STATISTICAL ANALYSIS OF ELECTRICITY ENERGY CONSUMPTION WITH RESPECT TO PROVINCES IN TURKEY

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

BY

MEHMET ÇAKMAK

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

DECEMBER 2014

Approval of the thesis:

STATISTICAL ANALYSIS OF ELECTRICITY ENERGY CONSUMPTION WITH RESPECT TO PROVINCES IN TURKEY

submitted by **MEHMET ÇAKMAK** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department**, **Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. İnci Batmaz
Head of Department, **Statistics** _______________

Assist. Prof. Dr. Ceylan Yozgatlıgil
Supervisor, **Statistics Dept., METU** _______________

Assoc. Prof. Dr. Özlem İlk Dağ
Co-Supervisor, **Statistics Dept., METU** _______________

Examining Committee Members:

Prof. Dr. İnci Batmaz
Statistics Dept., METU _______________

Assist. Prof. Dr. Ceylan Yozgatlıgil
Statistics Dept., METU _______________

Assoc. Prof. Dr. Özlem İlk Dağ
Statistics Dept., METU _______________

Assist. Prof. Dr. B. Burçak Başbuğ Erkan
Statistics Dept., METU _______________

Dr. Cem Baş
Economic and Social Indicators Dept., TUIK _______________

Date: 05.12.2014

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: MEHMET ÇAKMAK

Signature :

ABSTRACT

STATISTICAL ANALYSIS OF
ELECTRICITY ENERGY CONSUMPTION
WITH RESPECT TO
PROVINCES IN TURKEY

Çakmak, Mehmet
M.S., Department of Statistics
Supervisor: Assist. Prof. Dr. Ceylan Yozgatlıgil
Co-Supervisor: Assoc. Prof. Dr. Özlem İlk Dağ

December 2014, 89 pages

In recent years, the economic developments in Turkey, rapid increase in population and industrialization and such factors have led to an increment in the demand for electricity in Turkey. Therefore; the accurate estimation of electricity consumption will be important in determining the country's energy strategy. The purpose of this study is to forecast the future electricity energy consumption by acquiring the most consistent and accurate forecast models of the provinces in Turkey by using the fixed effects, random effects and dynamic panel data analysis methods with electricity consumption values of provinces of Turkey between the years of 1999-2011. Forecasting results suggest that the random effects panel data analysis is the best forecasting model among three methods providing the most accurate results.

Keywords: Electricity Consumption, Fixed Effects Panel Data Analysis, Random Effects Panel Data Analysis, Dynamic Panel Data Analysis, Forecasting

ÖZ

TÜRKİYE'DEKİ ELEKTRİK ENERJİSİ TÜKETİMİNİN
İLLER BAZINDA
İSTATİSTİKSEL ANALİZİ

Çakmak, Mehmet
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Ceylan Yozgatlıgil
Ortak Tez Yöneticisi: Doç. Dr. Özlem İlk Dağ

Aralık 2014, 89 sayfa

Son yıllarda Türkiye'de yaşanan ekonomik gelişmeler, nüfus ve sanayileşmenin hızlı bir şekilde artması vb. etkenler, Türkiye'de elektriğe olan talebin hızla artmasına neden olmuştur. Bu yüzden, elektrik tüketiminin doğru tahmin edilmesi, ülkemizin enerji stratejisinin belirlenmesinde önem arz edecektir. Bu çalışmanın amacı; sabit etkiler, tesadüfi etkiler ve dinamik panel veri analizi metotları ile Türkiye'deki illerin 1999-2011 yılları arasında gerçekleşen elektrik tüketim değerlerini kullanarak en tutarlı ve doğru tahmin modelini elde ederek Türkiye'deki illerin gelecek elektrik enerjisi tüketimlerini tahmin etmektir. Öngörü sonuçları, tesadüfi etkiler panel veri analizi tahmin modelinin bu üç yöntem arasında en doğru sonucu verdiğini göstermektedir.

Anahtar Kelimeler: Elektrik Tüketimi, Sabit Etkiler Panel Veri Analizi, Tesadüfi Etkiler Panel Veri Analizi, Dinamik Panel Veri analizi, Öngörü

To My Family

ACKNOWLEDGMENTS

I would like to thank to my supervisor Assist. Prof. Dr. Ceylan Yozgatlıgil for her continuous guidance, motivation and encouragements throughout the all stages of this thesis study. I feel myself extremely privileged to have her as my advisor.

I would like to express my deepest gratitude to my co-advisor Assoc. Prof. Dr. Özlem İlk Dağ. From the first day she was very supportive and I am sure, without her encouragement and guidance this thesis study would not have been completed.

I would like to thank the examining committee members, Prof. Dr. İnci Batmaz, Assist. Prof. Dr. Ceylan Yozgatlıgil, Assoc. Prof. Dr. Özlem İlk Dağ, Assist. Prof. Dr. Burçak Başbuğ Erkan and Dr. Cem Baş for their valuable suggestions, corrections and comments.

I also wish to express my sincere appreciation to my colleague Hamidullah Tarhan for his ideas and supports.

Finally, I would like to express my appreciation for my family for their continuous support, love and patience.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

ABBREVIATIONS

ABC	Artificial Bee Colony
AIS	Artificial Immune System
ANN	Artificial Neural Networks
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
BO	Build-Operate
BOT	Build-Operate-Transfer
ELR	Electricity Leakage Rate
ETS	Exponential Smoothing Method
FGLS	Feasible General Least-Squares
GA	Genetic Algorithm
GDP	Gross Domestic Product
GEE	Generalized Estimating Equations
GNP	Gross National Product
GRP	Gross Regional Product
GWh	Gigawatt hour
LR	Linear Regression
LSE	Least Squares Estimation
MAPE	Mean Absolute Percentage Error
MLE	Maximum Likelihood Estimation
MWh	Megawatt hour
NLR	Nonlinear Regression
PSO	Particle Swarm Optimization

RMLE	Restricted Maximum Likelihood Estimation
RMSE	Root of Mean Square Error
SGK	Social Security Institution
SVD	Singular Value Decomposition
TEIAS	Turkish Electricity Transmission Company
TOR	Transfer of Operating Rights
TUIK	Turkish Statistical Institute
TWh	Terawatt hour
VIF	Variance Inflation Factor

$1\ TWh = 10^3 GWh = 10^6 MWh$

CHAPTER 1

INTRODUCTION

Almost everything in our world is powered by electricity. Electricity is very important because it is the most common energy we consume and we depend on it heavily in our daily lives. Energy consumption increases all around the world due to the rise of economy, population and industrialization. Turkey has also carried out a serious economic and social growth in last decades and this growth caused an increase in energy demand especially in electricity demand. International Energy Agency (2011) mentioned that the demand for electricity has increased very rapidly over the last 25 years all over the world and electricity demand is expected to have the most rapidly increasing rate compared to all the end-user energy forms. Therefore, in order to meet this increasing demand it has become vital to analyze and control reliability of electricity supply.

The rapid increase in electricity demand motivated the privatization movements in all over the world as well as in Turkey. In 2001, Turkish government enacted No. 4628 Electricity Market Law. The main goal of this law is to supply electricity with low cost and high quality to the purchaser. After the entry into force of this law, studies related to electricity consumption estimation gained more importance because the problem with the electricity supply and demand balance was primarily because of the wrong policies due to wrong demand forecasts performed previously. Keleş (2005) stated that the models used by government in the past did not perform well and forecasted demand more than the actual consumption because of technical problems and bad assumptions, resulting with excess capacity, wrong investments like Build-Operate-Transfer (BOT), Build-Operate (BO) and Transfer of Operating Rights (TOR) projects, external dependence, higher electricity prices and an uncompetitive environment.

1.1. Objective of the Thesis

Basically, countries would like to see the future energy consumption in order to supply their future energy needs. In this thesis study, we will try to estimate Turkey's provinces electricity energy consumption using panel data analysis. The aim of the thesis study is to suggest the most accurate model for the electricity demand with respect to provinces of Turkey, not specifying relationship between electricity consumption and price and income elasticity and find out the relationship between electricity consumption and population, industrial enterprise and households of Turkish provinces. To our knowledge, modeling of annual Turkish electricity consumption with respect to provinces has not been studied so far. We believe that modeling the Turkish electricity consumption with respect to provinces will provide more accurate estimation of total electricity consumption of Turkey. Beside this, with the help of provincial electricity consumption forecasts, both the electricity consumption forecasts of provinces and the electricity consumption forecasts of geographical regions would be obtained. Past studies in the literature generally focus on only the annual electricity consumption forecasts of Turkey. These studies do not provide provincial and geographic regional estimate models. Therefore; this study will be an original study and will provide provinces electricity demand model in addition to Turkey's total electricity demand model and will provide more enlightening forecasts in order to take measures on future electricity consumption.

We performed marginal (fixed) effects panel data analysis, random effects panel data analysis and dynamic (transition) panel data analysis techniques to get the most accurate forecasting model of the provinces.

All the estimation and forecasting results of the models are provided by programming in R statistical package.

1.2. Organization of the Thesis

This thesis study is organized in five chapters. Chapter 1 briefly introduces the study and objective of the thesis. Chapter 2 explains electricity consumption forecasting methods in the literature. Chapter 3 gives the general review to the panel data and explains fixed effects panel data analysis, random effects panel data analysis and dynamic panel data analysis methodology in detail. Also panel data estimation and exponential smoothing forecasting method are expressed in this chapter. Chapter 4 describes the data, shows empirical results and forecast results of the models and compares the results. Finally; Chapter 5 presents the summary of the study and conclusions based on the forecast results. Moreover, it gives some perspectives about future investigation.

CHAPTER 2

LITERATURE REVIEW

Various methods have been used in the literature for modeling and forecasting electricity energy demand.

In most of the studies, the goal is to measure the effect of economic activity and energy prices on energy demand. In these studies, the main concern is estimating the effect of income and price elasticities on electricity consumption. Generally, soft computing methods, time series estimation methods, panel data estimation methods and multivariable regression methods are used in the literature in order to forecast electricity consumption. In the literature, gross domestic product (GDP), gross national product (GNP), urbanization rate, electricity prices, temperature values, import and export figures and income are seen as primary determinants of electricity demand.

In this section, we review the modelling studies which analyzed the total electricity demand for Turkey and other countries, so we restricted our interest to the modelling studies generally.

2.1. Electricity Modeling Studies in Turkey

The studies on modeling electricity consumption which are conducted in Turkey, are mostly the soft computing methods such as artificial neural networks (ANNs) methods, genetic algorithm (GA) methods, particle swarm optimization etc. and time series estimation methods, cointegration analyzing methods and some other methods often used in the literature in order to forecast Turkish electricity consumption.

2.1.1. Models

Erdoğdu (2007) developed an electricity consumption model by using cointegration analysis and partial adjustment methods in order to provide estimates of electricity demand in Turkey and forecast values. He used quarterly net electricity consumption per capita for the period between 1984 and 2004 in the estimation process. He selected quarterly real electricity prices and real gross domestic product (GDP) per capita of Turkey as independent variables. The appropriate model is suggested as below.

$\ell n(E_t) = \beta_0 + \beta_1 \ell n(P_t) + \beta_2 \ell n(Y_t) + \beta_3 \ell n(P_{t-2}) + \beta_4 t + \beta_5 \ell n(E_{t-2}) + \varepsilon_t$ where;

E_t is the electricity consumption per capita at time t

P_t is the price of electricity at time t

Y_t is the GDP per capita at time t

After fitting the model, Erdoğdu forecasted yearly electricity demand in Turkey by using autoregressive integrated moving average (ARIMA) modelling for 2005-2014 using the yearly electricity consumption data between the year 1923 and 2004. According to the study results, Erdoğdu reported that the current official electricity demand projections overestimate the electricity demand and he used cointegration analysis and partial adjustment methods. However, when compared to the actual values, forecast results of this study underestimate the electricity consumption. For example, Erdoğdu forecasted the net electricity energy consumption of Turkey in 2012 as 158.15 GWh while the actual value is 194.92 GWh.

Demir and Taşkın (2011) fitted an electricity consumption model by using quasi-Newton method. Using the previous years' electricity consumption amounts for Turkey, they forecasted the demand of electricity consumption. In the quasi-Newton optimization process, only coefficients are selected to optimize the model. Best fitted model function is obtained whenever the fitted curve is compatible with the original electricity consumption curve. In this method, independent variables are not used while developing the model. After developed the model, Demir and Taşkın forecasted the electricity consumption of Turkey. However, when compared to the actual values, forecast results of this study highly overestimate the electricity consumption. They forecasted the electricity energy consumption of Turkey in 2012 as 279.7 GWh while the actual value is 194.92 GWh.

2.1.2. Soft Computing Methods

Soft computing techniques are used to find a solution to problems which cannot be modeled or are too hard to model mathematically. This method has attracted much attention of researchers recently, because it is useful for the nonlinear modeling of large multivariate data sets.

Soft Computing Methods such as Artificial Neural Network, Genetic Algorithms, Swarm intelligence and etc. became very popular in the literature.

2.1.2.1. Artificial Neural Networks (ANN) Method

Recently, ANN method is used as another electricity demand forecasting model in the literature. The advantage of ANN is that it allows forecasting with smaller number of data.

Hamzacebi (2007), Bilgili (2009), Kavaklioglu et al. (2009) used ANN method in order to develop electricity demand forecasting model.

Hamzaçebi (2007) used ANN method to get Turkey's yearly net electricity energy consumption demand on sectorial basis. He used annual data between 1970 and 2004. Hamzaçebi selected the ANN method because this method is superior to forecast future values of more than one variable at the same time and to model the nonlinear relation in the data structure. After developed the fitted model, Hamzaçebi also forecasted the yearly electricity demand in Turkey until 2020. However; when compared to the actual values, forecast results of his study highly overestimate the electricity consumption. For example, Hamzaçebi forecasted the net electricity energy consumption of Turkey in 2012 as 246.52 GWh while the actual value is 194.92 GWh.

Bilgili (2009) employed linear regression method (LR), nonlinear regression (NLR) and ANN methods and compare these methods in order to obtain proper electricity consumption model for Turkey. He selected installed capacity, gross electricity generation, population and total number of subscribers of electricity of Turkey as independent variables and used the dataset between 1990 and 2007. This study showed that ANN method is the best model and it provides best fitted values of electricity consumption of Turkey when compared to LR and NLR.

Kavaklioglu et al. (2009) developed electricity consumption model using ANN method, too. They used electricity consumption amounts for Turkey between 1975 and 2006 along with other economic indicators. They selected GNP, population, import and export figures of Turkey as independent variables. Based on absolute and percentage mean square error, they compared different models. They forecasted electricity consumption of Turkey until 2027. According to the empirical results, they concluded that electricity consumption can be modeled using ANN method and the models can be used to forecast future electricity consumption. When the forecast results are compared to the other studies, it is seen that their fitted model is accurate. Again if we compare 2012 forecasting values, Kavaklioglu et al. forecasted the net electricity energy consumption of Turkey in 2012 as 195.37 GWh while the actual value is 194.92 GWh.

2.1.2.2. Genetic Algorithm Method

Genetic Algorithm (GA) method is also used often as electricity demand forecasting model in the literature.

Öztürk and Ceylan (2005) and Yiğit (2011) used GA approach in order to develop electricity demand forecasting model.

Öztürk and Ceylan (2005) employed GA approach to estimate the total electricity consumption of Turkey using the dataset between the years of 1980 and 2003. They selected GNP, population, import and export figures of Turkey as covariates. Exponential form of genetic algorithm electricity demand ($GAED_{\exp}$), quadratic form of genetic algorithm electricity demand ($GAED_{quad}$) and the mixture form of genetic algorithm electricity demand ($GAED_{mix}$) models are developed for obtaining the best fitted model. They compared the empirical results with actual values and concluded that quadratic form of GAED model is the best fitted model and they created the forecast values using this fitted model.

Yiğit (2011) also implemented the GA approach to estimate the net electricity consumption of Turkey until the year 2020 using the dataset between the years of 1979 and 2009. He selected GDP, population, import and export figures of Turkey as covariates. Linear form of genetic algorithm electricity demand ($GAED_{linear}$) and quadratic form of genetic algorithm electricity demand ($GAED_{quadratic}$) models are developed for obtaining the best fitted model. After developing the fitted model, he created the forecast values using both $GAED_{linear}$ and $GAED_{quadratic}$ models.

Dilaver and Hunt (2011) investigated the relationship between Turkish industrial electricity consumption, industrial value added and electricity prices in order to forecast future Turkish industrial electricity demand by applying the structural time series technique to the yearly data between the years of 1960 and 2008. They forecasted that Turkish industrial electricity demand will be between 97 and 148 TWh by 2020.

Dilaver and Hunt (2011) also investigated the relationship between Turkish total electricity consumption, GDP and electricity prices in order to forecast future Turkish total electricity demand by applying the structural time series technique to the yearly data between the years of 1960 and 2008. According to the study results

GDP and electricity prices have an important role on Turkish electricity demand. They forecasted that Turkish total electricity demand will be between 259 and 368 TWh in 2020.

Kıran et al. (2012) developed two new models based on artificial bee colony (ABC) and particle swarm optimization (PSO) techniques in order to estimate electricity energy demand in Turkey. He used GDP, population, import and export figures of Turkey as covariates. He proposed the models as linear and quadratic form. According to their empirical results the quadratic form provides better-fit solutions than linear form due to fluctuations of the socio-economic indicators. They also forecasted the Turkey's electricity energy demand until 2025 based on three different scenarios.

Kavaklioglu (2014) used multivariable regression method in order to model Turkey's electricity consumption through a nonlinear relationship using the annual data from 1970 to 2011. He selected GDP per capita, population, import and export figures of Turkey as covariates. He applied the Singular Value Decomposition (SVD) method to reduce the dimensionality of the problem and to provide robustness to the estimations. According to the results, Turkey's electricity consumption can be robustly modeled using SVD.

2.2. Modeling Studies on Electricity Consumption in the World

Azadeh et al. (2014) have analyzed the integrated algorithm for forecasting annual electrical energy consumption based on Artificial Immune System (AIS), GA, PSO and computer simulation. They have applied these methods to 16 countries from the years between 1980 and 2006 using the country's GDP and populations as explanatory variables. Their main aim was to compare these methods in order to obtain the best fitted model for forecasting purpose. They compared the empirical results by using mean absolute percentage error (MAPE) and reported that the AIS is the best method.

Narayan and Smyth (2005) used bounds testing approach to cointegration and Autoregressive Distribution Lag Model to estimate annual residential electricity demand in Australia. They selected income, temperature, electricity price and natural gas price as covariates for the two models. In the first model price data are used and in the second model the ratio of electricity price to natural gas is used. Their suggested models are represented as below;

$$\ell nEC_t = \alpha_0 + \alpha_1 \ell nY_t + \alpha_2 \ell nEP_t + \alpha_3 \ell nGP_t + \alpha_4 \ell nTM_t + \varepsilon_t \quad \text{for model 1.}$$

$$\ell nEC_t = \alpha_0 + \alpha_1 \ell nY_t + \alpha_2 \ell nRP_t + \alpha_3 \ell nTM_t + \mu_t \quad \text{for model 2.}$$

where,

EC is the per capita residential electricity consumption (kWh per capita),

Y is the real per capita income,

EP is the real residential electricity price ($A/kWh),

GP is the real price of natural gas ($A/kWh),

RP is the ratio of the real price of electricity to the real price of natural gas,

TM is the temperature variable.

Finally, ε_t and μ_t are error terms assumed to be white noises and normally and identically distributed.

According to the results, long run relationship between income and electricity consumption is found, and short and long run coefficients are estimated since the variables are found to be cointegrated. Natural gas price is insignificant in the first model, but in the second model, the price ratio is significant. Temperature has positive and significant effect on the electricity consumption only in the first model.

Mohamed and Bodger (2005) analyzed the impact of the economic and demographic variables on the yearly electricity consumption in New Zealand. They used GDP, average price of electricity and population of New Zealand as independent variables for each of the domestic and non-domestic sectors and total consumption using 35 years of data from 1965 to 1999 for each of the variables. They used multiple linear regression analysis. Based on the empirical results, electricity consumption of New Zealand is affected by the GDP, electricity price and the population. After constructing the proper model, they forecasted the electricity consumption and compared these forecasts with some available national forecasts and they concluded that electricity consumption can be modeled using multiple linear regression analysis and the models can be used to forecast future electricity consumption.

Bianco et al. (2009) also used different multiple linear regression models for estimating the total electricity consumption in Italy using the annual data from the time period between 1970 and 2007. They used GDP, GDP per capita and population as covariates. Firstly, domestic and non-domestic price elasticities are analyzed with respect to both short run and long run. They found domestic and non-domestic price elasticity as inelastic and indicate that electricity energy is a necessity and a big increase in the price of electricity will reduce very little amount of electricity demand. After analyzing the price elasticity, they proposed different regression models based on co-integrated or stationary data to forecast electricity consumption of Italy. Based on the empirical results, they compared their forecast results with national forecasts and concluded that the developed regressions are compatible with the official projections.

Inglesi (2010) employed the Engle–Granger methodology for cointegration and Error Correction models to forecast electricity demand of South Africa by using yearly data for the period between 1980 and 2005. He used real GDP, average electricity price, real disposable income and population as independent variables. According to the study result, in the long run, there is a relationship between electricity consumption and price and also with income. In the short run, the demand for electricity is explained by the GDP and the population of the country. In addition to the elasticity analyzes, Inglesi also forecasted the electricity demand of South Africa using two different scenarios until 2030.

Atakhanova and Howie (2007) performed fixed effects, random effects, and feasible general least-squares (FGLS) panel data methods in order to estimate Kazakhstan's total electricity demand, electricity demand in the industrial, service, and residential sectors using the regional dataset between the years of 1994 and 2003. They used real income, real electricity prices, population, industrial share in the total gross regional product (GRP) and efficiency in the industrial sector as covariates. Total electricity demand model by using random effects panel data method is developed for the following model;

$$\Delta Q_{it} = \phi_0 + \phi_1 \Delta Y_{it} + \phi_2 (\Delta Y_{it} \times D99_t) + \phi_3 \Delta P_{it} + \phi_4 \Delta POP_{it} + \phi_5 \Delta Share_{it} \\ + \phi_6 \Delta Eff_{it} + \mu_{it} + \varepsilon_{it}$$

where;

Δ shows the amount of change in the variable,

Q_{it} is quantity of electricity demanded,

$D99_t$ is a dummy variable which takes the value of one for the period 1999–2003,

Y_{it} is the real income,

P_{it} is the real price of electricity,

POP_{it} is the population,

$Share_{it}$ is industry's share in gross regional product (GRP),

Eff_{it} is the efficiency in the industrial sector,

ε_{it} is the error term with a variance of Σ.

Based on the estimation and specification test results, the random effects model is the best fitted model for the total electricity demand. Random effects demand model is also used for forecasting purposes. They forecasted the electricity demand for years 2010 and 2015 under medium, high and low economic growth scenarios. It is concluded that electricity demand may grow at either 3% or 5% per year depending on rates of economic growth, government policy regarding price increases and promotion of efficiency.

As it can be seen, various methods are used to estimate electricity consumption or demand in different countries. In the light of these studies, it is seen that the effect of different provinces on electricity consumption of Turkey has not been studied so far. We decide to use different panel data analysis methods to develop the electricity demand model of Turkey using its provinces. We discuss modeling issues in the next section.

CHAPTER 3

METHODOLOGY

As stated earlier, the aim of the thesis study is to suggest the best fitted model and forecast of electricity consumption of Turkey with respect to provinces. In the literature, various methods have been used for modeling and forecasting electricity energy demand such as time series, panel cointegration, soft computing methods and so on. In this thesis study, we will try to obtain the best fitted model and forecast by using panel data analysis methods.

3.1. Panel Data Analysis

Panel data analysis is a statistical method in which we observe repeated cross-sections of the same individuals. It is widely used in the area of social science, epidemiology, and econometrics. In panel data analysis, same entities (panels) such as individuals, companies, firms, countries etc. are observed at multiple time points.

Table 3.1 points out a sample of panel data set. As can be seen from the table, panel data are used whenever information is requested for both units and periods. However; cross sectional data give information about only a period of several units and time series data give information only for one unit according to the period.

Table 3.1: A Sample of Panel Data Set

Unit ID	Year	Y	X1	X2	X3
1	2010	8.8	7.8	5.8	1.8
1	2011	7.4	0.6	7.9	4.8
1	2012	9.4	2.1	5.4	1.3
2	2010	9.1	1.3	6.7	4.1
2	2011	8.3	0.9	6.6	5.0
2	2012	0.6	6.9	0.7	7.2
…					
300	2010	9.1	0.2	2.6	6.4
300	2011	4.8	5.9	3.2	3.2
300	2012	9.1	5.2	6.9	2.1

Advantages of Panel Data

We can explain the main advantages of panel data as following;

- ✓ Panel data provide more accurate inference of model parameters when compared to a single cross-section or time series data.
- ✓ Panel data are more informative and involve more variability, less collinearity
- ✓ Panel data provide more degrees of freedom than a cross-sectional data.
- ✓ Panel data allow us to control for variables that we cannot observe or measure.
- ✓ Panel data give information on the time-ordering of events.
- ✓ Panel data allow us to identify individual and time effects which cannot be identified by pure cross-sectional or time series data.

3.1.1. Fixed Effects Panel Data Model (Marginal Model)

Fixed effects panel data method is used when we are analyzing the effects of variables on response. In other words, in fixed effects panel data, dependent variable is modelled as a function of independent variables, while taking the within-subject correlation into account. In fixed effects panel data, subjects are assumed to be independent to each other. The primary aim of fixed effects panel data is to compare groups like male/female but not compare to the individuals. Fixed effects panel data explore the relationship between dependent variables and independent variable within an entity. This entity can be a country, a person or a company and affects the covariates.

The equation for the fixed effects model becomes:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + ... + \beta_k X_{kit} + u_{it}$$

where $u_{it} \sim \mathrm{N}(0, \Sigma)$

Y_{it} is the dependent variable where i =1,2,...,N stands for entity and t=1,2,...,T corresponds to time

β_k is the coefficient of X_{kit}

X_{kit} represents independent variables for entity i =1,2,...,N and time t = 1,2,...,T

u_{it} is the error term with a variance of Σ. Some possible structures for Σ include autoregressive or Toeplitz structures.

3.1.2. Random Effects Panel Data Model

Unlike the fixed effects model, in random effects model, the variation across entities is assumed to be random and uncorrelated with the independent variables included in the model. The main difference between fixed effects model and random effects model is that in random effects model, the omitted variables are uncorrelated with the independent variables while omitted variables are correlated with the independent variables in fixed effects model.

Random effects model allows model parameters to vary from one subject to another. Therefore, it provides heterogeneity among individuals. Random effects model is used when the differences across entities have some impact on your dependent variable.

The equation for the random effects model becomes:

$$Y_{it} = \beta_0 + \beta_1 \ X_{1it} + \beta_2 \ X_{2it} + ... + \beta_k \ X_{kit} + \alpha_{0i} + \alpha_{1i} Z_{1it} + ... + \alpha_{\ell i} Z_{\ell it} + u_{it}$$

where $u_{it} \sim \mathrm{N}(0, \Sigma)$

Y_{it} is the dependent variable where i =1,2,…,N stands for entity and t=1,2,...,T corresponds to time

β is the coefficient of X_{it}

X_{it} represents independent variables for entity i =1,2,...,N and time t = 1,2,...,T

α_{0i} is the unknown intercept for each entity where i =1,2,...,N

$\underline{\alpha_i} = (\alpha_{1i}, ..., \alpha_{\ell i}) \sim N(\underline{0}, \underline{D})$ are random slope terms for each entity

Z_{it} is a subset of X_{it}

u_{it} is the error term with a variance of Σ. Some possible structures for Σ include autoregressive or Toeplitz structures.

In contrast to fixed effects model, random effects model contains both between entity error and within entity error.

3.1.3. Dynamic Panel Data Model (Transition Model)

In Dynamic (Transition) Panel Data Model, dependent variable at time t is modelled depending on the independent variables at time t and the response at the previous time points. The aim of dynamic panel data is to learn from history. Dynamic panel data can be used only for balanced data. The advantage of dynamic panel data over cross-sectional data is that dynamic panel data provide much sufficient knowledge about the past time periods. Dynamic panel data models can be seen as a special case of fixed effects models. Unlike fixed effects model, dynamic panel models include past values of dependent variable in addition to other covariates in the model.

The equation for the dynamic model becomes:

$$Y_{it} = \beta_0 + \beta_1\ X_{1it} + \beta_2\ X_{2it} + \ldots + \beta_k\ X_{kit} + \alpha_1 Y_{it-1} + \ldots + \alpha_m Y_{it-m} + u_{it}$$

where;

Y_{it} is the dependent variable where i =1,2,...,N stands for entity and t=1,2,...,T corresponds to time

β is the coefficient of X_{it}

X_{it} represents independent variables for entity i =1,2,...,N and time t = 1,2,...,T

α_m is the coefficient of lag variables

$Y_{i,t-m}$ is the lag variable where i =1,2,...,N stands for entity, t = 1,2,...,T corresponds to time and m=1 for lag-1 model

u_{it} is the error term with a variance of Σ, which is an identity matrix.

In chapter 4, we will perform marginal (fixed) effects panel data analysis, random effects panel data analysis and dynamic (transition) panel data analysis techniques separately and compare these methods.

3.2. Estimation

In fixed effects panel data analysis and dynamic panel data analysis; we use generalized estimating equations (GEE) technique for the estimation and in random effects panel data analysis; we use restricted maximum likelihood maximum likelihood estimation (RMLE) method.

GEE method is used to estimate the marginal regression parameters for correlated responses and provides a general approach for analyzing correlated responses. These responses can be discrete or continuous. The main idea behind GEE is to generalize the usual likelihood equations with a univariate response by incorporating the covariance matrix of the vector of responses. Dunlop (1994), Diggle et al (1994) and Liang and Zeger (1995) showed that GEE is simpler and theoretically superior to its competition with least squares estimation (LSE) and MLE. However, the classical GEE method yields biased results when there are missing cases that are not completely at random. Extensions of GEE technique is proposed for such situations.

RMLE is also widely used method in estimating panel data models. In general, MLE selects the set of values of the model parameters and maximizes the likelihood function for a fixed set of data and the model. It provides accurate coefficient estimates whether or not correlation structure choice is correct. MLE is sufficient estimator. It is also asymptotically consistent and asymptotically efficient method which converges to the true values and provides the most precise estimates when sample size gets larger.

3.3. Forecasting

Forecasting can be described as predicting what will happen in the future based on historical data. Forecasting is based on the assumption that the past predicts the future. There are three methods for time series models for forecasting which are namely naive method, moving average method and exponential smoothing method. These methods help to provide reliable results, but cannot provide completely accurate results. In this study, exponential smoothing method (ETS) is used as forecasting method since we don’t have a long time series data.

3.3.1. Exponential Smoothing Method (ETS)

The main idea of exponential smoothing method is that forecasting the future based on mostly the most recent observation and forecast is provided from an exponentially weighted average of past observations. In other words; exponential smoothing method smooths the original series and use the smoothed series in forecasting future values of the variable based on the most recent observation. Hyndman (2002) showed the exponential smoothing equations as following;

$$F_t = \alpha A_{t-1} + (1-\alpha)F_{t-1}$$

where;

A_t is the actual value at time t,

F_t is the forecasted value at time t,

α is the smoothing constant which ranges from 0 to 1.

In order to forecast current period F_{t-1} , it is written in the following fashion;

$$F_{t-1} = \alpha A_{t-2} + (1-\alpha)F_{t-2}.$$

If we substitute the equations, we obtain the following;

$$F_t = \alpha A_{t-1} + (1-\alpha)\left[\alpha A_{t-2} + (1-\alpha)F_{t-2}\right].$$

We define the F_{t-2} as the following;

$$F_{t-2} = \alpha A_{t-3} + (1-\alpha)F_{t-3}.$$

Finally if we substitute all equations, we get the following formula

$$F_t = \alpha A_{t-1} + \alpha(1-\alpha)A_{t-2} + \alpha(1-\alpha)^2 A_{t-3} + \alpha(1-\alpha)^3 A_{t-4} + \alpha(1-\alpha)^4 A_{t-5} + ...$$

Therefore; if we increase those decimal weights, the values declines exponentially.

Exponential smoothing method is very practical method. The reason why we choose exponential smoothing method is that it uses less storage space for data. It is also extremely accurate forecasting method compared to other methods and it has little calculation complexity.

Hyndman et al. (2002) giving 30 models in total, described how each exponential smoothing method (ETS) corresponds to two state space models. They also discussed an automatic algorithm for identifying a proper exponential smoothing model in a general class of state space models. They showed that all exponential smoothing methods (including non-linear methods) are optimal forecasts from innovations state space giving a total of fifteen methods as shown in Table 3.2.

Table 3.2: Exponential Smoothing Methods (Source Hyndman et. al 2002)

	Seasonal Component		
Trend Component	N (None)	A (Additive)	M (Multiplicative)
N (None)	N,N	N,A	N,M
A (Additive)	A,N	A,A	A,M
A_d (Additive damped)	A_d,N	A_d,A	A_d,M
M (Multiplicative)	M,N		M,A
M_d (Multiplicative damped)	M,M		
	M_d,N		M_d,A
	M_d,M		

Note: In the R output, there are three letters to explain the models where the first one is for the error term being additive or multiplicative and other letters as shown in the table.

To be able to forecast the response variable, conditional expectation of future values of response given the past values of all the variables and forecast of explanatory variables are used. The ℓ-step ahead forecast of the response variable is shown as the following;

$$\hat{y}_n(\ell) = \boldsymbol{E}\left[y_{n+\ell} \middle| y_1, \cdots, y_{\boldsymbol{n}}, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{\boldsymbol{n}}, \hat{\boldsymbol{x}}_{\boldsymbol{n}}(\ell)\right]$$

where;

$y_1,\cdots,y_{\boldsymbol{n}}$ are observed values of response variable,

$\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{\boldsymbol{n}}$ are observed values of independent variable,

$\hat{\boldsymbol{x}}_{\boldsymbol{n}}(\ell)$ is a vector of forecasted covariates obtained by ETS.

In chapter 4, we will perform exponential smoothing method (ETS) to our independent variables and forecast these values. After forecasting independent variables, we will obtain forecasted electricity consumptions of Turkey and its provinces until 2015.

CHAPTER 4

DATA DESCRIPTION AND EMPRICAL RESULTS

4.1. Data Description

We have performed our analysis on a dataset of annual electricity consumption of Turkey's provinces from 1999 to 2011. The last year (2011) is allocated for forecast evaluation. The data used in this study covered annual electricity consumption of provinces, annual population of provinces, annual number of industrial enterprise of provinces / annual population of provinces and finally annual number of household of provinces / annual population of provinces.

After investigating the literature, it is considered that electricity consumption is affected by population, GDP, urbanization rate, total number of household and industrial enterprise and air temperatures of provinces. In the study, it was also planned to use GDP or industrialization index of provinces as covariate. Unfortunately, there is no such index available, and GDP is available for only Turkey not for provinces between 2002 and 2010. Although it is believed that these covariates will be very helpful to explain the variation in electricity consumption of Turkey, they cannot be used. We also tried to use temperature variable as covariate. We created several temperature variables such as average seasonal temperatures of provinces, the annual minimum and maximum temperatures of provinces. However, none of them have a significant effect on the electricity consumption of Turkey.

Annual electricity consumption of Turkey's provinces 1999-2011 (EC) is received from the Turkish Electricity Transmission Company (TEIAS). Annual population of Turkey's provinces between 1999 and 2011 (POP) is obtained from Turkish Statistical Institute (TUIK). Here, the data of annual population of provinces in 2000 and after the year 2007 are official data and the rest are projection data obtained from TUIK. Annual number of industrial enterprise of provinces is obtained from Social Security Institution (SGK) and some portion of the annual number of household of provinces is obtained from Turkish Statistical Institute (TUIK). TUIK has announced the number of household of provinces only in 1984, 2000 and after the year 2010. Our data require the number of household of provinces between 1999 and 2011. TUIK has the building permit statistics collected yearly. We obtained the annual occupancy permit from TUIK. Then, we added the annual occupancy permit to household of provinces between the years of 2000 and 2010 and finally we obtained number of household of provinces between the years of 2000 and 2010.

Table 4.1 shows a portion of our data. In the table, ID indicates the province codes of provinces. EC represents annual electricity consumption in thousand MWh. POP shows annual population in thousand. INDUSTRY states the annual number of industrial enterprise of provinces over annual population of provinces and HOUSEHOLD states the annual number of household of provinces over annual population of provinces.

Table 4.1: A Portion of Original Data

ID	TIME	EC (Thousand MWh)	POP (Thousand)	INDUSTRY	HOUSEHOLD
1	1999	2898.691	1842.446	0.002162	0.239147
1	2000	3029.431	1868.986	0.002027	0.240809
1	2001	2894.674	1890.021	0.001811	0.240955
1	2002	2951.605	1908.789	0.001839	0.239536
1	2003	3092.526	1927.074	0.001926	0.238116
1	2004	3709.263	1946.322	0.002053	0.236849
1	2005	3887.871	1966.258	0.002129	0.236514
1	2006	3907.265	1986.629	0.002228	0.237473
1	2007	4135.636	2006.65	0.002340	0.238717
1	2008	4112.09	2026.319	0.002334	0.240662

Table 4.1 (continued)

1	2009	3874.304	2062.226	0.002389	0.241153
1	2010	4003.428	2085.225	0.002449	0.242844
2	1999	506.67	561.194	0.000581	0.185904
2	2000	518.534	569.278	0.000591	0.183811
2	2001	524.052	571.915	0.000542	0.183199
2	2002	545.072	573.77	0.000525	0.182823
2	2003	573.353	575.388	0.000499	0.182770
2	2004	593.185	577.19	0.000632	0.182736
2	2005	703.72	579.109	0.000722	0.182951
2	2006	777.385	581.057	0.000773	0.183835
2	2007	832.558	582.762	0.000849	0.186413
2	2008	844.709	585.067	0.001231	0.187309
2	2009	835.84	588.475	0.001237	0.189135
2	2010	900.381	590.935	0.001362	0.191585
…					
81	1999	244.986	292.203	0.002310	0.239855
81	2000	326.275	296.412	0.002329	0.237134
81	2001	322.406	300.426	0.002250	0.235407
81	2002	298.717	304.096	0.002269	0.236917
81	2003	366.052	307.706	0.002385	0.247448
81	2004	376.059	311.489	0.002661	0.248829
81	2005	458.083	315.396	0.003041	0.249690
81	2006	515.468	319.392	0.003106	0.250026
81	2007	570.58	323.328	0.003356	0.249027
81	2008	615.789	328.611	0.003484	0.249424
81	2009	642.961	335.156	0.003419	0.254635
81	2010	725.679	338.188	0.003735	0.260182

4.2. Exploratory Analysis of Data

In order to summarize and visualize the important characteristics and develop intuition about our data set, we made exploratory data analysis.

4.2.1. Descriptive Statistics of the Data

Descriptive statistics of all the variables in the study are examined by using R software and are reported in Table 4.2.

Table 4.2: Descriptive Statistics

	EC(Thousand) MWh	POP(Thousand)	INDUSTRY	HOUSEHOLD
Min Value	32.83	74.41	0.000064	0.1259
1st Quartile	268.65	272.24	0.001006	0.2111
Median	598.36	476.11	0.001933	0.2406
Mean	1577.17	846.42	0.002080	0.2356
3rd Quartile	1480.63	886.50	0.002880	0.2675
Max Value	30525.03	13255.68	0.007488	0.3377
Std. Deviation	3183.34	1439.90	0.001363	0.0453

According to Table 4.2, the mean of electricity consumption is 1.577 GWh. However, it is clearly seen that the min and the max value of the electricity consumption is very far from the mean. Also, the standard deviation of EC is very high. They represent that electricity consumptions of the provinces are very different from each other. Since mean of EC is greater than median of EC, it shows right skewed distribution. Therefore, a transformation might be needed to satisfy the normality of the data. The difference between minimum and maximum values of population is also quite big. This is because the provinces of Turkey are different from each other with respect to population, industry, development and etc. Turkey has big provinces such as İstanbul, Ankara, İzmir and Bursa and also small provinces like Bayburt, Tunceli and Ardahan. Therefore, when forecasting the electricity consumption of Turkey, the estimation of electricity consumption with respect to provinces is very important. In our study, we will try to model the electricity consumption of Turkey with respect to the provinces.

As can be seen in Figure 4.1, the median of electricity consumption varies according to years. As the year increases, electricity consumption also increases.

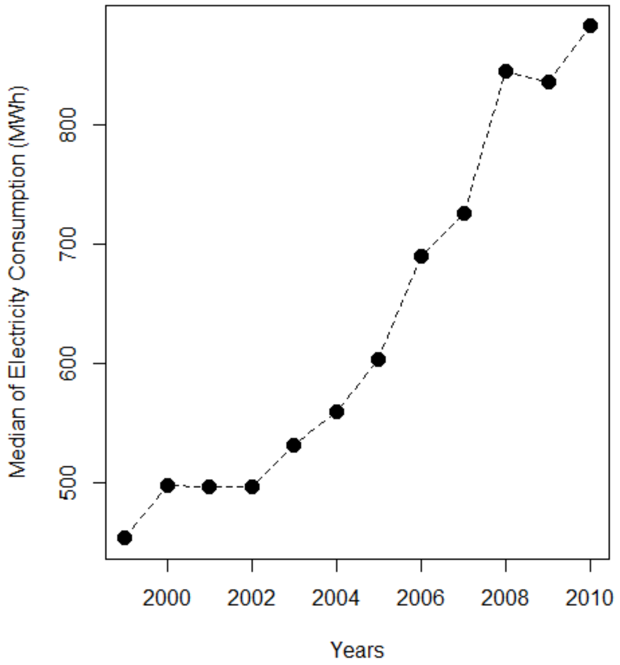


Figure 4.1: Annual Median of Electricity Consumption

4.2.2. Normality Checks

Statistical analysis should satisfy the assumption that the population follows a normal distribution with a common variance and additive error structure. When the relevant theoretical assumptions are approximately satisfied, the usual procedures can be applied in order to make inferences about unknown parameters of interest.

Before starting the confirmatory analysis, we have created histogram and normal QQ plot of the annual electricity consumption in order to have an idea about normality of data.

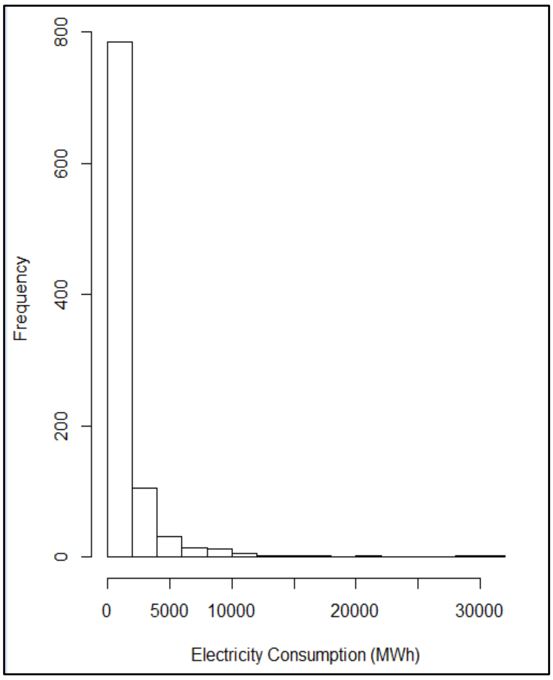


Figure 4.2: Histogram of Annual Electricity Consumption (MWh)

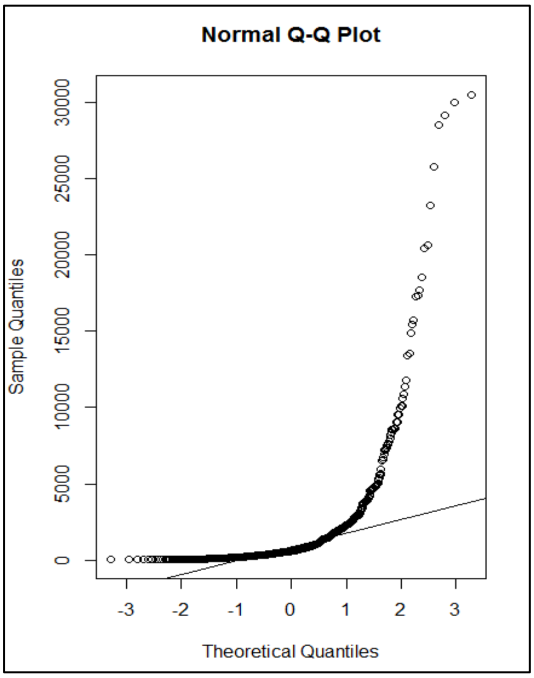


Figure 4.3: Normal QQ Plot of Annual Electricity Consumption (MWh)

Both the histogram and Q-Q plot show that annual electricity consumption are not distributed as normal distribution. Therefore, we have to apply transformation to the data.

4.2.3. Box-Cox Transformation

When the normality assumption is seriously violated, Box-Cox transformation can be used in order to hold normality assumption.

The aim of the Box-Cox transformation is to guarantee the normality assumption for linear models. The original form of the Box-Cox transformation is shown in the following form:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0; \\ \log y, & \text{if } \lambda = 0. \end{cases}$$

We decided to apply Box-Cox transformation to data in order to satisfy the normality assumption. Figure 4.4 shows the ranked log likelihood values and corresponding lambda value. After computing the log likelihood and lambda values, we conclude that Box-Cox transformation lambda value should be close to 0.2626. We take the λ value as 0.2 for the simplicity and transformed our data.

	lamda	lik
[1,]	**0.26262626**	-3105.954
[2,]	0.22222222	-3106.039
[3,]	0.30303030	-3108.774
[4,]	0.18181818	-3108.975
[5,]	0.34343434	-3114.562
[6,]	0.14141414	-3114.721
[7,]	0.10101010	-3123.246
[8,]	0.38383838	-3123.419
[9,]	0.06060606	-3134.523
[10,]	0.42424242	-3135.454
[11,]	0.02020202	-3148.526
[12,]	0.46464646	-3150.823
[13,]	-0.02020202	-3165.230
[14,]	0.50505051	-3169.718
[15,]	-0.06060606	-3184.607
[16,]	0.54545455	-3192.356
[17,]	-0.10101010	-3206.623
[18,]	0.58585859	-3219.020
[19,]	-0.14141414	-3231.242
[20,]	0.62626263	-3250.013
[21,]	-0.18181818	-3258.422
[22,]	0.66666667	-3285.678
[23,]	-0.22222222	-3288.116

Figure 4.4: Log Likelihood Values

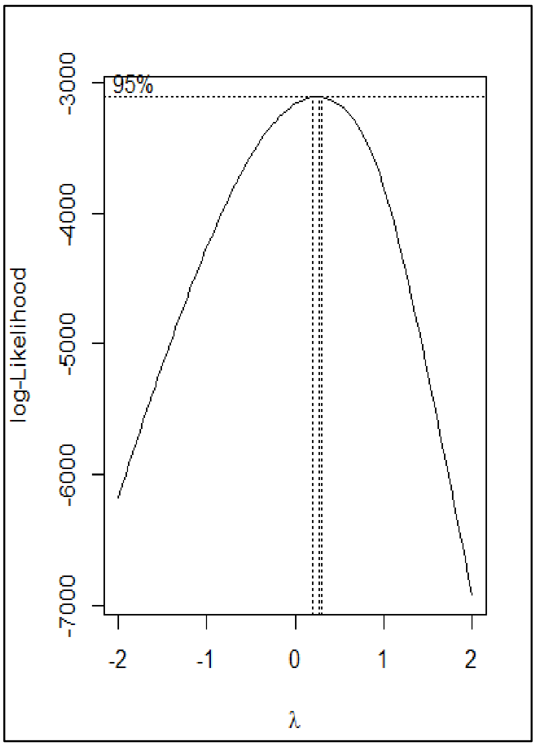
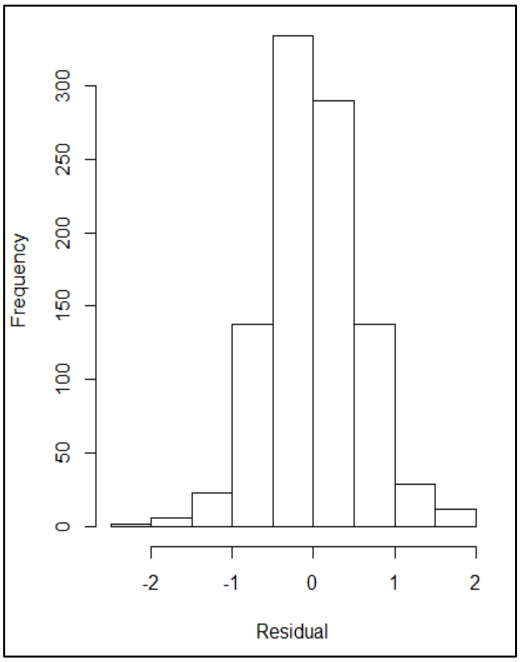


Figure 4.5: Proposed Lambda Value (λ)

After applying the Box-Cox transformation to our data, the new histogram and QQ normal plot are checked. Figure 4.6 and Figure 4.7 indicate that normality assumption is approximately satisfied.



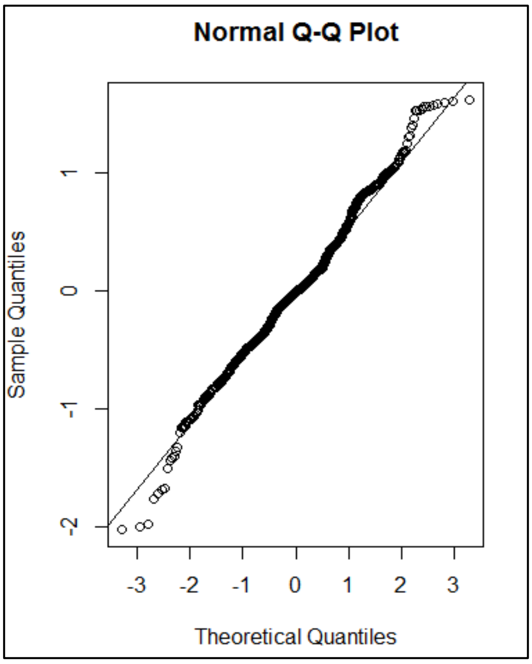


Figure 4.6: Transformed Histogram of Annual Electricity Consumption (MWh)

Figure 4.7: Transformed Normal QQ Plot of Annual Electricity Consumption (MWh)

Finally, in the estimation model; population of provinces, total number of industrial enterprise of provinces and total number of household of provinces are used as covariates. Total number of industrial enterprise and total number of household are not used directly. They are placed in the model as number of industrial enterprise of province / population of province and number of household of province / population of province. The reason behind that is to prevent multicollinearity between two variables.

4.2.4. Correlation of Variables

To be able to see whether the chosen covariates are highly related with our response variable, we construct a pairwise correlation matrix given in Table 4.3.

Table 4.3: Pairwise Correlation Matrix

	ID	TIME	EC	POP	INDUSTRY	HOUSEHOLD
ID	1.0000000	0.00000000	-0.2707378	-0.17581139	-0.1472733	-0.1025839
TIME	0.0000000	1.00000000	0.1611058	0.02591038	0.2945041	0.1404950
EC	-0.2707378	0.16110576	1.0000000	0.71425354	0.6996164	0.4393419
POP	-0.1758114	0.02591038	**0.7142535**	1.00000000	0.4647330	0.1875142
INDUSTRY	-0.1472733	0.29450411	**0.6996164**	0.46473297	1.0000000	0.7570116
HOUSEHOLD	-0.1025839	0.14049502	**0.4393419**	0.18751420	0.7570116	1.0000000

As can be seen clearly from this table, the most correlated independent variable to electricity consumption is population with coefficient 0.714. It is followed by industry and household covariates. In addition to the correlation matrix, cross correlation matrix is controlled whether to see how last year's values affected the current electricity consumption of provinces.

Table 4.4: Cross Correlation Matrix

	ID	TIME	EC	POP	INDUSTRY	HOUSEHOLD
ID	1.00000000	0.00000000	-0.2695278	-0.17601534	-0.1479756	-0.09860064
TIME	0.00000000	1.00000000	0.1512388	0.02345494	0.2593072	0.11751947
EC	-0.26952776	0.15123876	1.0000000	0.71414138	0.7033747	0.44941793
POP	-0.17601534	0.02345494	**0.7141414**	1.00000000	0.4726291	0.19813318
INDUSTRY	-0.14797556	0.25930717	**0.7033747**	0.47262910	1.0000000	0.75665007
HOUSEHOLD	-0.09860064	0.11751947	**0.4494179**	0.19813318	0.7566501	1.00000000

Table 4.3 and Table 4.4 reveal that the cross-correlation matrix is very similar to correlation matrix. So, it is unnecessary to use cross-correlation case in the model because it causes to lose extra one year of data.

4.2.5. Linearity between Dependent Variable and Covariates

Before developing the estimation model, linearity between dependent variable and covariates are tested. In order to see linearity, scatter plots of variables are constructed. According to scatter plot in Figure 4.8, there is an exponential correlation between electricity consumption and population. In order to make it linear, logarithm of population is taken into account. Figure 4.9 indicates that after taken logarithm of population, linear relation between electricity consumption and population is obtained.

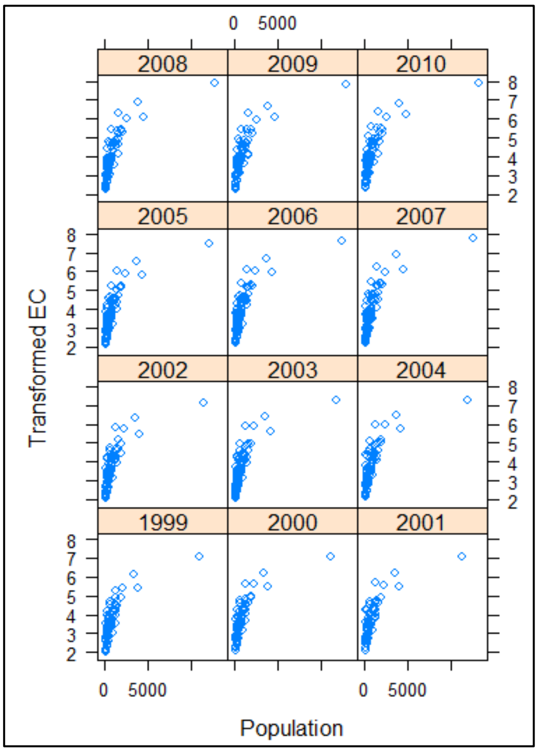


Figure 4.8: Scatter Plot of EC vs Population

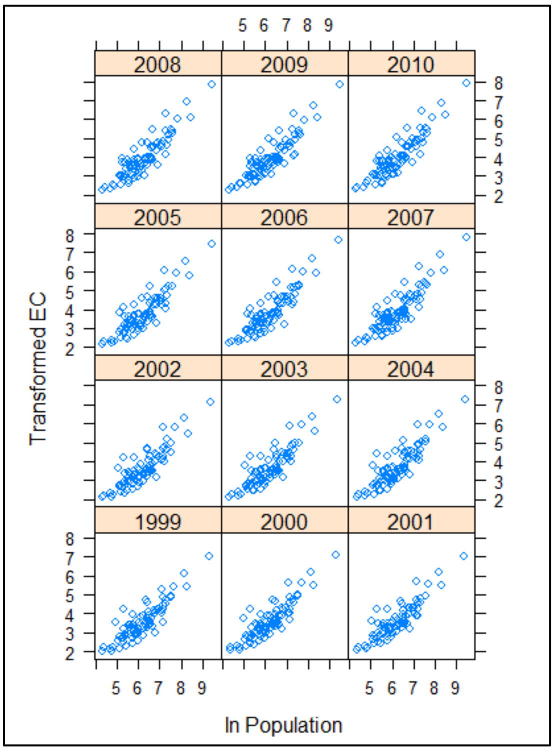


Figure 4.9: Scatter Plot of EC vs ℓn Population

In order to see linearity between electricity consumption of provinces and number of industrial enterprise of provinces, Figure 4.10 is constructed. This figure indicates that there is a linear correlation between electricity consumption and the number of industrial enterprise. Thus, this covariate does not require any transformation.

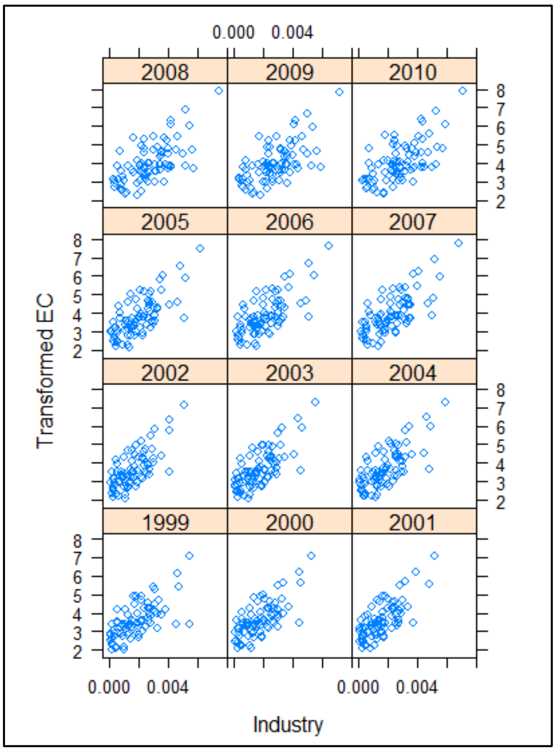


Figure 4.10: Scatter Plot of EC vs Industry

Finally, to figure out linearity between electricity consumption of provinces and number of household of provinces, Figure 4.11 is created. Figure 4.11 points out some non-linearity. Based on the form of the scatter plots, in the power transformation, the power has to be greater than 1. After trying several power values, it is decided to use the fifth degree transformation in order to solve this problem. Thus, new scatter plot is created by using fifth degree of household in Figure 4.12.

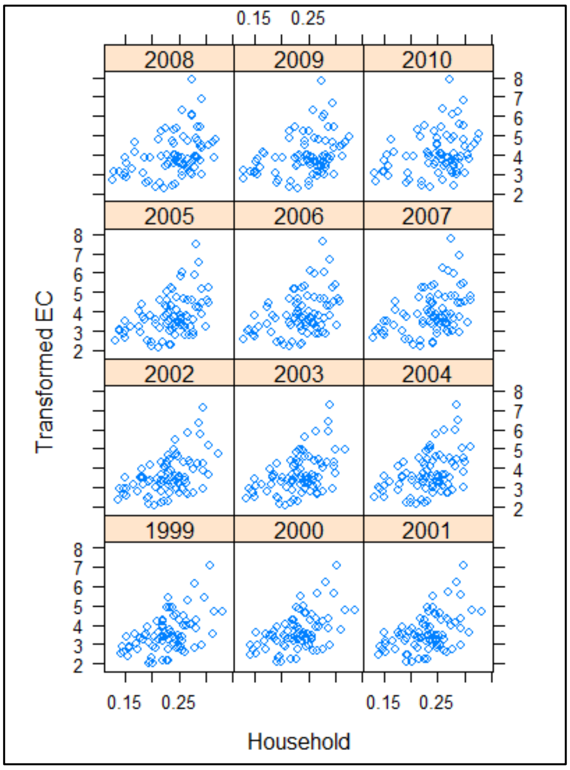


Figure 4.11: Scatter Plot of EC vs Household

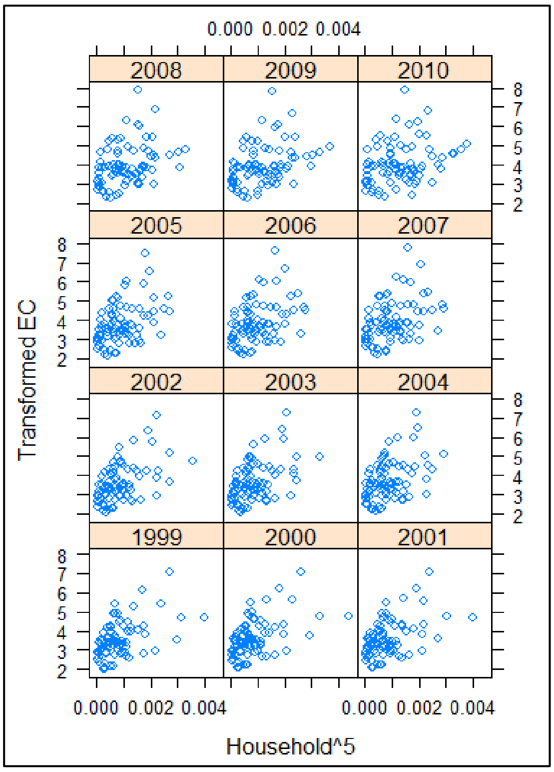


Figure 4.12: Scatter Plot of EC vs Household^5

4.3. Models

The fixed effects model is created in the light of all these processes as following;

$$EC^{Tr}{}_{it} = \beta_0 + \beta_1 TIME_{it} + \beta_2 \ell n(POP)_{it} + \beta_3 INDUSTRY_{it} + \beta_4 HOUSEHOLD^5 + u_{it}$$

where ;

$i = 1,2,\ldots,81,$

$u_{it} \sim \mathrm{N}(0,\Sigma)$, and

t = 0,1,…,11 (we subtract the original time values from 1999)

The other models, random effects model and dynamic model, are developed in a similar fashion. Specifically, the same covariates are used with these transformations.

4.4. Empirical Results

In this thesis study, we used panel data analysis method which is also known as longitudinal. We developed fixed effects, random effects and dynamic panel data models separately and compare the results of these models in order to get best estimation model.

4.4.1. Results of Fixed Effects Panel Data Model

First, we construct the model below which satisfies the linearity assumption.

$$EC^{Tr}{}_{it} = \beta_0 + \beta_1 TIME_{it} + \beta_2 \ell n(POP)_{it} + \beta_3 INDUSTRY_{it} + \beta_4 HOUSEHOLD^5 + u_{it}$$

where;

$t = 0,1,\ldots,11,$

$i = 1,2,\ldots,81,$

$u_{it} \sim \mathrm{N}(0,\Sigma)$.

In order to decide on the correlation structure of marginal model, we compute auto-correlations at different lags. Lag1 and lag2 correlations are found to be 0.9964 and 0.9942 respectively. In the light of these values, a working correlation of AR(2) process is chosen. For marginal AR(1) results see Appendix A.1 Table 1,2,3.

Table 4.5: Residuals of Marginal AR(2) Model

```
Summary of Residuals:
        Min          1Q       Median          3Q          Max
-1.05939180 -0.27558374 -0.04226765  0.17517783  1.43094948
```

We expect the residual values to be between the values of -2 and +2. Table 4.5 indicates that there are no outliers in the residuals because the residuals are between -

1.06 and 1.4. Median is almost zero. Therefore, it can be said that the average of residuals is close to zero.

Table 4.6: Results of Marginal AR(2) Model

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	-2.43788275	0.319886471	-7.621087	0.306751438	-7.947421
TIME	0.03012469	0.002182264	13.804324	0.002465033	12.220805
$\ell n(POP)$	0.92154610	0.047009365	19.603458	0.049391858	18.657854
INDUSTRY	55.23851535	8.828518657	6.256827	13.614898657	4.057211
HOUSEHOLD^5	0.79675168	0.431400110	1.846897	0.557644803	1.428780

Estimated Scale Parameter: 0.167488
Number of Iterations: 7

Intuitively, electricity consumption of provinces is expected to be positively related to time, population of provinces, number of industrial enterprise of provinces and number of household of provinces. Therefore, all independent variable's coefficients of estimate values are expected to be positive. Table 4.6 shows that the coefficients of estimate values are rational. When robust z values are checked, HOUSEHOLD variable seems to be insignificant (p-value = 0.07) at five percent significance level. However; at ten percent significance level, it is significant. Moreover; we consider that HOUSEHOLD is an important covariate and we held it in the model. Also, coefficient of estimate and coefficient of standard error of INDUSTRY variable seems to be high. This might be due to very small observed values for INDUSTRY. The standardization of all variables would solve this problem. However, the interpretation would get harder in this case, since perceiving one unit increase, for instance, in the standardized ln-population is challenging. Therefore, variables are kept unstandardized.

Table 4.7 provides the estimated correlation structure with a marginal AR(2) model. Correlation values are very high, mimicking the correlation structure estimated from real data. Thus, it seems logical to choose correlation structure as AR(2) process.

Table 4.7: Working Correlation of Marginal AR(2) Model

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
[1,]	1.0000000	0.9843627	0.9764379	0.9667510	0.9576032	0.9484343	0.9393793	0.9304045	0.9215169	0.9127138	0.9039950	0.8953594
[2,]	0.9843627	1.0000000	0.9843627	0.9764379	0.9667510	0.9576032	0.9484343	0.9393793	0.9304045	0.9215169	0.9127138	0.9039950
[3,]	0.9764379	0.9843627	1.0000000	0.9843627	0.9764379	0.9667510	0.9576032	0.9484343	0.9393793	0.9304045	0.9215169	0.9127138
[4,]	0.9667510	0.9764379	0.9843627	1.0000000	0.9843627	0.9764379	0.9667510	0.9576032	0.9484343	0.9393793	0.9304045	0.9215169
[5,]	0.9576032	0.9667510	0.9764379	0.9843627	1.0000000	0.9843627	0.9764379	0.9667510	0.9576032	0.9484343	0.9393793	0.9304045
[6,]	0.9484343	0.9576032	0.9667510	0.9764379	0.9843627	1.0000000	0.9843627	0.9764379	0.9667510	0.9576032	0.9484343	0.9393793
[7,]	0.9393793	0.9484343	0.9576032	0.9667510	0.9764379	0.9843627	1.0000000	0.9843627	0.9764379	0.9667510	0.9576032	0.9484343
[8,]	0.9304045	0.9393793	0.9484343	0.9576032	0.9667510	0.9764379	0.9843627	1.0000000	0.9843627	0.9764379	0.9667510	0.9576032
[9,]	0.9215169	0.9304045	0.9393793	0.9484343	0.9576032	0.9667510	0.9764379	0.9843627	1.0000000	0.9843627	0.9764379	0.9667510
[10,]	0.9127138	0.9215169	0.9304045	0.9393793	0.9484343	0.9576032	0.9667510	0.9764379	0.9843627	1.0000000	0.9843627	0.9764379
[11,]	0.9039950	0.9127138	0.9215169	0.9304045	0.9393793	0.9484343	0.9576032	0.9667510	0.9764379	0.9843627	1.0000000	0.9843627
[12,]	0.8953594	0.9039950	0.9127138	0.9215169	0.9304045	0.9393793	0.9484343	0.9576032	0.9667510	0.9764379	0.9843627	1.0000000

According to the empirical result, the proper fixed effects panel model is following;

$$\hat{EC}^{Tr}_{it} = -2.44 + 0.03TIME_{it} + 0.92\ell n(POP)_{it} + 55.24\,INDUSTRY_{it} + 0.8\,HOUSEHOLD^{5}_{it}$$

We take the coefficients of estimate in Table 4.6 while creating the fixed model. Since λ is taken as 0.2 in Box-Cox transformation, to obtain fitted electricity consumption values, we apply back transformation on EC as $EC^{(1/0.2)}$.

According to the fixed model;

A year increase in Time variable causes a 0.03^5 increase on electricity consumption (EC) when all other variables are held constant.

A unit increase in ℓn Population (POP) variable causes a 0.92^5 increase on electricity consumption (EC) when all other variables are held constant.

A unit increase in INDUSTRY variable causes a 55.24^5 increase on electricity consumption (EC) when all other variables are held constant.

Finally, a unit increase in $HOUSEHOLD^5$ variable causes a 0.8^5 increase on electricity consumption (EC) when all other variables are held constant.

Based on our fitted fixed effects panel data model, we can estimate the electricity consumption of provinces. For example, we know that the population of Ankara in 2010 is 4,890,893. The number of industrial enterprise of Ankara in 2010 is 21,728 and finally the number of household of Ankara in 2010 is 1,435,174. When we put these values to the fitted fixed effects model, the electricity consumption of Ankara in 2010 is found to be 9,396,164 MWh.

Variance inflation factor (VIF) values are checked because we assume that there is no multicollinearity problem among independent variables. O'Brien (2007) stated that VIF values help us to detect multicollinearity problem among variables. According to Table 4.8, all the VIF values are smaller than 10 and this indicates that there is no multicollinearity problem.

Table 4.8: Variance inflation factor (VIF) values of Fixed Effects Model

TIME	ℓn (POP)	INDUSTRY	$HOUSEHOLD^5$
1.142	1.377	3.453	2.662

4.4.2. Results of Random Effects Panel Data Model

In order to test time and intercept effects, random model is established with respect to time and intercept respectively as figured in Table 4.9. This new model is given below.

$$EC^{Tr}{}_{it} = \beta_0 + \beta_1 TIME_{it} + \beta_2 \ell n(POP)_{it} + \beta_3 INDUSTRY_{it} + \beta_4 HOUSEHOLD_{it}^5 + \alpha_{0i} + \alpha_{1i} TIME_{it} + u_{it}$$

where;

$t = 0,1,\ldots,11,$

$i = 1,2,\ldots,81,$

$u_{it} \sim \mathrm{N}(0, \Sigma)$.

According to the results, random model with respect to time effect is found significant (p-value<2.2*10^{-16}) and random model of each 81 provinces are constituted separately with respect to time effect (see Appendix A.2 Table 4).

Table 4.9: Results of Random Effects Model

```
Models:
random_intecept: EC ~ TIME + (HOUSEHOLD^5) +ℓn(POP)+ (INDUSTRY) + (1 | ID)
random_Time:     EC ~ TIME + (HOUSEHOLD^5) +ℓn(POP)+ (INDUSTRY) + (TIME | ID)
                 Df     AIC     BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
random_intecept  7 -1155.9 -1121.8 584.95  -1169.9
random_Time      9 -1517.0 -1473.1 767.52  -1535.0 365.13      2  < 2.2e-16 ***

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.8108 -0.4564  0.0092  0.4749  4.1089

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 ID       (Intercept) 0.1656239 0.40697
           TIME       0.0004305 0.02075  -0.28
 Residual             0.0061826 0.07863
Number of obs: 972, groups: ID, 81

Fixed effects:
             Estimate Std. Error t value
(Intercept) -2.387540   0.318981  -7.485
TIME         0.030605   0.002701  11.331
HOUSEHOLD^5  0.732049   0.477154   1.534
ℓn(POP)      0.910594   0.046936  19.401
INDUSTRY    69.395217   9.636166   7.202
```

We expect the residual values between the value -2 and +2. Table 4.9 reveals that there are outliers in the residuals because the residuals are between -5.8 and 4.1. However; median of residuals is almost zero. This can be the effect of the large variation in electricity consumption of provinces. While some them use high electricity such as Istanbul, some of them use very little such as Bayburt.

As stated earlier, electricity consumption of provinces is expected to be positively related to time, population of provinces, number of industrial enterprise of provinces and number of household of provinces. Table 4.9 shows that the coefficients of estimate values are rational.

When *t* values are checked, HOUSEHOLD variable seems to be insignificant (p-value = 0.06) at five percent significance level. However; at ten percent significance level, it is significant. Moreover; we consider that HOUSEHOLD is an important covariate and we held it in the model. Also, coefficient of estimate and coefficient of standard error of INDUSTRY variable seems to be high as in fixed effects model.

Table 4.10: Intercept and time slopes for the first 25 provinces

	(Intercept)	Time
1	0.140357711	-0.0052847719
2	-0.111368207	0.0038080330
3	-0.248231311	0.0010738656
4	-0.480120103	-0.0187037679
5	-0.020604350	-0.0139903975
6	-0.115225133	0.0178643011
7	0.148656604	0.0112231529
8	0.184351022	-0.0088665738
9	-0.267372155	0.0016173465
10	-0.147045376	0.0003390554
11	1.480588357	-0.0622031712
12	-0.432464753	-0.0065785243
13	-0.556385595	0.0069326331
14	0.294214761	-0.0026937666
15	-0.174701229	0.0411046389
16	0.461824608	-0.0023535417
17	0.178284799	0.0745921095
18	0.064510880	-0.0149090327
19	-0.307873434	-0.0052639075
20	0.071718656	-0.0043116790
21	-0.554768615	-0.0058136578
22	0.179105918	-0.0097305464
23	0.120415417	-0.0258894797
24	-0.020427332	-0.0145316538
25	-0.525859230	0.0110619030

Intercept and time slopes of each provinces are obtained and these values are presented in Table 4.10 for the first 25 provinces.

According to the empirical results, the proper random effects panel model for Adana (i=1) is as following;

$$\hat{EC}^{Tr}_{it} = -2.39 + 0.14 + (0.03 - 0.005)TIME_t + 0.91\ell n(POP)_t + 69.39 INDUSTRY_t + 0.73 HOUSEHOLD_t^5$$

We take the estimates of fixed coefficients in Table 4.9 and random coefficients of Adana in Table 4.10 while creating the random model. Since λ is taken as 0.2 in Box-Cox transformation, to obtain fitted electricity consumption values, we apply back transformation on EC as $EC^{(1/0.2)}$.

According to the random model for Adana;

A year increase in Time variable causes a 0.025^5 increase on electricity consumption (EC) when all other variables are held constant.
A unit increase in ℓn Population (POP) variable causes a 0.91^5 increase on electricity consumption (EC) when all other variables are held constant.
A unit increase in INDUSTRY variable causes a 69.39^5 increase on electricity consumption (EC) when all other variables are held constant.
And finally, a unit increase in $HOUSEHOLD^5$ variable causes a 0.73^5 increase on electricity consumption (EC) when all other variables are held constant.

Based on our fitted random effects panel data model, we can estimate the electricity consumption of provinces. . For example, we know that the population of Ankara in 2010 is 4,890,893. The number of industrial enterprise of Ankara in 2010 is 21,728 and finally the number of household of Ankara in 2010 is 1,435,174. When we put these values to the fitted random effects model, the electricity consumption of Ankara in 2010 is found to be 10,223,943 MWh.

4.4.3. Results of Dynamic Panel Data Model (Transition Model)

Finally, dynamic panel data method is implemented to this data. In order to perform dynamic panel data model, we take lag-1 model stated below which satisfies the linearity assumption. As shown below, the model has ylag1 variable. This variable points out to the response at the previous time point.

$$EC^{Tr}{}_{it} = \beta_0 + \alpha_1 EClag_1 + \beta_1 TIME_{it} + \beta_2 \ell n(POP)_{it} + \beta_3 INDUSTRY_{it} + \\ + \beta_4 HOUSEHOLD_{it}^{5} + u_{it}$$

where;

$t = 0,1,\ldots,11,$

$i = 1,2,\ldots,81,$

$u_{it} \sim \mathrm{N}(0,\Sigma)$.

We expect the residual values between the value -2 and +2. Table 4.11 points out that there are no outliers in the residuals because the residuals are between -0.5 and 0.5. The average of residuals is close to zero.

Table 4.11: Residuals of Dynamic Model

Summary of Residuals:

Min	1Q	Median	3Q	Max
-0.51846014	-0.03758863	0.00089233	0.03342151	0.54758761

Again we check all independent variable's coefficients of estimate values and they are found to be positive as expected. When robust z values are checked, TIME and INDUSTRY covariates seem to be insignificant. However; when p-value of these variables are checked, they are found as 0.03 and 0.04 respectively at five percent significance level which reveals that these covariates are significant. Also, it is clear that ylag1 variable is extremely significant.

Table 4.12: Results of Dynamic Model

```
Coefficients:
                  Estimate    Naive S.E.     Naive z  Robust S.E.   Robust z
(Intercept) -0.137595347 0.0395183033  -3.481813 0.0347575800 -3.958715
TIME         0.001693146 0.0009659194   1.752885 0.0009208218  1.838733
ylag1        0.967479524 0.0088652344 109.131861 0.0118044528 81.958863
HOUSEHOLD^5  0.201616168 0.1051216663   1.917932 0.0905678555  2.226134
ℓn(POP)      0.037866094 0.0080443842   4.707146 0.0093956254  4.030183
INDUSTRY     6.097734990 4.4320067713   1.375841 3.5826221236  1.702031

Estimated Scale Parameter:  0.0070239
Number of Iterations:  1
```

According to the empirical result, the proper dynamic panel model is as following;

$$\hat{EC}^{Tr}{}_{it} = -0.14 + 0.002 TIME_t + 0.97 EC_{t-1} + 0.04 \ell n(POP)_t + 6.1\, INDUSTRY_t + 0.2\, HOUSEHOLD^5{}_t$$

We take the coefficients of estimate in Table 4.12 while creating the dynamic model. Since λ is taken as 0.2 in Box-Cox transformation, to obtain fitted electricity consumption values, we apply back transformation on EC as $EC^{(1/0.2)}$.

According to the dynamic model;

A year increase in Time variable causes a 0.002^5 increase on electricity consumption (EC) when all other variables are held constant.
A year increase in ylag1 variable causes a 0.97^5 increase on electricity consumption (EC) when all other variables are held constant.
A unit increase in ℓn Population (POP) variable causes a 0.04^5 increase on electricity consumption (EC) when all other variables are held constant.
A unit increase in INDUSTRY variable causes a 6.1^5 increase on electricity consumption (EC) when all other variables are held constant.
And finally, a unit increase in $HOUSEHOLD^5$ variable causes a 0.2^5 increase on electricity consumption (EC) when all other variables are held constant.

Based on our fitted dynamic panel data model, we can estimate the electricity consumption of provinces. For example, we know that the population of Ankara in 2010 is 4,890,893. The electricity consumption of Ankara in 2009 (ylag1) is 8,611,587 MWh. The number of industrial enterprise of Ankara in 2010 is 21,728 and finally the number of household of Ankara in 2010 is 1,435,174. When we put these values to the fitted dynamic model, the electricity consumption of Ankara in 2010 is found to be 10,252,280 MWh.

Variance inflation factor (VIF) values are again checked for dynamic model. Table 4.13 indicates that all the VIF values except for ylag1 are smaller than 10 and this indicates that there is no multicollinearity problem. Since ylag1 VIF value is approximately equal to 10 and because this variable is of direct interest, it is also included in the analysis.

Table 4.13: Variance inflation factor (VIF) values of Dynamic Model

TIME	ylag1	ℓn (POP)	INDUSTRY	HOUSEHOLD[5]
1.140	10.106	6.585	4.010	2.253

4.5. Model Comparison

After fixed effects, random effects and dynamic panel data analysis are implemented, we have compared the actual values with fitted values of electricity consumption of provinces. In order to compare models, İstanbul, Kocaeli, Bayburt and Mardin provinces are selected. The main reason for choosing these provinces is that İstanbul is the Turkey’s biggest city and consequently is the most electricity consuming province of Turkey. In contrast to İstanbul, Bayburt is the smallest city of Turkey and also is the least electricity consuming province. The median value of the electricity consumption is the Mardin’s consumption. As for Kocaeli, it is the third most electricity consuming province of Turkey despite the fact that the population and the number of household of Kocaeli are not in the top ten. Therefore, the figures of these provinces are thought to give us more enlightening information about the comparison.

Firstly, we checked Turkey's total electricity consumption in order to decide which fitted model is the best. Figure 4.13 displays that random effects panel data model and the dynamic panel data model are the best fitted models for Turkey's aggregate electricity consumption because they seem to be closer to the actual values. However, it is very hard to specify the best model from the figure. Therefore, we need to check model comparison measures of models to determine the best fitted model.

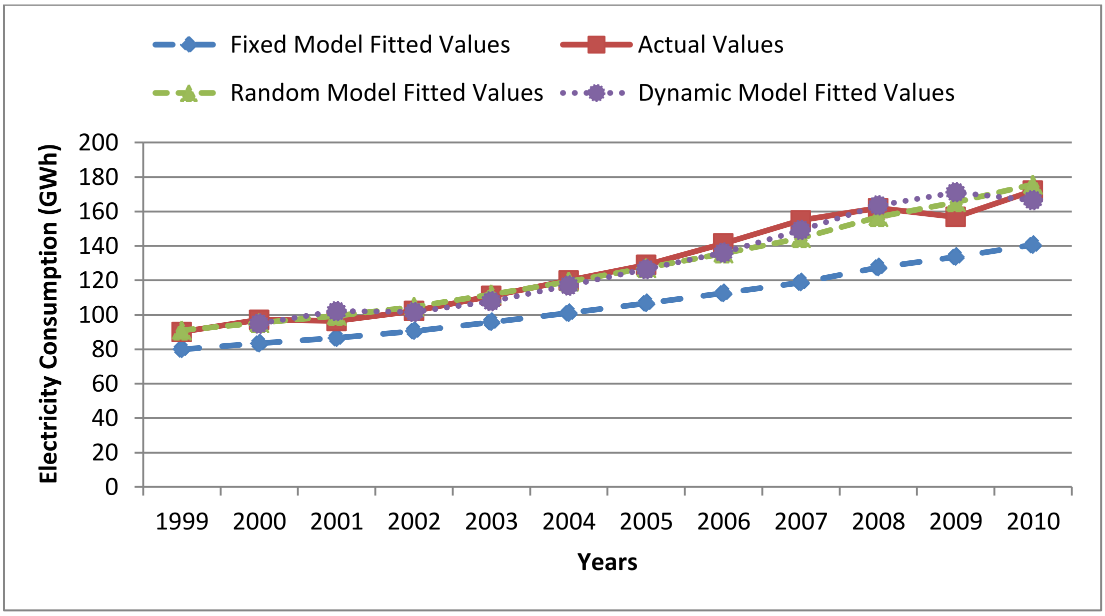


Figure 4.13: Comparison of Models for Turkey

We used Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and correlation value (r) as model comparison measurements. They are obtained as follows;

Root Mean Squared Error (RMSE): $\sqrt{\dfrac{\sum (f_t - a_t)^2}{n}}$

Mean Absolute Percentage Error (MAPE): $\frac{1}{n}\sum\frac{|a_t - f_t|}{a_t}$,

where a_t is the actual value and f_t is the forecast value.

Correlation value (r): Indicates linear relationship between actual electricity consumption values and fitted electricity consumption values.

Table 4.14 indicates that random effects panel data model is the best for Turkey, because RMSE and MAPE are smaller when compared to other model's comparison measures. Although correlation value of fixed effects model and random effects model is very close to each other, plot of fitted values and model comparison measures show that random effects model is the best. As can be seen from table, random effects model test values are close to zero and this indicates reliability of our fitted values.

Table 4.14: Accuracy Results of Models for Turkey

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	1342.14	44.34	0.98
Random Effects Model	205.93	6.96	0.98
Dynamic Model	250.04	7.42	0.97

In addition to comparing the models for Turkey, we also compare the panel data models for different provinces mentioned above. For Bayburt, Figure 4.14 is formed in order to decide which fitted model is the best. This figure indicates that random effects panel data model is the best fitted model for Bayburt because it seems to be closer to the actual values. To demonstrate this, we also check model comparison measures in Table 4.15.

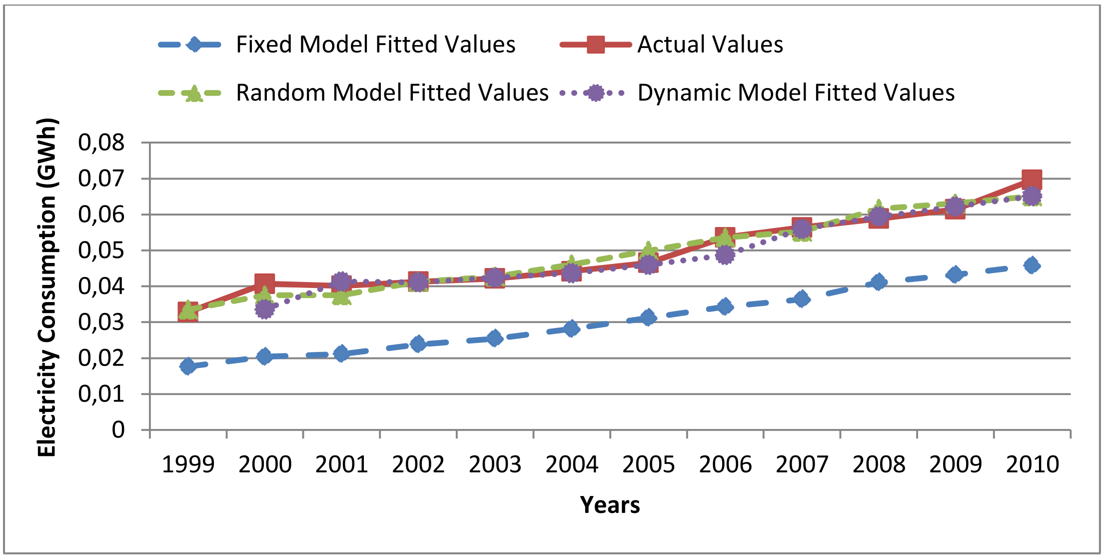


Figure 4.14: Comparison of Models for Bayburt

Table 4.15 shows that random effects panel data model is the best for Bayburt, because model comparison measures are smaller when compared to other model's comparison measures. Also, these accuracy measures show that our fitted values are almost the same as the actual values since they are very close to zero. The reason behind is that Bayburt is the least electricity consuming province of Turkey and it is expected to develop fitted values which are very close to the actual values.

Table 4.15: Accuracy Results of Models for Bayburt

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	18.42	38.37	0.98
Random Effects Model	2.33	3.79	0.97
Dynamic Model	3.06	3.88	0.96

When we observe the model comparison plot of Istanbul in Figure 4.15, we can see that random effects panel data model is again the best fitted model for Istanbul because it seems to be closer to the actual values according to the figure. In order to justify this, we need to check model comparison measures for Istanbul.

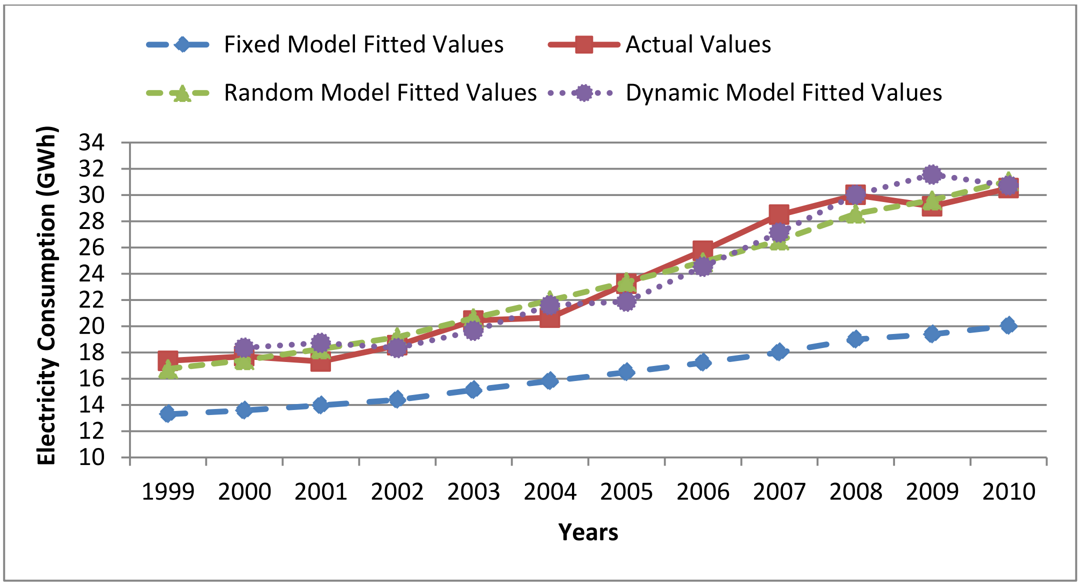


Figure 4.15: Comparison of Models for İstanbul

Table 4.16 shows that random effects panel data model is the best for Istanbul, because model comparison measures are smaller when compared to other model's comparison measures. Here, accuracy measures of Istanbul are not close to zero since it is the most electricity consuming province of Turkey and it is too difficult to develop fitted values which are close to the actual values.

Table 4.16: Accuracy Results of Models for İstanbul

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	7464.66	40.74	0.98
Random Effects Model	943.05	3.36	0.98
Dynamic Model	1169.88	4.16	0.97

Visual comparison of models is shown in Figure 4.16 for Kocaeli. Random effects panel data model and dynamic panel data model again seem to be the best fitted model for Kocaeli based on the figure. Here, fixed effects model highly underestimate the electricity consumption of Kocaeli. This is because Kocaeli is an industry province and its electricity energy consumption is extremely high. Fixed effects panel data model assume that there is a fixed effects of covariates on the response and give same intercept and slope term for each provinces. As stated above Kocaeli is an unusual province and it requires unique slope and intercept. Thus fixed effects panel data model highly underestimate the electricity consumption of Kocaeli. Now, we check the accuracy measures to decide which model is the best for Kocaeli.

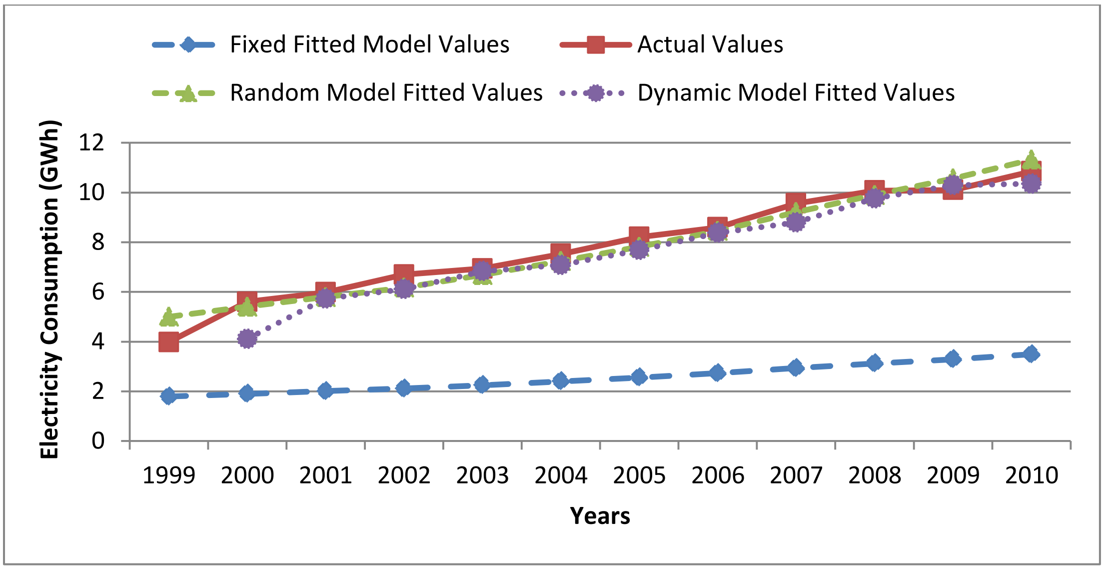


Figure 4.16: Comparison of Models for Kocaeli

Although correlation values of models are very close to each other, prediction plots and other model comparison measures which are shown in Table 4.17 demonstrate that random effects model is again the best fitted model.

Table 4.17: Accuracy Results of Models for Kocaeli

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	5498.23	66.86	0.97
Random Effects Model	438.16	5.68	0.97
Dynamic Model	613.25	7.84	0.98

In order to make more reliable comments, we also take Mardin which is median province in the list of the electricity consuming provinces. Figure 4.17 points out as demonstrated above examples that random effects model and the dynamic panel data model are best fitted models for Mardin because they seem to be closer to the actual values. It is very hard to specify the best model from the figure. Therefore, we need to check model comparison measures in order to decide which model is the best.

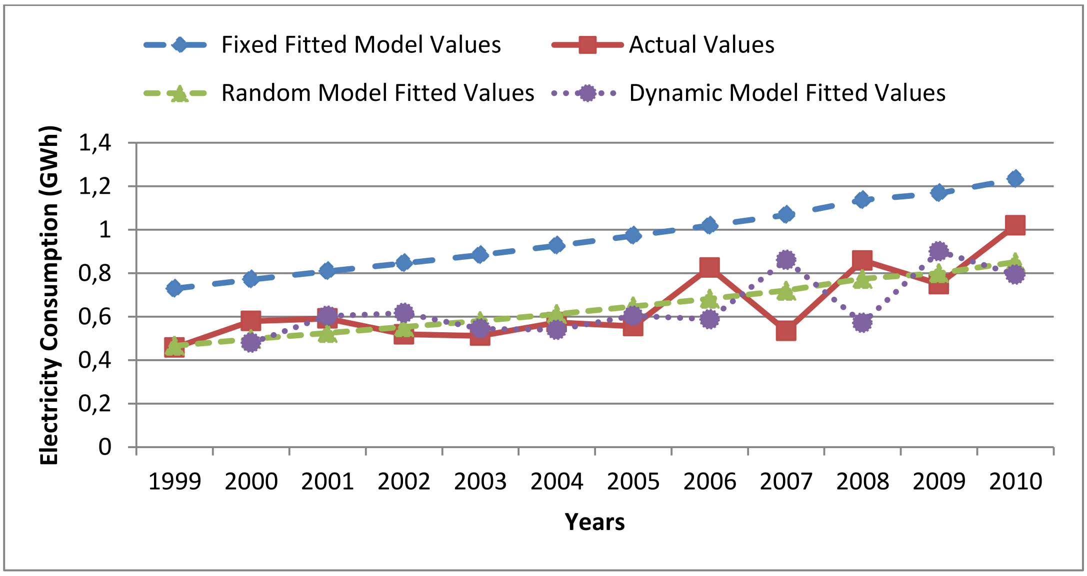


Figure 4.17: Comparison of Models for Mardin

Table 4.18 shows that random effects model is again the best fitted model. As can be seen from the below table, random effects model comparison measures are close to zero and this indicates reliability of our fitted values.

Table 4.18: Accuracy Results of Models for Mardin

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	332.02	52.82	0.79
Random Effects Model	99.82	12.93	0.81
Dynamic Model	176.71	20.41	0.31

Both the electricity consumption models of selected provinces Bayburt, İstanbul, Kocaeli and Mardin and electricity consumption models of Turkey state that random effects panel data analysis is the best forecasting model compared to fixed effects panel data analysis and dynamic panel data analysis model and it provides the most accurate results. Dynamic panel data model is also relatively reliable model and it also provides accurate results. Fixed effects panel data model is not reliable model and it's results don't overlap the actual result. We conclude from the model comparison analysis that fixed effects panel data model underestimate the electricity consumption of developed provinces such as İstanbul, Kocaeli, İzmir and Ankara and the developed geographical regions such as Marmara, Aegean, Mediterranean and Central Anatolian while it is overestimating the electricity consumption of least developed provinces such as Bayburt, Şırnak, Hakkari and the least developed geographical regions such as Southeastern Anatolia, Eastern Anatolia and Black Sea. Because it assumes that there is a fixed effect of covariates on the response and give same intercept and slope term for each province while random effects panel model gives different intercept and random slope for each province.

4.6. Electricity Leakage and Losses in Turkey

As mentioned before, when the proposed models compared for both provinces and geographical regions of Turkey, it is clearly seen that fixed effects panel data model overestimates the eastern and southeastern region of Turkey and their provinces while fixed effects panel data model is underestimating or nearly giving close estimates for the western and central region of Turkey and their provinces. Figure 4.18 and Figure 4.19 show the plots of fitted values and actual values for the Southeastern and Eastern Anatolian Regions, respectively. Logically, fixed effects panel data model is expected to overestimate the northern, eastern and southeastern region of Turkey because these regions have least developed provinces and consequently consumes less electricity. According to these graph, fixed effects panel data model highly overestimate the electricity consumption values while random effects model and dynamic model estimate it properly. This situation brought doubt us that the eastern and southeastern regions of Turkey could have high electricity leakage.

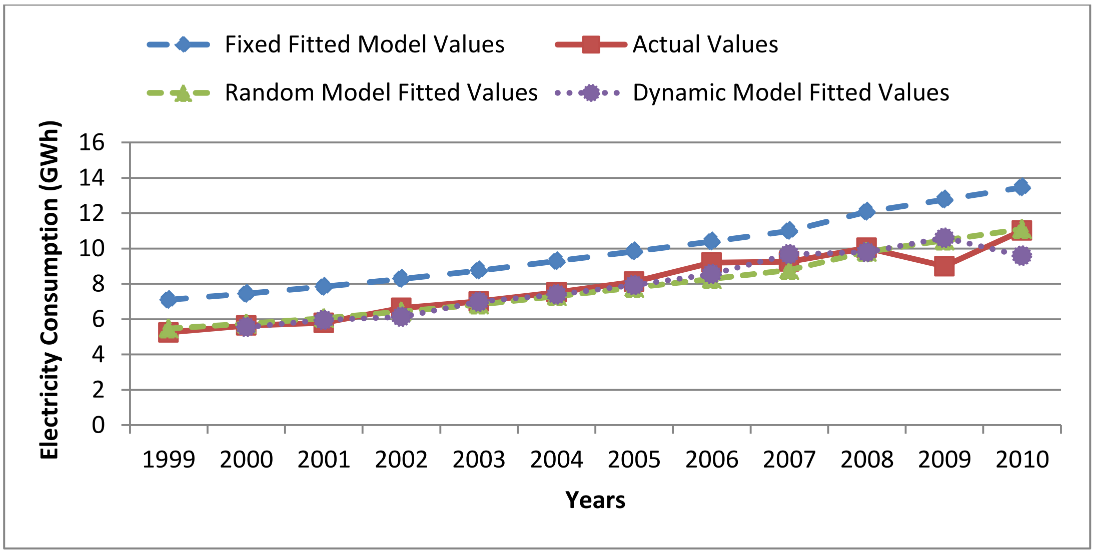


Figure 4.18: Comparison of Models for Southeastern Anatolia Region

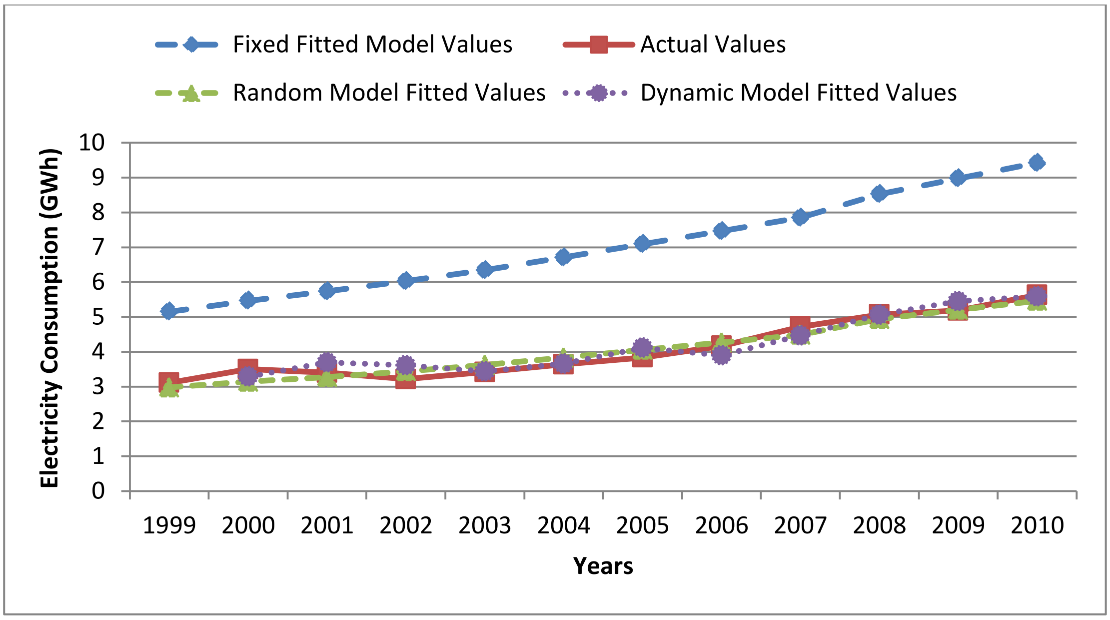


Figure 4.19: Comparison of Models for Eastern Anatolia Region

As stated in the methodology part, fixed effects panel data takes into account of all provinces' independent variables, that is, population, number of household and number of industrial enterprise of provinces, and assumes that there is a fixed effect on our dependent variable electricity consumption. However, random effects panel data analysis develops the model assuming that the differences across entities have some impact on our dependent variable. So, it provides different intercept and different slope for each entity. The reason behind the fixed effects panel data model being highly overestimating the electricity consumption in the eastern and southeastern regions might be due to high illegal usage of electricity in these provinces. Their official values are lower than the actual ones because of the high electricity leakage. Turkish Electricity Transmission Company (TEIAS) has announced the losses and leakage rates of Turkey's provinces in 2012 and it is shown in Table 4.19. Actually, we have losses and leakage rates of Turkey's provinces only for the years 2009 and 2012. Since data of 2012 is closer to our dataset, we prefer to use it. Table 4.19 verifies our doubt. As can be clearly seen in the table, leakage rates of eastern and southeastern region provinces are very high. Only the provinces whose leakage rate is greater than 10 are shown in the table.

Table 4.19: Losses and leakage rates of Turkey's provinces in 2012 (Source TEİAŞ)

Rank	Province	Leakage Rate (%)
1	ŞIRNAK	78.62
2	MARDİN	76.02
3	DİYARBAKIR	73.27
4	HAKKARİ	70.85
5	BATMAN	69.59
6	Ş.URFA	63.62
7	AĞRI	61.98
8	MUŞ	54.11
9	VAN	49.84
10	BİTLİS	45.68
11	SİİRT	41.42
12	IĞDIR	35.72
13	BİNGÖL	29.50
14	KARS	21.72
15	HATAY	15.37
16	BAYBURT	14.12
17	GİRESUN	13.64
18	ZONGULDAK	13.21
19	GAZİANTEP	13.18
20	TUNCELİ	13.17
21	MERSİN	11.78
22	ARTVİN	11.75

Table 4.19 (continued)

23	MALATYA	11.67
24	KASTAMONU	11.54
25	ERZURUM	11.43
26	YALOVA	11.30
27	ADANA	11.20
28	SAKARYA	11.05
29	ADIYAMAN	10.89
30	NİĞDE	10.60
31	KARAMAN	10.49
32	ANTALYA	10.38
33	AKSARAY	10.12
34	TOKAT	10.07

According to the table, Şırnak, Mardin, Diyarbakır and Hakkari are the provinces with the highest electricity leakage. When the plot of fitted values for these provinces is analyzed, we can see that fixed effects panel data model is highly overestimating the electricity consumption of these provinces as illustrated in Figures 4.20, 4.21 and 4.22.

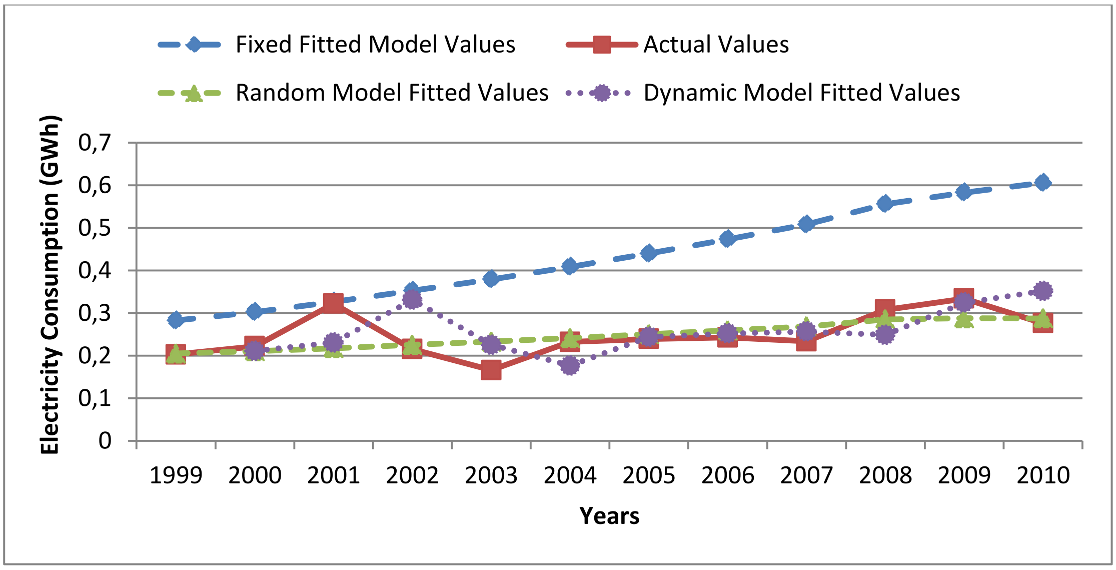


Figure 4.20: Comparison of Models for Şırnak

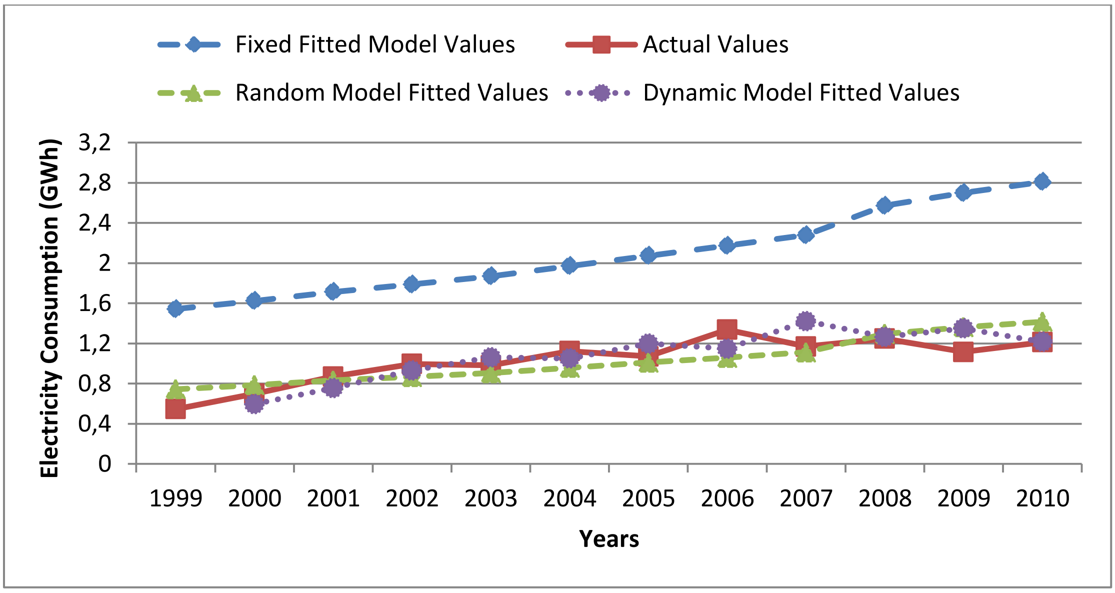


Figure 4.21: Comparison of Models for Diyarbakır

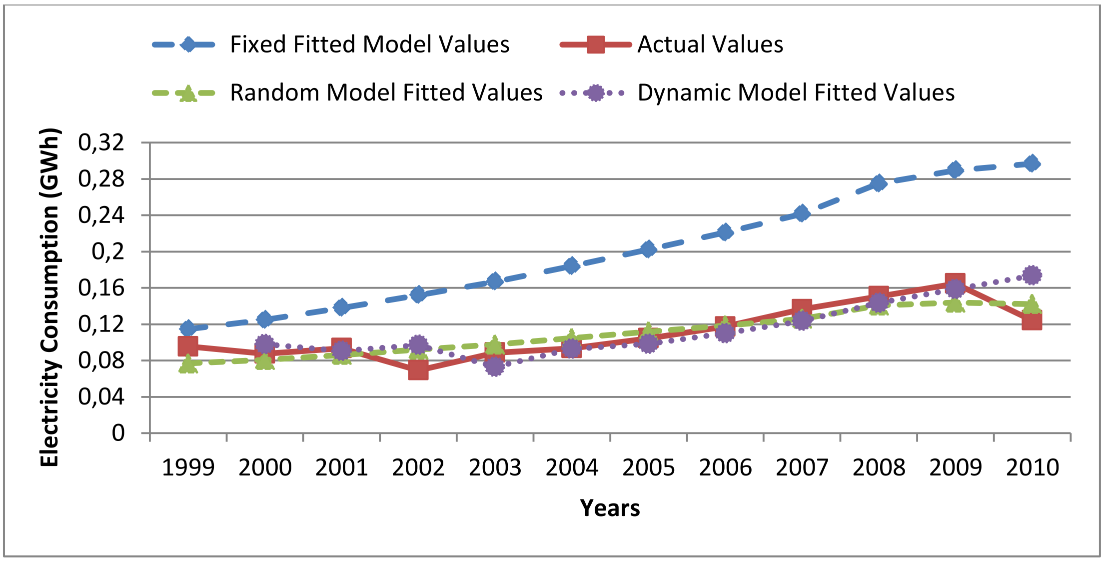


Figure 4.22: Comparison of Models for Hakkari

We found that fixed effects panel data model overestimated all the provinces of Southeastern Anatolia Region except for Kilis. Figure 4.23 reveals the graph of fitted values for each model for Kilis. It is seen that fixed effects panel data model underestimates the electricity consumption of Kilis. The leakage rate of Kilis in 2012 is 7.19 which is very low when compared to the other provinces of the region.

Therefore, fixed effects panel data model does not overestimate the electricity consumption of Kilis.

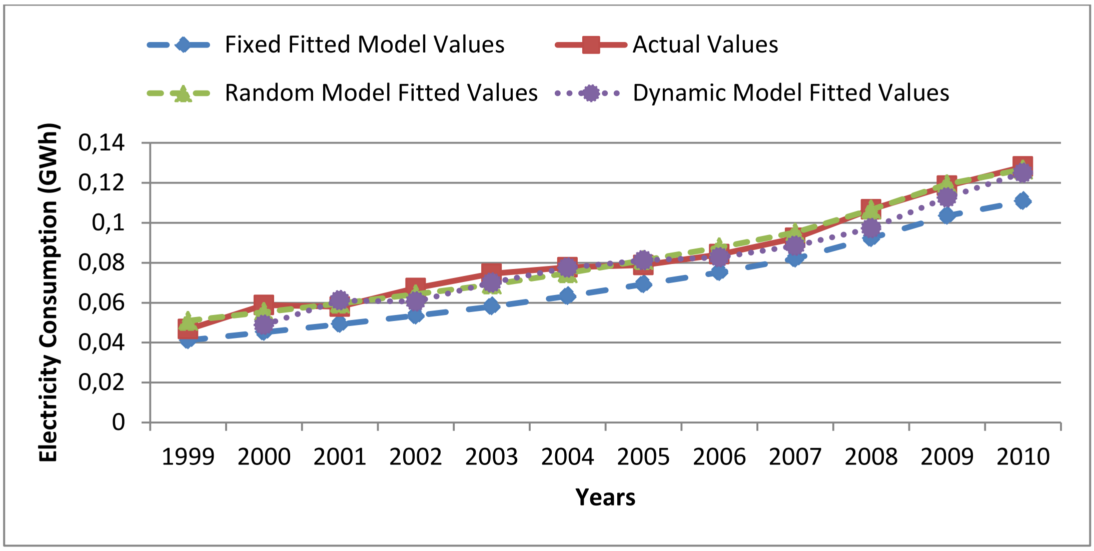


Figure 4.23: Comparison of Models for Kilis

Again we found that fixed effects panel data model overestimated all the provinces of Eastern Anatolia Region except for Tunceli. Figure 4.24 shows the graph of prediction values for each model for Tunceli. It is seen that fixed effects panel data model underestimates the electricity consumption of Tunceli. The leakage rate of Tunceli in 2012 is 13.17 which is very low when compared to the other provinces of the region. Therefore, fixed effects panel data model does not overestimate the electricity consumption of Tunceli.

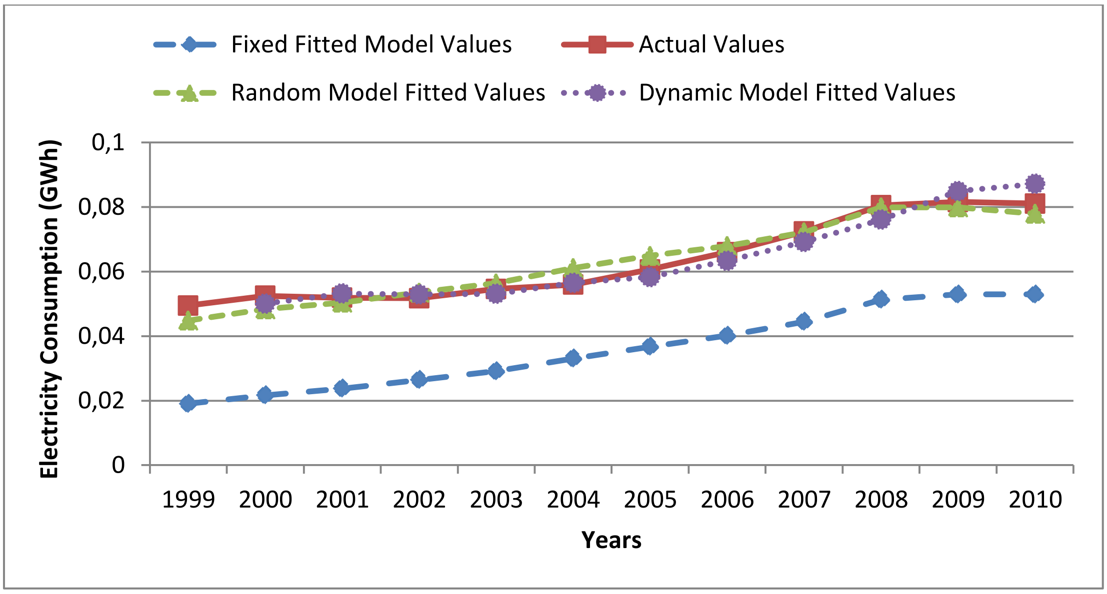


Figure 4.24: Comparison of Models for Tunceli

4.6.1. Analysis and Discussion on Electricity Leakage

In this thesis study, we try to make the best forecasting using the official data. We have realized that the eastern and southeastern regions of Turkey and their provinces have high electricity leakage and official data don't include these electricity leakage amounts. It is understood from Section 4.5, the fixed effects panel data analysis gives close values to the original data not the official one. Therefore; taking the electricity leakage rates (ELR) into account, we try to remodel the data. We create dummy variables, namely X5 and X6, and include these to the model in addition to the existing independent variables. The new variables are specified as the following;

$$\begin{pmatrix} \mathit{If}\ 40 < \mathrm{ELR} < 100 \\ \mathit{If}\ 10 < \mathrm{ELR} \leq 40 \\ \mathit{otherwise} \end{pmatrix} \quad \text{we set} \quad \begin{bmatrix} \underline{X5} & \underline{X6} \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

where ELR denotes the electricity leakage rates of provinces.

Table 4.20: A Portion of X5 and X6 Variables

Province	**Leakage Rate (%)**	**X5**	**X6**
ADANA	11.20	0	1
ADIYAMAN	10.89	0	1
AFYON	9.36	0	0
AĞRI	61.98	1	0
AKSARAY	10.12	0	1
AMASYA	8.48	0	0
ANKARA	8.23	0	0
ANTALYA	10.38	0	1
ARDAHAN	7.94	0	0
ARTVİN	11.75	0	1
AYDIN	9.37	0	0
BALIKESİR	8.87	0	0
BARTIN	8.98	0	0
BATMAN	69.59	1	0
BAYBURT	14.12	0	1
BİLECİK	3.96	0	0
BİNGÖL	29.50	0	1
BİTLİS	45.68	1	0
BOLU	4.91	0	0
BURDUR	8.08	0	0
BURSA	6.56	0	0
ÇANAKKALE	5.85	0	0
ÇANKIRI	7.63	0	0
ÇORUM	5.66	0	0
DENİZLİ	5.97	0	0
DİYARBAKIR	73.27	1	0
DÜZCE	6.98	0	0
EDİRNE	7.31	0	0

As can be seen from Table 4.20, X5 and X6 variables are created for the provinces within the framework of the given conditions. After creating new independent variables, the new fixed effects panel data model is fitted as illustrated below;

$$EC^{Tr}{}_{it} = \beta_0 + \beta_1 TIME_{it} + \beta_2 \ell n(POP)_{it} + \beta_3 INDUSTRY_{it} + \beta_4 HOUSEHOLD^5 \\ + \beta_5 X5_{it} + \beta_6 X6_{it} + u_{it}$$

where;

$t = 0,1,\ldots,11,$

$i = 1,2,\ldots,81,$

$u_{it} \sim \mathrm{N}(0, \Sigma)$.

By using this new fixed effects panel data model, we obtained the electricity consumptions of provinces taking electricity leakage into account. Based on the new fixed effects panel data electricity consumptions, model comparison graph and model comparison measures of provinces and regions are constructed again.

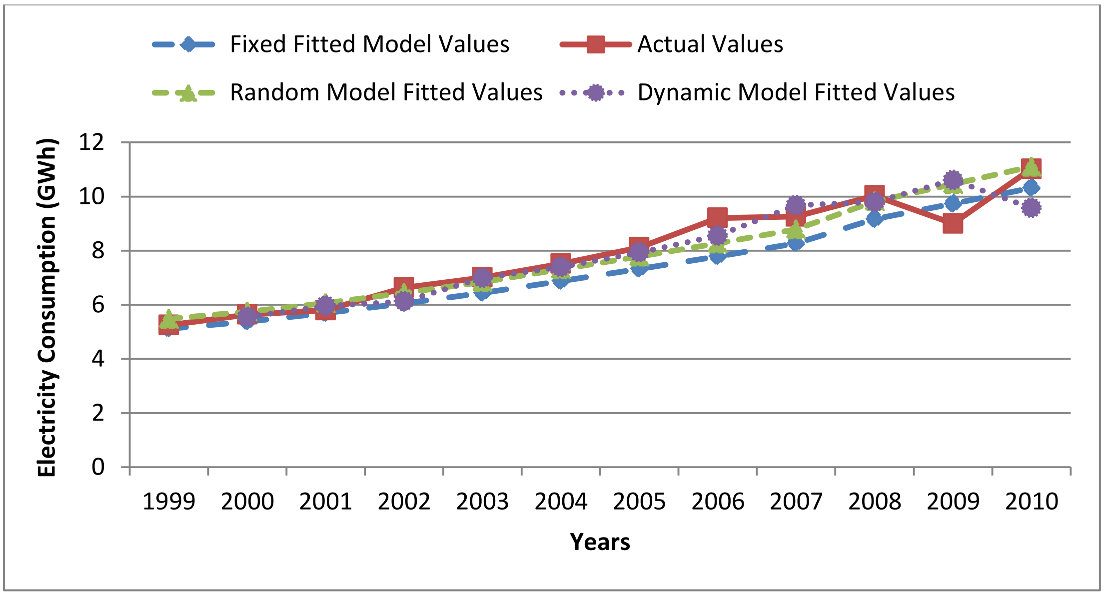


Figure 4.25: Comparison of Models for Southeastern Anatolia Region Using New Fixed Model

Figure 4.25 reveals the model comparison graph of Southeastern Anatolia Region using new fixed effects panel data model. It is seen that fixed effects panel data model seem to be close to the actual consumption values and don't overestimate the electricity consumptions. As it is remembered from Figure 4.18 that when electricity leakage of provinces are not considered, the fixed effects panel data model highly overestimated the electricity consumptions of Southeastern Anatolia Region. However; we can see that the fixed effects panel data model seems to be close to the actual consumption values when electricity leakage of the provinces are considered.

To determine which model is the best, comparison measures of models are checked in Table 4.21.

Table 4.21: Accuracy Results of Models for Southeastern Anatolia Region Using New Fixed Model

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	740.74	7.77	0.95
Random Effects Model	552.06	4.81	0.95
Dynamic Model	718.06	5.60	0.91

Based on the table, random effects panel data model is still the best fitted model, because model comparison measures RMSE and MAPE are smaller when compared to other model's comparison measures.

Figure 4.26 points out the model comparison graph of Eastern Anatolia Region using new fixed effects panel data model. In the figure, new fixed effects panel data model does not seem to be close to the actual consumption values and overestimate the electricity consumptions. As it is remembered from the Figure 4.19 that when the electricity leakage of provinces are not considered, the fixed effects panel data model highly overestimates the electricity consumptions of Eastern Anatolia Region. Although we have considered the electricity leakage of the provinces, the fixed effects panel data model does not produce accurate result.

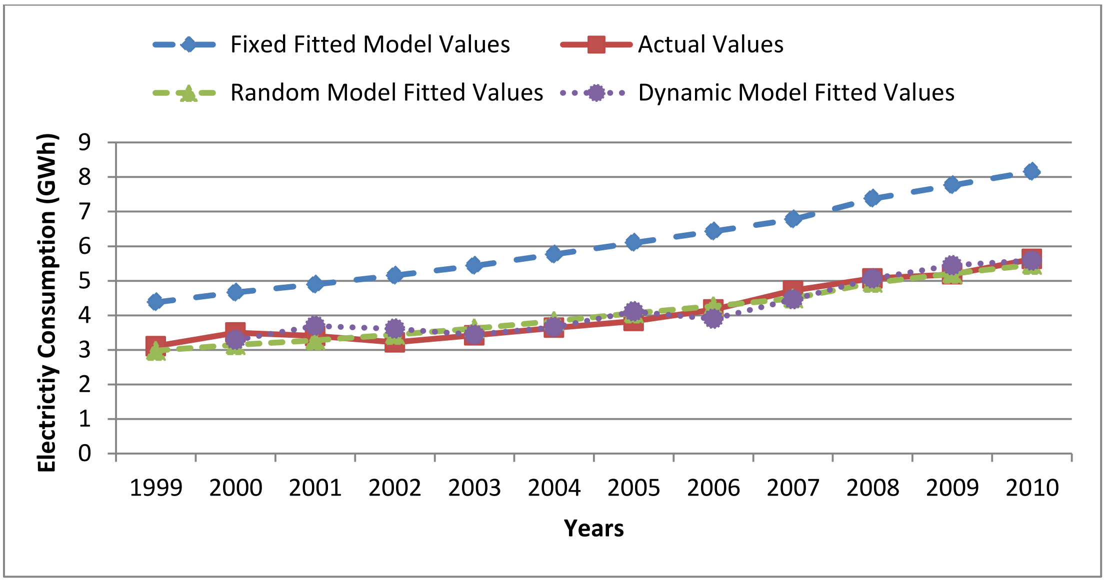


Figure 4.26: Comparison of Models for Eastern Anatolia Region Using New Fixed Model

Model comparison measures in Table 4.22 also reveals that fitted electricity consumption values of new fixed effects panel data model are not close to the actual consumption values and random effects panel data model is the best fitted model again.

Table 4.22 Accuracy Results of Models for Eastern Anatolia Region Using New Fixed Model

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	2049.51	49.42	0.96
Random Effects Model	191.92	4.58	0.97
Dynamic Model	228.31	4.82	0.96

As it can be seen from the above part, Şırnak, Diyarbakır and Hakkari are the provinces with the highest electricity leakage. When figures for these provinces were analyzed, it was seen that the fixed effects panel data model highly overestimates the electricity consumption of this provinces. We compare the models of these provinces again while taking the electricity leakage of these provinces into account. Using the new fixed effects panel data model, Hakkari, Şırnak and Diyarbakır's graphs are shown in Figure 4.27, 4.28 and 4.29, respectively.

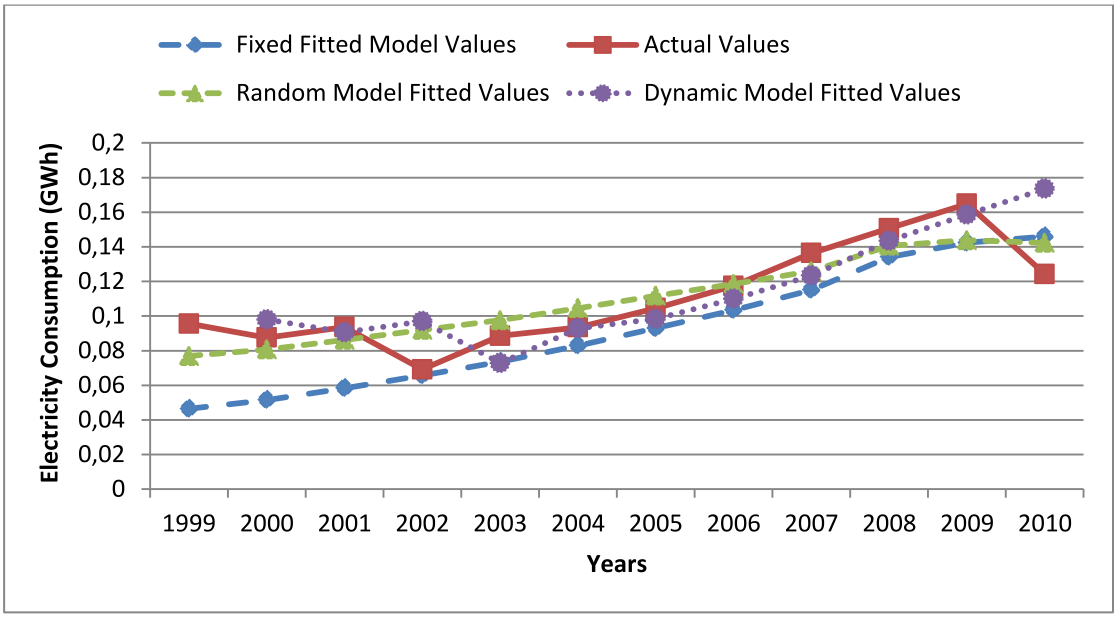


Figure 4.27: Comparison of Models for Hakkari Using New Fixed Model

Figure 4.27 shows that all the models are close to the actual values. When model comparison measures are checked in Table 4.23, it is seen that random effects model is the best model.

Table 4.23 Accuracy Results of Models for Hakkari Using New Fixed Model

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	24.74	20.35	0.86
Random Effects Model	13.56	11.66	0.87
Dynamic Model	18.88	13.08	0.79

For Şırnak, the new fixed effects model and the other models are very close to actual electricity consumption values as illustrated in Figure 4.28.

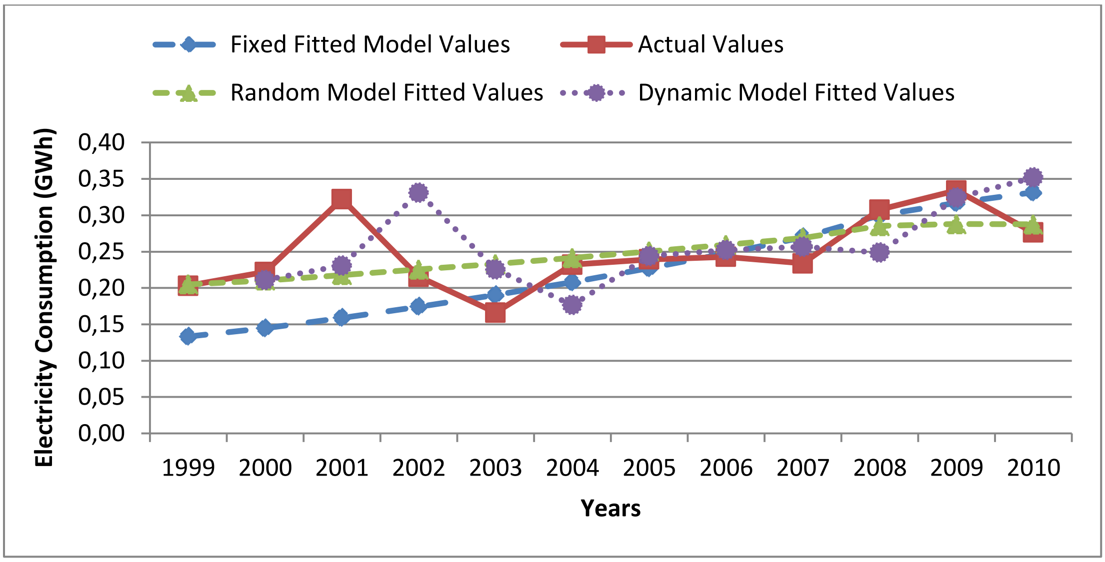


Figure 4.28: Comparison of Models for Şırnak Using New Fixed Model

When model comparison measures are checked in Table 4.24, we figure out that fitted electricity consumption values of new fixed effects panel data model are close to the actual consumption values, but random effects model is the best fitted model.

Table 4.24 Accuracy Results of Models for Şırnak Using New Fixed Model

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	61.42	17.83	0.54
Random Effects Model	40.94	11.59	0.54
Dynamic Model	59.10	19.24	0.31

The new fixed effects model is also close to actual electricity consumption values as illustrated in Figure 4.29.

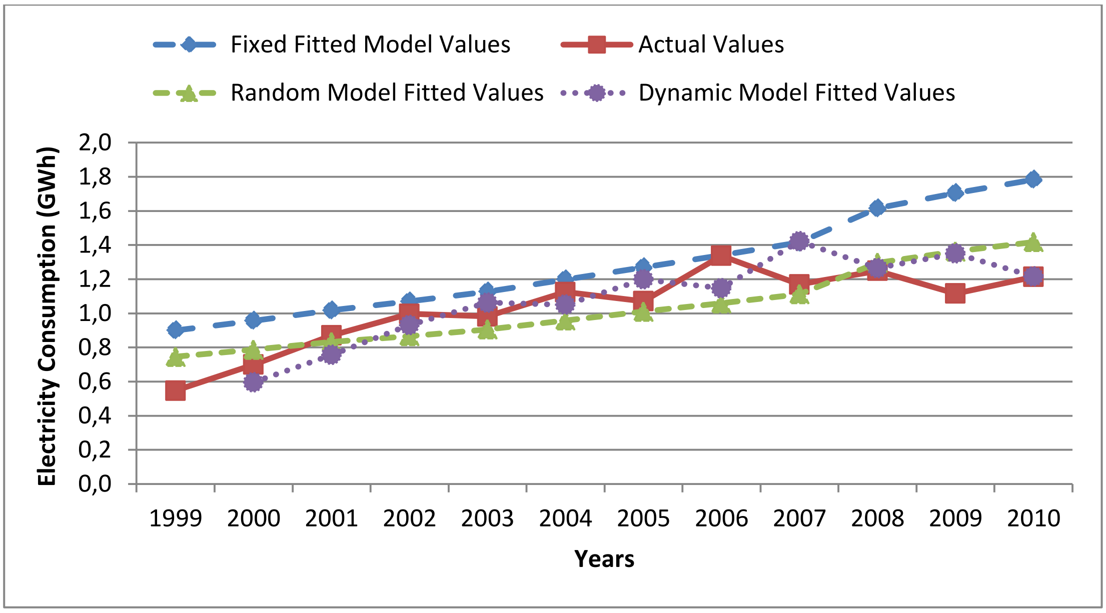


Figure 4.29: Comparison of Models for Diyarbakır Using New Fixed Model

When model comparison measures are checked in Table 4.25, we figure out that fitted electricity consumption values of new fixed effects panel data model is slightly close to the actual consumption values, and again random effects model is the best fitted model.

Table 4.25 Accuracy Results of Models for Diyarbakır Using New Fixed Model

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	309.82	26.31	0.77
Random Effects Model	154.85	13.55	0.75
Dynamic Model	138.72	10.87	0.82

As can be seen form these figures, the new fixed effects panel data model seems to be close to the actual consumption values and don't overestimate the electricity consumptions. When the electricity leakages of these provinces were not considered, fixed effects panel data model highly overestimated the electricity consumptions of these provinces. However; we can see that the fixed effects panel data model seems to be close to the actual consumption values when electricity leakage of the provinces are considered.

Again as it can be seen from the above part, the fixed effects panel data model overestimates all the provinces of Southeastern and Eastern Anatolia Region except for the provinces Tunceli and Kilis. The fixed effects panel data model underestimates these provinces electricity consumptions because these provinces have lower leakage rates when compared to the other provinces of the regions. Figure 4.30 reveals that the new fixed effects panel data model of Kilis seems to be close to the actual consumption values when electricity leakages of the provinces are considered.

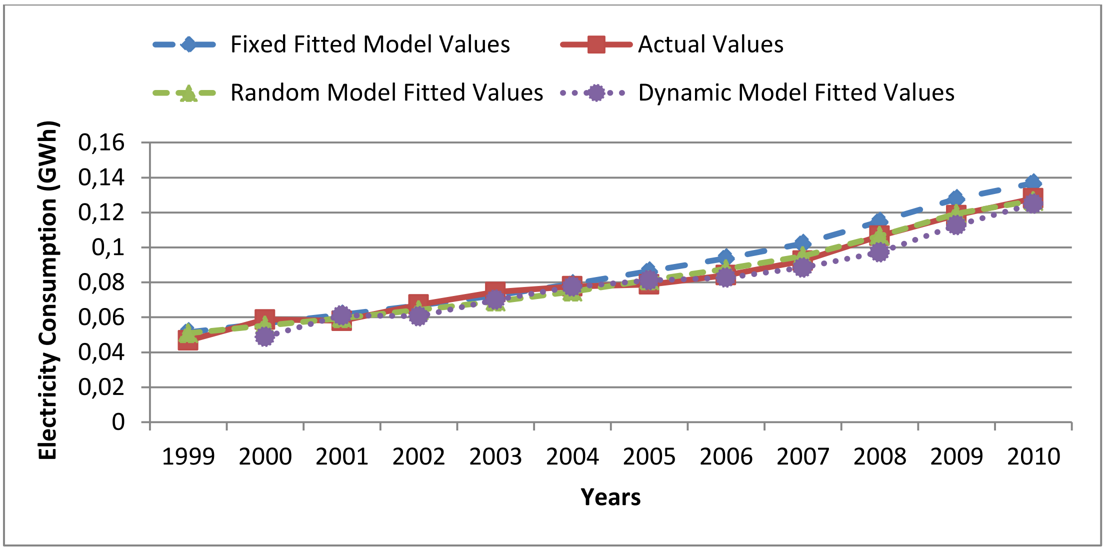


Figure 4.30: Comparison of Models for Kilis Using New Fixed Model

In order to demonstrate this, model comparison measures are checked for Kilis in Table 4.26. According to the table, all the fitted models are reliable, but random effects model is the best fitted model.

Table 4.26 Accuracy Results of Models for Kilis Using New Fixed Model

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	6.59	6.57	0.99
Random Effects Model	2.99	3.78	0.99
Dynamic Model	5.46	5.77	0.98

However, Figure 4.31 shows that although we have considered the electricity leakage of the Tunceli, the new fixed effects panel data model still underestimates the electricity consumptions.

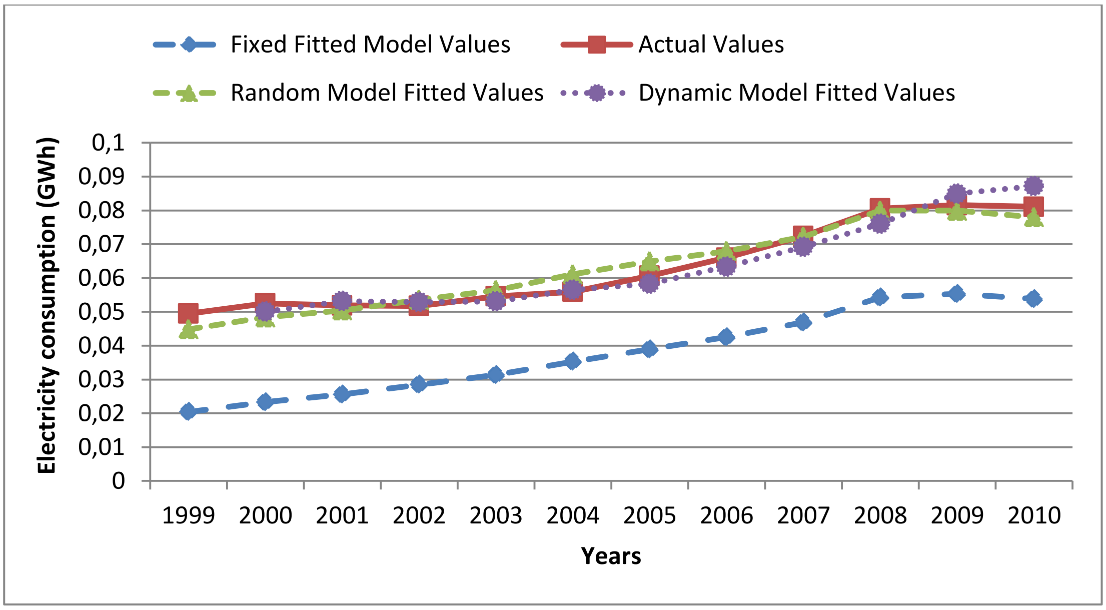


Figure 4.31: Comparison of Models for Tunceli Using New Fixed Model

When model comparison measures are checked in Table 4.27, the new fixed effects panel data model does not produce accurate results while random effects and dynamic fitted models are very compatible to the actual values.

Table 4.27 Accuracy Results of Models for Tunceli Using New Fixed Model

	RMSE	MAPE	Correlation (r)
Fixed Effects Model	25.31	41.20	0.97
Random Effects Model	2.99	4.38	0.97
Dynamic Model	3.07	3.90	0.97

We can conclude from the discussion part that remodeling fixed effects panel data taking the electricity leakage into account provides more accurate results and seems to be closer the actual values. However, it is again seen that the random effects panel data model is the best fitted model and therefore we use this model for forecasting.

If we detect the real net electricity consumption of Turkey, leakage rates of each province should be added to the official consumption values of each province and obtained the best fitted model. However; this thesis study aims to model electricity consumption of Turkey and provinces in the light of official data. Moreover, the leakage records are not available for each year.

4.7. Forecast Results

In this thesis study, exponential smoothing method (ETS) is used as forecasting method to obtain the forecasts of covariates using the data between 1999 and 2010. Using ETS method, we get the forecasted values of population of provinces, total number of industrial enterprise of provinces and total number of household of provinces between the years 2011 and 2015. Using these forecasts, total electricity energy consumption of Turkey and its provinces are created until 2015 with the best fitted model which is the random effects panel data model.

As stated at the beginning of the thesis, the last year of the data (2011) was allocated for forecast evaluation. We compare this actual electricity energy consumption of 2011 with the forecasted electricity energy consumption of 2011 to understand the efficiency of our forecasting results.

Table 4.28: Comparison of Actual Values with Forecasted Values for Some Provinces

Province Name	2011 Forecast Values (MWh)	2011 Actual Values (MWh)	Difference (MWh)
Nevşehir	579,334	579,221	112
Amasya	495,422	493,593	1,829
Zonguldak	2,665,009	2,656,927	8,082
Muğla	2,184,285	2,208,244	23,959
Malatya	1,231,864	1,185,758	46,106
Denizli	2,695,385	2,627,401	67,983
Giresun	443,378	515,382	72,004
Sakarya	2,274,486	2,200,799	73,687
Sivas	1,227,808	1,152,140	75,667
Ankara	10,132,942	10,259,178	126,236
Bursa	9,153,961	8,975,143	178,817
Diyarbakır	1,513,114	1,327,781	185,333
İstanbul	32,474,364	32,672,285	197,921
Samsun	2,189,808	2,435,978	246,170
İzmir	16,730,829	16,442,561	288,269
Kocaeli	12,067,627	11,500,479	567,147
TURKEY	**184,845,601**	**186,099,551**	**1,253,950**

According to Table 4.28, total electricity consumption of Turkey in 2011 is 186,099,551 MWh while it is forecasted as 184,845,601 MWh with the random effects panel data model. This indicates that the random effects model is an accurate fitted model and compatible with the official values. Generally; the difference between forecasted values and actual values of provinces are very close to each other except for Kocaeli. As stated earlier, Kocaeli is an unusual province because it is an industry province thus consumes electricity energy extremely high.

We also forecasted the total electricity energy consumption of Turkey and its provinces using the random effects panel data fitted model until 2015. Table 4.29 points out the forecasted total electricity energy consumption values of Turkey and its provinces until 2015. Turkish Electricity Transmission Company (TEIAS) has announced the total net electricity energy consumption of Turkey in 2012 as 194,923,349 MWh while it is forecasted as 194,714,839 MWh with our forecast. The difference between actual total electricity consumption and forecasted total electricity consumption of Turkey in 2012 is just 208,510 MWh. This is also shows the efficiency of our forecasting.

Table 4.29: Forecast Results with Random Effects Model

Province Name	**2011 Forecast Values (MWh)**	**2012 Forecast Values (MWh)**	**2013 Forecast Values (MWh)**	**2014 Forecast Values (MWh)**	**2015 Forecast Values (MWh)**
Adana	4,479,194	4,626,088	4,776,383	4,930,130	5,087,378
Adıyaman	979,316	1,029,165	1,081,029	1,134,971	1,191,055
Afyon	1,154,867	1,207,141	1,261,470	1,317,927	1,376,589
Ağrı	308,981	318,125	327,466	337,006	346,749
Amasya	495,422	510,047	525,064	540,482	556,313
Ankara	10,132,942	10,724,483	11,344,080	11,992,780	12,671,670
Antalya	5,487,619	5,853,188	6,237,988	6,642,762	7,068,282
Artvin	278,532	290,816	303,587	316,861	330,659
Aydın	1,822,292	1,914,548	2,010,267	2,109,538	2,212,446
Balıkesir	2,424,857	2,530,349	2,639,339	2,751,907	2,868,138
Bilecik	1,001,209	962,110	924,243	887,578	852,085
Bingöl	149,960	160,016	170,794	182,368	194,827
Bitlis	231,542	247,701	264,666	282,475	301,164
Bolu	873,017	915,987	960,313	1,006,043	1,053,223
Burdur	881,558	966,046	1,056,890	1,154,444	1,259,072
Bursa	9,153,961	9,517,495	9,892,491	10,279,211	10,677,934
Çanakkale	3,794,740	4,229,710	4,703,735	5,219,397	5,779,394
Çankırı	274,467	286,596	299,096	311,976	325,241
Çorum	733,773	768,196	804,317	842,238	882,067
Denizli	2,695,385	2,803,528	2,915,243	3,030,632	3,149,802
Diyarbakır	1,513,114	1,625,306	1,751,018	1,892,929	2,054,398
Edirne	1,051,364	1,092,844	1,136,264	1,181,742	1,229,409
Elazığ	885,152	905,424	926,324	947,884	970,138
Erzincan	257,431	264,320	271,357	278,542	285,879
Erzurum	961,055	1,018,288	1,078,375	1,141,435	1,207,595
Eskişehir	2,251,694	2,374,833	2,503,302	2,637,272	2,776,917
Gaziantep	4,949,936	5,268,532	5,603,593	5,955,766	6,325,712
Giresun	443,378	450,593	457,919	465,357	472,910
Gümüşhane	128,179	134,374	140,826	147,545	154,542
Hakkari	151,241	160,585	170,737	181,855	194,134

Table 4.29 (continued)

Hatay	5,300,302	5,726,997	6,184,122	6,673,819	7,198,410
Isparta	1,035,651	1,060,146	1,085,103	1,110,527	1,136,426
Mersin	3,198,819	3,326,668	3,458,572	3,594,629	3,734,931
İstanbul	32,474,364	33,994,824	35,571,721	37,206,609	38,901,052
İzmir	16,730,829	17,609,807	18,520,904	19,464,879	20,442,511
Kars	287,050	302,573	318,791	335,730	353,417
Kastamonu	660,611	687,344	714,936	743,406	772,776
Kayseri	2,988,694	3,131,217	3,279,184	3,432,760	3,592,105
Kırklareli	1,882,339	1,943,337	2,006,052	2,070,526	2,136,805
Kırşehir	353,386	370,714	388,869	407,892	427,829
Kocaeli	12,067,627	12,845,934	13,663,894	14,522,997	15,424,785
Konya	5,110,409	5,341,034	5,579,637	5,826,408	6,081,548
Kütahya	1,167,387	1,230,872	1,297,208	1,366,498	1,438,852
Malatya	1,231,864	1,285,184	1,340,216	1,396,997	1,455,565
Manisa	3,040,407	3,211,013	3,389,263	3,575,417	3,769,741
K.Maraş	3,173,576	3,363,179	3,561,734	3,769,559	3,986,974
Mardin	886,504	923,016	960,722	999,650	1,039,829
Muğla	2,184,285	2,266,029	2,350,324	2,437,235	2,526,832
Muş	282,367	299,261	316,893	335,297	354,503
Nevşehir	579,334	584,181	589,086	594,049	599,072
Niğde	884,052	924,727	966,920	1,010,679	1,056,049
Ordu	976,480	1,030,190	1,086,160	1,144,476	1,205,229
Rize	624,177	653,935	684,801	716,806	749,979
Sakarya	2,274,486	2,447,899	2,631,731	2,826,446	3,032,518
Samsun	2,189,808	2,320,246	2,456,987	2,600,274	2,750,352
Siirt	340,910	354,868	369,215	383,956	399,098
Sinop	270,923	285,380	300,466	316,202	332,610
Sivas	1,227,808	1,320,186	1,418,142	1,521,937	1,631,838
Tekirdağ	6,093,554	6,550,281	7,033,993	7,545,867	8,087,117
Tokat	648,164	675,762	704,342	733,933	764,566
Trabzon	1,055,535	1,127,552	1,203,502	1,283,550	1,367,869
Tunceli	81,588	86,817	92,675	99,280	106,773
Şanlıurfa	2,177,116	2,305,442	2,439,911	2,580,746	2,728,177
Uşak	1,001,343	1,056,628	1,114,235	1,174,259	1,236,801
Van	695,531	746,168	799,726	856,332	916,113
Yozgat	579,513	596,930	614,802	633,145	651,969
Zonguldak	2,665,009	2,720,903	2,777,730	2,835,502	2,894,232
Aksaray	586,183	626,121	668,204	712,520	759,158
Bayburt	69,224	73,289	77,571	82,079	86,826

Table 4.29 (continued)

Karaman	555,276	592,961	632,684	674,528	718,578
Kırıkkale	535,868	554,016	572,715	591,983	611,835
Batman	478,882	497,692	517,089	537,086	557,696
Şırnak	295,533	303,546	311,614	319,736	327,911
Bartın	289,979	302,161	314,727	327,683	341,040
Ardahan	86,492	93,967	101,981	110,567	119,761
Iğdır	126,542	132,605	138,821	145,193	151,725
Yalova	780,121	786,746	793,416	800,131	806,892
Karabük	869,355	910,159	952,487	996,383	1,041,892
Kilis	135,617	144,988	154,848	165,213	176,103
Osmaniye	899,559	1,010,204	1,131,494	1,264,184	1,409,061
Düzce	768,989	822,707	879,387	939,148	1,002,114
TURKEY	**184,845,601**	**194,714,839**	**205,057,777**	**215,895,761**	**227,251,568**

4.8. Comparison with Previous Studies

When the comparison with previous studies in the literature is made, we obtain the following table:

Table 4.30: Comparison of Forecasted Electricity Consumption Values with Previous Studies (GWh)

Year	**Actual Values**	**Random Effects Panel Data Model**	**Yiğit (2011)**	**Kavaklioglu et al. (2009)**	**Hamzaçebi (2007)**	**Erdoğdu (2007)**
2008	161.94	156.86		165.94	173.59	146.37
2009	156.89	165.38		175.04	189.47	145.14
2010	172.05	175.68	203.58	182.68	206.83	155.67
2011	186.09	185.93	214.51	189.32	225.8	156.01
2012	194.92	194.72	226.12	195.37	246.52	158.15
2013		205.06	238.49	201.09	269.19	169.21
2014		215.9	251.71	206.67	293.96	160.09
2015		227.25	265.85	212.17	321.05	

According to the table, Erdoğdu (2007) underestimates the electricity consumption while Yiğit (2011) and Hamzaçebi (2007) highly overestimate the net electricity consumption of Turkey. Kavaklıoğlu et al. (2009) developed an accurate fitted model. When we control their forecasted values, it is seen that their model is a proper model. However; the random effects panel data model is the best fitted model among the previous studies in the literature because its forecasted electricity consumption values are the closest to the official electricity consumption values and it is the most compatible to the official electricity consumption values. Therefore; our study is crucial for the authorities to take measures on future electricity consumption.

CHAPTER 5

CONCLUSION AND FURTHER RESEARCH

Countries would like to forecast the future energy consumption in order to supply their future energy needs. Therefore; it is vital to analyze and control the reliability of electricity supply. In this thesis study, we try to estimate Turkey's provinces electricity energy consumption using panel data analysis. We consider that modeling the Turkish electricity consumption with respect to provinces provide more accurate estimation of total electricity consumption of Turkey. Besides this, with the help of provincial electricity consumption forecasts, both the electricity consumption forecasts of provinces and the electricity consumption forecasts of geographical regions would be obtained. Thus; our study will be crucial for the authorities and will provide more enlightening forecasts in order to take measures on future electricity consumption.

In this thesis study, we try to acquire the most consistent and accurate electricity energy consumption model for Turkey and its provinces by using the fixed effects, random effects and dynamic panel data analysis methods with electricity consumption values of provinces of Turkey between the years of 1999-2010. By comparing the MAPE, RMSE values and correlation between actual electricity consumption values and fitted electricity consumption values, we concluded the random effects panel data analysis is the best fitted model among three methods and providing the most accurate results. After that, by using ETS method we obtain the forecasted values of population of provinces, total number of industrial enterprise of provinces and total number of household of provinces between the years 2011 and 2015. By using these forecasted values of covariates, we calculate the forecasts of the total electricity consumption values of Turkey and its provinces until 2015 using the random effects panel data fitted model. All the empirical results of the study are compared with the official electricity consumption amounts and found compatible with the official values.

For future research, it is suggested to develop Turkey's aggregate electricity energy consumption model by using the other energy elements such as natural gas, petrol, coal, renewable energy etc. with multivariate panel data analysis.

In another future research, it can be concentrated on just one crucial geographical region in terms of electricity consumption. For example, Marmara Region can be considered. Because this region has very little amount of electricity leakage, is very developed region, has a highly developed industry and consequently is the most electricity consuming region of Turkey. However, one might face estimation problems in such analysis due to small number of observations.

REFERENCES

Azadeh A, Taghipour M, Asadzadeh SM, Abdollahi M. Artificial immune simulation for improved forecasting of electricity consumption with random variations. Electrical Power and Energy Systems 2014; 55: 205–224.

Atakhanova Z, Howie P. Electricity demand in Kazakhstan. Energy Policy 2007; 35: 3729–3743.

Baltagi BH. Econometric analysis of panel data. John Wiley: Chichester; 2005.

Baltagi BH. Econometric analysis of panel data. 4th ed. John Wiley: West Sussex; 2008.

Bianco V, Manca O, Nardini S. Electricity consumption forecasting in Italy using linear regression models. Energy 2009; 34: 1413–1421.

Bolker B. Linear mixed-effects models using Eigen and S4; accessible at June 2, 2014; retrieved from: http://cran.r-project.org/web/packages/lme4/lme4.pdf.

Bilgili M. Estimation of net electricity consumption of Turkey. Journal of Thermal Science and Technology 2009; 29.2: 89-98.

Brian Ripley. Generalized estimation equation solver. R cran package accessible at July 12, 2014; retrieved from: http://cran.r-project.org/web/packages/gee/gee.pdf.

Demir AS, Taskin H. The modeling of the electric consumption of Turkey and estimation using a Quasi-Newton algorithm up to the year 2012. Energy Sources 2011; Part B, 6: 106–110.

Dilaver Z, Lester CH. Industrial electricity demand for Turkey: A structural time series analysis. Energy Economics 2011: 426- 436.

Dilaver Z, Lester CH. Turkish aggregate electricity demand: An outlook to 2020. Energy 2011; 36: 6686–6696.

Dilaver, Z. Residential electricity demand for Turkey: A structural time series analysis. Surrey Energy Economics Centre (SEEC), Department of Economics, University of Surrey, UK. Working paper retrieved from: http://www.aaee.at/2009-IAEE/uploads/fullpaper_iaee09/P_140_Dilaver_Zafer%20_29-Jun 2009, % 2016: 51.pdf.

Douglas B. Computational methods for mixed models. Department of Statistics University of Wisconsin Madison. R cran package 2014; retrieved from: http://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf.

Erdogdu E. Electricity demand analysis using cointegration and ARIMA modelling: a case study of Turkey. Energy Policy 2007; 35: 1129–1146.

EÜAŞ, Turkish Electricity Generation Company, Annual Report 2011, retrieved on 20 September 2014 from: http://www.euas.gov.tr.

Hamzacebi C. Forecasting of Turkey's net electricity energy consumption on sectoral bases. Energy Policy 2007; 35: 2009–16.

Halicioglu F. Residential electricity demand dynamics in Turkey. Energy Economics 2007; 29(2): 199–210.

Hsiao C. Panel data analysis—advantages and challenges. Xiamen University, Wang Yanan Institute for Studies in Economics; 2007. Working paper retrieved from:

http://wise.xmu.edu.cn/Master/Download/UploadFiles%5C2007101220281470554751 1776.pdf.

Hsiao C. Panel data analysis – advantages and challenges. Xiamen University, Wang Yanan Institute for Studies in Economics; 2006. IEPR Working Paper 06.49.

Hyndman RJ, Koehler AB, Snyder RD, Grose S. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 2002; 18(3): 439-454.

Hyndman RJ, Koehler AB, Ord JK, Snyder RD. Forecasting with exponential smoothing: The state space Approach. 8th ed. Springer-Verlag, Berlin, Heidelberg; 2008.

Inglesi R. Aggregate electricity demand in South Africa: Conditional forecasts to 2030. Applied Energy 2010; 87: 197–204.

International Energy Agency. World energy outlook 2011. retrieved on 10 August 2014 from: http://www.iea.org/Textbase/npsum/weo2011sum.pdf.

Brüderl J. Panel data analysis. University of Mannheim 2005; retrieved from: http://www.sowi.unimannheim.de/lessm/veranst/Panelanalyse.pdf.

Kavaklioglu K. Robust electricity consumption modeling of Turkey using Singular Value Decomposition. Electrical Power and Energy Systems 2014; 54: 268–276.

Kavaklioglu K, Ceylan H, Ozturk HK, Canyurt OE. Modeling and prediction of Turkey's electricity consumption using Artificial Neural Networks. Energy Conversion and Management 2009; 50: 2719–2727.

Keles SM. Electricity demand forecasts and their effects on Turkish economy. Republic of Turkey Prime Ministry under Secretariat of Treasury Ankara; 2005.

Kıran MS, Özceylan E, Gündüz M, Paksoy T. Swarm intelligence approaches to estimate electricity energy demand in Turkey. Knowledge-Based Systems 2012; 36: 93–103.

Küçükbahar, D. Modeling monthly electricity demand in Turkey for 1990- 2006, Master Thesis, Industrial Engineering Department, Middle East Technical University, Ankara; 2008.

Maddala, GS. Introduction to Econometrics. Third ed. New York: Wiley; 2001.

Mohamed Z, Bodger P. Forecasting electricity consumption in New Zealand using economic and demographic variables. Energy 2005; 30: 1833–1843.

Narayan PK, Smyth R. The residential demand for electricity in Australian application of the bounds testing approach to cointegration. Energy Policy 2005; 33: 467–474.

O'Brien RM. A caution regarding rules of thumb for variance inflation factors. Quality & Quantity 2007; 41: 673–690.

Oscar TR. Panel data analysis fixed and random effects using stata. University of Princeton, lecture notes accessible at February 25, 2014; retrieved from: http://www.princeton.edu/~otorres/Panel101.pdf.

Ozpala P. Statistical modeling of hourly electricity load series in Turkey, Master Thesis, Economics Department, Middle East Technical University, Ankara; 2013.

Ozturk HK, Ceylan H. Forecasting total and industrial sector electricity demand based on genetic algorithm approach: Turkey case study. International Journal of Energy Research 2005; 29: 829–840.

Pengfei L. Box-Cox Transformations: An Overview. 2005; retrieved from: http://www.ime.usp.br/~abe/lista/pdfm9cJKUmFZp.pdf.

Retrieved from website http://www.r-project.org/conferences/useR-2006/Abstracts/Hyndman.pdf, accessible at January 17, 2014.

Statistics Database of TUIK Official Website: http://www.tuik.gov.tr/, March 2014.

Wikipedia. Panel Data, accessible at August 8, 2014; retrieved from: http://en.wikipedia.org/wiki/Panel_data.

Wikipedia. Box-Cox Transformation, accessible at August 13, 2014; retrieved from: http://en.wikipedia.org/wiki/Power_transform.

Yigit V. Genetik algoritma ile Türkiye net elektrik enerjisi tüketiminin 2020 yılına kadar tahmini. International Journal of Engineering Research and Development 2011; 3: 37-41.

APPENDIX A

MODEL RESULTS

A.1 Results of Marginal AR(1) Model

Table 1: Working Correlation of Marginal AR(1) Model

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	1.0000000	0.9834910	0.9672546	0.9512863	0.9355815	0.9201361	0.9049456	0.8900059	0.8753128	0.8608623
[2,]	0.9834910	1.0000000	0.9834910	0.9672546	0.9512863	0.9355815	0.9201361	0.9049456	0.8900059	0.8753128
[3,]	0.9672546	0.9834910	1.0000000	0.9834910	0.9672546	0.9512863	0.9355815	0.9201361	0.9049456	0.8900059
[4,]	0.9512863	0.9672546	0.9834910	1.0000000	0.9834910	0.9672546	0.9512863	0.9355815	0.9201361	0.9049456
[5,]	0.9355815	0.9512863	0.9672546	0.9834910	1.0000000	0.9834910	0.9672546	0.9512863	0.9355815	0.9201361
[6,]	0.9201361	0.9355815	0.9512863	0.9672546	0.9834910	1.0000000	0.9834910	0.9672546	0.9512863	0.9355815
[7,]	0.9049456	0.9201361	0.9355815	0.9512863	0.9672546	0.9834910	1.0000000	0.9834910	0.9672546	0.9512863
[8,]	0.8900059	0.9049456	0.9201361	0.9355815	0.9512863	0.9672546	0.9834910	1.0000000	0.9834910	0.9672546
[9,]	0.8753128	0.8900059	0.9049456	0.9201361	0.9355815	0.9512863	0.9672546	0.9834910	1.0000000	0.9834910
[10,]	0.8608623	0.8753128	0.8900059	0.9049456	0.9201361	0.9355815	0.9512863	0.9672546	0.9834910	1.0000000
[11,]	0.8466504	0.8608623	0.8753128	0.8900059	0.9049456	0.9201361	0.9355815	0.9512863	0.9672546	0.9834910
[12,]	0.8326730	0.8466504	0.8608623	0.8753128	0.8900059	0.9049456	0.9201361	0.9355815	0.9512863	0.9672546

	[,11]	[,12]
[1,]	0.8466504	0.8326730
[2,]	0.8608623	0.8466504
[3,]	0.8753128	0.8608623
[4,]	0.8900059	0.8753128
[5,]	0.9049456	0.8900059
[6,]	0.9201361	0.9049456
[7,]	0.9355815	0.9201361
[8,]	0.9512863	0.9355815
[9,]	0.9672546	0.9512863
[10,]	0.9834910	0.9672546
[11,]	1.0000000	0.9834910
[12,]	0.9834910	1.0000000

Table 2: Residuals of Marginal AR(1) Model

Summary of Residuals:

Min	1Q	Median	3Q	Max
-1.04148065	-0.27377636	-0.04176986	0.17105411	1.44841807

Table 3: Results of Marginal AR(1) Model

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	-2.60298463	0.315355472	-8.254129	0.296849409	-8.768704
TIME	0.03047489	0.002625976	11.605168	0.002499695	12.191446
$\ell n(POP)$	0.93310354	0.045927751	20.316770	0.045760235	20.391144
INDUSTRY	51.11086844	9.550261699	5.351777	13.300985185	3.842638
HOUSEHOLD^5	1.22069022	0.490081459	2.490790	0.581656278	2.098645

Estimated Scale Parameter: 0.1616639
Number of Iterations: 7

A.2 Intercept and Time Values of Provinces in Random Effects Model

Table 4: Intercept and time values of all provinces

	(Intercept)	t						
1	0.140357711	-0.0052847719	28	-0.238720069	-0.0219843006	55	-0.378661711	0.0166255338
2	-0.111368207	0.0038080330	29	-0.100601865	-0.0018277421	56	0.067441281	-0.0173350110
3	-0.248231311	0.0010738656	30	-0.182298581	-0.0216575726	57	-0.098001654	-0.0057862256
4	-0.480120103	-0.0187037679	31	0.282343584	0.0274762454	58	-0.384866987	0.0266860873
5	-0.020604350	-0.0139903975	32	0.187895257	-0.0118174001	59	0.780055713	0.0237691458
6	-0.115225133	0.0178643011	33	-0.038678830	-0.0029711368	60	-0.488230053	-0.0031441297
7	0.148656604	0.0112231529	34	0.310532532	0.0290871315	61	-0.536814710	0.0135799674
8	0.184351022	-0.0088665738	35	0.606704390	0.0221128318	62	0.342187447	-0.0161951397
9	-0.267372155	0.0016173465	36	-0.291517363	-0.0007476799	63	-0.267354429	-0.0002600790
10	-0.147045376	0.0003390554	37	-0.070396041	-0.0053034513	64	0.047446010	0.0054751484
11	1.480588357	-0.0622031712	38	-0.007496552	-0.0060049249	65	-0.899578502	0.0097830055
12	-0.432464753	-0.0065785243	39	0.820977792	-0.0056059694	66	-0.281076518	-0.0039538556
13	-0.556385595	0.0069326331	40	0.004328484	-0.0068405371	67	0.734132393	-0.0127182331
14	0.294214761	-0.0026937666	41	1.020841206	0.0289155454	68	-0.254796529	0.0069934439
15	-0.174701229	0.0411046389	42	0.039927645	0.0071892594	69	0.240263823	-0.0092337503
16	0.461824608	-0.0023535417	43	-0.236111360	0.0091045770	70	0.121687913	0.0066029118
17	0.178284799	0.0745921095	44	-0.166726732	-0.0035519852	71	0.174381388	-0.0111718902
18	0.064510880	-0.0149090327	45	-0.280954602	0.0218994046	72	-0.131055538	-0.0248293601
19	-0.307873434	-0.0052639075	46	0.140151069	0.0175177488	73	-0.168969446	-0.0286267479
20	0.071718656	-0.0043116790	47	-0.289824077	0.0008932264	74	0.078775607	-0.0114255136
21	-0.554768615	-0.0058136578	48	0.201182666	-0.0180941447	75	-0.098627444	0.0050705514
22	0.179105918	-0.0097305464	49	-0.561771275	0.0049291210	76	-0.189239641	-0.0103515794
23	0.120415417	-0.0258894797	50	0.397743350	-0.0305795385	77	1.104887027	-0.0547189130
24	-0.020427332	-0.0145316538	51	0.249331644	-0.0036879783	78	0.434400316	-0.0025084751
25	-0.525859230	0.0110619030	52	-0.434060163	0.0032306123	79	0.100678396	-0.0031877937
26	0.057320679	0.0045995247	53	-0.006824437	-0.0020871156	80	-0.581808316	0.0480257225
27	0.060586226	0.0156522704	54	-0.231549201	0.0258458631	81	-0.071173123	0.0086507272

APPENDIX B

R CODES

B.1 R Codes for Exploratory Analysis

```
library(epicalc)
library(lattice)
require(car)

#Data Inputting
datalast<- read.csv2(file.choose(), header=T)
dim(datalast)
head(datalast)

#Variance inflation factor (VIF) values
t=datalast[,2]-1999
vif(lm(EC~ t + log(POP) + INDUSTRY+HOUSEHOLD^5, data=datalast))
vif(lm(EC~ t + log(POP) + INDUSTRY+HOUSEHOLD, data=datalast))

#Box-Cox Transformation
b=boxcox(EC ~ t + POP + INDUSTRY+ HOUSEHOLD, data=datafirst)
lamda=b$x
lik=b$y
 bc=cbind(lamda,lik)
bc[order(-lik),]

# Pairwise Correlation
cor(datalast,use="pairwise.complete.obs")
```

```
# Linearity between Dependent Variable and Covariates
time=c("1999",2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010)

xyplot((datalast[,3])~(datalast[,4])|time,ylab="Transformed
EC",xlab="Population",aspect=1)

xyplot((datalast[,3])~log(datalast[,4])|time,ylab="Transformed EC",xlab="log
Population",aspect=1)

xyplot((datalast[,3])~(datalast[,5])|time,ylab="Transformed
EC",xlab="Industry",aspect=1)

xyplot((datalast[,3])~(datalast[,6])|time,ylab="Transformed
EC",xlab="Household",aspect=1)

xyplot((datalast[,3])~log(datalast[,6])|time,ylab="Transformed EC",xlab="log
Household",aspect=1)

xyplot((datalast[,3])~exp(datalast[,6])|time,ylab="Transformed EC",xlab="exp
Household",aspect=1)

xyplot((datalast[,3])~(datalast[,6])^2|time,ylab="Transformed EC",xlab="Square of
Household",aspect=1)

xyplot((datalast[,3])~(datalast[,6])^5|time,ylab="Transformed
EC",xlab="Household^5",aspect=1)

# Determining Lag Values
data.EC=matrix(c(rep(0,81*12)),ncol=12)
for(i in 1:12)
```

```
data.EC[,i]=datalast[datalast[,2]==(i+1998),3]
head(data.EC)
new1=NULL
for(i in 1:11)
new1=c(new1,data.EC[,i])
new2=NULL
for(i in 2:12)
new2=c(new2,data.EC[,i])
ac1=cor(new1,new2,use="pairwise.complete.obs")
ac1
new1=NULL
for(i in 1:10)
new1=c(new1,data.EC[,i])
new2=NULL
for(i in 3:12)
new2=c(new2,data.EC[,i])
ac2=cor(new1,new2,use="pairwise.complete.obs")
ac2
```

B.2 R Codes for Modeling

```
# Fixed Effects Model

library(gee)
fixedmodel <- gee((datalast[,3])~ t + log(datalast[,4])+(datalast[,5])+ (datalast[,6]^5),
id = datalast[,1], family = gaussian, corstr = "AR-M",Mv=2)

summary(fixedmodel)
```

```
# Random Effects Model

library(lme4)
random_lmer1 <- lmer(EC ~ t + (HOUSEHOLD^5) + log(POP)  + (INDUSTRY) +
(1 | ID), data = datalast, na.action = na.omit)
random_lmer2 <- lmer(EC ~ t + (HOUSEHOLD^5) + log(POP)  + (INDUSTRY) +
(t | ID), data = datalast, na.action = na.omit)
anova(random_lmer1, random_lmer2)

summary(random_lmer2)
fixef(random_lmer2)
ranef(random_lmer2)

# Dynamic Model

datalag1<- read.csv2(file.choose(), header=T)
dim(datalag1)
head(datalag1)

t=datalag1[,2]-2000

lag1model <- gee(y ~ t  + ylag1 +(HOUSEHOLD)^5 + log(POP)+ INDUSTRY ,
data = datalag1, id = ID, family = gaussian, corstr = "independence")
summary(lag1model)
```

B.3 R Codes for Accuracy Checking

```
library(Metrics)
x=(fixedmodel$fitted.v)^5
head(x)
z=(fixedmodel$y)^5
```

```
head(z)
accuracy(x, z, test=NULL)

a=(fitted(random_lmer2))^5
head(a)

accuracy(a, z, test=NULL)

m=(lag1model$fitted.v)^5
head(m)

g=(lag1model$y)^5
head(g)

accuracy(m, g, test=NULL)
```

B.4 R Codes for Forecasting

for Adana;

```
library(forecast)
k1=ets((data1[,4]))
k2=ets((data1[,5]))
k3=ets((data1[,6]))
forecast(k1,5)
forecast(k2,5)
forecast(k3,5)
```