ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

EDUCATION AND INCOME INEQUALITY IN TURKEY: NEW EVIDENCE FROM PANEL DATA

M.A. THESIS

Esra ÖZTÜRK

Department of Economics

Economics Programme

MAY, 2015

ISTANBUL TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

EDUCATION AND INCOME INEQUALITY IN TURKEY: NEW EVIDENCE FROM PANEL DATA

M.A. THESIS

Esra ÖZTÜRK
(412131007)

Department of Economics

Economics Programme

Thesis Advisor: Asst. Prof. Ayşegül Kayaoğlu-Yılmaz
Thesis Co-Advisor: Asst. Prof. Christopher Hannum

MAY,2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ SOSYAL BİLİMLER ENSTİTÜSÜ

TÜRKİYE'DE EĞİTİM VE GELİR EŞİTSİZLİĞİ: PANEL DATA ÜZERİNDEN YENİ BULGULAR

YÜKSEK LİSANS TEZİ

Esra ÖZTÜRK
(412131007)

İktisat Anabilim Dalı

İktisat Programı

Tez Danışmanı: Yard. Doç. Dr. Ayşegül Kayaoğlu-Yılmaz
Tez Ek Danışmanı: Yard. Doç. Dr. Christopher Hannum

MAYIS, 2015

Esra Öztürk, a **M.A.** student of ITU **Institute of Social Sciences** student ID **412131007**, successfully defended the **thesis** entitled **"EDUCATION AND INCOME INEQUALITY IN TURKEY: NEW EVIDENCE FROM PANEL DATA"**, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Asst. Prof. Ayşegül KAYAOĞLU YILMAZ** İstanbul Technical University	....
Co-advisor :	**Asst. Prof. Christopher HANNUM** İstanbul Technical University	....
Jury Members :	**Prof. Dr. Fuat ERDAL** İstanbul Technical University	....
	Assoc. Prof. Raziye SELİM İstanbul Tecnical University	....
	Asst. Prof. Bengi İLHAN İstanbul Kemerburgaz University	....

Date of Submission : May 4, 2015
Date of Defense : May 29, 2015

FOREWORD

First of all, I would like to thank to my advisor Asst. Prof. Ayşegül Kayaoğlu-Yılmaz and co-advisor Asst. Prof. Christopher Hannum for their supports and advices all through the thesis preparation process.

I would also like to thank to Asst. Prof. Bengi Yanık-İlhan for her encouragement and guidance from the beginning of my graduate studies till the end.

In addition, I appreciate to all my friends for being a part of my living and brightening up my life with their existence.

My special thanks go to my family who supported and encouraged me in all periods of my life. I would not have achieved my goals without their supports and patiences. I will be grateful to them all through my life.

April, 2015 Esra ÖZTÜRK

TABLE OF CONTENTS

Page

ABBREVIATIONS

2SLS	**:** Two Stage Least Square
ABPRS	**:** Address Based Population Registration System
GDP	**:** Gross Domestic Product
GLS	**:** Generalized Least Square
GMM	**:** Generalized Method of Moments
NUTS	**:** Nomenclature of Territorial Units for Statistics
OECD	**:** Organisation of Economic Co-operation and Development
OLS	**:** Ordinary Least Square
R&D	**:** Research and Development
TURKSTAT	**:** Turkish Statistical Institution
UNDP	**:** United Nations Development Programme
UNESCO	**:** United Nations Educational, Scientific and Cultural Organization
WB	**:** World Bank

LIST OF TABLES

LIST OF FIGURES

EDUCATION AND INCOME INEQUALITY IN TURKEY: NEW EVIDENCE FROM PANEL DATA

SUMMARY

In this study, income and education inequalities in Turkey are analyzed by using panel data approach in provincial level between 2008 and 2013. The differences of income and education inequalities between provinces and regions are presented. A general picture of Turkey according to these topics is provided.

The Gini index of both income and education is used to measure the inequalities. The income Gini coefficient is provided from Turkish Statistical Institute (TURKSTAT) in Nuts1 level. It shows that income inequality has a downward trend for this period in Turkey. The east part of Turkey has higher income distortions, while Marmara and Black Sea regions have relatively lower income Gini coefficients. Since the economic activities are located in western regions of Turkey, it is expected to have lower income Gini coefficient in this side of the country.

On the other side, the education Gini index is calculated by using the completed education level macro data obtained from TURKSTAT in 8 levels: illiterate, literate without diploma, primary school, secondary school, high school, university degree, master and doctorate. The education Gini is computed in provincial level from 2008 to 2013. Like income Gini index, the education Gini coefficient is higher in Eastern and Southeastern Anatolia regions. Some provinces in these regions have the education Gini coefficients that are higher than 50%. The proportion of illiterate population is relatively larger in eastern side than the west part of Turkey. The provinces that have more equal education distribution are generally located in Marmara and Central Anatolia regions.

The final goal of the thesis is to examine the impact of education inequality on income inequality in Turkey. To evaluate an econometric analysis, a dataset including income and education Gini coefficients, per capita value added, labor force participation ratio, number of students in secondary school, per capita budget expenditure of local authorities on education and demographic variables is organized in provincial level. In the econometric analysis, static and dynamic panel data methods are estimated. All the variables are used in logarithmic form, so that the coefficients of the variables indicate the elasticity for income Gini coefficient. Pooled OLS, fixed effects and random effects models are applied in static form. Since the results indicate the serial correlation between income Gini index and error term, the analysis continues with dynamic models. OLS, fixed effects, random effects and first-difference models are estimated. However, there is reverse causality between income Gini and education Gini coefficients. Therefore, Anderson-Hsiao model is used to control this causality. To control both serial correlation and endogeneity, Arellano-Bond model is applied. The results of all these methods indicate that education Gini has a negative impact on income Gini coefficient. Therefore, it can be said that the results are robust.

Key words: Income Inequality, Education Inequality, Panel Data Analysis
JEL Codes: O15, I24, C33

TÜRKİYE'DE EĞİTİM VE GELİR EŞİTSİZLİĞİ:

PANEL DATA ÜZERİNDEN

YENİ BULGULAR

ÖZET

Bu tezde, Türkiye'deki eğitim ve gelir eşitsizlikleri panel veri kullanılarak il bazında analiz edilmiştir. Bu analiz, 2008 ve 2013 yılları arasını kapsamaktadır. İller ve bölgeler arasındaki gelir ve eğitim eşitsizliği farkları sunulmuştur. Türkiye'nin bu konudaki yapısına genel bir bakış sağlanmıştır.

Eşitsizlik ölçütü olarak eğitim ve gelir Gini katsayıları kullanılmıştır. Gelir için Gini katsayısı Türkiye İstatistik Kurumu (TÜİK)'den birinci derece İBBS düzeyinde elde edilebilmektedir. Elde edilen gelir Gini katsayısı, Türkiye'de gelir eşitsizliğinin azalan bir trende sahip olduğunu göstermektedir. Buna göre, tezde kullanılan dönemler içinde gelir Gini indeksinin genel olarak her bölge için azaldığı görülmektedir. Bölge bazında bakılacak olursa, Türkiye'nin doğu bölgelerinde gelir dağılımının daha adaletsiz olduğu, bu bölgelerde gelir eşitsizliğinin daha yüksek olduğu görülmektedir. Buna karşın, Marmara ve Karadeniz bölgelerinde görece daha düşük gelir Gini katsayısı gözlenmektedir. Bu bölgelerde ekonomik faaliyetlerin yoğunlaşması burada daha düşük gelir eşitsizliğinin görülmesinin nedenleri arasında sayılabilir.

Eğitim Gini katsayısı ise yine TÜİK'den elde edilen bitirilen eğitim düzeyi makro verisi kullanılarak 8 düzey için hesaplanmıştır. Hesaplamada kullanılan düzeyler okuma-yazma bilmeme, okuma-yazma bilme ama diploma sahibi olmama, ilkokul, ortaokul, lise, üniversite, yüksek lisans ve doktora olarak belirlenmiştir. Katsayı, 2008 ve 2013 yılları arasında il bazında hesaplanmıştır. Gelir Gini indeksinde görüldüğü gibi eğitim Gini katsayısı da Doğu ve Güneydoğu Anadolu bölgelerinde daha yüksek olarak saptanmıştır. Bu bölgelerdeki bazı illerde eğitim Gini katsayısının %50'den daha yüksek olduğu gözlenmiştir. Bu bölgelerde okuma-yazma bilmeyen nüfus oranının görece diğer bölgelerden daha yüksek olması eğitim eşitsizliğinin doğuda batı bölgelerine göre daha yüksek çıkmasına neden olmaktadır. Diğer yandan, daha düşük eğitim Gini indeksine sahip iller genellikle Marmara ve Orta Anadolu bölgelerinde yerleşmişlerdir. Bu bölgelerde okullaşma oranları, okul sayıları daha yüksek olduğundan ve özellikle üniversitelerin büyük çoğunluğu bu bölgelerde (İstanbul ve Ankara gibi) yer aldığından, bu beklenen bir sonuçtur.

Tezin nihai amacı olarak, eğitim eşitsizliğinin gelir eşitsizliği üzerine etkisi analiz edilmiştir. Ekonometrik analiz yapmak için, gelir ve eğitim Gini katsayıları, kişi başına düşen katma değer, işgücüne katılım oranı, ortaokula kayıtlı öğrenci sayısı, kişi başına düşen merkezi yönetim kümülatif bütçe giderleri ve demografik verilerden oluşan bir veri seti hazırlanmıştır. Ekonometrik analizde statik ve dinamik modeller uygulanmıştır. Bütün değişkenler logaritmik formda kullanılmıştır, bu nedenle modellerde yer alan değişkenlerin katsayıları elastikiyetleri göstermektedir. Öncelikle En Küçük Kareler, Fixed Effects ve Random Effects metotları statik formda tahmin edilmiştir. Ancak gelir Gini katsayısı ve hata terimleri arasında korelasyon olduğundan dinamik modellere geçilmiştir. Bu modeller Nickell sapmasını kontrol etmek için de daha uygun yöntemlerdir. Dinamik modellerden ilk olarak En Küçük Kareler, Fixed effects, Random Effects ve First-difference modelleri kullanılmıştır. Eğitim ve gelir Gini indeksleri arasında ters nedensellikten şüphelenildiği için, bu durumu kontrol altına almak adına Anderson-Hsiao modeli uygulanmıştır. Hem gelir Gini indeksi ile hata terimleri arasında korelasyonu hem de endojeniteyi göz önünde bulundurduğundan son olarak Arellano-Bond model tahmin edilmiştir. Tahmin edilen bütün yöntemler, gelir Gini indeksi ile eğitim Gini indeksi arasında negatif etki olduğu sonucunu göstermişlerdir. Bu nedenle elde edilen sonuçların robust (berk) olduğu görülmüş olur.

Anahtar Kelimeler: Gelir Eşitsizliği, Eğitim Eşitsizliği, Panel Veri Analizi

1. INTRODUCTION

1.1. The Aim of Thesis

In this thesis, a general view of income and education inequality in Turkey is presented and education Gini coefficient is calculated to measure inequality for each provinces. The main purpose of this study is to determine the impact of educational inequality on income distribution at provincial level in Turkey for the period between 2008 and 2013.

1.2. The Importance of Thesis and Its Method

Education is a widely-studied subject in economics literature. Its key role in the economic and social development processes of societies, as a crucial ingredient of growth dynamic through its effect on human capital level of a country, establish it to be a very important topic in the literature.

It is not only its direct effect on the economic development and growth but also indirect spillover effect of it on social welfare makes the analysis of education (educational inequality, especially) very important for all the countries. Education makes people more active in social life and enhances the quality of their life. From an economic perspective, increase in education level provides people to have better skills. Thus, labor force becomes better trained and productive. More skilled and productive labor force induces improvements in human capital accumulation and creates a significant impact on economic growth. The study of Lopez, Thomas and Wang (1998) present the positive effect of the stock of human capital on the economic growth for 12 countries between 1970 and 1994. Moreover, the contributions of division of higher education to economic growth are searched for the period between 1965 and 2000 for Taiwan. Engineering/science, business/social science and agricultural sciences have a positive and significant effect on economic development. One additional year in education provides approximately 19% increase in real output (Lin, 2004).

The rise of the education level is one of the most important factors of economic development and reduction of poverty. If the education level of a population increases, this would affect the earnings of the labor force. The higher earnings directly reduce the number of people under the poverty line. The conclusions of the research that search for the impact of education on poverty reduction in Pakistan imply the negative relationship between higher education and poverty. This study also shows that the increase in the educational achievement reduces the possibility of people being poor (Awan *et al.*, 2011).

As well as the level of education, the distribution of education has an emphasis in economics. While the distortions of education dispersion have a negative impact on economy, they also cause huge gaps socially between and within societies. A higher level of educational inequality induces the economic growth negatively and gives rise to poverty. Castello and Domenech (2002) suggest a linkage between human capital inequality and economic growth. They explain the negative relationship between human capital inequality which is measured by education Gini, and economic development by the association of higher education inequality with lower investment rates and as a consequence, lower economic growth.

Despite the unfavorable effect of education inequality on per capita income or income growth, its impact on income inequality is not stated clearly in the literature. Economics literature presents both negative and positive relationship between income inequality and education inequality. A cross-country analysis for 59 countries suggests that larger dispersion of education of labor force brings greater income inequality (Park, 1996). The other study that shows the positive effect of education inequality on income distribution is Gregorio and Lee (2002)'s research of panel data from 1965 to 1990. This study indicates that decrease in educational dispersion by one standard deviation reduces the income inequality. These are the examples that represent the education inequality with standard deviation of years of schooling.

On the other hand, the studies that measure the education inequality with education Gini index imply negative relationship between education inequality and income inequality. Checchi (2001)'s research is one of them with U-shaped relationship between income Gini and education Gini coefficients for 94 countries. Földvari and Van Leeuwen (2014) also find out the negative and U-shaped effect of education Gini index on income Gini.

In this concept, this thesis is prepared to analyze the income and education inequality of Turkey. Calculation of education Gini index is indicated and education Gini is constructed for each provinces. A panel data analysis is preferred to find out the impact of education inequality on income distribution in Turkey in provincial level. Through this analysis, the thesis aims to answer that how the unclear results that are stated in economics literature conclude in Turkey. The data between 2008 and 2013 is used in the thesis. Static and dynamic panel data estimation methods are applied in the econometric analysis part. All the variables are converted to logarithmic form to make the interpretation easier, so that the coefficients of variables indicate the elasticity. Pooled OLS, fixed effects and random effects models are estimated in static form. Since the income Gini coefficient and error term have serial correlation, dynamic models are used to control this characteristic. In dynamic model part, OLS, fixed effects, random effects and first-difference methods are applied. It is suspected to become reverse causality between education and income Gini coefficients. Therefore Anderson-Hsiao and Arellano-Bond models are included to control for this causality and serial correlation. All these models show the negative effect of education Gini coefficient on income Gini. According to results of econometric analysis that are prepared by using static and dynamic models, it is discussed about the policies that should be applied in Turkey and policy implications about education are presented.

1.3. The Structure of Thesis

Section 2 introduces the concept of income inequality and educational inequality and presents the general picture of these two concepts regarding Turkey through descriptive statistics. Basic assumptions, definitions and explanations about these inequalities are presented in this section. Measurement technics of both income and educational inequality are shown and calculation methods are indicated in detail. Literature review about the relationship between educational inequality and income distribution, inequality calculation and econometric methods is located at the end of this section.

Section 3 describes the data and the methodology in detail. Econometric models used for the regression analysis and theoretical background of the models are explained in this section. The sources and the calculation of inequality coefficients are presented and the changing process of these inequality coefficients is shown by figures and tables year by year.

Section 4 presents the findings of econometric analysis of the thesis. The results of the model are analyzed and the question of how much impact has the educational inequality on income distribution is answered through the regression outcomes.

Section 5 concludes the thesis where the findings of the econometrical analysis evaluated in sum. By considering the inferences of econometric analysis, the necessities of improvement of educational inequality are discussed and some policy suggestions are provided.

2. INCOME AND EDUCATION INEQUALITY

2.1. Income Inequality

2.1.1. Income inequality

Income inequality is the most significant evidence of distinction in life standards within each country. It can be defined as differences in income for the whole population or the distortions in income distribution in a country (Litchfield, 1999).

High income inequality may be caused by many reasons but the most important one is changes in labor force participation decision. Directing human capital into the labor force with an efficient way has the most equalizing effect on income distribution and it can prevent the waste of human resources. The problem of low-paid and low-skilled worker would be diminished with an effective guidance. Besides the problem of effective dispersion of human capital, inequality in earnings and wages are other factors that increase the income inequality and worsen the economic development of a country. According to fluctuations in the earnings of individuals, the income of the household changes and income inequality across the country increases (Duman, 2008 and OECD, 2014).

2.1.1.1. Income inequality in Turkey

Turkey is one of the countries that have relatively higher income inequality. Although it seems a chronic problem of Turkey's economy, income inequality starts to decrease recently.

There are many different ways to measure income inequality. The share of income quintiles is the one of them that is calculated and published by World Bank (WB) for most of the countries. The income share of bottom quintile is given in the Figure 2.1 for some countries. As it can be seen from the figure that Turkey, shown with the thick black line, has an upward trend. This means that population at the bottom income share gets larger proportion from total income and it can be thought as evidence that shows the decrease of income inequality. However, Turkey is far behind even from developing countries with its relatively lower income share of bottom quintile. Even if it has higher values than Latin American countries, the proportion is still lower and the inequality is still larger than most of the developing countries such as Georgia, Romania, Belarus.

The increase in the income of bottom quintile is resulted with a decrease in the share of top quintile. The Figure 2.2 that is prepared with the share of top quintile data of some countries obtained from WB shows that Turkey has a downward trend until 2008. After the global crisis, the income share of top quintiles fluctuates with minor changes. In contrast to bottom quintile, the higher income share of top quintile indicates higher income inequality. Thus, Turkey has still larger distortions in the income distribution. Most of the developing countries have relatively lower income shares of top quintile and as a result lower income inequality than Turkey.

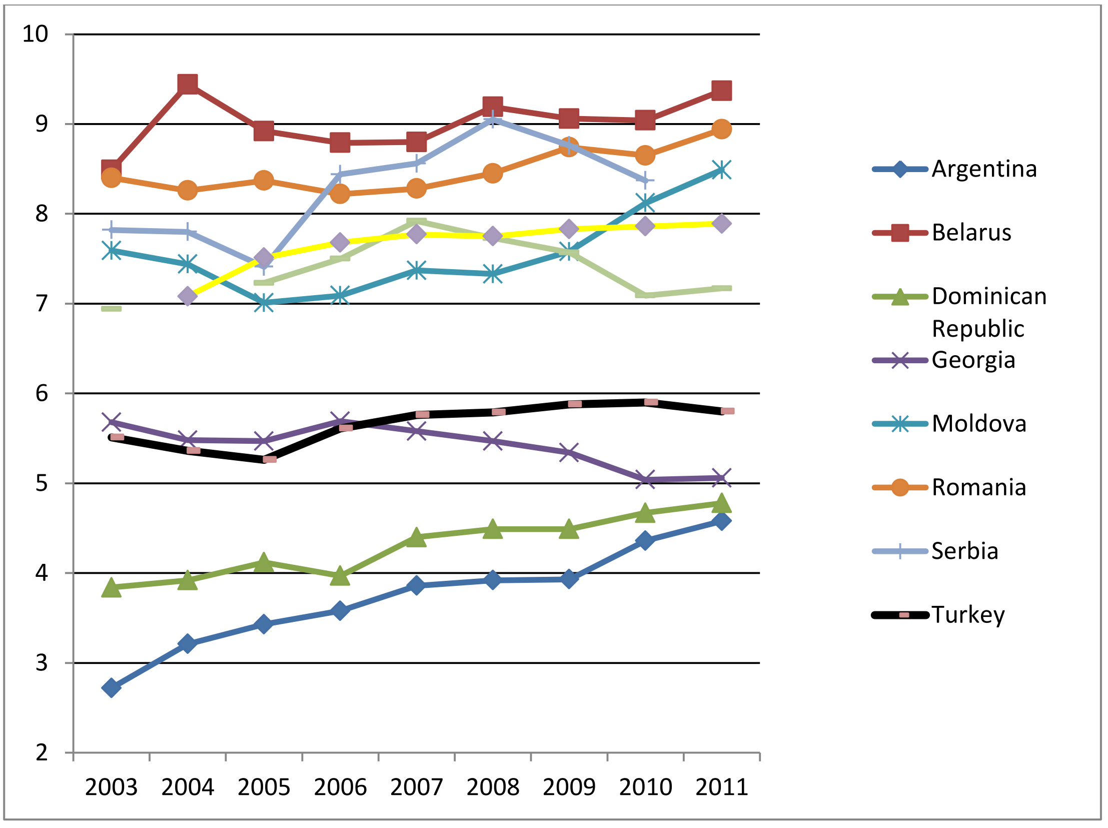

Figure 2.1: The income share of bottom quintile for some countries
Source: World Bank, Development Research Group

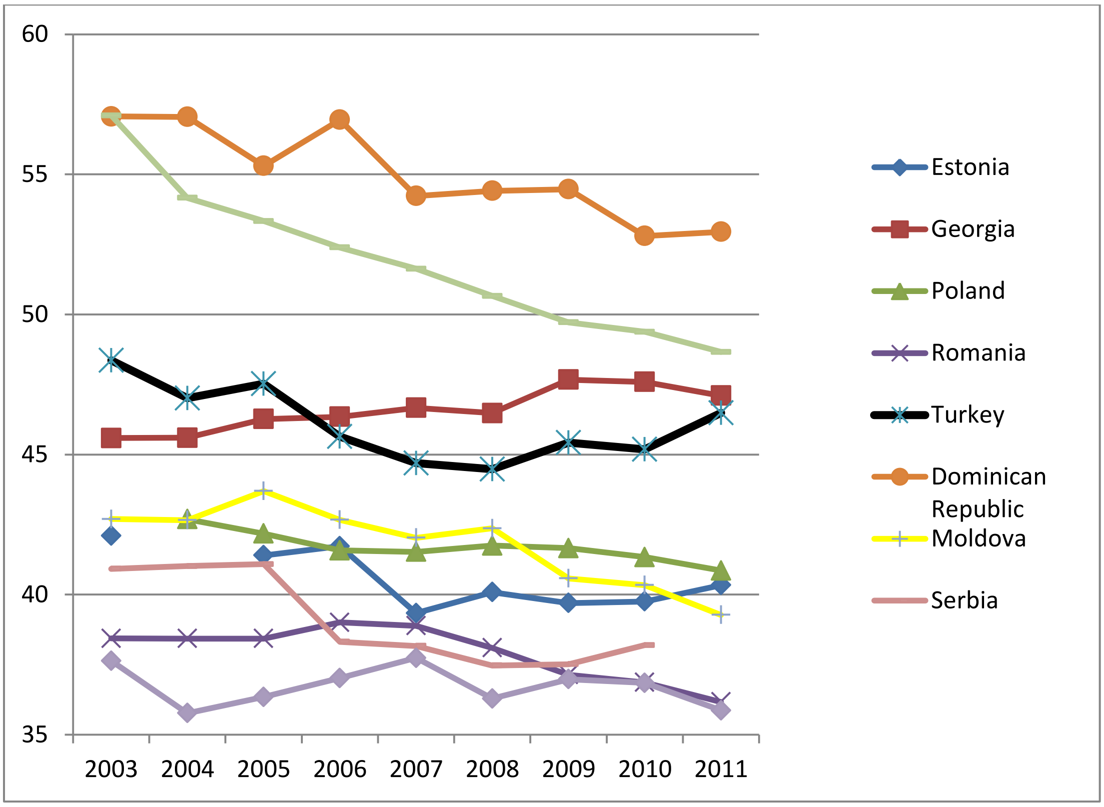

Figure 2.2: The income share of top quintile for some countries
Source: World Bank, Development Research Group

The upward trend in the income share of bottom quintile and the decrease in the share of top quintile are the evidences of improvements in income distribution of Turkey. However, Turkey has still relatively higher income inequality compared to most of OECD countries and developing economies (OECD Factbook, 2014). Recent researches and statistics show that Turkey is third country in the rankings of unequal income distributions after Chile and Mexico in 2011 (Selim, Günçavdı and Bayar, 2014). This means there is still a lot to do to equalize the income distribution of Turkey.

2.1.2. Income Gini coefficient

Besides the distribution of income by quintiles, the most common inequality measurement is Gini coefficient recently. The Gini index is originally developed by the Italian statistician Corrado Gini (1912). It provides the opportunity to compare income inequality levels of different units easily, since it indicates the inequality levels with a real number (Selim *et al*, 2014).

The Gini index is mainly based on the comparisons of cumulative percentages of the population against cumulative percentages of their income. It ranges between 0 and 1. The Gini index is 0 where all the households have equal incomes and 1 where one person has all the income and the others have nothing (OECD, 2014).

The income Gini indexes of some countries are shown in the Figure 2.3 below. The indexes are obtained from WB database. Turkey has again relatively worse view of income distribution when it is compared with the other developing countries. Although it has generally a downward trend until 2008, the global crisis affects the decrease of income inequality of Turkey and it starts to increase.

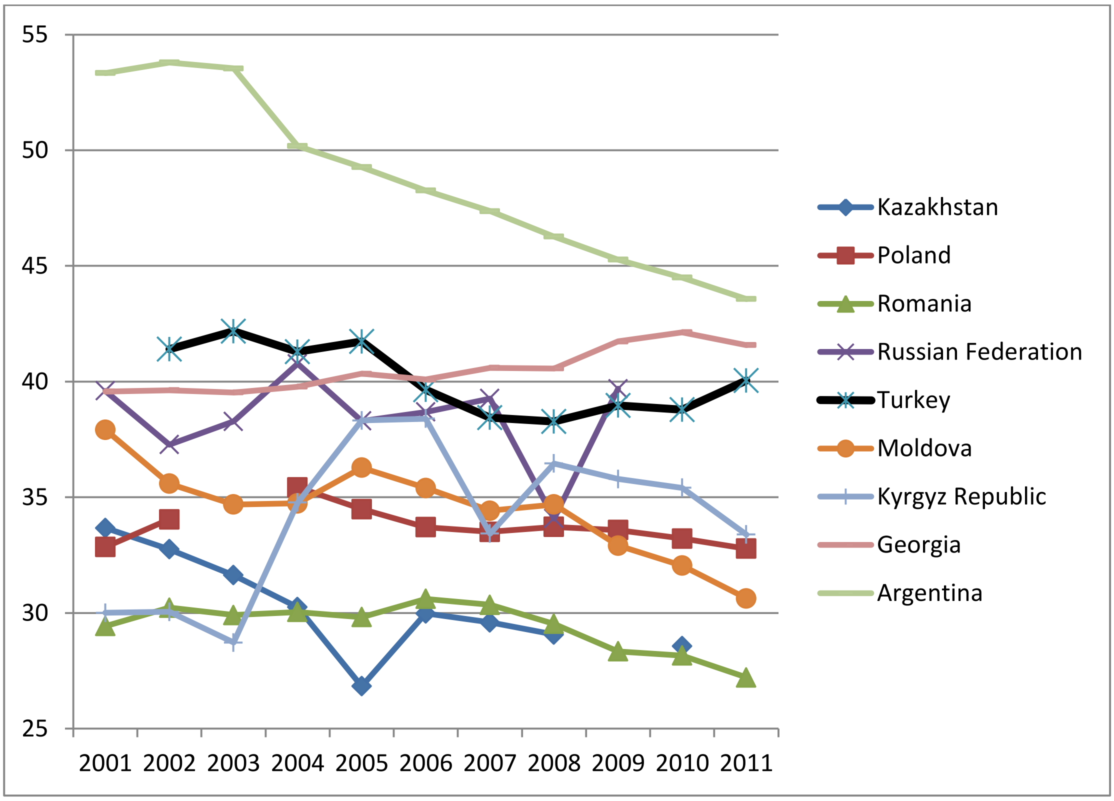

Figure 2.3: The income Gini coefficients of some countries
Source: World Bank, Research Development Group

Income Gini coefficient can be calculated by two different methods: the direct method and the indirect method. These methods will be introduced in two subsections below.

2.1.2.1. The direct method of Gini calculation

The direct method presents a mathematical approach to the Gini coefficient calculation. It is defined as "the ratio to the mean of half of the average over all pairs of the absolute deviations between people" (Deaton, 1997).

The formula used for the calculation of Gini index in the direct method is given below:

$$G = \frac{1}{\mu \, N \, (N-1)} \sum_{i>j} \sum_j |y_i - y_j| \qquad (\mathbf{2.1})$$

where G is the Gini coefficient, μ is the mean of income, N is the total number of observations and y_i and y_j are the incomes of the individuals (Thomas *et al.*, 2000).

2.1.2.2. The indirect method of Gini calculation

In the indirect method of Gini calculation, firstly the Lorenz curve is constructed using cumulative proportion of income on the vertical axis and the cumulative proportion of population on the horizontal axis. A forty-five degree line which presents the perfect equality of income distribution is drawn as egalitarian line.

The Gini coefficient is defined as the ratio of the area between Lorenz Curve and the egalitarian line to the area of egalitarian triangle. The Lorenz Curve is shown in Figure 2.4 below (Thomas, Wang and Fan, 2000).

$$G = \frac{Area\ of\ A\ (between\ egaliterian\ line\ and\ Lorenz\ Curve)}{Area\ of\ OWQ\ (Egaliterian\ triangle)} \qquad (\mathbf{2.2})$$

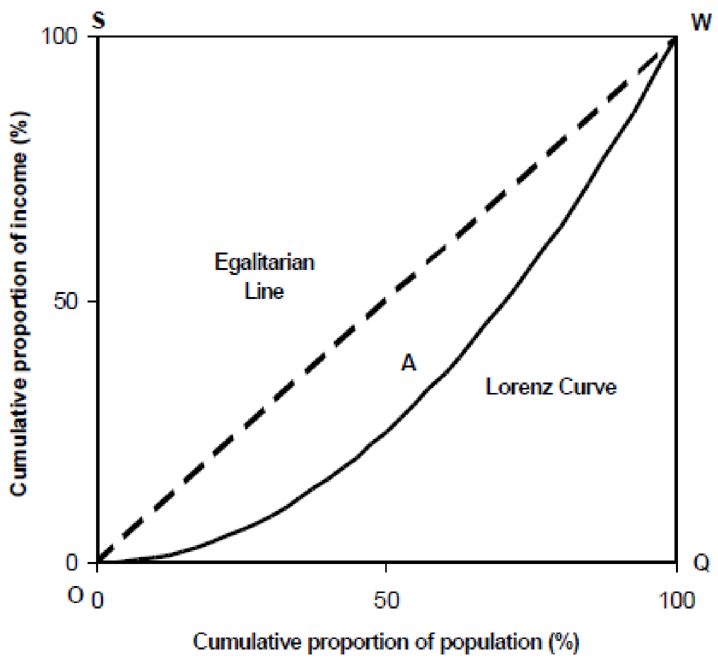

Figure 2.4: The Lorenz Curve
Source: Thomas, Wang and Fan (2001)

2.2. Educational Inequality

2.2.1. Educational inequality

Education has a key role with its contributions to development process of human capital and economic growth in economics literature. But the unequal opportunities for schooling causes distortions in the distribution of education and recent studies show that education inequality is a great deal in most of the developing countries, especially with its direct effect on human capital and income growth (Castello and Domenech, 2002 and Lopez *et al.*, 1998).

Standard deviation of schooling is extensively preferred measurement of education inequality (Park, 1996, Gregorio and Lee, 2002). But it determines the education distribution in absolute terms. For relative measurement of education inequality, a relatively new indicator, developed by evaluating income Gini concept, education Gini should be preferred.

2.2.1.1. Education inequality in Turkey

Turkey is one of the developing country suffers from higher education inequality problem. This means that Turkey cannot use its whole capacity of human capital accumulation. Redistribution of education opportunities is necessary to remove idle capacity problem of Turkey.

Average years of schooling data of some countries obtained from United Nations Development Programme (UNDP) which is one of the data that is used for education Gini calculation to present general view of Turkey and its status in the world. Since the orginal database shows the same average years of schooling value for the countries after 2010, cross section analysis of values is shown in the Figure 2.5 for only 2010. While most of the developed countries such as United States, Switzerland, Japan and Finland have very high average years of schooling with more than 10 years, Turkey has relatively lower value, 7,8. This result shows how insufficient is the education level of Turkey. In this sense, the Gini index which is calculated with average years of schooling data is expected to be higher in Turkey and the other countries which have lower average years of schooling values.

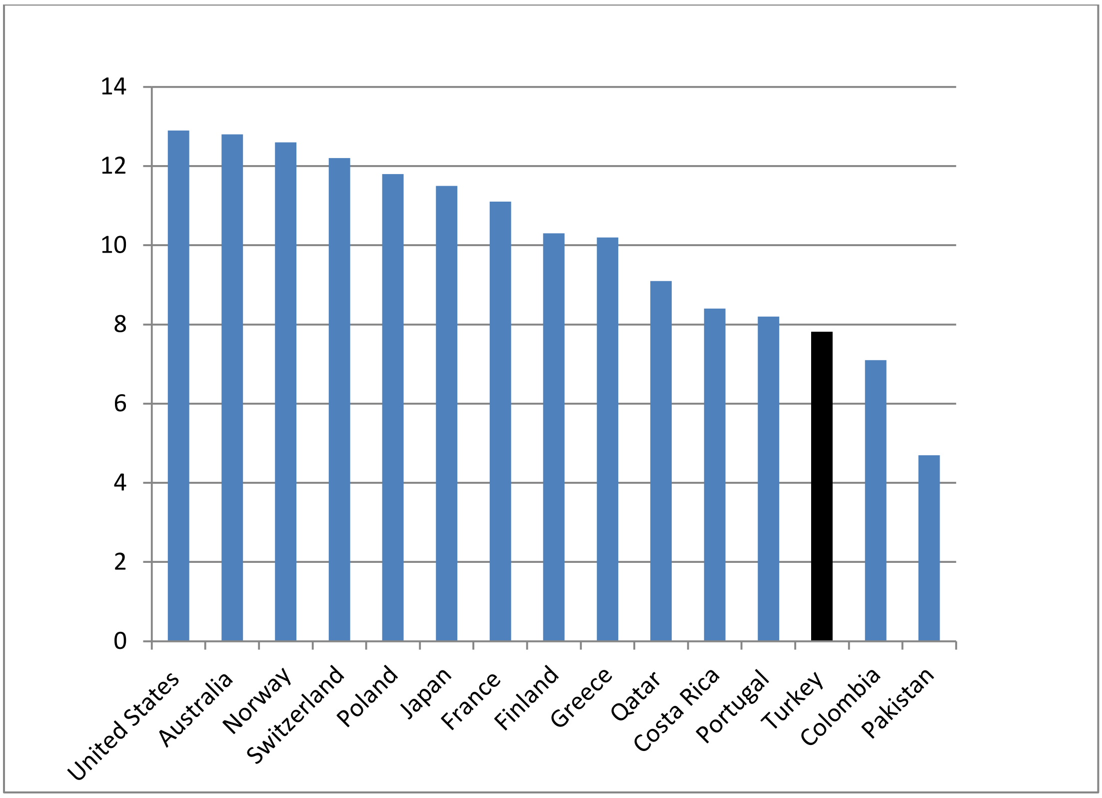

Figure 2.5: Average years of schooling of some countries
Source: UNESCO Institute for Statistics

2.2.2. Educational Gini coefficient

Education Gini index is the most commonly used inequality measurement recently. It is mainly evaluated from income Gini concept. It ranges from 0 which shows the perfect equality to 1 that means perfect inequality.

Education Gini coefficient can be calculated using different kinds of data according to availability of data types. Although average years of schooling is the most commonly used data to compute education Gini index, attainment level of schooling and number of years of schooling data are also appropriate for Gini calculation. Because of the lack of proper data in Turkey, completed level of education over aged 15 macro data obtained from TURKSTAT is chosen to be convenient for application of Gini index (Földvari and Leeuwen, 2014, Thomas *et al.*, 2000 and Checchi, 2001).

Similarly, the calculation methods of income Gini can be applied to education Gini index, but the main way preferred in this thesis is based on the Thomas *et al.* (2000)'s calculations.

The education Gini formula used in this thesis is given in the following:

$$G_E = \left(\frac{1}{\mu}\right) \sum_{i=2}^{n} \sum_{j=1}^{i-1} p_i \left| y_i - y_j \right| p_j \qquad \textbf{(2.3)}$$

where G_E is the education Gini coefficient based on completed levels of education aged over 15 data, μ is the average years of schooling for the concerned population, p_i and p_j are proportions of population with certain level of education, y_i and y_j are the years of schooling at different completed levels of education and n is the number of levels in data.

The number of levels can change according to available data types but it is generally preferred in 4 levels in economics literature: no-schooling, primary, secondary and higher education (Földvari and Leewen, 2014, Castello and Domenech, 2002 and Checchi, 2001). But in this thesis, 8 levels of education are preferred to make use of the available data which includes all levels separately. The levels are illiterate, literate without diploma, primary school (5 years), secondary school (8 years), high school, university degree, master and doctorate.

2.3. Literature Review

In this part of the thesis, the previous studies in the economics literature related to inequality measurements and the analysis of the relationship between education inequality and income inequality will be introduced. The impact of education inequality on income distribution will be presented from perspective of economics literature. The literature background of the method used in econometric analysis part of the thesis will be represented.

The measurement of inequality has different methods in the economics literature. While Gregorio and Lee (2002) use the standard deviation of the education distribution among those over 15 years of age to measure educational inequality, Park (1996) prefers the relative dispersion of educational attainment. However, the educational Gini is the most commonly used coefficient in recent studies (Checchi, 2001, Castello and Domenech, 2002 and Földvari and Leeuwen, 2014). An educational Gini is evaluated from the income Gini concept and calculated with different kinds of data such as average years of schooling, attainment level of schooling or years of schooling according to availability. For Turkey, education Gini coefficient is generally calculated by using completed education levels data (Yanık, 2004, Güngör, 2010 and Tomul, 2011) . The educational Gini is the most convenient coefficient used in recent economics literature because it is very easy to evaluate and it reflects inequality in a strong way using levels of education.

The relationship between education and income is widely-studied area in economics literature especially after realization of significant effect of human capital on output and income. Lopez, Thomas and Wang (1998) search for the effect of human capital on GDP growth for 12 countries between 1970 and 1994 and they consider average schooling as human capital. The results show that under free market conditions, an increase in the human capital stock promotes the growth but in closed or semi-closed economies, the impact of average schooling is zero or barely significant. Castello and Domenech (2002) try to explain this effect of education on income with using the human capital Gini which is calculated with schooling years aged over 15 and growth of per capita income for 108 countries over five-year intervals from 1960 to 2000. This paper proves that human capital inequality has a negative effect on economic growth rates as expected and it decreases the acceleration of growth of economies.

O'Neill (1995) handles the relationship between education and income from a different aspect. He especially focuses on the impact of convergence in education levels on income inequality with lagged version of school enrollment ratios and as alternative average years of schooling as a measurement of human capital. The paper shows that in Europe and developed countries, convergence in education levels leads a decrease in income inequality. But the same pattern does not emerge for world as a whole. The results imply that industrialized countries keep ahead of less developed countries and the whole world. On the other side, like converting income Gini into educational Gini, Thomas, Wang and Fan (2000) also convert Kuznets Curve into the Education Kuznets Curve. The relationship between standard deviation of schooling and average years of schooling is searched and an inverse U-shape is obtained.

Apart from the article of Thomas, *et al*, Kuznets Curve is applied in different concepts with different data in recent studies. A different approach to Kuznets Curve is to add educational inequality measures to the income Kuznets Curve. Park (1996) does this analysis for 59 countries with cross-section data. Income Kuznets Curve is constructed with income Gini and relative dispersion of educational attainment is added to the analysis as an educational inequality measure. The results show that the analysis with educational variables suffers from lack of robustness. The inverse U-shape is obtained when only income variables are used, but when education variables are added, it is not valid anymore. But the important parts of the results are higher level of education attainment which is measured by average years of schooling results with more equally distributed income distribution and income inequality gets greater when the dispersion of education attainment becomes larger among labor force.

The main purpose of the thesis is to look at the relationship between income inequality and education inequality which is another way of explaining the impact of education on income. Different forms of analysis are done to display this effect with Gini coefficients of both income and education. Gregorio and Lee (2002) examine the relationship between income distribution and education inequality with income Gini and standard deviation of educational distribution over 15 data as inequality measurements and they also add educational attainment data to see the effect on the income inequality. A panel data analysis with data for large range of countries from 1965 to 1990 shows that the higher education attainment leads to more equal income distribution and it lowers the income Gini coefficient. In addition, reduction in educational dispersion by one standard deviation decreases income inequality by 0.02. This means education inequality is positively correlated with income Gini index like Park's (1996) results which is computed with the dispersion of education attainment data as inequality measurement.

On the other hand, Földvari and Leeuwen (2014) do this analysis for a large period of time between 1870 and 2000. A Kuznet-type relationship is constructed with using only Gini coefficients and unexpected results are obtained. The data is separated into two categories as before 1950 and after 1950 and the categories are analyzed one by one. The results of two categories are very different from each other. Before 1950, there is a positive relationship between educational Gini and income Gini coefficients and inverted U-curve is obtained. After 1950, the relationship changes into normal U-curve and the results show a negative relationship between two Gini variables. This change in the direction is explained as a result of increased skill premium caused by an increase demand for skill after 1950s.

Different combinations of analysis done using inequality measurements are searched to determine the relationship between education and income. Similar to the articles of Földvari and Leeuwen (2014) and Gregorio and Lee (2002), the correlation between the distribution of education and the distribution of incomes is searched in Checchi's article with education Gini index (2001). The panel data analysis is constructed on the educational and income Gini coefficients of 94 countries calculated from 1960 to 1995. The article concludes that education inequality is negatively correlated with income inequality. An increase in the education inequality causes a decrease in income inequality and results with a more equally distributed income distribution. Furthermore, the analysis indicates that there is a U-shaped relationship between income inequality and average years of schooling with a turning point at 6.5 years. An average increase in education by one year in the population reduces the income Gini by more than 1 point.

Turkish literature about education inequality has relatively recent studies. Education inequality is generally measured by using education Gini coefficient and analyzed in the regional level. These researches show that east part of the country has more unequal education distribution, while western regions have lower education distortions. On the other side, gender differences in education inequality is another major problem that becomes one of the main research subject in the literature. The results of the studies indicates that there is a huge gap between female and male education inequalities. The average years of schooling is relatively much lower for female than male, especially in eastern regions of Turkey. In contrast, the results of analysis present higher education inequality for this group (Yanık, 2004 and Tomul, 2011).

Yanık-İlhan and Aydıner-Avşar (2013) analyze the education inequality among working age population (aged 15-64) in Turkey with a birth-cohort analysis for the period between 1988 and 2011. They use the Household Labor Force Survey for cohort analysis and calculated average years of schooling and education Gini coefficient in provincial level for this period. The results show the huge gap between female and male education inequality and average years of schooling in Turkey. Male population has higher education attainment and lower education inequality than female in the working age population. The cohort analysis results indicate that education inequality is lower for younger birth cohorts for all age groups. The inequality gap between younger and older cohorts is not large for men, contrary to this, it gets higher for women.

The relationship between human capital inequality which is defined as education inequality and measured with education Gini coefficient and economic growth is analyzed for the provinces of Turkey for the period between 1975 and 2000. The completed schooling level data obtained from Census of Population is used for the calculation of education Gini coefficient. The econometric results implies the negative effect of education inequality on economic growth in the provinces with lower education Gini indexes. However, in the provinces that has more equal education distribution, education inequality has positive impact on output growth. This shows the U-shaped relationship between these two variables (Güngör, 2010).

The econometric methods which are suitable for the dataset used in this thesis are searched in the econometrics literature. Since the dataset is small panel with 6 years and has reverse causality and endogeneity problem, dynamic panel methods are thought to be applicable. An earlier study about the estimation of dynamic model is published by Arellano and Bond (1991). They estimated a dynamic model with both generated and real panel data by GMM. The dataset contains a sample of United Kingdom companies. 140 manufacturing firms' unbalanced panel data for the time period between 1979 and 1884 is analyzed and a model for employment is applied. The empirical results imply that GMM estimators have a smaller downward bias than OLS and within-group estimations in Monte Carlo simulation which is applied for 100 units, 7 time-periods and 2 parameters. The variances of GMM estimators are also smaller. In the employment model, GMM estimation suited well than the other estimation models. The only problem of the estimation technic is downward bias of the standard errors that is observed in both Monte Carlo simulation and employment model.

Blundell and Bond (1998) studied on the efficient initial conditions and moment restrictions of dynamic models by constructing a Monte Carlo study with 1000 Monte Carlo replications. They used the same dataset with Arellano and Bond (1991). The results of Monte Carlo simulation indicate great downward bias and imprecise estimates for first-differences generalized method of moments (GMM). But the system GMM estimates have smaller bias and improved sensitivity. In addition, the coefficient of lagged dependent variable is also estimated higher with system GMM method than that with first-differenced GMM. Since the variances of system GMM estimators are lower in this simulation, the article concludes that system GMM estimators are more efficient than non-linear GMM estimators.

Another research from Blundell and Bond (2000) works on the estimation of a Cobb-Douglas production function using panel data for 8 years period. The data of 509 US manufacturing firms which invest Research and Development departments is collected for this study. The system GMM generates higher valued and better determined estimators of dependent variable than first-differenced GMM. Also system GMM accepts additional instruments as valid and illuminating variables. In this concept, system GMM is more appropriate for applications of dynamic models.

An empirical study that searches for the reasons of difference between R&D decisions of German and British firms by applying both static and dynamic models for each country. More than 200 R&D performing firms' data from both Germany and United Kingdom for the time period between 1987 and 1996 is used in this research (Bond, Harhoff and Van Reenen, 2003). In contrast to OLS, within group estimation and first-differenced GMM methods, the system GMM fits well with the specification of the Cobb-Douglas production function containing R&D expenditures as third input in addition to labor and capital.

Windmeijer (2005) searched for the solution for the download bias of the standard errors of two-step GMM estimators. A Monte Carlo simulation is created with 10000 replications by applying two different time periods, T=4 and T=8. In this concept, the coefficients of lagged dependent variables are estimated %50 larger with system GMM than differenced GMM. This means that system GMM estimator fixes the downward bias of the differenced GMM estimator by using more instruments which improves the efficiency of model. Also, the standard errors of system GMM estimators seems smaller than differenced GMM model estimators, this proves that system GMM estimators have higher efficiency.

3. DATA AND METHODOLOGY

3.1. Data

In this section, the data employed in the thesis will be introduced. Descriptive features of the data will be explained using tables and figures together with the sources and the measurement technics.

A panel dataset including income Gini and education Gini indexes, per capita value added, labor force participation ratio, number of students in secondary school, population, total budget expenditure of public authorities on education and some demographic variables such as male population ratio, crude divorce rate, number of births, crude marriage rate and crude suicide rate, will be used in the thesis. Tables A.1 to A.6 in the Appendix A provides the descriptive statistics for each variable. While income Gini index is in Nuts1 level and per capita value added data is in Nuts2 level, the other variables are in provincial level. The data which are in Nuts1 or Nuts2 level are expanded to provincial level by considering the region the provinces belong to.[1] The dataset contains yearly macro data between 2008 and 2013 and there are officially 81 provinces in that period of time in Turkey.

The income Gini is the first inequality measurement to introduce. It is calculated by Turkish Statistical Institute (TURKSTAT) in Nuts1 level (12 regions). The income Gini macro data between 2008 and 2013 is used in this thesis. Since the other data used in model is in provincial level, income Gini data is expanded to provincial level by using the same Gini value for the provinces that belong to the region categorized in Nuts1 level.

When the descriptive statistics of income Gini coefficient given below in the Table 3.1 is analyzed, it can be seen that income Gini tends to decrease year by year.

[1] Extension from Nuts levels to provincial level is done by considering the regions that provinces belong to. Each province has the same Gini value as the other provinces which they categorized in the same region with. For example, if the West Marmara region in Nuts1 level has 0.337 income Gini value, all the provinces located in this region such as Tekirdağ, Edirne, Kırklareli, Balıkesir, Çanakkale, are thought to have the same value, 0.337.The same process is applied to the provinces in the Nuts2 level.

While North East Anatolia region has the highest income Gini value, 0.436, in 2008, Mediterranean region has relatively lower but still the highest value, 0.399, in 2013. It is difficult to say which exact region has more unequal income distribution from the Figure 3.1, but Mediterranean and East Anatolia regions have the greater income inequality levels in general. However, the most important part of the statistics is that although the average income Gini has a downward trend, there is an increase in the average value of it in 2009. This increase may be the result of 2008 economic crisis which affected the whole world's economy.

Income Gini index is generally high in the east part of the Turkey. East Anatolian regions' income distribution has larger distortions than Marmara and Black Sea regions. Central Anatolia fluctuates on the average board. The average Gini index of Turkey which is bold black line seems relatively higher. This may be caused by the effect of Marmara region which is more populated region than others. Selim *et al*'s (2014) report also claims that İstanbul has the highest welfare level between 2005 and 2010. This explains why it is the most significant migration-receiving region in Turkey. The earlier study of Başlevent and Dayıoğlu (2005) also mentions the within-region decline of income inequality in İstanbul between 1994 and 2003. But the Figure 3.1 shows the increase of income Gini of İstanbul after 2009.

Table 3.1: Descriptive statistics of income Gini coefficient by year

Income Gini	Observation	Mean	Standard Dev	Min	Max
2008	81	0,376	0,030	0,331	0,436
2009	81	0,390	0,019	0,359	0,415
2010	81	0,375	0,028	0,327	0,417
2011	81	0,373	0,033	0,326	0,427
2012	81	0,366	0,026	0,309	0,407
2013	81	0,359	0,029	0,315	0,399

Source: TURKSTAT Income and Living Conditions Survey

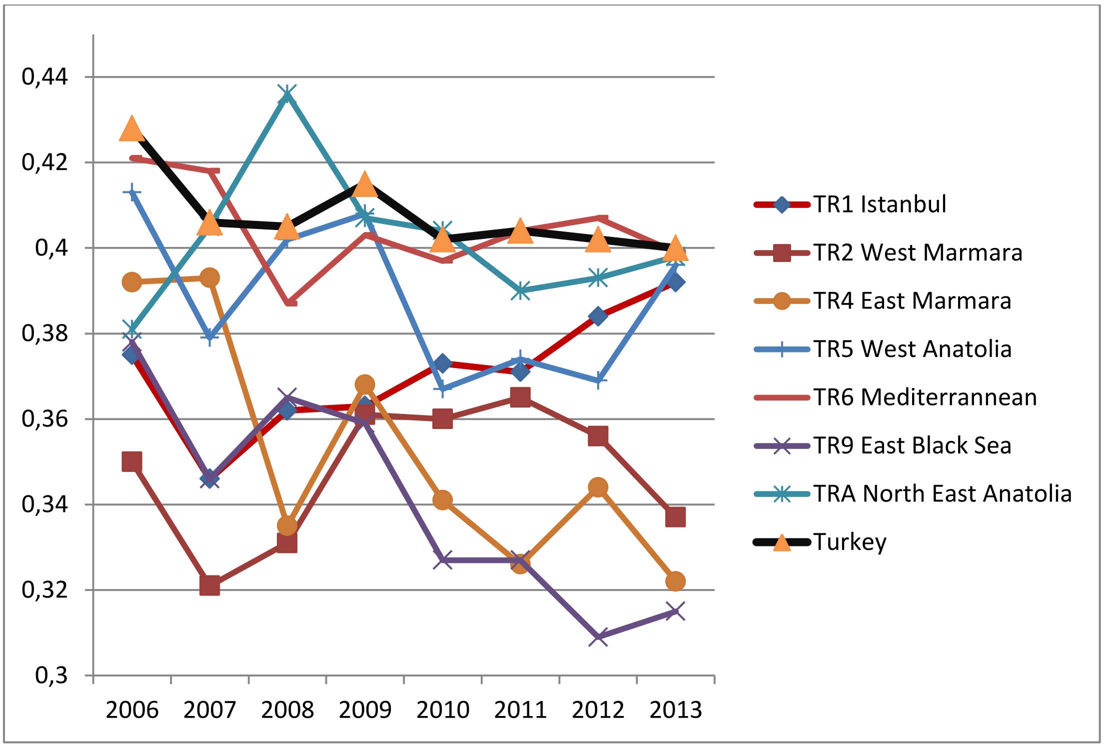

Figure 3.1: Income Gini of Nuts1 level from 2006 to 2013
Source: TURKSTAT Income and Living Conditions Survey

As a measurement of inequality in education, educational Gini coefficient is calculated. Different data types are used for calculation of educational Gini in economics literature. While Földvari and Leeuwen (2014), Castello and Domenech (2002) and Checchi (2001) used average years of schooling data in 4 levels of education (namely; no-schooling, primary, secondary and higher education), Thomas *et al.* (2000) prefer attainment level of schooling data in 7 levels, no-schooling, partial primary, complete primary, partial secondary, complete secondary, partial tertiary and complete tertiary education for Gini calculation. Moreover, Castello and Domenech (2002) use the data of population aged over 15. For Turkey, Yanık-İlhan (2004) calculated education Gini coefficient by using Census of Population data in 6 levels for 1975, 1980, 1985, 1990 and 2000 in provincial level. Another research is calculated education Gini coefficient by using Census of Population data for individuals aged over 25 for the period between 1975 and 2000 in provincial level (Tomul, 2011). In this study, education Gini index is calculated for female and male seperately and gender differences in education inequality are analyzed. The education levels that are preferred in calculation are illeterate, literate without diploma, primary school, secondary school, high school and university degree. Because of the difficulties in finding data, educational Gini coefficient used in this thesis is computed using completed education levels for population aged over 15. Data is again obtained from TURKSTAT for the periods between 2008 and 2013.

The calculation of educational Gini is based on the calculation method of Thomas *et al* (2000). The formula for the calculation is given below:

$$G_E = \left(\frac{1}{\mu}\right) \sum_{i=2}^{n} \sum_{j=1}^{i-1} p_i \left|y_i - y_j\right| p_j \qquad \textbf{(3.1)}$$

where G_E is the Educational Gini, μ is the average years of schooling for the concerned population, p_i and p_j stand for the proportions of population with given levels of schooling, y_i and y_j are the years of schooling at different education levels, n is the number of levels/categories in schooling data. In this thesis, the number of levels in education is provided in 8 categories based on the data obtained from TURKSTAT. Those levels are illiterate, literate without diploma, primary school (5 years), secondary school (8 years), high school, university degree, master and doctorate.

Descriptive statistics of educational Gini coefficients for each region is provided in Table 3.2 below. As it can be seen from the table educational Gini has a downward trend between 2008 and 2013. The average of educational Gini for the period of analysis is 0.334. The downward trend of the educational Gini can also be understood by looking at the average values for each year. While the average educational Gini value is 0.367 in 2008, it decreases to 0.312 in 2013. However, as it can be seen from the Figure 3.2, some provinces generally located in Eastern and Southeastern Anatolia regions like Van, Şanlıurfa, Siirt, Muş, Diyarbakır, Ağrı have very high Gini value for education. In 2008, Şırnak has the highest Gini coefficient with its 0.536 value. On the other hand, in 2013, Ağrı gets the highest value, 0.4114. Furthermore, the lowest values of educational Gini generally belong to provinces in Marmara and relatively Central Anatolian regions such as Ankara, Bilecik, Bursa, Eskişehir, Kırklareli, Konya, Sakarya, Yalova. In 2008, Eskişehir has the most equally distributed education distribution with the lowest Gini value 0.288. But Ankara, the capital city of Turkey, has the lowest educational Gini value in 2013. When İstanbul, the most populated city in Turkey, is analyzed, the educational Gini of it seems relatively lower than the average value of the country with its average value of 6 years, being 0.286. Even, in 2013, it has one of the lowest values, 0.273. These results are in line with Tomul (2011)'s findings for the period between 1975 and 2000. The article finds out that western regions of Turkey (İstanbul, Ankara, İzmir, Bursa) have the lowest education inequality in contrast to eastern regions (Mardin, Şanlıurfa, Ağrı, Van, Gaziantep) in these years.

Table 3.2: Descriptive statistics of educational Gini by year

Education Gini	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
2008	81	0,367	0,064	0,288	0,536
2009	81	0,359	0,058	0,287	0,511
2010	81	0,332	0,045	0,270	0,447
2011	81	0,320	0,039	0,265	0,420
2012	81	0,314	0,037	0,262	0,413
2013	81	0,313	0,036	0,261	0,411

Source: Author's calculation based on ABPRS data

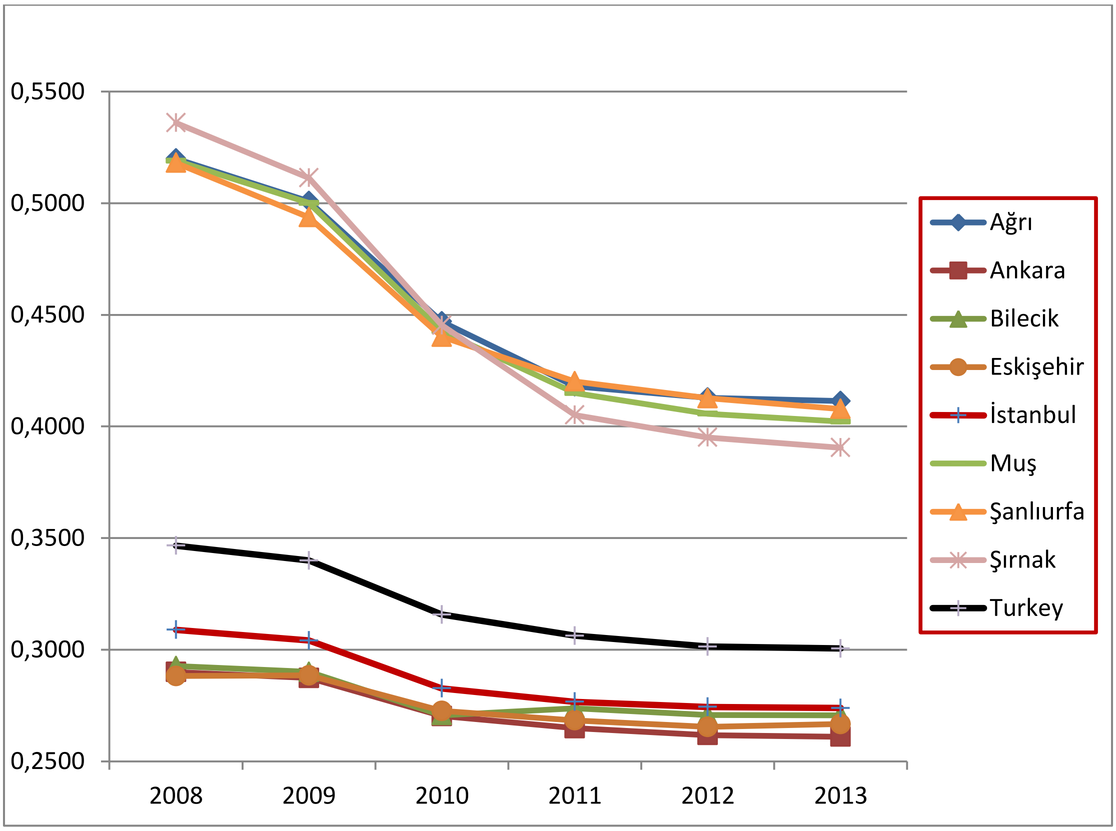

Figure 3.2: Educational Gini of some provinces between 2008 and 2013
Source: Author's calculation based on ABPRS data

Although per capita income is one of the most significant data used in the literature such as Park (1996), Gregorio and Lee (2002) and Checchi (2001); in this thesis, per capita value added data is used as a proxy instead of per capita GDP, because per capita GDP is not available at any statistics center in Turkey. However, per capita value added is only available for the years from 2008 to 2011 from TURKSTAT. Therefore, the data of 2012 and 2013 are not included in the dataset. Per capita value added is measured in million TL and the sectorial discrimination is not included in data. Thus, it only shows the total per capita value added in all of the sectors. The original value added data is in Nuts2 level (26 regions). For that reason the data is expanded to provincial level by using the value for the provinces depending on the Nuts2 region they belong to.

Labor force participation ratio is another variable which is thought to effect income (see Lopez, Thomas and Wang, (1998) and O'Neill, (1995) among others) and; therefore, it is included in the data assuming that it has an impact on income Gini as well. The data for this variable is in provincial level for the period of interest and it is provided in percentages in the TURKSTAT's website.

While Castello and Domenech (2002) add the initial total years of schooling of population over 15 to their dataset, Park (1996) and Gregorio and Lee (2002) choose the educational attainment values for analyzing the effect of educational attainment on income Gini. By following the literature given above, number of students in secondary school is added to dataset in provincial level.[2]

Social expenditure is another significant issue that can affect income inequality. Income distribution seems more equally distributed in regions where social expenditure of government is larger (Gregorio and Lee, 2002). Instead of social expenditure, budget expenditure of local authorities on education is added to the dataset. It is measured in thousand TL and this data is obtained from Ministry of Finance, General Directorate of Public Accounts from 2008 to 2013. Per capita budget expenditure is preferred to remove collinearity problem between population and budget expenditure.

---

[2] Although investment share in GDP is a very important data that must be included into the dataset because of its great effect on both income and inequality, finding any data about investment or even a proxy for investment is very difficult in Turkey (O'Neill, 1995, Castello and Domenech, 2002). Because of this difficulty, it is not added to the dataset.

Some demographic variables of provinces such as male population ratio, crude divorce rate, number of births, crude marriage rate and crude suicide rate are included in the dataset in provincial level to avoid the omitted variable bias as much as possible in addition to control for some province specific factors other than those that will be captured by province level fixed effects. All of these data are from TURKSTAT from 2008 to 2013 at province level.

3.2. Model

In this section, the econometric model that is chosen according to the features of dataset introduced above is explained. The reasons that are considered when deciding to apply the model are explained.

The analysis is mainly constructed to find the impact of education inequality on income distribution. Higher education attainment and income are two different concepts which actually affect each other (Bils and Klenow, 2000). Higher education attainment is directly related with education inequality, because if the education attainment gets higher, this would have an equalizing effect on education distribution. Therefore, it can be inferred that education inequality and income inequality also affect each other. This means there is a reverse causality between dependent variable, income Gini index, and explanatory variable, education Gini.

The reverse causality of education and income inequality measurements creates simultaneity problem in the model. Simultaneity causes the violation of the assumption that a variable must be uncorrelated with error term in a statistical model, zero conditional mean assumption and the variable violates this assumption is called endogenous variable. In this sense, specification an instrument variable that is uncorrelated with error term but highly correlated with endogenous explanatory variable is the best way to get rid of simultaneity problem (Baum, 2006).

The dataset consists of panel data for 6 years, so that it is called short panel. The applied method should consider the short panel feature of the data. As a result, all these characteristics of the panel dataset are analyzed and the econometric models that are applied in the thesis are chosen to cover these features step by step.

The empirical analysis starts with basic static models of panel data methods. Pooled OLS and modified for yearly and regional fixed effects OLS are the simplest models. Then fixed effects and random effects models in static form are applied to the dataset. To control the serial correlation between dependent variable and error terms, the analysis is continued by using dynamic models. Dynamic OLS, fixed effects, and first-difference methods are estimated. Since there is a reverse causality between income and education Gini indexes, Anderson-Hsiao and Arellano-Bond models are preferred as the most appropriate methods to eliminate simultaneity and serial correlation problems of the data.

3.2.1. Static models

3.2.1.1. Pooled OLS

This part of the analysis begins with a basic model which ignores the panel structure of the data. Each observation is thought as a cross section data and applied the OLS method. The pooled OLS panel data regression is:

$$y_{it} = \beta x_{it} + u_{it} \qquad \textbf{(3.2)}$$

where t=1,…,T and I=1,…,N.

A pooled regression assumes that there exists neither a correlation across individuals, nor across time periods for any individual. So that, this would ignore the individual effect which generates correlation between error terms for each individual i. Another assumption of the method is that errors are homoscedastic.

Pooled regression does not make the best use of the data. Therefore, under appropriate conditions that error term is uncorrelated with regressors, pooled estimation gives unbiased but inefficient results. However, if error term is correlated with independent variables, the results are also biased (Johnston and Dinardo, 1997).

3.2.1.2. Fixed effects model

The Fixed Effects method modestly relaxes the assumption that the estimation function does not change over time and space. In this model, each cross-sectional unit can have its own constant term, but the slope estimate does not vary across individual (Baum, 2006).

The fixed effects model is mainly based on the idea of removing the unobserved effect and time-invariant explanatory variables. Consider a simple model:

$$y_{it} = \beta x_{it} + \alpha_i + u_{it} \quad \textbf{(3.3)}$$

where t=1,…,T and I=1,…,N. When the equation (3.3) is averaged for each i:

$$\bar{y}_i = \beta \bar{x}_i + \alpha_i + \bar{u}_i \quad \textbf{(3.4)}$$

where $\bar{y}_i = 1/T \sum_{t=1}^{T} y_{it}$ and so on. When the equations (3.3) and (3.4) are subtracted for each t, the fixed effects transformation is obtained:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i \text{ for } t=1,\dots,T$$

or,

$$\ddot{y}_{it} = \beta \ddot{x}_{it} + \ddot{u}_{it} \text{ for } t=1,\dots,T \quad \textbf{(3.5)}$$

Here, $\ddot{y}_{it} = y_{it} - \bar{y}_i$, $\ddot{x}_{it}$ and $\ddot{u}_{it}$ are time-demeaned data on *y*, *x* and *u*. Since the unobserved effect, α_i is removed from the equation, the pooled OLS method would be used to estimate the model. The pooled OLS estimator of time-demeaned variables are called fixed effects estimator. The model also can be constructed by adding extra explanatory variables:

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk} + \alpha_i + u_{it} \text{ for } t=1,\dots,T \quad \textbf{(3.6)}$$

The time-demeaning version of this equation is given below for each individual *i*:

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it1} + \beta_2 \ddot{x}_{it2} + \dots + \beta_k \ddot{x}_{itk} + \ddot{u}_{it} \text{ for } t=1,\dots,T \quad \textbf{(3.7)}$$

which is estimated by pooled OLS again.

The fixed effects model only uses the within (over time) variation, so that the model must be constructed by using variables with sufficient variation over time. The model can only estimate coefficients on time-varying regressors. The homoscedasticity and serially uncorrelatedness assumptions of pooled OLS are also valid for fixed effects estimation. Pooled OLS method provides consistent estimators of within-transformed data. If the idiosyncratic error u_{it} is uncorrelated with the explanatory variables, then the fixed effects estimator is unbiased (Wooldridge, 2012).

3.2.1.3. Random effects model

Instead of the fixed effects model that consider the individual-specific intercept as fixed effect of that individual, random effects model assumes that α_i is random and uncorrelated with all other regressors (Cameron and Trivedi, 2010 and Baum, 2006)

Random effects method is begun with the same unobserved effects model like fixed effects:

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_k x_{itk} + \alpha_i + u_{it} \qquad \textbf{(3.8)}$$

But here a constant term $\beta_0$is added to the equation to make zero mean assumption about α_i. The given equation above becomes random effects model if the unobservable effects are uncorrelated with each independent variables:

$$Cov(x_{itj}, \alpha_i) = 0 \text{ , } t=1,\ldots,T; j=1,\ldots,k \qquad \textbf{(3.9)}$$

Random effects assumptions include all the assumptions of fixed effects model and additionally the strict assumption that α_i is uncorrelated with all the independent variables in all time periods.

The given equation of random effects can be estimated by pooled OLS method but this method ignores the composite error term characteristic of the model. The error term is defined as $v_{it} = \alpha_i + u_{it}$. Since α_i is a component of the error term for each period, v_{it} is serially correlated across time. For this reason, generalized least square (GLS) method can be used to solve the serial correlation problem of random effects. GLS method gives better results especially when the panel data is short panel with large sample. To transform the equation, the constant $\theta$is defined as:

$$\theta = 1 - [\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}]^{1/2} \qquad \textbf{(3.10)}$$

which have values between 0 and 1. By using this constant, the equation is transformed to the form given below:

$$y_{it} - \theta\bar{y}_i = \beta_0(1-\theta) + \beta_1(x_{it1} - \theta\bar{x}_{i1}) + \cdots$$

$$+\beta_k(x_{itk} - \theta\bar{x}_{ik}) + (v_{it} - \theta\bar{v}_i) \quad \textbf{(3.11)}$$

Here over bars symbol the time average values of variables. This equation involves the quasi-demeaned data on each variable (Wooldridge, 2012).

The advantage of random effects model is that it allows the time-invariant explanatory variables to be in the estimation. Therefore, the marginal effects of all variables can be estimated by this model. Furthermore, if the random effects assumptions are held, the estimators are consistent but not unbiased. On the other hand, random effects model estimates inconsistent coefficients if fixed effects model is appropriate for the data (Cameron and Trivedi, 2010 and Wooldridge, 2012).

3.2.2. Dynamic models

3.2.2.1. Dynamic pooled OLS

Dynamic version of pooled OLS model has the same characteristics like static pooled OLS except it involves the lagged value of dependent variable, y_{it}. The new model is:

$$y_{it} = \rho y_{i,t-1} + \beta x_{it} + u_{it} \quad \textbf{(3.12)}$$

The dynamic pooled OLS which is the simplest model of dynamic panel data methods has the same assumptions as static pooled OLS. As it is mentioned above in the pooled OLS subsection, this method estimates inefficient coefficients.

3.2.2.2. Dynamic fixed effects model

The dynamic fixed effects transformation of simple model with first lagged of dependent variable can be shown as:

$$y_{it} - \bar{y}_i = \rho\left(y_{i,t-1} - \bar{y}_i\right) + \beta_1(x_{it1} - \bar{x}_{i1}) + \cdots + \beta_k(x_{itk} - \bar{x}_{ik}) + u_{it} - \bar{u}_i$$

or,

$$\ddot{y}_{it} = \rho\ddot{y}_{i,t-1} + \beta_1\ddot{x}_{it1} + \beta_2\ddot{x}_{it2} + \cdots + \beta_k\ddot{x}_{itk} + \ddot{u}_{it} \text{ for } t=1,\dots,T \quad \textbf{(3.13)}$$

The dynamic fixed effects model is also estimated by pooled OLS method and it uses the same assumptions like static fixed effects model.

3.2.2.3. Dynamic random effects model

Similarly, a dynamic random effects model is involves the lagged values of dependent variable. The modified equation of the model is:

$$y_{it} = \beta_0 + \rho y_{i,t-1} + \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_k x_{itk} + \alpha_i + u_{it} \qquad \textbf{(3.14)}$$

The fixed effects assumption and the uncorrelatedness of unobserved effects with all regressors are also valid for dynamic random effects model. The equation is estimated by GLS method as static random effects.

3.2.2.4. Dynamic first-differenced model

In the dynamic panel methods, the lagged dependent variable is correlated with error terms, especially in the within transformations. This correlation creates bias in the estimator of lagged dependent variable which does not decrease by increasing *N*, the number of individuals. In addition, if the independent variables are correlated with the lagged dependent variable, their estimators would be affected by the bias. Since u_i error component is included in every value of dependent variable in random effects model, the lagged dependent variable is correlated with composite error term. Therefore, random effects model is also affected by Nickell bias (Nickell, 1981 and Baum, 2006)

A possible solution to Nickell bias is first-difference method. This model is based on the idea that differencing the simple equation and subtracting differenced version from original equation. Consider a simple dynamic model equation:

$$y_{it} = \rho y_{i,t-1} + \beta x_{it} + \alpha_i + u_{it} \quad t=1,\ldots,T;\ i=1,\ldots,N \qquad \textbf{(3.15)}$$

The first-differenced transformation equation is:

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \beta \Delta x_{it} + \Delta u_{it} \qquad \textbf{(3.16)}$$

Apart from static model, OLS estimators in dynamic first-differenced method are inconsistent, since $\Delta y_{i,t-1}$ is correlated with lagged error terms. First-differencing method assumes that first difference of idiosyncratic errors is not serially correlated and OLS assumptions are also valid here. In this method, the coefficients of time-invariant independent variables are not identified because of the difference (Baum, 2006, Cameron and Trivedi, 2010 and Wooldridge, 2001).

3.2.2.5. Anderson–Hsiao model

Another method to solve Nickell bias problem is suggested by Anderson and Hsiao (1981). In this model, the simple equation is first-differenced to eliminate the unobserved effect, α_i. Therefore, the equation used in this model is given as:

$$y_{it} - y_{i,t-1} = \rho(y_{i,t-1} - y_{i,t-2}) + \beta(x_{it} - x_{i,t-1}) + (u_{it} - u_{i,t-1})$$

or,

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \beta \Delta x_{it} + \Delta u_{it} \tag{3.17}$$

Anderson-Hsiao model proposes to use $y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$. Thus, the instrument of lagged dependent is not correlated with lagged error terms unless error terms are serially correlated. If the model includes other lagged dependent variables in its structure, they can be used as instruments for themselves.

Anderson-Hsiao model is an instrumental variable estimation and its estimators are consistent but not efficient, since it does not make use of all the available information (Cameron and Trivedi, 2010, Baltagi, 2005 and Anderson and Hsiao, 1981).

3.2.2.6. Arellano - Bond model

Arellano and Bond (1991) propose to use generalized method of moments (GMM) estimation to produce more efficient estimators than Anderson-Hsiao model. They argue that if the orthogonality conditions between lagged values of dependent variable and error terms are taken into consideration, additional instruments can be obtained.

Arellano-Bond model creates a system that is based on the orthogonality of second lagged values of dependent variable and error terms. In each period, an extra instrument is added to the system, so that the set of valid variables at time T becomes $(y_{i1}, y_{i2}, \dots, y_{i,T-2})$. In this sense, the number of instruments differ in each time period. For instance, for t=3, the Arellano-Bond equation becomes:

$$y_{i3} - y_{i2} = \rho(y_{i2} - y_{i1}) + \beta(x_{i3} - x_{i2}) + (u_{i3} - u_{i2}) \quad \textbf{(3.18)}$$

Here, $y_{i1}$is a valid instrument for $(y_{i2} - y_{i1})$ which is uncorrelated with $(u_{i3} - u_{i2})$ if the u_{it} are not serially correlated. In the next step, for t=4, the equation changes into:

$$y_{i4} - y_{i3} = \rho(y_{i3} - y_{i2}) + \beta(x_{i4} - x_{i3}) + (u_{i4} - u_{i3}) \quad \textbf{(3.19)}$$

In this period, y_{i1} and $y_{i2}$are valid instruments for $(y_{i3} - y_{i2})$. Thus, at time T, there are T-2 valid instruments for the equation.

Arellano-Bond model is estimated in a GMM context. It creates consistent and more efficient estimators in this concept (Arellano and Bond, 1991, Baltagi, 2005, Cameron and Trivedi, 2010 and Baum, 2006).

4. ECONOMETRIC ANALYSIS

In this section, econometric analysis about the impact of education inequality on income inequality is presented. Both static and dynamic panel data estimations are shown in the section and the final model and tests are provided.

First of all, all the variables in the models are in logarithmic form. Since all of them have positive values, the dataset can be converted to logarithmic form. In this sense, the coefficients can be estimated as elasticity. This would make easier to interpret the estimators.

The panel data analysis is started with static model estimations. The results of different forms of static models are applied to the data explained in the Data section. Since static models provide only the equitemporaneous effects of explanatory variables on dependent variable, the interpretation of the estimators is done by taking into consideration this characteristic of the model.

The variables used in the models are education Gini coefficient, number of students in secondary school, budget expenditure of local authorities on education, per capita value added, labor force participation ratio and suicide rate as explanatory variables and income Gini coefficient as dependent variable.

Table 4.1: Static Model Estimations

Variable	Pooled OLS	Pooled OLS (2)	Fixed Effects	Fixed Effects (2)	Random Effects
ln(education Gini)	-0.022	0.018	-0.094	-0.094	-0.002
	(0.050)	(0.033)	(0.095)	(0.095)	(0.068)
ln(secschoolstudents)	0.023***	0.002	-0.020	-0.020	0.028***
	(0.003)	(0.002)	(0.044)	(0.044)	(0.005)
ln(percapexpenditure)	0.034**	0.006	0.031	0.031	0.032*
	(0.012)	(0.012)	(0.033)	(0.033)	(0.015)
ln(pcvalueadded)	-0.139***	-0.019	-0.231***	-0.231***	-0.159***
	(0.016)	(0.016)	(0.049)	(0.049)	(0.023)
ln(laborforce)	-0.001	0.093***	0.316***	0.316***	0.122***
	(0.027)	(0.024)	(0.045)	(0.045)	(0.032)
ln(suiciderate)	0.048***	0.014**	0.023***	0.023***	0.032***
	(0.007)	(0.004)	(0.005)	(0.005)	(0.005)
Constant	0.022	-1.195***	0.009	0.009	-0.283
	(0.164)	(0.154)	(0.643)	(0.643)	(0.213)
Region fixed effects	NO	YES	NO	NO	NO
Year fixed effects	NO	YES	NO	NO	NO
N	323	323	323	323	323
r2	0.470	0.861	0.296	0.296	

Note: * p<0.05, ** p<0.01 and *** p<0.001

The first model shown in the Table 4.1 is done by using pooled OLS method. The model specifications in all static models are formed with the robustness option, so that the results are robust. In this model, education Gini coefficient and labor force participation ratio are insignificant. Since the aim of thesis is to examine the impact of education inequality on income inequality, the model must be changed to capture this relationship.

The second model is specified by adding dummies for regions in Nuts1 level and years to first estimation, since income Gini coefficient is originally in Nuts1 level. In this case, education Gini is still insignificant. Moreover, other variables such as number of students in secondary school, per capita budget expenditure and per capita value added turns to become insignificant. However, most of the regional dummies are significant except West Anatolia and Central Anatolia regions compared to region 1, İstanbul.

The within-group regression results are shown in the third column. The education Gini coefficient, number of students in secondary school and per capita budget expenditure on education variables have insignificant estimators in this model. The serial correlation of error terms are tested for this specification. Wooldridge test's null hypothesis is given below:

$H_0 = no\ first\ order\ autocorrelation$

H_0 is rejected for fixed effect estimation and the serial correlation is corrected by clustering the provinces. The forth model in the table gives the results for corrected within-group regression. After preventing the serial correlation in standard errors, it seems that the education Gini coefficient is still insignificant and the correction does not change the results.

The last column of the Table 4.1 represents the results of random effect regression. The education Gini coefficient is not significant in this model. Breusch and Pagan test is applied to check the significance of individual effects. The null hypothesis of the test is:

$H_0: variance\ of\ the\ individual\ effects(u) = 0$

Test result rejects the null hypothesis. This means that individual effects are significant and the homoscedasticity assumption of the random effect estimation is violated.

To decide between fixed and random effects models, Hausman test is run. The null hypothesis of Hausman test is in the following:

$H_0: difference\ in\ coefficients\ not\ systematic$

The Hausman test indicates that the null hypothesis cannot be rejected, so that there are not significant differences between fixed effects and random effects coefficients. Therefore, random effects estimation seems appropriate in this case.

Since there is serial correlation between dependent variable and error term, this violates the assumption of static models. Therefore, dynamic models are used to remove this correlation. By applying dynamic estimations, Nickell bias can also be controlled. The results of the different forms of dynamic models are shown in Table 4.2.

Table 4.2: Dynamic Model Estimations

Variable	Dynamic OLS	Dynamic Fixed Effects	Dynamic Random Effects	First-Differences	Anderson -Hsiao	Arellano -Bond
lagged ln(incomeGini)	0.704*** (0.064)	0.041 (0.070)	0.681*** (0.065)			-0.284*** (0.057)
Lagged diff ln(incomeGini)				-0.135** (0.049)	-0.471*** (0.134)	
ln(educationGini)	-0.092* (0.037)	-0.428*** (0.119)	-0.094* (0.038)			-0.386** (0.118)
ln(secschoolstudents)	0.003 (0.003)	-0.063 (0.054)	0.004 (0.003)			-0.086 (0.050)
ln(percapexpenditure)	-0.012 (0.011)	-0.053 (0.053)	-0.011 (0.011)			-0.059 (0.041)
ln(pcvalueadded)	-0.046** (0.016)	-0.068 (0.087)	-0.050** (0.017)			0.012 (0.086)
ln(laborforce)	-0.099*** (0.023)	0.071 (0.062)	-0.099*** (0.023)			0.101* (0.047)
ln(suiciderate)	0.028*** (0.005)	0.018*** (0.004)	0.029*** (0.005)			0.011*** (0.003)
Year fixed effects	YES	YES	YES			YES
diff ln(educationGini)				-0.407*** (0.115)	-0.490*** (0.142)	
diff ln(secschoolstudents)				-0.063 (0.049)	-0.106 (0.055)	
diff ln(expenditurepc)				-0.056 (0.042)	-0.059 (0.044)	
diff ln(pcvalueadded)				0.022 (0.081)	-0.106 (0.119)	
diff ln(laborforce)				0.092 (0.046)	0.143* (0.057)	
diff ln(suiciderate)				0.012*** (0.003)	0.011** (0.003)	
diff year fixed effects	NO	NO	NO	YES	YES	
Second lagged ln(incomeGini)					-0.447*** (0.058)	
Constant	0.377** (0.124)	-0.501 (0.706)	0.347** (0.128)			-1.409* (0.647)
N	242	242	242	161	161	161
r2	0.746	0.468		0.482	.	

Note: * p<0.05, ** p<0.01 and *** p<0.001.

First of all, the simplest dynamic model, dynamic OLS is applied to the data. First degree lagged value of income Gini variable is added to the models. Pooled OLS estimation results indicates that lagged income Gini variable is significant and has positive effect on income Gini coefficient. Education Gini coefficient is also significant and negatively correlated with income Gini coefficient.

Since dynamic pooled OLS is the basic dynamic panel data method and cannot capture the individual effects, dynamic fixed effects model is applied. In the dynamic fixed effects model, all the variables seem insignificant except the education Gini. Lagged income Gini is also insignificant. Education Gini has negative effect on income Gini in this model.

The dynamic random effects model is given in the third column of the table. The method has nearly the same results with pooled OLS model. While the number of students in secondary school and per capita budget expenditure on education variables are insignificant, education Gini is in negative association with income Gini coefficient. First lagged value of income Gini has positive effect on income Gini.

The fourth model is estimated by using first-differenced method. Differenced values of each variable are included in the model instead of the logarithm of real values. First-differenced model removes the individual effects by differencing it out. In this model, both lagged income Gini and education Gini is significant and have negative impact on income Gini. First-differenced method has correlation between error terms. Therefore, the model is specified as clustered in provinces. The coefficient of education Gini is interpreted as 1% change in education Gini coefficient declines the income Gini index 0.4%, when the other variables remain constant.

Since education Gini index is suspected to be endogenous and there is a reverse causality between education Gini and income Gini coefficients, Anderson-Hsiao model is applied to control both of them. The lagged first difference of dependent variable is an explanatory variable and second lag in levels is used as an explanatory variable. First stage results of Anderson-Hsiao model show that the second lag of the income Gini is a good predictor of the lagged first difference. Thus, the first condition for it to be a valid instrument is satisfies. The instrumental variables regression results indicate that Anderson-Hsiao method estimates higher coefficients compared to first-difference model. However, the direction of the effects of both lagged difference of income Gini and differenced education Gini indices is the same as first-differenced method and they are both negative. In addition, Anderson-Hsiao estimator is consistent but inefficient, as it does not make use of all the available information.

Lastly, Arellano-Bond model is applied to control both serial correlation between income Gini coefficient and error term and endogeneity of education Gini index. Only the lagged income Gini coefficient is added to the model. Education Gini has again negative impact on income Gini and the lagged variable is also negatively correlated with it. The elasticity of education Gini for income Gini is -0.386. This means 1% change in education Gini cause -0.3% change in income Gini when the other variables remain constant. The model also shows that lagged income Gini is significant and elasticity of it is -0.284. When the other explanatory variables are analyzed, the number of students in secondary school, per capita budget expenditure on education and per capita value added variables seem insignificant. The labor force participation ratio is significant and its elasticity is 0.101. This is explained as 1% increase in the labor force participation ratio causes 0.101% increase in income Gini coefficient. The suicide rate variable is also significant and positively correlated with income Gini.

To test the overidentification, Sargan test is applied to Arellano-Bond model. The null hypothesis of the test is:

$H_0: overidentifying\ restrictions\ are\ valid$

Sargan test indicates that H_0 cannot be rejected and this means some of the instruments are not correlated with the error term.

In the econometric analysis, static models are estimated at first. Simple OLS, fixed effects and random effects models are applied. However, because of the serial correlation between income Gini and error term, the analysis continues with dynamic models. These models also provide the control of Nickell bias. Dynamic fixed effects, random effects and first-difference models are estimated. Since it is suspected that education Gini is endogenous variable and there is reverse causality between education Gini and income Gini, Anderson-Bond model is applied. In the next step, Arellano-Bond model is used to control both endogeneity and serial correlation. All these models mentioned above indicate the negative association between education Gini and income Gini coefficients. This proves the robustness of the models. But none of these models are perfectly eliminates the serial correlation and endogeneity. Therefore, the results only indicate the effect of education inequality on income distribution.

5. CONCLUSION

First of all, the main aim of the thesis is to present a general picture of income and education inequality in Turkey, to calculate the education Gini index and to examine the impact of education inequality on income inequality in provincial level for the period between 2008 and 2013.

Income inequality which is measured by using income Gini index obtained from TURKSTAT in Nuts1 level decreases year by year. This means income distribution of Turkey becomes more equal. However, east side of the country has higher Gini coefficients than the other parts. This may be the result of the inadequacy of economic activities in these regions. Besides, Marmara and Black Sea regions generally have lower income Gini coefficients and more equal income distributions.

Education Gini coefficient is calculated to measure education inequality by using completed education level data in provincial level for the period between 2008 and 2013. In this calculation, population separated into 8 education levels: illiterate, literate without diploma, primary education, secondary education, high school, university degree, master and doctorate. Education Gini is also has a downward trend during this period. Even in regions that have the highest education Gini values such as Eastern and Southeastern regions, education distribution tends to become more equal. Marmara and Central Anatolia regions have relatively lower education Gini coefficients, because they have higher schooling rate and most of the universities are located in these regions.

As a final goal, the impact of the education inequality on income inequality is analyzed. Besides the education Gini index, per capita value added, labor force participation ratio, number of students in secondary school, budget expenditure of local authorities on education and demographic variables are included in the model as regressors. All the variables used in the models are in logarithmic form. In the econometric analysis, static and dynamic models are used. Simple OLS, fixed effects and random effects methods is applied in static form. Since there is serial correlation between income Gini and error term, the analysis continues with dynamic models. In this part, OLS, fixed effects, random effects and first-difference methods are estimated. After that, because of the suspicion of reverse causality between education and income Gini coefficients, Anderson-Hsiao model is used. Finally, it is thought that Arellano-Bond model is more suitable for the control of both endogeneity and serial correlation.

All the models mentioned above indicate the negative effect of education Gini coefficient on income Gini. Since all of the estimations show the same direction of effect, it can be said that the results are robust. According to results of Arellano-Bond model, the elasticity of education Gini coefficient for income Gini is -0.284. This means that 1% change in education Gini index decreases income Gini 0.284%, when the other variables remained constant. When the results of the whole Arellano-Bond model are analyzed, however the number of students in secondary school, per capita budget expenditure on education and per capita value added variables are insignificant in the model, the directions of their impacts is mentioned in this section. Number of students in secondary school is negatively correlated with income Gini. It can be said that education attainment has negative effect on income inequality. Per capita budget expenditure on education is also negatively associated with income Gini and the other insignificant variable, per capita value added has positive impact on income Gini. The labor force participation ratio is positively correlated with income Gini index. As the labor force participation ratio increases 1%, the income Gini increases 0.101% as well, when the other variables remain constant. The demographic variable, suicide rate is also significant and has positive effect on income Gini.

The negativity of the impact of education Gini on income Gini gives some ideas about the possible policy suggestions that should be applied in the future. Since the average schooling level of Turkey is very low with 7.1 years in 2008, income distribution has higher distortions. The highly educated people get higher salaries and the others with low education levels earn relatively lower salaries. This conditions increase the income inequality across the whole population. When the labor force becomes more educated, technological innovations would increase and this would accelerate the creation of more skilled jobs. In this sense, more people earn higher salaries and income inequality would decline. Therefore, the major policy concern for decrease of income inequality would be the increase of average schooling of the population. Then the labor force would become more educated and this would provide the decrease of income inequality.

Data constraints make the application of Kuznet Curve in Turkey impossible. GDP per capita data is unavailable in Turkey in provincial level. However, futher researches for the explanation of the main reasons of income and education inequality will be done in the future. Besides, it would be interesting to analyze not only the quantity of the educational investments but also the effect of the quality of education on income inequality if data would be available. This kind of a research would have changed the policy suggestions with respect to decrease in income inequality in Turkey.

REFERENCES

Anderson T. W. and Hsiao C. (1982). Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics.* 18, 47-82.

Arellano M. and Bond S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies.* Vol. 58, No. 2, pp. 277-297.

Awan M. S., Malik N., Sarwar H. and Waqas M. (2011). Impact of Education on Poverty Reduction. Munich Personal RePEc Archive Paper No. 31826.

Baltagi B. H. (2005). *Econometric Analysis of Panel Data (Third Edition)*. New York, NY: Wiley.

Başlevent C. and Dayıoğlu M. (2005). A Household Level Examination of Regional Income Disparity in Turkey. *METU Studies in Development Journal*, 32 (December), 2005, 275-302.

Baum C. F. (2006). *An Introduction to Modern Econometrics Using Stata.* Texas: Stata Press.

Bils M. and Klenow P. J. (2000). Does Schooling Cause Growth? *American Economic Review*, 90(5), 1160-1183.

Blundell R. and Bond S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics.* 87, pp. 115-143.

Blundell R. and Bond S. (2000). GMM Estimation with Persistent Panel Data: An Application to Production Functions. *Econometric Reviews.* 19 (3), pp. 321-340.

Bond S., **Harhoff D. and Van Reenen J.** (2003). Corporate R&D and Productivity in Germany and the United Kingdom. Centre for Economic Performance, London School of Economics and Political Science, London.

Cameron A. C. and Trivedi P. K. (2010). *Microeconometrics Using Stata (Revised Edition).* Texas: Stata Press.

Castelló A. and Doménech R. (2002). Human Capital Inequality and Economic Growth: Some New Evidence. *The Economic Journal.* Vol. 112, No. 478, Conference Papers (Mar., 2002) pp. C187–C200.

Checchi D. (2001). Education, Inequality and Income Inequality. Distributional Analysis Research Programme Discussion Paper. No: DARP 52. The Toyota Centre, London.

Deaton A. (1997). The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. *John Hopkins University Press.* Baltimore and London

Duman A. (2008). Education and Income Inequality in Turkey: Does Schooling Matter? *Financial Theory and Practice.* 32 (3), pp. 369-385.

Földvari P. and Van Leeuwen B. (2014). Educational and Income Inequality in Europe, ca. 1870-2000. *Cliometrica.* 8, pp.271-300.

Gregorio J. D. and Lee J.-W. (2002). Education and Income Inequality: New Evidence from Cross-Country Data. *Review of Income and Wealth.* Series. 48, No. 3.

Güngör N. D. (2010). Education, Human Capital Inequality and Economic Growth: Evidence from Turkey. *Regional and Sectoral Economic Studies,* Vol. 10-2 (2010).

Johnston J. and Dinardo J. (1996). *Econometric Methods (Fourth Edition).* New York: Mc-Graw Hill.

Lin T. **C.** (2004). The Role of Higher Education in Economic Development: An Empirical Study of Taiwan Case. *Journal of Asian Economics,* 15, 355-371.

Litchfield J. (1999). Inequality: Methods and Tools. STICERD, London School of Economics.

Lopez R., Thomas V. and Wang Y. (1998). Addressing the Education Puzzle: The Distribution of Education and Economic Reform. Policy Research Working Paper No: 2031. The World Bank, Washington, D.C.

Nickell S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica.* Vol. 49, No. 6, 1417-1426.

OECD (2014). OECD Factbook 2014: Economic, Environmental and Social Statistics. *OECD Publishing*. OECD, Paris.

O'Neill D. (1995). Education and Income Growth: Implications for Cross-Country Inequality. *Journal of Political Economy.* Vol. 103, No. 6, pp. 1289–1301.

Park K. H. (1996). Educational Expansion and Educational Inequality on Income Distribution. *Economics of education Review.* Vol. 15, No. 1, pp. 51-58.

Roodman D. (2009). How to Do xtabond2: An Introduction to Difference and System GMM in Stata. *The Stata Journal.* 9, No.1, pp. 86-136.

Selim R., **Günçavdı Ö. And Bayar A.A.** (2014). Türkiye'de Bireysel Gelir Dağılımı Eşitsizlikleri: Fonksiyonel Gelir Kaynakları ve Bölgesel Eşitsizlikler. TÜSİAD Report No: TÜSİAD-T/2014-06/554. TÜSİAD, İstanbul.

Thomas V., Wang Y. and Fan X. (2001). Measuring Education Inequality: Gini Coefficients of Education. Policy Research Working Paper No: 2525. The World Bank, Washington, D.C.

Tomul E. (2011). Education Inequality in Turkey: An Evolution by Gini Index. *Education and Science,* 2011, Vol. 36, No 160.

Windmeijer F. (2005). A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimator. *Journal of Econometrics.* 126, pp. 25-51.

Wooldridge J. M. (2001). *Econometric Analysis of Cross Section and Panel Data (Third Edition).* London: MIT Press.

Wooldridge J. M. (2012). *Introductory Econometrics: A Modern Approach (Fifth Edition).* Ohio: South Western Publication.

World Bank (2014). Introduction to Poverty Analysis. Working Paper No: 90288. World Bank, Washington, D.C.

Yanık B. (2004). Türkiye'de Eğitim Eşitsizliği. *Master Thesis,* ITU, İstanbul, MA.

Yanık-İlhan B. and Aydıner-Avşar N. (2013). Education Inequality Among Working Age Population in Turkey: A Cohort Analysis. In *Emerging Patterns of Work and Turkish Labour Market Challenges under Globalization*, Netherlands: Wolters Kluwer.

Zhang X. and Kanbur R. (2001). What Difference Do Polarisation Measures Make? An Application to China. *Journal of Development Studies.* Vol. 37, Issue. 3, pp. 85-98.

Turkish Statistical Institute (2015). < *http://www.turkstat.gov.tr/Start.do*>

United Nations Development Programme Database (2015). <*http://www.undp.org/*>

World Bank Database (2015). < *http://data.worldbank.org/*>

OECD Database (2015). < *https://data.oecd.org/*>

APPENDICES

APPENDIX A **:** Descriptive Statistics

APPENDIX A

Table A.1: Descriptive statistics of all variables in 2008

2008	Obs	Mean	Std. Dev.	Min	Max
IncomeGini	81	0,377	0,031	0,331	0,436
EducationGini	81	0,367	0,064	0,288	0,536
percapitaValueAdded	81	9659	3892	4379	18689
Laborforce	81	47,209	8,616	26,900	66,300
DivorceRate	81	1,163	0,589	0,120	2,690
Births	81	15978	27568	487	225910
MarriageRate	81	9,198	1,344	6,580	13,360
SuicideRate	81	4,496	1,887	1,540	14,330
MalePopRatio	81	0,504	0,012	0,488	0,575
SecSchoolPopRatio	81	0,030	0,007	0,019	0,051
PercapitaExp	81	1,280	0,493	0,714	3,597

Table A.2: Descriptive statistics of all variables in 2009

2009	Obs	Mean	Std. Dev.	Min	Max
IncomeGini	81	0,390	0,019	0,359	0,415
EducationGini	81	0,359	0,058	0,287	0,511
percapitaValueAdded	81	9737	3593	4846	18300
Laborforce	81	48,779	7,746	30,400	65,800
DivorceRate	81	1,313	0,646	0,140	2,740
Births	81	15618	25660	948	210170
MarriageRate	81	8,341	1,166	6,280	11,670
SuicideRate	81	4,346	1,707	1,330	12,980
MalePopRatio	81	0,504	0,011	0,492	0,569
SecSchoolPopRatio	81	0,032	0,008	0,018	0,059
PercapitaExp	81	1,513	0,639	0,800	5,329

Table A.3: Descriptive statistics of all variables in 2010

2010	Obs	Mean	Std. Dev.	Min	Max
IncomeGini	81	0,375	0,029	0,327	0,417
EducationGini	81	0,332	0,045	0,270	0,447
percapitaValueAdded	81	10972	3850816	5575	20149
Laborforce	81	49,409	6,198	31,800	61,800
DivorceRate	81	1,340	0,659	0,140	2,820
Births	81	15534	26008	873	213378
MarriageRate	81	7,969	1,034	5,820	10,740
SuicideRate	81	4,298	2,030	0,610	16,310
MalePopRatio	81	0,504	0,012	0,490	0,570
SecSchoolPopRatio	81	0,035	0,009	0,020	0,068
PercapitaExp	81	1,760	0,825	0,898	6,803

Table A.4: Descriptive statistics of all variables in 2011

2011	Obs	Mean	Std. Dev.	Min	Max
IncomeGini	81	0,374	0,033	0,326	0,427
EducationGini	81	0,320	0,039	0,265	0,420
percapitaValueAdded	81	12639	4671669	5894	23247
Laborforce	81	50,683	6,667	30,600	62,500
DivorceRate	81	1,306	0,658	0,110	2,820
Births	81	15366	25962	891	212241
MarriageRate	81	7,912	0,965	6,090	10,360
SuicideRate	80	3,882	1,586	0,620	11,270
MalePopRatio	81	0,505	0,013	0,490	0,576
SecSchoolPopRatio	81	0,034	0,008	0,021	0,065
PercapitaExp	81	2,027	0,831	1,113	5,948

Table A.5: Descriptive statistics of all variables in 2012

2012	Obs	Mean	Std. Dev.	Min	Max
IncomeGini	81	0,366	0,027	0,309	0,407
EducationGini	81	0,314	0,037	0,262	0,413
percapitaValueAdded	-	-	-	-	-
Laborforce	81	50,141	6,542	26,900	60,900
DivorceRate	81	1,332	0,636	0,130	2,730
Births	81	15887	27493	913	225393
MarriageRate	81	7,906	0,967	6,270	10,290
SuicideRate	81	4,503	1,579	1,600	8,800
MalePopRatio	81	0,504	0,013	0,493	0,580
SecSchoolPopRatio	81	0,035	0,008	0,021	0,065
PercapitaExp	81	2,279	0,958	1,168	6,932

Table A.6: Descriptive statistics of all variables in 2013

2013	Obs	Mean	Std. Dev.	Min	Max
IncomeGini	81	0,359	0,030	0,315	0,399
EducationGini	81	0,313	0,036	0,261	0,411
percapitaValueAdded	-	-	-	-	-
Laborforce	81	50,527	5,511	36,200	62,800
DivorceRate	81	1,343	0,615	0,140	2,700
Births	81	15840	27703	945	227162
MarriageRate	81	7,740	0,936	6,140	10,120
SuicideRate	81	4,405	1,407	0,740	9,330
MalePopRatio	81	0,504	0,010	0,494	0,558
SecSchoolPopRatio	81	0,038	0,007	0,026	0,063
PercapitaExp	81	2,596	1,040	1,240	7,823

CURRICULUM VITAE

Name Surname: Esra Öztürk

Place and Date of Birth: İstanbul/1990

E-Mail: ozturk.esra@outlook.com / ozturkesr@itu.edu.tr

B.Sc.: İstanbul Technical University/Mathematical Engineering

Hochschule für Technik Stuttgart/Mathematical Engineering/Erasmus Student

Professional Experience and Rewards:

- TUBITAK 1001 Project Scholar
  "Correction for Institutional Population in Household Surveys" coordinated by Asst. Prof. Bengi Yanık İlhan from İstanbul Kemerburgaz University
  January 2014 – June 2015 (18 Months)

Conferences

- Presented the paper "Do Local Elderly Care Facilities Have an Impact on Home Care, The Case of Turkey" (joint with Bengi Yanık İlhan), ***EconAnadolu 2015***, Anadolu University, Eskişehir, 10-12 June 2015.

Language:

- Turkish (Native), English (Advanced), German (Beginner)

Computer Skills:

- MS Office, STATA, SPSS, Mathematica, Eviews