

LONGITUDINAL MODELS FOR LIFE EXPECTANCY AT BIRTH IN
TURKEY

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



BY
ÖZLEM ASKER

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

JANUARY 2021

Approval of the thesis:

**LONGITUDINAL MODELS FOR LIFE EXPECTANCY AT BIRTH IN
TURKEY**

submitted by **ÖZLEM ASKER** in partial fulfillment of the requirements for the
degree of **Master of Science in Statistics, Middle East Technical University** by,

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

Prof. Dr. Ayşen Dener Akkaya
Head of the Department, **Statistics**

Prof. Dr. Özlem İlk Dağ
Supervisor, **Statistics, METU**

Examining Committee Members:

Assoc. Prof. Dr. Ceylan Yozgatlıgil
Statistics, METU

Prof. Dr. Özlem İlk Dağ
Statistics, METU

Prof. Dr. Serpil Aktaş Altunay
Statistics, Hacettepe University

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

Assist. Prof. Dr. Tuğba Adalı
Institute of Population Studies, Hacettepe University

Date: 15.01.2021



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 : Özlem Asker

Signature :

ABSTRACT

LONGITUDINAL MODELS FOR LIFE EXPECTANCY AT BIRTH IN TURKEY

Asker, Özlem
Master of Science, Statistics
Supervisor : Prof. Dr. Özlem İlk Dağ

January 2021, 123 pages

World Health Organization (WHO) gives the definition of life expectancy at birth as “the average number of years that a newborn is expected to live if current mortality rates continue to apply”. Life expectancy at birth is an important indicator of a country's overall health status and its economic and social level of development. In order for governments to set their policies correctly, it is important to determine factors that affect life expectancy at birth and how they affect it. In this thesis, longitudinal data analyses are used to find out the factors that are related with the life expectancy at birth. Specifically, random effects model, marginal model and transition model are used for these analyses. The results suggest that the most suitable model for the data set that is used is marginal model. According to the constructed marginal model, Life Expectancy at Birth is negatively related with Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand) and Infant Mortality Rate (Per Thousand). In addition, when measurement error models are fitted under different assumptions of measurement error levels for these four variables, it is seen that only the significance of Infant Mortality Rate changes.

Keywords: Longitudinal Data, Marginal Models, Measurement Error, Random
Effects Models, Transition Models



ÖZ

TÜRKİYE’DE DOĞUŞTA BEKLENEN YAŞAM SÜRESİ İÇİN UZUNLAMASINA MODELLER

Asker, Özlem
Yüksek Lisans, İstatistik
Tez Yöneticisi: Prof. Dr. Özlem İlk Dağ

Ocak 2021, 123 sayfa

Dünya Sağlık Örgütü doğumda beklenen yaşam süresi tanımını, “mevcut ölüm oranları geçerliiyken bir yenidoğanın yaşaması beklenen ortalama yıl sayısı” olarak vermektedir. Doğumda beklenen yaşam süresi, bir ülkenin genel sağlık durumu ile ekonomik ve sosyal gelişmişlik düzeyinin önemli bir göstergesidir. Devletlerin politikalarını doğru bir şekilde belirleyebilmeleri için, doğumda beklenen yaşam süresini etkileyen faktörleri ve bu faktörlerin doğumda beklenen yaşam süresini nasıl etkilediğini belirlemek önemlidir. Bu tezde, doğumda beklenen yaşam süresiyle ilişkili faktörleri tespit etmek için uzunlamasına veri analizi kullanılmıştır. Spesifik olarak, analiz için rassal etkiler modeli, marjinal model ve geçiş modeli kullanılmıştır. Sonuçlar, kullanılan veri seti için en uygun modelin marjinal model olduğunu göstermektedir. Oluşturulan marjinal modele göre, Doğumda Beklenen Yaşam Süresi; Kaba Doğum Hızı (Binde), Kaba Evlenme Hızı (Binde), Kaba Ölüm Hızı (Binde) ve Bebek Ölüm Hızı (Binde) gibi faktörlerin artışından olumsuz etkilenmektedir. Ayrıca, bu dört değişken için, farklı ölçüm hata seviyesi varsayımları altında, ölçüm hatası modelleri kurulduğu zaman, sadece Bebek Ölüm Hızı'nın anlamlılığının değiştiği görülmektedir.

Anahtar Kelimeler: Geçiş Modeli, Marjinal Model, Ölçüm Hatası, Rassal
Etkiler Modeli, Uzunlamasına Veriler





to my families

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor Prof. Dr. Özlem İlk Dağ for her endless support, patience, understanding, guidance and encouragement through this learning experience. I am very grateful to her for accepting me as her thesis student and providing me to benefit from her profound knowledge.

I would also like to thank to my examining committee members, Assoc. Prof. Dr. Ceylan Yozgatlıgil, Prof. Dr. Serpil Aktaş Altunay, Assist. Prof. Dr. Fulya Gökalp Yavuz and Assist. Prof. Dr. Tuğba Adalı for their valuable suggestions and feedbacks.

I would not be able to graduate with my master's degree without my beloved friends' endless support and guidance. Ozancan Özdemir, Celal Oğuz Özdemir and Ecem Ünal were always there for me when I felt like I was going to give up on everything. They provided me with the support and encouragement I needed. I will always be grateful to them.

I would like to thank my manager, Emine İffet Şahin, for her understanding, patience, encouragement and support through my thesis studies. I would not be able to get my master's degree if she was not that understanding and encouraging. I also want to thank my co-workers Alev İlhan, Zeynep Unutulmaz and Zeynep Nurdağ for being with me at least eight hours a day and easing my stress on the days when I was very stressful.

For their friendship and limitless psychological support, I would like to thank my dearest friends Zeynep Güven and Elif Günal. Whenever I needed their guidance, support and love, they were always with me without asking twice.

I want to thank my mom (Fatoş), dad (Hasan) and my little brother (Emre) for their unconditional love. No matter what, they always stand by me and support me in my decisions. I would also like to thank my second mom (Figen) and dad (Kemal) for

believing me and supporting me all the time through my thesis writing process. I cannot thank them enough.

My fiancé, Berksu, is exactly the source of love in my life. I always felt his endless love and support with me even when he was far away from me. Thank you for loving me, being my family and letting me live my happiest years with you. Thank you so much for introducing me to Cicikuş, also. He has been the joy of my life since the day I met him.



TABLE OF CONTENTS

ABSTRACT	v
ÖZ	vii
ACKNOWLEDGMENTS	x
TABLE OF CONTENTS	xii
LIST OF TABLES	xiv
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS	xvii
CHAPTERS	
1 INTRODUCTION	1
2 LITERATURE REVIEW	3
3 METHODOLOGY	13
3.1 Random Effects Model	13
3.2 Marginal Model	14
3.3 Transition Model.....	15
3.4 Measurement Error Model.....	16
3.5 Data Description	16
4 RESULTS.....	19
4.1 Exploratory Analysis	19
4.2 Results of the Constructed Models	24
4.2.1 Random Effects Models	24
4.2.2 Marginal Models	33
4.2.3 Transition Models.....	38
4.2.4 Models with Dummy Variables.....	43

4.3	Measurement Error Model	44
5	CONCLUSION	49
	REFERENCES	53
APPENDICES		
A.	R Codes	57
B.	Correlation Matrix of Numeric Variables	76
C.	Regression Diagnostic Plots of Final Random Effects Model	80
D.	Results and Regression Diagnostics of Random Effects Models with Dummy Variables	85
E.	Results of Marginal Models that are Constructed before Final Marginal Model	91
F.	Regression Diagnostic Plots of Final Marginal Model	93
G.	Results and Regression Diagnostics of Marginal Models with Dummy Variables	97
H.	Results of Transition Models that are Constructed before Final Transition Model	102
I.	Regression Diagnostic Plots of Final Transition Model	104
J.	Results and Regression Diagnostics of Transition Models with Dummy Variables	106
K.	Scatterplots of Response Variable and Significant Covariates	111
L.	IDs of the 81 Variance Combinations in Measurement Error Part	118
M.	Variance Combinations where Infant Mortality Rate is Found to be Insignificant in Measurement Error Part.....	121

LIST OF TABLES

TABLES

Table 4.1 Summary Statistics of the Variables	19
Table 4.2 Results of the Shapiro Wilk Normality Test Conducted on the Response Variable for Each Year	22
Table 4.3 Correlations between Life Expectancies at Birth in Different Years	22
Table 4.4 Results of Pesaran CD Test for Life Expectancy at Birth for Each Year	24
Table 4.5 Estimated Regression Coefficients for the Initial Random Effects Model	25
Table 4.6 VIF Values of Variables	26
Table 4.7 Estimated Regression Coefficients for the Reduced Random Effects Model	27
Table 4.8 Estimated Regression Coefficients for the Final Random Effects Model	28
Table 4.9 Estimated Regression Coefficients for the Random Effects Model with Year in the Random Part	32
Table 4.10 Estimated Regression Coefficients for the Final Marginal Model with Exchangeable Correlation Structure	34
Table 4.11 Estimated Regression Coefficients for the Final Marginal Model with AR-1 Correlation Structure	34
Table 4.12 Estimated Regression Coefficients for the Final Transition Model	39
Table B.1 Correlation Table of Numeric Variables	76
Table D.2 VIF Values of the Variables in Random Effects Model with Dummy Variables	86
Table D.3 Estimated Regression Coefficients for the Reduced Random Effects Model with Dummy Variables	86
Table D.4 Estimated Regression Coefficients for the Final Random Effects Model with Dummy Variables	87

LIST OF FIGURES

FIGURES

Figure 4.1. Histogram of Life Expectancy for Each Year	21
Figure 4.2. Scatterplot of Life Expectancy at Birth versus Time (Year)	23
Figure 4.3. Scatterplot of Standardized Residuals of Random Effects Model versus Time (Year)	29
Figure 4.4. Scatterplot of Fitted Response Values of Random Effects Model versus Life Expectancy at Birth	31
Figure 4.5. Scatterplot of Standardized Residuals of Marginal Model versus Time (Year)	36
Figure 4.6. Scatterplot of Fitted Response Values of Marginal Model versus Life Expectancy at Birth	37
Figure 4.7. Scatterplot of Standardized Residuals of Transition Model versus Time (Year)	40
Figure 4.8. Scatterplot of Fitted Response Values of Transition Model versus Life Expectancy at Birth	41
Figure 4.9. Scatterplot of Estimated Intercept Values versus IDs of Variance Combinations	46
Figure 4.10. Scatterplot of Estimated Coefficients of Crude Birth Rate versus IDs of Variance Combinations	46
Figure 4.11. Scatterplot of Estimated Coefficients of Crude Marriage Rate versus IDs of Variance Combinations	47
Figure 4.12. Scatterplot of Estimated Coefficients of Crude Death Rate versus IDs of Variance Combinations	47
Figure 4.13. Scatterplot of Estimated Coefficients of Infant Mortality Rate versus IDs of Variance Combinations	48
Figure C.1. Scatterplot of Standardized Residuals of Random Effects Model versus Crude Birth Rate	80
Figure C.2. Scatterplot of Standardized Residuals of Random Effects Model versus Crude Marriage Rate	81

Figure C.3. Scatterplot of Standardized Residuals of Random Effects Model versus Crude Death Rate	82
Figure C.4. Scatterplot of Standardized Residuals of Random Effects Model versus Infant Mortality Rate	83
Figure C.5. Scatterplot of Standardized Residuals of Random Effects Model versus Number of Students Per Classroom	84



LIST OF ABBREVIATIONS

ABBREVIATIONS

AF: Agricultural Field

CBR: Crude Birth Rate (Per Thousand)

CDR: Crude Death Rate (Per Thousand)

CMR: Crude Marriage Rate (Per Thousand)

IMR: Infant Mortality Rate (Per Thousand)

LEAB: Life Expectancy at Birth

NM: Net Migration

NMR: Net Migration Rate

NOCH: Number of Cinema Halls

NOFHHS: Number of First Hand Housing Sales

NOHS: Number of Housing Sales

NOSHHS: Number of Second Hand Housing Sales

NOSPC: Number of Students Per Classroom

NOTH: Number of Theater Halls

PD: Population Density

CHAPTER 1

INTRODUCTION

According to the definition of World Health Organization (2006), life expectancy at birth is “the average number of years that a newborn is expected to live if current mortality rates continue to apply”. In other words, it reflects the overall mortality level of a population. Thus, life expectancy at birth gives an insight on health statuses of countries, since the higher life expectancy means the better health status for a country.

One of the main purposes of governments is maintaining and improving the health status of their citizens. To contribute to the health statuses of countries, factors that have positive and negative effects on the health of populations should be determined and governments should make policies accordingly. Thus, factors that are related with the life expectancy at birth should be determined in order to contribute to the health of populations.

Several economic, social and environmental factors are expected to be related to the life expectancy at birth. The aim of this research is determining some of these factors and understanding how much effect they have on the life expectancies of 81 provinces of Turkey.

In medical field, longitudinal data is one of the most commonly used data types. In this thesis, analyzes are done on a longitudinally collected data set for the 81 provinces of Turkey. The data set is consolidation of a couple of different data sets, which are retrieved from the website of Turkish Statistical Institute. It includes life expectancies at birth and several different economic, social and environmental attributes (which may be related with the life expectancy at birth) of 81 provinces of Turkey.

Many different economic, social and environmental factors that were not used in previous research are used in this research. Moreover, unlike the previous research in this field, life expectancies at birth are examined based on the provinces of Turkey in this thesis. Random effects, marginal and transitional models are constructed on the data set. In addition, since there may be measurement errors in health and socio-economic variables, measurement error models are fitted.

Organization of the thesis is specified as follows. Chapter 2 includes information about literature review of research done both around the world and in Turkey about life expectancy. Chapter 3 covers the explanations of random effects, marginal, transition and measurement error models, which are used in the analysis of the data set and gives data description. Chapter 4 gives exploratory analysis of the data set, results of random effects, marginal and transition models that are constructed on data and the investigation of measurement errors in covariates. In Chapter 5, summary of the results and conclusion of the analysis are given.

CHAPTER 2

LITERATURE REVIEW

In this chapter, summaries and discussions of some of the research done both around the world and in Turkey about life expectancy are given.

Asiskovitch (2010) investigated the effect of healthcare systems on the life expectancies of women and men in 19 OECD (Organization for Economic Co-operation and Development) countries. OECD Health Data 2009 was used in the analysis. The data set contains health and healthcare indicators of 30 member states of OECD for the period 1960-2008. Since the data set has some missing values for some years and countries, analysis was done for the period 1990-2005 and for 19 countries, namely Australia, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and United States. Some countries were not included to the analysis due to their regimes and sizes.

Multilevel panel data regression analysis was used in the analysis. The dependent variables are life expectancies at birth and at age 65 for women and life expectancies at birth and at age 65 for men. The independent variables and their measurement units are listed below.

- GDP: US\$ 2000 purchasing power parity rates
- Public and mandatory private expenditure for social policies, excluding healthcare (old-age, survivors, incapacity-related, unemployment benefits, active labor market programs, family benefits and services, and housing policies): % of GDP per capita
- National funding for healthcare: % of GDP per capita

- Public funding for healthcare: % of GDP per capita
- Private funding for healthcare: % of GDP per capita
- Prevalence of smoking among women: % of females 15+ who are daily smokers
- Prevalence of smoking among men: % of males 15+ who are daily smokers
- Prevalence of overweight or obesity among women: % of relevant population
- Prevalence of overweight or obesity among men: % of relevant population
- Nitrogen oxide emissions (air quality): Kilograms per capita
- Level of education in a given society: Expected years of schooling, entire population

No multicollinearity problem was found between these independent variables. Four sets of four panel data regressions were constructed for the analysis, for life expectancies at birth and at age 65 for women and life expectancies at birth and at age 65 for men. In the first two sets, national funding for healthcare was used, and the level of education variable was included in one set, while it was excluded in the other set. In the second two sets, public and private funding for healthcare were used, and the level of education variable was included in one set, while it was excluded in the other set. The results of the models in the first set showed that the coefficients of national funding for healthcare were negative and insignificant for both women and men for the models where the response variable was life expectancy at birth. However, the coefficients of national funding for healthcare were positive and significant for both women and men for the models where the response variable was life expectancy at age 65. The results of the models in the second set showed that the coefficients of public funding for healthcare were positive but insignificant for both women and men for the models where the response variable was life expectancy at birth, while it was positive and significant

when the life expectancy at age 65 is investigated. When the response variable was life expectancy at age 65, the coefficients of private funding were significant only if level of education was not in the model and the coefficients were positive for both women and men.

Zha et al. (2019) investigated the effects of 13 factors of socio-economy, ecological environment, geological environment and geographical environment on life expectancies of residents in Tibet in 2010. The results of the analysis showed that gross regional product, tectonic division, vegetation regionalization, annual average solar radiation, vegetation type and gross regional product per capita variables had important effects on life expectancy in Tibet. However, stratum lithology, annual average temperature and annual average precipitation affected life expectancy mildly. Moreover, moisture index, geomorphic type, elevation and soil type had little effects on life expectancy.

Lin et al. (2012) examined four socioeconomic and political factors' contributions in increasing life expectancy at birth in less developed countries between 1970 and 2004. The list of factors that were used and their explanations are listed below.

- Political Regime: "Measured by the regime score from the Polity IV database"
- Economy: "Measured by log10 gross domestic product per capita at purchasing power parity"
- Nutritional Status: "Measured by the proportion of undernourished people in the population"
- Educational Environment: "Measured by the literacy rate of the adult population aged 15 years and over"

The data set used in the analysis was consisted of annual data between the years 1970 and 2004 of 119 less developed countries on life expectancy at birth and on the four factors that are listed above. Linear mixed models were used for the analysis. With a lag of 0 to 10 years, longitudinal effects of four factors on life expectancy at birth was examined. Adjustments were done on regional and time

correlations. All the four factors were found to be positively related with life expectancy at birth.

Delavari et al. (2016) investigated some key socioeconomic factors' effects on life expectancy at birth in Iran between 1985 and 2013. The list of socioeconomic factors that were used in the analysis are listed below.

- Total Fertility Rate
- GDP Per Capita
- Food Availability
- Literacy Rate
- Number of Doctors Per 10,000 Population
- CO_2 Emission
- Degree of Urbanization
- Inflation Rate

Time series analysis was used for the analysis. Findings showed that GDP per capita, food availability, literacy rate and number of doctors per 10,000 population had positive effects on life expectancy at birth, total fertility rate had negative effect on life expectancy at birth, CO_2 emission, degree of urbanization and inflation rate had no effects on life expectancy at birth.

Bayati et al. (2013) examined factors that affect life expectancy at birth in Eastern Mediterranean Region. Based on literature review, they determined explanatory variables as GDP, health expenditure, food production index (food availability), percentage of total population over age 15 that has been employed, education level (education index of Human Development Index), the percentage of one-year olds immunized with Measles-Containing Vaccine, percentage of the urban population and CO_2 emission. The results showed that GDP per capita, food production index, percentage of total population over age 15 that has been employed, education level and percentage of urban population were significantly and positively related with life expectancy at birth, while health expenditure, the percentage of one-year olds

immunized with Measles-Containing Vaccine and CO_2 emission were not significantly related with life expectancy at birth.

Bilas et al. (2014) wanted to investigate factors that affect life expectancy at birth in 28 European Union countries from 2001 to 2011. The list of factors that were used in the analysis are listed below.

- GDP Per Capita
- GDP Growth Rate
- Education Enrolment
- Level of Education Attained
- Population Growth Rate

Panel data analysis approach was used for the analysis. Findings showed that only GDP per capita and level of education attained affected life expectancy at birth. GDP per capita had positive effect and level of education attained had negative effect on life expectancy at birth.

Shahbaz et al. (2015) examined the determinants of life expectancy from 1972-2012 (in the presence of economic misery). The list of factors that were used in the analysis are listed below.

- Public Health Spending Per Capita
- Food Supply Proxies by Agricultural Food Supply Index
- Economic Misery Proxies by Interaction between Inflation and Unemployment
- Illiteracy Rate
- Urbanization Per Capita

Time series analysis was used for the analysis. Findings showed that public health spending, food supply and urbanization had positive effects on life expectancy at birth, economic misery had negative effect on life expectancy at birth, illiteracy had no effects on life expectancy at birth.

Kröger et al. (2015) discussed random measurement error in subjective health and socio-economic status variables when applying a cross-lagged panel design. It was found from literature reviews that in subjective health there is a measurement error of 25-65%, and in socio-economic variables there is a measurement error of 5-30%. It was also found from the analysis that the coefficients of the variables with higher measurement error were higher than the coefficients of other variables. For solution, sensitivity analysis, multiple imputation for measurement error, measurement models, instrumental variables etc. were suggested. An example of a Bayesian model, which takes measurement error into account, was given as a more recently developed solution to the problem. The Bayesian model was applied to a simulated data which had measurement error problem. The results of that analysis showed that taking measurement error into account changed the results significantly. Thus, it was noted that if the measurement error is not taken into account in the analyses when health and socio-economic variables are used, the results will be significantly biased in a direction that is unknown.

Huque et al. (2016) argued that spatial data, which is a common data type in health research, often have measurement error. Thus, coefficients of spatially defined covariates will be biased in the analysis if the measurement error is ignored. Measurement error in the covariates leads to the false acceptance of the null hypothesis (the hypothesis that coefficients are zero) when significance of the coefficients of covariates are tested. It was also tested using simulation study if the problem of having biased coefficient existed or not when covariates had measurement error. The results of that simulation study confirmed the problem. For solution, semiparametric regression approach was suggested. The semiparametric bias correction method applied on the simulated data worked well. The method was also illustrated on Ischemic Heart Disease data.

Bayın (2016) investigated the factors affecting life expectancies of women and men at birth and at age 65 (since in developed countries, most of the deaths occur after the age 65). For the analysis, a data set which consisted of life expectancies of 34 OECD member countries at birth and at age 65 in 2013 (the response variable) and

the related covariates were used. Covariates were determined by using literature reviews and health-related statistical data for OECD countries. The related covariates were categorized as economic, demographic, health resources, health service utilization, perceived health status and urbanization. In order to determine the effect of the related variables on life expectancy of women and men, multiple linear regression analysis was used on complete data.

Covariates used in the research and their explanations are listed below.

- Perceived health status for women and men: Percentage of people over age 15 who perceive their health status as good / very good (data are taken from OECD)
- Number of patient beds: Per 1000 people (data are taken from OECD)
- Per capita income: In terms of USD (data are taken from OECD)
- Per capita health expenditure: In terms of USD (data are taken from OECD)
- Mother mortality rate: Per 100,000 live births (data are taken from WHO)
- Baby mortality rate: Per 1,000 live births (data are taken from WHO)
- Number of doctor visits: Average number of doctor visits per person per year (data are taken from OECD)
- Number of days of hospital stay: Annual average number of days of hospital stay (data are taken from OECD)
- Drug consumption: In terms of USD (data are taken from OECD)
- Urban population rate: Percentage of people living in urban areas (data are taken from WHO)

Research results showed that baby mortality rate was the variable that had the most significant effect on life expectancy at birth both for women and for men. Life expectancies at birth for both women and men were negatively related with baby

mortality rate. In fact, it was found that baby mortality rate was the only variable that was related with the life expectancy at birth for women. In addition to the baby mortality rate, it was found that urban population rate and perceived health status had positive effect on life expectancy at birth for men.

Moreover, they showed that per capita health expenditure was the variable that had the most significant and positive effect on life expectancy at age 65 both for women and for men. In addition to the per capita health expenditure, it was found that drug consumption had negative effect on life expectancy at age 65 both for women and men. In addition, it was found that the number of days of hospital stay had positive effect on life expectancy at age 65 for women, and urban population rate had positive effect on life expectancy at age 65 for men.

Teker et al. (2012) determined the socio-economic factors related with the life expectancy of women and men. Covariates were determined using literature reviews. The analysis was done on the data set which consisted of variables dated from 1995 to 2009. Unit root test, cointegration test and error correction model were applied to the data set. Life expectancies for women and men were examined separately like it was in the research of Bayın (2016), so there were two different response variables.

Covariates used in the research are listed below.

- Dependence rate of elderly population to the young population
- Number of beds per 1,000 patients
- Ratio of health expenses to national income
- Number of doctors per 1,000 patients

The relationship between covariates and the life expectancy of women and men were examined separately by using regression analysis. Regression results showed that there were significant and strong relationships. In addition, all covariates had similar effects on both of the response variables.

It was found that there were long-term relationships between factors since all factors were found to be integrated of order 1. In addition, the results of vector error correction models showed that, for both women and men, there was a negative relationship between life expectancy and dependence rate of elderly population to the young population, and there were positive relationships between life expectancy and other covariates. For men, dependence rate of elderly population to the young population had the most powerful effect on life expectancy. For women, number of doctors per 1,000 patients had the most powerful effect on life expectancy.

Teker et al. (2012) found that dependence rate of elderly population to the young population, number of beds per 1,000 patients, ratio of health expenses to national income and number of doctors per 1,000 patients had effects on life expectancy for women and men by conducting time series analysis on data. Bayın (2016) extended the research by dividing the response variable into four, as life expectancy of women and men at birth and at age 65. She found that baby mortality rate had effect on life expectancy at birth for women, while baby mortality rate, urban population rate and perceived health status had effect on life expectancy at birth for men. Moreover, per capita health expenditure, drug consumption and number of days of hospital stay had effect on life expectancy at age 65 for women, whereas per capita health expenditure, drug consumption and urban population rate had effect on life expectancy at age 65 for men by conducting multiple linear regression analysis on data. Two of the factors in data sets of two research were similar. Teker et al. (2012) found that there was a positive relationship between number of beds per 1,000 patients and life expectancy for both women and men, whereas, Bayın (2016) found no relationship between life expectancy at birth and at age 65 and number of patient beds for both of the genders. Teker et al. (2012) found that there was a positive relationship between ratio of health expenses to national income and life expectancy for both women and men. Likewise, Bayın (2016) found that there was a positive relationship between life expectancy at age 65 and per capita health expenditure for both genders.

In this thesis, many different covariates (several economic, social and environmental factors) will be examined, and the focus will be on life expectancy at birth as the response variable. Panel data analysis will be applied on the data set (which is one of the most preferable analysis techniques in medical field). Life expectancies will be examined based on the 81 provinces of Turkey. In addition, as Kröger et al. (2015) and Huque et al. (2016) discussed in their research, there may be measurement error in health, socio-economic and spatial variables. Thus, the covariates used in this thesis will also be examined for the possible existence of measurement error.



CHAPTER 3

METHODOLOGY

Panel data is a type of data which includes information on same individuals over different time points. Panel data distinguishes from other data types (time series data and cross-sectional data) by its features. It is different from cross-sectional data due to the dependence structure among repeated observations. Moreover, the dimension of panel data is different from the dimension of the time series data: there are usually more subjects, but fewer time points in panel data. Panel data contains information about both individual behaviors and common behaviors of groups. It is widely used in econometrics, social science, epidemiology and health statistics. In this thesis, the most common three panel data models are used for analysis. These three models, namely random effects model, marginal model and transition model, are briefly reviewed in this chapter. Finally, a model with possible measurement error in the covariates is discussed in the last section. In the analysis; MASS, plm and lme4 packages are used in the construction part of random effects models, car package is used for checking multicollinearity between covariates by getting Variance Inflation Factor (VIF) values of the covariates, MuMIn package is used for getting QIC values, Metrics package is used for getting Root Mean Square Error (RMSE) values of models, gee package is used in the construction part of marginal and transition models, swgee package is used for checking possible measurement error in covariates. plm package is also used for checking cross-sectional dependence in the data set.

3.1 Random Effects Model

Hedges (2012) defines random effects models as “statistical models in which some of the parameters (effects) exhibit some form of random variation”. Some

parameters of random effects model are allowed to differ among individuals, providing heterogeneity between subjects. Hence, the variance of the response variable varies from one subject to another. Higher variance of the response variable indicates larger differences between subjects. The most common and one of the simplest random effects model is $y_{it} = \beta_0 + \beta_1 * x_{it} + b_{i0} + b_{i1} * x_{it} + \varepsilon_{it}$, where y is the dependent variable, x is the covariate, ε is the error term, β_0 and β_1 are the coefficients for non-random terms, b_{i0} and b_{i1} are the random coefficients which vary from one individual to another, i is the indicator for individuals ($i=1, 2, \dots, N$), t is the indicator for time points ($t=1, 2, \dots, T$). Here, ε_{it} is normally distributed with mean 0 and variance σ^2 , and b_i is normally distributed with mean 0 and variance σ_b^2 . Even a simpler model, a random intercept model, allows only the intercept to be random and be subject specific.

3.2 Marginal Model

Marginal model is the most commonly used panel data model. This model considers within subject correlation and models dependent variables as a function of covariates. It focuses on differences between subgroups and groups, but not individuals (unlike the random effects model). The most common and the simplest marginal model is $y_{it} = \beta_0 + \beta_1 * x_{it} + \varepsilon_{it}$, where y is the dependent variable, x is the covariate, ε is the error term, β_0 and β_1 are the coefficients, i is the indicator for individuals, t is the indicator for time points. Unlike cross sectional data models, responses in different time points for the same subject are allowed to be dependent in marginal models. Thus, variances and covariances of error terms in marginal models form a $T \times T$ matrix ($\Sigma_{T \times T}$). Covariance among error terms is not expected to be 0, which is the case in cross sectional data models. There are commonly used possible structures for $\Sigma_{T \times T}$ as exchangeable, unstructured, toeplitz and AR(1) (autoregressive, order 1). Examples of those possible $\Sigma_{T \times T}$ structures are given below for $T=4$.

Exchangeable:

$$\begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ & & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ & & & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

Unstructured:

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_{22}^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_{33}^2 & \sigma_{34} \\ & & & \sigma_{44}^2 \end{bmatrix}$$

Toeplitz:

$$\sigma^2 \begin{bmatrix} 1 & \sigma_1 & \sigma_2 & \sigma_3 \\ & 1 & \sigma_1 & \sigma_2 \\ & & 1 & \sigma_1 \\ & & & 1 \end{bmatrix}$$

AR(1):

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$$

3.3 Transition Model

Transition model uses data from past in order to make inferences about data set. Transition model models dependent variable at time t as a function of independent

variables at time t and dependent variables at former time points. The most common and the simplest (using only lag 1 of the response variable) transition model is $y_{it} = \beta_0 + \beta_1 * x_{it} + \alpha_1 * y_{i,t-1} + \varepsilon_{it}$, where y is the dependent variable, x is the covariate, ε is the error term, β_0 and β_1 are the coefficients, α_1 is the coefficient for the lagged part of the model and measures the association between the present and past, i is the indicator for individuals, t is the indicator for time points. The estimate of α_1 is expected to be positive and usually significant. However, the higher order lags are not always statistically significant.

3.4 Measurement Error Model

Measurement error is defined as the difference between observed values of the covariates (or the response) and their true values. Carroll et al. (2006) claims that “covariate measurement error frequently arise in longitudinal studies and they present considerable challenges in statistical inference about such data”. A marginal model with measurement error in the covariate can be defined as $y_{it} = \beta_0 + \beta_1 * x_{it} + \varepsilon_{it}$, where $x_{it} = w_{it} - e_{it}$. Here, w is the observed value of the covariate x . In other words, although the true value of the covariate is x , we can only observe it with an error, e . As usual, i is the indicator for individuals, t is the indicator for time points. The e is normally distributed with mean equal to 0 and covariance matrix $\sum_e$ (it is a $p \times p$ matrix, where p is the number of covariates). Magnitudes of variance terms in $\sum_e$ give the degrees of measurement errors in each variable.

3.5 Data Description

The data set is a combination of a couple of data sets which are retrieved from the website of Turkish Statistical Institute. Data set consists of 19 different attributes of the 81 provinces in Turkey, from 2013 to 2015. Those attributes are; Life Expectancy at Birth (response variable), Population Density, Population, Number

of First Hand Housing Sales, Number of Second Hand Housing Sales, Number of Housing Sales, GDP, Net Migration, Net Migration Rate, Number of Theater Halls, Number of Cinema Halls, Agricultural Field, Export (TL), Import (TL), Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand), Infant Mortality Rate (Per Thousand) and Number of Students Per Classroom. Among the variables; Number of Theater Halls and Number of Cinema Halls are considered as proxies of expenses which are made on cultural activities. Moreover, Number of First Hand Housing Sales, Number of Second Hand Housing Sales and Number of Housing Sales are considered as proxies of expenses which are made on real estate sector. Thus, it can be said that those variables are indicators of welfare levels of provinces. Some explanations on some of the variables are given below.

- Population Density: Population per unit area (population divided by total area)
- GDP: Gross Domestic Product (total value of goods produced and services provided in a country during a specified year)
- Net Migration: Difference between the number of immigrants and the number of emigrants throughout a specified year
- Net Migration Rate: Net migration through a specified year divided by the estimated mid-year population and multiplied by 1,000
- Agricultural Field: Specified in decares
- Crude Birth Rate (Crude Birth Rate): Number of live births during a specified year divided by the estimated mid-year population and multiplied by 1,000 (according to the definition given by the website of OECD Glossary of Statistical Terms on 18th of April 2013, the Crude Birth rate is “the number of live births occurring among the population of a given geographical area during a given year, per 1,000 mid-year total population of the given geographical area during the same year”)

- Crude Marriage Rate (Crude Marriage Rate): Number of marriages occurred during a specified year divided by the estimated mid-year population and multiplied by 1,000 (according to the definition given by the website of OECD Glossary of Statistical Terms on 18th of April 2013, the Crude Marriage rate is “the number of marriages occurring among the population of a given geographical area during a given year, per 1,000 mid-year total population of the given geographical area during the same year”)
- Crude Death Rate (Crude Death Rate): Number of deaths occurred during a specified year divided by the estimated mid-year population and multiplied by 1,000 (according to the definition given by the website of OECD Glossary of Statistical Terms on 18th of April 2013, the Crude Death rate is “the number of deaths occurring among the population of a given geographical area during a given year, per 1,000 mid-year total population of the given geographical area during the same year”)
- Infant Mortality Rate: according to the definition given by the website of OECD Glossary of Statistical Terms on 4th of January 2006, the infant mortality rate is “the number of deaths under one year of age occurring among the live births in a given geographical area during a given year, per 1,000 live births occurring among the population of the given geographical area during the same year”.
- Number of Students Per Classroom: Represents only the number of primary school students

2015 observation of Export variable is missing for Kars. This value is imputed by taking the average of 2013 and 2014 observations of Export variable of this province.

CHAPTER 4

RESULTS

In this chapter, information on variables in the data set, results of exploratory analysis of data set, results of constructed models and measurement error models for some of the covariates are given.

4.1 Exploratory Analysis

Some descriptive statistics and plots are given in this section.

Table 4.1 Summary Statistics of the Variables

	<i>LEAB</i>	<i>PD</i>	<i>Population</i>	<i>NOFHHS</i>	<i>NOSHHS</i>
Min.	74.90	11.50	75,620	91	18
1st Qu.	77.50	46.75	284,142	1,506	1,208
Median	78.10	60.10	512,873	2,898	2,532
Mean	78.19	122.96	959,279	6,870	7,994
3rd Qu.	78.80	110.05	986,071	7,193	7,418
Max.	80.70	2,820.90	14,657,434	112,491	130,936

	<i>NOHS</i>	<i>GDP</i>	<i>NM</i>	<i>NMR</i>	<i>NOTH</i>
Min.	138	1,116,852	-20,084	-45.600	0
1st Qu.	2,860	5,288,962	-4,764	-9.750	2
Median	5,368	8,715,419	-1,167	-2.800	4
Mean	14,864	25,484,883	0	-3.104	8.263
3rd Qu.	15,439	20,855,202	2,057	3.650	7
Max.	239,767	722,567,040	66,321	42.200	224

Table 4.1 (continued)

	<i>NOCH</i>	<i>AF</i>	<i>Export (TL)</i>	<i>Import (TL)</i>	<i>CBR</i>
Min.	0	129,294	1.094×10^4	4.469×10^3	10.20
1st Qu.	4	1,042,063	1.001×10^8	6.521×10^7	12.60
Median	10	2,504,902	4.653×10^8	1.928×10^8	14.80
Mean	27.28	2,949,788	4.224×10^9	5.654×10^9	16.73
3rd Qu.	25.50	3,655,100	1.639×10^9	1.428×10^9	19.05
Max.	778	19,239,668	2.093×10^{11}	3.206×10^{11}	34.70

	<i>CMR</i>	<i>CDR</i>	<i>IMR</i>	<i>NOSPC</i>
Min.	5.760	2.800	3.80	13
1st Qu.	6.905	4.400	8.25	18
Median	7.650	5.900	9.70	21
Mean	7.606	5.774	10.53	22.88
3rd Qu.	8.140	6.900	12.50	27
Max.	10.490	10.000	24.60	46

Table 4.1 gives minimum, 1st quartiles, median, mean, 3rd quartiles and maximum values of the numerical variables in the data set. Mean and median values of Life Expectancy at Birth being approximately equal shows that Life Expectancy at Birth may have symmetrical distribution. Life Expectancy at Birth values range between 74.90 and 80.70 with an average of 78.19. There is a big difference between the most crowded province's population and the least crowded province's population (14,581,814 people). Dispersion of the values of second hand housing sales is greater than the dispersion of the values of first hand housing sales. GDP values range between 1,116,852 and 722,567,040 with an average of 25,484,883. Some provinces do not have cinema or theater halls. Export (TL) values range between 1.094×10^4 and 2.093×10^{11} with an average of 4.224×10^9 and Import (TL) values range between 4.469×10^3 and $3.206 \times$

10^{11} with an average of 5.654×10^9 . Mean value of Crude Birth Rate being greater than mean value of Crude Death Rate means that Population has an increasing trend in general. Mean value of Import being greater than mean value of Export means there are more import activities being held than export activities in the country. In addition, it is seen from the data set that young age structure leads to higher crude marriage rates, lower crude death rates and higher crude birth rates for provinces.

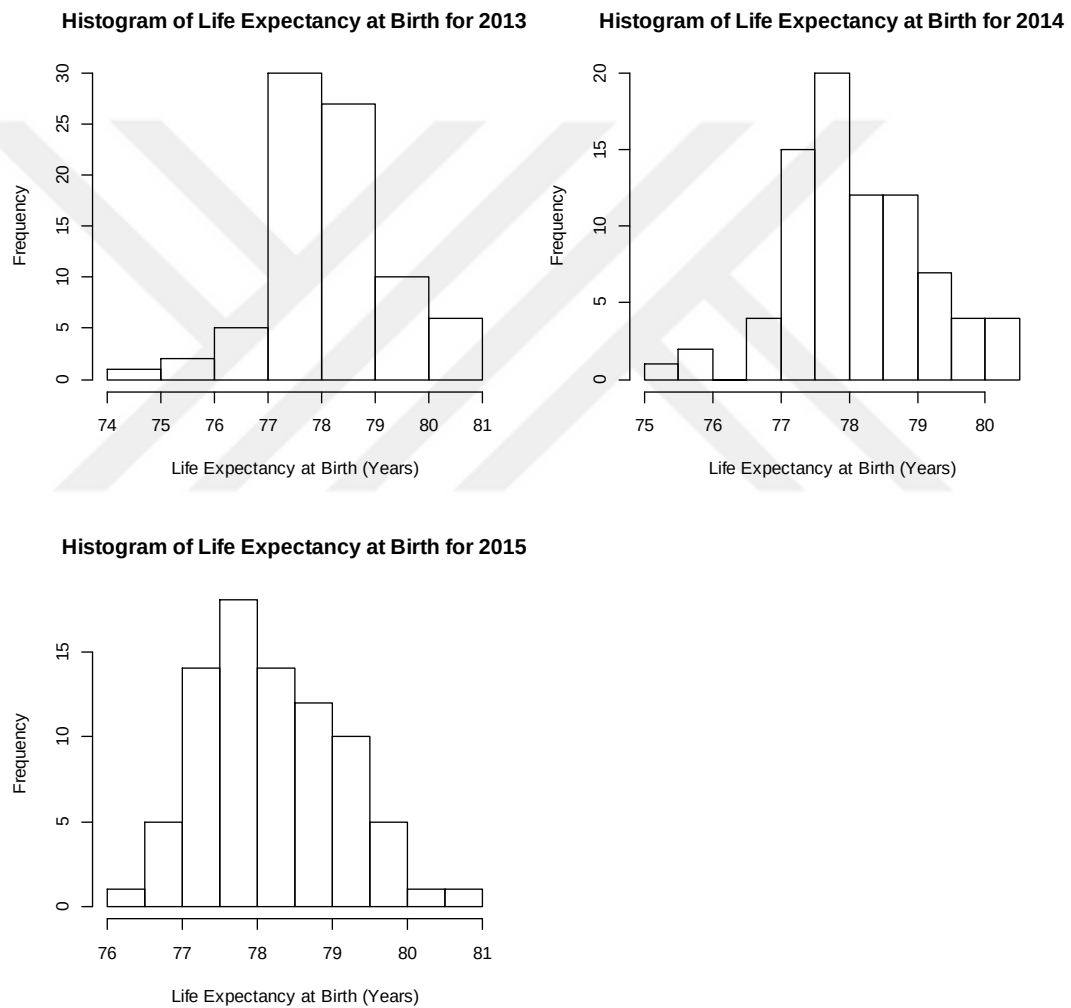


Figure 4.1. Histogram of Life Expectancy for Each Year

Figure 4.1 shows that shapes of the histograms of the response variable have approximately normal distribution. To strengthen our check of the assumption that

the variable comes from normal distribution, Shapiro Wilk Normality Test is conducted on response variable for each year. Results of the test are given in Table 4.2. From these results, it can be seen that response variable does not depart from normal distribution in any year.

Table 4.2 Results of the Shapiro Wilk Normality Test Conducted on the Response Variable for Each Year

Year	<i>W Statistic</i>	<i>p-value</i>
2013	0.973	0.088
2014	0.97	0.054
2015	0.986	0.536

In order to check the relationship of life expectancies at birth in different years, a correlation table is given below. As it can be seen in Table 4.3, the correlations between life expectancies at birth in different years are extremely high. Moreover, as it can be seen from the scatterplot in Figure 4.2., on the average, the response variable does not change with time. Thus, we can infer that the life expectancy at birth values in three years are close to each other.

Table 4.3 Correlations between Life Expectancies at Birth in Different Years

	<i>LEAB in 2013</i>	<i>LEAB in 2014</i>	<i>LEAB in 2015</i>
LEAB in 2013	1	0.98	0.87
LEAB in 2014	0.98	1	0.90
LEAB in 2015	0.87	0.90	1

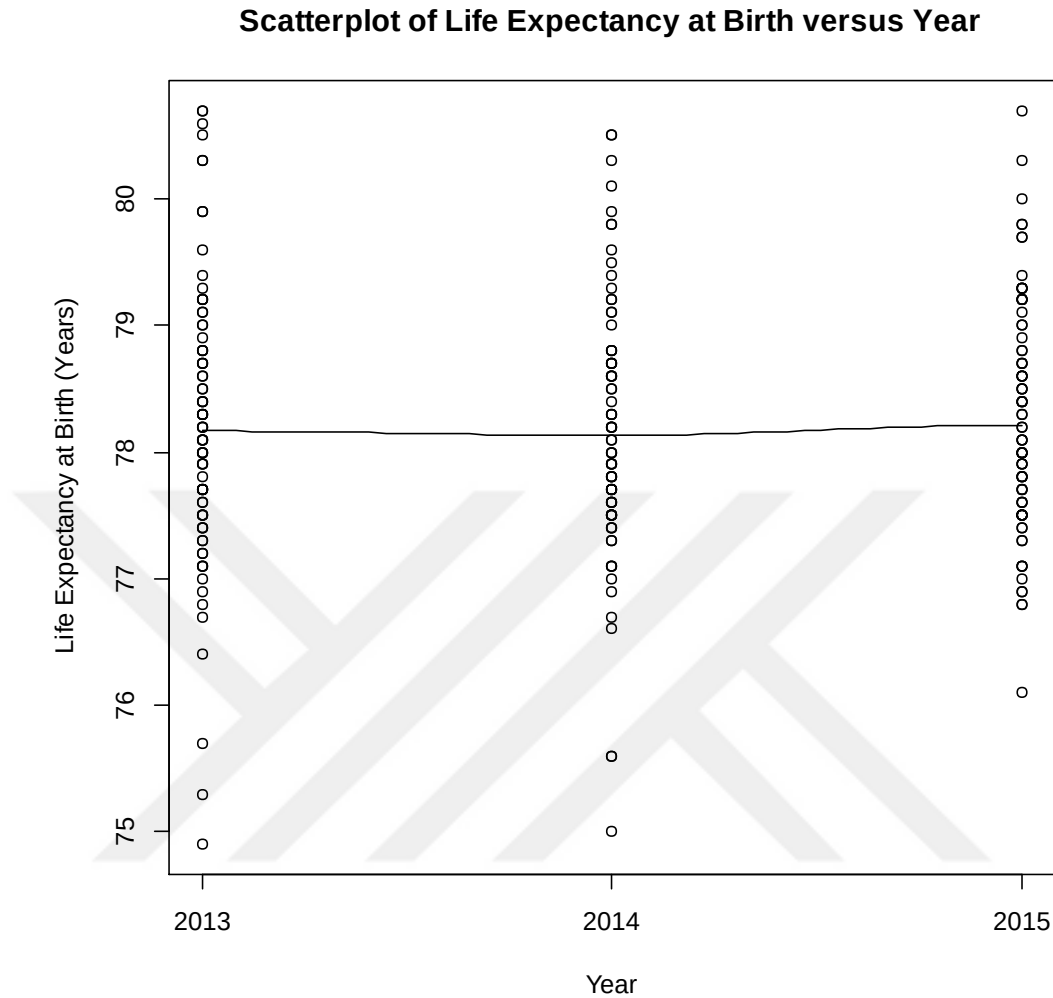


Figure 4.2. Scatterplot of Life Expectancy at Birth versus Time (Year)

From the correlation table among covariates, provided in Appendix B, it may be understood that there may be multicollinearity problem in the data set. For instance, there are very high correlations (higher than 0.95) between Population Density and some of the other covariates, namely, GDP, Import, Export, Number of Theater Halls and Number of Cinema Halls. Therefore, Variance Inflation Factor (VIF) values of the variables should be checked and some measures for multicollinearity can be taken during model construction part.

In order to check the cross-sectional dependence in the data set, Pesaran CD Test is used (Torres-Reyna, 2010). Life Expectancy at Birth values are checked for any

possible cross-sectional dependence for each of the three years. It is seen from Table 4.4 that there is no significant dependence in the data set.

Table 4.4 Results of Pesaran CD Test for Life Expectancy at Birth for Each Year

Year	<i>Z Statistic</i>	<i>p-value</i>
2013	0.644	0.519
2014	0.916	0.36
2015	1.316	0.188

4.2 Results of the Constructed Models

Regression analyses are conducted using the covariates Year, Population Density, Number of Housing Sales, GDP, Net Migration Rate, Number of Theater Halls, Number of Cinema Halls, Agricultural Field, Export (TL), Import (TL), Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand), Infant Mortality Rate (Per Thousand), Number of Students Per Classroom and response variable Life Expectancy at Birth. Random effects, marginal effects and transition models are constructed. In all the models, $t = 2013, 2014, 2015$ and $i = 1, 2, \dots, 81$. Then, using the significant variables in these models, a measurement error model is fitted.

4.2.1 Random Effects Models

The initial random effects model, given in equation 4.1, is constructed using the covariates Year, Population Density, Number of Housing Sales, GDP, Net Migration Rate, Number of Theater Halls, Number of Cinema Halls, Agricultural Field, Export (TL), Import (TL), Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand), Infant Mortality Rate (Per Thousand) and Number of Students Per Classroom.

$$\begin{aligned}
& \text{Life Expectancy at Birth}_{it} = \beta_0 + \beta_1 * t + \beta_2 * \text{Population Density}_{it} \\
& + \beta_3 * \text{Number of Housing Sales}_{it} + \beta_4 * \text{GDP}_{it} + \beta_5 * \\
& \text{Net Migration Rate}_{it} + \beta_6 * \text{Number of Theater Halls}_{it} + \beta_7 * \\
& \text{Number of Cinema Halls}_{it} + \beta_8 * \text{Agricultural Field}_{it} + \beta_9 * \\
& \text{Export (TL)}_{it} + \beta_{10} * \text{Import (TL)}_{it} + \beta_{11} * \\
& \text{Crude Birth Rate (Per Thousand)}_{it} + \beta_{12} * \\
& \text{Crude Marriage Rate (Per Thousand)}_{it} + \beta_{13} * \\
& \text{Crude Death Rate (Per Thousand)}_{it} + \beta_{14} * \\
& \text{Infant Mortality Rate (Per Thousand)}_{it} + \beta_{15} * \\
& \text{Number of Students Per Classroom}_{it} + b_i + \epsilon_{it}
\end{aligned} \tag{4.1}$$

Table 4.5 Estimated Regression Coefficients for the Initial Random Effects Model

	<i>Estimate</i>	<i>Standard Error</i>	<i>Z-value</i>	<i>p-value</i>
Intercept	-36.3	69.7	-0.52	0.6
Year	0.06	0.03	1.78	0.07
PD	-2.22×10^{-3}	1.67×10^{-3}	-1.33	0.18
NOHS	-4.31×10^{-6}	1.24×10^{-5}	-0.35	0.73
GDP	-5.40×10^{-9}	9.69×10^{-9}	-0.56	0.58
NMR	-1.80×10^{-3}	2.33×10^{-3}	-0.78	0.44
NOTH	1.36×10^{-3}	7.75×10^{-3}	0.17	0.86
NOCH	6.77×10^{-3}	5.87×10^{-3}	1.15	0.25
AF	-2.26×10^{-8}	3.89×10^{-8}	-0.58	0.56
Export (TL)	-6.52×10^{-12}	3×10^{-11}	-0.22	0.83
Import (TL)	1.94×10^{-11}	2.52×10^{-11}	0.77	0.44
CBR	-0.14	0.03	-5.55	2.89×10^{-8}
CMR	-0.24	0.08	-2.95	0.32×10^{-2}
CDR	-0.71	0.07	-10.1	$< 2.2 \times 10^{-16}$
IMR	-0.04	0.01	-3.75	0.18×10^{-3}
NOSPC	-0.02	0.01	-1.72	0.09

Adjusted R^2 value of the model is 0.39, meaning that covariates can explain 39% of the variability in the life expectancy. The p-value of the test for the significance of the model is less than 2.2×10^{-16} , meaning that the model is significant, and at least one of the covariates is significant.

After constructing the model, VIF values of the variables are checked for multicollinearity.

Table 4.6 VIF Values of Variables

	<i>VIF Value</i>
Year	3.45
PD	31.1
NOHS	18.7
GDP	88.1
NMR	1.09
NOTH	5.53
NOCH	31.3
AF	1.48
Export (TL)	72.1
Import (TL)	94
CBR	2.89
CMR	1.73
CDR	2.48
IMR	1.24
NOSPC	2.64

From Table 4.6, it can be seen that VIF values of Population Density, Number of Housing Sales, GDP, Number of Cinema Halls, Export (TL) and Import (TL) are bigger than 10. This is in accordance with what we already observed from correlation matrix. Thus, those variables should be omitted from the model.

Next, the random effect model given in equation 4.2 is constructed using the covariates Year, Net Migration Rate, Number of Theater Halls, Agricultural Field, Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand), Infant Mortality Rate (Per Thousand) and Number of Students Per Classroom. Those covariates are the ones whose VIF values are less than 10.

$$\begin{aligned}
& \text{Life Expectancy at Birth}_{it} = \beta_0 + \beta_1 * t + \beta_2 * \\
& \text{Net Migration Rate}_{it} + \beta_3 * \text{Number of Theater Halls}_{it} + \beta_4 * \\
& \text{Agricultural Field}_{it} + \beta_5 * \text{Crude Birth Rate (Per Thousand)}_{it} + \\
& \beta_6 * \text{Crude Marriage Rate (Per Thousand)}_{it} + \beta_7 * \\
& \text{Crude Death Rate (Per Thousand)}_{it} + \beta_8 * \\
& \text{Infant Mortality Rate (Per Thousand)}_{it} + \beta_9 * \\
& \text{Number of Students Per Classroom}_{it} + b_i + \epsilon_{it}
\end{aligned} \tag{4.2}$$

Table 4.7 Estimated Regression Coefficients for the Reduced Random Effects Model

	<i>Estimate</i>	<i>Standard Error</i>	<i>Z-value</i>	<i>p-value</i>
Intercept	3.17	61.9	0.05	0.96
Year	0.04	0.03	1.36	0.17
NMR	-2.15×10^{-3}	2.26×10^{-3}	-0.95	0.34
NOTH	-1.79×10^{-3}	3.5×10^{-3}	-0.51	0.61
AF	-1.58×10^{-8}	3.38×10^{-8}	-0.47	0.64
CBR	-0.14	0.02	-5.53	3.18×10^{-8}
CMR	-0.26	0.08	-3.23	1.26×10^{-3}
CDR	-0.7	0.07	-10.3	$< 2.2 \times 10^{-16}$
IMR	-0.04	0.01	-3.82	1.34×10^{-4}
NOSPC	-0.02	0.01	-2.21	0.03

Adjusted R^2 value of the model is 0.40, meaning that covariates can explain 40% of the variation in dependent variable. The p-value of the test for the significance of the model is less than 2.2×10^{-16} , meaning that the model is significant.

Finally, random effect model is constructed by omitting the insignificant covariates from the previous model.

$$\begin{aligned}
 \text{Life Expectancy at Birth}_{it} &= \beta_0 + \beta_1 * \\
 \text{Crude Birth Rate (Per Thousand)}_{it} &+ \beta_2 * \\
 \text{Crude Marriage Rate (Per Thousand)}_{it} &+ \beta_3 * \\
 \text{Crude Death Rate (Per Thousand)}_{it} &+ \beta_4 * \\
 \text{Infant Mortality Rate (Per Thousand)}_{it} &+ \beta_5 * \\
 \text{Number of Students Per Classroom}_{it} &+ b_i + \epsilon_{it}
 \end{aligned} \tag{4.3}$$

Table 4.8 Estimated Regression Coefficients for the Final Random Effects Model

	<i>Estimate</i>	<i>Standard Error</i>	<i>Z-value</i>	<i>p-value</i>
Intercept	87.4	0.82	107	$< 2.2 \times 10^{-16}$
CBR	-0.11	0.02	-5.62	1.9×10^{-8}
CMR	-0.3	0.07	-3.95	7.69×10^{-5}
CDR	-0.65	0.06	-10.5	$< 2.2 \times 10^{-16}$
IMR	-0.04	0.01	-3.92	9×10^{-5}
NOSPC	-0.03	8.58×10^{-3}	-4.09	4.37×10^{-5}

Adjusted R^2 value of the model is 0.40, meaning that covariates can explain 40% of the variation in dependent variable. The p-value of the test for the significance of the model is less than 2.2×10^{-16} , meaning that the model is significant. In addition, all regression coefficients seem to be significant.

After constructing the final model, diagnostics of the model should be checked. For this purpose, scatterplots of standardized residuals versus covariates of final

random effects model are examined for any kind of pattern. After examining the scatterplots, which are given in Appendix C, it is seen that there is no distinguishable pattern in the plots. This means that standardized residuals do not change much as covariates change and the model seems appropriate.

Secondly, scatterplot of standardized residuals of final random effects model versus time is examined for any possible pattern. As it can be seen from Figure 4.3, there is no distinguishable pattern. This means that standardized residuals do not change much as time changes.

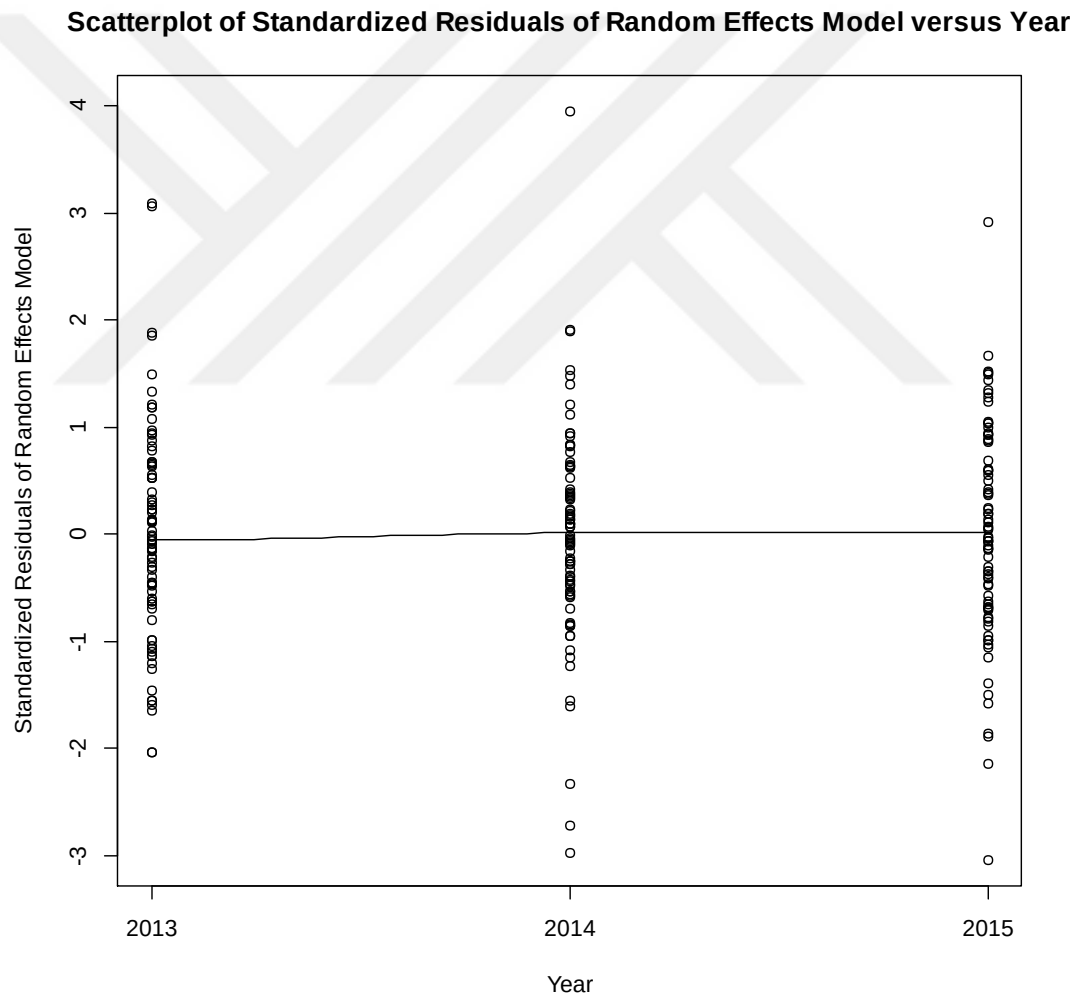


Figure 4.3. Scatterplot of Standardized Residuals of Random Effects Model versus Time (Year)

The final model:

$$\begin{aligned}
\text{Estimated Life Expectancy at Birth}_{it} = & 87.4 - 0.11 * \\
& \text{Crude Birth Rate (Per Thousand)}_{it} - 0.3 * \\
& \text{Crude Marriage Rate (Per Thousand)}_{it} - 0.65 * \\
& \text{Crude Death Rate (Per Thousand)}_{it} - 0.04 * \\
& \text{Infant Mortality Rate (Per Thousand)}_{it} - 0.03 * \\
& \text{Number of Students Per Classroom}_{it} + \hat{b}_i
\end{aligned} \tag{4.4}$$

In the following comments, we consider that $\hat{b}_i = 0$ for all i values (for all provinces). If Crude Birth Rate increases one unit in a province, the expected decrease in Life Expectancy at Birth will be 0.11 unit. If Crude Marriage Rate increases one unit, expected decrease in Life Expectancy at Birth will be 0.3 unit. If Crude Death Rate increases one unit, expected decrease in Life Expectancy at Birth will be 0.65 unit. If Infant Mortality Rate increases one unit, expected decrease in Life Expectancy at Birth will be 0.04 unit. If Number of Students per Classroom increases one unit, expected decrease in Life Expectancy at Birth will be 0.03 unit. Since the covariates cannot be equal to zero, there is no practical interpretation of the intercept term here. In addition, estimated variance of b_i is 0.61359, which is a small value. That means, there is not much difference between the provinces in terms of Life Expectancies at Birth.

Scatterplot of Fitted Response Values of Random Effects Model versus Life Expectancy at Birth

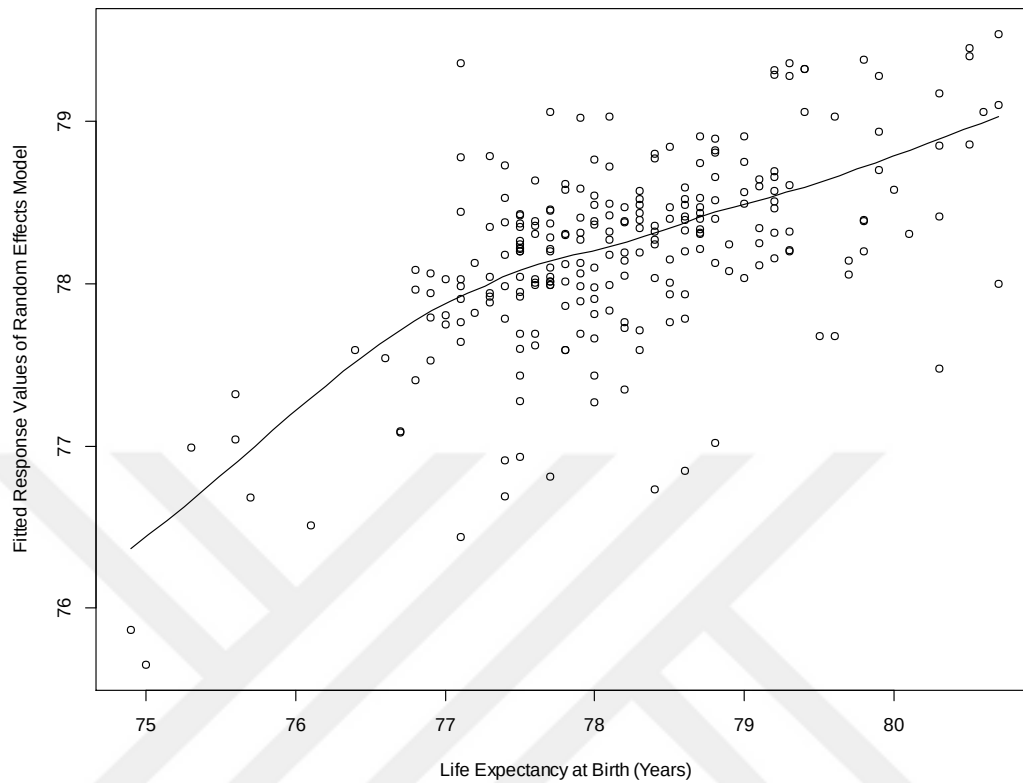


Figure 4.4. Scatterplot of Fitted Response Values of Random Effects Model versus Life Expectancy at Birth

As it can be seen from Figure 4.4, it can be said that until the Life Expectancy at Birth value of 78, the model slightly overestimates and after the Life Expectancy at Birth value of 78, the model slightly underestimates. However, fitted response values of the random effects model are mostly concordant with the observed response values. The curve in the plot is a smoothing curve that indicates the pattern of the concordance between real Life Expectancy at Birth values and fitted Life Expectancy at Birth values.

After the determination of the significant variables for random effects model, it is suspected that the Year variable may also be in the random part of the random effects model. In other words, we wonder if the response changes over the years

are different for different provinces. Thus, a model is constructed accordingly as in equation 4.5.

$$\begin{aligned}
 \text{Life Expectancy at Birth}_{it} = & \beta_0 + \beta_1 * t + \beta_2 * \\
 & \text{Crude Birth Rate (Per Thousand)}_{it} + \beta_3 * \\
 & \text{Crude Marriage Rate (Per Thousand)}_{it} + \beta_4 * \\
 & \text{Crude Death Rate (Per Thousand)}_{it} + \beta_5 * \\
 & \text{Infant Mortality Rate (Per Thousand)}_{it} + \beta_6 * \\
 \text{Number of Students Per Classroom}_{it} = & b_{i0} + b_{i1} * t + \epsilon_{it}
 \end{aligned} \tag{4.5}$$

Table 4.9 Estimated Regression Coefficients for the Random Effects Model with Year in the Random Part

	<i>Estimate</i>	<i>Standard Error</i>	<i>t-value</i>
Intercept	173.05	135.61	1.28
Year	-0.04	0.07	-0.63
CBR	-0.06	0.02	-2.76
CMR	-0.42	0.07	-5.67
CDR	-0.55	0.07	-8.04
IMR	-0.09	0.03	-3.52
NOSPC	-0.04	0.01	-3.08

All regression coefficients except the intercept and coefficient of Year seem to be significant. However, estimate of the variance of random slope is basically 0. Thus, it can be said that there is no difference between provinces in terms of their changes in the expected life expectancy over the years. Therefore, there is no need to consider Year variable in the random part of the random effects model for this data set.

Root Mean Square Error (RMSE) of the final random effects model is calculated and found to be 0.23, which is close to zero. Thus, it can be said that the model has

a good fit. Since the function that is used to create the model is not a likelihood-based function, AIC and BIC values cannot be computed for the final random effects model.

4.2.2 Marginal Models

Second part of the data analysis includes constructing a marginal model for the data set. In the creation of the models, all of the possible correlation structures for marginal model are examined separately. As a result of this examination, exchangeable and AR(1) correlation structures are used for the analysis of final marginal model, since those correlation structures are the only ones which give reasonable results. The procedure used in this section is similar to the procedure used in Section 4.2.1. Firstly, full model is constructed using the covariates Year, Population Density, Number of Housing Sales, GDP, Net Migration Rate, Number of Theater Halls, Number of Cinema Halls, Agricultural Field, Export (TL), Import (TL), Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand), Infant Mortality Rate (Per Thousand) and Number of Students Per Classroom. Secondly, covariates whose VIF values are bigger than 10 (Population Density, Number of Housing Sales, GDP, Number of Cinema Halls, Export (TL) and Import (TL)) are omitted from the full model and a new model is created with the rest of the covariates (reduced model). Finally, insignificant covariates are omitted from the reduced model, and the final model is created using the covariates Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand) and Infant Mortality Rate (Per Thousand). The final model is given in equation 4.6.

$$\begin{aligned}
 & \text{Life Expectancy at Birth}_{it} = \beta_0 + \beta_1 * \\
 & \text{Crude Birth Rate (Per Thousand)}_{it} + \beta_2 * \\
 & \text{Crude Marriage Rate (Per Thousand)}_{it} + \beta_3 * \\
 & \text{Crude Death Rate (Per Thousand)}_{it} + \beta_4 *
 \end{aligned}
 \tag{4.6}$$

$$\text{Infant Mortality Rate (Per Thousand)}_{it} + \varepsilon_{it}$$

Table 4.10 Estimated Regression Coefficients for the Final Marginal Model with Exchangeable Correlation Structure

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	86.8	1.27	68.2
CBR	-0.12	0.02	-5.45
CMR	-0.37	0.1	-3.56
CDR	-0.57	0.07	-8.04
IMR	-0.05	0.01	-3.72

Table 4.11 Estimated Regression Coefficients for the Final Marginal Model with AR-1 Correlation Structure

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	85.8	1.13	75.6
CBR	-0.11	0.02	-5.33
CMR	-0.29	0.08	-3.38
CDR	-0.53	0.07	-8
IMR	-0.04	0.01	-4.16

All regression coefficients seem to be significant for both models. The results of both models are really close to each other in terms of estimated regression coefficients, robust standard errors, robust Z-values, QIC goodness of fit statistics (244.3 for marginal model with AR-1 correlation structure and 250.2 for marginal model with exchangeable correlation structure) and closeness of the working correlations in the output of both models to the correlation structure obtained from

the data. Thus, both marginal models can be used. Here, we proceed with the marginal model with exchangeable correlation structure.

For diagnostics purposes, scatterplots of standardized residuals versus covariates of final marginal model are examined for any kind of pattern. After examining the scatterplots, which are given in Appendix F, it is seen that there is no distinguishable pattern in the plots. This means that standardized residuals do not change much as covariates change.

Secondly, scatterplot of standardized residuals of final marginal model versus time are examined for any possible pattern. As it can be seen from Figure 4.5, there is no distinguishable pattern in the plot. This means that standardized residuals do not change much as time changes.

Scatterplot of Standardized Residuals of Marginal Model versus Year

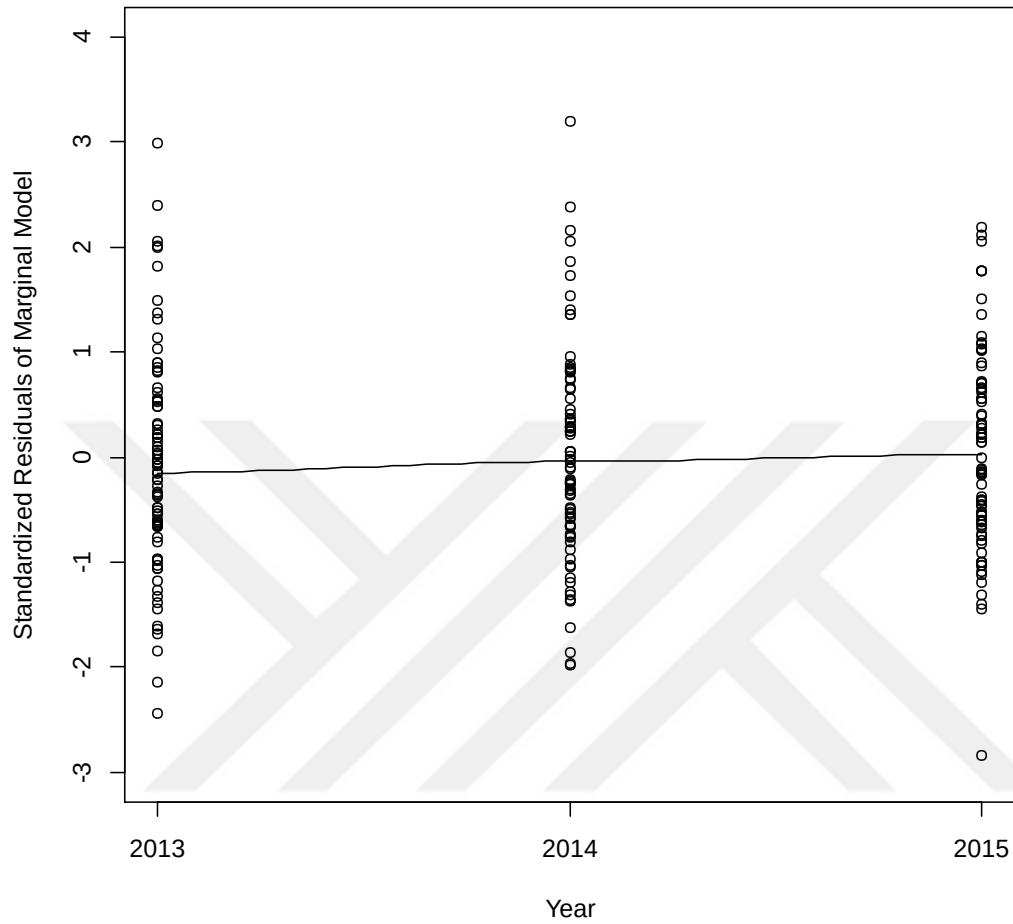


Figure 4.5. Scatterplot of Standardized Residuals of Marginal Model versus Time (Year)

The final model:

$$\begin{aligned}
 \text{Expected Life Expectancy at Birth}_{it} = & 86.8 - 0.12 * \\
 & \text{Crude Birth Rate (Per Thousand)}_{it} - 0.37 * \\
 & \text{Crude Marriage Rate (Per Thousand)}_{it} - 0.57 * \\
 & \text{Crude Death Rate (Per Thousand)}_{it} - 0.05 * \\
 & \text{Infant Mortality Rate (Per Thousand)}_{it}
 \end{aligned}
 \tag{4.7}$$

There is a negative relationship between the life expectancy at birth and all of the covariates in this model. If Crude Birth Rate increases one unit, expected decrease in Life Expectancy at Birth will be 0.12 unit. If Crude Marriage Rate increases one unit, expected decrease in Life Expectancy at Birth will be 0.37 unit. If Crude Death Rate increases one unit, expected decrease in Life Expectancy at Birth will be 0.57 unit. If Infant Mortality Rate increases one unit, expected decrease in Life Expectancy at Birth will be 0.05 unit. Since the covariates cannot be equal to zero, there is no practical interpretation of the intercept term here.

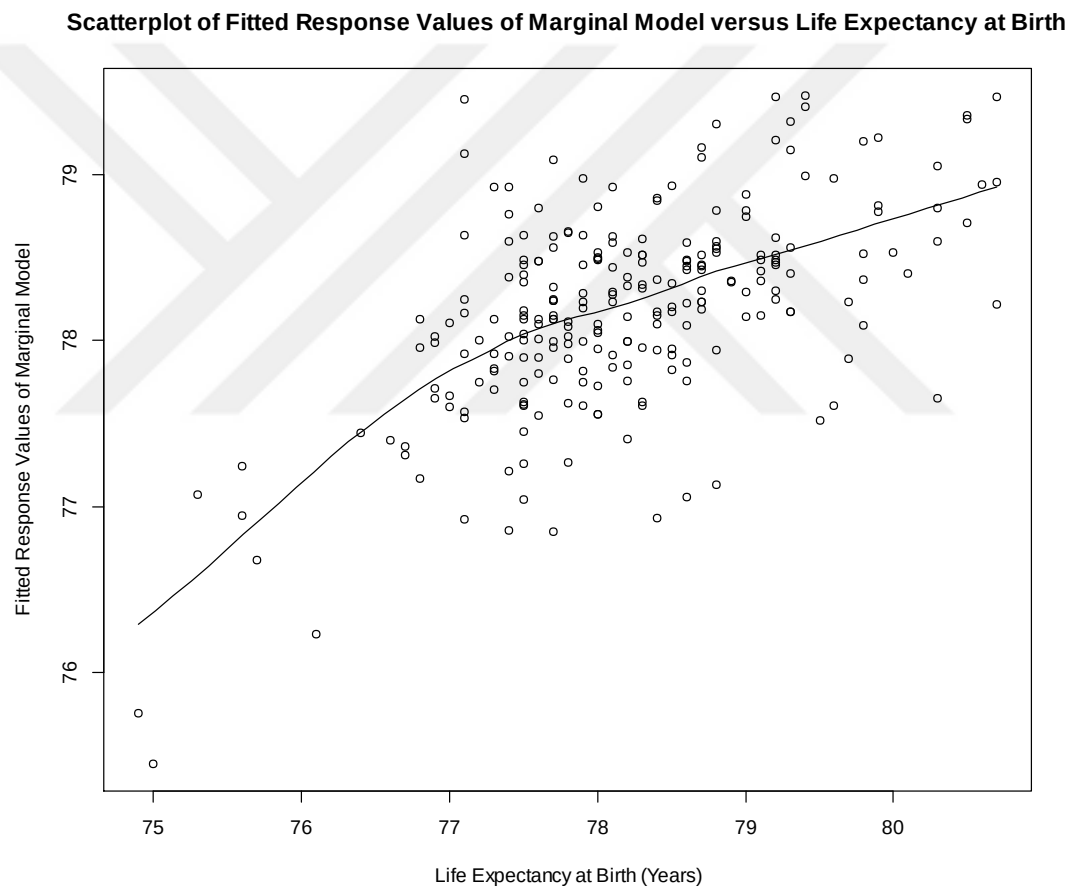


Figure 4.6. Scatterplot of Fitted Response Values of Marginal Model versus Life Expectancy at Birth

As it can be seen from Figure 4.6, it can be said that until the Life Expectancy at Birth value of 78, the model overestimates and after the Life Expectancy at Birth

value of 78, the model slightly underestimates. However, fitted response values of the marginal model are mostly concordant with the observed response values. The curve in the plot is a smoothing curve that indicates the pattern of the concordance between real Life Expectancy at Birth values and fitted Life Expectancy at Birth values.

Root Mean Square Error (RMSE) of the final marginal model is calculated and found to be 0.83, which is close to zero. Thus, it can be said that the model has a good fit. Note that, according to RMSE values, the final random effects model has a better fit than the final marginal model. However, random effects model requires the estimation of 81 extra parameters (b_i), and it is considered to be unnecessary when these 81 provinces are similar to each other. In addition, Hausman Test is used to compare the final random effects model and the final marginal model and the result shows that marginal model should be used for this data set, since the p-value of the test is found to be less than the significance level 0.05. Since the function that is used to create the model is not a likelihood-based function, AIC and BIC values cannot be computed for the final marginal model (Pan, 2001).

4.2.3 Transition Models

Third part of the data analysis includes constructing a transition model for the data set. The procedure used in this section is similar to the procedure used in Section 4.2.1. Firstly, full model is constructed using the covariates Year, Population Density, Number of Housing Sales, GDP, Net Migration Rate, Number of Theater Halls, Number of Cinema Halls, Agricultural Field, Export (TL), Import (TL), Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand), Infant Mortality Rate (Per Thousand), Number of Students Per Classroom and Lag 1 of Life Expectancy at Birth. Secondly, covariates whose VIF values are bigger than 10 (Population Density, Number of Housing Sales, GDP, Number of Cinema Halls, Export (TL) and Import (TL)) are omitted from the full model, and a new model is created with the rest of the

covariates (reduced model). Finally, insignificant covariates are omitted from the reduced model and final model is created using the covariates Year, Crude Death Rate (Per Thousand) and Life Expectancy at Birth Lag 1.

The final model is given in equation 4.8. The results of the fitted model are given in Table 4.12. All regression coefficients seem to be significant.

$$\begin{aligned} \text{Life Expectancy at Birth}_{it} = & \beta_0 + \beta_1 * t + \beta_2 * \\ & \text{Crude Death Rate (Per Thousand)}_{it} + \beta_3 * \\ & \text{Life Expectancy at Birth}_{i,t-1} + \varepsilon_{it} \end{aligned} \quad (4.8)$$

Table 4.12 Estimated Regression Coefficients for the Final Transition Model

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	-324	98	-3.3
Year	0.17	0.05	3.41
CDR	-0.04	0.01	-2.73
LEAB Lag 1	0.85	0.02	43.1

After constructing the final model, diagnostics of the model should be checked. For this purpose, scatterplots of standardized residuals versus covariates of final transition model are examined for any kind of pattern firstly. After examining the scatterplots, which are given in Appendix I, it is seen that there is no distinguishable pattern in the plots. This means that standardized residuals do not change much as covariates change.

Secondly, scatterplot of standardized residuals of final transition model versus time are examined for any possible pattern. As it can be seen from Figure 4.7, it can be said that there is no distinguishable pattern in the plot. This means that standardized residuals do not change much as time changes.

Scatterplot of Standardized Residuals of Transition Model versus Year

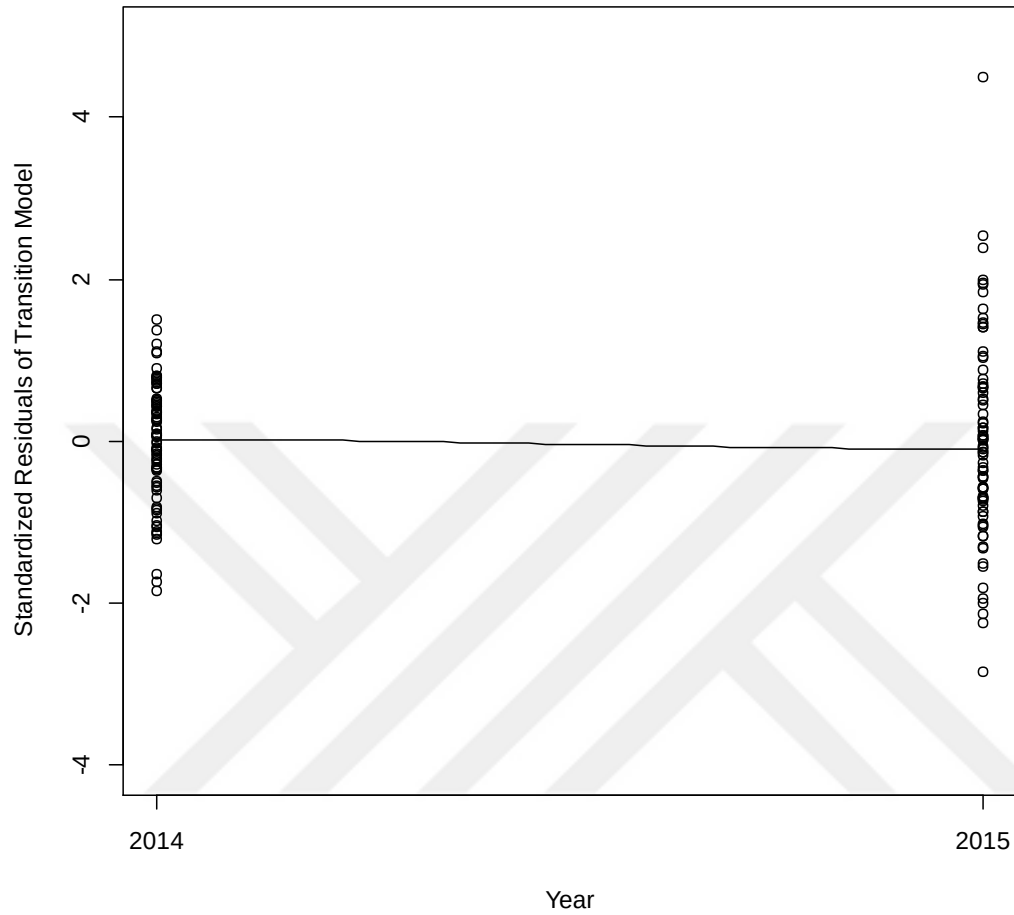


Figure 4.7. Scatterplot of Standardized Residuals of Transition Model versus Time (Year)

The final fitted model:

$$\begin{aligned} \text{Estimated Life Expectancy at Birth}_{it} = & -324 + 0.17 * t - 0.04 * \\ & \text{Crude Death Rate (Per Thousand)}_{it} + 0.85 * \\ & \text{Life Expectancy at Birth}_{i,t-1} \end{aligned} \quad (4.9)$$

Life Expectancy at Birth increases slightly over the years. If Crude Death Rate increases one unit, expected decrease in Life Expectancy at Birth is 0.04 unit. Life

Expectancy at Birth observed in the last year is positively and significantly related with the Life Expectancy at Birth this year, as expected. Since the covariates cannot be equal to zero, there is no practical interpretation of the intercept term here.

Scatterplot of Fitted Response Values of Transition Model versus Life Expectancy at Birth

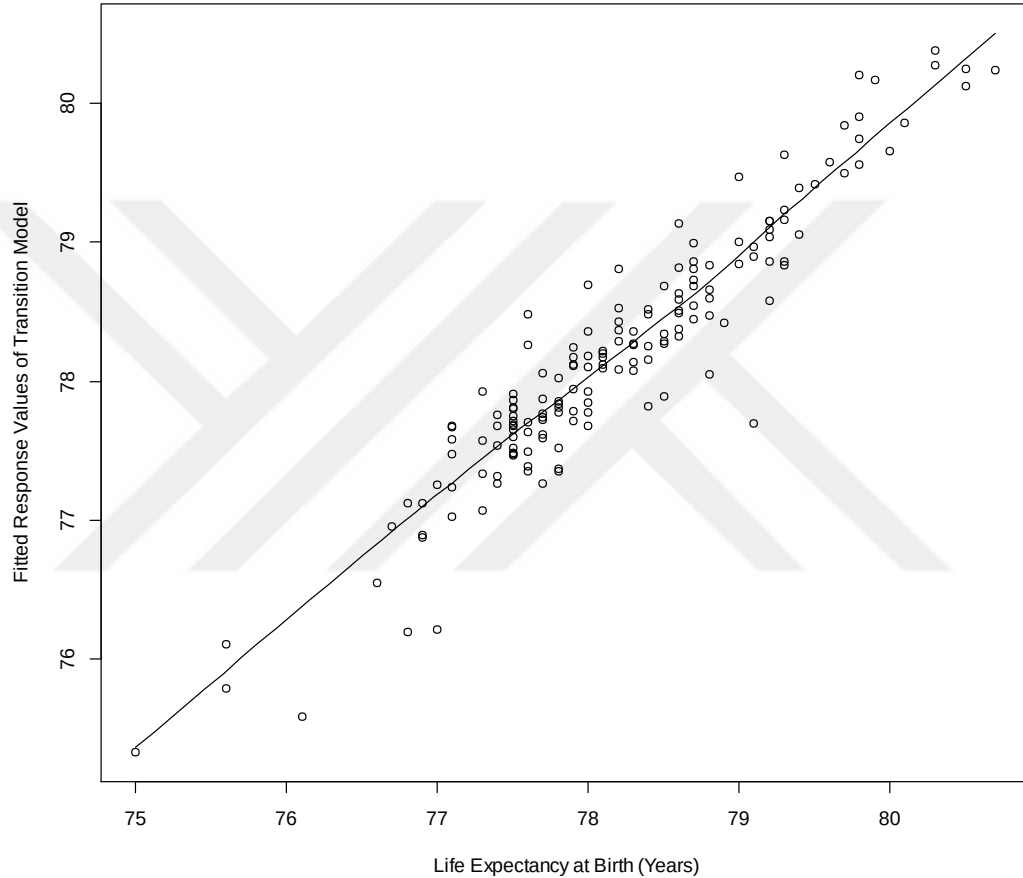


Figure 4.8. Scatterplot of Fitted Response Values of Transition Model versus Life Expectancy at Birth

As it can be seen from Figure 4.8, the fitted response values are very close to the real response values in the transition model.

Root Mean Square Error (RMSE) of the final transition model is calculated and found to be 0.31, which is close to zero. Thus, it can be said that the model has a good fit. According to RMSE values, the final random effects model has a better fit

than the final transition model and the final transition model has a better fit than the final marginal model. However, all of the RMSEs are very close to each other. Hence, we picked the best model according to interpretability and the number of useful parameters: Random effects model requires the estimation of many random effect parameters, which are unjustifiable for this data set, since provinces are similar to each other in terms of the Life Expectancy at Birth. Transition model, although providing the most compatible fits, highly depends on the last year's Life Expectancy at Birth. Its success roots from the fact that the autocorrelations are extremely high for this data set. Any errors in the measurement of Life Expectancy at Birth in the first year would result in considerable errors in the prediction of Life Expectancy at Birth in the later years with these models. Marginal model, on the other hand, borrows strength from other covariates in the data. Since the function that is used to create the model is not a likelihood-based function, AIC and BIC values cannot be computed for the final transition model.

After getting the final models, signs of coefficients of the covariates are checked by examining the scatterplots of the response variable versus significant covariates. After examining the scatterplots, which are given in Appendix K, relationships between covariates and response variables are observed as below.

There is a negative relationship between Life Expectancy at Birth and the following covariates:

- Crude Birth Rate,
- Crude Marriage Rate,
- Crude Death Rate,
- Infant Mortality Rate,
- Number of Students Per Classroom

There is a positive relationship between Life Expectancy at Birth and the following covariates:

- Year,
- Lag 1 of Life Expectancy at Birth

Thus, all the signs of coefficients of the covariates are found to be as expected for all three final models.

4.2.4 Models with Dummy Variables

From the scatterplot of Life Expectancy at Birth versus Crude Birth Rate, which is in Appendix K, it can be seen that the pattern of the observation points is different where Crude Birth Rate is below 17 than the pattern of those points where Crude Birth Rate is above 17. A similar pattern is observed with Crude Marriage Rate. Thus, in an attempt to improve our models, these information are considered to construct dummy variables for Crude Birth and Marriage Rates, and in addition to the previous analyses, new models are fitted with these two dummy variables added.

Dummy variables that are created for the variables Crude Birth Rate and Crude Marriage Rate are as below.

Crude Birth Rate (Per Thousand)_{it} is:

- 0 if *Crude Birth Rate (Per Thousand)_{it}* is greater than 17,
- 1 if *Crude Birth Rate (Per Thousand)_{it}* is less than or equal to 17.

Crude Marriage Rate (Per Thousand)_{it} is:

- 0 if *Crude Marriage Rate (Per Thousand)_{it}* is less than or equal to 8,
- 1 if *Crude Marriage Rate (Per Thousand)_{it}* is greater than 8.

After creating the dummy variables, the interaction between Crude Birth Rate and its dummy variable, the interaction between square of Crude Birth Rate and its dummy variable, the interaction between Crude Marriage Rate and its dummy

variable and dummy variable of Crude Marriage Rate are added to each three type of models, since the patterns of the observations in the scatterplot of Life Expectancy at Birth versus Crude Birth Rate and in the scatterplot of Life Expectancy at Birth versus Crude Marriage Rate show small amount of non-linearity. However, none of those newly added terms are found to be significant. On the contrary, adding those terms cause insignificance of some of the variables which are found to be significant previously. In addition, adding those terms did not make any improvements on the diagnostics of the models. Thus, it is decided to omit those terms back from the analysis. The results for the models with dummy variables for three types of models are given in Appendices D, G and J.

4.3 Measurement Error Model

According to the definition given in the Memobust Handbook on Methodology of Modern Business Statistics which was released on 26th of March 2014, “measurement error is the difference between the observed value of a variable and the true, but unobserved, value of that variable”. Missing observations and covariate measurement error frequently arise in longitudinal studies and they present considerable challenges in statistical inference about such data (Carroll et al., 2006; Yi, 2008). It has been well documented that ignoring missing responses and covariate measurement error may lead to severely biased results, thus leading to invalid inferences (Fuller, 1987; Carroll et al., 2006). Due to the natures of the covariates in our data, it is suspected that there may be measurement error in the covariates. The measurements involve quantities in provinces with millions of population, and hence it is challenging to measure them without considerable error. The “swgee” package, which was written by Juan Xiong and Grace Y. Yi on 1st June 2019, is used for the analysis of data with measurement error and possibly missing cases. We would like to illustrate the use of models under missing data as well in case there are missing cases in data set. Thus, by doing random selection, variables of Ardahan in 2014 are made missing intentionally. All significant

covariates of the final marginal model (Crude Birth Rate, Crude Marriage Rate, Crude Death Rate and Infant Mortality Rate) are checked for measurement error. Following Yi (2008), variance terms in Σ_e (the covariance matrix of e_{it} in the measurement error equation $w_{it} = x_{it} + e_{it}$) are specified as 0, 0.5 and 1 respectively for each covariate to illustrate the cases with different degrees of measurement error in each covariate. Thus, we get $3^4=81$ different combinations of variance terms in Σ_e , and hence, 81 different models are fitted by this way. Each variance combination is given an ID (Identification Number) from 1 to 81. IDs of the variance combinations are listed in the table in Appendix L. Four variances that are listed belong to Crude Birth Rate, Crude Marriage Rate, Crude Death Rate and Infant Mortality Rate respectively.

The 81 values of the estimated coefficients for the intercept and the four covariates are separately given in the plots below. The points specified with filled circle show the significant coefficients and the points specified with cross show the insignificant covariates.

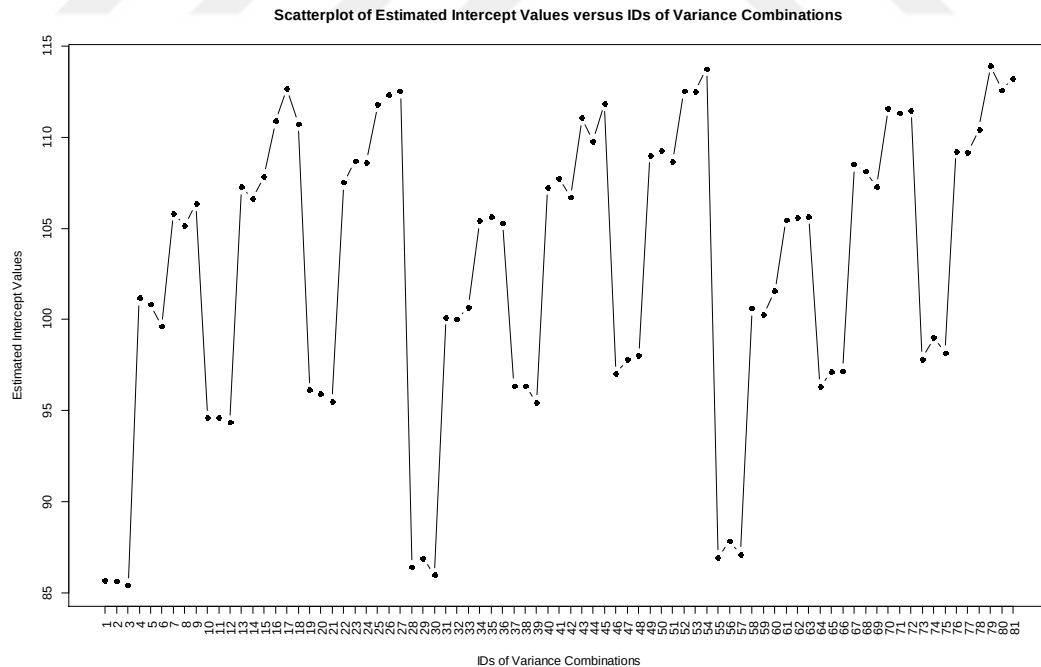


Figure 4.9. Scatterplot of Estimated Intercept Values versus IDs of Variance Combinations

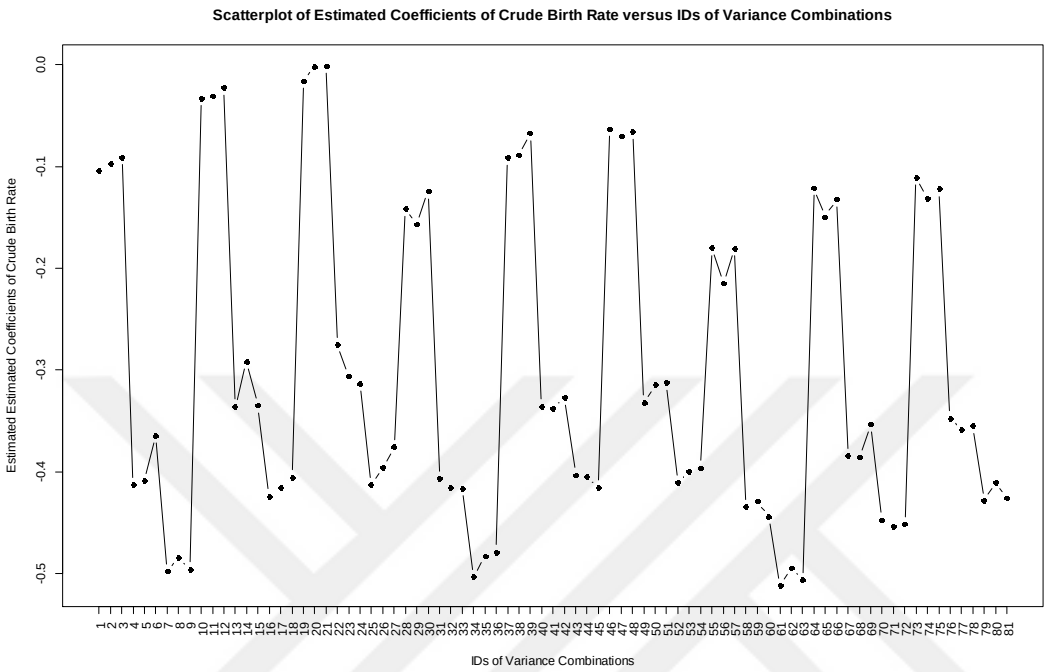


Figure 4.10. Scatterplot of Estimated Coefficients of Crude Birth Rate versus IDs of Variance Combinations

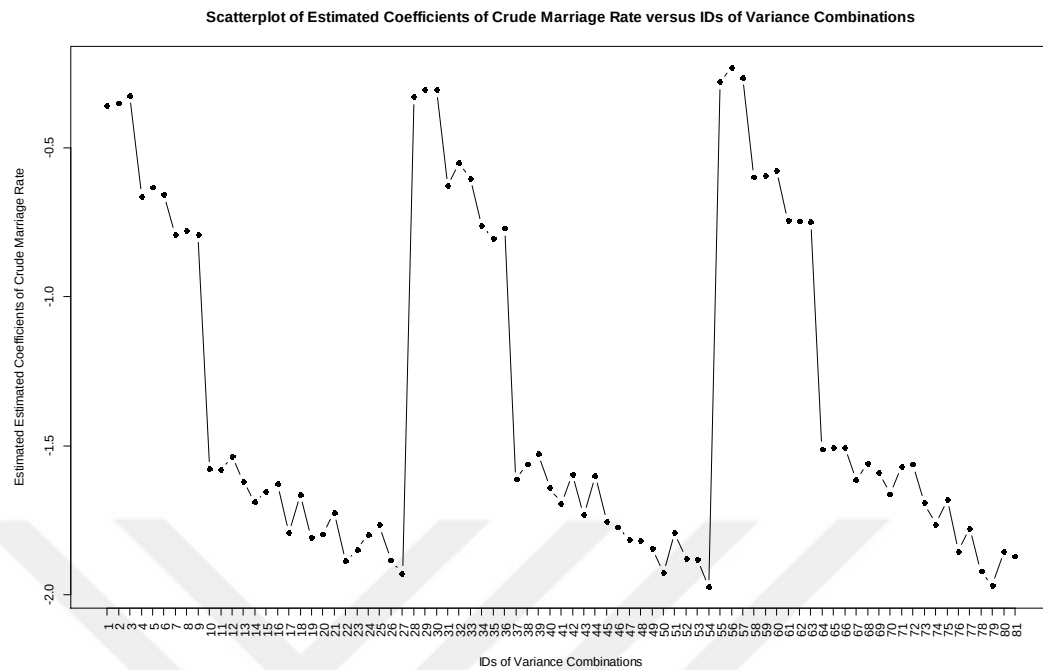


Figure 4.11. Scatterplot of Estimated Coefficients of Crude Marriage Rate versus IDs of Variance Combinations

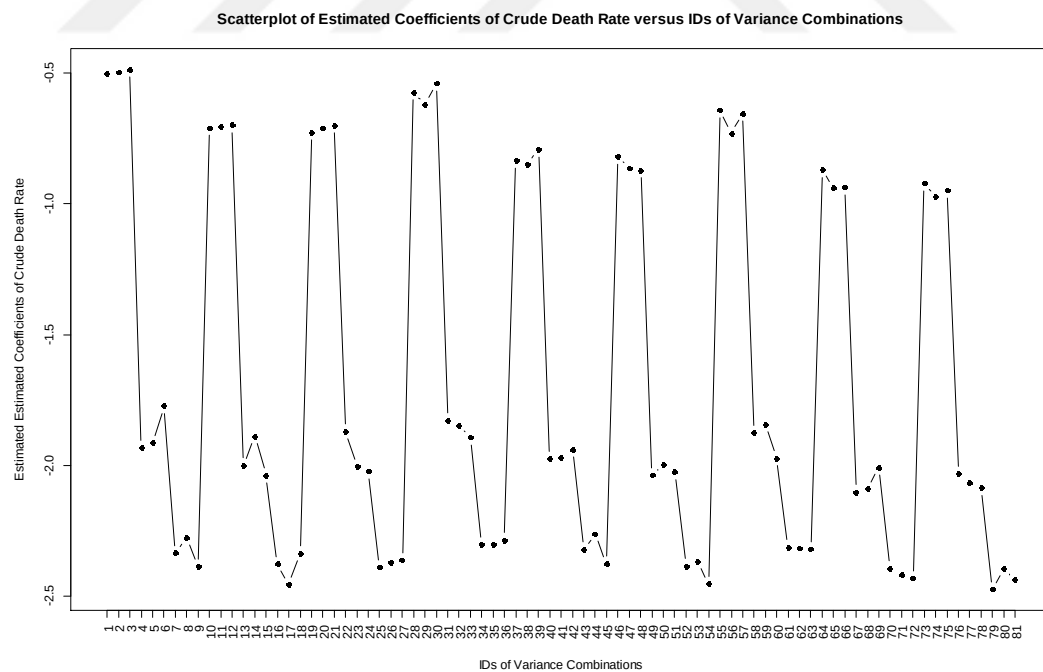


Figure 4.12. Scatterplot of Estimated Coefficients of Crude Death Rate versus IDs of Variance Combinations

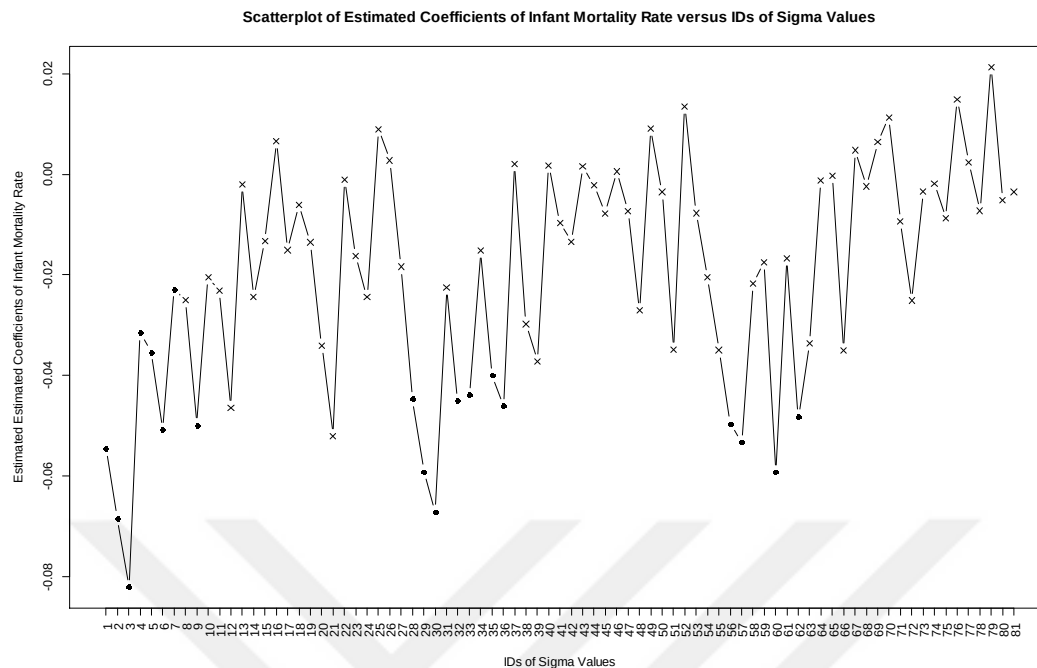


Figure 4.13. Scatterplot of Estimated Coefficients of Infant Mortality Rate versus IDs of Variance Combinations

As it can be seen from Figures 4.9 - 4.13, only the significancies of the estimated coefficients of Infant Mortality Rate change as the degrees of measurement errors change. To be more clear, a list of variance combinations where Infant Mortality Rate is insignificant is given in Appendix M.

From this table, it can be understood that it is not really clear that changes in the degree of measurement error in which variable changes the significancy of Infant Mortality Rate. However, it can be said that as different combinations and degrees of measurement errors are considered for four variables, the effect of Infant Mortality Rate on Life Expectancy at Birth changes.

CHAPTER 5

CONCLUSION

Life expectancy at birth is one of the most important indicators of a country's economical and social level of development and its overall health status. Thus, it is really crucial to find out factors that affect life expectancy at birth and how they affect it in order to provide governments with the necessary information to set their policies accordingly.

Panel data modeling is one of the most commonly used modelling methods in health statistics. In this thesis the most common three panel data models are used for analysis and their results are given.

Firstly, random effects model is constructed. The final random effects model is found to be $Expected\ Life\ Expectancy\ at\ Birth_{it} = 87.4 - 0.11 * Crude\ Birth\ Rate\ (Per\ Thousand)_{it} - 0.3 * Crude\ Marriage\ Rate\ (Per\ Thousand)_{it} - 0.65 * Crude\ Death\ Rate\ (Per\ Thousand)_{it} - 0.04 * Infant\ Mortality\ Rate\ (Per\ Thousand)_{it} - 0.03 * Number\ of\ Students\ Per\ Classroom_{it} + \hat{b}_i$. According to this model, Life

Expectancy at Birth is negatively affected by Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand), Infant Mortality Rate (Per Thousand) and Number of Students Per Classroom. Thus, according to this model appropriate government policies to increase life expectancy at birth would be encouraging birth control, healing health system in the country so that there will be less deaths occurring at every age and building more schools. The negative relation between marriage rate and life expectancy could be due to many reasons. First, married women might be at higher risk of death due to pregnancies and giving births. Secondly, some provinces might have younger populations, and

hence lower marriage rate and higher life expectancies. The estimated variance of b_i being a small value indicates low heterogeneity between provinces. This means there is no need to use random effects model.

Secondly, marginal model is constructed. The final marginal model is found to be

$$\begin{aligned} \text{Expected Life Expectancy at Birth}_{it} &= 86.8 - 0.12 * \\ \text{Crude Birth Rate (Per Thousand)}_{it} &- 0.37 * \\ \text{Crude Marriage Rate (Per Thousand)}_{it} &- 0.57 * \\ \text{Crude Death Rate (Per Thousand)}_{it} &- 0.05 * \end{aligned}$$

$\text{Infant Mortality Rate (Per Thousand)}_{it}$. According to this model, Life Expectancy at Birth is negatively affected by Crude Birth Rate (Per Thousand), Crude Marriage Rate (Per Thousand), Crude Death Rate (Per Thousand) and Infant Mortality Rate (Per Thousand). Thus, according to this model appropriate government policies to increase life expectancy at birth would be encouraging birth control and healing health system in the country so that there will be less deaths occurring at every age. Constructing marginal model is found to be the most appropriate choice for this data set since the model gives good regression diagnostics, and avoids the estimation of many random effect parameters.

Thirdly, transition model is constructed. The final transition model is found to be

$$\begin{aligned} \text{Estimated Life Expectancy at Birth}_{it} &= -324 + 0.17 * t - 0.04 * \\ \text{Crude Death Rate (Per Thousand)}_{it} &+ 0.85 * \text{Life Expectancy at Birth}_{i,t-1}. \end{aligned}$$

According to this model, Life Expectancy at Birth is negatively affected by Crude Death Rate (Per Thousand) and it is positively affected by time and lag 1 of Life Expectancy at Birth. As it can be seen, the life expectancies in the consecutive years are highly related with each other, and the life expectancy is slightly increasing over the years. However, coefficient of only one of the covariates, together with year and lag, are found to be significant and this is not highly informative for explaining life expectancy at birth. Thus, transitional model is not chosen as the best model for this data set.

According to the final marginal model that is constructed, the highest estimated average value of Life Expectancy at Birth for three years belongs to Ankara and the lowest average value belongs to Kilis. The observed values of Life Expectancy at Birth show that the highest average value of Life Expectancy at Birth for three years belongs to Tunceli and the lowest average value belongs to Kilis.

The best model is picked according to interpretability and the number of useful parameters. Random effects model is not picked since estimation of many random effects parameters is unnecessary because of the similarity of Life Expectancy at Birth values of provinces. Transition model is not picked since it depends on previous years' Life Expectancy at Birth values and errors in the prediction of Life Expectancy at Birth values of the previous years highly affect the Life Expectancy at Birth values of the later years. Thus, marginal model is found to be the best model for this data set.

In addition to the modelling part, all significant covariates of the final marginal model (Crude Birth Rate, Crude Marriage Rate, Crude Death Rate and Infant Mortality Rate) are checked for measurement error. A total of 81 models with different degrees of measurement error (0, 0.5 and 1) in each covariate are examined. As a result, it is seen that as different combinations and degrees of measurement errors are considered for four covariates, the significance of Infant Mortality Rate changes. Nonetheless, it is not really clear that changes in the degree of measurement error of which variable changes the significance of Infant Mortality Rate.

In this thesis, the data set has spatial feature. Provinces that are geographically close to each other might also have observations that are close to each other in value. Thus, it would be useful to add locational information (such as longitude and latitude) of the provinces to the constructed models as covariates. In addition, since the data set includes only three time points and this number is insufficient for doing forecast on data, forecasting of Life Expectancy at Birth for 81 provinces cannot be done. We hope this thesis will also shed a light on government agencies and show

the importance and need of more and reliable data collection. Those topics are left
as a future work.



REFERENCES

- Asiskovitch, S. (2010). Gender and health outcomes: The Impact of Healthcare Systems and their Financing on Life Expectancies of Women and Men. *Social Science & Medicine*, 70, 886–895.
- Bayati, M., Akbarian, R. & Kavosi, Z. (2013). Determinants of Life Expectancy in Eastern Mediterranean Region: A Health Production Function. *International Journal of Health Policy and Management*, 1 (1), 57-61.
- Bayın, G. (2016). Doğuştta ve İleri Yaşta Beklenen Yaşam Sürelerine Etki Eden Faktörlerin Belirlenmesi. *Türkiye Aile Hekimliği Dergisi*, 20 (3), 93-103.
- Bilas, V., Franc, S. & Bosnjak, M. (2014). Determinant Factors of Life Expectancy at Birth in the European Union Countries. Retrieved January 18, 2021, from <https://core.ac.uk/download/pdf/20339573.pdf>.
- Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed., pp. 1-5). Boca Raton, Florida: Chapman & Hall/CRC.
- Delavari, S., Zandian, H., Rezaei, S., Moradinazar, M., Delavari, S., Saber, A. & Fallah, R. (2016). Life Expectancy and its Socioeconomic Determinants in Iran. Retrieved January 18, 2021, from [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5133029/#:~:text=Gulis%20\(9\)%20has%20examined%20the,main%20determinants%20of%20life%20expectancy](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5133029/#:~:text=Gulis%20(9)%20has%20examined%20the,main%20determinants%20of%20life%20expectancy).
- Fuller, W. A. (1987). *Measurement Error Models*. Wiley Series in Probability and Statistics. New York: John Wiley & Sons.
- Hedges, L. V. (2012). ‘Random-Effects Models’, in Salkind, N. J. (ed.). *Encyclopedia of Research Design* (pp. 1198-1201). Thousand Oaks, California: SAGE Publications, Inc.

- Huque, M. H., Bondell, H. D., Carroll, R. J. & Ryan, L. M. (2016). Spatial Regression with Covariate Measurement Error: A Semiparametric Approach. *Journal of the International Biometric Society*, 72 (3), 678-686.
- Kröger, H., Hoffmann, R. & Pakpahan, E. (2015). Consequences of Measurement Error for Inference in Cross-Lagged Panel Design—The Example of the Reciprocal Causal Relationship between Subjective Health and Socio-Economic Status. *Journal of the Royal Statistical Society*, 179, 607–628.
- Lin, R., Chen, Y., Chien, L. & Chan, C. (2012). Political and Social Determinants of Life Expectancy in Less Developed Countries: A Longitudinal Study. Retrieved January 18, 2021, from <https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-85>.
- Organisation for Economic Co-operation and Development (2006, 2013). OECD Glossary of Statistical Terms. Retrieved April 4, 2020, from <https://stats.oecd.org/glossary/>.
- Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, 57 (1), 120-125.
- Shahbaz, M., Loganathan, N., Mujahid, N., Ali, A. & Nawaz, A. (2015). Determinants of Life Expectancy and its Prospects under the Role of Economic Misery: A Case of Pakistan. Retrieved January 18, 2021, from https://mpira.ub.uni-muenchen.de/67167/1/MPRA_paper_67167.pdf.
- Statistics Netherlands (2014). 'Memobust Glossary', in Laar, R. V. D. (ed.). *Memobust Handbook on Methodology of Modern Business Statistics*.
- Teker, D., Teker, S. & Sönmez, M. (2012). Ekonomik Değişkenlerin Kadın ve Erkeğin Yaşam Süresine Etkisi. *İşletme Araştırmaları Dergisi*, 4 (3), 118-126.

Torres-Reyna, O. (2010). *Getting Started in Fixed/Random Effects Models using R*. Retrieved January 28, 2021 from <https://www.princeton.edu/~otorres/Panel101R.pdf>.

Turkish Statistical Institute (n.d.). Temel İstatistikler. Retrieved November 19, 2019, from <http://www.tuik.gov.tr/UstMenu.do?metod=temelist>.

Turkish Statistical Institute (n.d.). Veritabanları. Merkezi Dağıtım Sistemi. Retrieved November 19, 2019, from <http://www.tuik.gov.tr/PreTabloArama.do?metod=search&araType=vt>.

World Health Organization (2006). WHO Statistical Information System. Retrieved June 22, 2020, from <https://www.who.int/whosis/whostat2006DefinitionsAndMetadata.pdf>.

Xiong, J. & Yi, G. Y. (2019). swgee: An R Package for Analyzing Longitudinal Data with Response Missingness and Covariate Measurement Error. *The R Journal*, 11 (1), 416-426.

Yi, G. Y. (2008). A Simulation-based Marginal Method for Longitudinal Data with Dropout and Mismeasured Covariates. *Biostatistics*, 9 (3), 501–512.

Zha, X., Tian, Y., Gao, X., Wang, W. & Yu, C. (2019). Quantitatively Evaluate the Environmental Impact Factors of the Life Expectancy in Tibet, China. *Environmental Geochemistry and Health*, 41, 1507–1520.



APPENDICES

A. R Codes

#reading the data into R:

```
data=read.table(file="clipboard",sep="\t",header=TRUE)
```

```
data$Year=as.numeric(data$Year)
```

```
attach(data)
```

```
summary(data)
```

#correlation matrix of numeric variables:

```
data1=data[,-(1:2)]
```

```
cor(data1)
```

#scatterplot of life expectancy at birth versus year:

```
scatter.smooth(data$Year,data$Life.Expectancy.at.Birth,main="Scatterplot of Life  
Expectancy at Birth versus Year",xlab="Year",ylab="Life Expectancy at Birth  
(Years)",xaxt='n')
```

```
axis(side = 1, at=2013:2015)
```

```
y1=data[which(data$Year==2013),]
```

```
y2=data[which(data$Year==2014),]
```

```
y3=data[which(data$Year==2015),]
```

```
par(mfrow=c(2,2))
```

```
hist(y1$Life.Expectancy.at.Birth,main="Histogram of Life Expectancy at Birth for  
2013",xlab="Life Expectancy at Birth (Years)")
```

```
hist(y2$Life.Expectancy.at.Birth,main="Histogram of Life Expectancy at Birth for 2014",xlab="Life Expectancy at Birth (Years)")
```

```
hist(y3$Life.Expectancy.at.Birth,main="Histogram of Life Expectancy at Birth for 2015",xlab="Life Expectancy at Birth (Years)")
```

```
shapiro.test(y1$Life.Expectancy.at.Birth)
```

```
shapiro.test(y2$Life.Expectancy.at.Birth)
```

```
shapiro.test(y3$Life.Expectancy.at.Birth)
```

```
#correlation matrix of years:
```

```
matrix=cbind(y1$Life.Expectancy.at.Birth,y2$Life.Expectancy.at.Birth,y3$Life.Expectancy.at.Birth)
```

```
cor(matrix)
```

```
#checking for cross-sectional dependence:
```

```
library(plm)
```

```
pcdtest(plm(y1$Life.Expectancy.at.Birth~1,data=data,model="pooling"))
```

```
pcdtest(plm(y2$Life.Expectancy.at.Birth~1,data=data,model="pooling"))
```

```
pcdtest(plm(y3$Life.Expectancy.at.Birth~1,data=data,model="pooling"))
```

```
#Random Effects Models:
```

```
library(MASS)
```

```
library(plm)
```

```
re=plm(Life.Expectancy.at.Birth~Year+Population.Density+Number.of.Housing.Sales+GDP+Net.Migration.Rate+Number.of.Theater.Halls+Number.of.Cinema.Halls+Agricultural.Field+Export..TL.YTL.+Import..TL.YTL.+Crude.Birth.Rate..Per.Thousand.+Crude.Marriage.Rate..Per.Thousand.+Crude.Death.Rate..Per.Thousand.+
```

```
Infant.Mortality.Rate..Per.Thousand.+Number.of.Students.Per.Classroom,data=dat  
a,model="random",index=c("Province"))
```

```
summary(re)
```

```
library(car)
```

```
vif(re)
```

```
#model with variables that have VIF values less than 10:
```

```
re75=plm(Life.Expectancy.at.Birth~Year+Net.Migration.Rate+Number.of.Theater.  
Halls+Agricultural.Field+Crude.Birth.Rate..Per.Thousand.+Crude.Marriage.Rate..  
Per.Thousand.+Crude.Death.Rate..Per.Thousand.+Infant.Mortality.Rate..Per.Thous  
and.+Number.of.Students.Per.Classroom,data=data,model="random",index=c("Pro  
vince"))
```

```
summary(re75)
```

```
vif(re75)
```

```
#model with significant variables:
```

```
re1=plm(Life.Expectancy.at.Birth~Crude.Birth.Rate..Per.Thousand.+Crude.Marria  
ge.Rate..Per.Thousand.+Crude.Death.Rate..Per.Thousand.+Infant.Mortality.Rate..P  
er.Thousand.+Number.of.Students.Per.Classroom,data=data,model="random",inde  
x=c("Province"))
```

```
summary(re1)
```

```
#getting residuals:
```

```
res3=resid(re1)
```

```
stdres3=scale(res3)
```

```
#getting rmse:
```

```
library(Metrics)
```

```

y3=data$Life.Expectancy.at.Birth

est3=y3-res3

rmse(y3,est3)

#standardized residuals vs. year (random effects model):

scatter.smooth(data$Year,stdres3,main="Scatterplot of Standardized Residuals of
Random Effects Model versus Year",xlab="Year",ylab="Standardized Residuals of
Random Effects Model",ylim=c(-3,4),xaxt='n')

axis(side = 1, at=2013:2015)

#standardized residuals vs. covariates (random effects model):

scatter.smooth(data$Crude.Birth.Rate..Per.Thousand.,stdres3,main="Scatterplot of
Standardized Residuals of Random Effects Model versus Crude Birth
Rate",xlab="Crude Birth Rate (Per Thousand)",ylab="Standardized Residuals of
Random Effects Model",ylim=c(-3,4))

scatter.smooth(data$Crude.Marriage.Rate..Per.Thousand.,stdres3,main="Scatterplo
t of Standardized Residuals of Random Effects Model versus Crude Marriage
Rate",xlab="Crude Marriage Rate (Per Thousand)",ylab="Standardized Residuals
of Random Effects Model",ylim=c(-3,4))

scatter.smooth(data$Crude.Death.Rate..Per.Thousand.,stdres3,main="Scatterplot of
Standardized Residuals of Random Effects Model versus Crude Death
Rate",xlab="Crude Death Rate (Per Thousand)",ylab="Standardized Residuals of
Random Effects Model",ylim=c(-3,4))

scatter.smooth(data$Infant.Mortality.Rate..Per.Thousand.,stdres3,main="Scatterplo
t of Standardized Residuals of Random Effects Model versus Infant Mortality
Rate",xlab="Infant Mortality Rate (Per Thousand)",ylab="Standardized Residuals
of Random Effects Model",ylim=c(-3,4))

```

```
scatter.smooth(data$Number.of.Students.Per.Classroom,stdres3,main="Scatterplot
of Standardized Residuals of Random Effects Model versus Number of Students
Per Classroom",xlab="Crude Marriage Rate (Per Thousand)",ylab="Standardized
Residuals of Random Effects Model",ylim=c(-3,4))
```

```
#fitted response values vs. real response values (random effects model):
```

```
scatter.smooth(data$Life.Expectancy.at.Birth,est3,main="Scatterplot of Fitted
Response Values of Random Effects Model versus Life Expectancy at
Birth",xlab="Life Expectancy at Birth (Years)",ylab="Fitted Response Values of
Random Effects Model")
```

```
#random effects model with year variable in the random part:
```

```
library(lme4)
```

```
m1=lmer(Life.Expectancy.at.Birth~Year+Crude.Birth.Rate..Per.Thousand.+Crude.
Marriage.Rate..Per.Thousand.+Crude.Death.Rate..Per.Thousand.+Infant.Mortality.
Rate..Per.Thousand.+Number.of.Students.Per.Classroom+(1|Year),data=data,REM
L=FALSE)
```

```
summary(m1)
```

```
#defining dummy variables:
```

```
dre1=ifelse(data$Crude.Birth.Rate..Per.Thousand.<=17,1,0)
```

```
dre2=ifelse(data$Crude.Marriage.Rate..Per.Thousand.>8,1,0)
```

```
#model with dummy variables and significant variables:
```

```
re12=plm(Life.Expectancy.at.Birth~Crude.Birth.Rate..Per.Thousand.+Crude.Birth.
Rate..Per.Thousand.*dre1+I(Crude.Birth.Rate..Per.Thousand.^2)*dre1+Crude.Mar
riage.Rate..Per.Thousand.+Crude.Marriage.Rate..Per.Thousand.*dre2+dre2+Crude.
Death.Rate..Per.Thousand.+I(Crude.Death.Rate..Per.Thousand.^2)+Infant.Mortalit
y.Rate..Per.Thousand.+Number.of.Students.Per.Classroom,data=data,model="rand
om",index=c("Province"))
```

```
summary(re12)
```

```
vif(re12)
```

```
#model with variables that have VIF values less than 10:
```

```
re175=plm(Life.Expectancy.at.Birth~Crude.Marriage.Rate..Per.Thousand.+Infant.  
Mortality.Rate..Per.Thousand.+Number.of.Students.Per.Classroom,data=data,mode  
l="random",index=c("Province"))
```

```
summary(re175)
```

```
vif(re175)
```

```
#model with significant variables:
```

```
re13=plm(Life.Expectancy.at.Birth~Crude.Marriage.Rate..Per.Thousand.+Infant.M  
ortality.Rate..Per.Thousand.,data=data,model="random",index=c("Province"))
```

```
summary(re13)
```

```
#getting residuals:
```

```
res31=resid(re13)
```

```
stdres31=scale(res31)
```

```
#standardized residuals vs. year (random effects model with dummies):
```

```
scatter.smooth(data$Year,stdres31,main="Scatterplot of Standardized Residuals of  
Random Effects Model versus Year",xlab="Year",ylab="Standardized Residuals of  
Random Effects Model",ylim=c(-3,4),xaxt='n')
```

```
axis(side = 1, at=2013:2015)
```

```
#standardized residuals vs. covariates (random effects model with dummies):
```

```
scatter.smooth(data$Crude.Marriage.Rate..Per.Thousand.,stdres31,main="Scatterpl  
ot of Standardized Residuals of Random Effects Model versus Crude Marriage
```

```
Rate",xlab="Crude Marriage Rate (Per Thousand)",ylab="Standardized Residuals  
of Random Effects Model",ylim=c(-3,4))
```

```
scatter.smooth(data$Infant.Mortality.Rate..Per.Thousand.,stdres31,main="Scatterplot  
of Standardized Residuals of Random Effects Model versus Infant Mortality  
Rate",xlab="Infant Mortality Rate (Per Thousand)",ylab="Standardized Residuals  
of Random Effects Model",ylim=c(-3,4))
```

```
#fitted response values vs. real response values (random effects model with  
dummies):
```

```
scatter.smooth(data$Life.Expectancy.at.Birth,est31,main="Scatterplot of Fitted  
Response Values of Random Effects Model versus Life Expectancy at  
Birth",xlab="Life Expectancy at Birth (Years)",ylab="Fitted Response Values of  
Random Effects Model")
```

```
#Marginal Models:
```

```
ordered=order(as.integer(data$Province))
```

```
data2=data[ordered,]
```

```
library(gee)
```

```
mar=gee(Life.Expectancy.at.Birth~Year+Population.Density+Number.of.Housing.  
Sales+GDP+Net.Migration.Rate+Number.of.Theater.Halls+Number.of.Cinema.Ha  
lls+Agricultural.Field+Export..TL.YTL.+Import..TL.YTL.+Crude.Birth.Rate..Per.  
Thousand.+Crude.Marriage.Rate..Per.Thousand.+Crude.Death.Rate..Per.Thousand.  
+Infant.Mortality.Rate..Per.Thousand.+Number.of.Students.Per.Classroom,data=da  
ta2,id=Province,family=gaussian,corstr="exchangeable")
```

```
summary(mar)
```

```
#model with variables that have VIF values less than 10:
```

```
mar75=gee(Life.Expectancy.at.Birth~Year+Net.Migration.Rate+Number.of.Theate  
r.Halls+Agricultural.Field+Crude.Birth.Rate..Per.Thousand.+Crude.Marriage.Rate.
```

```
.Per.Thousand.+Crude.Death.Rate..Per.Thousand.+Infant.Mortality.Rate..Per.Thou  
sand.+Number.of.Students.Per.Classroom,data=data2,id=Province,family=gaussian  
,corstr="exchangeable")
```

```
summary(mar75)
```

```
#model with significant variables:
```

```
mar1=gee(Life.Expectancy.at.Birth~Crude.Birth.Rate..Per.Thousand.+Crude.Marri  
age.Rate..Per.Thousand.+Crude.Death.Rate..Per.Thousand.+Infant.Mortality.Rate..  
Per.Thousand.,data=data2,id=Province,family=gaussian,corstr="exchangeable")
```

```
summary(mar1)
```

```
armar1=gee(Life.Expectancy.at.Birth~Crude.Birth.Rate..Per.Thousand.+Crude.Ma  
rriage.Rate..Per.Thousand.+Crude.Death.Rate..Per.Thousand.+Infant.Mortality.Rat  
e..Per.Thousand.,data=data2,id=Province,family=gaussian,corstr="AR-M")
```

```
summary(armar1)
```

```
unstrmar1=gee(Life.Expectancy.at.Birth~Crude.Birth.Rate..Per.Thousand.+Crude.  
Marriage.Rate..Per.Thousand.+Crude.Death.Rate..Per.Thousand.+Infant.Mortality.  
Rate..Per.Thousand.,data=data2,id=Province,family=gaussian,corstr="unstructured  
")
```

```
summary(unstrmar1)
```

```
#calculating QIC values of final marginal models:
```

```
library(MuMIn)
```

```
QIC(mar1)
```

```
QIC(armar1)
```

```
#getting residuals:
```

```
res1=mar1$resid
```

```
stdres1=scale(res1)
```

```
#getting rmse:
```

```
library(Metrics)
```

```
y1=data2$Life.Expectancy.at.Birth
```

```
est1=y1-res1
```

```
rmse(y1,est1)
```

```
#standardized residuals vs. year (marginal model):
```

```
scatter.smooth(data2$Year,stdres1,main="Scatterplot of Standardized Residuals of  
Marginal Model versus Year",xlab="Year",ylab="Standardized Residuals of  
Marginal Model",ylim=c(-3,4),xaxt='n')
```

```
axis(side = 1, at=2013:2015)
```

```
#standardized residual vs. covariates (marginal model):
```

```
scatter.smooth(data2$Crude.Birth.Rate..Per.Thousand.,stdres1,main="Scatterplot  
of Standardized Residuals of Marginal Model versus Crude Birth  
Rate",xlab="Crude Birth Rate (Per Thousand)",ylab="Standardized Residuals of  
Marginal Model",ylim=c(-3,4))
```

```
scatter.smooth(data2$Crude.Marriage.Rate..Per.Thousand.,stdres1,main="Scatterpl  
ot of Standardized Residuals of Marginal Model versus Crude Marriage  
Rate",xlab="Crude Marriage Rate (Per Thousand)",ylab="Standardized Residuals  
of Marginal Model",ylim=c(-3,4))
```

```
scatter.smooth(data2$Crude.Death.Rate..Per.Thousand.,stdres1,main="Scatterplot  
of Standardized Residuals of Marginal Model versus Crude Death  
Rate",xlab="Crude Death Rate (Per Thousand)",ylab="Standardized Residuals of  
Marginal Model",ylim=c(-3,4))
```

```
scatter.smooth(data$Infant.Mortality.Rate..Per.Thousand.,stdres1,main="Scatterplo  
t of Standardized Residuals of Marginal Model versus Infant Mortality
```

```
Rate",xlab="Infant Mortality Rate (Per Thousand)",ylab="Standardized Residuals  
of Marginal Model",ylim=c(-3,4))
```

```
#fitted response values vs. real response values (marginal model):
```

```
scatter.smooth(data2$Life.Expectancy.at.Birth,est1,main="Scatterplot of Fitted  
Response Values of Marginal Model versus Life Expectancy at Birth",xlab="Life  
Expectancy at Birth (Years)",ylab="Fitted Response Values of Marginal Model")
```

```
#finding max. and min. estimated life expectancy at birth values:
```

```
estimates=cbind(data2$Province,data2$Year,est1)
```

```
estimates
```

```
#defining dummy variables:
```

```
dmar1=ifelse(data2$Crude.Birth.Rate..Per.Thousand.<=17,1,0)
```

```
dmar2=ifelse(data2$Crude.Marriage.Rate..Per.Thousand.>8,1,0)
```

```
#model with dummy variables and significant variables:
```

```
mar12=gee(Life.Expectancy.at.Birth~Crude.Birth.Rate..Per.Thousand.+Crude.Birt  
h.Rate..Per.Thousand.*dmar1+I(Crude.Birth.Rate..Per.Thousand.^2)*dmar1+Crud  
e.Marriage.Rate..Per.Thousand.+Crude.Marriage.Rate..Per.Thousand.*dmar2+dma  
r2+Crude.Death.Rate..Per.Thousand.+I(Crude.Death.Rate..Per.Thousand.^2)+Infan  
t.Mortality.Rate..Per.Thousand.,data=data2,id=Province,family=gaussian,corstr="e  
xchangeable")
```

```
summary(mar12)
```

```
#model with variables that have VIF values less than 10:
```

```
mar175=gee(Life.Expectancy.at.Birth~Crude.Marriage.Rate..Per.Thousand.+Infant  
.Mortality.Rate..Per.Thousand.,data=data2,id=Province,family=gaussian,corstr="ex  
changeable")
```

```
summary(mar175)
```

```

#getting residuals:

res11=mar175$resid

stdres11=scale(res11)

y11=data2$Life.Expectancy.at.Birth

est11=y11-res11

#standardized residuals vs. year (marginal model with dummies):

scatter.smooth(data2$Year,stdres11,main="Scatterplot of Standardized Residuals
of Marginal Model versus Year",xlab="Year",ylab="Standardized Residuals of
Marginal Model",ylim=c(-3,4),xaxt='n')

axis(side = 1, at=2013:2015)

#standardized residuals vs. covariates (marginal model with dummies):

scatter.smooth(data2$Crude.Marriage.Rate..Per.Thousand.,stdres11,main="Scatter
plot of Standardized Residuals of Marginal Model versus Crude Marriage
Rate",xlab="Crude Marriage Rate (Per Thousand)",ylab="Standardized Residuals
of Marginal Model",ylim=c(-3,4))

scatter.smooth(data$Infant.Mortality.Rate..Per.Thousand.,stdres11,main="Scatterpl
ot of Standardized Residuals of Marginal Model versus Infant Mortality
Rate",xlab="Infant Mortality Rate (Per Thousand)",ylab="Standardized Residuals
of Marginal Model",ylim=c(-3,4))

#fitted response values vs. real response values (marginal model with dummies):

scatter.smooth(data2$Life.Expectancy.at.Birth,est11,main="Scatterplot of Fitted
Response Values of Marginal Model versus Life Expectancy at Birth",xlab="Life
Expectancy at Birth (Years)",ylab="Fitted Response Values of Marginal Model")

#Transition Models:

datatrans=read.table(file="clipboard",sep="\t",header=TRUE)

```

```

osub=order(as.integer(datatrans$Province))

datatrans=datatrans[osub,]

library(gee)

trans=gee(Life.Expectancy.at.Birth~Year+Population.Density+Number.of.Housing
.Sales+GDP+Net.Migration.Rate+Number.of.Theater.Halls+Number.of.Cinema.H
alls+Agricultural.Field+Export..TL.YTL.+Import..TL.YTL.+Crude.Birth.Rate..Per.
Thousand.+Crude.Marriage.Rate..Per.Thousand.+Crude.Death.Rate..Per.Thousand.
+Infant.Mortality.Rate..Per.Thousand.+Number.of.Students.Per.Classroom+Life.E
xpectancy.at.Birth_lag1,data=datatrans,id=Province,family=gaussian,corstr="indep
endence")

summary(trans)

#model with variables that have VIF values less than 10:

trans75=gee(Life.Expectancy.at.Birth~Year+Net.Migration.Rate+Number.of.Theat
er.Halls+Agricultural.Field+Crude.Birth.Rate..Per.Thousand.+Crude.Marriage.Rat
e..Per.Thousand.+Crude.Death.Rate..Per.Thousand.+Infant.Mortality.Rate..Per.Tho
usand.+Number.of.Students.Per.Classroom+Life.Expectancy.at.Birth_lag1,data=da
tatrans,id=Province,family=gaussian,corstr="independence")

summary(trans75)

#model with significant variables:

trans1=gee(Life.Expectancy.at.Birth~Year+Crude.Death.Rate..Per.Thousand.+Life
.Expectancy.at.Birth_lag1,data=datatrans,id=Province,family=gaussian,corstr="ind
ependence")

summary(trans1)

#getting residuals:

res2=trans1$resid

```

```

stdres2=scale(res2)

#getting rmse:

library(Metrics)

y2=datatrans$Life.Expectancy.at.Birth

est2=y2-res2

rmse(y2,est2)

#standardized residuals vs. year (transition model):

scatter.smooth(datatrans$Year,stdres2,main="Scatterplot of Standardized Residuals
of Transition Model versus Year",xlab="Year",ylab="Standardized Residuals of
Transition Model",ylim=c(-4,5),xaxt='n')

axis(side = 1, at=2014:2015)

#standardized residuals vs. covariates (transition model):

scatter.smooth(datatrans$Year,stdres2,main="Scatterplot of Standardized Residuals
of Transition Model versus Year",xlab="Year",ylab="Standardized Residuals of
Transition Model",ylim=c(-3,4),xaxt='n')

axis(side = 1, at=2014:2015)

scatter.smooth(datatrans$Crude.Death.Rate..Per.Thousand.,stdres2,main="Scatterp
lot of Standardized Residuals of Transition Model versus Crude Death
Rate",xlab="Crude Death Rate (Per Thousand)",ylab="Standardized Residuals of
Transition Model",ylim=c(-3,4))

scatter.smooth(datatrans$Life.Expectancy.at.Birth_lag1,stdres2,main="Scatterplot
of Standardized Residuals of Transition Model versus Life Expectancy at Birth Lag
1",xlab="Life Expectancy at Birth Lag 1 (Years)",ylab="Standardized Residuals of
Transition Model",ylim=c(-3,4))

#fitted response values vs. real response values (transition model):

```

```

scatter.smooth(datatrans$Life.Expectancy.at.Birth,est2,main="Scatterplot of Fitted
Response Values of Transition Model versus Life Expectancy at Birth",xlab="Life
Expectancy at Birth (Years)",ylab="Fitted Response Values of Transition Model")

#defining dummy variables:

dtrans1=ifelse(datatrans$Crude.Birth.Rate..Per.Thousand.<=17,1,0)

dtrans2=ifelse(datatrans$Crude.Marriage.Rate..Per.Thousand.>8,1,0)

#model with dummy variables and significant variables:

trans12=gee(Life.Expectancy.at.Birth~Year+Crude.Death.Rate..Per.Thousand.+I(C
rude.Death.Rate..Per.Thousand.^2)+Crude.Birth.Rate..Per.Thousand.+Crude.Birth.
Rate..Per.Thousand.*dtrans1+I(Crude.Birth.Rate..Per.Thousand.^2)*dtrans1+Crud
e.Marriage.Rate..Per.Thousand.+Crude.Marriage.Rate..Per.Thousand.*dtrans2+dtr
ans2+Life.Expectancy.at.Birth_lag1,data=datatrans,id=Province,family=gaussian,c
orstr="independence")

summary(trans12)

#model with variables that have VIF values less than 10:

trans175=gee(Life.Expectancy.at.Birth~Year+Crude.Marriage.Rate..Per.Thousand.
+Life.Expectancy.at.Birth_lag1,data=datatrans,id=Province,family=gaussian,corstr
="independence")

summary(trans175)

#model with significant variables:

trans13=gee(Life.Expectancy.at.Birth~Year+Life.Expectancy.at.Birth_lag1,data=d
atrans,id=Province,family=gaussian,corstr="independence")

summary(trans13)

#getting residuals:

res21=trans13$resid

```

```
stdres21=scale(res21)
```

```
y21=datatrans$Life.Expectancy.at.Birth
```

```
est21=y21-res21
```

```
#standardized residuals vs. year (transition model with dummies):
```

```
scatter.smooth(datatrans$Year,stdres21,main="Scatterplot of Standardized  
Residuals of Transition Model versus Year",xlab="Year",ylab="Standardized  
Residuals of Transition Model",ylim=c(-4,5),xaxt='n')
```

```
axis(side = 1, at=2014:2015)
```

```
#standardized residuals vs. covariates (transition model with dummies):
```

```
scatter.smooth(datatrans$Year,stdres21,main="Scatterplot of Standardized  
Residuals of Transition Model versus Year",xlab="Year",ylab="Standardized  
Residuals of Transition Model",ylim=c(-3,4),xaxt='n')
```

```
axis(side = 1, at=2014:2015)
```

```
scatter.smooth(datatrans$Life.Expectancy.at.Birth_lag1,stdres21,main="Scatterplot  
of Standardized Residuals of Transition Model versus Life Expectancy at Birth Lag  
1",xlab="Life Expectancy at Birth Lag 1 (Years)",ylab="Standardized Residuals of  
Transition Model",ylim=c(-3,4))
```

```
#fitted response values vs. real response values (transition model with dummies):
```

```
scatter.smooth(datatrans$Life.Expectancy.at.Birth,est21,main="Scatterplot of  
Fitted Response Values of Transition Model versus Life Expectancy at  
Birth",xlab="Life Expectancy at Birth (Years)",ylab="Fitted Response Values of  
Transition Model")
```

```
#response vs. covariates:
```

```
scatter.smooth(Crude.Birth.Rate..Per.Thousand.,Life.Expectancy.at.Birth,main="Scatterplot of Life Expectancy at Birth versus Crude Birth Rate (Per Thousand)",ylab="Life Expectancy at Birth (Years)")
```

```
scatter.smooth(Crude.Marriage.Rate..Per.Thousand.,Life.Expectancy.at.Birth,main="Scatterplot of Life Expectancy at Birth versus Crude Marriage Rate",xlab="Crude Marriage Rate (Per Thousand)",ylab="Life Expectancy at Birth (Years)")
```

```
scatter.smooth(Crude.Death.Rate..Per.Thousand.,Life.Expectancy.at.Birth,main="Scatterplot of Life Expectancy at Birth versus Crude Death Rate",xlab="Crude Death Rate (Per Thousand)",ylab="Life Expectancy at Birth (Years)")
```

```
scatter.smooth(Infant.Mortality.Rate..Per.Thousand.,Life.Expectancy.at.Birth,main="Scatterplot of Life Expectancy at Birth versus Infant Mortality Rate",xlab="Infant Mortality Rate (Per Thousand)",ylab="Life Expectancy at Birth (Years)")
```

```
scatter.smooth(Number.of.Students.Per.Classroom,Life.Expectancy.at.Birth,main="Scatterplot of Life Expectancy at Birth versus Number of Students Per Classroom",xlab="Number of Students Per Classroom",ylab="Life Expectancy at Birth (Years)")
```

```
scatter.smooth(datatrans$Life.Expectancy.at.Birth_lag1,datatrans$Life.Expectancy.at.Birth,main="Scatterplot of Life Expectancy at Birth versus Life Expectancy at Birth Lag 1",xlab="Life Expectancy at Birth Lag 1 (Years)",ylab="Life Expectancy at Birth (Years)")
```

```
scatter.smooth(datatrans$Year,datatrans$Life.Expectancy.at.Birth,main="Scatterplot of Life Expectancy at Birth versus Year",xlab="Year",ylab="Life Expectancy at Birth (Years)",xaxt='n')
```

```
axis(side = 1, at=2014:2015)
```

```
#Measurement Error part:
```

```

data3=read.table(file="clipboard",sep="\t",header=TRUE)

data3$Life.Expectancy.at.Birth=as.integer(data3$Life.Expectancy.at.Birth)

data3$Year=as.numeric(data3$Year)

attach(data3)

library(swgee)

sigmaVals <- c(0,0.5,1)

for(i in 1:3){
  for(j in 1:3){
    for(k in 1:3){
      for(m in 1:3){
        sigma=diag(c(sigmaVals[i],sigmaVals[j],sigmaVals[k],sigmaVals[m]))

        output =
        swgee(Life.Expectancy.at.Birth~Crude.Birth.Rate..Per.Thousand.+Crude.Marriage.
        Rate..Per.Thousand.+Crude.Death.Rate..Per.Thousand.+Infant.Mortality.Rate..Per.
        Thousand.,data=data3,id=Province,family=gaussian,corstr="exchangeable",missin
        gmodel=Missing~Life.Expectancy.at.Birth,SIMEXvariable=c("Crude.Birth.Rate..P
        er.Thousand.","Crude.Marriage.Rate..Per.Thousand.","Crude.Death.Rate..Per.Thou
        sand.","Infant.Mortality.Rate..Per.Thousand."),SIMEX.err=sigma,repeated=FALS
        E,lambda=seq(0,1,0.5))

        fileName=paste(c(sigmaVals[i],sigmaVals[j],sigmaVals[k],sigmaVals[m]),collapse
        ="_")

        sink(paste(fileName,".txt",collapse="_"))

        print(summary(output))

        sink()

```

```
}
```

```
}
```

```
}
```

```
}
```

```
coeff=read.table(file="clipboard",sep="\t",header=TRUE)
```

```
attach(coeff)
```

```
plot(ID,Beta_0,type="b",main="Scatterplot of Estimated Intercept Values versus  
IDs of Variance Combinations",xlab="IDs of Variance  
Combinations",ylab="Estimated Intercept Values",pch=16,xaxt='n')
```

```
axis(side = 1, at=1:81,las=2)
```

```
plot(ID,Beta_1,type="b",main="Scatterplot of Estimated Coefficients of Crude  
Birth Rate versus IDs of Variance Combinations",xlab="IDs of Variance  
Combinations",ylab="Estimated Estimated Coefficients of Crude Birth  
Rate",pch=16,xaxt='n')
```

```
axis(side = 1, at=1:81,las=2)
```

```
plot(ID,Beta_2,type="b",main="Scatterplot of Estimated Coefficients of Crude  
Marriage Rate versus IDs of Variance Combinations",xlab="IDs of Variance  
Combinations",ylab="Estimated Estimated Coefficients of Crude Marriage  
Rate",pch=16,xaxt='n')
```

```
axis(side = 1, at=1:81,las=2)
```

```
plot(ID,Beta_3,type="b",main="Scatterplot of Estimated Coefficients of Crude  
Death Rate versus IDs of Variance Combinations",xlab="IDs of Variance  
Combinations",ylab="Estimated Estimated Coefficients of Crude Death  
Rate",pch=16,xaxt='n')
```

```
axis(side = 1, at=1:81,las=2)
```

```

sym=rep(16,nrow(coeff))

sym[c(8,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,31,34,37,38,39,40,
41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,58,59,61,63,64,65,66,67,68,69,70,71
,72,73,74,75,76,77,78,79,80,81)]=4

plot(ID,Beta_4,type="b",main="Scatterplot of Estimated Coefficients of Infant
Mortality Rate versus IDs of Sigma Values",xlab="IDs of Sigma
Values",ylab="Estimated Estimated Coefficients of Infant Mortality
Rate",pch=sym,xaxt='n')

axis(side = 1, at=1:81,las=2)

```

B. Correlation Matrix of Numeric Variables

Table B.1 Correlation Table of Numeric Variables

	<i>LEAB</i>	<i>PD</i>	<i>Population</i>	<i>NOFHHS</i>	<i>NOSHHS</i>
LEAB	1	0.055	0.068	0.108	0.128
PD	0.055	1	0.92	0.869	0.82
Population	0.068	0.92	1	0.974	0.951
NOFHHS	0.108	0.869	0.974	1	0.988
NOSHHS	0.128	0.82	0.951	0.988	1
NOHS	0.119	0.844	0.964	0.996	0.998
GDP	0.088	0.951	0.977	0.963	0.932
NM	0.215	0.532	0.608	0.728	0.753
NMR	0.271	0.126	0.143	0.227	0.239
NOTH	0.115	0.948	0.954	0.925	0.896
NOCH	0.096	0.959	0.972	0.954	0.922
AF	-0.096	-0.081	0.199	0.219	0.228
Export (TL)	0.056	0.980	0.93	0.885	0.833
Import (TL)	0.056	0.984	0.921	0.873	0.822
CBR	-0.27	0.011	0.047	-0.047	-0.071
CMR	-0.361	0.073	0.107	0.045	0.039
CDR	0.015	-0.179	-0.249	-0.19	-0.182
IMR	-0.414	-0.094	-0.087	-0.171	-0.194
NOSPC	-0.17	0.326	0.438	0.367	0.345

Table B.1 (continued)

	<i>NOHS</i>	<i>GDP</i>	<i>NM</i>	<i>NMR</i>	<i>NOTH</i>
LEAB	0.119	0.088	0.215	0.271	0.115
PD	0.844	0.951	0.532	0.126	0.948
Population	0.964	0.977	0.608	0.143	0.954
NOFHHS	0.996	0.963	0.728	0.227	0.925
NOSHHS	0.998	0.932	0.753	0.239	0.896
NOHS	1	0.949	0.744	0.234	0.912
GDP	0.949	1	0.638	0.154	0.977
NM	0.744	0.638	1	0.609	0.601
NMR	0.234	0.154	0.609	1	0.137
NOTH	0.912	0.977	0.601	0.137	1
NOCH	0.939	0.992	0.615	0.138	0.976
AF	0.225	0.084	0.021	-0.028	0.02
Export (TL)	0.859	0.973	0.54	0.11	0.963
Import (TL)	0.846	0.965	0.519	0.092	0.96
CBR	-0.061	-0.024	-0.318	-0.401	-0.053
CMR	0.042	0.044	-0.205	-0.363	0.026
CDR	-0.186	-0.174	0.043	0.205	-0.132
IMR	-0.185	-0.137	-0.384	-0.388	-0.144
NOSPC	0.356	0.324	0.061	-0.078	0.292

Table B.1 (continued)

	<i>NOCH</i>	<i>AF</i>	<i>Export (TL)</i>	<i>Import (TL)</i>	<i>CBR</i>
LEAB	0.096	-0.096	0.056	0.056	-0.27
PD	0.959	-0.081	0.98	0.984	0.011
Population	0.972	0.199	0.93	0.921	0.047
NOFHHS	0.954	0.219	0.885	0.873	-0.047
NOSHHS	0.922	0.228	0.833	0.822	-0.071
NOHS	0.939	0.225	0.859	0.846	-0.061
GDP	0.992	0.084	0.973	0.965	-0.024
NM	0.615	0.021	0.54	0.519	-0.318
NMR	0.138	-0.028	0.11	0.092	-0.401
NOTH	0.976	0.02	0.963	0.96	-0.053
NOCH	1	0.078	0.971	0.973	-0.026
AF	0.078	1	-0.039	-0.051	0.131
Export (TL)	0.971	-0.039	1	0.994	0.0006
Import (TL)	0.973	-0.051	0.994	1	-0.004
CBR	-0.026	0.131	0.0006	-0.004	1
CMR	0.037	0.238	0.05	0.043	0.64
CDR	-0.165	-0.142	-0.156	-0.143	-0.842
IMR	-0.131	0.043	-0.096	-0.093	0.767
NOSPC	0.315	0.239	0.298	0.278	0.705

Table B.1 (continued)

	<i>CMR</i>	<i>CDR</i>	<i>IMR</i>	<i>NOSPC</i>
LEAB	-0.361	0.015	-0.414	-0.17
PD	0.073	-0.179	-0.094	0.326
Population	0.107	-0.249	-0.087	0.438
NOFHHS	0.045	-0.19	-0.171	0.367
NOSHHS	0.039	-0.182	-0.194	0.345
NOHS	0.042	-0.186	-0.185	0.356
GDP	0.044	-0.174	-0.137	0.324
NM	-0.205	0.043	-0.384	0.061
NMR	-0.363	0.205	-0.388	-0.078
NOTH	0.026	-0.132	-0.144	0.292
NOCH	0.037	-0.165	-0.131	0.315
AF	0.238	0.142	0.043	0.239
Export (TL)	0.05	-0.156	-0.096	0.298
Import (TL)	0.043	-0.143	-0.093	0.278
CBR	0.64	-0.842	0.767	0.705
CMR	1	-0.609	0.549	0.486
CDR	-0.609	1	-0.56	-0.759
IMR	0.549	-0.56	1	0.51
NOSPC	0.486	-0.759	0.51	1

C. Regression Diagnostic Plots of Final Random Effects Model

Scatterplot of Standardized Residuals of Random Effects Model versus Crude Birth Rate

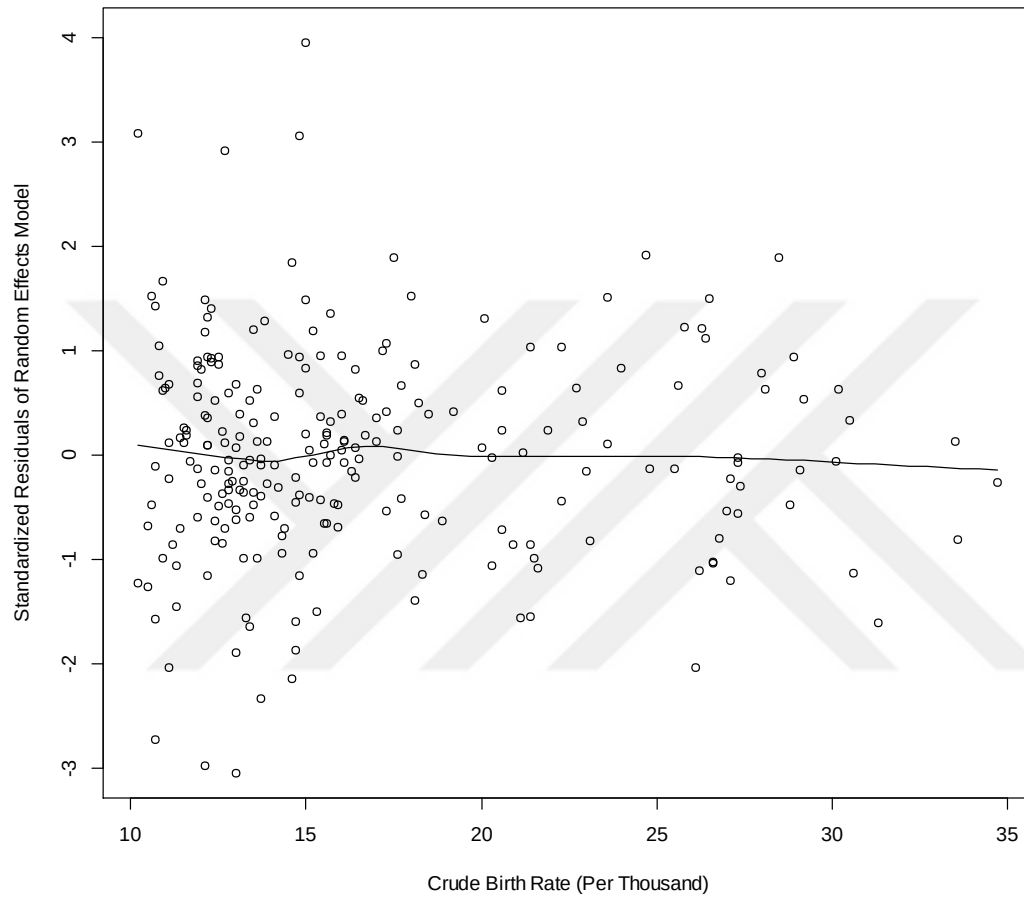


Figure C.1. Scatterplot of Standardized Residuals of Random Effects Model versus Crude Birth Rate

Scatterplot of Standardized Residuals of Random Effects Model versus Crude Marriage Rate

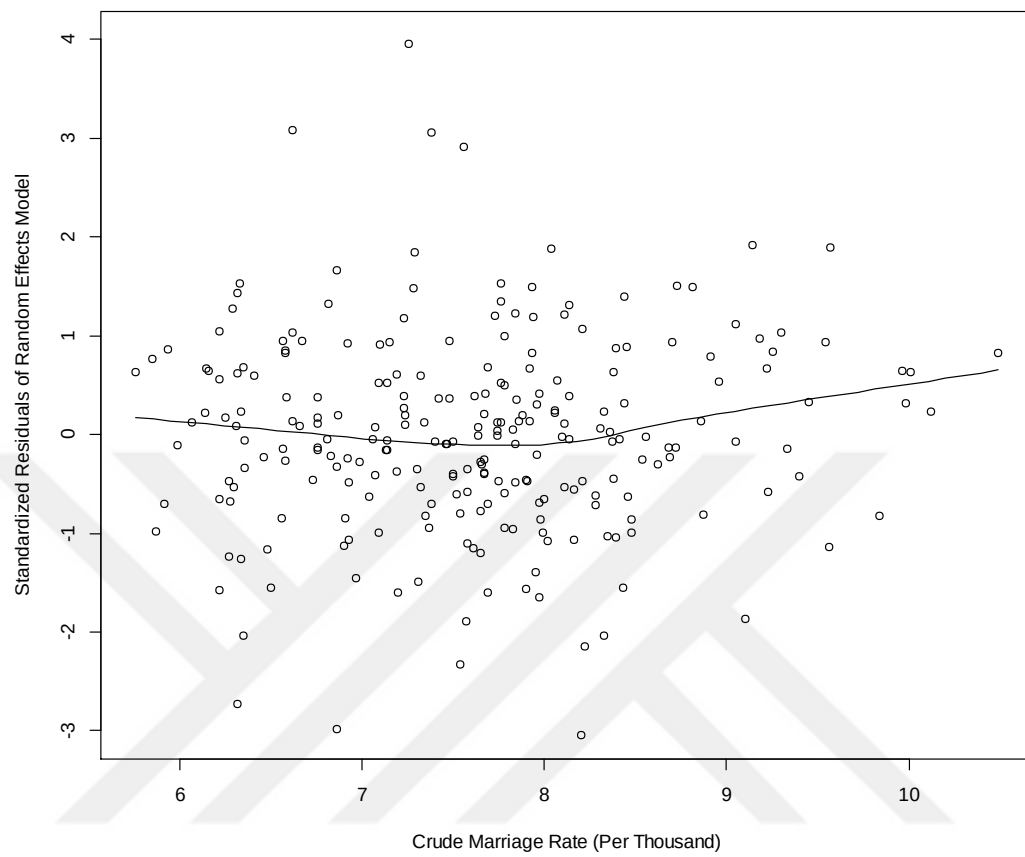


Figure C.2. Scatterplot of Standardized Residuals of Random Effects Model versus Crude Marriage Rate

Scatterplot of Standardized Residuals of Random Effects Model versus Crude Death Rate

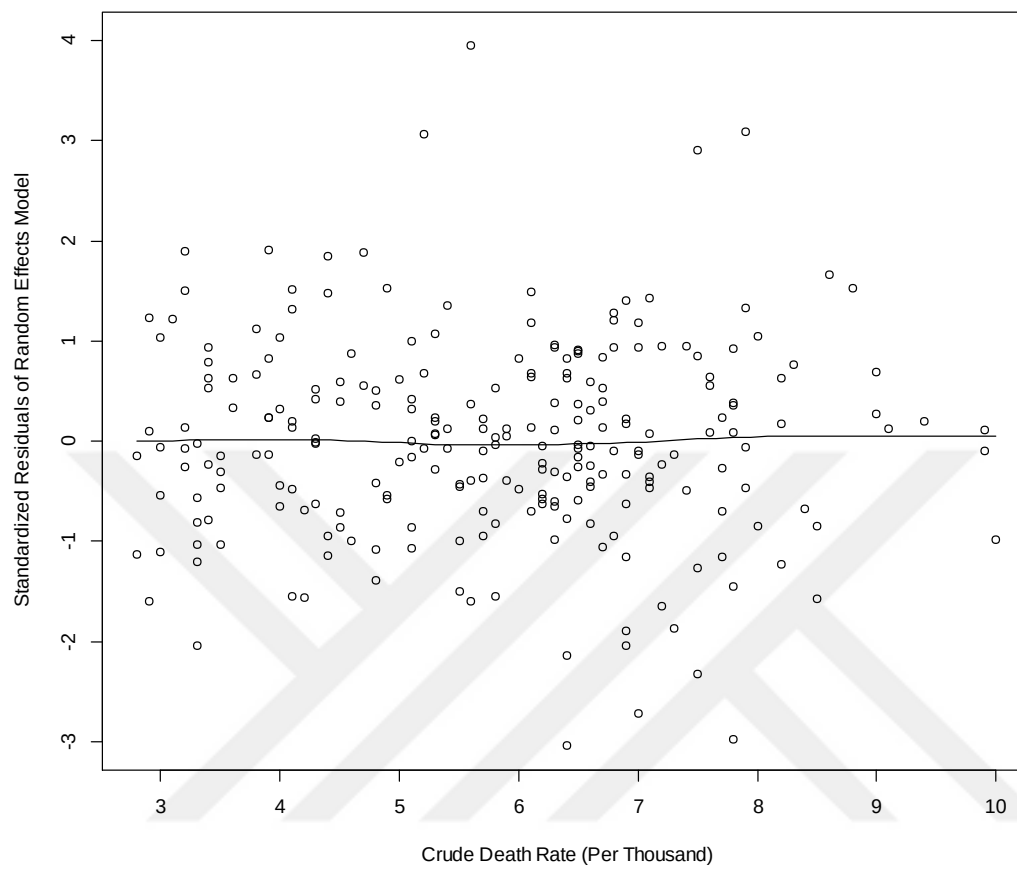


Figure C.3. Scatterplot of Standardized Residuals of Random Effects Model versus Crude Death Rate

Scatterplot of Standardized Residuals of Random Effects Model versus Infant Mortality Rate

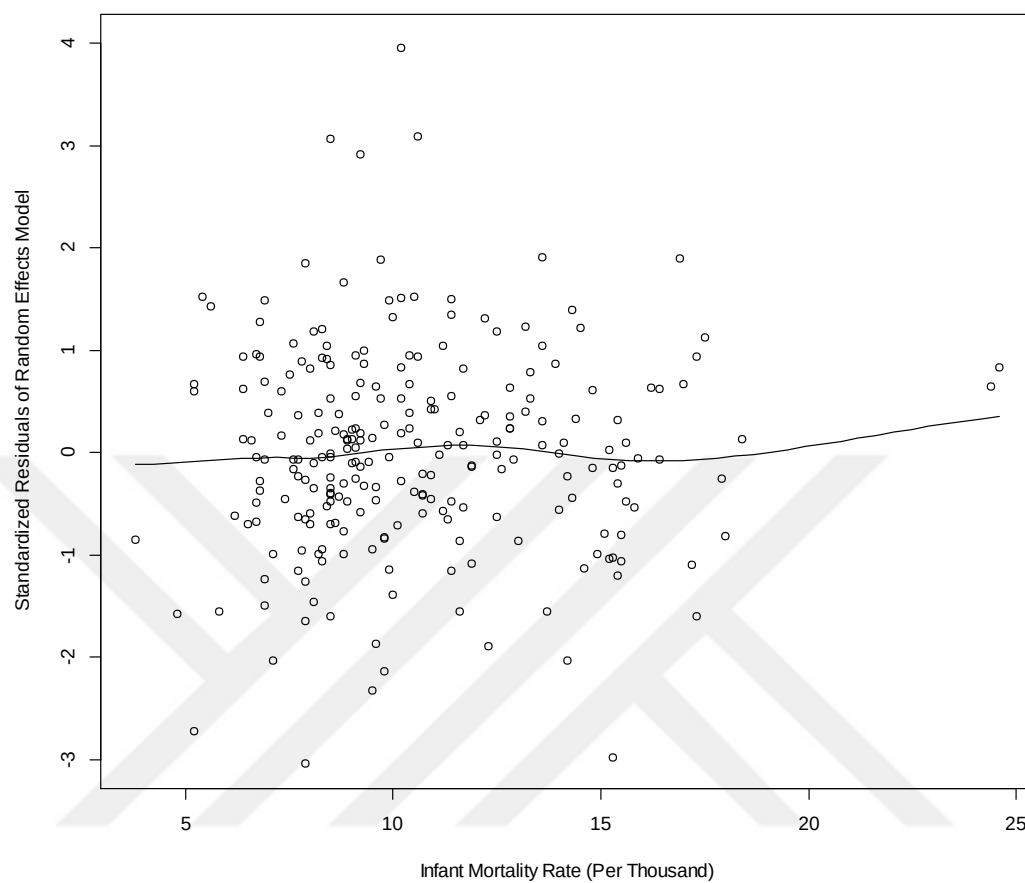


Figure C.4. Scatterplot of Standardized Residuals of Random Effects Model versus Infant Mortality Rate

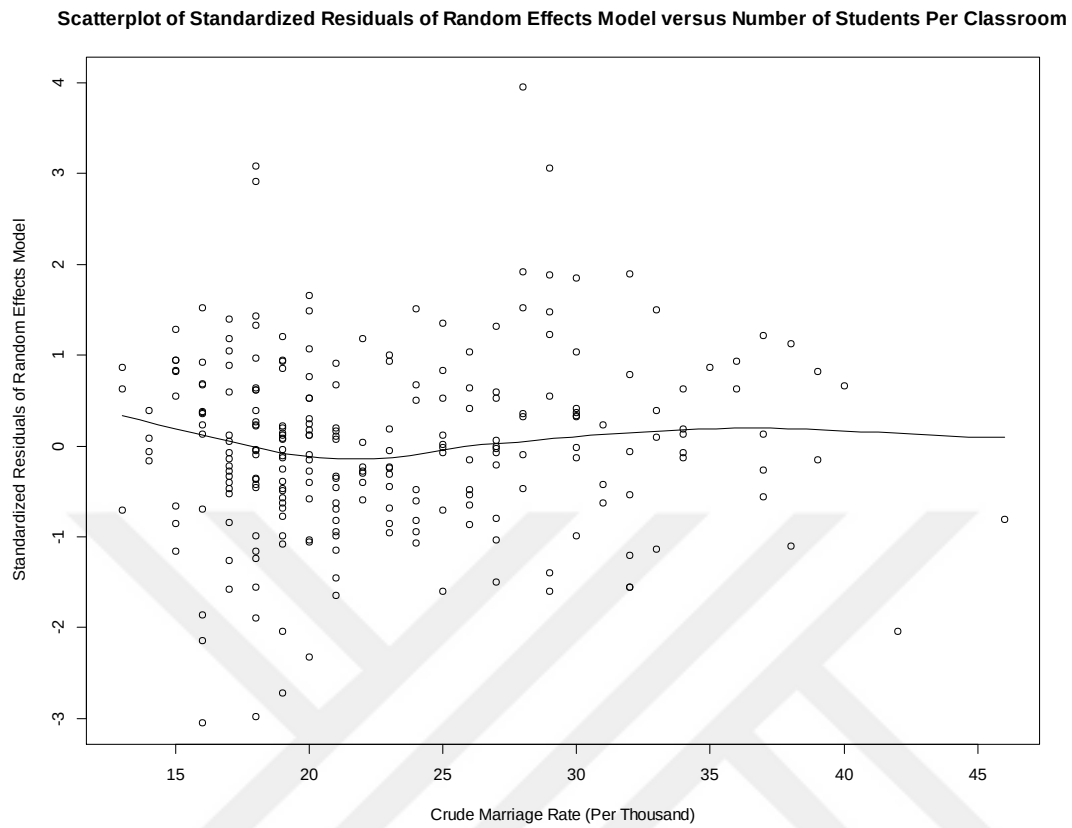


Figure C.5. Scatterplot of Standardized Residuals of Random Effects Model versus Number of Students Per Classroom

D. Results and Regression Diagnostics of Random Effects Models with Dummy Variables

Table D.1 Estimated Regression Coefficients for the Initial Random Effects Model with Dummy Variables

	<i>Estimate</i>	<i>Standard Error</i>	<i>Z-value</i>	<i>p-value</i>
Intercept	90.3	1.63	55.2	$< 2.2 \times 10^{-16}$
CBR	-0.16	0.11	-1.44	0.15
Dummy Variable of CBR	-0.33	0.46	-0.71	0.47
CBR ²	3×10^{-4}	2.5×10^{-3}	0.13	0.89
CMR	-0.26	0.08	-3.33	8.8×10^{-4}
Dummy Variable of CMR	-0.05	0.36	-0.14	0.88
CRD	-1.45	0.28	-5.01	5.4×10^{-7}
CDR ²	0.05	0.02	2.82	4.7×10^{-3}
IMR	-0.04	0.01	-3.32	9×10^{-4}
NOSPC	-0.04	8.8×10^{-3}	-4.66	3.1×10^{-6}
Interaction of CBR and its Dummy Variable	0.04	0.05	0.89	0.37
Interaction of CBR ² and its Dummy Variable	-1.2×10^{-3}	1.2×10^{-3}	-1.03	0.3
Interaction of CMR and its Dummy Variable	7.1×10^{-3}	0.05	0.15	0.88

Table D.2 VIF Values of the Variables in Random Effects Model with Dummy Variables

	<i>VIF Value</i>
CBR	51
Dummy Variable of CBR	167.6
CBR ²	47.7
CMR	1.68
Dummy Variable of CMR	78.5
CDR	42.3
CDR ²	35.9
IMR	1.24
NOSPC	1.54
Interaction of CBR and its Dummy Variable	568.6
Interaction of CBR ² and its Dummy Variable	158.8
Interaction of CMR and its Dummy Variable	77

Table D.3 Estimated Regression Coefficients for the Reduced Random Effects Model with Dummy Variables

	<i>Estimate</i>	<i>Standard Error</i>	<i>Z-value</i>	<i>p-value</i>
Intercept	80.5	0.56	142.5	$< 2.2 \times 10^{-16}$
CMR	-0.2	0.08	-2.44	0.015
IMR	-0.05	0.014	-3.91	9.35×10^{-5}
NOSPC	-8.1×10^{-3}	9.8×10^{-3}	-0.83	0.41

Table D.4 Estimated Regression Coefficients for the Final Random Effects Model with Dummy Variables

	<i>Estimate</i>	<i>Standard Error</i>	<i>Z-value</i>	<i>p-value</i>
Intercept	80.5	0.56	143.2	$< 2.2 \times 10^{-16}$
CMR	-0.23	0.08	-3.01	2.6×10^{-3}
IMR	-0.05	0.014	-4.04	5.25×10^{-5}

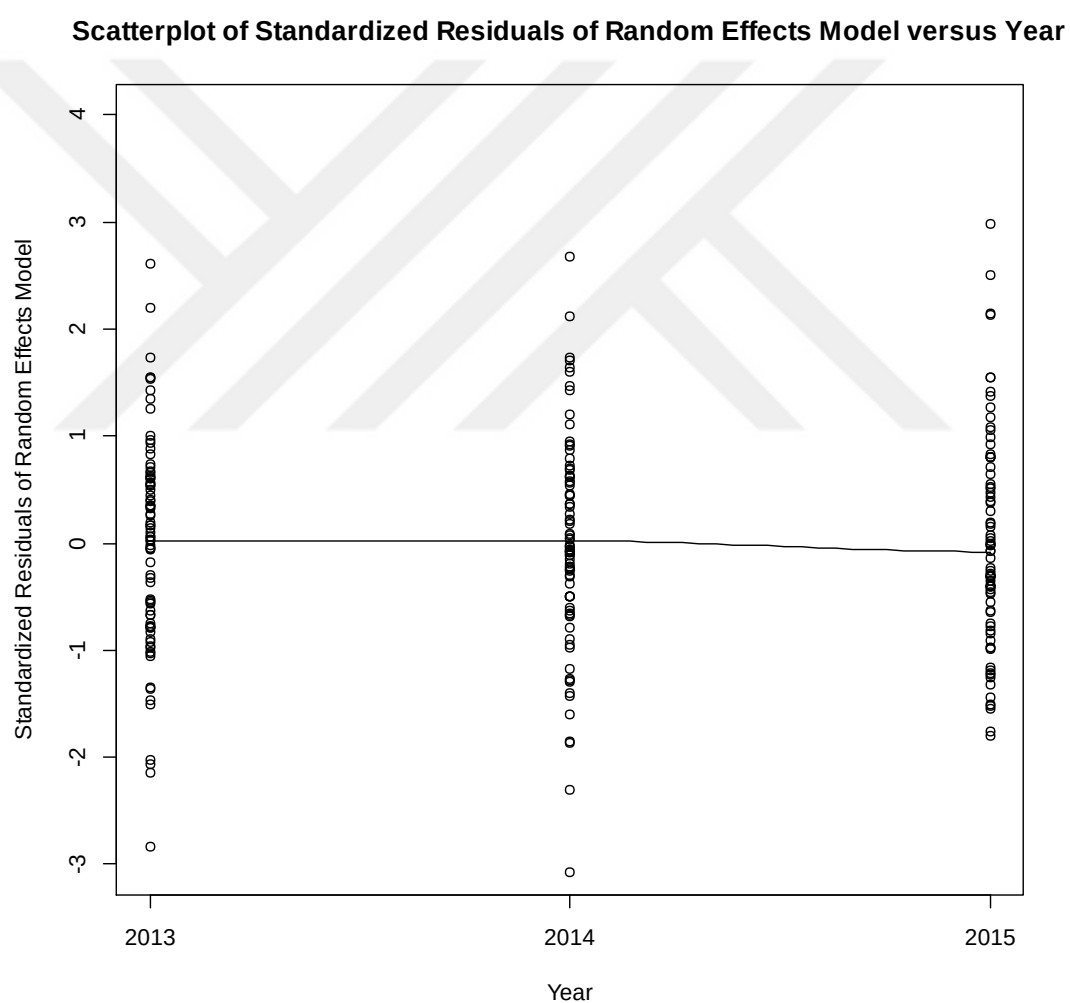


Figure D.1. Scatterplot of Standardized Residuals of Random Effects Model with Dummy Variables versus Time (Year)

Scatterplot of Standardized Residuals of Random Effects Model versus Crude Marriage Rate

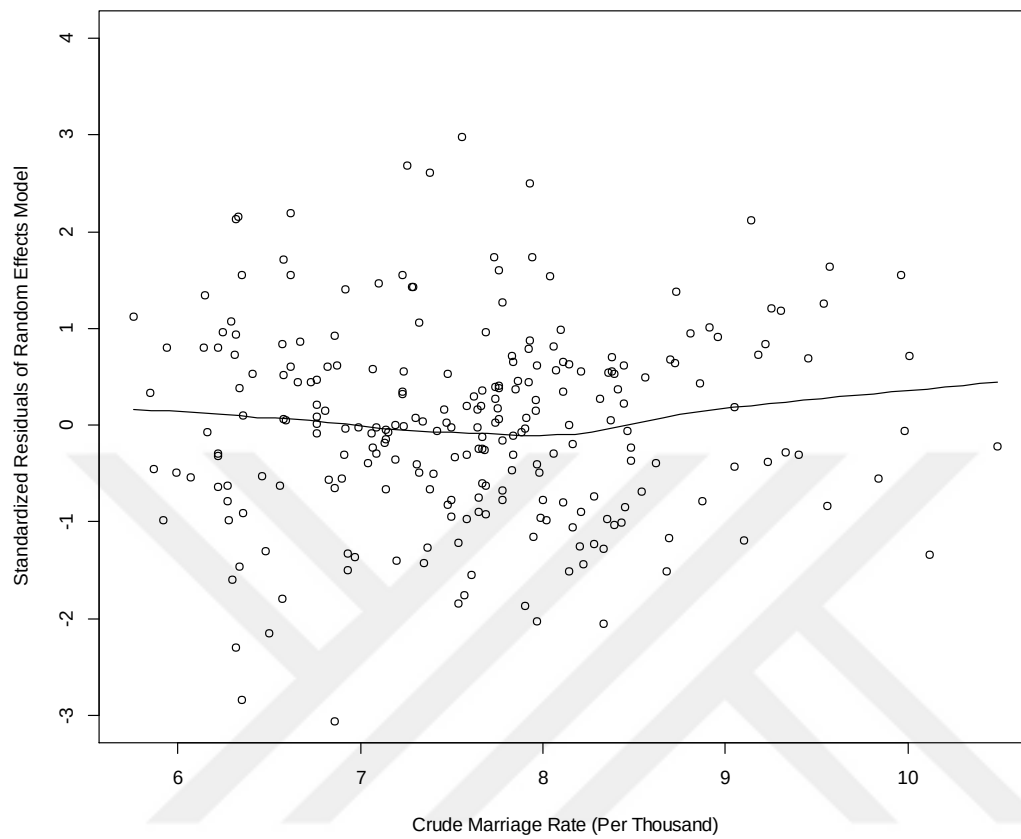


Figure D.2. Scatterplot of Standardized Residuals of Random Effects Model with Dummy Variables versus Crude Marriage Rate

Scatterplot of Standardized Residuals of Random Effects Model versus Infant Mortality Rate

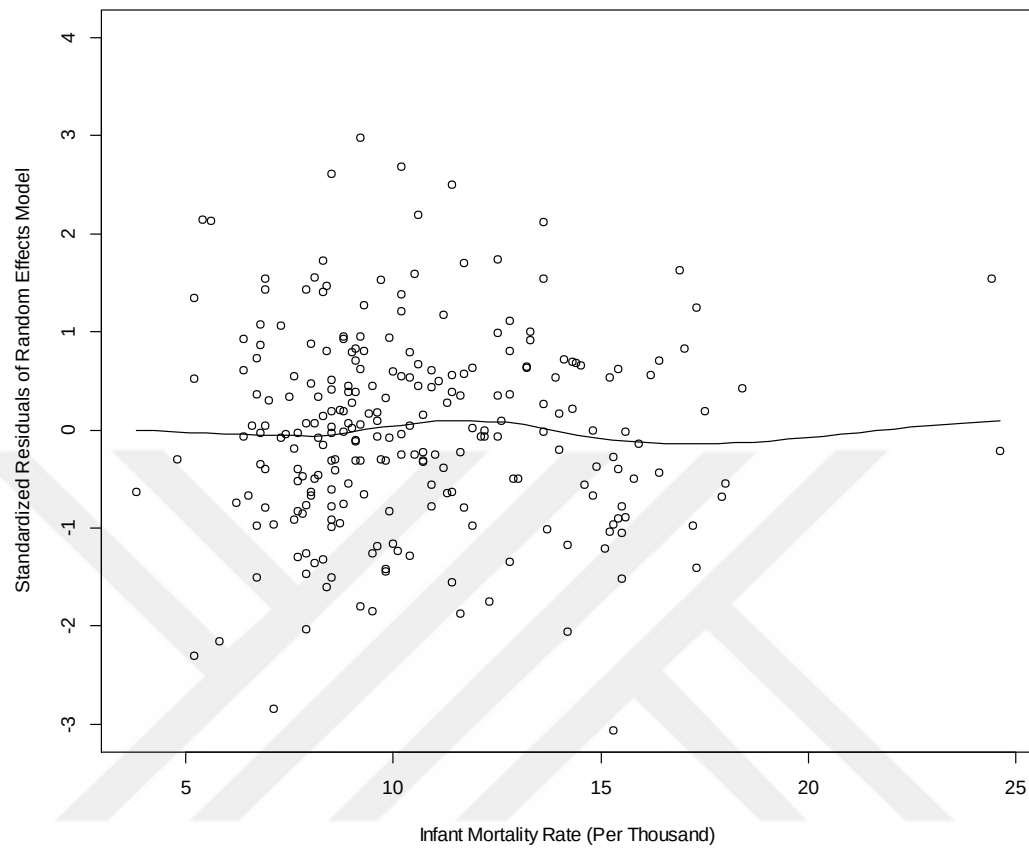


Figure D.3. Scatterplot of Standardized Residuals of Random Effects Model with Dummy Variables versus Infant Mortality Rate

Scatterplot of Fitted Response Values of Random Effects Model versus Life Expectancy at Birth

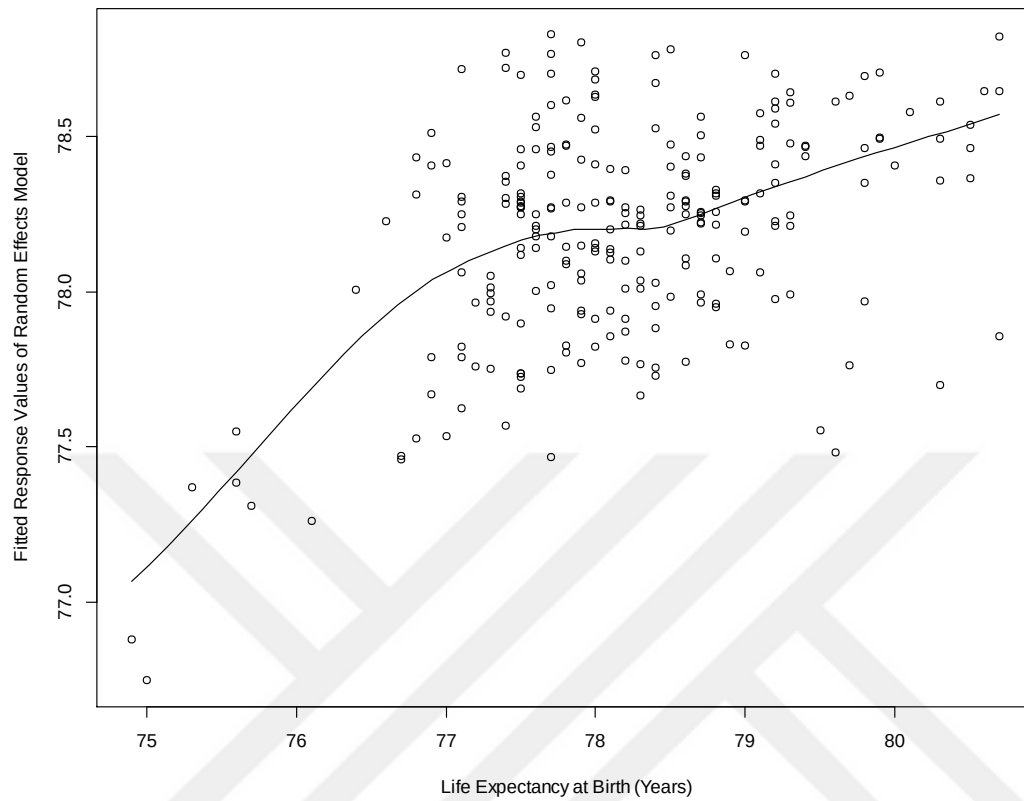


Figure D.4. Scatterplot of Fitted Response Values of Random Effects Model with Dummy Variables versus Life Expectancy at Birth

E. Results of Marginal Models that are Constructed before Final Marginal Model

Table E.1 Estimated Regression Coefficients for the Initial Marginal Model

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	-3.61	91.8	-0.39
Year	6.15	0.05	1.34
PD	-2.21×10^{-3}	1.42×10^{-3}	-1.56
NOHS	-4.31×10^{-6}	9.01×10^{-9}	-0.48
GDP	-5.40×10^{-9}	6.1×10^{-9}	-0.88
NMR	-1.80×10^{-3}	3.63×10^{-3}	-0.5
NOTH	1.38×10^{-3}	5.55×10^{-3}	0.24
NOCH	6.78×10^{-3}	4.82×10^{-3}	1.41
AF	-2.25×10^{-8}	2.62×10^{-8}	-0.86
Export (TL)	-6.49×10^{-12}	2.2×10^{-11}	-0.29
Import (TL)	1.94×10^{-11}	2.3×10^{-11}	0.84
CBR	-0.14	0.03	-4.32
CMR	-0.24	0.09	-2.5
CDR	-0.71	0.1	-7.28
IMR	-0.04	0.01	-3.61
NOSPC	-0.02	0.02	-1.03

Table E.2 Estimated Regression Coefficients for the Reduced Marginal Model

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	9.36	79.5	0.12
Year	0.04	0.04	0.98
NMR	-2.12×10^{-3}	3.65×10^{-3}	-0.58
NOTH	-1.68×10^{-3}	1.58×10^{-3}	-1.06
AF	-1.51×10^{-8}	2.14×10^{-8}	-0.71
CBR	-0.13	0.03	-4.4
CMR	-0.26	0.1	-2.74
CDR	-0.69	0.09	-7.33
IMR	-0.04	0.01	-3.54
NOSPC	-0.02	0.02	-1.39

F. Regression Diagnostic Plots of Final Marginal Model

Scatterplot of Standardized Residuals of Marginal Model versus Crude Birth Rate

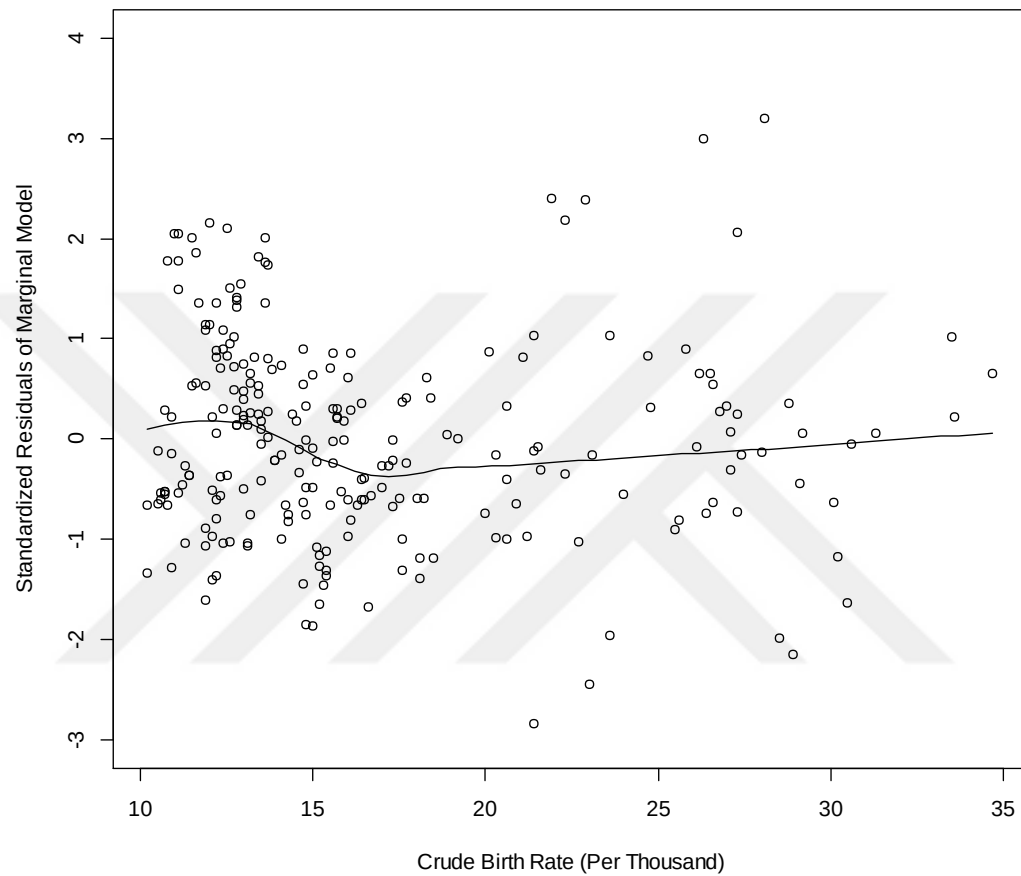


Figure F.1. Scatterplot of Standardized Residuals of Marginal Model versus Crude Birth Rate

Scatterplot of Standardized Residuals of Marginal Model versus Crude Marriage Rate

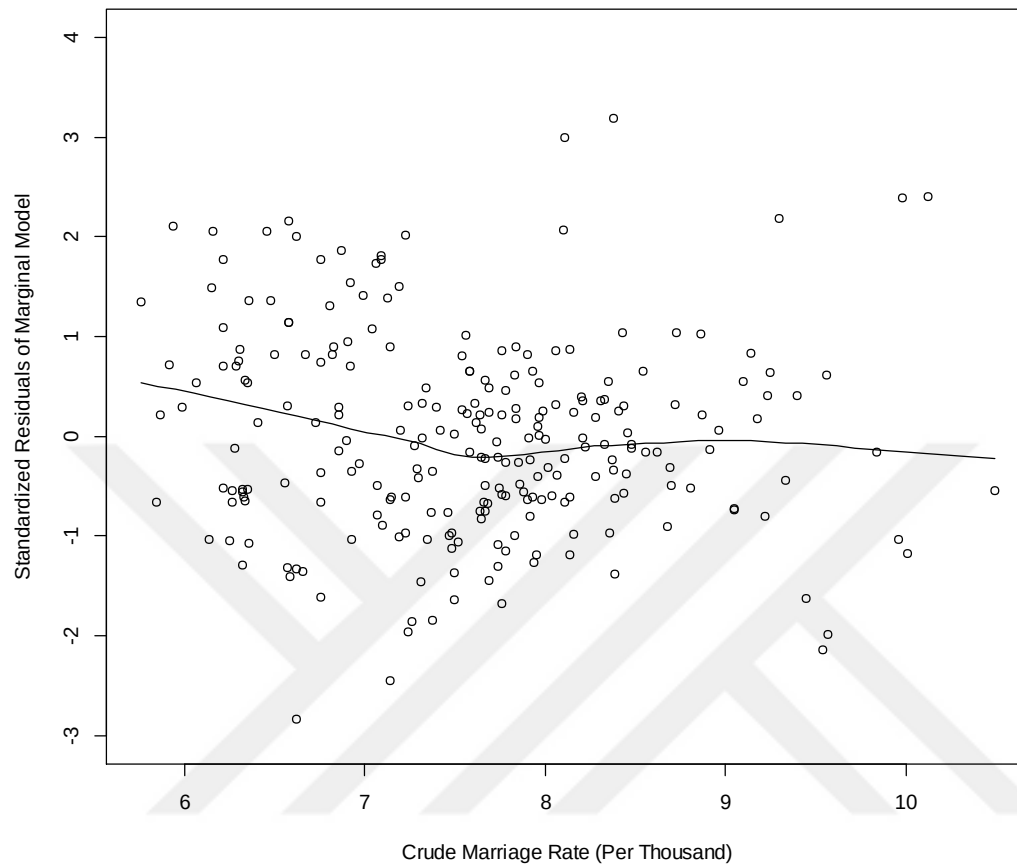


Figure F.2. Scatterplot of Standardized Residuals of Marginal Model versus Crude Marriage Rate

Scatterplot of Standardized Residuals of Marginal Model versus Crude Death Rate

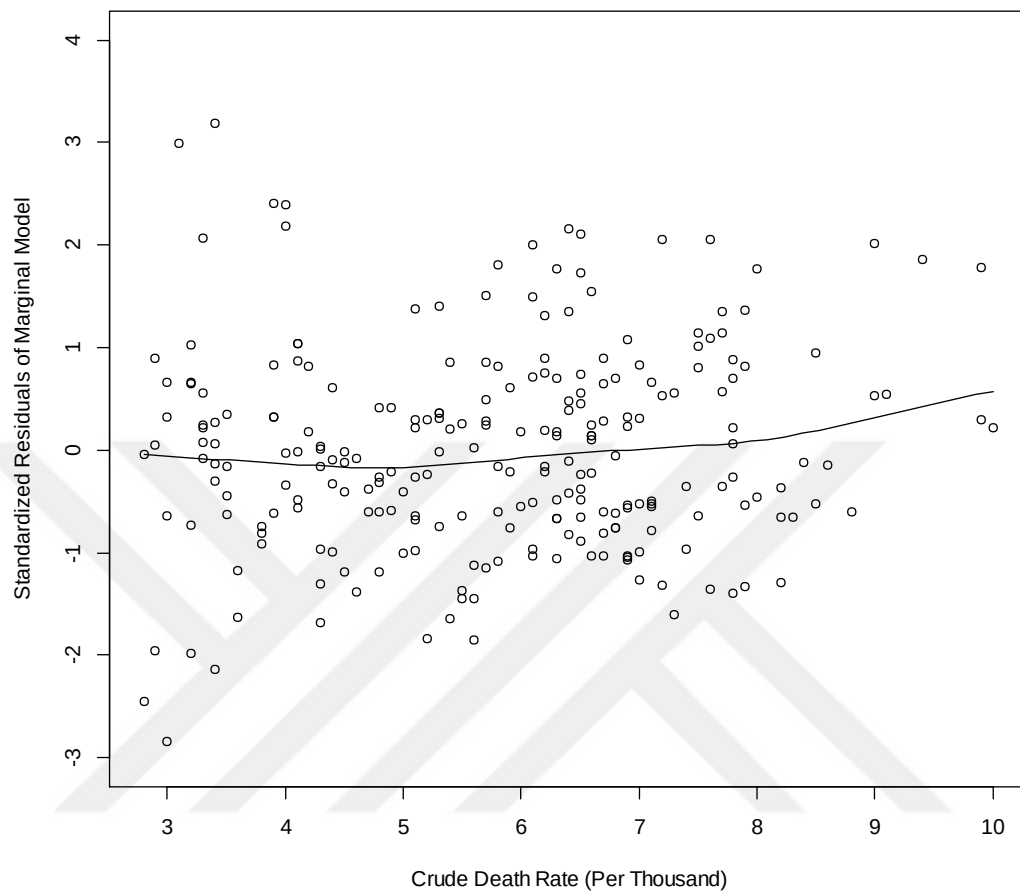


Figure F.3. Scatterplot of Standardized Residuals of Marginal Model versus Crude Death Rate

Scatterplot of Standardized Residuals of Marginal Model versus Infant Mortality Rate

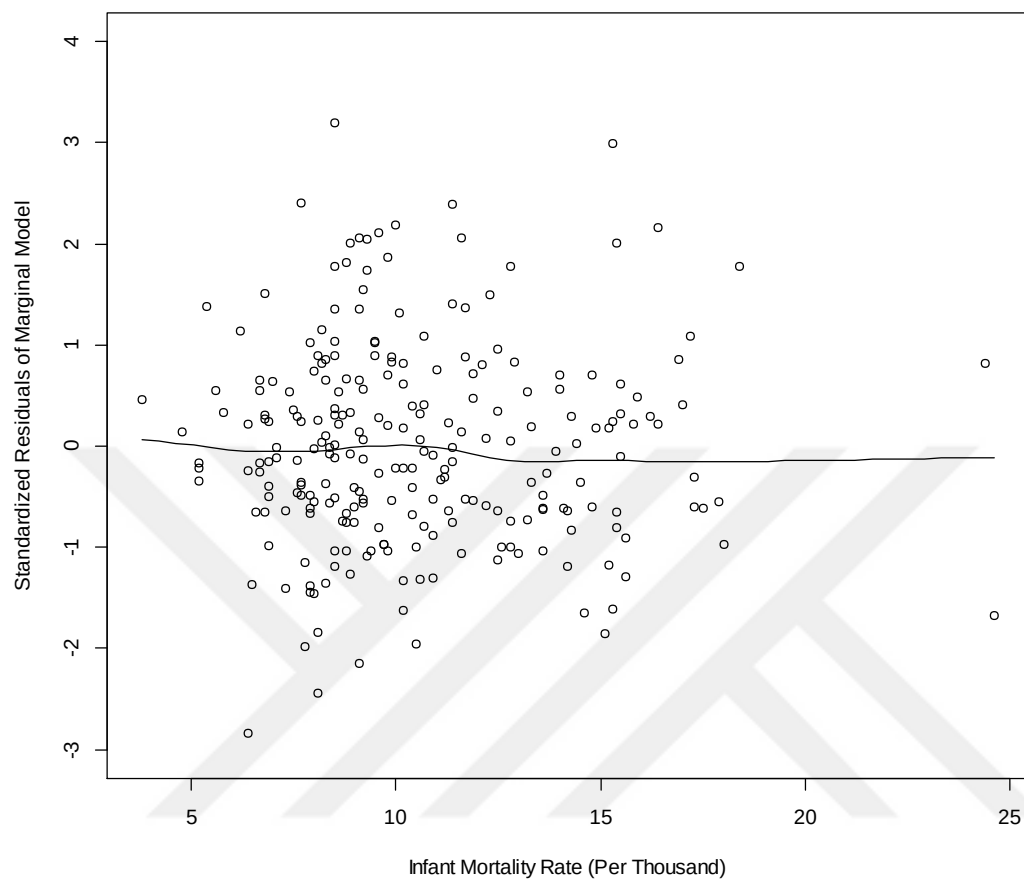


Figure F.4. Scatterplot of Standardized Residuals of Marginal Model versus Infant Mortality Rate

G. Results and Regression Diagnostics of Marginal Models with Dummy Variables

Table G.1 Estimated Regression Coefficients for the Initial Marginal Model with Dummy Variables

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	86	3.51	24.5
CBR	0.07	0.25	0.29
Dummy Variable of CBR	-8.05	4.12	-1.95
CBR ²	-4.3 x 10 ⁻³	5.3 x 10 ⁻³	-0.8
CMR	-0.3	0.13	-2.32
Dummy Variable of CMR	1.34	1.78	0.75
CDR	-1.18	0.34	-3.44
CDR ²	0.05	0.02	2.05
IMR	-0.05	0.01	-4.23
Interaction of CBR and its Dummy Variable	1.38	0.53	2.59
Interaction of CBR ² and its Dummy Variable	-0.05	0.02	-2.98
Interaction of CMR and its Dummy Variable	-0.16	0.22	-0.74

Table G.2 Estimated Regression Coefficients for the Final Marginal Model with Dummy Variables

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	80.5	0.79	101.8
CMR	-0.23	0.11	-2.15
IMR	-0.05	0.01	-3.74

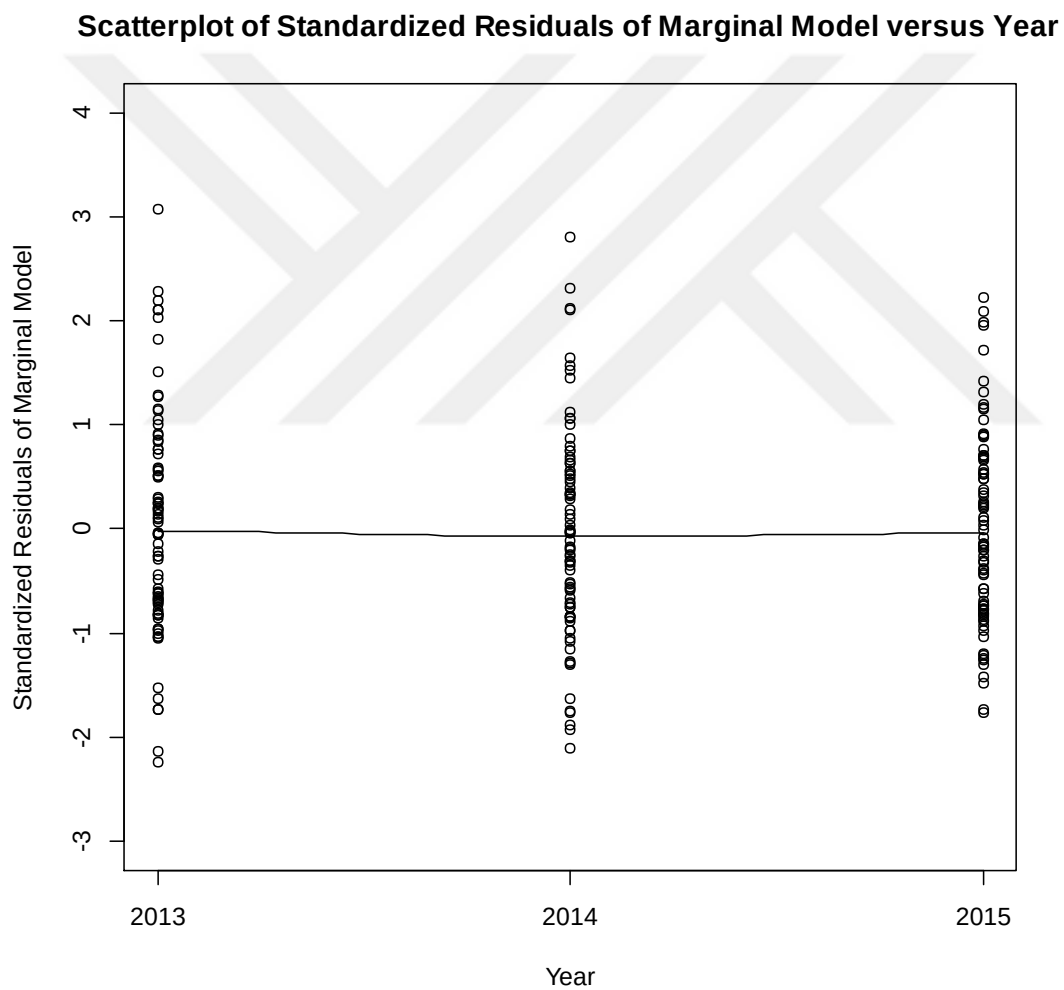


Figure G.1. Scatterplot of Standardized Residuals of Marginal Model with Dummy Variables versus Time (Year)

Scatterplot of Standardized Residuals of Marginal Model versus Crude Marriage Rate

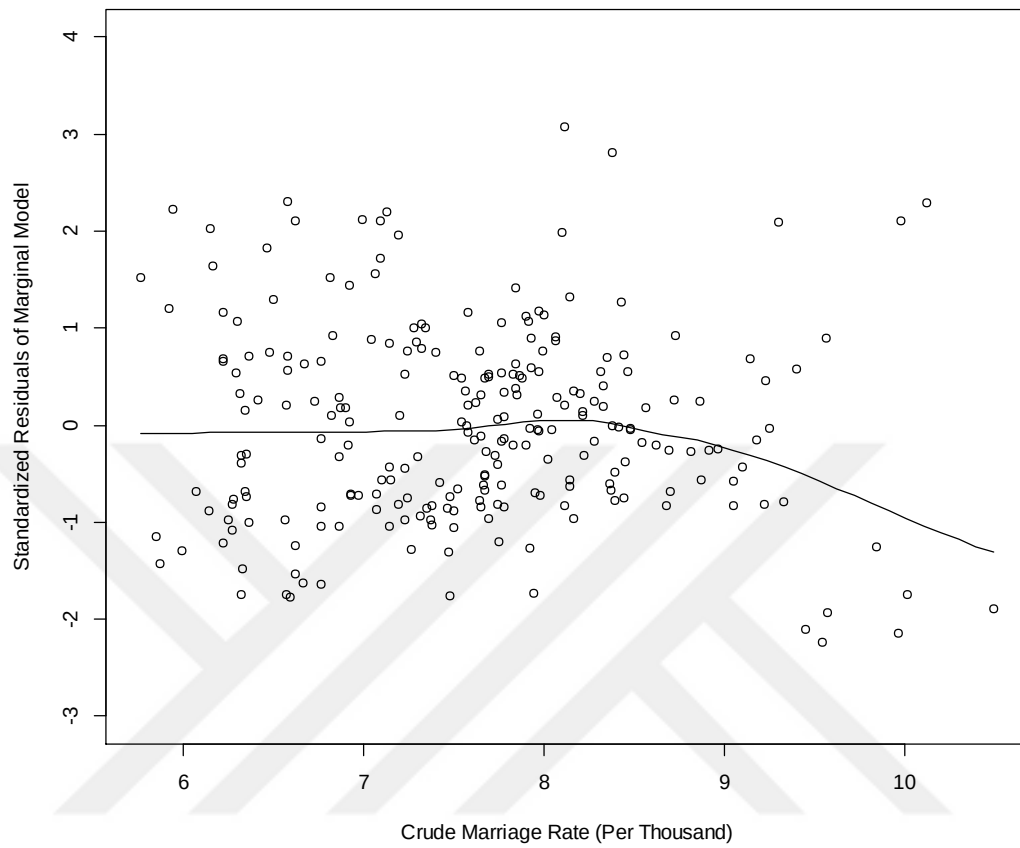


Figure G.2. Scatterplot of Standardized Residuals of Marginal Model with Dummy Variables versus Crude Marriage Rate

Scatterplot of Standardized Residuals of Marginal Model versus Infant Mortality Rate

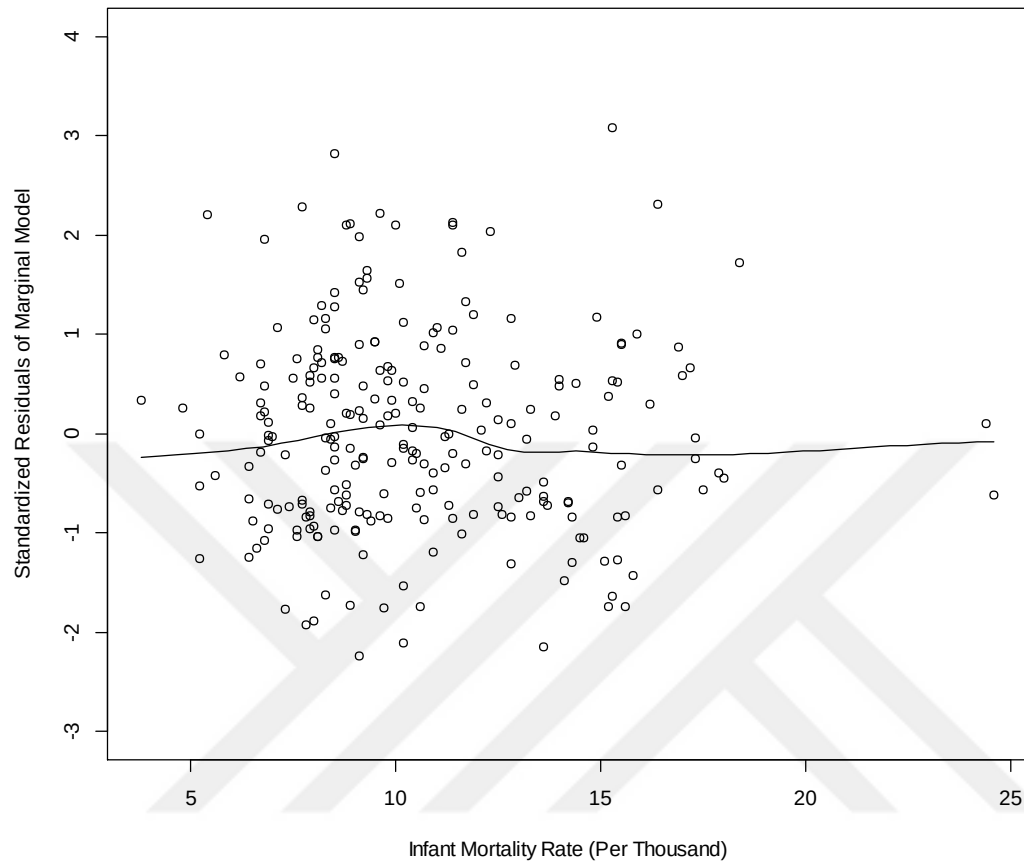


Figure G.3. Scatterplot of Standardized Residuals of Marginal Model with Dummy Variables versus Infant Mortality Rate

Scatterplot of Fitted Response Values of Marginal Model versus Life Expectancy at Birth

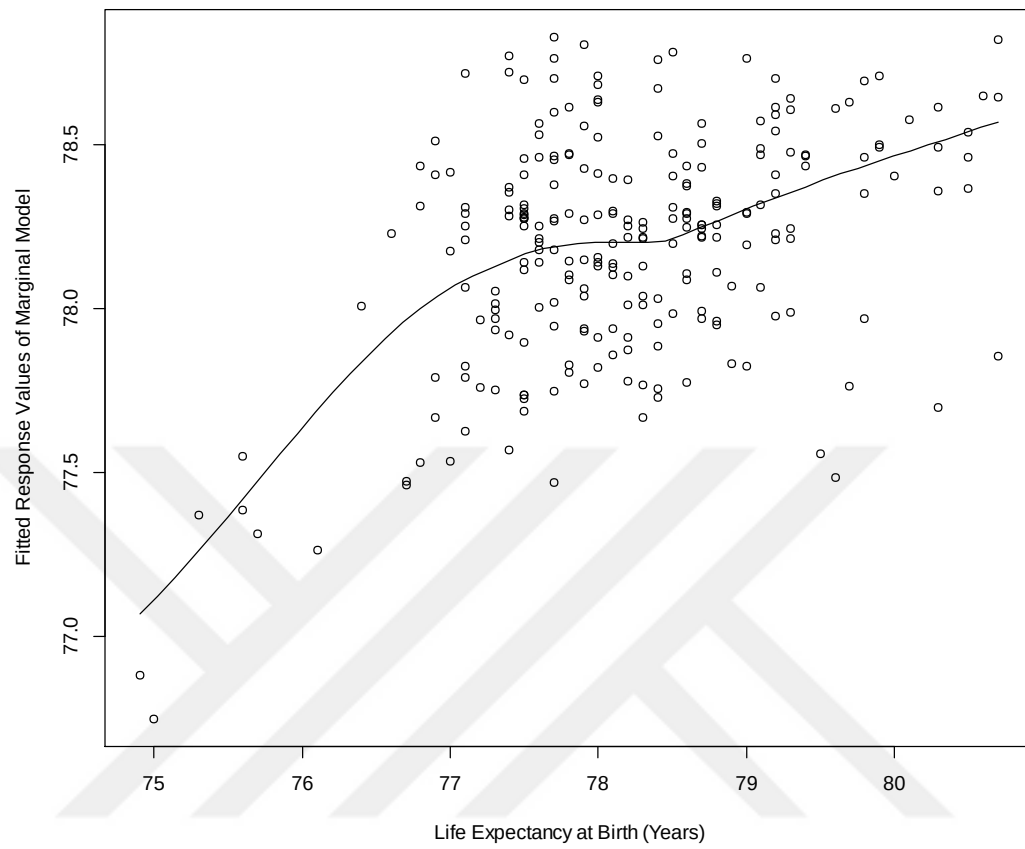


Figure G.4. Scatterplot of Fitted Response Values of Marginal Model with Dummy Variables versus Life Expectancy at Birth

H. Results of Transition Models that are Constructed before Final Transition Model

Table H.1 Estimated Regression Coefficients for the Initial Transition Model

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	-250.8	87.9	-2.85
Year	0.13	0.04	3.02
PD	-1.05×10^{-3}	5.62×10^{-4}	-1.87
NOHS	-8.01×10^{-7}	6.69×10^{-6}	-0.12
GDP	-5.85×10^{-9}	6.25×10^{-9}	-0.94
NMR	2.45×10^{-4}	3.38×10^{-3}	0.07
NOTH	5.58×10^{-3}	4.98×10^{-3}	1.11
NOCH	5.72×10^{-3}	3.72×10^{-3}	1.54
AF	-7.68×10^{-9}	8×10^{-9}	-0.96
Export (TL)	8.41×10^{-12}	1.25×10^{-11}	0.67
Import (TL)	-3.09×10^{-13}	1.03×10^{-11}	-0.03
CBR	-4.55×10^{-3}	0.01	-0.31
CMR	-0.03	0.06	-0.51
CDR	-0.12	0.05	-2.51
IMR	-0.02	0.01	-1.46
NOSPC	-8.47×10^{-3}	8.73×10^{-3}	-0.97
Lag 1 of LEAB	0.8	0.03	28.96

Table H.2 Estimated Regression Coefficients for the Reduced Transition Model

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	-238.9	86.5	-2.76
Year	0.13	0.04	2.92
NMR	-1.25×10^{-3}	2.85×10^{-3}	-0.44
NOTH	1.88×10^{-4}	6.06×10^{-4}	0.31
AF	2.74×10^{-9}	6.45×10^{-9}	0.42
CBR	-2.47×10^{-3}	0.01	-0.19
CMR	-0.05	0.05	-1.01
CDR	-0.11	0.04	-2.65
IMR	-0.01	0.01	-0.98
NOSPC	-0.01	7×10^{-3}	-1.93
Lag 1 of LEAB	0.81	0.03	30.15

I. Regression Diagnostic Plots of Final Transition Model

Scatterplot of Standardized Residuals of Transition Model versus Crude Death Rate

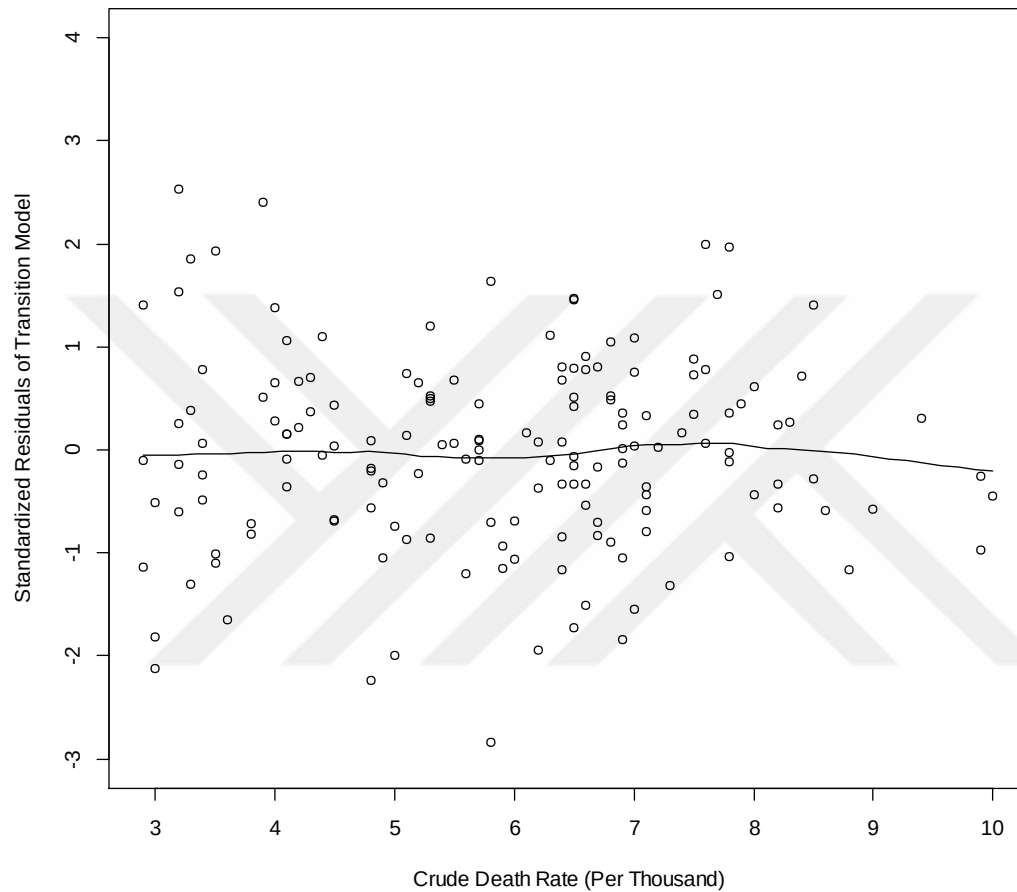


Figure I.1. Scatterplot of Standardized Residuals of Transition Model versus Crude Death Rate

Scatterplot of Standardized Residuals of Transition Model versus Life Expectancy at Birth Lag 1

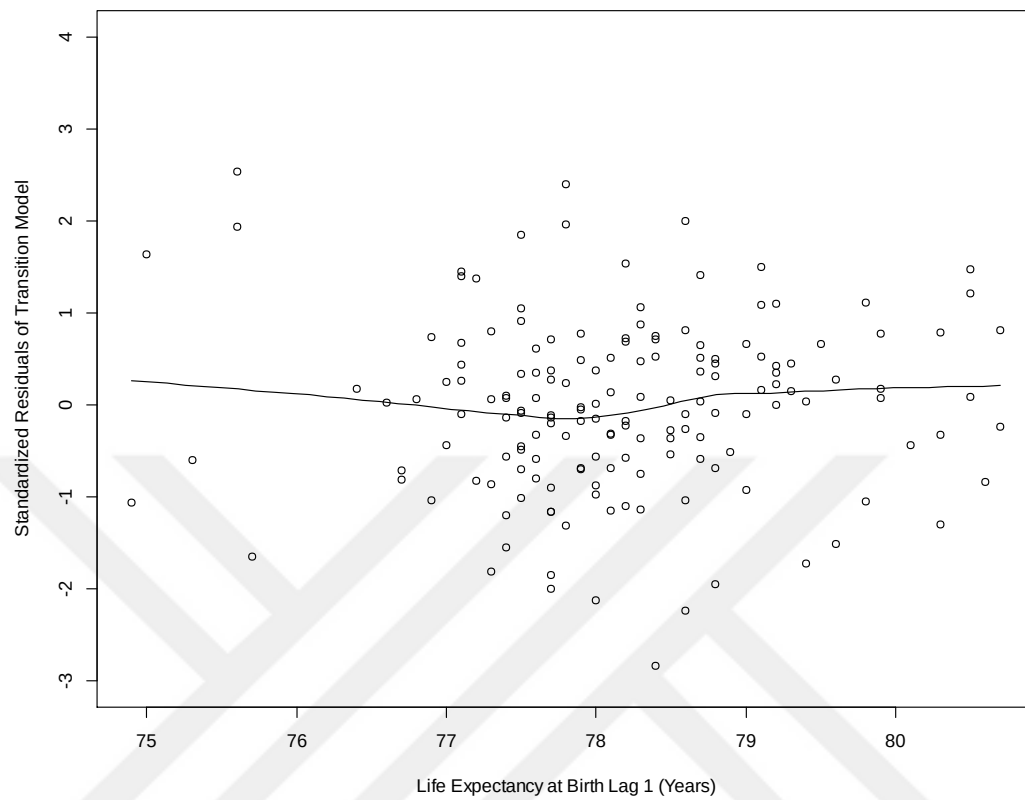


Figure I.2. Scatterplot of Standardized Residuals of Transition Model versus Lag 1 of Life Expectancy at Birth

J. Results and Regression Diagnostics of Transition Models with Dummy Variables

Table J.1 Estimated Regression Coefficients for the Initial Transition Model with Dummy Variables

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	-300.6	93.7	-3.21
Year	0.16	0.05	3.36
CBR	0.1	0.103	0.998
Dummy Variable of CBR	0.08	2.44	0.03
CBR ²	-2.44×10^{-3}	2.2×10^{-3}	-1.13
CMR	-0.13	0.104	-1.26
Dummy Variable of CMR	-0.79	0.93	-0.84
CDR	-0.14	0.197	-0.71
CDR ²	1.7×10^{-3}	0.013	0.13
Interaction of CBR and its Dummy Variable	0.09	0.31	0.29
Interaction of CBR ² and its Dummy Variable	-5.55×10^{-3}	0.01	-0.52
Interaction of CMR and its Dummy Variable	0.11	0.12	0.97
Lag 1 of LEAB	0.81	0.03	27.9

Table J.2 Estimated Regression Coefficients for the Reduced Transition Model with Dummy Variables

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	-321	98.2	-3.27
Year	0.16	0.05	3.37
CMR	0.02	0.03	0.73
Lag 1 of LEAB	0.85	0.02	36.6

Table J.3 Estimated Regression Coefficients for the Final Transition Model with Dummy Variables

	<i>Estimate</i>	<i>Robust Standard Error</i>	<i>Robust Z-value</i>
Intercept	-312.6	98.4	-3.18
Year	0.16	0.05	3.28
Lag 1 of LEAB	0.85	0.02	40.3

Scatterplot of Standardized Residuals of Transition Model versus Year

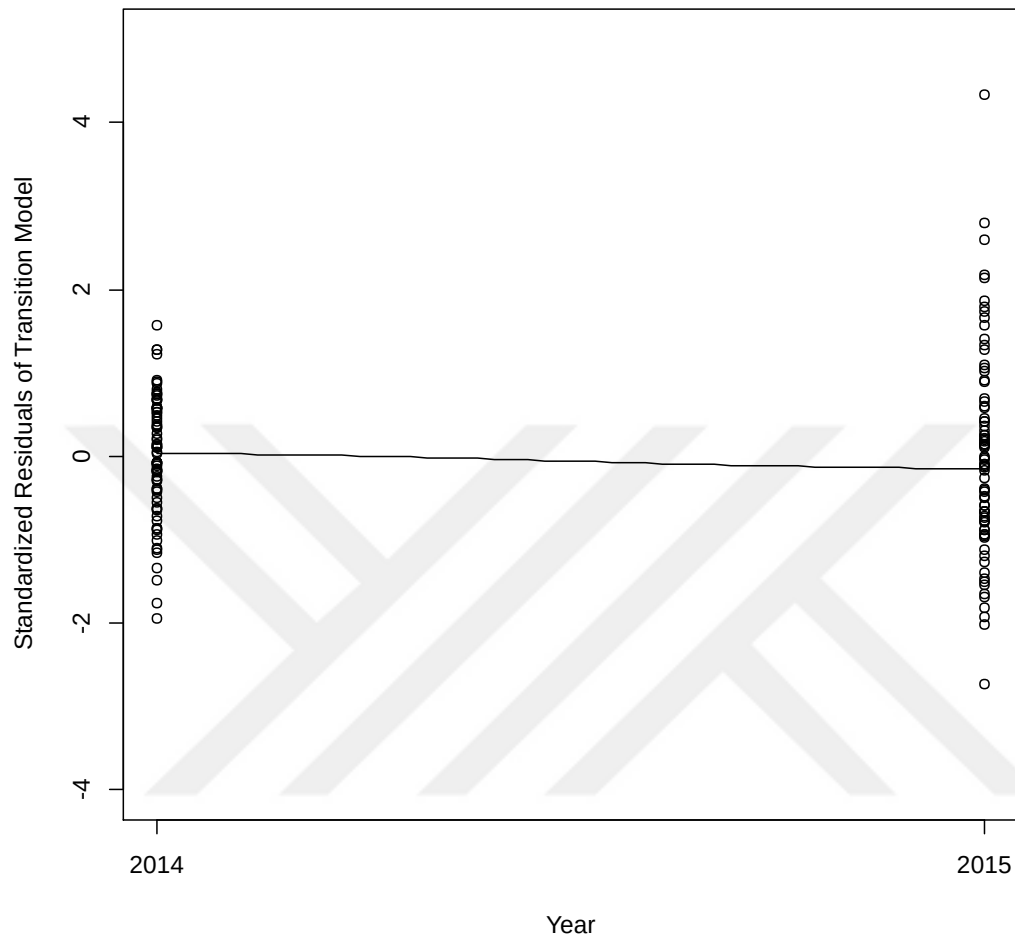


Figure J.1. Scatterplot of Standardized Residuals of Transition Model with Dummy Variables versus Time (Year)

Scatterplot of Standardized Residuals of Transition Model versus Life Expectancy at Birth Lag 1

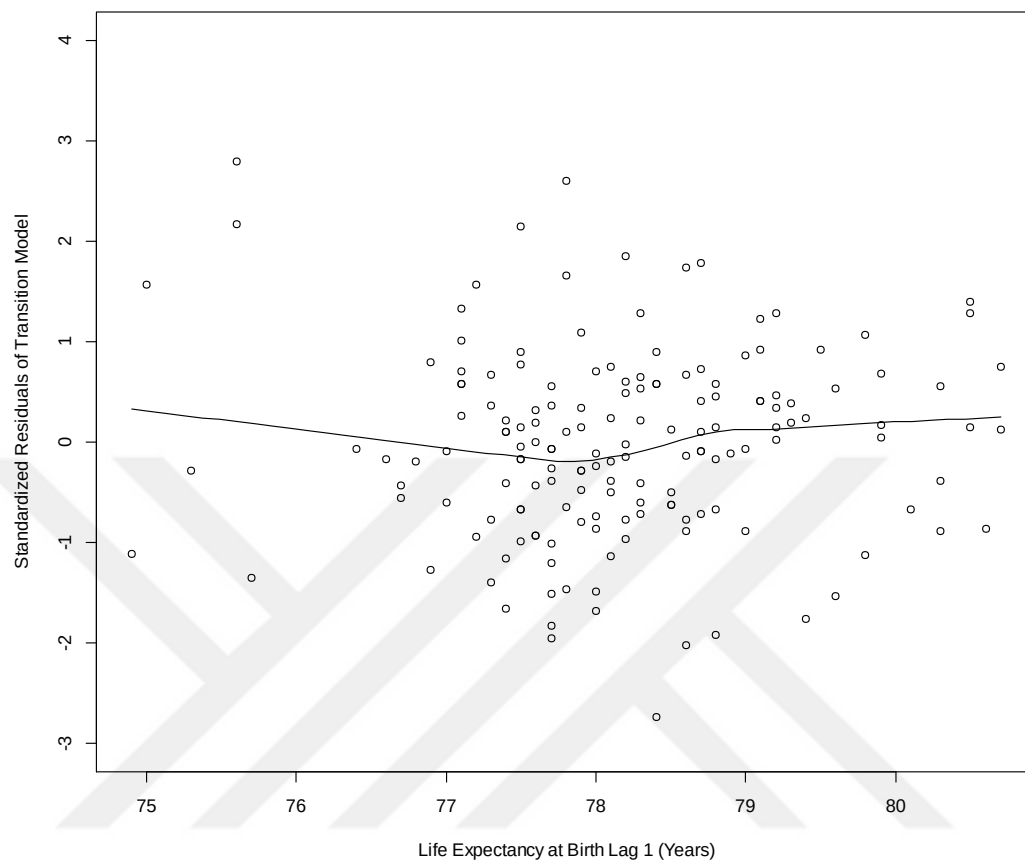


Figure J.2. Scatterplot of Standardized Residuals of Transition Model with Dummy Variables versus Lag 1 of Life Expectancy at Birth

Scatterplot of Fitted Response Values of Transition Model versus Life Expectancy at Birth

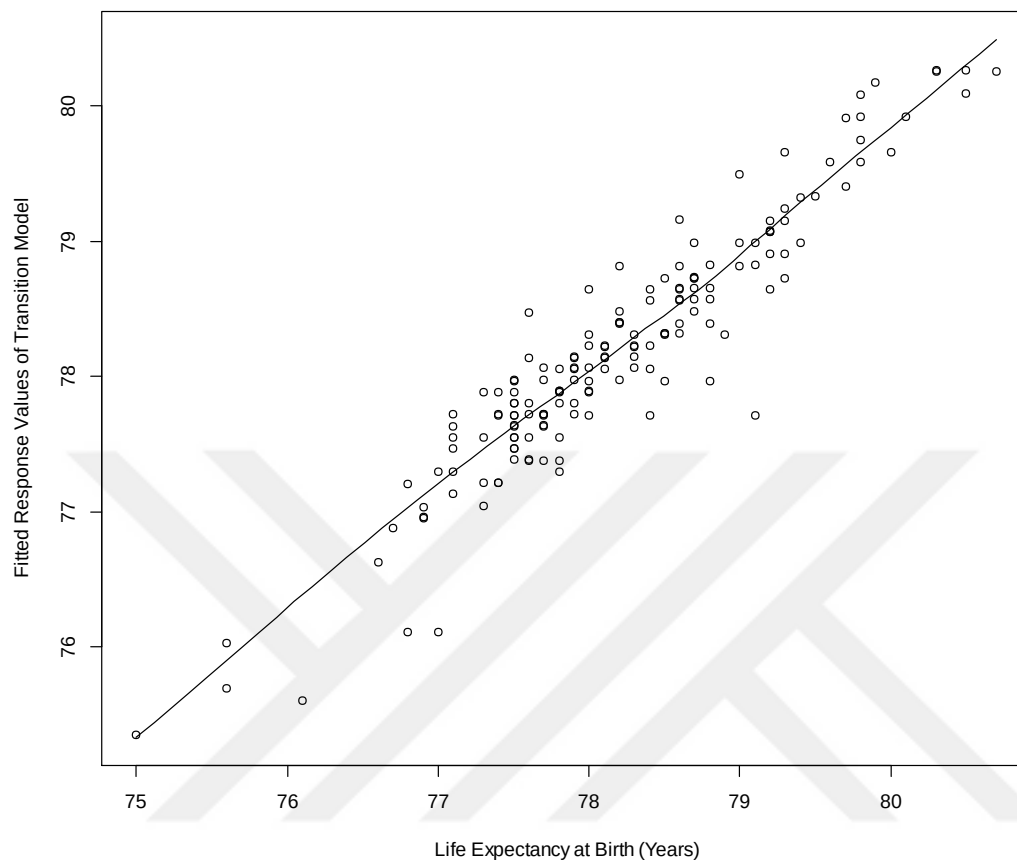


Figure J.3. Scatterplot of Fitted Response Values of Transition Model with Dummy Variables versus Life Expectancy at Birth

K. Scatterplots of Response Variable and Significant Covariates

Scatterplot of Life Expectancy at Birth versus Crude Birth Rate

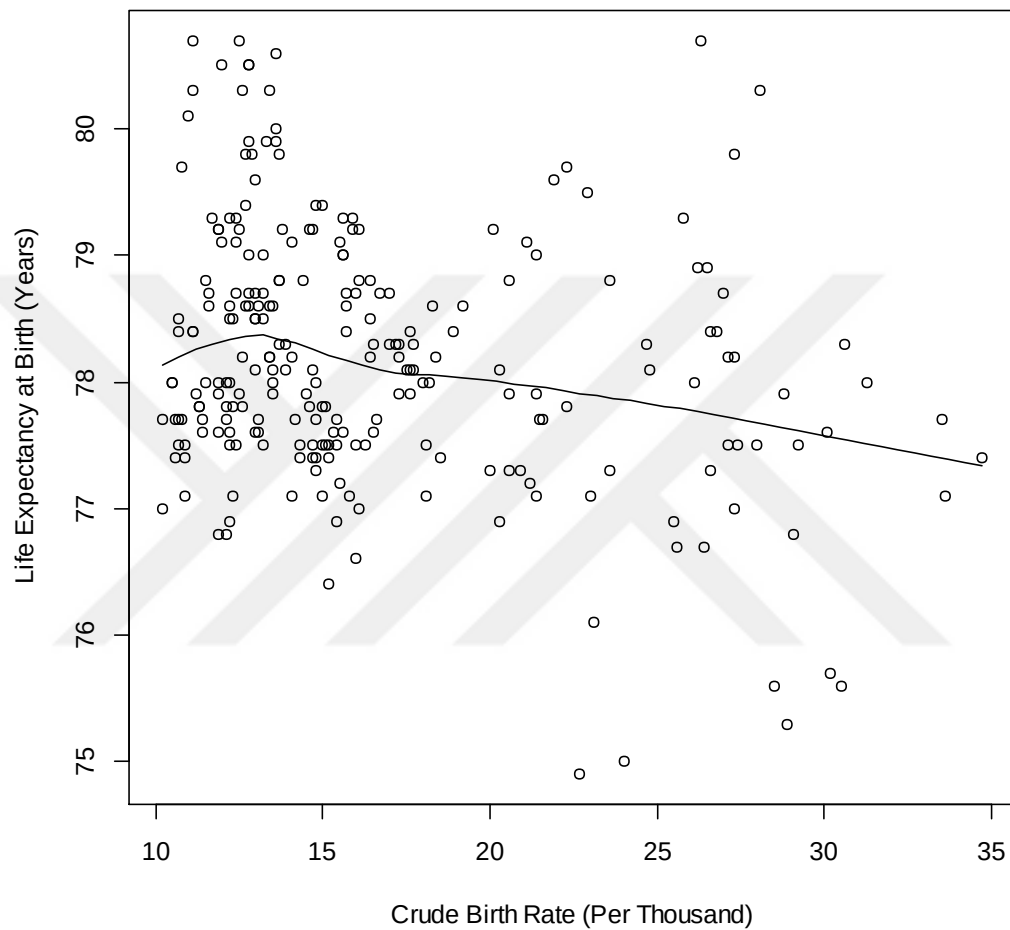


Figure K.1. Scatterplot of Life Expectancy at Birth versus Crude Birth Rate

Scatterplot of Life Expectancy at Birth versus Crude Marriage Rate

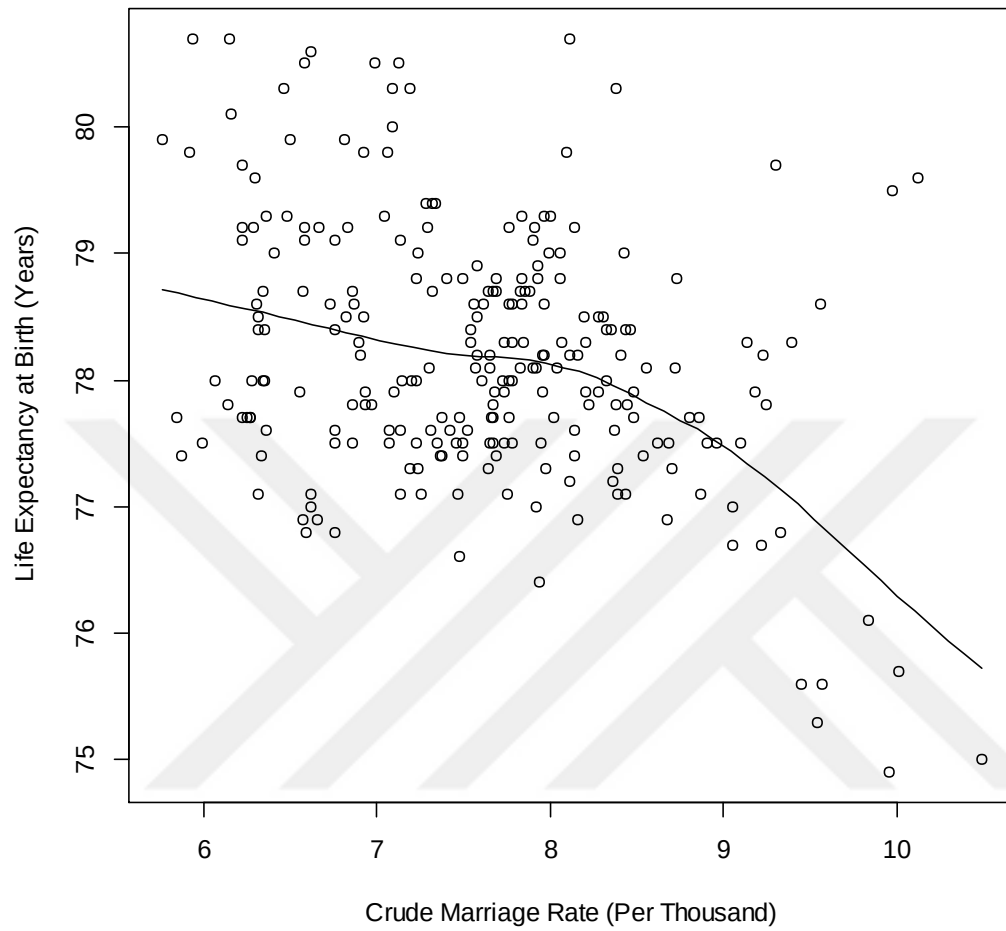


Figure K.2. Scatterplot of Life Expectancy at Birth versus Crude Marriage Rate

Scatterplot of Life Expectancy at Birth versus Crude Death Rate

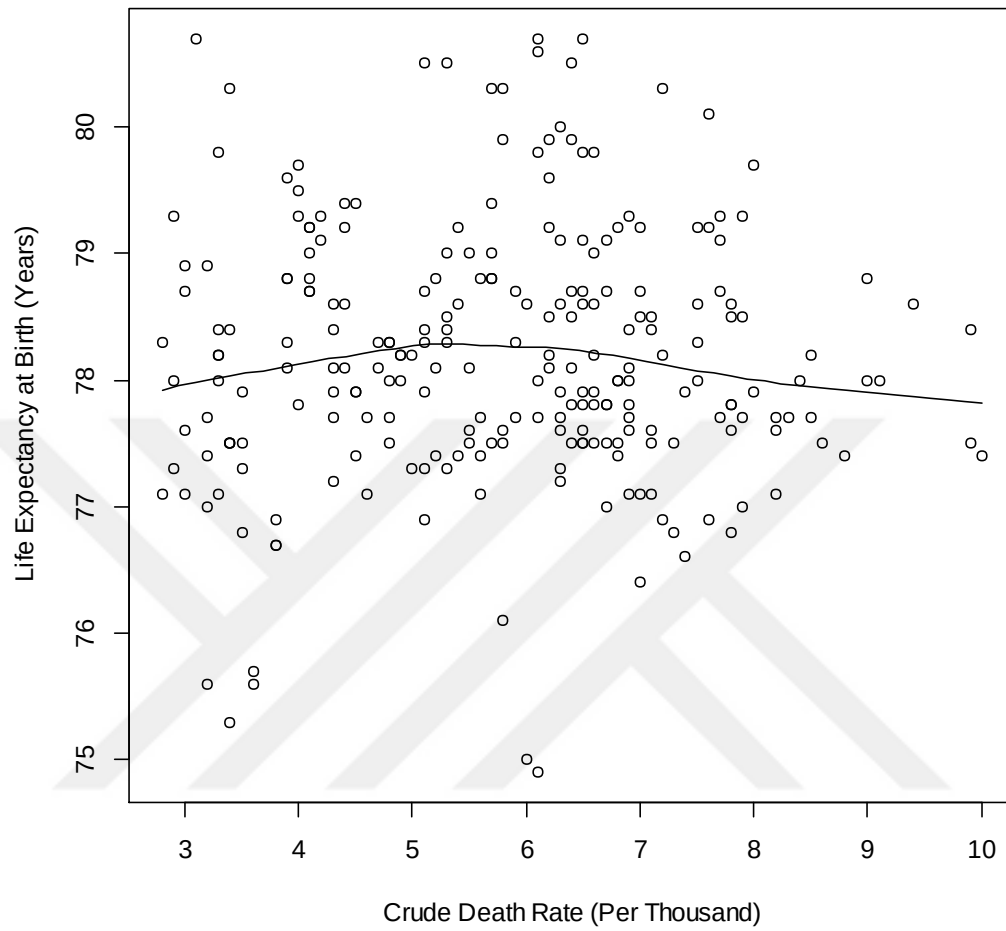


Figure K.3. Scatterplot of Life Expectancy at Birth versus Crude Death Rate

Scatterplot of Life Expectancy at Birth versus Infant Mortality Rate

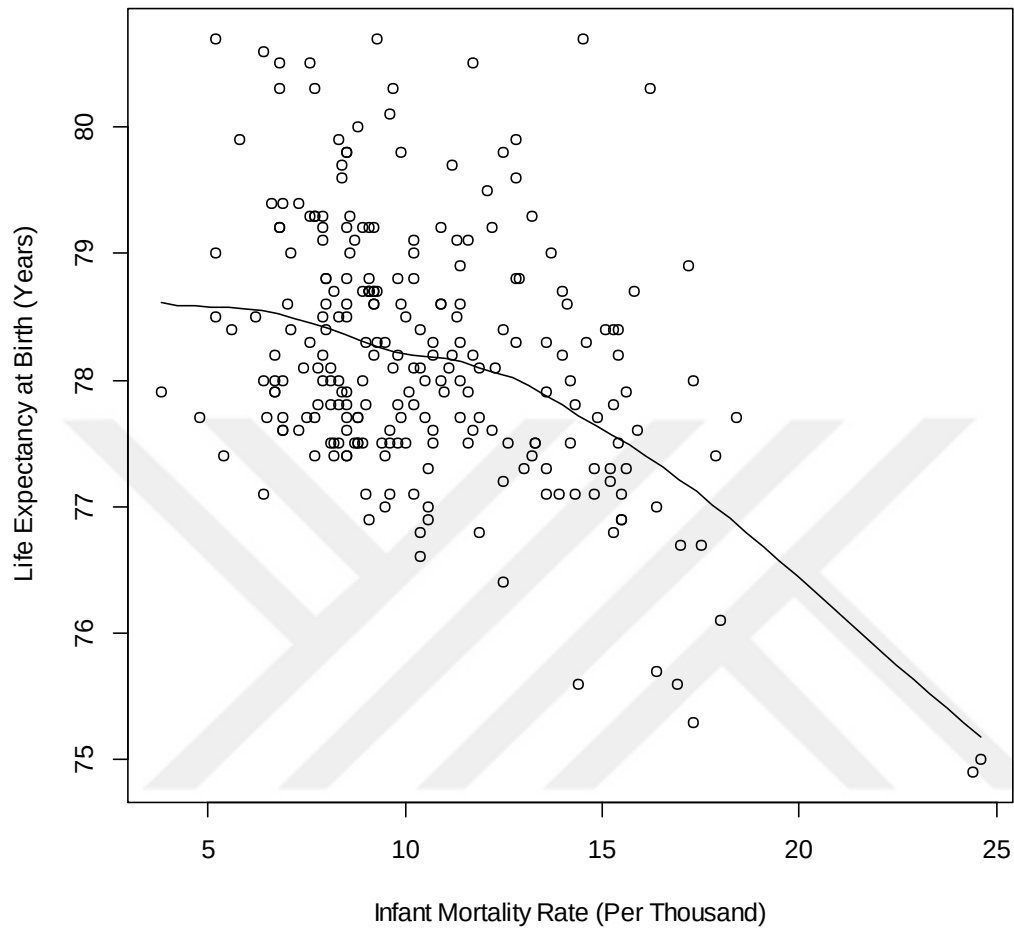


Figure K.4. Scatterplot of Life Expectancy at Birth versus Infant Mortality Rate

Scatterplot of Life Expectancy at Birth versus Number of Students Per Classroom

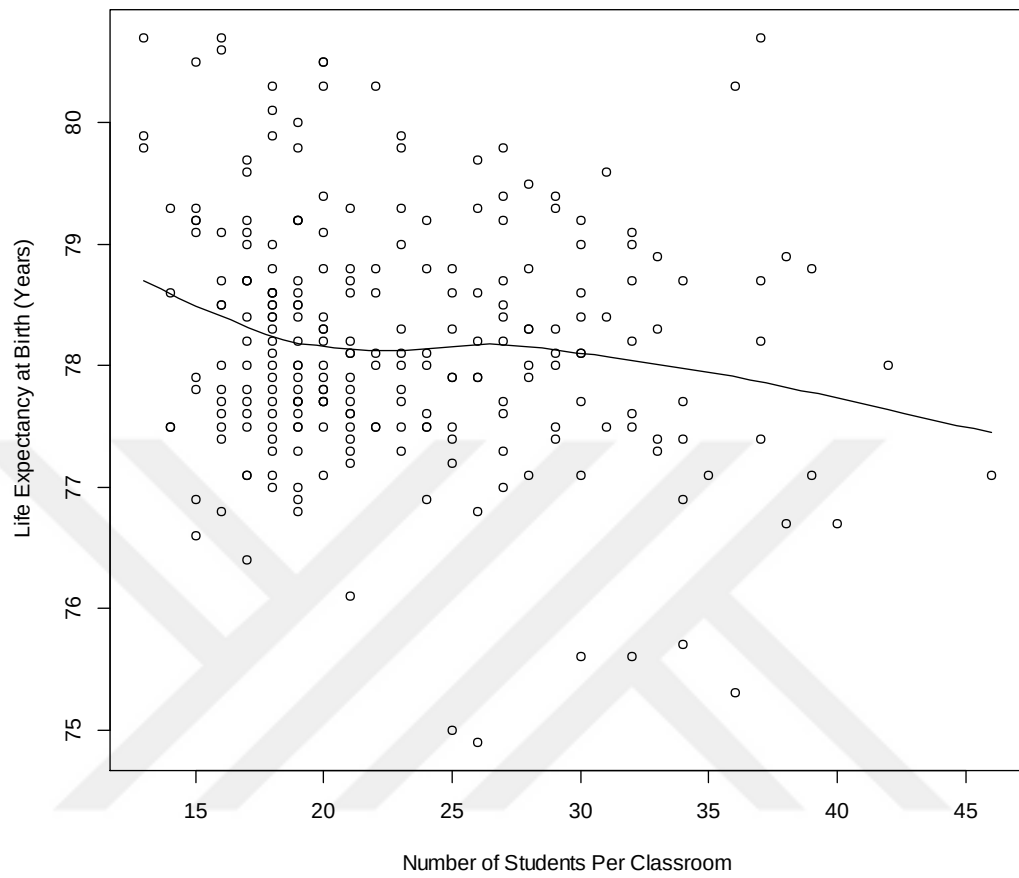


Figure K.5. Scatterplot of Life Expectancy at Birth versus Number of Students Per Classroom

Scatterplot of Life Expectancy at Birth versus Life Expectancy at Birth Lag 1

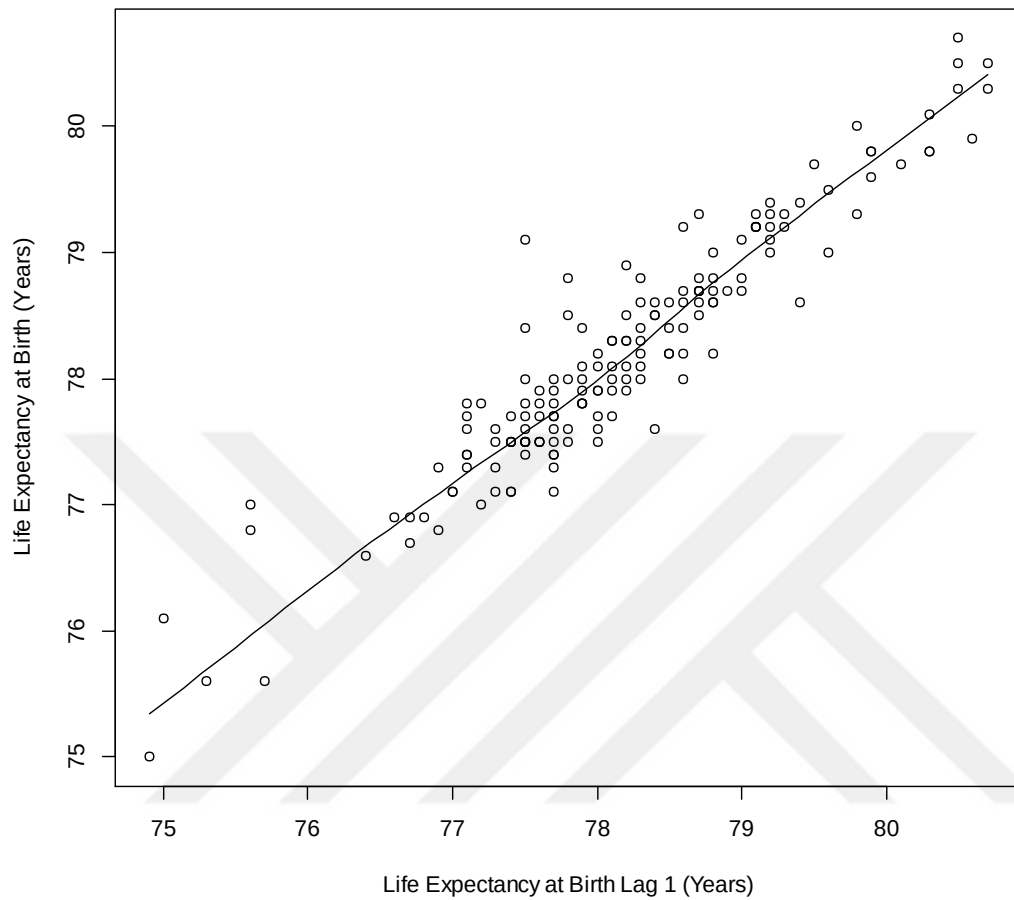


Figure K.6. Scatterplot of Life Expectancy at Birth versus Lag 1 of Life Expectancy at Birth

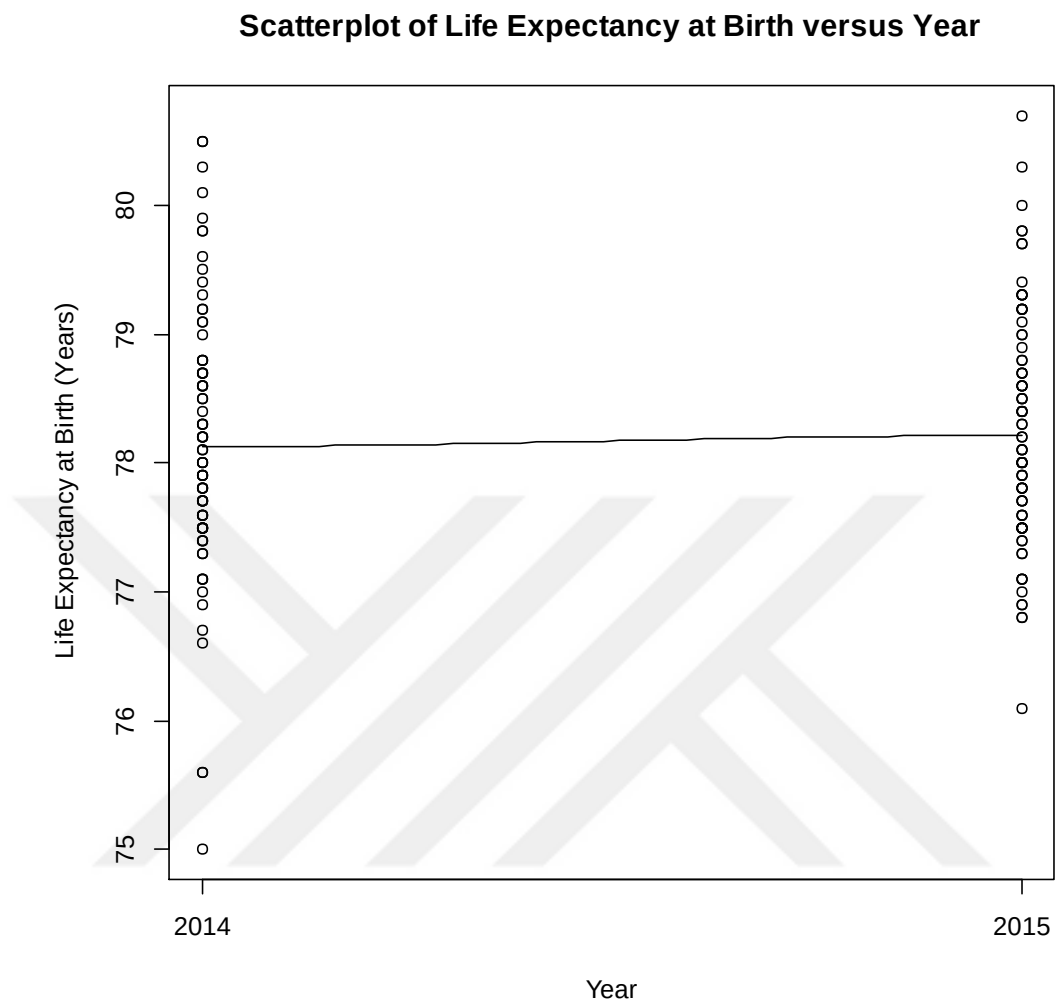


Figure K.7. Scatterplot of Life Expectancy at Birth versus Year

L. IDs of the 81 Variance Combinations in Measurement Error Part

Table L.1 IDs of the 81 Variance Combinations

<i>ID</i>	<i>Variance Combination</i>
1	0, 0, 0, 0
2	0, 0, 0, 0.5
3	0, 0, 0, 1
4	0, 0, 0.5, 0
5	0, 0, 0.5, 0.5
6	0, 0, 0.5, 1
7	0, 0, 1, 0
8	0, 0, 1, 0.5
9	0, 0, 1, 1
10	0, 0.5, 0, 0
11	0, 0.5, 0, 0.5
12	0, 0.5, 0, 1
13	0, 0.5, 0.5, 0
14	0, 0.5, 0.5, 0.5
15	0, 0.5, 0.5, 1
16	0, 0.5, 1, 0
17	0, 0.5, 1, 0.5
18	0, 0.5, 1, 1
19	0, 1, 0, 0
20	0, 1, 0, 0.5
21	0, 1, 0, 1
22	0, 1, 0.5, 0
23	0, 1, 0.5, 0.5
24	0, 1, 0.5, 1
25	0, 1, 1, 0
26	0, 1, 1, 0.5
27	0, 1, 1, 1

Table L.1 (continued)

<i>ID</i>	<i>Variance Combination</i>
28	0.5, 0, 0, 0
29	0.5, 0, 0, 0.5
30	0.5, 0, 0, 1
31	0.5, 0, 0.5, 0
32	0.5, 0, 0.5, 0.5
33	0.5, 0, 0.5, 1
34	0.5, 0, 1, 0
35	0.5, 0, 1, 0.5
36	0.5, 0, 1, 1
37	0.5, 0.5, 0, 0
38	0.5, 0.5, 0, 0.5
39	0.5, 0.5, 0, 1
40	0.5, 0.5, 0.5, 0
41	0.5, 0.5, 0.5, 0.5
42	0.5, 0.5, 0.5, 1
43	0.5, 0.5, 1, 0
44	0.5, 0.5, 1, 0.5
45	0.5, 0.5, 1, 1
46	0.5, 1, 0, 0
47	0.5, 1, 0, 0.5
48	0.5, 1, 0, 1
49	0.5, 1, 0.5, 0
50	0.5, 1, 0.5, 0.5
51	0.5, 1, 0.5, 1
52	0.5, 1, 1, 0
53	0.5, 1, 1, 0.5
54	0.5, 1, 1, 1
55	1, 0, 0, 0
56	1, 0, 0, 0.5
57	1, 0, 0, 1

Table L.1 (continued)

<i>ID</i>	<i>Variance Combination</i>
58	1, 0, 0.5, 0
59	1, 0, 0.5, 0.5
60	1, 0, 0.5, 1
61	1, 0, 1, 0
62	1, 0, 1, 0.5
63	1, 0, 1, 1
64	1, 0.5, 0, 0
65	1, 0.5, 0, 0.5
66	1, 0.5, 0, 1
67	1, 0.5, 0.5, 0
68	1, 0.5, 0.5, 0.5
69	1, 0.5, 0.5, 1
70	1, 0.5, 1, 0
71	1, 0.5, 1, 0.5
72	1, 0.5, 1, 1
73	1, 1, 0, 0
74	1, 1, 0, 0.5
75	1, 1, 0, 1
76	1, 1, 0.5, 0
77	1, 1, 0.5, 0.5
78	1, 1, 0.5, 1
79	1, 1, 1, 0
80	1, 1, 1, 0.5
81	1, 1, 1, 1

M. Variance Combinations where Infant Mortality Rate is Found to be Insignificant in Measurement Error Part

Table M.1 Variance Combinations where Infant Mortality Rate is Insignificant

<i>Variance Combinations</i>
0, 0, 1, 0.5
0, 0.5, 0, 0
0, 0.5, 0, 0.5
0, 0.5, 0, 1
0, 0.5, 0.5, 0
0, 0.5, 0.5, 0.5
0, 0.5, 0.5, 1
0, 0.5, 1, 0
0, 0.5, 1, 0.5
0, 0.5, 1, 1
0, 1, 0, 0
0, 1, 0, 0.5
0, 1, 0, 1
0, 1, 0.5, 0
0, 1, 0.5, 0.5
0, 1, 0.5, 1
0, 1, 1, 0
0, 1, 1, 0.5
0, 1, 1, 1
0.5, 0, 0.5, 0
0.5, 0, 1, 0
0.5, 0.5, 0, 0
0.5, 0.5, 0, 0.5
0.5, 0.5, 0, 1
0.5, 0.5, 0.5, 0
0.5, 0.5, 0.5, 0.5
0.5, 0.5, 0.5, 1

Table M.1 (continued)

<i>Variance Combinations</i>
0.5, 0.5, 1, 0
0.5, 0.5, 1, 0.5
0.5, 0.5, 1, 1
0.5, 1, 0, 0
0.5, 1, 0, 0.5
0.5, 1, 0, 1
0.5, 1, 0.5, 0
0.5, 1, 0.5, 0.5
0.5, 1, 0.5, 1
0.5, 1, 1, 0
0.5, 1, 1, 0.5
0.5, 1, 1, 1
1, 0, 0, 0
1, 0, 0.5, 0
1, 0, 0.5, 0.5
1, 0, 1, 0
1, 0, 1, 1
1, 0.5, 0, 0
1, 0.5, 0, 0.5
1, 0.5, 0, 1
1, 0.5, 0.5, 0
1, 0.5, 0.5, 0.5
1, 0.5, 0.5, 1
1, 0.5, 1, 0
1, 0.5, 1, 0.5
1, 0.5, 1, 1
1, 1, 0, 0
1, 1, 0, 0.5
1, 1, 0, 1
1, 1, 0.5, 0

Table M.1 (continued)

<i>Variance Combinations</i>
1, 1, 0.5, 0.5
1, 1, 0.5, 1
1, 1, 1, 0
1, 1, 1, 0.5
1, 1, 1, 1

