, cognitive domain impacts on the science scores. In the study of Carnoy et al. (2015), they try to

compare the TIMSS and PISA performance of Russian students from different level of family academic resources with the neighbors of the country in a cross-country comparison with descriptive statistics without any modeling approach.

We aim to decrease the separate modeling for each indicator to reduce the error arising from the modeling and so we combine and include several indicators in the model in a nested structure. Besides, it is aimed to use a technique that can model the dependency structure among students in the same school and that can consider the nested structure of the data. For these purposes, the 2019 TIMSS data is examined primarily in terms of explanatory variables, and the large number of variables are reduced by correlation and dimension reduction analysis. Finally, the obtained data structure is analyzed with several algorithm techniques of LMM.

The following chapter covers methodologies and technical background of LMM and iterative estimation algorithms. Chapter 4 introduces the TIMSS data set used for modelling. Chapter 5 includes descriptive statistics, data analysis with linear model, LMM types methods and algorithm implementations and results. Chapter 6 examines the achievement by considering the country effect. The last chapter which is numbered as 7 summarizes the study and give a conclusion about the modelling part.

CHAPTER 3

METHODOLOGY

3.1 Linear Mixed Models

Linear mixed model, introduced by Laird and Ware (1982), is an extended form of linear models with a random effect term. They are mainly applied for longitudinal and clustered data sets. The data has repeated measures over the same individual in logitudinal data. Datasets in health sciences can be shown as a suitable set of longitudinal data. Most of the time, doctor-patient example is another longitudinal data example. For example, assume that a group of patients routinely visit a clinic to check their blood values such as glicose level or hormone level. Assume that the clinic has 10 doctors and 10 patients from each doctor are randomly selected. In such cases, it is crucial to consider variation between patients and within patient groups for the same doctor. By doing so, we need some parameters to explain such kind of variability. On the other hand, a clustered data set is a data structure in which a group of individuals/observations can be considered together. As in the example of TIMSS 2019, each school represents a cluster; while students are the elements within the cluster. In this type of data, the dependency structure among the elements within a cluster should be considered while modeling the data.

In addition to the response variable, coefficients, parameters and error terms in a linear model, LMM has one extra parameter which is the random effect. That is why LMMs are also known as linear models with random effects. The general formula of LMM is shown below:

$$y_i = X_i\beta + Z_i u_i + \epsilon_i, \qquad i = 1, \ldots, n, \tag{3.1}$$

where y_i is $(n_i \times 1)$ vector of response, X_i is $(n_i \times p)$ design matrix for the fixed effects of the i-th subject, β is $(p\ x\ 1)$ unknown fixed effect vector, ϵ_i is $(n_i \times 1)$ vector of within-subject error for the i-th subject, Z_i is $(n_i \times q)$ design matrix for the random effects of the i-th subject, u_i is $(q\ x\ 1)$ unknown random effect vector.

The random effects are normally distributed as:

$$u_i \sim N(0, D), \tag{3.2}$$

where D is the positive-definite covariance matrix. The error terms are normally distributed as:

$$\epsilon_i \sim N(0, R_i), \tag{3.3}$$

where R_i is the covariance matrix. Also, y_i is normally distributed as:

$$y_i \sim N(X_i\beta, Z_i D Z_i' + R_i). \tag{3.4}$$

3.2 Algorithms

To make an inference about the data of interest, it is necessary to know the model coefficients. There are several parameter estimation techniques in statistical modeling methods. For example, estimation techniques such as MLE, least square estimation and method of moment estimation are commonly used. These methods are mainly applied for classical regression techniques. However, for LMM, those are not suitable due to the complexity of LMM. Firstly, LMM has no closed form solutions. Hence, the derivation process for parameter estimations is unique for each parameter. Secondly, parameter estimation in LMM needs iterative estimation algorithms such as the Expectation-Maximization algorithm or the Newton-Raphson algorithm since each parameter estimation includes some other unknown parameters. Therefore, iterative algorithms assign some initial values and run all the process until the convergence occurs. In the following subsections, the corresponding iterative algorithms are given in detail.

3.2.1 Expectation-Maximization Algorithm

Expectation-Maximization (EM) algorithm, which is introduced by A.P. Dempster et al. (1977), is an iterative method for implementing maximum likelihood parameter estimations in incomplete datasets. The EM algorithm consists of two main steps. The first one is the expectation step that calculates the expected probability of the corresponding function. The second one is the maximization step that tries to maximize the parameter values from the first step. This cycle is repeated until convergency.

Let us define $L(\theta\,;\ X)$ as the log likelihood function of the unknown parameter θ:

$$L(\theta; X) = p(X|\theta) = \int p(X, Z|\theta)\, dZ \tag{3.5}$$

where X and Z represent the observed and the missing data, respectively. Dempster et al. (1977) describe the two steps of EM as follows:

Expectation Step (E-Step): The E-step of the EM algorithm aims to estimate the expectation of $L(\theta; X)$.

$$Q(\,\theta \mid \theta^t\,) = E_{Z\,|X,\theta^t}[\log \mathrm{L}(\theta; X, Z)] \tag{3.6}$$

Maximization Step (M-Step): The M-Step of the EM algorithm aims to compute unknown parameter θ by using outputs from E-step.

$$\theta^{t+1} = argmax_\theta\ Q(\theta|\theta^t) \tag{3.7}$$

For example, X is a set of observable measures such as $X = (x_1, x_2, \dots, x_n)$ while Z is a set of unobservable measures such as $Z = (z_1, z_2, \dots, z_n)$. Also, Y denotes a joint set of X and Z.

In LMM, the set Z represents unknown random effects. In E-Step of the EM algorithm, the expected value of these random effects cannot be computed. Therefore, its conditional expectations are estimated with respect to the unknown parameters such as β_i, D and R_i. Since the log-likelihood function is not solved

easily, the conditional expectation is used instead of individual statistics. In E-Step the expectation of conditional case is computed. In M-Step the parameters are updated by maximizing the function.

Let θ= (β_i, D, R_i) be a set of parameters.

In the E-Step of the EM Algorithm,

$$E[(u_i \mid Y_i; \theta^t)] \; and \; E(u_i u_i' \mid y; \theta^t),$$

are computed.

M-Step of the EM Algorithm:

$$\beta^{(t+1)} = (X'X)X'\left(y - ZE(u|y; \theta^t)\right) \tag{3.8}$$

$$D^{(t+1)} = \frac{1}{N}\sum_{i=0}^{N} E(u_i u_i'|y; \theta^t) \tag{3.9}$$

$$R^{(t+1)^2} = \frac{1}{n}\left(\left(\left||y - X\beta^{(t+1)}\right|\right)^2 + \sum_{i=0}^{N} Tr\left(Z_i' Z_i E(u_i u_i' \mid y; \theta^t)\right) - 2\sum_{0}^{N}\left(y_i - X_i\beta^{(t+1)}\right)' Z_i E(u|y; \theta^t)\right) \tag{3.10}$$

θ^{t+1}= (β_i, D, R_i) is estimated until the convergency of $L(\theta^{t+1} \mid X) - L(\theta^t \mid X)$ occurs (Meng & Rubin, 1993).

E-Step and M-Step are iterative. The loop continues until θ reaches convergency.

The EM algorithm has become prominent thanks to its certain features such as adaptation, convergency and being able to handle with missing data. Wu (1983) mentions that EM algorithm is used easily for many problems because it has a nice form of the complete-data likelihood function. Moreover, Dempster et al. (1981)

report that EM method guarantees the increase in the likelihood function in each step of the iteration.

On the other hand, it is important to keep in mind that the EM algorithm is so sensitive to initial points. Wu (1983) indicates that if the log-likelihood function has more than one maximum global or maximum local points, the convergence of the algorithm depends on the selected starting values. Therefore, it is advisable to run the EM algorithm several times at different starting points and observe the result.

3.2.2 Expectation-Conditional Maximization Algorithm

Expectation-Conditional Maximization which is known as ECM in short is introduced by Meng and Rubin (1993). The EM algorithm is popular due to its simplicity in implementations and stability in convergency. However, in the case of complicated MLEs, the EM may lose its power due to the unattractive structure of the M-Step. According to them, if the data has missing observations the ECM algorithm is more suitable than the EM algorithm.

In case of random effects (u_i) are known, the least square estimation technique could be used for the parameter estimation in LMM. However, due to unknown random effects, extended estimation techniques should be used, instead. In this section, the ECM algorithm which is a Generalized Expectation-Maximization algorithm is used for LMM. While E-step of the algorithm is the same with E-step of EM, maximization step (M-step) is a little bit different. M-step is replaced with several CM-steps. In some cases, computing conditional states of the parameters by ECM is relatively simple because it is more complicated to estimate with EM algorithm. CM step of ECM is completed when the convergency is satisfied. Convergency properties like increasing the likelihood of ECM are the same as the EM.

Assume that θ is a set of parameters and S is a total number of cycles occurred in CM-steps. s^{th} cycle of the iteration t+1, E-step computes the following:

$$Q(\theta|\theta^{\{t+(s-1)/S\}}) = \int L(\theta|Y) f(Y_{mis}|Y_{obs}, \theta = \theta^{\{t+(s-1)/S\}}) dY_{mis} \quad (3.11)$$

Moreover, CM-step of ECM computes $\theta^{(t+s/S)}$in order to maximize the following:

$$Q(\theta^{(t+s/S)}|\theta^{\{t+(s-1)/S\}} \geq Q(\theta|\theta^{\{t+(s-1)/S\}}) \quad (3.12)$$

for all $\theta \epsilon \Theta_s(\theta^{\{t+(s-1)/S\}}$.

The parameter set θ_{t+1} is estimated until the convergency of

$$Q\left(\theta^{(t+1)} \mid \theta^{(t)}\right) \geq Q\left(\theta^{(t)} \mid \theta^{(t)}\right) \quad (3.13)$$

occurs (Meng and Rubin, 1993).

In LMM, random effects are assumed as missing values. Let $\theta = (\beta, D, \sigma^2)$ is a set of parameters and $u = (u_1', u_2', \ldots, u_k')' isavector$ of missing values. θ^t is the estimation of parameter θ at t[th] iteration. $Q\left(\theta|\hat{\theta}^{(t)}\right) = E_{u|y,\theta=\theta^t}$ represents the conditional expectation of u given observation y at iteration t. In E-step of ECM, the aforementioned conditional expectations which are necessary are computed. $\hat{b}_i^{(k)} = E(b_i|Y, \hat{\theta}^{(k)})$, $\hat{R}_{b_i}^{(k)} = E(b_i b_i^T|Y, \hat{\theta}^{(k)})$, $\hat{e}_i^{(k)} = E(e_i|Y, \hat{\theta}^{(k)})$ are such conditional expectations.

The number of CM steps of ECM belongs to the number of parameters. For example, we have 3 parameters in parameter space θ. Therefore, CM-steps are conducted for each parameter, separately. This process goes until the convergency satisfies (Wang and Fan, 2010).

3.2.3 Newton-Raphson Algorithm

Newton-Raphson Algorithm, also known as Newton's Method, which is introduced by Isaac-Newton and Joseph Raphson is an algorithm that enables to find the roots of a mathematical equation. Assume that there is a curve as $y = f(x)$. Firstly, the tangent line of the function at x_0 which means slope of the function is obtained. Here, x_0 is a random initial point which affects the convergency speed. Later, the point touches upon the x-axis is defined as new x_1. The loop continues until convergency occurs.

Due to the geometric explanation and the Taylor Series Expansion, the algorithm's formula has become as the following:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}. \tag{3.14}$$

And the Newton-Raphson General Formula is as the following:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{3.15}$$

The starting value for x_0 decides the convergency speed. Sometimes, the function may not reach convergency.

3.2.4 Newton-Raphson and Fisher Scoring Algorithm

The study of Jennrich and Schluchter (1986) focuses primarily on how incomplete or unbalanced repeated-measures data is analyzed. In doing so, Jennrich and Schluchter (1986) use Newton Raphson algorithm, Fisher algorithm and Hybrid EM Scoring algorithm to compute the maximum likelihood estimates.

In order to use given algorithms, it is needed to know the score vector, $\boldsymbol{s}$ and the Hessian matrix, $\boldsymbol{H}$. The s vector is defined as the derivatives of log-likelihood, λ of the given data and it is calculated as:

$$s = \begin{bmatrix} s_\beta \\ s_\theta \end{bmatrix} = \begin{bmatrix} \partial\lambda/\,\partial\beta \\ \partial\lambda/\,\partial\theta \end{bmatrix}. \tag{3.16}$$

On the basis of the s vector, Hessian matrix is computed as follows:

$$H = \begin{bmatrix} H_{\beta\beta} H_{\beta\theta} \\ H_{\theta\beta} H_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \partial^2\lambda/\partial\beta\partial\beta & \partial^2\lambda/\partial\beta\partial\theta \\ \partial^2\lambda/\partial\theta\partial\beta & \partial^2\lambda/\partial\theta\partial\theta \end{bmatrix}. \tag{3.17}$$

From that point, Newton Raphson and Fisher algorithms are taken into account. The new parameters are calculated by using the formula below:

$$\begin{bmatrix} \tilde{\beta} \\ \tilde{\theta} \end{bmatrix} = \begin{bmatrix} \beta \\ \theta \end{bmatrix} - \begin{bmatrix} H_{\beta\beta} H_{\beta\theta} \\ H_{\theta\beta} H_{\theta\theta} \end{bmatrix}^{-1} \begin{bmatrix} s_\beta \\ s_\theta \end{bmatrix}. \tag{3.18}$$

The algorithm suggests to replace the Hessian matrix by its expectation. Therefore, it equals

$$E[H_{\theta\theta}] = -\sum_{i=1}^{n} X_i' \, \Sigma_i^{-1} X_i, \tag{3.19}$$

Due to $E[H_{\beta\theta}]$=0, the new parameters β and θ are computed by solving different equations. Finally, the final parameter β can be computed from the following:

$$\tilde{\beta} = \left(\sum_{i=1}^{n} X' \Sigma^{-1} X \right) \left(\sum_{i=1}^{n} X' \Sigma^{-1} y_i \right), \tag{3.20}$$

Additionally, the final parameter θ can be obtained from the equation below until convergency is satisfied.

$$\tilde{\theta} = \theta + I_{\theta\theta}^{-1} s_\theta \tag{3.21}$$

3.2.5 Hybrid EM Scoring Algorithm

Jennrich and Schluchter (1986) define the Hybrid EM Scoring algorithm for incomplete data-oriented models. According to the Jennrich and Schlucter (1986), the algorithm has an advantage over the Newton Raphson and Scoring algorithms because of being able to fit covariance matrices with large number of parameters. In this algorithm, there are two main parts. In the first part, β is calculated using the equation below:

$$\tilde{\beta} = \left(\sum_{i=1}^{n} X'\Sigma^{-1} X\right)\left(\sum_{i=1}^{n} X'\Sigma^{-1} y_i\right). \tag{3.22}$$

At the beginning of the second part, e_i^* and R_i are estimated with the help of updated estimates β and the current estimates of θ, where

$$e_i^* = E(e_i^* \mid e_i) \tag{3.23}$$

and

$$R_i = cov(e_i^* \mid e_i). \tag{3.24}$$

At the middle of the second part, e_i^* and R_i are estimated by using the formula below:

$$S = \frac{1}{n}\sum_{i=1}^{n}(e_i^* e_i^{*\prime} + R_i). \tag{3.25}$$

At the final step of the second part, the updated θ is computed as:

$$\Delta\theta = I^{-1}s, \tag{3.26}$$

where

$$[s]_r = \frac{1}{2} tr \sum^{-1}(S - \Sigma)\sum^{-1}\Sigma, \tag{3.27}$$

and

$$[I]_{rs} = \frac{1}{2} tr\Sigma^{-1}\dot{\Sigma}_r\Sigma^{-1}\dot{\Sigma}_s \tag{3.28}$$

where

$$\dot{\Sigma}_{ir} = \partial\Sigma_i/\partial\theta_r \tag{3.29}$$

This process continues until convergency occurs (Jennrich & Schluchter, 1986).

3.2.6 ECME and FAST Algorithms

Schafer (1998) offers new improvements for the derivation of MLE which are easy to implement and their requirements are similar with the EMs. In the continuation of these derivations, three different algorithms are developed for LMM. The first application is the combination of the EM and Fisher Scoring algorithms. The second one is for correcting empirical Bayes interval estimation which is irrelevant for this study. The last one is Markov Chain Monte Carlo algorithm for Bayesian posterior simulation. Schafer (1998) uses SAS, S-PLUS, MLn and HLM for expressing the algorithms.

By applying such algorithms, it is assumed that each V_i which is error variance is identity matrix and the subunits in the matrix include exchangeable errors. Additionally, it is assumed that V_is are known.

Remind that, the likelihood function below needs maximization to estimate the parameters $\beta, \sigma^2 \; and \; \varepsilon$.

$$L_0(\beta, \sigma^2, \varepsilon) \propto (\sigma^2)^{-\frac{N}{2}} \prod_{i=1}^{m} |W_i|^{\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}(y_i - X_i\beta)^T W_i (y_i - X_i\beta)\right\}, \tag{3.30}$$

where $N = \sum_{i=1}^{m} n_i$, $W = (Z_i \varepsilon Z_i^T + V_i)^{-1}$ and $\varepsilon = \sigma^{-2}\varphi$.

Schafer (1998) proposes to add new function which is the inverse of ε. The additional function includes free parameters such as unknown covariance parameters and known symmetric matrices with dimension q × q. The error term seems like

$$\varepsilon^{-1} = \sum_{j=1}^{g} w_j G_j \tag{3.31}$$

where $w = (w_1, w_2, \ldots, w_g)^T$ represents a vector of unknown covariance matrices and $g = q(q+1)/2$.

Using some other relationships, logarithm of the L_0 is shown as the following equation:

$$l_0 = \frac{N}{2}\log\tau - \frac{m}{2}\log|\varepsilon| + \frac{1}{2}\sum_{i=1}^{m}\log|U_i| - \frac{\tau}{2}\sum_{i=1}^{m}(y_i - X_i\beta)^T W_i (y_i - X_i\beta) \tag{3.32}$$

The first and the second derivatives give the following functions below:

$$\frac{\partial^2 l_0}{\partial\beta\partial\tau} = -\Gamma^{-1}(\beta - \tilde{\beta}), \tag{3.33}$$

$$\frac{\partial^2 l_0}{\partial\beta\partial w_j} = -\sigma^{-2}\left(\sum_{i=1}^{m}\gamma_i^T U_i G_j U_i \gamma_i\right)(\beta - \tilde{\beta}). \tag{3.34}$$

These two functions are crucial for the algorithm, since FS algorithm needs the expectations of the second derivatives of the function y for fixed parameters. $\tilde{\beta}$ is an

unbiased estimator of β. Therefore, the expectations of them are equal to zero. Additionally, we need to know that $E(\hat{b}_i) = 0$ and $E(\hat{b}_i \hat{b}_i^T) = \sigma^2(\varepsilon - U_i)$.

The EM algorithm in LMM assumes random effects as missing data. Moreover, when ε is constant L_0 is proportional to the following expression:

$$(\sigma^2)^{-\frac{N}{2}} exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - X_i\beta)^T W_i(y_i - X_i\beta), \tag{3.35}$$

where W_i is fixed.

By using this function, it is possible to maximize L_0 in terms of (β, σ^2) while ε is fixed at its current value. This algorithm is known as ECME.

ECME formulas for MLE are explained by Schafer (1998). The current estimations are updated by using the followings:

$$U_i^{(t)} = \left(\varepsilon^{(t)^{-1}} + Z_i^T V_i^{-1} Z_i\right)^{-1}, \tag{3.36}$$

$$W_i^{(t)} = V^{-1} - V^{-1} Z_i U_i^{(t)} Z_i^T V_i^T, \tag{3.37}$$

$$\beta^{(t)} = \left(\sum_{i=1}^{m} X_i^T W_i^{(t)} X_i\right)^{-1} \left(\sum_{i=1}^{m} X_i^T W_i^{(t)} y_i\right), \tag{3.38}$$

$$\sigma^{2^{(t+1)}} = \frac{1}{N}\sum_{i=1}^{m}\left(y_i - X_i\beta^{(t)}\right)^T W_i^{(t)}\left(y_i - X_i\beta^{(t)}\right), \tag{3.39}$$

$$\hat{b}_i^{(t)} = U_i^{(t)} Z_i^T V_i^{-1}\left(y_i - X_i\beta^{(t)}\right), \tag{3.40}$$

$$\varepsilon^{(t+1)} = \frac{1}{m}\sum_{i=1}^{m}(\sigma^{-2^{(t)}}\hat{b}_i^{(t)}\hat{b}_i^{(t)T} + U_i^{(t)}. \tag{3.41}$$

The log-likelihood function of the algorithm at each iteration after equation 3.39 can be obtained with almost zero-additional cost (Schafer, 1998).

A type of NR algorithm is the FS algorithm. In NR, the loglikelihood function l is tried to maximize at $\theta = \theta^{(t)}$ by iteratively solving the equation:

$$C\theta^{(t+1)} = d, \tag{3.42}$$

where $C = -\partial^2 l/\partial\theta\partial\theta^T$, $d = C\theta^{(t)} + \partial l/\partial\theta$. While applying the process, there is no need to use the second derivatives because C is replaced as $C = -\partial^2 l/\partial\theta\partial\theta^T + R$ and the convergency properties are still satisfied. In new representation C, R means $o_p(n)$ where n is proportionally sample size. If C is equal to $-E(\partial^2 l/\partial\theta\partial\theta^T)$, then, the algorithm is called as FS. Using the expectation of the value rather than the exact value of the second derivatives, FS algorithm similarly converges with NR in terms of speed.

The algorithm FS computes $\beta^{(t)}$ similar to the ECME does. Therefore, it is said that the ECME already implements the scoring algorithm for β. Variance components which are obtained via scoring algorithm are closer to the value obtained from MLE rather than the ECME parts (Schafer, 1998). Unfortunately, scoring algorithm does not precisely guarantee to increase the loglikelihood function at each cycle. However, ECME does. Therefore, $\sigma^{2^{(t+1)}} = \sigma_{SCORE}^{2(t+1)}$ and $\varepsilon^{(t+1)} = \varepsilon_{SCORE}^{(t+1)}$ are set and the ECME estimations are stored.

So, the loglikelihood function becomes

$$l_0(\beta^{(t+1)}, \sigma^{2^{(t+1)}}, \varepsilon^{(t+1)}) \tag{3.43}$$
$$= -\frac{N}{2}\log \sigma^{2^{(t+1)}}$$
$$-\frac{m}{2}\log|\varepsilon^{(t+1)}| + \frac{1}{2}\sum_{i=1}^{m}\log|U_i^{(t+1)}| - \frac{N}{2}\left(\frac{\sigma_{ECME}^{2^{(t+2)}}}{\sigma^{2^{(t+1)}}}\right).$$

In case that $l_0$decreases, ECME estimates get involved to the process. For example, $\sigma^{2^{(t+1)}}$ is replaced with $\sigma_{ECME}^{2^{(t+1)}}$ and $\varepsilon^{(t+1)}$ is replaced with $\varepsilon_{ECME}^{(t+1)}$. Then, parameter estimation formulas are rerun. However, the solution of $C\eta = d$ may fall into outside of the parameter space. Moreover, if the matrix C does not seem positive definite, then, Schafer (1998) offers ignoring the scoring part for the related cycle and considering the ECME estimates.

Scoring algorithm is faster than other known EM algorithms. Scoring algorithm reaches convergency in 10-15 iterations, but the EM needs hundreds or thousands cycles for the same dataset. On the other hand, in the absence of large number of variance parameters, these new algorithms are 1.5 times higher than EM algorithm in terms of per-iteration cost (Schafer, 1998).

3.2.7 MCMC Algorithm

Markov Chain Monte Carlo (MCMC) algorithm is a Bayesian technique that simulates draws from the posterior distribution. In LMM, MCMC randomly generates a sample from the conditional posterior distributions for each unknown parameter. For example, assume that the parameter space is $(\beta, \sigma^2, \varphi, B)$ where β, σ^2, φ and B represent fixed effects parameters, variance of the error terms, variance of the random effects and random terms, respectively. The current parameters $(\beta^{(t)}, \sigma^{2^{(t)}}, \varphi^{(t)})$ and random effects $B^{(t)}$ are updated in turn by considering the following rules.

$$\sigma^{2(t+1)} \sim P\left(\sigma^2 \middle| y, \beta^{(t)}, \varphi^{(t)}, B^{(t)}\right), \tag{3.44}$$

$$\beta^{(t+1)} \sim P\left(\beta \middle| y, \sigma^{2(t+1)}, \varphi^{(t)}, B^{(t)}\right), \tag{3.45}$$

$$B^{(t+1)} \sim P\left(B \middle| y, \beta^{(t+1)}, \sigma^{2(t+1)}, \varphi^{(t)}\right), \tag{3.46}$$

$$\varphi^{(t+1)} \sim P\left(\varphi \middle| y, \beta^{(t+1)}, \sigma^{2(t+1)}, B^{(t+1)}\right). \tag{3.47}$$

The rules are determined by Gibbs sampling algorithm which is a subset of MCMC algorithm. The function of $\left(\beta^{(t)}, \sigma^{2(t)}, \varphi^{(t)}, B^{(t)}\right)$ reaches convergency and becomes $P(\beta, \sigma^2, \varphi, B|y)$ as $t \to \infty$. However, when n is large and random effects are badly estimated, Gibbs sampler which slowly converges becomes much slower. But, still, it is accepted that Gibbs sampler has easy implementation.

Schafer (1998) proposes new MCMC method which reaches convergency faster and runs the process using Restricted Maximum Likelihood Estimation (RMLE). The new technique covers Metropolis-Hasting (MH) algorithm that is one of the MCMC methods.

Assume that $\eta^{(t)}$ is a current version of variance parameters $\eta = \left(\tau, w_1, \dots, w_g\right)^T$ and a value $\eta^{\dagger}$ is drawn from the density function $h(\eta^{\dagger})$ by considering $P(\eta|y) \propto L_1(\eta)\pi(\eta)$, where $\pi(\eta)$ is the prior density and L_1 is the RML. Then, the acceptance ratio is calculated as

$$R^{(t)} = \frac{P(\eta^{\dagger}|y)}{P\left(\eta^{(t)}|y\right)} \frac{h\left(\eta^{(t)}\right)}{h(\eta^{\dagger})} \tag{3.48}$$

The rule of updated estimates is set as

$$\eta^{(t+1)} = \begin{cases} \eta^{\dagger} \; if \; u \leq R^{(t)}, \\ \eta^{(t)} \; if \; u > R^{(t)}, \end{cases} \tag{3.49}$$

where $u \sim U(0,1)$.

Running the process repeatedly creates a chain and its stationary distribution is $P(\eta|y)$. In case of good approximation to the $P(\eta|y)$ from h, convergency occurs rapidly. Schafer (1998) approximates $P(\eta|y)$ by using multivariate t-distribution. This is calculated by modifying FS algorithm for RMLE.

The distribution $P(\eta|y)$ may be evaluated by some iteration cycles of modified Gibbs sampler. In some cases, a value from MH algorithm may be rejected. Therefore, we compute Gibbs updated versions ${\sigma^2_{GIBBS}}^{(t+1)}$ and ${\varepsilon^2_{GIBBS}}^{(t+1)}$ given the current simulated values σ^2 and ε^2 . At the same time, ${\sigma^2_{MH}}^{(t+1)}$ and ${\varepsilon^2_{MH}}^{(t+1)}$ are taken into account by drawing a new candidate value. After all, $P(\eta^{(t+1)}|y)$ is drawn and $R^{(t)}$ is calculated. According to the calculation of $R^{(t)}$, continuation of the process is decided. If the process retains, then, we obtained ${\sigma^2_{GIBBS}}^{(t+2)}$ and ${\varepsilon^2_{GIBBS}}^{(t+2)}$ at the end of the cycle.

The combination of MH and Gibbs is more attractive and effective. The main advantage of these hybrid algorithms for LMM is that the loglikelihood function can be easily obtained. However, for LMM with non-normal errors, the algorithms may be more complicated because the likelihood function cannot be easily evaluated (Schafer, 1998).

CHAPTER 4

DATA: TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY (TIMSS)

4.1 Trends in International Mathematics and Science Study (TIMSS)

The Trends in International Mathematics and Science Study (TIMSS) is a comprehensive study that aims to evaluate the fourth and eighth grade students achievement in the area of Mathematics and Science (Fishbean et al., 2019). The survey is conducted approximately in 70 countries by International Association for the Evaluation of Educational Achievement every four years since 1995. TIMSS 2019 is the seventh assessment and it is open-source. The data-base provides information about students' home, teacher, school and national context beside students' achievement scores. By mixing all the information, TIMSS produces the comparable perspectives to the students' mathematics and science achievement scores.

4.2 TIMSS Dataset

TIMSS has four main datasets as student, home, teacher and school. The teacher data set is excluded from the study, since its missing values are highly prodominant for some countries of interest of this study. Student dataset is on a student basis and comprise personal information, socioeconomic measures for each student and tendency to mathematics and science of participants etc. Home dataset is also on a student basis and mainly covers early learning information of the students. The school data set is on a school basis. The data set includes school based information such as the number of books in the school and the thoughts of parents about the school. Each dataset contains the unique key column like student id or school id.

4.3 Turkey Achievement Example

In TIMSS 2019, Turkey participates with fifth grade and eight grade students to mathematics and science achievement surveys. Turkey ranked 23rd among 58 participating countries with an average score of 523 in mathematics achievement at the fourth-grade level (even Turkey works with fifth grades for TIMSS 2019, the term is used as fourth-grade for the ease of explanation). With this performance, Turkey is above the TIMSS scale midpoint (500 points) and outperforms many participating countries according to the 2019 results.

In this study, we try to understand the effective factors on mathematics achievement of fifth grade students in Turkey. To explain these factors, we use LMM, which takes into account the inter-school variability in addition to the several coefficients related to schools and students. To construct the model, mathematics (math) score is used as a response variable. However, TIMSS does not provide the real scores of the student participants. It supplies five plausible values for each area and they are drawn from the resulting posterior distribution for each observation (Fishbean et al., 2019). TIMSS 2019 Technical Report particularly highlights that these plausible values are not individual test scores and they only should be used for group-level analyses. Similarly, Programme for International Student Assesment (PISA) has an analogous way to present achievement scores of the individuals. For example, PISA divides the one long booklet into sub booklets in order to shorten the test length (Uysal, 2015). So, the booklets that participants have are not totally the same. To make fair competition, the test scores are not evaluated by the traditional methods. It is represented by plausible values which are imputations of latent variables. As it is mentioned before, plausible values are randomly drawn from a probability distribution for each individual. Additionally, for TIMSS, these values are drawn from a conditional normal distribution. Plausible values are represented with column names as *ASMMAT01* to *ASMMAT05* in the database. To make a reliable inference, TIMSS (2011) recommends that average of five plausible values can be used.

Considering this suggestion, we perform the analysis by using average of them as a dependent variable.

In preprocessing step of the analysis, student, home and school data sets are merged via unique columns which are student id and school id. Therefore, we have one clustered data set with students are nested in schools. The set has 180 different schools and 181 different classes. Additionally, it includes 4028 observations. However, as each school includes one class (except one of them), it is not possible to add class level effects in the model and see the differences between classes within schools.

4.3.1 Preprocessing Step

4.3.1.1 Reducing the Variables

The merged TIMSS data has approximately 450 columns such as gender of student, whether student has an internet connection or own room, whether student likes science or mathematics, the number of books that student has, etc. While some of them are related with science achievement; others are related with mathematics achievement which we interest. This means that student responses a booklet for both mathematics and science achievement. Therefore, all answers take places in one dataset. Some columns have higher number of missing values because the questions which may be irrelevant with the participants or maybe participants do not want to answer the questions. Additionally, there may be a correlation between some variables. The correlation plot (Figure 1) is added to examine the extent of the correlations.

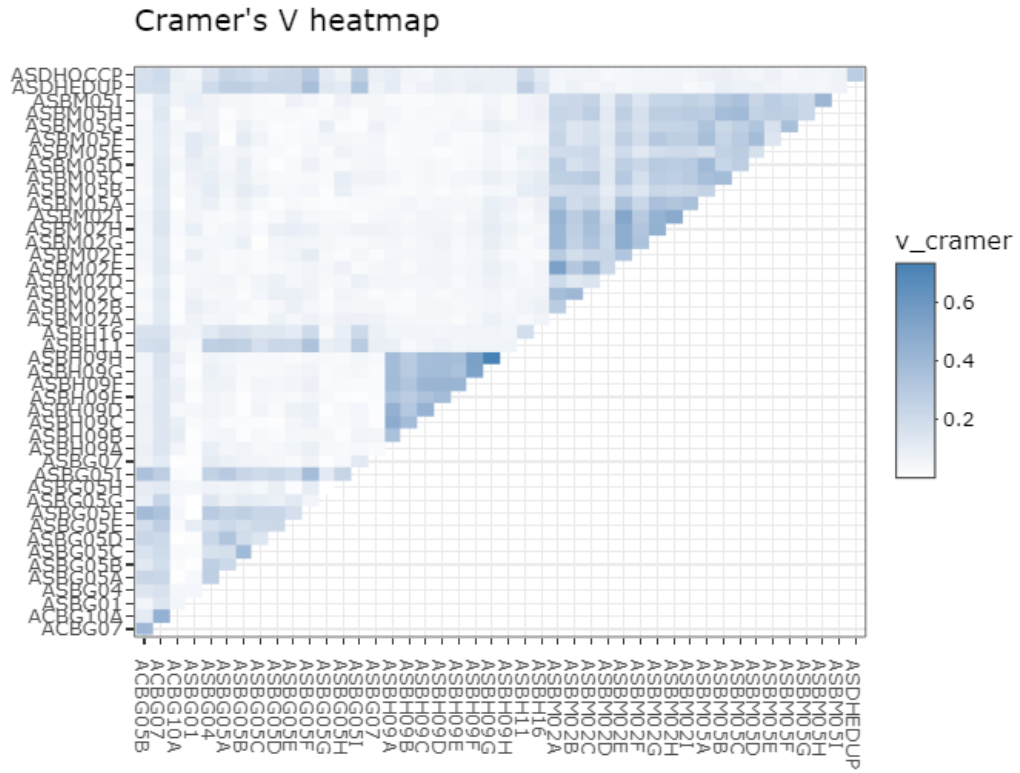

Figure 1 Correlation plot (heatmap) of explanatory variables

According to the plot above, we drop the variable ASBH09H which is more associated with science achievement due to the high (0.73) correlation. As a result, the following variables given at Table 1 are used for the analysis.

Table 1: TIMSS variables of the study

	Variable Name	**Variable Explanation**	**Encoding**
Student Dataset	IDSCHOOL	School ID	
	IDSTUD	Student ID	
	ASBG01	Gender of Student	1: Girl | 2: Boy
	ASBG04	Number of Books in Your Home	1: 0-10 | 2: 11-25 | 3: 26-100 | 4: 101-200 | 5: More than 200
	ASBG05A	A Computer or Tablet	

Table 1 (continued)

ASBG05B	Study Desk/Table for Your Use	Do you have any of these things at your home? 1: Yes | 2: No
ASBG05C	Your Own Room	
ASBG05D	Internet Connection	
ASBG05E	Your Own Mobile Phone	
ASBG05F	Central Heating	
ASBG05G	Air Conditioning	
ASBG05H	Washing Machine	
ASBG05I	Dishwasher	
ASBG07	Were You Born in Turkey?	1: Yes | 2: No | 3: I do not know | 4: Not applicable
ASBM02A	I enjoy learning mathematics	1: Agree a lot | 2: Agree a little | 3: Disagree a little | 4: Disagree a lot
ASBM02B	I wish I did not have to study mathematics	
ASBM02C	Mathematics is boring	
ASBM02D	I learn many interesting things in mathematics	
ASBM02E	I like mathematics	
ASBM02F	I like any schoolwork that involves numbers	
ASBM02G	I like to solve mathematics problems	
ASBM02H	I look forward to mathematics lessons	
ASBM02I	Mathematics is one of my favorite subjects	
ASBM05A	I usually do well in mathematics	1: Agree a lot | 2: Agree a little | 3: Disagree a little | 4: Disagree a lot
ASBM05B	Mathematics is harder for me than for many of my classmates	

Table 1 (continued)

	ASBM05C	I am just not good at mathematics	
	ASBM05D	I learn things quickly in mathematics	
	ASBM05E	Mathematics makes me nervous	
	ASBM05F	I am good at working out difficult mathematics problems	
	ASBM05G	My teacher tells me I am good at mathematics	
	ASBM05H	Mathematics is harder for me than any other subject	
	ASBM05I	Mathematics makes me confused	
Home Dataset	IDSTUD	Student ID	
	ASBH09A	My child's school does a good job including me in my child's education	1: Agree a lot | 2: Agree a little | 3: Disagree a little | 4: Disagree a lot
	ASBH09B	My child's school provides a safe environment	
	ASBH09C	My child's schoolcares about my child's progress in school	
	ASBH09D	My child's school does a good job informing me of his/her progress	
	ASBH09E	My child's school promotes high academic standards	
	ASBH09F	My child's school does a good job in helping him/her become better in reading	

Table 1 (continued)

ASBH09G	My child's school does a good job in helping him/her become better in mathematics	
ASBH09H	My child's school does a good job in helping him/her become better in science	
ASBH11	Amounts of Books for Children at Home	1: 0-10 | 2: 11-25 | 3: 26-50 | 4: 51-100 | 5: More than 100
ASDHEDUP	Parents' Highest Education Level	1: Finished some primary or lower secondary or did not go to school | 2: Finished Lower Secondary | 3: Finished Upper Secondary | 4: Finished Post-Secondary Education | 5: Finished University or Higher
ASBH16	How far in his/her education do you expect your child to go?	1: Finish Lower Secondary Education | 2: Finish Upper Secondary Education | 3: Finish Post-Secondary Vocational Courses | 4: Finish Short-cycle Tertiary Education | 5: Finish Bachelor's Level | 6: Finish Master's or Doctor
ASDHOCCP	Parent's Highest Occupation Level	1: Has never worked outside home for pay,

Table 1 (continued)

			general laborer, or semi-professional (skilled agricultural or fishery worker, craft or trade worker, plant or machine operator) | 2: Clerical (clerk or service or sales worker) | 3: Small Business Owner | 4: Professional (Corporate Manager or Senior Official, Professional, or Technician or Associate Professional)
School Dataset	IDSCHOOL	School ID	
	ACBG05B	Which best describes the immediate area in which your school is located?	1: Urban-Densely Populated | 2: Suburban-On fringe or outskirts of urban area | 3: Medium Size City or Large Town | 4: Small Town or Village | 5: Remote Rural
	ACBG07	How many computers (including tablets) does your school have for use by fourth grade students?	
	ACBG10A	Does your school have library?	1: Yes | 2: No

4.3.1.2 Derived Variables

For the ease of interpretation and more reliable results, it may be preferable to use transformed variables when it is possible and coherent. Most of the time, converted variables are more explanative than raw variables. For example, only height value or only weight value is meaningless about obesity. However, the combination of the two creates the body mass index which gives an insight into the measure of body fat. Similarly, it will be conceptually more beneficial and explanatory to present a combination of some information, rather than giving the information about a particular concept separately. For example, SES scores are created in order to interpret the socio-economic status of individuals as a whole with the help of an index, instead of interpreting the several variables separately. Hence, we prefer such explanatory variables for our analysis. In addition to that, to obtain reliable outputs all variable encodings are modified. For instance, while 1 means yes in the original form, 1 represents no in the modified form. Similarly, 2 shows us yes for variables that have positive meaning. More clearly, for the variable whether student has own room, 1 means no and 2 is yes which is expected to give good effects for students. Also, emotional columns are coded as 1, 2, 3 and 4, like Likert scale, from agree a lot to disagree a lot. To illustrate, for the variable that he or she enjoys learning mathematics, 1 is changed with 4. On the other hand, for the variable that he or she wishes he/she did not have to study mathematics, 1 is remained the same. Therefore, we think if coding is 4, then, it is expected to have a good effect to the student achievements. In short, the raw Likert scale is modified according to whether the column is expected to affect student achievement positively or negatively.

4.3.1.2.1 SES Score

The data has certain columns that represent socio-economic status of a student. Some of the status variables change from country to country while some of them remain constant. For instance, the data includes nine columns in terms of wealth status. Five

of them are the same for all countries. They are whether a student has a computer or tablet, study desk, own room, internet connection and mobile phone. The other four columns also represent socio-economic status indicators for a specific country. For example, they are central heating, air conditioning, washing machine and dishwasher for Turkey while they are gaming system, smartphone, luxury watch and luxury car for United Arab Emirates. If a student has the related indicator, then, he or she answers as 2 which represents having it. Otherwise, he or she responses as 1 which means not having it. We prefer to add them up to create SES score for Turkey. When SES score increases, it is accepted that wealth status of the student also increases. Hence, the wealth status is represented via only one derived variable.

4.3.1.2.2 Family Attitude Score

One of the important factors for students' mathematics achievement is family attitude. If families support their children, it is expected that the student has higher achievement score. In TIMSS example, supporting is represented by the number of books at home, the number of books for children at home and the highest level of education expected by families. In addition to these indicators, parent's highest education level and parent's highest occupation level are regarded as family attitudes. By adding them up, we create a new family attitude score for the analysis.

4.3.1.2.3 School Status

To measure interschool variability, the variables that describe schools are needed. These variables in our analysis are selected as the existence of school library and the total number of computers per student in the school. The existence of school library has two categories which are yes (1) and no (2) and its raw data is found in the dataset. However, the total number of computers per student in the school should be derived by using other variables. The data has the total number of computers. Dividing it into the total number of students per school gives the indicator that is the

number of computers per student. By doing this, we use more homogenized variable for school. Assume that while school 1 has 50 computers and 25 students, school 2 has 50 computers and 10 students. If the number of computers was used directly without taking into account the number of students in the corresponding school, the misguided indicator would be taken into account. For this case, considering only total number of computers gives the same results. However, we know the total number of students and we can say that school 2 has more advantages in terms of the number of computers. As while school 1 has 2 computers per student, school 2 has 5 ones per student.

After calculating the total number of computers per student in school, it is coded as 2 for the values greater than median 0.6. Otherwise, it is marked as 1. The existence of library coding is interchanged for logical integrity. So, the response yes which has positive meaning for existing library is represented with 2. In the last step, these two categorical indicators are summed up.

CHAPTER 5

DATA ANALYSIS

5.1 Exploratory Data Analysis

In this step, the main characteristics of the response and explanatory variables are summarized. This gives an idea of the data and the motivation for the application methods. The data has 180 different schools. Different number of students were selected from these schools. For example, the highest number of students comes from the school whose id number is 5122 with 34 students. The second school with the highest rate is the school numbered as 5148 with 33 students. On the contrary, the schools numbered 5013, 5039 and 5048 have only 3 representatives.

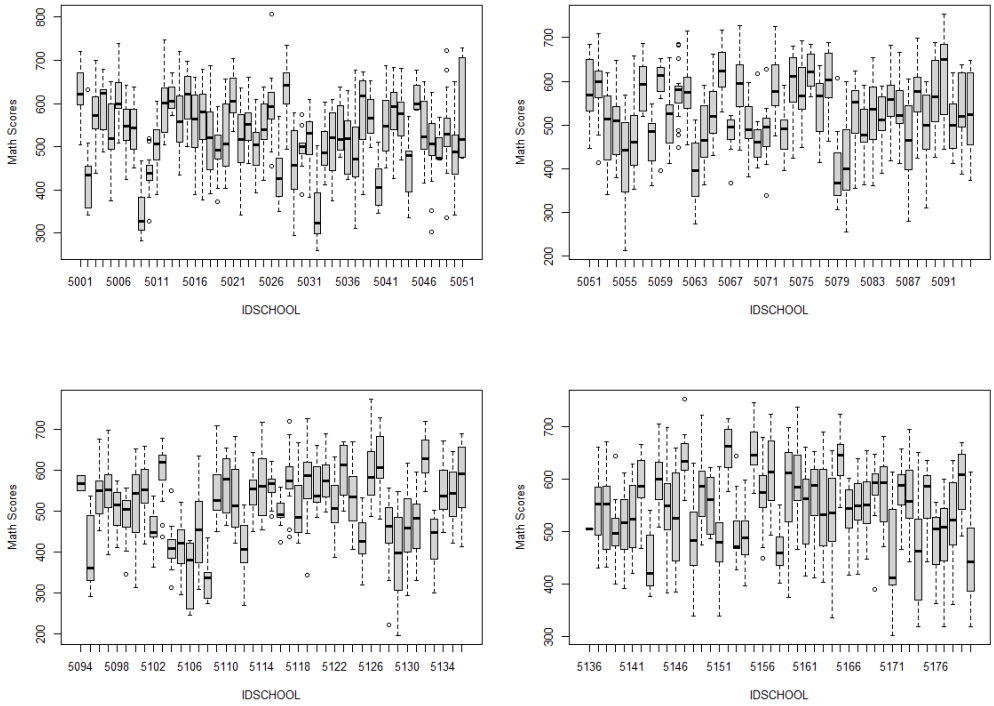

Figure 2 Box-plots of math scores by schools

The box-plots (Figure 2) depict the math scores by schools. It indicates the difference amongst schools in terms of math scores. Moreover, the variation between schools may have a crucial role in terms of modelling. Therefore, mixed-effect modelling design is suitable for the case. In addition to that, we observe some outliers in the data.

There are 3052 responses from 1639 girls and 1413 boys. 2971 of them were born in Turkey. The variable immediate area which represents the area near the school has 4 categories. 1451 students' schools belong to the area urban densely populated while 871 of them are in a medium size city or a large town. According to the 460 students, their schools are located in a small town or a village. Only 270 of the schools belong to the suburban which means on fringe or outskirts of urban area.

The factors which are emotional, mathematical tendency and school for family are continuous indicators (the more detailed explanation related to factor analysis will be given at the following section). Emotional factor takes values between -4.86 and 1.73. Its median value is greater than the mean value. So, it is expected that the variable has left-skewed distribution shape. On the other hand, school for family factor has the median value of -0.44. Its maximum and minimum values are 5.44 and -0.96, respectively. So, its range is 6.39. Because the mean of the variable is 0 and the median is less than the mean, probably, distribution shape of the variable looks right-skewed. Additionally, mathematical tendency factor has a mean of 0 and a median of 0.2. It takes values between -3.25 and 2.86. Distribution shapes of these three variables can be seen at Figure 3.

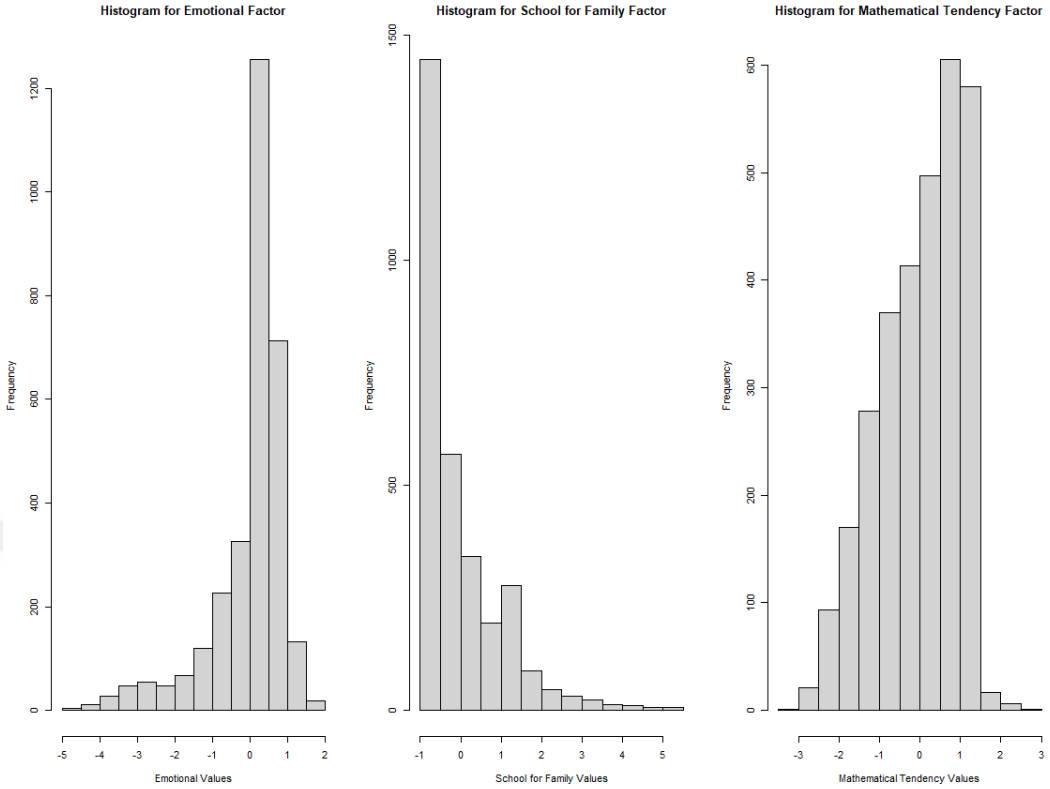

Figure 3 Distribution shapes of the factor scores

As it is said before, the distribution shape of the emotional factor looks strongly left-skewed. The highest frequency belongs to the value which is between 0 and 1. On the other hand, the shape of the variable school for family factor seems strongly right skewed as expected. The values are mostly gathered around -1 and -0.5.

The variables SES score, family attitude and school status are variables which take discrete values. SES score values are in between 9 and 18. The highest number of frequencies belongs to the value 15. Family attitude has median of 17. The maximum and minimum values are 27 and 8, respectively. School status takes the values of 2, 3 and 4. The value 3 and 4 have the highest frequencies as 1247 and 1223, respectively.

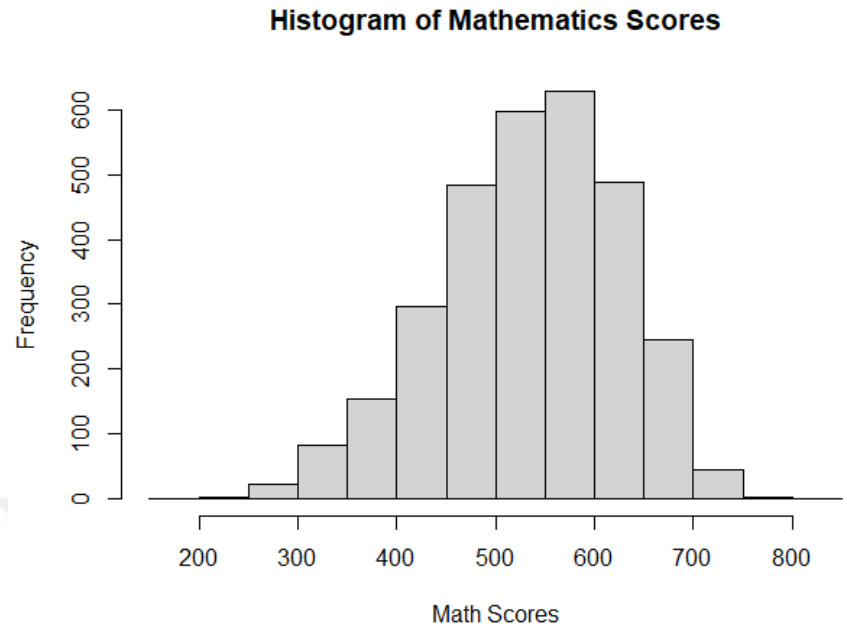

Figure 4: Histogram of mathematics scores

Figure 4 depicts the distribution shape of mathematics scores. They take values between 200 and 800. The highest frequency belongs to the scores between 550-600.

5.2 Factor Analysis

Factor analysis is a data reduction technique and it is suitable for Likert scale type of analysis. Therefore, it successfully fits with the TIMSS data. Some columns have Likert scale type of answers. Also, they give the same meanings around the same ideas which provide a conceptual integrity regarding the variables. For example, a group of columns is related with student's tendency to mathematics. Another group of columns is about parent's thoughts about children's schools. Overall, 25 columns are represented with fewer indicators by the help of the factor analysis.

After the implementation of factor analysis with promax rotation (which gives the most cognitively logical results), three main groups are gathered around three

different factors, separately. These factor scores are used at the rest of the analysis. The table below shows us the factor loadings.

Table 2: Factor loadings

Variable Name	Factor 1 (Emotional Factor)	Factor 2 (School for Family Factor)	Factor 3 (Mathematical Tendency Factor)
ASBM02A	0.848		-0.102
ASBM02B	0.373		0.224
ASBM02C	0.548		0.163
ASBM02D	0.467		-0.125
ASBM02E	0.946		-0.105
ASBM02F	0.675		-0.138
ASBM02G	0.786		
ASBM02H	0.795		
ASBM02I	0.827		
ASBM05A	0.410		0.309
ASBM05B	-0.171		0.815
ASBM05C			0.731
ASBM05D	0.348		0.258
ASBM05E			0.515
ASBM05F	0.296		0.316
ASBM05G	0.248		0.296
ASBM05H			0.813
ASBM05I			0.777
ASBH09A		0.763	
ASBH09B		0.638	
ASBH09C		0.772	
ASBH09D		0.766	
ASBH09E		0.707	

ASBH09F	0.805
ASBH09G	0.738

The following sections provide with more detailed information about the factor results.

5.2.1 Emotional Factor

The variables from ASBM02A to ASBM02I are about students' thoughts towards mathematics. The question with these columns is how much a student agrees with given statements about learning mathematics. For example, ASBM02A says that "I enjoy learning mathematics". On the other hand, ASBM02B stands for "I wish I did not have to study mathematics". According to Table 2, we understand that these columns are gathered around Factor 1. Since these questions related to the emotions of students, Factor 1 is entitled as **emotional factor**.

5.2.2 School for Family Factor

The survey includes not only students' perspectives but also parents' perspectives. Questions about their children's schools are asked to parents and results are included in the dataset. These are represented from ASBH09A to ASBH09G. The related question for these columns is what parents think of their children's schools. The columns stand for the idea like "School provides a safe environment", "School cares about child's progress in school" and "School promotes high academic standards". Parents give a score from 1 for strongly disagreeing to 4 for strongly agreeing. Factor analysis reveals that these statements can be represented with Factor 2 and it is entitled as **school for family factor**.

5.2.3 Mathematical Tendency Factor

The variables from ASBH05A to ASBH05I give an idea about student's mathematics anxiety. The question asked to students for these columns is how much a student agrees with the given statements about mathematics. To illustrate, the columns represent the ideas such as "I usually do well in mathematics", "I am good at working out difficult mathematics problems" and "Mathematics makes me nervous". These statements are mainly collected around Factor 3. So, this factor is entitled as **mathematical tendency factor**.

5.3 Analyses

In this part, different modelling designs are built to compare and to find the most suitable design for our case. In doing so, raw variables, that are gender, whether a student born in Turkey or not, school immediate area, derived variables, which are SES score, family attitude and school status and factor scores, that are emotional factor, mathematical tendency factor and school for family factor are used. In total, we have 12 variables including dependent variable math score and unique keys which are student id and school id. Before modelling, the data is divided into train and test sets with rates 0.8 and 0.2 for cross-validation, respectively.

5.3.1 Linear Model Analysis

The analysis part starts with a linear model. Linear model is built by using *lm* in R with all the variables. The variables that are gender, born in Turkey and immediate area are informed to the model as factor type. After the modelling, outputs show that school for family factor is statistically insignificant at the 5% confidence level. Additionally, immediate area has 4 levels and level 1 is defined as reference. According to the level 1, while level 4 is statistically significant, levels 2 and 3 are statistically insignificant. However, the variable immediate area is included for the

rest of the analysis. On the other hand, school for family factor is statistically insignificant. So, it is dropped from the model. The table below shows us parameter estimations, standard error terms and p-values obtained from linear model.

Table 3: Outputs of LM

Variable Name	**Estimate**	**Std. Error**	**Pr(>|t|)**	**Sign.**
Intercept	249.88	19.62	$< 2*10^{-16}$	***
Gender (2)	10.666	3.04	0.0005	***
Born_in_Turkey (2)	55.341	9.57	$8.26*10^{-09}$	***
Immediate_Area (2)	2.651	5.68	0.641	
Immediate_Area (3)	-7.292	3.61	0.044	*
Immediate_Area (4)	-51.426	4.85	$< 2*10^{-16}$	***
EmotionalFactor	7.265	1.51	$1.52*10^{-06}$	***
MathematicalTendencyFactor	38.214	1.55	$< 2*10^{-16}$	***
SES	11.684	0.81	$< 2*10^{-16}$	***
Family_Attitude	2.117	0.63	0.00085	***
School_Status	7.522	2.06	0.000267	***
Signif. Codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**				

Intercept and coefficients for gender, born in Turkey, immediate area, emotional factor, mathematical tendency factor, SES score, family attitude, and school status are 249.88, 10.666, 55.341, 2.651, -7.292, -51.426, 7.265, 38.214, 11.684, 2.117 and 7.522, respectively. These coefficients belong to the model that does not include school for family factor. Also, by using the test set, mean squared error is calculated as 67.74. According to the model outputs in table 3, it can be said that while some variables have a positive effect on math scores, some others negatively effect math scores. For example, being boy and being Turkey-born increase math score 10.666 units and 55.341 units, respectively. If the school is located in medium size city or

small town, then, it negatively affects math scores of the students. Also, when other situations remain constant, one unit increase in emotional factor increases math scores by 7.265 units. Similarly, one unit change in mathematical tendency factor increases or decreases math scores by 38.214 units. Moreover, SES score, family attitude and school status have also positive effects to math scores in case of one-unit increase.

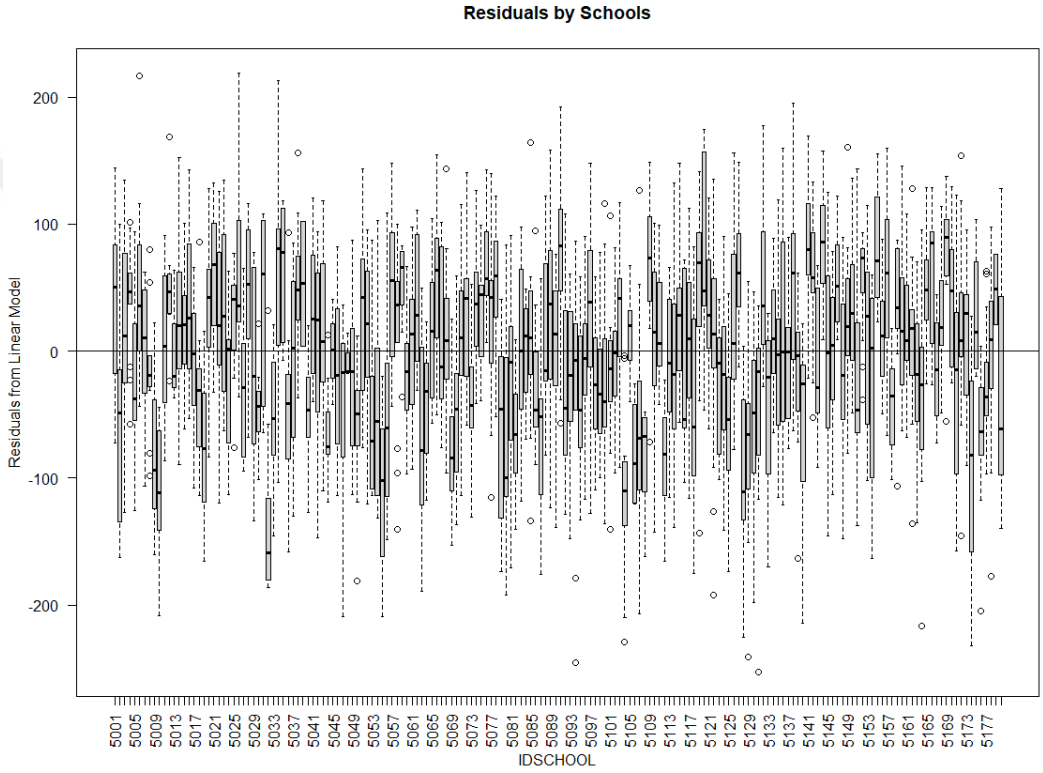

Figure 5: Residuals by schools

The boxplot above (Figure 5) shows us the residuals obtained from the linear model for each school. While x-axis represents id numbers of schools, y-axis represents the residuals. By looking at the plot, it can be seen that many schools look like outliers. Even their boxes do not touch the line at y=0. This means that the residuals do not scatter randomly around zero. Before, we stated that there is a difference between schools by looking at the box-plot for each school (Figure 2). And, we added that

linear model is not suitable for TIMSS data because it does not take within subject correlation into account. Now, residual plot is another reason to use LMM.

5.3.2 Linear Mixed Model Analysis Using Functions lme and lmer in R

LMM can be built in R by using the functions **"lme"** and **"lmer"**. The function lme is gathered from the nlme package while the function lmer is located in package lme4. Jose Pinheiro and Douglas Bates (2021) explain the updated nlme package. The article "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data" written by Lindstrom, M.J. and Bates, D.M. (1988) forms the computational framework of the lme function. Additionally, the formulation of the LMM is based on the article of Laird and Ware (1982). On the other hand, lmer function is found in package lme4 (Bates et al., 2021). The package is updated and released in 2021.

As linear model regression, gender, born in Turkey and immediate area are introduced as factors in LMM. Similarly, the variable school for family factor is statistically insignificant. Therefore, it is dropped from the model.

In our data, two types of random effects are possible. The first one is adding only one random effect which is school to be able to model the variations amongst schools. The second model option is adding a random effect for students nested in schools, besides schools. So, we can consider both school and student variations. However, the second one is a little bit complex than the first one. Therefore, the necessity to add the student nested in school model should be checked before moving to a more complex model.

ANOVA helps to decide which model is statistically more preferable. Assume that model 1 is linear model. Model 2 is LMM with random effect school and model 3 is LMM with random effects school and student nested in school. ANOVA for LM (model 1) and LMM with random effect school (model 2) says that LMM is more suitable for this dataset. It is understood by looking at the AIC and BIC terms in table

4. We know that the model with smaller AIC and BIC terms are preferable. The analysis says that there is statistically significant evidence to use the LMM. Then, ANOVA for LMM with random effect school (model2) and LMM with random effects school and student nested in school (model 3) is conducted. After implementing the ANOVA function that subtracts -2 ML log-likelihood value for the reference model from the nested model, we found the unsignificant result of this test ($p > 0.05$). That is, no need to use complex model which is LMM with random effects school and student nested school (model3) because the values of AIC and BIC for more complex LMM (model 3) are greater than the values of the LMM with random effect school (model 2). It says that the second model is well enough for our modelling design. Therefore, we retain the unnested model which is LMM with random effect school (model 2). See the model structures at below:

Model 1: y~covariates (LM)

R Syntax:

```
model=lm(y~factor(Gender)+factor(Born_in_Turkey)+factor(Immediate_Area)+E
motionalFactor+MathematicalTendencyFactor+ses+family_attitude+
school_status,data=train)
```

Model 2: y~covariates + school as random effect (LMM)

R Syntax:

```
model2=lme(y~factor(Gender)+factor(Born_in_Turkey)+factor(Immediate_Area)
+EmotionalFactor+MathematicalTendencyFactor+ses+family_attitude+
school_status, random = ~ 1| IDSCHOOL, train, na.action = "na.omit", method =
"ML")
```

Model 3: y~covariates + school and student nested school as random effects (LMM)

R Syntax:

model=lme(y~factor(Gender)+factor(Born_in_Turkey)+factor(Immediate_Area)+ EmotionalFactor+MathematicalTendencyFactor+ses+family_attitude+ school_status, random = ~ 1| IDSCHOOL/IDSTUD, train, na.action = "na.omit", method = "ML")

Table 4: Outputs of ANOVA

Model	df	AIC	BIC	Sign.
Model 1	12	27988.36	28057.97	✓
Model 2	13	27312.95	27388.35	
Model 2	13	27312.95	27388.35	X
Model 3	14	27314.95	27396.15	

The unnested mixed model intercept and coefficients for gender, born in Turkey, immediate area, emotional factor, mathematical tendency factor, SES score, family attitude, and school status are 386.10, 12.732, 40.698, -9.293, -15.179, -69.960, 10.640, 35.681, 3.614, 1.494 and 10.958. While two coefficients for immediate area are not statistically significant, p-values of all other variables are less than 0.05, that is, they are statistically significant at the 5% confidence level. Additionally, its mean square error value is equal to 60.39. These results are obtained by using both functions lme and lmer, separately. These two functions give the same results as expected. Outputs of LMM can be seen from table 5 which is below.

Table 5: Outputs of LMM

Variable Name	Estimate	Std. Error	Pr(>|t|)	Sign.
Intercept	386.10	24.47	0.00	***
Gender (2)	12.732	2.62	0.00	***
Born_in_Turkey (2)	40.698	8.04	0.00	***
Immediate_Area (2)	-9.293	13.73	0.499	
Immediate_Area (3)	-15.179	9.07	0.096	.
Immediate_Area (4)	-69.960	10.94	0.00	***
EmotionalFactor	10.640	1.299	0.00	***
MathematicalTendencyFactor	35.681	1.306	0.00	***
SES	3.614	0.786	0.00	***
Family_Attitude	1.494	0.546	0.006	***
School_Status	10.958	5.25	0.04	***
Signif. Codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**				

According to the LMM outputs, all the variables except immediate area have a positive effect on math scores in case of one-unit increase. For example, one unit increase in the variable born in Turkey increases math scores by 40.698 units. Similarly, higher mathematical tendency gives higher math score when other situations remain constant. Additionally, family attitude is significant but, it does not give points as highly as other positive effects in case of one unit increase.

5.3.3 Linear Mixed Model Analysis Using Different Algorithms in R

Schafer J.L. (1998) describes how certain algorithms for LMM can be implemented using R. ECME algorithm, Fast ECME algorithm and Fast MCMC algorithm are examples of these certain algorithms. These functions are called as ecmeml, fastml and fastmcmc in R, respectively. Ecmeml and fastml are implementations of the

MLE method. However, it is possible to use restricted maximum likelihood estimation such as ecmerml and fastrml.

R Syntax:

pred=cbind(int,final$Gender,final$Born_in_Turkey,final$Immediate_Area,final$EmotionalFactor,final$MathematicalTendencyFactor,final$ses,final$family_attitude,final$school_status)

xcol=1:dim(pred)[2]

zcol=1

R Syntax of ECME Algorithm:

e=ecmeml(final$y,final$IDSCHOOL,pred,xcol,zcol)

Table 6: Outputs of ecme algorithm

Variable Name	**Estimate**
Intercept	343.73
Gender	12.771
Born_in_Turkey	45.212
Immediate_Area	-18.197
EmotionalFactor	9.977
MathematicalTendencyFactor	35.578
SES	3.103
Family_Attitude	1.883
School_Status	12.199

By looking at table 6, it can be seen that model intercept and coefficients for gender, born in Turkey, immediate area, emotional factor, mathematical tendency factor, SES score, family attitude, and school status are obtained via the function ecmeml as 343.73, 12.771, 45.212, -18.197, 9.977, 35.578, 3.103, 1.883 and 12.199,

respectively. These parameter estimations are calculated with 11 iterations. Moreover, the last iteration has the loglikelihood value of -14214.08.

The function fastml reaches the parameter estimations for gender, born in Turkey, immediate area, emotional factor, mathematical tendency factor, SES score, family attitude, and school status in 6 iterations. The outputs of fast algorithm are shown in table 7 which is below. These estimation values are 343.73, 12.771, 45.212, -18.197, 9.977, 35.578, 3.103, 1.883 and 12.199, respectively. -14214.08 is the loglikelihood value for the last iteration.

R Syntax of Fast Algorithm:

f=fastml(final$y,final$IDSCHOOL,pred,xcol,zcol)

Table 7: Outputs of fast algorithm

Variable Name	**Estimate**
Intercept	343.73
Gender	12.771
Born_in_Turkey	45.212
Immediate_Area	-18.197
EmotionalFactor	9.977
MathematicalTendencyFactor	35.578
SES	3.103
Family_Attitude	1.883
School_Status	12.199

The algorithm Fast MCMC which is applied via the function fastmcmc in R needs some prior information. However, in our case, it is not so sensitive to priors. Therefore, arbitrary initial points are used. As a result, we have parameter estimations for gender, born in Turkey, immediate area, emotional factor, mathematical tendency factor, SES score, family attitude, and school status as 343.73, 12.774, 45.186, -18.206, 9.982, 35.576, 3.089, 1.882 and 12.206,

respectively. Unfortunately, number of iterations and loglikelihood value are not supplied from the function fastmcmc.

R Syntax of Fastmcmc Algorithm:

prior <- list(a=1,b=2,c=3,Dinv=4)

mcmc=fastmcmc(final$y,final$IDSCHOOL,pred,xcol,zcol,prior,seed=1)

Table 8: Outputs of fastmcmc algorithm

Variable Name	**Estimate**
Intercept	343.989
Gender	12.774
Born_in_Turkey	45.186
Immediate_Area	-18.206
EmotionalFactor	9.982
MathematicalTendencyFactor	35.576
SES	3.089
Family_Attitude	1.882
School_Status	12.206

By building LMM, the functions lme and lmer work in similar way. Therefore, parameter estimations give the same results. The coding designs are also the same and easy to implement. For example, lme and lmer functions need to identify dependent variable, first. After dependent variable, "~" is added. Explanatory variables are written using "+". It is crucial to define random term. For our case, random term identified as "random=~1|IDSCHOOL". Then, name of dataset, na action and method are stated.

Table 9: Comparable table for LM and LMM

Variable Name	**LM (lm)**	**LMM (lme)**	**LMM (lmer)**
Intercept	249.88	386.10	386.10

Gender (2)	10.666	12.732	12.732
Born_In_Turkey(2)	55.341	40.698	40.698
Immediate_Area (2)	2.651	-9.293	-9.293
Immediate_Area(3)	-7.292	-15.179	-15.179
Immediate_Area(4)	-51.426	-69.960	-69.960
Emotional Factor	7.265	10.640	10.640
Mathematical Tendency Factor	38.214	35.681	35.681
SES	11.684	3.614	3.614
Family Attitude	2.117	1.494	1.494
School Status	7.522	10.958	10.958
RMSE	67.74	60.388	60.388
Algorithm	NA	EM	Nelder-Mead and BOBYQA

Table 10 shows that parameter estimations of the functions ecme and fast algorithms give the same results. Also, estimations obtained from the function fastmcmc are very close to them. The number of iterations of ecme algorithm equals to 11 while the number of iterations for fast algorithm is 6. That is, fast algorithm reaches the convergency faster than ecme algorithm. On the other hand, fastmcmc algorithm does not give the number of iterations.

Table 10: Comparable table for different algorithms

Variable Name	**func. ecmeml**	**func. fastml**	**func. fastmcmc**
Intercept	343.73	343.73	343.989
Gender	12.771	12.771	12.774
Born_In_Turkey	45.212	45.212	45.186

Immediate_Area	-18.197	-18.197	-18.206
Emotional Factor	9.977	9.977	9.982
Mathematical Tendency Factor	35.578	35.578	35.576
SES	3.103	3.103	3.089
Family Attitude	1.883	1.883	1.882
School Status	12.199	12.199	12.206
Iterations	11	6	NA
Algorithm	ECME	ECME and Fisher Scoring	MCMC

5.4 Model Evaluation

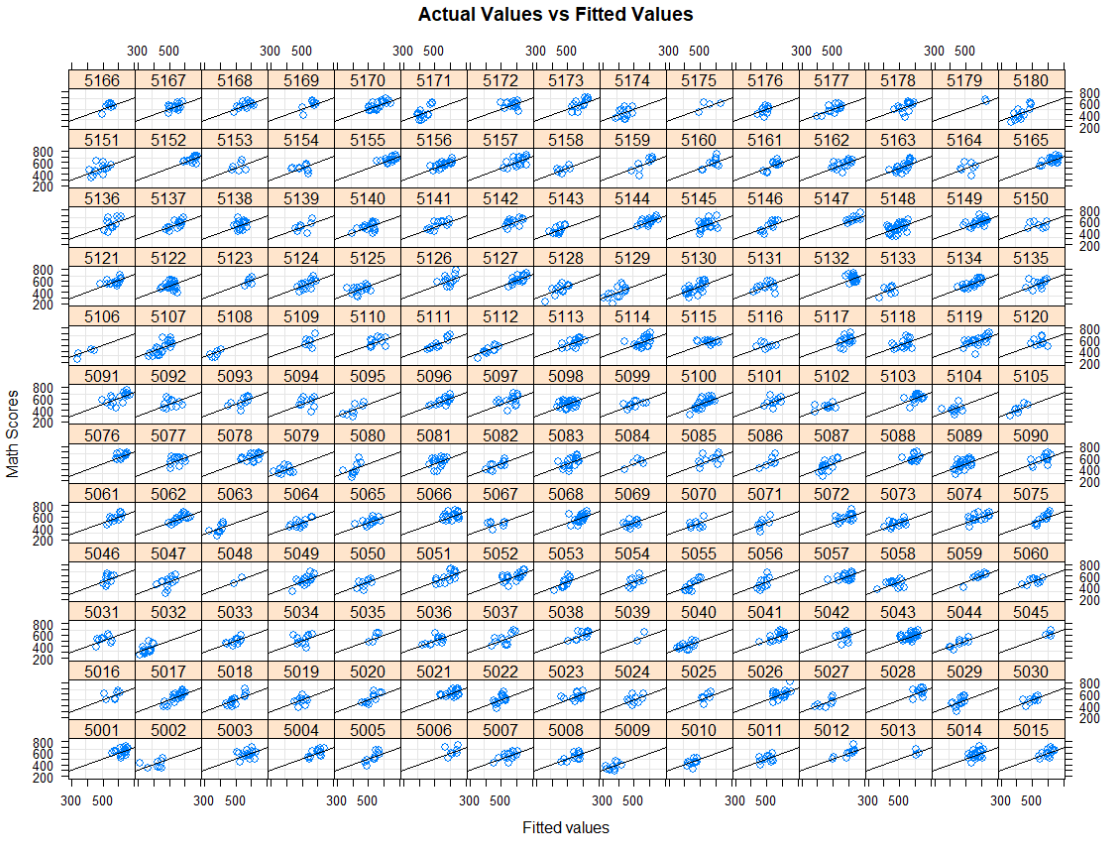

Figure 6: Actual values vs fitted values

Figure 6 shows actual math scores and fitted math scores obtained from LMM for each school. The horizontal axis represents fitted values while the vertical axis shows actual values. We see that each school has reliable actual and fitted values as their values gather around the line. This means that LMM is suitable analysis for TIMSS data.

CHAPTER 6

COUNTRY COMPARISON

In this chapter, the comparison of the factors that affect the students' mathematics achievement for three different countries are evaluated using TIMSS 2019. These three countries are Turkey, England and South Africa. There are several reasons why these countries were chosen as a sample. Turkey's mathematical achievement score is higher than the scale mid-point; while it has the highest standard deviation, which shows a greater fluctuation throughout students. We choose two countries also to compare with Turkey depending on three factors:

1. Having countries in the most successful 10 and the least successful 10.
2. Including a developed country and a developing country.
3. They tested the same grade as Turkey (grade 5).

England and South Africa satisfy the given factors. However, the proportion of missing observations is quite high for England. If all variables used in chapter 5 are taken into account, some columns for England do not even have inputs. When missing observation patterns are examined, it turned out that most of the missing data for England belong to the home dataset. We can conclude that the home dataset does not exist for England, so it is not included in analysis for country comparisons.

In addition to that, in case of merging the datasets student and school, the major part of the data is still missing. When it is applied row-wise deletion to the data, Turkey dataset decreases to 3334 rows from 4028 rows. England loses 1760 observations out of 3396 rows. The number of rows of South Africa declines to 6931 from 11891 observations. These numbers are computed for each case. If all variables are deleted at the end of the merging datasets, 38% of the data is missing. The crucial point is that the majority of the missing values belongs to the school dataset. Additionally,

the correlation between dependent variable math score and the variables come from school dataset are very close to 0 for each case. That is, these variables do not have any relations. When the school dataset is ignored, the missingness rate approximately decreases to 30%. Therefore, we decide not to add school dataset. As a result, we have 13602 observations from three different countries. The corresponding variable names, explanations and encodings are found at Table 11 or country comparison analysis.

Table 11: Data dictionary for country comparison

	Variable Name	**Variable Explanation**	**Encoding**
Student Dataset	IDCOUNTRY	Country ID	
	IDSCHOOL	School ID	
	IDSTUD	Student ID	
	ASBG01	Sex of Student	1: Girl | 2: Boy
	ASBG04	Number of Books in Your Home	1: 0-10 | 2:11-25 | 3: 26-100 | 4: 101-200 | 5: More than 200
	ASBG05A	A Computer or Tablet	Do you have any of these things at your home? 1: Yes | 2: No
	ASBG05B	Study Desk/Table for Your Use	
	ASBG05C	Your Own Room	
	ASBG05D	Internet Connection	
	ASBG05E	Your Own Mobile Phone	
	ASBG05F	Central Heating (T) Your own television (E)	

Table 11 (continued)

	Dictionary (SA)	
ASBG05G	Air Conditioning (T) Bicycle (E) Electricity (SA)	
ASBG05H	Washing Machine (T) Musical Instrument (E) Running Tap Water (SA)	
ASBG07	Were You Born in The Related Country?	1: Yes | 2: No | 3: I do not know | 4: Not applicable
ASBM02A	I enjoy learning mathematics	1: Agree a lot | 2: Agree a little | 3: Disagree a little | 4: Disagree a lot
ASBM02B	I wish I did not have to study mathematics	
ASBM02C	Mathematics is boring	
ASBM02D	I learn many interesting things in mathematics	
ASBM02E	I like mathematics	
ASBM02F	I like any schoolwork that involves numbers	
ASBM02G	I like to solve mathematics problems	
ASBM02H	I look forward to mathematics lessons	

ASBM02I	Mathematics is one of my favorite subjects	

Table 11 (continued)

ASBM05A	I usually do well in mathematics	1: Agree a lot | 2: Agree a little | 3: Disagree a little | 4: Disagree a lot
ASBM05B	Mathematics is harder for me than for many of my classmates	
ASBM05C	I am just not good at mathematics	
ASBM05D	I learn things quickly in mathematics	
ASBM05E	Mathematics makes me nervous	
ASBM05F	I am good at working out difficult mathematics problems	
ASBM05G	My teacher tells me I am good at mathematics	
ASBM05H	Mathematics is harder for me than any other subject	
ASBM05I	Mathematics makes me confused	

The dataset has 32 explanatory variables and 1 dependent variable. Some variables in table 11 includes certain letters like T, E and SA. This means that the variable has different type of questions and it is specified with these letters. T represents Turkey. E means England while SA represents South Africa.

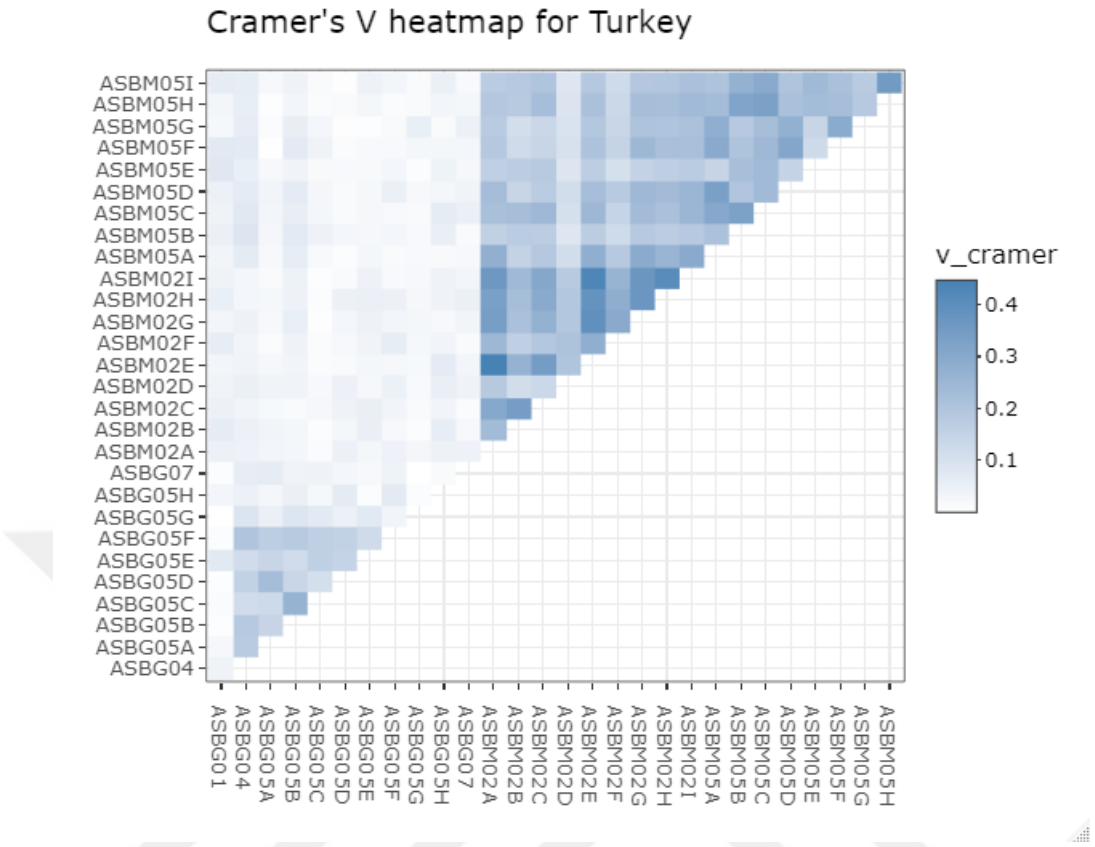

Figure 7: Heatmap for Turkey

The plot above shows the correlation between the variables for Turkey. It seems that there are no highly correlated variables.

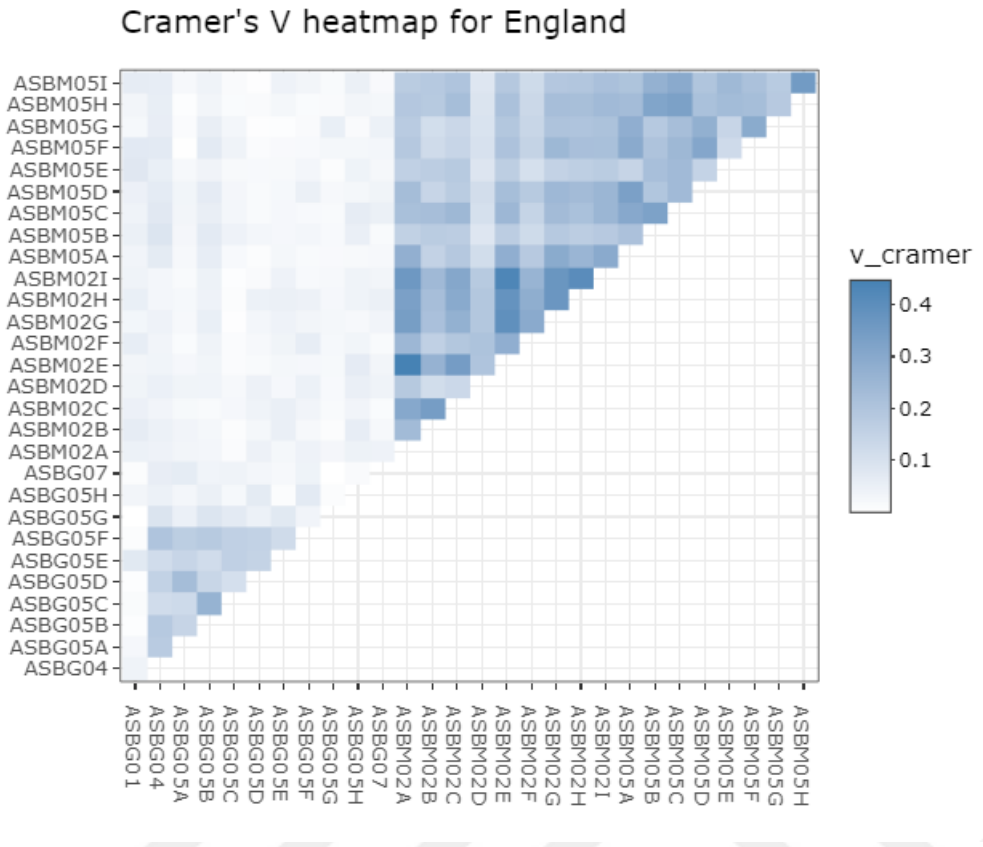

Figure 8: Heatmap for England

Similar with Turkey, the dataset for England does not have highly correlated variables. The highest correlation is almost 0.45.

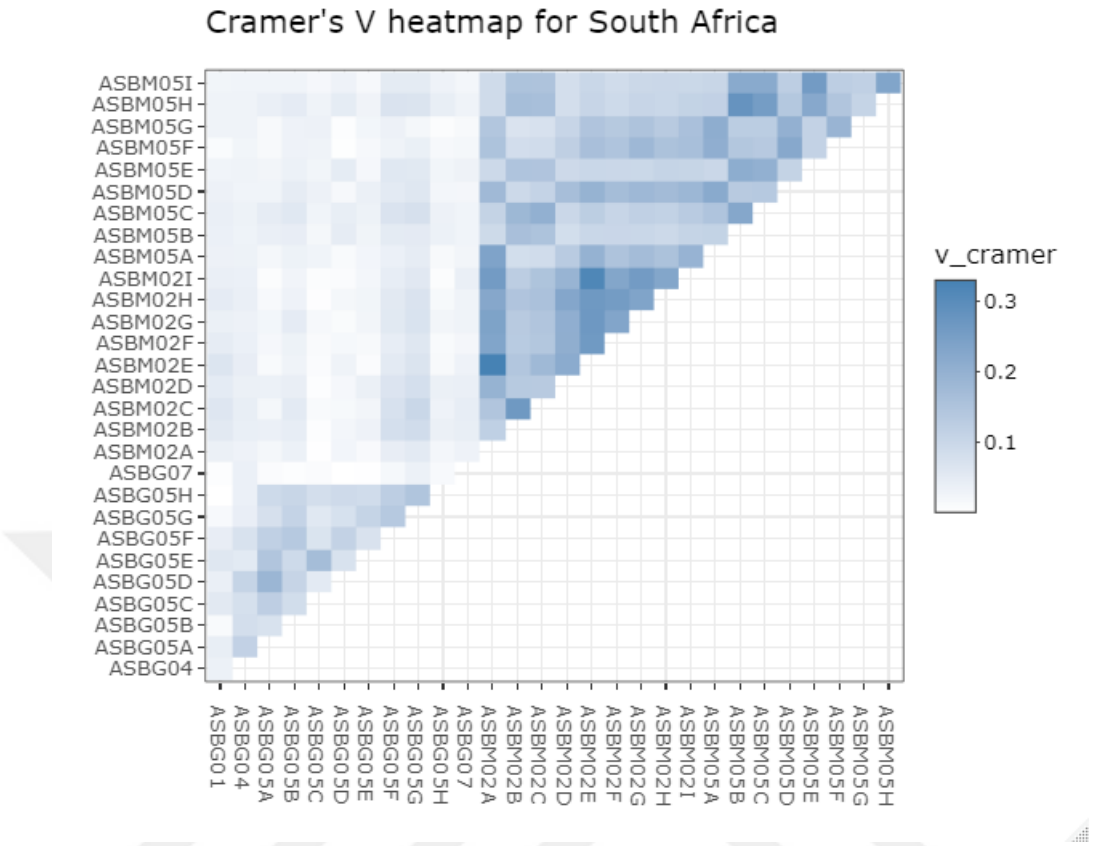

Figure 9: Heatmap for South Africa

The heatmap for South Africa shows that there are no highly correlated variables. The highest correlation is approximately 0.33.

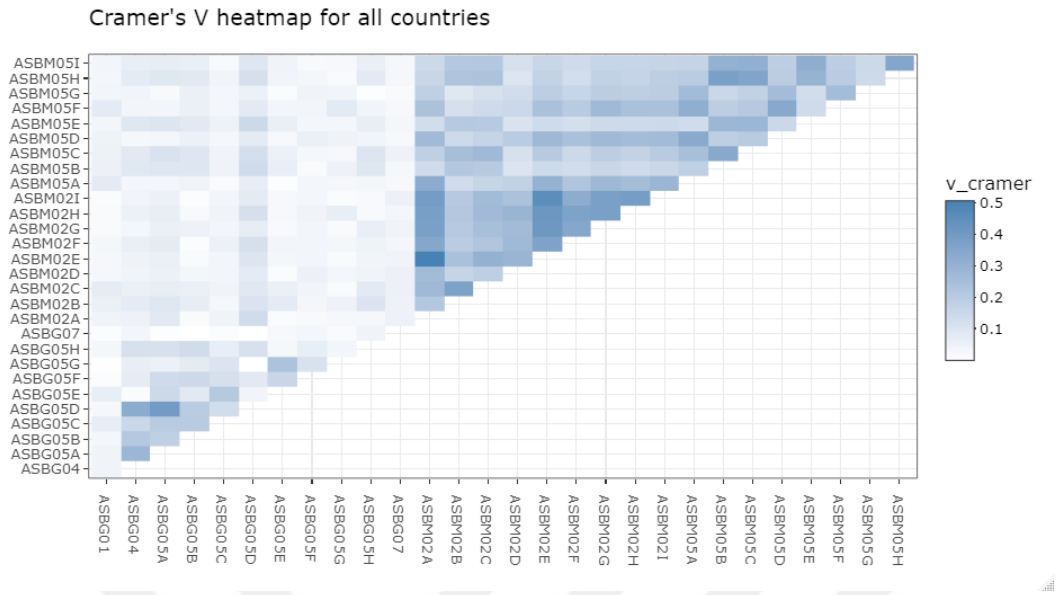

Figure 10: Heatmap for all countries

The heatmap for all countries shows that there is no high correlation between the variables. Therefore, we can continue with these columns.

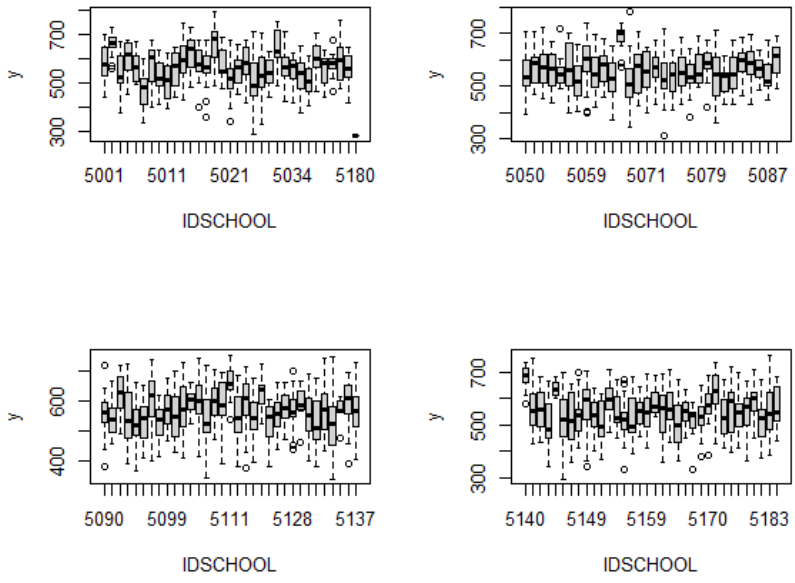

Figure 11: Math scores by school for England

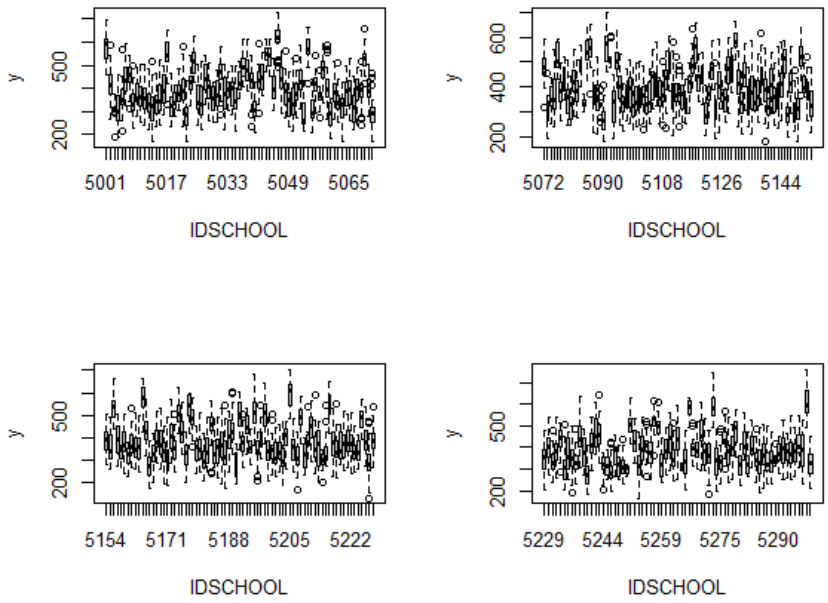

Figure 12: Math scores by school for South Africa

The plots above represent math scores of the students by school ID for England and South Africa, respectively. It is seen that both box-plots have a variation of math scores. Therefore, it is meaningful to use LMM in order to express this variance. For Turkey, the related plot is given in Figure 2.

We report only lme results in this section, since all algorithms give highly similar parameter estimation results.

6.1 Analyses

Before pioneering modelling, we first implement the factor analysis to some scale variables. These are the variables from *ASBM02A* to *ASBM02I* and from *ASBM05A* to *ASBM05I*. The columns give similar meaning about students' mathematical interest. Therefore, the factor analysis makes them a group and represents similar meaning with fewer plausible columns. Now, we have 2 factor scores which are mainly separated. Similar to Chapter 5.2., the first score is named as **emotional**

factor; while the second score is entitled as **mathematical tendency factor**. The factor loadings are shown at Table 12.

Table 12: Factor loadings for all countries

Variable Name	Factor 1 (Emotional Factor)	Factor 2 (Mathematical Tendency Factor)
ASBM02A	0.779	
ASBM02B	0.137	0.467
ASBM02C	0.271	0.422
ASBM02D	0.525	
ASBM02E	0.836	
ASBM02F	0.710	-0.101
ASBM02G	0.726	
ASBM02H	0.743	
ASBM02I	0.775	
ASBM05A	0.544	
ASBM05B		0.718
ASBM05C		0.680
ASBM05D	0.521	
ASBM05E	-0.125	0.641
ASBM05F	0.497	
ASBM05G	0.397	

ASBM05H	0.761
ASBM05I	0.686

After the factor analysis, SES score is created using the indicators related to the economic status of students. The indicators from ASBG05A to ASBG05H are aggregated by row and the derived variable shows SES scores for each student.

Final dataset includes the variables country ID, school ID, student ID, gender, born status in the related country, number of books, emotional factor, mathematical tendency factor, SES score and dependent variable, math score.

6.1.1 Linear Model Analysis

Apart from Chapter 5, the country effect is taken into account as a factor while conducting models, in this chapter. Moreover, the variable number of books is evaluated within family attitude score in Chapter 5. Now, it is evaluated individually because the variables used for family attitude score are not included to the current data due to overwhelming missing values. Additionally, unlike Chapter 5, interaction terms of country variable are added to the model because if they are not, only the intercept term would change from country to country. In the first implementation, the interaction term of country and mathematical tendency factor is statistically insignificant. So, it is dropped from the model. After dropping insignificant variables, the model is built again and the process is named as the second implementation. Output obtained from the second linear model is shown in Table 13.

Table 13: Outputs of LM for country comparison

Variable Name	**Estimate**	**Std. Error**	**Pr(>|t|)**	**Sign.**
Intercept	213.09	6.82	$< 2*10^{-16}$	***
IDCountry (Turkey)	53.55	13.56	$7.90*10^{-05}$	***

IDCountry (England)	279.49	18.42	$< 2*10^{-16}$	***
Gender (2)	-14.57	1.71	$<2*10^{-16}$	***
Born_in_Country (2)	30.34	3.29	$< 2*10^{-16}$	***
Number of Books	4.45	0.81	$3.48*10^{-08}$	***
Emotional Factor	15.28	0.92	$< 2*10^{-16}$	***
Mathematical Tendency Factor	40.06	0.70	$< 2*10^{-16}$	***
SES	11.84	0.47	$< 2*10^{-16}$	***
Turkey*Gender (2)	27.15	3.11	$<2*10^{-16}$	***
England*Gender (2)	15.87	3.51	$6.34*e^{-06}$	***
Turkey*Born_in_Country(2)	34.46	8.77	$8.54*10^{-05}$	***
England*Born_in_Country(2)	-31.26	5.896	$1.16*10^{-07}$	***
Turkey*Num_of_Books	15.51	1.51	$<2*10^{-16}$	***
England*Num_of_Books	15.98	1.497	$<2*10^{-16}$	***
Turkey*Emotional Factor	-15.18	1.75	$<2*10^{-16}$	***
England*Emotional Factor	-13.68	1.64	$<2*10^{-16}$	***
Turkey*SES	-1.42	0.899	0.113	
England*SES	-12.27	1.24	$<2*10^{-16}$	***

All variables except interaction of Turkey and SES score are statistically significant. Due to significancy of interaction of England and SES score, the interaction of countries and SES score is not excluded from the model. Additionally, the standard errors of the country variables seem high. In this model, the country South Africa is taken as reference factor. If the model coefficient has a positive sign, then, the variable has a positive effect on math score of the student. Otherwise, it has a negative effect on math score of the student. Various inferences can be made from this model regarding the relationship between the change in the math score variable and the explanatory variables. For example, in terms of math score, Turkey 53.55 units and England 279.49 units are better than South Africa. Similarly, one-unit

increase in the SES variable results in an 11.84 increase in the math score. On average, male students' success in this field is 14.57 units lower than female students. Also, checking the interactions with countries gives additional results in terms of gender variable. For example, being a male in Turkey provides 27.15 more math score on average than being male in South Africa. Also, even SES scores gives positive coefficient in the model, the interaction of this variable with countries are negative. We can say that one score increases in the SES score in Turkey comparing South Africa gives 1.42 less point in math score.

Additionally, the mean squared error for the linear model is 74.72. Even though, the results of the LM are not bad, the data is more suitable for LMM. As given in figures 2, 11 and 12, we see that math scores by schools have a variety. Moreover, Figure 13 shows residuals obtained from LM by each school. Some boxes are not even touch the line and the data has many outliers for this model. Also, there is a nested data with schools and students. Country factor may be considered as one of the layers. So, the next chapter shows the LMM results and comparisons.

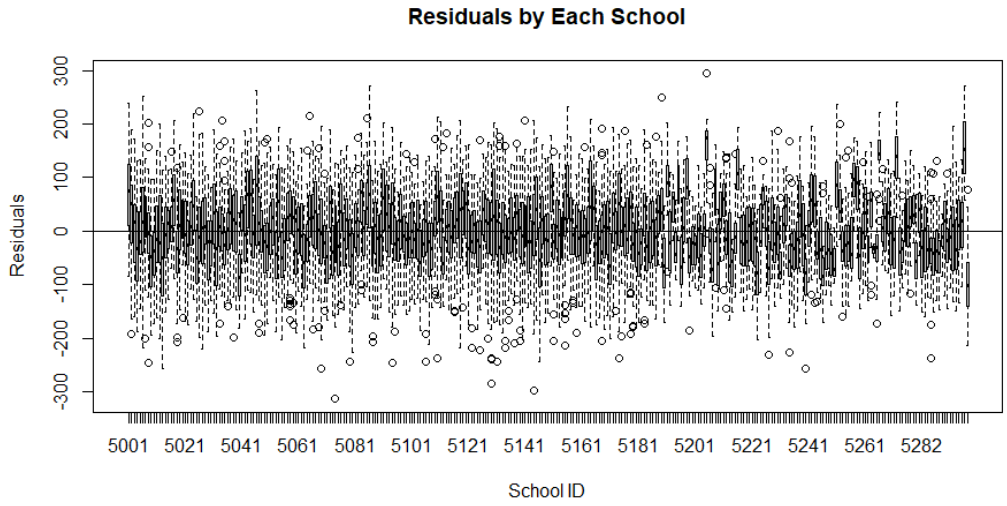

Figure 13: Residuals obtained from LM by each school

6.1.2 Linear Mixed Model Analysis Using Function lme in R

Two different models are conducted to explain mathematics achievement of students for Turkey, England and South Africa. In the model implementation, country is specified like dummy variables, not a random effect. The first model is built with country ID, gender, born status, number of books, emotional factor, mathematical tendency factor, SES score, interactions of the variables and country and random effect school ID. The model syntax in R format is given as

*model=lme(y~factor(IDCNTRY)+factor(Gender)+factor(Born_in_Country)+EmotionalFactor+MathematicalTendencyFactor+ses+factor(IDCNTRY)*factor(Gender)+factor(IDCNTRY)*factor(Born_in_Country)+factor(IDCNTRY)*MathematicalTendencyFactor+factor(IDCNTRY)*EmotionalFactor+factor(IDCNTRY)*ses,random = ~ 1| IDSCHOOL, final, na.action = "na.omit", method = "ML")*

The output of the first model is shown at the table below.

Table 14: Outputs of LMM for country comparison

Variable Name	Estimate	Std. Error	Pr(>|t|)	Sign.
Intercept	296.91	7.09	0.00	***
IDCountry (Turkey)	15.52	12.95	0.231	
IDCountry (England)	214.95	17.04	0.00	***
Gender (2)	-10.98	1.57	0.00	***
Born_in_Country (2)	21.27	3.04	0.00	***
EmotionalFactor	18.06	0.87	0.00	***
MathematicalTendencyFactor	36.38	0.82	0.00	***
SES	6.64	0.46	0.00	***
Turkey*Gender (2)	22.40	2.88	0.00	***
England*Gender (2)	7.46	3.20	0.02	***
Turkey*Born_in_Country	44.37	8.06	0.00	***

England*Born_in_Country	-17.19	5.44	0.002	***
Turkey*MathematicalTendencyFactor	4.91	1.597	0.002	***
England*MathematicalTendencyFactor	11.51	1.89	0.00	***
Turkey*Emotional Factor	-15.72	1.65	0.00	***
England*Emotional Factor	-17.91	1.58	0.00	***
Turkey*SES	3.76	0.82	0.00	***
England*SES	-4.37	1.15	0.00	***

All the variables except country dummy for Turkey are statistically significant for the LMM with a random effect school. Due to other significant term of the variables, country dummy for Turkey is not dropped from the model. However, the variable number of books is statistically insignificant in the first design. Therefore, it is excluded from the model. Additionally, it may be said that the differences between Turkey and South Africa are not statistically significant, because the p-value of dummy variable of Turkey is greater than significance level 0.05. Moreover, standard errors of the country variable are a bit high.

According to table 14, the model outputs say that England has really high math score rather than Turkey or South Africa when all situations remain the same. While being boy in Turkey has a positive effect on math score, being boy in England has a negative effect. Also, one unit increase in mathematical tendency factor in England increases math scores by 47.89 units. The same subject in Turkey increases math scores by 41.29 units. The same unit in South Africa rises math scores by 36.38. Therefore, it can be understood that when all situations are the same, mathematical tendency factor gives the highest score in England and the lowest score in South Africa. On the other hand, it is the opposite for the emotional factor. In South Africa, one unit increase in emotional factor gives the highest increment by 18.06 when all situations for all countries are the same. In Turkey, this value is 2.34. In England, it is equal to 0.15 which is very low. Moreover, when we evaluate the SES score, Turkey ranks first with 10.4 units in terms of giving the highest score when all the

conditions remain the same. South Africa ranks second with 6.64 units. England ranks third with 2.27.

The second LMM is built using the same variables but, this time, random effect is students nested within schools. R syntax of the model is

*model=lme(y~factor(IDCNTRY)+factor(Gender)+factor(Born_in_Country)+EmotionalFactor+MathematicalTendencyFactor+ses+factor(IDCNTRY)*factor(Gender)+factor(IDCNTRY)*factor(Born_in_Country)+factor(IDCNTRY)*EmotionalFactor+factor(IDCNTRY)*ses,random = ~ 1| IDSCHOOL/IDSTUD, final, na.action = "na.omit", method = "ML")*

The output of the second model is presented at Table 15.

Table 15: Outputs of LMM of student nested schools for country comparison

Variable Name	**Estimate**	**Std. Error**	**Pr(>|t|)**	**Sign.**
Intercept	291.69	7.03	0.00	***
IDCountry (Turkey)	23.01	12.84	0.073	**
IDCountry (England)	204.71	16.89	0.00	***
Gender (2)	-11.16	1.56	0.00	***
Born_in_Country (2)	22.87	3.02	0.00	***
Number of Books	8.96	0.56	0.00	***
EmotionalFactor	18.34	0.86	0.00	***
MathematicalTendencyFactor	36.4	0.82	0.00	***
SES	5.61	0.46	0.00	***
Turkey*Gender	22.68	2.86	0.00	***
England*Gender	10.05	3.18	0.002	***
Turkey*Born_in_Country(2)	41.41	7.99	0.00	***
England*Born_in_Country(2)	-20.55	5.39	0.00	***
Turkey*MathematicalTendencyFactor	3.28	1.59	0.039	***
England*MathematicalTendencyFactor	9.54	1.88	0.00	***

Turkey*Emotional Factor	-16.12	1.64	0.00	***
England*Emotional Factor	-18.08	1.56	0.00	***
Turkey*SES	3.06	0.82	0.00	***
England*SES	-4.07	1.14	0.00	***

According to the output above, all the variables used in the model are statistically significant at 5% confidence level. The outputs obtained from the first LMM and the second LMM are very close to each other in terms of the coefficients. However, in the second model, the variable number of books is added to the model. Similar model interpretations can be made as in the LMM with a random effect school.

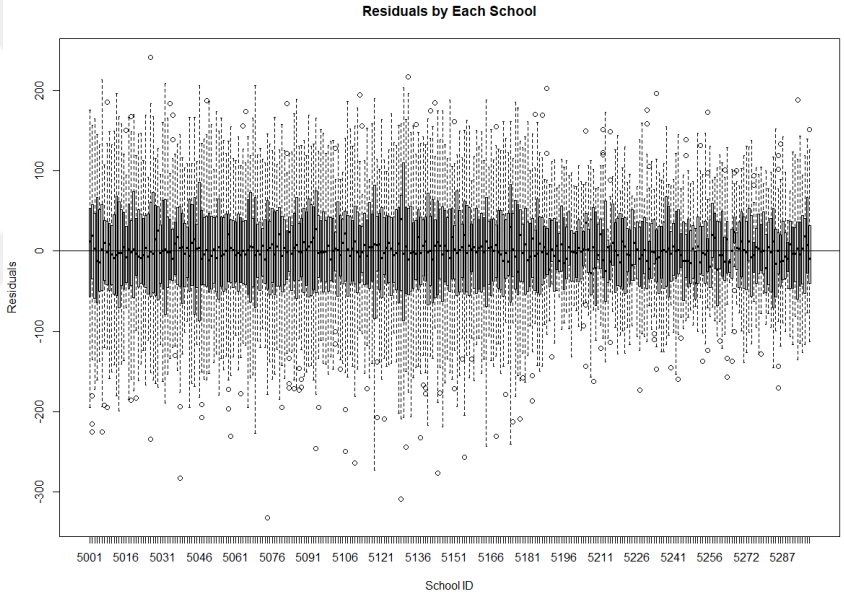

Figure 14: Residuals obtained from the first LMM by each school

The plot above shows residuals which is obtained from LMM by each school. It is seen that all the boxes touch the line and they are ordered regularly. By comparing the plots Figure 13 and 14, it is understood that the residuals from LMM is more reliable. Therefore, LMM is a good method to explain TIMSS 2019.

6.1.3 Model Decision

Up to now, we have one LM and two LMM. At the beginning of the chapter, it is stated that LMM is more suitable for the data because schools have no standard range in terms of math scores and variation is high among them. To make reliable reason for this idea, ANOVA which computes the value subtracting -2 REML loglikelihood from the corresponding value is applied to the models. The candidate models are given below in detail.

Model 1: y ~ covariates

Model 2: y ~ covariates + school as random effect (LMM)

Model 3: y ~ covariates + school as random effect + student nested school as random effect (LMM)

Model	df	AIC	BIC	Sign.
Model 1	20	155993.0	156143.4	✓
Model 2	20	154097.1	154247.5	
Model 2	20	154097.1	154247.5	X
Model 3	22	153845.7	154011.1	

According to the ANOVA test, there is an important difference between model 1 and model 2. Therefore, we continue with the model 2 for the first test result. The second ANOVA test also says that model 2 is enough for this dataset. There is no need to enhance the model with two random effects. So, we retain the second model which includes the covariates and the random effect school.

6.1.4 Model Evaluation

ANOVA says that LMM is more suitable than LM for the dataset TIMSS 2019. We have the model with 8 explanatory variables and one random effect which is school. The plot below presents actual values versus fitted values of the math scores for all countries.

By looking at the plot, it is seen that actual and fitted values are gathered like a line around the center. This means that the model fits good for the dataset.

CHAPTER 7

CONCLUSION / DISCUSSION

Modelling process shows that LMM is a reasonable technique for analyzing TIMSS data because it is a clustered data and the variation in between clusters and within clusters should be counted together in the model. After trying several types of model settings, ANOVA shows us the statistically meaningful model. So, the LMM as a subject based modelling is giving better results than the classical approach.

While building LMM, iterative estimation algorithms help to find parameter estimations. Because of some restrictions, it is not possible to compare all the models. Instead, we compare them within related groups. One of the algorithms is ECME which is a Generalized Expectation-Maximization algorithm for the implementation of LMM. Another method is Fast algorithm which consists of Fisher Scoring and ECME algorithms with faster convergency than ECME algorithm. In the implementations of this study, the function ecmeml has 11 iterations while the function fastml has only 6 iterations for LMM. However, both algorithms require numerical variables and they do not work with categorical variables, for now. Therefore, we assume categorical variables as numerical variables only for comparison aim.

There are several R packages implementing LMM with different algorithms. Although the nlme and lme4 packages are powerful in terms of the richness of the content of the outputs and the flexibility in the model definitions, the package lmm includes algorithms such as ecme and fast is more difficult to use than other packages in this sense. Additionally, the package lme4 includes up to date and effective linear algebra methods. Therefore, it is faster and more efficient than the package nlme. Also, lme4 has wider usage such as generalized linear mixed models and it can easily handle huge number of random effects. On the other hand, nlme provides more

flexibility about modelling. For example, in nlme, it is possible to define the variance-covariance matrix.

In this study, we state general perspectives of LMM and iterative estimation algorithms which are used in LMM. These iterative algorithms are detailed in terms of historical context, formulation and pros-cons. Also, explanations for situations suitable for LMM are practically supplied. In order to illustrate, one of the contemporary datasets, TIMSS 2019 is used. LM and two different types of LMM are built for Turkey implementation and country comparisons with England and south Africa, separately. The variables gender, born in Turkey, immediate area, emotional factor, mathematical tendency factor, SES score, family attitude, and school status seem to be statistically significant for Turkey implementation. Unlike Turkey implementation, the interaction terms are added to the model in country comparisons. According to the obtained result, England among these three countries has the highest mathematic score when other situations are the same and constant. The variables used for all countries are very similar and most of the variables are statistically significant. As a future study, we are planning to extend the model by adding more countries and by including outer variables such as GDP per capita and mean education year for each country.

REFERENCES

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*(2), 155–175.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*(105), 456-466.

Carnoy, M., Khavenson, T., & Ivanova, A. (2015). Using TIMSS and PISA results to inform educational policy: a study of Russia and its neighbours. (45), 248-271.

Conay, M., Khavenson, T., & Ivanova, A. (2013). Using TIMSS and PISA results to inform aducational policy: a study Russia and its neighbours. *Compare: A journal of Comparative and International Education*(45), 248-271.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*(39), 1-38.

Eisenhart, C. (1947). The Assumptions Underlying the Analysis of Variance. *Biometrics*(3), 1-21.

Fishbein, B., Foy, P., & Tyack, L. (2019). *Reviewing the TIMSS 2019 Achievement Item Statistics, Chapter 10.* TIMSS.

Galbraith, S., Daniel, J. A., & Vissel, B. (2010). A Study of Clustered Data and Approaches to Its Analysis. *Journal of Neuroscience*(30), 10601-10608.

Gokalp Yavuz, F., & Arslan, O. (2018). Linear Mixed Model with Laplace Distribution (LLMM). *Statistical Papers*(59), 271-289.

Gronmo, L. S., & Olsen, R. V. (2006). TIMSS versus PISA: The Case of Pure and Applied Mathematics. *2nd IEA International Research Conference.* Washington DC.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*(23), 17-27.

Haladyna, T. M., & Rodriguez, M. C. (2013). Developing and validating test items. *New York: Routledge*.

Henderson, C. R. (1950). Estimation of Genetic Parameters. *Ann. Math. Stat.*(21), 309-310.

Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced Repeated-Measures Models with Structured Covariance Matrices. *Biometrics*(42), 805-820.

Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 963-974.

Laird, N. M., Lange, N., & Stram, D. (1987). Maximum Likelihood Computations with Repeated Measures: Application of the EM Algorithm. *Journal of the American Statistical Association*(82), 97-105.

Langdon, D., McKittrick, G., Beede, D., Khan, B., & Doms, M. (2011). STEM: good jobs now and for the future. *ESA*, *03-11.* Washington DC.

Lindstrom, M. J., & Bates, D. M. (1988, December). Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association*(83), 1014-1022.

Liou, P. Y., & Bulut, O. (2020). The Effects of Item Format and Cognitive Domain on Students' Science Performance in TIMSS 2011. *Research in Science Education*, 99-121.

Liu, C., & Rubin, D. B. (1994). The ECME Algorithm: a Simple Extension of EM and ECM with Faster Monotone Convergence. *Biometrika*(81), 633-648.

Meng, X.-L., & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*(80), 267-278.

Mullis, I., Martin, M., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

OECD (Organization for Economic Cooperation and Development; PISA (Program for International Student Assessment). (2010). *PISA 2009 Results. What Students Know and Can Do: Student Performance in Reading, Mathematics, and Science.* Paris.

Pinheiro, J. C., Liu, C., & Wu, Y. N. (2001). Efficient Algorithms for Robust Estimation in Linear Mixed-Effects Models Using the Multivariate t-Distribution. *Journal of Computational and Graphical Statistics, 10*, 249-276.

Ramirez, M. J. (2006). Understanding the low mathematics achievement of Chilean students: A cross-national analysis using TIMSS data. *International Journal of Educational Research*(45), 102-116.

Schafer, J. L. (1998). Some improved procedures for linear mixed models.

Schmidt, W. H., & Burroughs, N. A. (2016). Influencing public school policy in the United States: the role of large-scale assessments. *Research Papers in Education, 31*, 567-577.

TIMSS. (2019). *Methods and Procedures: TIMSS 2019 Technical Report.*

TIMSS. (2019). *SUPPLEMENT 1: International Versions of the TIMSS 2019 Context Questionnaires.*

TIMSS. (2019). *SUPPLEMENT 2: National Adaptations of the International Context Questionnaires.*

TIMSS. (2019). *SUPPLEMENT 3: Variables Derived from the Student, Home,Teacher, and School Context Data.*

Uysal, S. (2015). Educational Research and Reviews Factors affecting the Mathematics achievement of Turkish students in PISA 2012. *Academic Journals, 10*(12), 1670-1678.

Wang, W., & Fan, T. (2010). ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. *Computational Statistics and Data Analysis*(54), 1328-1341.

West, B. T., & Galecki, A. T. (2011). An Overview of Current Software Procedures for Fitting Linear Mixed Models. *The American Statistician*(65), 274-282.

Wiberg, M. (2019). The relationship between TIMSS Mathematics achievements, grades, and national test scores. *Education Inquiry*(10), 328-343.

Wu, C. (1983). On The Convergence Properties of The EM Algorithm. *The Annals of Statistics, 11*(1), 95-103.