Essays On Inequality and Education

Atiyeh Vahidmanesh

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics

Djavad Salehi-Isfahani, Chair
Suqin Ge
Kwok Ping Tsang
Wen You

February 17, 2017
Blacksburg, Virginia

Keywords: Inequality of opportunity, Human Opportunity Index, Return to Education,
Education Attainment

Essays On Inequality and Education

Atiyeh Vahidmanesh

ABSTRACT

This dissertation provides evidence of the return to education in Iran as well as measurement of inequality of opportunity and the Human Opportunity Index using cross-section data of Trends in Mathematics and Science Studies and Harmonized Household Income and Expenditure Surveys of several Middle Eastern Countries. The first chapter studies the return to education and the effect of school availability on education attainment in Iran. The Census 2006 allows us to get closer to the district of schooling by focusing on non-migrants. We estimate the return to education and the effect of school availability both for migrant and non-migrant sub-samples. We employ school availability as an instrument to correct the ability bias. We find availability of school increases women's education attainment more than men's and it is higher among the non-migrant sample. Using instrumental variable, the return to education is 6.50% in 2012 suggesting an upward bias in OLS.

The second chapter provides estimates of Human Opportunity Index (HOI) in the Middle East and North Africa. Our estimates show the HOI improve over time in MENA region and compare favorably with similar measures computed for other regions, notably Latin America. Using Shapley decomposition, we find that parental background and place of living are the most important circumstances explaining inequality of opportunity to access in basic opportunities. Understanding the change in HOI and factors that influence it most complement existing analyses of inequality of opportunity in education, earning, and consumption for MENA countries because they focus on aspects of inequality of opportunity that are largely provided by the state.

The third chapter provides estimates of inequality of educational opportunity (IOP) using TIMSS dataset. We estimate the index of IOP using the ex-ante approach both for the fourth and eighth grade. The computed index of IOP shows that there is an improvement in IOP both for mathematics and science from grade four to eight. The investigations about relevant inputs suggest that there is a negative relationship between educational expenditure and the level of IOP. The relationship between the index of IOP and average economic growth as well as GDP per capita is positive.

Essays On Inequality and Education

Atiyeh Vahidmanesh

General Audience Abstract

This dissertation provides evidence of the return to education in Iran as well as measurement of inequality of opportunity and the Human Opportunity Index using cross-section data of Trends in Mathematics and Science Studies and Harmonized Household Income and Expenditure Surveys of several Middle Eastern Countries. The first chapter studies the return to education and the effect of school availability on education attainment in Iran. The results show school availability has a positive and significant effect on women education attainment. We also estimate the rates of return to education for men between 25-60 years old in Iran in 2006 and 2012. Our findings show that the returns to schooling decline in Iran between 2006 and 2012.

The second chapter provides estimates of Human Opportunity Index (HOI) in the Middle East and North Africa. Our estimates show the HOI improve over time in MENA region and compare favorably with similar measures computed for other regions, notably Latin America. The HOI evaluates the basic opportunities which mostly publicly provided. This makes the HOI relevant for understanding the effectiveness of MENA governments in providing children with an equal chance to succeed. Our findings show the HOI improves over time in MENA. The improvement over time is not surprising because as a service expands, especially if it starts inequitably, it generally covers more of the less advantaged households.

The third chapter provides estimates of inequality of educational opportunity (IOP) using TIMSS dataset. We estimate the index of IOP using variance decomposition method. Our estimates show that in most countries of our sample the index of IOP increases from fourth grade to eighth grade suggesting the circumstances out of the control of an individual play important role in higher level of education. We investigate the correlational relationship between the index of IOP and some economic characteristics. Our findings suggest that there is a negative relationship between educational expenditure and the level of IOP. The relationship between the index of IOP and average economic growth as well as GDP per capita is positive.

To my mother Zahra,

my husband Ali, and the soul of my father Mohammad

Acknowledgments

I am deeply indebted to a large number of people for their cooperation, encouragement, and critical assistance in completing this work, although it would be impossible to mention all. This thesis would not have been possible without the guidance, encouragement, and support of my committee members. I am grateful to my advisor and mentor, Professor Djavad Salehi-Isfahani for his support, invaluable suggestions, and intellectual generosity. Throughout this journey, I have learned a lot from him both academically and personal life. I learned from him how to think critically about social issues and apply my economic knowledge to solve them. I learned from him to work hard, be hopeful, optimistic, and patient in research. He also supported my fellowship application of Harvard Kennedy School in 2014.I am deeply thankful to him.

I would like to express my deep appreciation and gratitude to the other members of my dissertation committee: Professor Wen You, Kwok Ping Tsang, and Suqin Ge. I really appreciate their help, support, and valuable comments on my proposal as well as on the final dissertation. Many thanks to Dr. Tarek Masoud who supported me during my fellowship at the Middle East Initiative at Harvard Kennedy School of Government.

I am thankful to my family whose love and support allowed me to embark on this journey and encouraged me throughout the way. I want to thank my mother, Zahra, who has thought me, by example, the value of life-long learning and who has selflessly supported my intellectual endeavors without ever questioning my decisions; to my sisters, Parvin, Parvaneh, and Maryam whose kindness and never-ending support have inspired me; and to my nieces and nephew, Hanieh, Hoda, Sophia, and Ali whose unconditional love pushed me to move forward. To my father, may his soul rest in peace. I am also thankful to my mother-in-law, father-in-law, and sister-in-law for their patient and kindness.

I would like to thank my dear friends in Blacksburg, Leyla Nazhandali, Masoud Agah, Maryam Kamali, Reyhaneh Ojani, Fatemeh Hashemi, Nasibeh Azadeh fard, Hasti Ardakani, Mina Hejazi, and Asal Pilehvari. I feel so lucky to have them during my time at Virginia Tech. I am also thankful to my friends in Boston area who made it possible to live away from my family while I was research fellow there. I want to especially thank Maryam Aliakbarpour, Mitra Akhtari, Fereshte Khani, Sepideh Mahabadi, Hoda Koushyar, Ameneh Shamekhi, Fatemeh Kazemi, Narges Jahanseir, and Alaleh Chahargameh.

I choose economics for my graduate studies to find solutions for economic problems and to make lives for poor people better, particularly less advantaged people in Iran. I hope I can return their favor one day.

Lastly, my true love of my life, Ali. He kept me going, and without his support, sacrifice, and kindness this dissertation would not have been possible. I hope I can one day return the favor. I want to dedicate this work to him and my family.

Contents

List of Figures

List of Tables

Chapter 1

School Attainment and Return to Education in Iran

1.1 Introduction

Unemployment is one of the main economic problems in contemporary Iran. Since the 1979 revolution, Iran has made notable progress in raising of education among its citizen, yet, this educational expansion is not translated to economic prosperity and development. This is not the issue just for Iran as other MENA countries also suffer from low productivity of education (Salehi-Isfahani et al. 2009). After Mincer (1974), scholars try to estimate the return of schooling and investigate the causal relationship between education and labor market outcome. Mincer results are still helpful and informative. As Card (1999) noted, there is a potential "ability bias"' in the Mincer equation which confounds the causal relation between education and earnings. Griliches (1977) in his survey of the 60s and 70s literature deduced that ability biases should be quite small. Becker (1975) also concluded that ability biases were overemphasized in human capital literature. Despite this criticism, many scholars keep on investigating to find the true causal effect of education. Other than ability bias, "discount rate bias" is another source for being skeptical of OLS estimates. Having access to school census data in Iran gives me a unique opportunity to estimate the return to education using instrumental variables. As Card (2001) surveyed, geographical proximity is one of the instruments which was used in literature. In this paper, I first estimate the effect of school availability on education attainment and then I estimate the return to education using both OLS and IV identifications. For the latter part, I use both Household Income and Expenditure Surveys (HIES) and the Census data for Iran. The reason that I use Census data, is it allows us to get closer to district of schooling, by focusing on non-migrants. In HIES data we don't know where an individual lived when s/he was in school-age, however, in the Census we know where an individual lives when s/he was in school-age. We define non-migrants as those who live in their birth place at the time of survey. Those who live in cities or villages other than their birth place are considered as migrants in this study. Since we use availability of schools in sub-district neighborhood of an individual as instrumental variables, it makes more sense to focus on non-migrant sub-sample. As I said, this piece of information is not available in HIES data. Other than that, we can see the differences between migrants and non-migrants as well in terms of return to education and the effect of school availability. Using "Poverty Mapping"' technique, I simulate log of annual wage for census given information available in HIES data. Although there are some critics about matching methods, the results of this study are quite convincing. The paper continues as follows: In section 1.2 , I briefly review the literature behind education attainment and return to schooling. In section 1.3, I described the data I have used. Section 1.4 summarized the methodology and the identification strategy used in this paper. I presented my results in section 1.5 and section 1.6 provides concluding remarks.

1.2 Literature Review

There is an extensive literature on the effect of school availability, and more specifically gender of school, on education attainment. In addition, since Mincer (1974), in education economics literature, scholars have tried to estimate the return to schooling and see the effect of education on people well-being.

1.2.1 School Availability and Education Attainment

In this paper, we focuse on education attainment and more specifically, I want to see the effect of school supply on the probability of reaching a higher level of education in Iran. Duflo et al. (2001) focused on the labor market consequences of school construction in Indonesia. Based on her paper, there is an estimate of 0.12 to 0.19 years of education along with 1.5 to 2.7 percent increase in wages in Indonesia as a result of the construction of each primary school per 1000 children. She estimated 6.8% to 10.6% return to education for Indonesia. In this sense, Iran also experienced an extensive school construction between 1976 and 2004. After the 1979 revolution, the revolutionary government accelerated school constructions and also tried to build single-sex schools rather coeducational ones. In another paper, using 2SLS strategy, Duflo showed that "an increase of 10 percentage points in the proportion of primary school graduates in the labor force reduced the wages of the older cohorts by 3.8-10% and increased their formal labor force participation by 4-7%"' (Duflo (2004)). This result is important because, in contrast to usual assumptions, improvement in the rate of human capital accumulation does not translate to economic growth. Assaad and Saleh (2016) investigate the effect of improved local supply of schooling on intergenerational mobility in Jordan. Their findings suggest that the local availability of basic public schools improve -"intergenerational mobility in education"'- it is found that "an increase in the local supply of basic schools reduces intergenerational persistence in education for women three times more than it does for men. A one standard deviation increase in basic schools per 1000 people reduces the coefficient of intergenerational persistence by at least one-third for women and by one-fifth for men."'Neilson and Zimmerman (2014) investigates the effect of school supply (especially elementary and middle school) on home prices, academic achievement, and school attendance. Using the Difference-in-Differences strategy, they find powerful evidence that school construction led to improvement in reading scores for elementary and middle school students. Schultz (2004) evaluates the effect of school availability and conditional cash transfer on education attainment of poor kids. Although this paper is not directly investigated the educational impact of school supply, it shows the impact of subsidized schools on the level of enrollment for poor kids. Nishimura et al. (2008) investigate the effect of "Universal Primary Enrollment(UPE)"' policy on education attainment. Using probit models they estimate "the determinants of enrollment, delayed enrollment, and the completion of the fourth and fifth grades."'According to this study, the UPE decreases delayed enrolment in rural Uganda. This policy has positive effects on poor (and mostly on girls) about access to school.

1.2.2 Return to Education

In addition to the effect of school availability on education attainment, in this research, I investigatethe effect of human capital accumulation on the individual wage in the context of Iran. Salehi-Isfahani et al. (2009) presented a comparative study of private returns to schooling of urban men in Egypt, Iran, and Turkey. They showed in Turkey return to schooling improved over time, whereas, in Egypt and Iran, it increases between 1988 and 2000 but it declines after that. The current study is different from Salehi-Isfahani et al. (2009) study in several ways. In this paper, I first try to investigate the effect of school availability on education attainment which gives me the estimates which validate using of school availability as instruments. I also provide both OLS and 2SLS estimates of the return to education in Iran using both Household Surveys and Census. In addition, I used the "poverty mapping"' method to simulate the wage for individuals in Census data using Household survey data. Having the OLS estimates of the return to schooling for two years (2006 and 2012) provide me with the ingredient to see what happened in Iran in terms of education productivity over time. The literature on return to schooling is so rich and vast. Card (1994) reviews research which tries to estimate the causal relationship of wage and schooling. Most of the research tries to deal with the potential endogeneity problem in Mincer equation by using instrumental variable strategy. Angrist and Keueger (1991) use the quarter of an individual's birth as an instrument. Their analysis verifies that people born early in the year attain a lower level of schooling and hence earnings. Butcher and Case (1994) shows the effect of sibling composition on educational attainment. They show that "women's educational choices have been systematically affected by the sex composition of her siblings, and that men's choices have not."' On average, girls with sisters attain a lower level of education than girls with only brothers. Card (1993) used college accessibility as an instrument to disentangle the causal link between education and earnings. His analysis shows that men who were raised in a neighborhood with a four-year college nearby will acquire higher schooling and earnings. This pattern is even true for men with less educated parents. Card (2001) completes his previous papers in terms of statistical problems in the estimation of Mincer equation. Reviewing various studies, he concludes that OLS estimates of the return to schooling are typically higher than IV. This is a puzzle because due to "ability bias"' and "discount rate bias"' we expect to see an upward bias in OLS and higher estimates for IV could not explain this phenomenon. Several researchers try to explain these results. Some argue that ability bias in OLS, in fact, is relatively small and the downward pressure of measurement errors in schooling reflects in a downward bias in the OLS estimates. Card (2001) believes it is very hard to accept measurement errors as a reason which explains this gap between IV and OLS. There are other explanations including upward bias is even further in IV due to unobserved differences between the characteristics of the treatment and control groups implicitly framed in IV. In this study, this is not the case. Almost for all specifications, my IV estimates are lower than OLS estimates. This is similar to Duflo (2000). The reason for that is maybe the existence of larger ability bias in Iran. Another reason may be related to the validity of my instruments which correct existed bias in OLS.

1.3 Data

In this paper, we used four datasets to estimate the probability of school availability on education attainment and return to education in Iran. They are Household Income and Expenditure Surveys (HIES) (2006 and 2012), Census (2006), and the 2014 Iranian School Census. The latter dataset is a unique dataset of all schools in Iran which enables us to determine school availability for each individual at the sub-district level. The HIES (2006) sampled 30910 households and 135,270 individuals which are nationally representative. The dataset consists of rich information on household income and expenditure and through that, we can simply measure the per capita consumption or income. In addition, we can determine employed individuals both in public and private sectors which we can derive the log of annual wage for them. This latter variable benefits us to estimate the return to education equation. Other than variables on income and expenditures, the HIES has a wide range of variables on household and individual characteristics which enable us to use the two-sample estimation technique and match individual wage and salaries in HIES with the Census. In this study, we restricted ourselves to the individuals born in Iran who are aged 25 and 60 in 2006. The reason we use the 2006 samples instead of the more recent census of 2011 is that in 2006 we can identify the location of individuals at the sub-district level where we then can match them with School Census data while in 2011 census we can locate individuals only at the district level. The restricted sample of HIES for individuals who are aged 25 and 60 has 54841 individuals. Among them 28437 individuals are female and 26404 are male and about 28,816 of them are employed. For estimation of return to schooling, my sample would be employed men aged between 25 and 60 years old. This sample has 22913 observation. We also utilize HIES 2012 to estimate the return to schooling using both standard Mincer equation and instrumental variable strategy. HIES 2012 has 38192 households and 146,062 individuals. The restricted sample (Employed Men, 25-60 years old) has 26381 individuals. As Salehi-Isfahani et al. 2009 noted we focus on men because labor force participation rates for men are much higher than women in Iran. These higher participation rates give us more accurate results and decrease the risk of selection bias (see table 1.1). The age range as we mentioned above is between 25-60 years old. We use this age group because It is more likely for this group to be in labor market rather than school or military service, or retired. The participation rate is more stable in this age group leading us the lower problem with selection into the wage labor force.

The second dataset we used in this study is the two percent of 2006 Census data. The Census surveyed 1,367,310 individuals living in 345,799 households. For our studies, again we restricted the sample to individuals who are between 25 and 60 years old which reduced the sample to 577,719 individuals. Among them, 288,280 are female and 289,439 are male. The Census also has rich information on household and individual characteristics; however, it does not have any information on wage and consumption of the household. Using two sample estimation technique, we can simulate the wage of individuals in census using household survey information. In this paper, we used the method developed by Lanjouw et al. (2002) and utilizing Povmap2 software for our simulations. The third data set is the 2014 Iranian School Census dataset. This is a unique dataset of all schools in Iran based on gender, education type, and location. Each school has a geographical id namely space and building

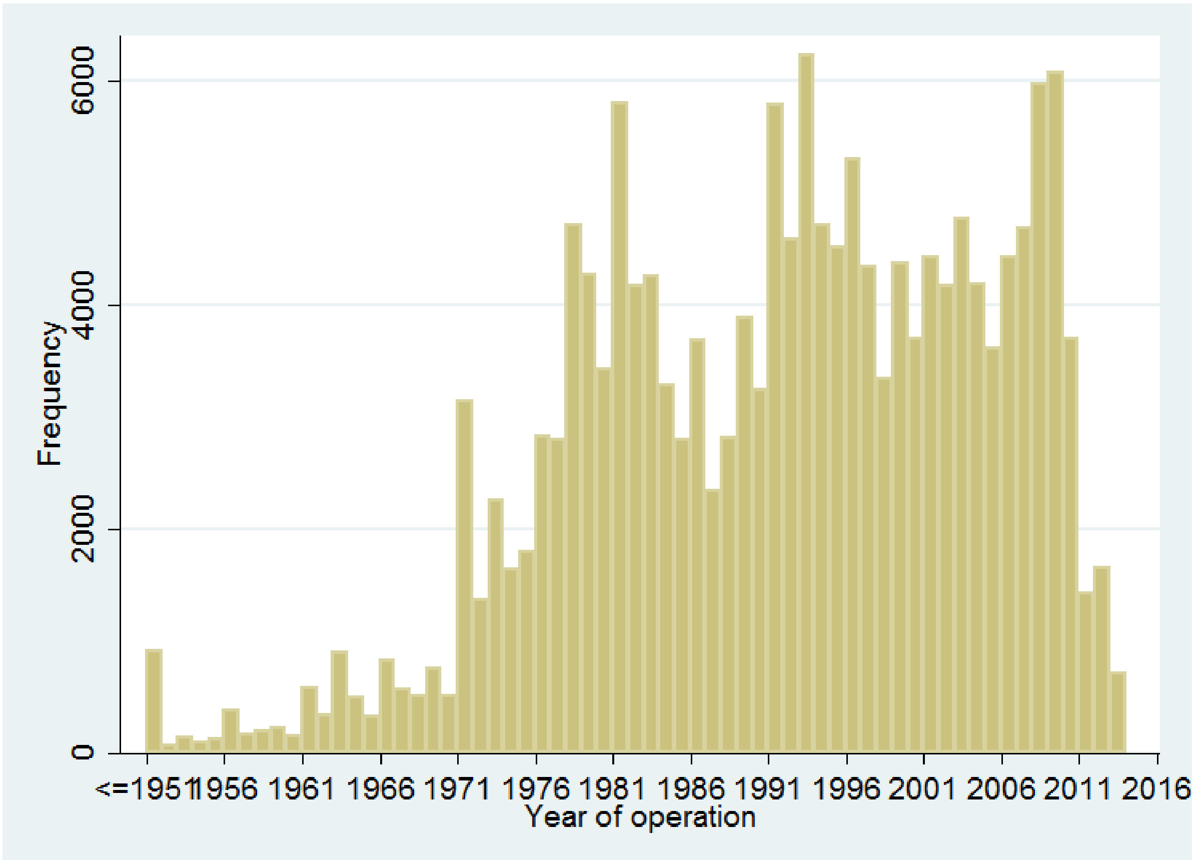
Table 1.1: Participation and Unemployment Rates

	2006	**2012**
Labor Force Participation Rate, ages +15		
Total	46.5	44.8
Male	73.3	73.1
Female	18.5	16.4
Labor Force Participation Rate, ages 15-64		
Total	48.3	46.8
Male	76	76.1
Female	19.4	17.4
Unemployment Rate (% Labor Force)		
Total	10.1	13.1
Male	11.6	11.5
Female	17.6	20.2

Source: WDI database, 2017

id which enables us to identify the geographical location of the school at the village or township level, however, since we can identify individual in census data at the sub-district level, we redefine the geographical id at the sub-district level to match it later on with census. There is information about the gender of school, the time that the school was built, and the education type of the school. Unfortunately, we do not have any data on which schools are resolved, however, due to the government policy which tries to improve the school supply in all towns and villages, we assume that if a school built, it never resolved. This assumption in the context of Iran makes sense. Iran has experienced a baby-boom from 1978 to 1985 and lots of schools were built from 1988 to 2004. Some of those schools, especially private ones may resolve within past few years because of reduction in school-aged population but this is not the case for the time frame of our study and as a result, it makes sense to assume those schools are not resolved in 2006. We don't have access to quality of school as well. In this study, we use availability of schools as instrumental variables to correct ability bias. One can argue that more able students may be sent to better school and as a result, for correcting ability bias problem, quality of school plays role too. Unfortunately, there is no information about the quality of school and we could not look at ability bias through the lens of school

Figure 1.1: Number of Schools in Iran by Year of Construction, 1951-2014

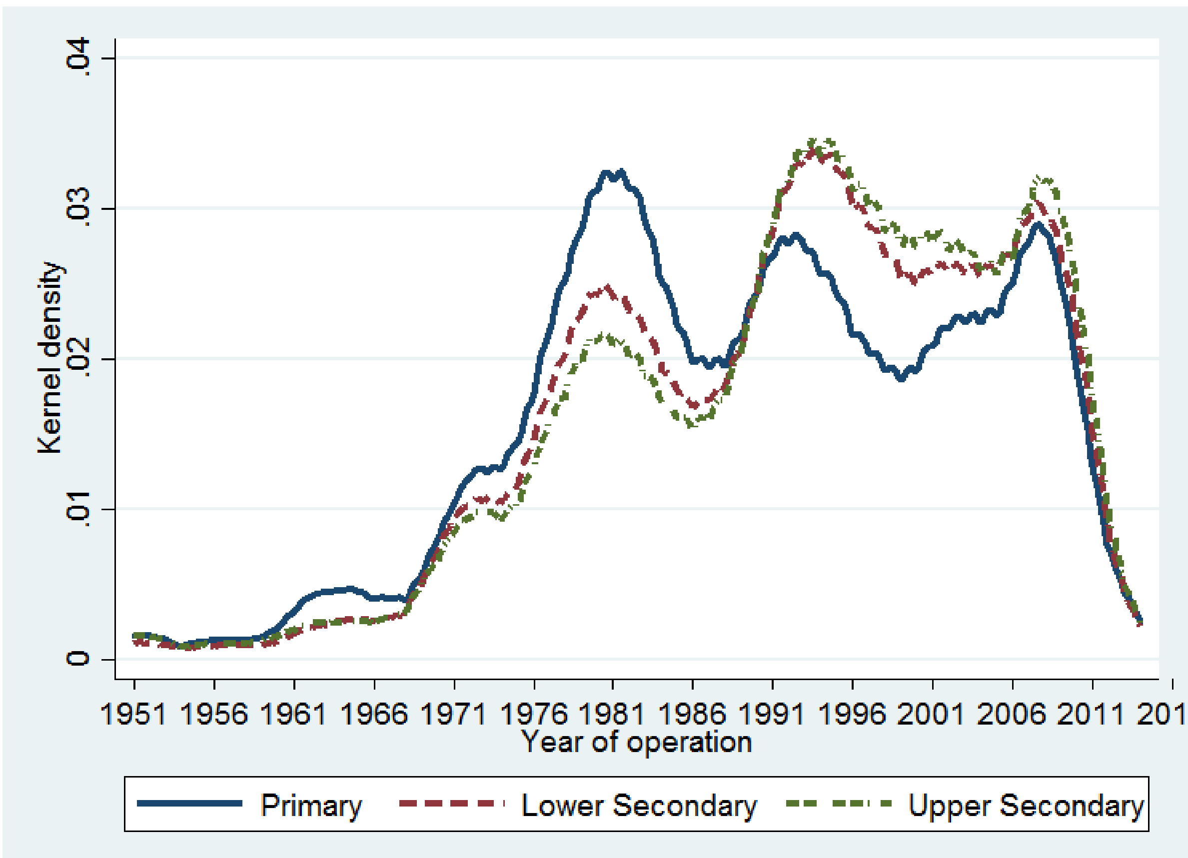


Source: School Census in Iran

quality. Based on space id and building id, we can identify about 170,000 schools in the country in 2011. Among them, 42.2% are only girl schools and 43.49% are only boy schools, about 14.31% are also mixed schools. Most of the mixed schools are primary schools and they are mostly located in rural area and small towns. Figure 1.1 shows the number of all schools by year of operationin in Iran since 1951.

As it is obvious in this figure, massive school construction in Iran started from early of the 1970s and it accelerated after the 1979 revolution and specifically during 1990s after the war and when the baby boom generation reached to school age. Figure 1.2 shows the distribution of all schools by year of operation and based on education type in Iran since 1951.

Figure 1.2: Distribution of all Schools by Operation year and based on Education Type, 1951-2014



Source: School Census in Iran

1.4 Methodology

In this paper, first I estimate the effect of school availability on the probability of acquiring the higher level of education and then I provide the estimates of the return to education in Iran. For the latter goal, I estimated the return to education in Iran both using Household Income and Expenditure Surveys (2006 and 2012) and Censu(2006). As I explained, for census estimation I used two samples estimation which I provide a brief summary of this method here.

1.4.1 Access to School and Education Attainment

There is a literature that shed light on the effect of school availability on education attainment and labor market outcome. One of the Millenium Development Goal of the United Nations is exactly about the expansion of school supply. Having access to the unique dataset of school census in Iran, we have enough information to estimate the effect of school constructions on education attainment. We utilize the 2006 Iranian Census for this exercise. The reasons we used Census data are first it has more observations with higher variation in terms of the education level of individuals. In addition, in Census data we can identify the immigration status of individuals and provide our estimates both for migrants vs non-migrants samples. However, in Census data we don't have access to income and wealth of individuals but using the rich information on household characteristics we build the wealth index using factor analysis. The model we use is as follow:

$$Y_i = \alpha + X'\beta + \gamma \times (SA) + \epsilon \tag{1.1}$$

where Y_i is one if the individual i finish high school and/or ever attend secondary school, X are some relevant covariates such as wealth index and place of living, SA is school availability meaning is the appropriate school (secondary or high school) available in the sub-district of individual i when that individual reach to secondary or upper secondary school age.ϵ is the error term. We estimate 1.1 for whole, migrants, and non-migrants samples. To check the robustness of our results, we use the cohort and province fixed effect to capture the effect of age and place.

1.4.2 Estimation of Return to Education

Model of earnings and return to education was started by Mincer (1974) which became the framework for studying of education in developing countries. The original Mincer model specifies:

$$ln[w(s,x)] = \alpha_0 + \gamma s + \beta_0 x + \beta_1 x^2 + \epsilon \tag{1.2}$$

where $w(s,x)$ is wage at the level of schooling s and work experience x. In this model γ is called the "rate of return to schooling"' and ϵ is a mean zero residual. As Trostel (2005) presents in his paper, there are evidence of nonlinearity in Mincer equation and he proposed a broader specification for Mincer equation. As of him, the extended version of Mincer equation can be written as:

$$ln[w(x,s)] = \alpha_0 + \sum \gamma_j S^j + \sum \beta_h E^h + \epsilon \tag{1.3}$$

In this paper, we estimated the quadratic form both experience and years of schooling.

1.4.3 Two Samples Estimation: Application of Micro-level Estimation of Welfare

As we pointed out above, we use the micro-level estimation of welfare to drive log of wage in census data. Basically, we estimated standard Mincer regression using Household Survey data where we have access to wage and salary data, however, for disentangling the effect of migrants vs non-migrants we use Census data which required using two samples estimation techniques. This technique has been applied in various studies measuring poverty and welfare inequalities in absence of income and consumption data and the estimates are quite convincing (see: Assaad et al. 2016 and Elbers et al. 2003). These econometric advances make it possible to overcome the limitations of the census by modeling consumption (or income) in Household Income and Expenditure Surveys and using such models to predict consumption (income). This method allows for mapping consumption (income) from survey to census data to provide estimates of poverty, consumption (income), and inequality for entire nations. (Hentschel et al. 2000, Lanjouw et al. 2002, and Elbers et al. 2003).

This method is developed by the World Bank with the goal of improvement of anti-poverty programs. This technique of consumption imputations from survey to census yielded ingredients for "poverty map"' and hence imputing consumption (income) to local levels was referred to as poverty mapping. The poverty map technique has been applied since then to a large number of developing countries in order to make available disaggregated estimates of poverty and inequality (Alderman et al. 2002, Bedi et al. 2007, Elbers et al. 2007). In this paper, we use data from HIES (2006) to predict wage onto Census 2006 in Iran. For doing that, we rely on the methods of Elbers et al. (2003) and restrict both samples to employed individuals who have wage and salaries. (Elbers, Lanjouw, and Lanjouw 2003) impute per capita expenditure as a measure of wellbeing but since we want to estimate the return to education, we use that method to impute log of annual wage. y_{ci} is the log of annual wage for individual i in cluster c in HIES. The first step of imputation is to estimate the following model based on covariates X_{ci} that are available in both samples:

$$ln(y_{ci}) = \beta X_{ci} + u_{ci} \tag{1.4}$$

where β are the k parameters to be estimated and u_{ci} is a vector of disturbances with distribution $F(0, \Sigma)$. As Elbers et al. (2003) noted, localities (clusters) are probably correlated with disturbances, u_{ci} can be decomposed into cluster effect, η_c and an idiosyncratic error, ϵ_{ci}.

$$u_{ci} = \eta_{ci} + \epsilon_{ci}$$

After estimate of β is derived from 1.4 using OLS or FGLS, the $\hat{u_{ci}}$ can be generated. There is a small number of clusters sampled within a survey and as a result the variance of the cluster effect could not be modeled with heteroskedasticity.[1] However, the ϵ_{ci} (idiosyncratic element) can be heteroskedastic as follows:

$$\hat{u_{ci}} = \hat{u_{c.}} + (\hat{u_{ci}} - \hat{u_{c.}}) = \hat{\eta_c} + e_{ci} \tag{1.5}$$

with $\hat{u_{c.}}$ indicating the average over cluster c. Since the goal of this paper is to estimation of

[1]Krafft, Caroline, Ragui Assaad, Hanan Nazier, Racha Ramadan, Atiyeh Vahidmanesh, and Sami Zouari. 2017. "Estimating Poverty and Inequality in the Absence of Consumption Data: An Application to the Middle East and North Africa (Forthcoming)."' Economic Research Forum Working Paper Series. Cairo, Egypt.

return to education, we don't aim to present all details of poverty map technique here. For sake of our estimation, we use PovMap2, the package developed by the World Bank Group, to estimate 1.4 and simulate y_{ci} for individuals in census. The simulated imputed values of $ln({y_{ci}}^r)$ are generated for the Census as Tarozzi and Deaton (2009):

$$ln({y_{ci}}^r) = \hat{\beta}^r \times X_{ci} + \hat{\eta_c}^r + e^r_{ci} \tag{1.6}$$

We ran the simulation for 100 times and used the average $ln({y_{ci}}^r)$ as the log of annual wage for individuals in census data. We then merged this variable with Census data and school data to build a dataset containing variables on individual characteristics, school availability at the sub-district level neighborhood of individuals and wage for employed people. We restrict our sample to men aged between 25 and 60 who are employed (either in public or private sector).

1.4.4 Identification Strategy: Instrumental Variables

There are some debates that the coefficient of schooling in traditional Mincer regression may not be consistent due to endogeneity. The problem is schooling choice may be correlated with wage and as a result, there will be a bias in γ. In fact the endogeneity arises because when we estimate the standard Mincer equation (1.2) we implicitly have the following equation as well:

$$S_i = {X_i}'\delta + u_i$$

We need to have $E(X_i\epsilon_i) = 0$ and $E(S_i\epsilon_i) = 0$ which means we need to have $E(u_i\epsilon_i) = 0$. Early research suggested on the issue of "ability bias". Griliches (1977) argued that this bias was, in reality, small. Another concern is about "discount bias"' which is reflected in the work of Card (1994). One way to solve this issue is using instrumental variable identification. For instance Angrist and Keueger (1991) use the quarter of an individual's birth as an instrument for education. The IV exercise in their paper leads to 28% higher return compare to OLS specification. Butcher and Case (1994) use sibling composition on educational attainment. Card (1993) use presence of a nearby college as an instrument for schooling. Maluccio (1998) also used distance to nearest high school as an instrument to estimate the return to schooling. In this research, I utilize the unique dataset of school census to build variables of primary, lower secondary, and high school availability in each sub-district. There are about 890 sub-district in Iran and using school census data I can determine is there an appropriate school available at the neighborhood of an individual when s/he turns to school age (i.e. 6, 10, and 14 for primary, lower secondary, and high school). These dummy variables can be used as instruments because we can assume that availability of school is not correlated with ability or any unobserved characteristics of individuals and it is correlated with schooling. If there is a school in a neighborhood of an individual, it is more likely that the parents send their kids to school. In this sense as Card (1993) and Maluccio (1998) noted, the school availability could be a valid instrument.

1.5 Results

In this section, we present our results both effect of school availability on education attainment and the estimates of the return to education.

1.5.1 Access to School and Education Attainment

As described in 1.1 I first estimate the effect of school availability on the probability of education attainment.I focus both on female and male samples. The hypothesis is school availability has a larger effect on women than men. In less developed areas, if there is no school, the parents won't simply send their girls to school while there is a chance for boys to go to schools in other villages. In addition, in presence of schools, some boys may go to work on lands while the girls may have the chance to attend school. Both mechanisms provide an intuition of higher effect of school availability on the probability of education attainment for women.

Table 1.2 presents the estimates of finishing high school given availability of upper secondary level schools in the neighborhood of individuals when the person turned in 14 years old (the age that students in Iran start high-school). Table 1.5 provides estimates of ever-attend in secondary school condition on the availability of lower secondary when the person turned in 10 years old.

Table 1.2: Probability of Finishing High School Condition on School Availability

	Dependent Variable: Finish High School (Whole Sample, $25 \leq Age \leq 60$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6
qwealth2	0.254***	0.277***	0.290***	0.249***	0.287***	0.311***	0.266***	0.279***	0.290***
	(0.016)	(0.016)	(0.016)	(0.021)	(0.022)	(0.022)	(0.016)	(0.016)	(0.015)
qwealth3	0.398***	0.443***	0.475***	0.400***	0.469***	0.510***	0.406***	0.433***	0.466***
	(0.019)	(0.020)	(0.020)	(0.023)	(0.025)	(0.028)	(0.019)	(0.020)	(0.020)
qwealth4	0.574***	0.666***	0.696***	0.559***	0.695***	0.732***	0.604***	0.669***	0.702***
	(0.019)	(0.020)	(0.019)	(0.025)	(0.028)	(0.028)	(0.019)	(0.019)	(0.018)
qwealth5	1.201***	1.355***	1.386***	1.150***	1.386***	1.419***	1.278***	1.395***	1.437***
	(0.032)	(0.032)	(0.018)	(0.040)	(0.038)	(0.024)	(0.028)	(0.027)	(0.019)
rural	-0.477***	-0.490***	-0.500***	-0.574***	-0.606***	-0.618***	-0.414***	-0.420***	-0.432***
	(0.025)	(0.027)	(0.018)	(0.035)	(0.038)	(0.023)	(0.021)	(0.022)	(0.017)
hschool	0.593***	0.278***	0.294***	0.906***	0.467***	0.455***	0.441***	0.212***	0.249***
	(0.032)	(0.048)	(0.034)	(0.053)	(0.076)	(0.052)	(0.028)	(0.040)	(0.033)
Cohort FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE (province)	No	No	Yes	No	No	Yes	No	No	Yes
Sample Size	Whole	Whole	Whole	Female Only	Female Only	Female Only	Male Only	Male Only	Male Only
Observations	536,655	536,655	536,655	267,901	267,901	267,901	268,754	268,754	268,754

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Census 2006 and School Census of Iran (2014)

As it is obvious in table 1.2, availability of high-school at the time when the individual reached high-school age has a significant and positive effect on the probability of completion of high-school. One may have a concern that this result is biased due to cohort and location effect. In other words, school availability is significant because we don't see the effect of cohort and location effect. In specification 2 and 3 I first add cohort fixed effect and then location fixed effect. Part of the effect of school availability would be captured by cohort and location effect. As Greene (2002) noted there is a risk for biasedness because of incidental parameter problem in a non-linear model with fixed effect specification. He proposed to use logit instead of probit model. I estimated logit fixed effect as well and the results are quite similar. In fact, we know that logit coefficients are about 1.6 higher than probit coefficients and the results of my model using logit function follow this rule to some extent. Tables 6 and 7 in appendix A summarized my results for the whole sample (both female and male). The trend is pretty similar for migrant and non-migrant sub-samples.

It is important to interpret the coefficients in probit model correctly. In contrast with the simple linear regression, in probit model, we could not simply interpret the results. The marginal value of the coefficient is as follow:

$$\frac{\partial P[Y_i = 1|X_{1i}, ...X_{Ki}; \beta_0, ..., \beta_K]}{\partial X_{ki}} = \beta_k \phi(\beta_0 + \sum_{k=1}^{K} \beta_k X_{ki})$$

For the first regression in table 1.2, the marginal effect of availability of high school is 0.15 which means having high-school available in the neighborhood of an individual increases the chance of finishing high school by 0.15. Using cohort and province fixed effect, decreases the marginal value of high school availability. However, it is still positive and significant. In this specification, school availability increases the probability of finishing high school by 0.07. The results for female and male samples will be 0.09 and 0.06 respectively.

We run similar regressions for both non-migrant and migrant samples and their estimates are presented in tables 1.3 and 1.4. The reason we did this exercise is first of all we can see the difference between these two sub-samples. Second, since we know the place of living of non-migrants, we can talk exactly about the effect of school availability on education attainment. In other words, for non-migrants, we know whether they have access to schools (lower secondary or high school) when they are in school-age and as a resutl, we can judge better about the effect of school availability on education attainment. In addition, we will use district of schooling as instrumental variables, so working with non-migrant sub-sample would give us the estimates on the return to schooling corrected for potential endogeneity. The trend is quite similar for non-migrants and migrants samples. It is important to note that we could not conclude that much about the migrant sample. The problem is, we don't know at what age the individual migrated. There is a variable about the immigration status of individuals in 1996 (ten years before this census were collected), however, since we restricted the sample to those older than 25 years old and younger than 60 years old, this variable does not help that much. The marginal effect of high-school availability for migrant and non-migrant samples with FE specification are 0.068 for both of them. For the simple specification (without FE) this effect is 0.14 and 0.15 for migrants and non-migrants respectively. The trend for female and male samples are also quite similar to male and female of the whole sample. The effect is much larger for female compared to male.

Table 1.3: Probability of Finishing High School Condition on School Availability, Non-Migrants

	Dependent Variable: Finish High School (Non Migrate Sample, $25 \leq Age \leq 60$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6
qwealth2	0.259***	0.279***	0.295***	0.249***	0.282***	0.314***	0.272***	0.283***	0.295***
	(0.019)	(0.019)	(0.016)	(0.027)	(0.028)	(0.025)	(0.018)	(0.019)	(0.017)
qwealth3	0.418***	0.455***	0.485***	0.420***	0.482***	0.523***	0.422***	0.442***	0.471***
	(0.023)	(0.024)	(0.023)	(0.031)	(0.033)	(0.032)	(0.023)	(0.024)	(0.023)
qwealth4	0.580***	0.658***	0.682***	0.565***	0.688***	0.723***	0.602***	0.654***	0.679***
	(0.024)	(0.025)	(0.023)	(0.033)	(0.036)	(0.035)	(0.023)	(0.024)	(0.022)
qwealth5	1.176***	1.307***	1.316***	1.118***	1.336***	1.347***	1.246***	1.336***	1.357***
	(0.044)	(0.039)	(0.023)	(0.059)	(0.050)	(0.038)	(0.036)	(0.032)	(0.021)
rural	-0.549***	-0.557***	-0.559***	-0.721***	-0.750***	-0.754***	-0.455***	-0.458***	-0.464***
	(0.031)	(0.032)	(0.018)	(0.043)	(0.044)	(0.023)	(0.024)	(0.024)	(0.017)
hschool	0.617***	0.303***	0.301***	0.995***	0.536***	0.509***	0.440***	0.234***	0.250***
	(0.038)	(0.056)	(0.035)	(0.065)	(0.094)	(0.059)	(0.031)	(0.046)	(0.033)
Cohort FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE (province)	No	No	Yes	No	No	Yes	No	No	Yes
Sample Size	Whole	Whole	Whole	Female Only	Female Only	Female Only	Male Only	Male Only	Male Only
Observations	332,089	332,089	332,089	159,609	159,609	159,609	172,480	172,480	172,480

Robust standard errors in parentheses
*** p< 0.01, ** p<0.05, * p<0.1

Source: Census 2006 and School Census of Iran (2014)

Table 1.4: Probability of Finishing High School Condition on School Availability, Migrants

	Dependent Variable: Finish High School (Migrated individuals, $25 \leq Age \leq 60$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6
qwealth2	0.242***	0.270***	0.290***	0.246***	0.288***	0.309***	0.252***	0.270***	0.295***
	(0.022)	(0.022)	(0.023)	(0.027)	(0.028)	(0.030)	(0.026)	(0.026)	(0.026)
qwealth3	0.358***	0.416***	0.459***	0.358***	0.434***	0.474***	0.375***	0.418***	0.468***
	(0.025)	(0.025)	(0.026)	(0.030)	(0.031)	(0.035)	(0.030)	(0.029)	(0.028)
qwealth4	0.548***	0.664***	0.705***	0.523***	0.675***	0.712***	0.599***	0.697***	0.746***
	(0.027)	(0.025)	(0.026)	(0.032)	(0.032)	(0.034)	(0.030)	(0.028)	(0.027)
qwealth5	1.218***	1.408***	1.452***	1.165***	1.421***	1.454***	1.318***	1.489***	1.557***
	(0.027)	(0.029)	(0.025)	(0.030)	(0.032)	(0.030)	(0.030)	(0.031)	(0.029)
rural	-0.370***	-0.384***	-0.405***	-0.371***	-0.398***	-0.419***	-0.312***	-0.318***	-0.338***
	(0.029)	(0.032)	(0.029)	(0.034)	(0.039)	(0.033)	(0.033)	(0.034)	(0.032)
hschool	0.555***	0.249***	0.275***	0.785***	0.384***	0.372***	0.438***	0.179***	0.237***
	(0.037)	(0.051)	(0.045)	(0.052)	(0.068)	(0.057)	(0.039)	(0.052)	(0.048)
Cohort FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE (province)	No	No	Yes	No	No	Yes	No	No	Yes
Sample Size	Whole	Whole	Whole	Female Only	Female Only	Female Only	Male Only	Male Only	Male Only
Observations	202,589	202,589	202,589	107,356	107,356	107,356	95,233	95,233	95,233

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Census 2006 and School Census of Iran (2014)

Other than high school we also looked at the probability of everattend in lower secondary school condition on availability of secondary school in sub-district of individuals. Tables 1.5,1.6, and 1.7 show the results for whole, non-migrant, and migrant samples respectively. The results are quite similar to tables 1.2, 1.3, and 1.4. Being in higher quantile of income distribution and availability of schools have positive and significant effect on probability of ever-attend in lower secondary school. It is interesting to note that availability of high school also has positive and significant effect. We can say that when an individual lives in a neighborhood that both lower secondary and high school are available, it is more likely for him/her to be sent to school. The idea is parents in this situation have more incentive to send their kids to lower secondary school because they know their kids can continue their education after that. It is also interesting to note that living in rural area has negative effect both on probability of finishing highschool and ever attending in lower secondary school. The marginal effect for availability of lower secondary school given FE specifiacation is about 0.011 meaning if there is a lower secondary school available in sub-district of an individual, the chance for that individual to ever attend in lower secondary school increases by 0.011 points. It is interesting to note that the marginal effect for availability of high school is 0.016.

Table 1.5: Probability of Ever Attend to Lower Secondary School Condition on School Availability

	Dependent Variable: Ever Attend Secondary School (Whole Sample, $25 \leq Age \leq 60$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5
qwealth2	0.180***	0.240***	0.229***	0.210***	0.276***	0.267***	0.167***	0.226***	0.217***
	(0.012)	(0.013)	(0.012)	(0.016)	(0.018)	(0.016)	(0.014)	(0.014)	(0.014)
qwealth3	0.181***	0.283***	0.293***	0.220***	0.322***	0.330***	0.160***	0.267***	0.281***
	(0.013)	(0.013)	(0.013)	(0.019)	(0.020)	(0.019)	(0.015)	(0.015)	(0.015)
qwealth4	0.177***	0.349***	0.360***	0.199***	0.387***	0.394***	0.173***	0.342***	0.359***
	(0.014)	(0.014)	(0.014)	(0.019)	(0.021)	(0.020)	(0.016)	(0.016)	(0.015)
qwealth5	0.087***	0.350***	0.350***	0.129***	0.433***	0.426***	0.067**	0.312***	0.321***
	(0.030)	(0.027)	(0.029)	(0.033)	(0.036)	(0.042)	(0.031)	(0.025)	(0.025)
rural	-0.171***	-0.203***	-0.201***	-0.263***	-0.302***	-0.303***	-0.103***	-0.135***	-0.132***
	(0.012)	(0.012)	(0.012)	(0.015)	(0.017)	(0.016)	(0.013)	(0.013)	(0.012)
hschool	0.341***	0.087***	0.116***	0.451***	0.169***	0.194***	0.288***	0.054**	0.086***
	(0.028)	(0.022)	(0.020)	(0.042)	(0.034)	(0.032)	(0.027)	(0.024)	(0.021)
lschool	0.579***	0.079***	0.085***	0.676***	0.113***	0.108***	0.526***	0.066***	0.080***
	(0.029)	(0.021)	(0.019)	(0.039)	(0.030)	(0.029)	(0.027)	(0.021)	(0.020)
Cohort FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE (province)	No	No	Yes	No	No	Yes	No	No	Yes
Sample Size	Whole	Whole	Whole	Female Only	Female Only	Female Only	Male Only	Male Only	Male Only
Observations	536,655	536,655	536,655	267,901	267,901	267,901	268,754	268,754	268,754

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Census 2006 and School Census of Iran (2014)

Table 1.6: Probability of Ever Attend to Lower Secondary School Condition on School Availability, Non-Migrant Sample

	Dependent Variable: Ever Attend Secondary School (Non Migrate Sample, 25≤ *Age* ≤ 60)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5
qwealth2	0.183***	0.238***	0.223***	0.223***	0.285***	0.276***	0.165***	0.219***	0.203***
	(0.015)	(0.015)	(0.013)	(0.021)	(0.022)	(0.020)	(0.015)	(0.016)	(0.015)
qwealth3	0.201***	0.293***	0.296***	0.233***	0.329***	0.336***	0.186***	0.280***	0.284***
	(0.016)	(0.016)	(0.015)	(0.024)	(0.025)	(0.022)	(0.017)	(0.017)	(0.017)
qwealth4	0.196***	0.349***	0.352***	0.211***	0.386***	0.388***	0.197***	0.342***	0.349***
	(0.018)	(0.017)	(0.016)	(0.026)	(0.027)	(0.024)	(0.019)	(0.019)	(0.019)
qwealth5	0.099***	0.334***	0.323***	0.101***	0.385***	0.369***	0.114***	0.322***	0.320***
	(0.029)	(0.032)	(0.035)	(0.034)	(0.041)	(0.048)	(0.029)	(0.030)	(0.031)
rural	-0.223***	-0.245***	-0.246***	-0.352***	-0.384***	-0.386***	-0.149***	-0.170***	-0.170***
	(0.014)	(0.014)	(0.014)	(0.018)	(0.020)	(0.020)	(0.015)	(0.015)	(0.015)
hschool	0.375***	0.104***	0.130***	0.480***	0.176***	0.197***	0.329***	0.079***	0.109***
	(0.033)	(0.028)	(0.026)	(0.052)	(0.048)	(0.046)	(0.031)	(0.028)	(0.027)
lschool	0.605***	0.068***	0.081***	0.682***	0.078**	0.084**	0.565***	0.067***	0.085***
	(0.032)	(0.026)	(0.025)	(0.043)	(0.040)	(0.038)	(0.030)	(0.026)	(0.026)
Cohort FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE (province)	No	No	Yes	No	No	Yes	No	No	Yes
Sample Size	Whole	Whole	Whole	Female Only	Female Only	Female Only	Male Only	Male Only	Male Only
Observations	332,089	332,089	332,089	159,609	159,609	159,609	172,480	172,480	172,480

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Census 2006 and School Census of Iran (2014)

Table 1.7: Probability of Ever Attend to Lower Secondary School Condition on School Availability, Migrant Sample

	Dependent Variable: Ever Attend Secondary School (Migrated Sample, $25 \leq Age \leq 60$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5
qwealth2	0.174***	0.241***	0.237***	0.190***	0.260***	0.247***	0.168***	0.237***	0.238***
	(0.021)	(0.023)	(0.022)	(0.022)	(0.024)	(0.024)	(0.030)	(0.031)	(0.028)
qwealth3	0.137***	0.257***	0.277***	0.189***	0.302***	0.308***	0.101***	0.235***	0.263***
	(0.024)	(0.023)	(0.023)	(0.026)	(0.027)	(0.028)	(0.031)	(0.030)	(0.028)
qwealth4	0.122***	0.328***	0.351***	0.159***	0.369***	0.377***	0.102***	0.319***	0.350***
	(0.024)	(0.024)	(0.024)	(0.027)	(0.028)	(0.030)	(0.029)	(0.028)	(0.026)
qwealth5	0.044	0.353***	0.362***	0.145***	0.481***	0.474***	-0.037	0.273***	0.294***
	(0.041)	(0.031)	(0.033)	(0.045)	(0.040)	(0.044)	(0.045)	(0.034)	(0.033)
rural	-0.123***	-0.148***	-0.143***	-0.147***	-0.182***	-0.175***	-0.057**	-0.070***	-0.069***
	(0.020)	(0.019)	(0.018)	(0.022)	(0.022)	(0.020)	(0.023)	(0.023)	(0.023)
hschool	0.283***	0.059*	0.093***	0.409***	0.156***	0.184***	0.210***	0.008	0.046
	(0.035)	(0.032)	(0.030)	(0.056)	(0.052)	(0.048)	(0.036)	(0.035)	(0.034)
lschool	0.528***	0.093***	0.079***	0.665***	0.161***	0.132***	0.447***	0.062**	0.062**
	(0.037)	(0.029)	(0.026)	(0.051)	(0.042)	(0.041)	(0.035)	(0.030)	(0.028)
Cohort FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE (province)	No	No	Yes	No	No	Yes	No	No	Yes
Sample Size	Whole	Whole	Whole	Female Only	Female Only	Female Only	Male Only	Male Only	Male Only
Observations	202,589	202,589	202,589	107,356	107,356	107,356	95,233	95,233	95,233

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Census 2006 and School Census of Iran (2014)

1.5.2 Return To Education

Return To Education Using Household Income and Expenditure Survey 2006

In this section, we are presenting our estimates on return to education in Iran. As I explained above, I start with the standard Mincer regression similar to 1.2 using the Household Income and Expenditure Survey (2006). Using Poverty Mapping technique, I present theses estimates both for migrant and non-migrant sub-samples using the Census (2006). I also use school availability as an instrument to check is there any bias in our initial results. Tables 1.8, 1.9, and 1.10 present estimates of return to education for HIES (2006). Table 1.8 shows the return for education given years of education as the independent variable. I then run the standard Mincer Regression when I put a limit on years of education. This exercise helps us to see how robust is my results and the idea behind that is since there is a very small variation for years of education above bachelor degree, I can replace years of education above that with 16. The estimates are presented in 1.9. I also run the regression with the level of education instead of years of schooling. The reason is the sample documents education code of individuals which can be transformed to the level of education very easily and precisely. In contrast with years of schooling, I face a smaller risk of measurement errors. Table 1.10 provides the estimates for the level of education. The same as traditional Mincer regression, the experience function is concave, however, the years of schooling function behave as a convex function. It means that initially as years of schooling increases the change in wage is decreasing but after a point, increasing in years of schooling translating into a positive change in wage. One can explain this phenomenon in a way that before the threshold maybe people are hired in low skilled jobs and for these kinds of jobs experience is much more important than schooling. As a result, additional years of schooling does not increase the wage (since the potential experience decline due to additional schooling). However, if the individual's education is beyond that threshold, then high skill level type of jobs are available for that person so additional schooling translates into higher wages. This is compatible with the inverse U-shape of wage with experience. We define experience as age-6-years of schooling so as experience goes up for a given age, the schooling should go down which means the marginal rate of wage would be negative. The benchmark specification in our exercise is the first regression where I include only exp^2 which is compatible with standard Mincer regression. In this case, the return for schooling is about 9.96% which is quite similar to other countries where the return to schooling are estimated between 10-15%. As Salehi-Isfahani et al. (2009) pointed out, the standard Mincer equation assumes a constant value of each year of schooling. There are various studies talking about non-linearity returns for schooling (see Trostel 2005). From this perspective, I estimated the quadratic function for $yeduc^2$. The estimates are presented in column 2 of table 1.8. The weighted mean for years of schooling in my sample (men, age between 25-60 years old) in 2006 was 6.75 years and at this value, the marginal value of return to schooling using column 2 specification would be:

$$\frac{\partial lnwage}{\partial yeduc} = \gamma_1 + 2\gamma_2 yeduc \tag{1.7}$$

calculating the above derivative at $\overline{yeduc}$ will give me the return equal to $0.047 + 2 \times 0.003 \times 6.75 = 0.0875$ or since I have log dependent variable, the private return to schooling will be about 9.14%. Columns three and four is the replication of our benchmark analysis using district level fixed effect. Adding district fixed effect, the results are quite robust and similar to benchmark. The only difference is I have better fitness (higher R^2) when I have district level fixed effect. The same trend can be seen in table 1.9 . I don't use quadratic specification because I fix *yeduc* to 16 for all people with the education above 16 years (or college degree). The private return to schooling is 10.1% and 9.75% respectively. The third model that I use to estimate the private return to education is using education level instead of years of schooling. As I explained above, in this specification, I face a lower risk of measurement error due to the fact that I have access to education code rather than years of schooling. The base level is no education. The estimates show that there is an increasing return to the level of education becomes higher. On average people with a university degree has higher return compare to upper secondary and lower level of education. This is compatible with the human capital theory. Considering the effect of location (district fixed effect), the coefficient for the university is 1.387 which means compare to illiterate the private rate of return for a person with a university degree is approximately 300% higher ($e^{1.387} \cong 4.00$).

Return to Education Using Census 2006

I did the same exercise using Census data. As mentioned above, working with census data gives me the advantage to see the effect of the return to schooling both on migrant and non-migrant samples. Using "Poverty Map"' technique, I can simulate log of annual wage for individuals in census data using HIES 2006. I ran the simulation for 100 times and used the average of that as my dependent variable. Although the PovMap2 software is basically designed for estimation of consumption model and deriving consumption per capita, however, my results for simulation of wage is quite convincing. There are some differences with HIES results. The return to education estimated by standard Mincer equation (1.2) from the census is lower than HIES, however, when we add $yeduc^2$ the estimates for $yeduc$ is quite similar to HIES. Columns 2 and 4 used equation 1.3, the coefficient for $yeduc^2$ is still positive but insignificant. Overall estimates from both HIES and the Census are quite comparable. Tables 1.11, 1.12, and 1.13 present the estimates of the return to education for years of schooling, years of schooling with limit on that, and level of education for the whole, non-migrant, and migrant sub-samples.

Table 1.8: Return to Education-Years of Schooling

Dependent Variable: Log Annual Wage				
	(1)	(2)	(3)	(4)
VARIABLES	lywage	lywage	lywage	lywage
yeduc	0.095***	0.047***	0.092***	0.042***
	(0.003)	(0.006)	(0.004)	(0.005)
$yeduc^2$		0.003***		0.003***
		(0.0004)		(0.0004)
exp	2.485***	3.114***	2.338***	2.981***
	(0.326)	(0.340)	(0.337)	(0.361)
exp^2	-1.994***	-3.506***	-1.808***	-3.364***
	(0.527)	(0.566)	(0.512)	(0.565)
Observations	16,622	16,622	16,622	16,622
R-squared	0.173	0.179	0.261	0.267
District Fixed Effect	No	No	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Sample: Men, 25-60 years old
Data: HIES-2006

Table 1.9: Return to Education-Limit on Years of Schooling

Dependent Variable: Log Annual Wage		
	(1)	(2)
VARIABLES	lywage	lywage
yeduc	0.096***	0.093***
	(0.003)	(0.004)
exp	2.355***	2.206***
	(0.325)	(0.334)
exp^2	-1.781***	-1.598***
	(0.525)	(0.507)
Observations	16,622	16,622
R-squared	0.169	0.258
District FE	No	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Sample: Men, 25-60 years old
Data: HIES-2006

Table 1.10: Return to Education by Education Level

Dependent Variable: Log Annual Wage		
	(1)	(2)
VARIABLES	lywage	lywage
Primary	0.210***	0.222***
	(0.030)	(0.026)
Lower Secondary	0.557***	0.534***
	(0.040)	(0.038)
Upper Secondary	0.897***	0.857***
	(0.043)	(0.045)
University	1.418***	1.387***
	(0.050)	(0.063)
Other	0.107	0.133*
	(0.080)	(0.076)
exp	3.244***	3.022***
	(0.329)	(0.342)
exp^2	-3.846***	-3.529***
	(0.542)	(0.533)
Observations	16,622	16,622
R-squared	0.172	0.262
District FE	No	Yes
Marginal Effects		
Lower Sec to Primary	0.347	0.312
Upper Sec to Primary	0.687	0.635
University to Primary	1.208	1.165
Upper Sec to Lower Sec	0.340	0.323
University to Lower Sec	0.861	0.853
University to Upper Sec	0.521	0.530

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Sample: Men, 25-60 years old

Data: HIES 2006

Note:Marginal returns are the difference between the regression coefficients for the levels.

Table 1.11: Return to Education-Years of Schooling (Census)

	Dependent Variable: Log Annual Wage (simulated)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	lywage	lywage	lywage	lywage	lywage	lywage	lywage	lywage	lywage	lywage	lywage	lywage
yeduc	0.052***	0.046***	0.056***	0.050***	0.062***	0.059***	0.063***	0.068***	0.041***	0.039***	0.046***	0.043***
	(0.003)	(0.005)	(0.005)	(0.003)	(0.003)	(0.007)	(0.005)	(0.005)	(0.004)	(0.005)	(0.006)	(0.004)
$yeduc^2$		0.0004		0.0003		0.0001		0.0002		0.0001		0.0002
		(0.0003)		(0.0004)		(0.0004)		(0.0004)		(0.0003)		(0.0003)
exp	1.999***	2.083***	2.489***	2.562***	2.263***	2.294***	2.850***	1.799***	1.530***	1.563***	1.932***	1.983***
	(0.288)	(0.344)	(0.331)	(0.413)	(0.308)	(0.372)	(0.312)	(0.556)	(0.301)	(0.335)	(0.346)	(0.416)
exp^2	-1.877***	-2.095***	-2.361***	-2.551***	-1.994***	-2.076***	-2.719***	-0.360	-1.641***	-1.723***	-1.849***	-1.975***
	(0.374)	(0.522)	(0.361)	(0.569)	(0.396)	(0.561)	(0.347)	(0.861)	(0.441)	(0.532)	(0.412)	(0.578)
Observations	107,768	107,768	107,768	107,768	59,680	59,680	59,680	59,680	47,675	47,675	47,675	47,675
R-squared	0.076	0.077	0.426	0.426	0.094	0.094	0.463	0.464	0.059	0.059	0.384	0.384
Sample	Whole	Whole	Whole	Whole	NM*	NM	NM	NM	M*	M	M	M
District FE	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Sample: Men, 25-60 years old

Data: Census 2006 ; NM: Non Migrant, M: Migrant

As you see in table 1.11, the return to education for the whole sample from model 1 is around 5.34%. Using district fixed effect, the return to education would be 5.76%. Columns 5 to 8 are our models for non-migrant sample and columns nine to twelve show the results for the migrant sample. Overall, the return for non-migrant is higher than the migrant. In the context of Iran, this result can be explained by the fact that on average poor people in the removte and rural area are more likely to go to larger cities seeking jobs and most of them will end up living in the suburb where they neither have access to decent jobs nor good education. The average years of schooling for migrants are lower than non-migrants and the percentage of illiterate people is higher among migrants. For both samples $yeduc^2$ is insignificant again. In the standard specification of Mincer regression, the return for non-migrants and migrants is 6.40% and 4.19% respectively. Adding district FE, the explanatory power of regression increases for both samples and the return to schooling becomes higher.

Table 1.12: Return to Education-Years of Schooling (Census)

	Dependent Variable: Log Annual Wage (simulated)					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	lywage	lywage	lywage	lywage	lywage	lywage
yeduc*	0.054***	0.058***	0.063***	0.065***	0.043***	0.048***
	(0.003)	(0.005)	(0.003)	(0.005)	(0.004)	(0.005)
exp	1.923***	2.413***	2.190***	2.778***	1.456***	1.853***
	(0.281)	(0.317)	(0.307)	(0.306)	(0.292)	(0.327)
exp^2	-1.699***	-2.172***	-1.822***	-2.544***	-1.468***	-1.654***
	(0.371)	(0.352)	(0.401)	(0.348)	(0.432)	(0.393)
Observations	107,768	107,768	59,680	59,680	47,675	47,675
R-squared	0.076	0.426	0.094	0.462	0.059	0.384
District FE	No	Yes	No	Yes	No	Yes
Sample	Whole	Whole	NM	NM	M	M

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Sample: Men, 25-60 years old

Data: Census 2006

Yeduc is limited to 16 years

Table 1.13: Return to Education by Education Level (Census)

	Dependent Variable: Log Annual Wage (simulated)					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	lywage	lywage	lywage	lywage	lywage	lywage
Primary	0.197***	0.220***	0.270***	0.277***	0.173***	0.193***
	(0.018)	(0.014)	(0.025)	(0.015)	(0.021)	(0.018)
Lower Secondary	0.353***	0.378***	0.459***	0.466***	0.297***	0.314***
	(0.027)	(0.022)	(0.036)	(0.020)	(0.030)	(0.029)
Upper Secondary	0.578***	0.616***	0.721***	0.729***	0.467***	0.510***
	(0.033)	(0.037)	(0.039)	(0.029)	(0.040)	(0.047)
University	0.825***	0.884***	0.988***	1.006***	0.671***	0.749***
	(0.050)	(0.070)	(0.043)	(0.059)	(0.060)	(0.078)
Other	0.280***	0.249***	0.354***	0.318***	0.242***	0.208***
	(0.041)	(0.025)	(0.059)	(0.036)	(0.048)	(0.030)
exp	2.150***	2.626***	2.346***	2.920***	1.590***	2.028***
	(0.323)	(0.366)	(0.360)	(0.366)	(0.323)	(0.370)
exp^2	-2.276***	-2.748***	-2.230***	-2.949***	-1.819***	-2.128***
	(0.485)	(0.491)	(0.544)	(0.517)	(0.507)	(0.507)
Observations	107,768	107,768	59,680	59,680	47,675	47,675
R-squared	0.078	0.427	0.096	0.463	0.060	0.385
District FE	No	Yes	No	Yes	No	Yes
Sample	Whole	Whole	NM	NM	M	M

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Sample: Men, 25-60 years old

Data: Census 2006

The same trends are seen here in tables 1.12 and 1.13. The return for non-migrants are higher than migrants and compare to HIES estimates, the census estimates are lower. One reason could be explained by the fact that the poverty mapping model is mostly used to estimate per capita consumption rather than the wage. When I use level of education, the trend is again quite similar to HIES estimates. In all regressions, as the level of education increases, the return to that also increases. In this sample and taking into account the district fixed effect in our model, the return to university level is about 142.05% higher than a person with no education. For non-migrants, this is about 173.46% and for migrants, it is about 111.49%.

Return To Education: Instrumnetal Variables Identification

As I explain earlier, I use school availability as an instrument to check the potential bias in my OLS results. Unfortunately, due to lack of information in HIES 2006 about the geographical location of individuals, I could not use that dataset to check my 2SLS results with OLS results. However, the HIES 2012 has a variable which I can define the sub-district for each individual and allow me to merge HIES data with school data. I also use the Census 2006 to compare my IV results with the OLS results. I use following specifications for the first stage and then using first stage results, I run the standard Mincer regression as my second stage.

$$\begin{aligned} yeduc &= \alpha + \beta_1 pschool + \beta_2 lschool + \beta_3 hschool + u_i \\ yeduc &= \alpha + \beta_1 pschool + \beta_2 lschool + \beta_3 hschool + \mu_i + u_i \end{aligned} \tag{1.8}$$

where *pschool*, *lschool*, and *hschool* are dummies show school availability in sub-district of the individual. μ is the district FE. Tables 1.14 and 1.15 present the iv-estimates of Census 2006 dataset.

As it is obvious in results, there are quite robust coefficients for years of schooling (*yeduc*) and they are very similar to OLS results. It seems the bias in standard Mincer equation when we apply it for Iran is quite small. For instance, in 2SLS results, the return for schooling is about 5.02% for standard Mincer equation and whole sample. This number for OLS is about 5.34%. Even adding district FE does not change the trend that much. Using OLS the rate of return for the whole sample is 5.75% whereas in 2SLS it is around 4.08%. The

Table 1.14: Effect of School Availability on Schooling (First Stage, Census Data)

Dependent Variable: Yeduc, First Stage Results						
	(1)	(2)	(3)	(4)	(5)	(6)
	one	one	one	one	one	one
VARIABLES	yeduc	yeduc	yeduc	yeduc	yeduc	yeduc
pschool	1.403***	1.381***	1.345***	1.265***	1.448***	1.448***
	(0.061)	(0.065)	(0.079)	(0.083)	(0.097)	(0.104)
lschool	0.889***	1.076***	0.921***	0.994***	0.845***	1.221***
	(0.059)	(0.062)	(0.074)	(0.077)	(0.095)	(0.100)
hschool	1.331***	1.470***	1.450***	1.472***	1.196***	1.414***
	(0.069)	(0.071)	(0.086)	(0.089)	(0.112)	(0.116)
Observations	107,701	107,701	59,619	59,619	47,669	47,669
R-squared	0.039	0.083	0.045	0.103	0.033	0.099
District FE	No	Yes	No	Yes	No	Yes
Sample	Whole	Whole	NM	NM	M	M

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Sample: Men, 25-60 years old

Data: Census 2006

Table 1.15: Estimates of Return to Education (Second Stage, Census Data)

Dependent Variable: Log Annual Wage (simulated), Second Stage						
	(1)	(2)	(3)	(4)	(5)	(6)
	two	two	two	two	two	two
VARIABLES	lywage	lywage	lywage	lywage	lywage	lywage
yeduc	0.049**	0.040***	0.048**	0.046***	0.056**	0.035***
	(0.022)	(0.004)	(0.020)	(0.004)	(0.024)	(0.005)
exp	1.891**	1.932***	1.819**	2.294***	2.140**	1.468***
	(0.876)	(0.217)	(0.766)	(0.208)	(1.057)	(0.293)
exp^2	-1.841***	-2.151***	-1.856***	-2.542***	-1.930***	-1.630***
	(0.526)	(0.286)	(0.460)	(0.280)	(0.712)	(0.358)
Observations	107,641	107,641	59,599	59,599	47,629	47,629
R-squared	0.076	0.421	0.090	0.457	0.052	0.381
District FE	No	Yes	No	Yes	No	Yes
Sample	Whole	Whole	NM	NM	M	M

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Sample: Men, 25-60 years old
Data: Census 2006

bias is larger for non-migrant sample compares to whole and migrant sample. At most, there is 1.80% gap between the return to schooling using OLS and 2SLS. The Sargan test shows the instruments are quite valid. Basically, for the whole sample the p-value is 0.05 which we can reject it only at 10% but for migrant and non-migrant samples the Sargan test failed to reject the null which I can say my instruments are valid (For Migrants Sargan score=4.29, and p-value=0.11 and for non-migrants Sargan score=1.05 and p-value=0.59). Tables 1.16 and 1.17 present the estimates of two stages least square for Household Income and Expenditure Survey(2012). The results show that the return to education is about 6.07% and 6.50% respectively (for standard specification and using district FE). The Sargan test shows the instruments for both specifications are valid (we fail to reject the null hypothesis) and the Wu-Hausman test shows that we have endogeneity problem in the first place (p-value=0.0005 which means we reject the null that *yeduc* is exogenous). Tables 1.18 and 1.19 provides OLS estimates of return to education using years of schooling and level of education for HIES 2012 dataset. Specifically, 1.18 gives us a framework to compare OLS results with 2SLS results. As we compare column 1 in table 1.18 with column 1 in table 1.17, we can see that there is an upward bias in OLS compare to 2SLS. This trend was seen when we use Census 2006 data but in HIES the gap is larger. For instance, the return reported in column 1 of those two tables are 9.20% and 6.08% respectively suggesting 3% gap between these two estimates. Having FE in our model, this gap declines to 2.5%. Overall, I can conclude that there is an upward bias in OLS compare to IV-estimates which can be explained by ability bias. Although I face with some bias in OLS estimates, I can still get some intuitions from comparing my OLS results between HIES 2006 and 2012. Comparing columns 1 and 3 of tables 1.8 and 1.18, one can conclude the rate of return in education drop in Iran between 2006 and 2012 (from 9.64% to 8.98% or 0.66 percentage point). Taking into account non-linear specification (i.e. $yeduc^2$),for average years of schooling ($\overline{yeduc}$) I can calculate return to schooling the same as . For 2006 and using FE speification (column 4):

$$\overline{yeduc} = 6.75$$
$$\frac{\partial(lnwage)}{\partial(yeduc)} = 0.042 + 2 \times 0.003 \times 6.75 = 0.0825$$
$$\text{Return to Schooling} = 8.59\%$$

For 2012 and taking into account the information in column 4 I have:

$$\overline{yeduc} = 7.60$$
$$\frac{\partial(lnwage)}{\partial(yeduc)} = 0.050 + 2 \times 0.002 \times 7.60 = 0.0804$$
$$\text{Return to Schooling}=8.37\%$$

For standard model without FE (; column 2) we have return to schooling equal 9.14% and 9.02% respectively. So between 2006 and 2012 the rate of return to schooling in Iran decline by about 0.66 percentage point in standard Mincer equation model and about 0.22 if we use extended Mincer equation. One explanation is due to the effect of sanction on the economy. Iran has experienced a baby boom between 1977 to 1985 and the baby-boom generation would be entering to job market on 2003. However, because of lack of jobs and especially high-skilled jobs, some of this generation continued their education at the higher level (graduate school and professional level). This fact could be reflected in $\overline{yeduc}$ between 2006 and 2012 where the average years of schooling increases about one year. As those graduates start to enter the job market, the western sanctions hit Iran economy in 2010-2011 and those sanctions along with some problematic policies cause economic hardship in Iran which reflected on -6% economic growth in 2012. This economic hardship ends up lots of university graduates to work in lower level jobs or stay unemployed which translate into declining of the return to education. The same trend can be seen when we use level of education as an independent variable. Tables 1.10 and 1.19 give us the ingredients to compare 2006 to 2012. As it is obvious in table 1.19 the trend is similar to 2006 but the magnitude of estimates are smaller, which confirms the previous results I have for years of schooling. The estimates to university degree compare to no-education is about 1.257 in 2012 compare to 1.387 in 2006 which means the return of university degree declines by 48.79 percentage point. Although there may be bias in our OLS estimates, which may lead me to be skeptical about this number, the decline of university return is compatible with the economic and the labor market context of Iran. Iran has one of the largest unemployment rate among individual with the university degree in the region. As a developing country, Iran's economy does not produce that much high-skilled jobs and a good portion of people with university degree end up to work in lower level jobs. This phenomenon could explain the decline of return of university degree between 2006 and 2012.

Table 1.16: Effect of School Availability on Schooling (First Stage, HIES Data)

Dependent Variable: Yeduc, First Stage		
	(1)	(2)
	one	one
VARIABLES	yeduc	yeduc
pschool	1.306***	1.413***
	(0.169)	(0.111)
lschool	1.625***	1.902***
	(0.159)	(0.105)
hschool	0.959***	1.107***
	(0.181)	(0.118)
Observations	16,480	32,872
R-squared	0.042	0.094
District FE	No	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Sample: Men, 25-60 years old
Data: HEIS 2012

Table 1.17: Estimates of Return to Education (Second Stage, HIES Data)

Dependent Variable: Log Annual Wage, second stage		
	(1)	(2)
	two	two
VARIABLES	lywage	lywage
yeduc	0.059***	0.063***
	(0.012)	(0.009)
exp	4.951***	5.166***
	(0.455)	(0.375)
exp^2	-6.634***	-6.789***
	(0.473)	(0.466)
Observations	16,477	16,477
R-squared	0.186	0.234
District FE	No	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Sample: Men, 25-60 years old
Data: HIES 2012

Table 1.18: Estimates of Return to Education (HIES 2012)

Dependent Variable: Log Annual Wage				
	(1)	(2)	(3)	(4)
VARIABLES	lywage	lywage	lywage	lywage
yeduc	0.088***	0.056***	0.086***	0.050***
	(0.002)	(0.005)	(0.002)	(0.005)
$yeduc^2$		0.002***		0.002***
		(0.0002)		(0.0002)
exp	5.827***	6.270***	5.832***	6.316***
	(0.270)	(0.272)	(0.273)	(0.276)
exp^2	-6.827***	-7.908***	-6.928***	-8.113***
	(0.473)	(0.481)	(0.475)	(0.481)
Observations	16,507	16,507	16,507	16,507
R-squared	0.206	0.209	0.246	0.250
District FE	No	No	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Sample: Men, 25-60 years old
Data: HIES 2012

Table 1.19: Estimates of Return to Education (HIES 2012)

	Dependent Variable: Log Annual Wage	
	(1)	(2)
VARIABLES	lywage	lywage
Primary	0.228***	0.207***
	(0.028)	(0.025)
Lower Secondary	0.519***	0.482***
	(0.032)	(0.029)
Upper Secondary	0.840***	0.795***
	(0.037)	(0.033)
University	1.297***	1.257***
	(0.039)	(0.037)
Other	0.011	0.002
	(0.111)	(0.111)
exp	6.580***	6.624***
	(0.273)	(0.276)
exp^2	-8.585***	-8.783***
	(0.483)	(0.481)
Observations	16,507	16,507
R-squared	0.207	0.248
District FE	No	Yes
Marginal Effects		
Lower Sec to Primary	0.291	0.275
Upper Sec to Primary	0.321	0.313
University to Primary	1.069	1.050
Upper Sec to Lower Sec	0.321	0.313
University to Lower Sec	0.778	0.775
University to Upper Sec	0.457	0.462

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Sample: Men, 25-60 years old

Data: HIES 2012

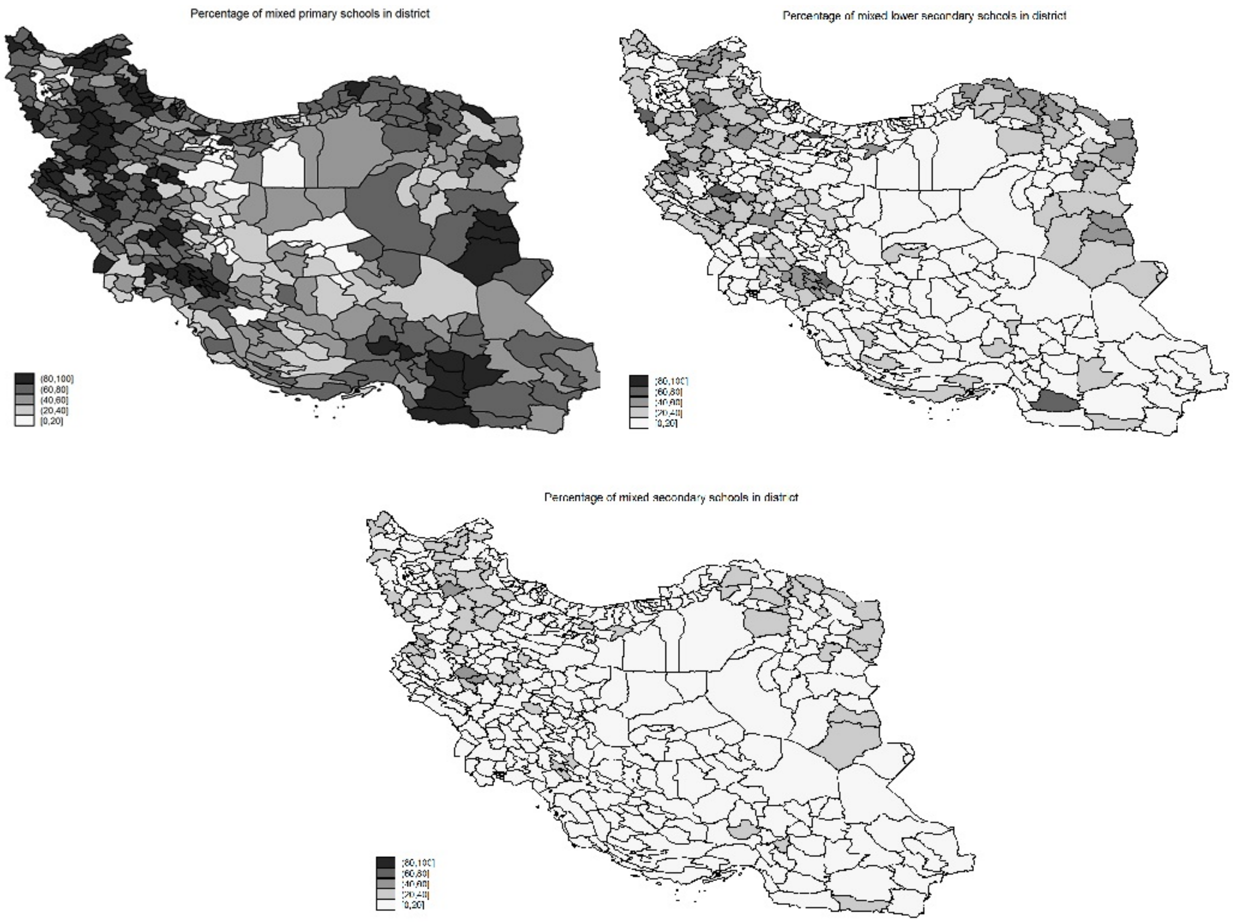
Note:Marginal returns are the difference between the regression coefficients for the levels.

1.6 Conclusions

In this paper, I try to shed light on the effect of school availability on education attainment as well as investigating the return to education in Iran. My results show that having high-school at the neighborhood of individuals would increase the chance of finishing high school and this effect is even larger for female than male. This is an interesting and important result in the context of Iran, especially from the policy point of view. There is a policy in Iran after the revolution to expand school supply in the country, especially in rural areas. In addition, the government pledged to segregate schools from the gender perspective. As figure 1.3 shows, most of the mixed schools in Iran are at the primary level. It seems gender segregated schools help girls to attend in the higher level of education. In other words, parents would more likely send girls to school if the school is gender-segregated. This is the hypothesis for future research and it is not in the scope of this paper. Expansion of school supply and access to girl-only schools can justify why the marginal effect of access to high school is larger for girls than boys. It is reasonable that when there is a high school in the neighborhood, it is more likely for girls to be in school while boys may stay out of school and work. This is compatible with Assaad et al. (2014) results where they estimated that the probability of not entering school is lower for most vulnerable girls compared to most vulnerable boys in Iran. In all regressions, availability of high school has a significant and positive effect on the probability of finishing high school. While the marginal effect is higher for non-migrant females compare to migrant ones, it is quite the same for males. I also estimate the probability of ever attend to lower secondary school condition on school availability and the patterns are quite similar. In the second part of this study, I estimate the return to education using the Mincer equation. I use both household surveys and Census data to present a complete picture of human capital in the labor market in Iran.

As (Card 2001) noted a standard solution for potential endogeneity problem in the Mincer regression, which gives us biased OLS results, is using instrumental variables. Among various instruments proposed in the literature, I use the geographic proximity of schools. It is reasonable to assume that availability of school is independent of students ability but it affects schooling choices. As I presented in first part of this study, there is a positive and significant relationship between probability of finishing high school or ever attend to lower secondary school and availability of high school and/or lower secondary school in a sub-district where an individual lives. In this sense, one can say school availability affects schooling choices but it is uncorrelated with unobserved ability factors. In most studies using IV, the IV estimates

Figure 1.3: Distribution of Mixed Schools by Education Level



Source: School Census in Iran

are larger than OLS estimates. However, there are some studies where IV estimates are smaller than OLS ones (see e.g (Duflo et al. 2001)). In this study, my IV estimates are lower than OLS estimates suggesting that there is an ability bias in my OLS estimates. In other words, those who have a higher level of education tend to have higher wages but part of that is not because of their education but because they have more mental and cognitive abilities. Correcting education with instrumental variables which are not correlated with ability may correct that upward bias. The Sargan test proposed that my instruments are valid. It is also interesting to note that overall the bias in OLS is not too large and the difference between IV and OLS is quite small. Using "poverty map" technique, I could apply my models (different specification of Mincer equation) to census data. This exercise gives me an opportunity to see the difference between the return among migrants and non-migrants. Overall, in this study, we have a higher return for non-migrant people than migrant people. The reason for this event could be explained by this hypothesis that in Iran, less educated and people in lower quantile of the income distribution are more likely to immigrate to the urban area to find jobs and most of them end up living in the suburb of large cities such as Tehran in a very low socio-economic status. This is a severe social problem in Iran and these results at least confirm this hypothesis.

I also apply Mincer equation for HIES 2012 dataset. This is the earliest household survey data available which I can locate individuals at the sub-district level and as a result I could do the IV exercise. Comparing IV estimates of HIES 2012 with Census 2006 and OLS estimates of HIES 2012 with HIES 2006 shows that the rate of return to schooling has experienced a decline over time in Iran. The IV estimates of HIES 2012 using district FE is 0.063 whereas it is 0.040 in Census 2006 (both are significant). Even if we neglect the Census data and just stick to household surveys, the OLS estimates proposed a decline around 0.006 which translates to 0.65 percentage point in return to schooling for standard Mincer equation. Using non-linear specification the rate of return to schooling would decline about 0.22 percentage point. Economic hardship and decline after 2010 in Iran which cause decline in demand for labor and especially high-skilled labors along with a positive shift in labor supply because of baby boom generation pushes an overall decline in the real wage. So at any given level of education, people make less money and this reflects in decline of return to education.

Chapter 2

Equality of Human Opportunities in the Middle East and North Africa

[1]This research has been supported by a grant from the ERF under the RIAD initiative.

2.1 Introduction

Since the Arab uprisings of 2011, there has been much attention on economic inequality in the Arab world as a possible source of popular dissatisfaction with the status quo. Curiously, the standard measures of inequality, such as the Gini index, do not place most Arab countries among the highly unequal developing countries, such as China or Latin American countries (Bibi and Nabli 2009, Belhaj-Hassine 2015). To reconcile the widely held perceptions of profound social and economic injustice in the region (Cammett et al. 2015) with low estimates of income inequality – the two parts of the "MENA inequality puzzle", as labeled by the Ianchovichina, Mottaghi, and Devarajan (2015) – researchers have looked for evidence of inequality beyond the standard indicators of income and consumption. Ianchovichina, Mottaghi, and Devarajan (2015) provides evidence of high levels of inequality in financial wealth and Salehi-Isfahani, Hassine, and Assaad (2014) and Assaad, Salehi-Isfahani, and Hendy (2013) find high levels of inequality of opportunity in education in several MENA countries. On the other hand, evidence of economic mobility from wage data in Egypt does not indicate particularly high levels of inequality of opportunity (Belhaj-Hassine 2012).

In this paper, we provide estimates for yet another dimension of inequality of opportunity, access of young children to basic services, such as electricity, water, sanitation, and schools, measured by the recently developed Human Opportunity Index (HOI). Although access to these services is sometimes listed as sources of inequality in the region, for example by Cammett et al. (2015, 3), our findings do not suggest that inequality of access to basic services is particularly high in the MENA countries that we study.

Of the various indicators of inequality, the HOI is particularly appealing because, unlike standard measures of inequality, such as the Gini index of income or consumption, it is forward looking in that it captures the inequality of opportunity faced by a rising generation. The HOI is closely connected with the concept of equality of opportunity as developed by Roemer (1998) and empirically implemented by Bourguignon, Ferreira, and Menéndez (2007) and Ferreira and Gignoux (2011), among others. It measures the extent to which access to basic services that matter for the productive future of a child depends on circumstances beyond his or her control. The circumstances are defined by the child's gender and the characteristics of the family and the community into which he or she is born.

Another appealing aspect of HOI is that the services it evaluates are for the most part publicly provided. This makes the HOI especially relevant for understanding the effectiveness of MENA governments in providing children from different backgrounds with an equal chance to succeed. MENA governments play a large role in their economies and do so often in the name of the poor. Since independence, many countries of the region have been ruled by dominant, populist states that promised an equitable distribution of the fruits of state-led economic development – and especially to provide equitable access to health and education – in exchange for political acquiescence. The question of how equitable is the provision of basic services is important for the debate on the relevance of the so-called "authoritarian bargain" (Desai, Olofsgård, and Yousef 2009) and the extent to which authoritarian MENA government have kept their end of the bargain. If the authoritarian bargain has empirical validity, we should expect that the type of inequality of opportunity measured by the HOI is less acute than, for example, inequality of opportunity in educational achievement, which depends greatly on parental resources and has been estimated to be high in several MENA countries Salehi-Isfahani et al. (2014).

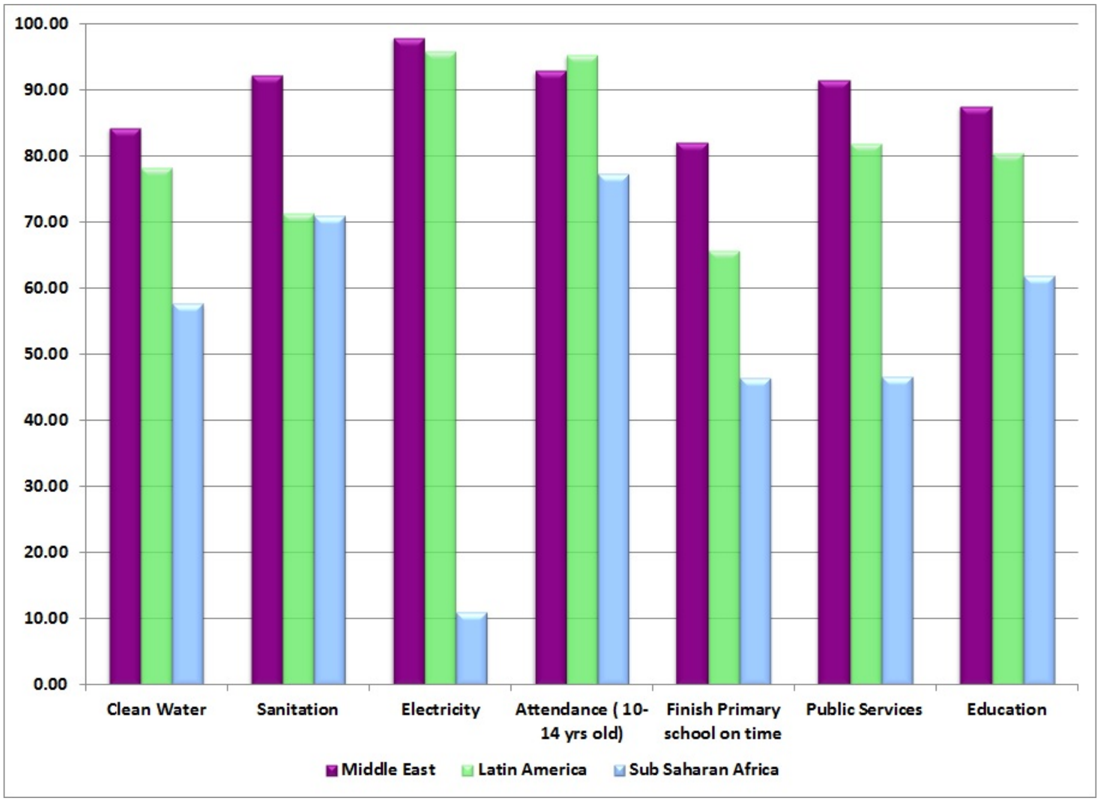
The main contribution of this paper is to provide estimates of HOI in a comparative framework, both over time and across countries in MENA. Existing estimates of HOI in the MENA region are confined to Egypt. Vélez et al. (2012) use Egyptian Household Income and Expenditure Surveys for 2000 and 2009, and Aran and Ersado (2013) utilize both HIES and the Demographic Health Surveys (DHS) of 2000 and 2008 to compute HOI for housing and education opportunities. El-Kogali and Krafft (2015) use the HOI methodology to estimate inequality of opportunity in early childhood development for a sample of 12 MENA countries. In this paper, we provide estimates of HOI for a sample of 10 MENA countries. For several countries for which we have more than one year of data, our estimates show how the HOI has changed over time. Since we are able to keep the set of circumstances constant over time, comparisons over time offer meaningful policy implications. As de Barros (2009) note, the main purpose of HOI is to measure progress over time in children's access to the basic opportunities that are publicly provided, both in the extent of coverage and equity in their distribution.

We also offer a comparative perspective with Latin America (de Barros 2009 and Molinas et al. 2010, Sub-saharan Africa (Dabalen, Narayan, Saavedra-Chanduvi, and Suarez 2014), and Asia (Son 2013), for which similar estimates exist. We follow recent advances in the measurement of equality of opportunity in access to services pioneered at the World Bank (de Barros 2009). As we see below, like the generalized Lorenz curve, the HOI is sensitive to both the mean, or coverage, and the variance of the distribution of access to a particular service. We provide a decomposition of how the change in HOI between two points in time is accounted for by changes in its coverage and equity its distribution. We also offer the contribution of different circumstances – e.g., family background vs. community characteristics – to total inequality in access to a service.

The empirical literature on inequality of opportunity, and in particular the approach of de Barros (2009), has been criticized for its limited policy implications. Kanbur and Wagstaff (2014) are concerned that statistics that show the contribution of a circumstance variables to inequality of opportunity is low, even with the warning that it is a lower bound, can mislead policy makers into thinking that inequality of opportunity is moderate or manageable. They argue that the conventional measures of inequality, such as the poverty rate, are less open to such ambiguity. This is a valid criticism of empirical studies of inequality of opportunity in which an outcome – such as income – has at least two types of influences, circumstances, and effort. HOI analysis is not strictly speaking a decomposition between circumstances and effort because access to publicly provided basic services depends fundamentally on their availability than an individual or family effort to access them. Of the services we study only reaching sixth grade on time can be said to depend in part on child and parent effort. Estimates of the extent of inequality in access to these services is therefore not a simple decomposition into circumstances and effort. Another reason why the critique of inequality of opportunity by Kanbur and Wagstaff (2014) does not reduce the value of our results is that we use a given set of circumstances, albeit incomplete, to measure access to the same basic service over time, and between MENA and Latin America. The fact that our estimates of inequality of opportunity are lower bounds does not reduce their value to policy makers who benefit from knowing how access to a service they provide has changed over time in their countries, and how it compares with the same in other countries in the region and outside.

Our main finding is that overall access to basic services in MENA countries, though still relatively inequitable, have improved and compared favorably with Latin America. The improvement over time is not surprising because as a service expands, especially if it starts inequitably, it generally covers more of the less advantaged households. In the limiting case, when coverage is complete, the distribution is necessarily equitable and the HOI reaches its maximum value of one. Our results also provide an interesting comparison between MENA and Latin America. Figure 2.1 depicts our estimates of the HOI for MENA for two periods roughly ten years apart (around the turn of the century and a decade later) and estimates computed by de Barros (2009) for Latin America. The average values at both points in times

Figure 2.1: Comparing HOI's for MENA and Latin America



Note: Simple averages are for groups of countries for which estimates are available for two years at least 5 years apart.
Source: Latin America: Molinas et al. (2010); MENA: authors' calculations.

are higher in MENA. Both regions experienced notable improvements. The improvement in MENA that is particularly noteworthy is in housing conditions. Over the decade this index in MENA increased from 64% to 83%, while in Latin America it increased from 58% to 69%. The education HOI, which was already quite high in MENA, increased from 81% to 86%, compared to 70% to 78% for Latin America.

The remainder of this paper proceeds as follows. The next section describes the harmonized HIES data we use. Section 2.3 describes the methodology of calculating the HOI. Section 2.4 presents the results, and section 2.5 concludes.

2.2 Data

Studies of inequality of human opportunities usually use surveys, such as HIES, that include information on household access to basic services, schooling of children 0-16 living with their parents, and other family characteristics. We take advantage of the harmonized income and expenditure survey for MENA countries provided by the ERF. We supplement these data with several surveys from Iran and the most recent (2012) household socio-economic survey from Iraq, available from the World Bank. Most countries in our study appear more than one year; we have multiple surveys from Iran (18) and Palestine (9), but only one survey each from Morocco, Sudan, Syria, and Yemen. We use a total of 43 surveys in this study but for brevity limit our discussion to surveys from around the beginning and the end of the decade 2000-2010 (see Table 8 for the complete list of the surveys and their dates). These surveys are well-suited for the purpose of cross-country analysis because they have harmonized definitions of variables measuring access to basic services and education. In almost all cases, access to a service is a binary variable.

Our working samples are constructed for children 0-16 years old with information on their own characteristics, those of their parents and the characteristics of the households in which they live. Limiting the child age to 16 ensures that more than 95% of the children are still living with their parents. The samples range in size from about 5,800 to 77,000 (see Table 9).

Our definitions of variables closely follow those in the harmonized data as well as those used by de Barros (2009). An important exception is the definition of sanitation, which can vary from country to country. In some surveys, sanitation is minimally defined as a latrine, which could be a hole in the ground with no running water. In the harmonized data the variable designating toilet facilities signifies a higher level of sanitation than a latrine and sometimes indicates a flush toilet. In Iran, sanitation is a binary variable for having a bath in the house, which usually includes a flush toilet and hot water, and thus sets a higher standard than we use in other countries in our sample. Because we use this variable as our sanitation variable, our estimates of housing HOI for Iran are underestimates. Iran's 2006 census reports the presence of both bath and toilet; 99.1% of families had access to a toilet compared to 84.8% to a bath.

2.3 Methodology

The HOI is based on a social welfare function first suggested by (Sen 1976), which defines social welfare as the multiple of the mean access to an advantage and a measure of equity in its distribution. The HOI is thus composed of two elements: (i) the coverage rate of a basic service; and (ii) how equitably that service is distributed relative to the distribution of circumstances. The index increases with the coverage rate and declines with the degree of inequality of access to a service (the details are explained in the technical note de Barros and Molinas Vega 2008). de Barros (2009) suggest that the HOI indices for each service can be averaged to arrive at a single index for inequality of opportunity.

To make this concept operational, two choices have to be made: what are the basic opportunities and what is the population of interest. From the equality of opportunity perspective, inequities that matter are those that result from the circumstances beyond individual control, which suggests the population of interest should be limited children for whom access to opportunities is beyond their control. One could include adults, but then the HOI would be closer to a static notion of inequality, like the Gini index.

The choice of what to include in the set of "basic opportunities" is less clear. In their pioneering study of HOI, done for Latin American countries, de Barros (2009) define basic opportunities as "all those essential to ensuring that today's children will have the potential, as adult, to better achieve the outcomes of their choosing." They limit the choice of basic opportunities to basic services, such as electricity, clean water, sanitation, and education. We consider access to basic services as access to opportunities for housing quality. Education opportunities are measured by school attendance at ages 10-14 and completion of sixth grade on time. The choice of opportunities to include is also determined by the availability of comparable data across time and countries. In this paper, we follow de Barros (2009) and consider opportunities that are associated with access to publicly provided basic services.

To get a bit more specific, let us define more precisely the two components of the HOI: the average level of access to a specific service, $\bar{p}$, and equity in access to that service. The latter is measured by a version of the "dissimilarity index (D)", which is widely used in sociology. The D-index measures the disparity of access rate for a service (such as clean water) by different groups of people (or types) defined by their circumstances, such as rural girls with illiterate parents. So, for example, if the access rate for a given service for a particular type is x% and the average access rate for the whole population is y%, then the dissimilarity between x and y influences the D-index. The D-index is constructed so as to range between 0 and 1, with $D = 0$ for perfect equality of opportunity and 1 for perfect inequality. The HOI is then defined as:

$$HOI = \bar{p}\,(1 - D)\,, \tag{2.1}$$

where $\bar{p}$ is the average rate of coverage of the service and $(1-D)$ measures equality in access.

D can be estimated through a variety of parametric, semi-parametric, or nonparametric methods. de Barros and Molinas Vega (2008) suggest to estimate it using a logistic regression in which the dependent variable, I, is binary, indicating access when $I = 1$ and zero otherwise. This methodology is implemented using the module hoi.

$$\ln(\frac{P(I=1|X)}{1-P(I=1|X)} = X\beta, \tag{2.2}$$

where X is a vector of circumstance variables (see Table 2.1):

Using the estimated coefficients from the logistic regression 2.2, this procedure computes each individual's predicted probability of access to an opportunity, $\hat{p}_i$ as:

$$\hat{p}_i = \frac{\exp(X_i\hat{\beta})}{1+\exp(X_i\hat{\beta})} \tag{2.3}$$

The estimates D is then computed using the distance between $\hat{p}_i$ and the average access rate, $\bar{p}$.

$$\hat{D} = \frac{1}{2\bar{p}}\sum_{i=1}^{n} w_i|\hat{p}_i - \bar{p}|, \tag{2.4}$$

where w_i are sampling weights, $\bar{p} = \sum_{i=1}^{n} w_i\hat{p}_i$, and n is the number of individuals. D is equal to zero if $\hat{p}_i = \bar{p}$ for all i, and equal to one if only one person has access.

2.4 Results

In this section, we discuss our results for housing and education.

Table 2.1: Specification of separable logistic regression function

List of circumstances	Specification
Child gender and age	dummy
Parents' education	quadratic
Per capita household expenditures	logarithmic
Number of siblings	linear
Presence of parents	dummy
Urban or rural	dummy
Gender of household head	dummy

2.4.1 Housing

The HOI for housing is a simple average of three separate indices for electricity, water, and sanitation. We report the estimates for all three components as well as their average in Table 2.2. Tables 2.3 and 2.4 report the estimates for the two parts of the HOI, the dissimilarity index, and coverage. For all our tabulated results, to avoid clutter, we present the estimates for two periods roughly corresponding to the years 2000 and 2010. The appendix B presents the graphs of all the years for which we have data.

The overall HOI for housing has improved over time, a result of both improved coverage as well as equity. The HOI varies widely across the region, from a high of 97.90 in Iran in 2010 to a low of 35.75 in Yemen, also in 2010 (see Table 2.2). Egypt, Iran, Iraq, Jordan, and Palestine all have housing HOI values above 90 percent. Unsurprisingly, Sudan and Yemen, the two poorest countries in our sample, have very low values. Morocco, also a relatively poor country by the region's standards, had a low value too (45.02%) in 2000 but we do not have an estimate for 2010. Jordan experienced a small drop in its housing HOI during 2006-2010, mostly the result of a water crisis that reduced access to clean water in 2010, especially in refugee areas. The surprising case here is Tunisia, with a housing HOI of 84.23% in 2010, which had dropped from 87.26 since 2005 despite a growing economy. One culprit in the case of Tunisia is the low value of HOI for piped water, itself a result of high inequality of access in Tunisia (Table 2.3).

We find the most favorable comparison between MENA and Latin America in the housing HOI. In MENA, in the later period (circa 2010), this index was 94.0%, up from 90.9% in the earlier period (circa 2000), compared to 69% in Latin America, up from 58.5% in the prior decade (differences are significant at the 5% level). The lowest housing HOI in MENA (Yemen 35.7%) was about the same as the lowest in Latin America (Nicaragua 33.5%). The MENA advantage is both in coverage and equity, as measured by the D-index. The average D-indices in Latin America for access to clean water, sanitation, and electricity are 14.1%, 18.3%, and 7.3%, respectively, whereas in MENA they are 7.3%, 2.5%, and 2.3% for these services, respectively.

Table 2.2: Human Opportunity Index for Housing Services (percent)

	Piped Water			Sanitation			Electricity			HOI_Housing	
Country	Circa 2000	Circa 2010	Change (%)	Circa 2000	Circa 2010	Change (%)	Circa 2000	Circa 2010	Change (%)	Circa 2000	Circa 2010
Egypt	75.15	84.92	0.75	98.25	99.29	0.08	97.67	99.73	0.16	90.36	94.65
	(0.29)	(0.41)		(0.09)	(0.18)		(0.11)	(0.06)			
Iran	86.92	94.67	0.70	57.43	88.87	2.86	96.56	99.13	0.23	80.30	94.22
	(0.16)	(0.13)		(0.21)	(0.18)		(0.09)	(0.11)			
Iraq	76.33	81.52	1.04	98.77	98.06	-0.14	99.82	98.94	-0.18	91.64	92.84
	(0.36)	(0.21)		(0.09)	(0.07)		(1.07)	(0.46)			
Jordan	95.59	92.00	-0.90	100.00	99.98	0.00	99.84	100.00	0.04	98.48	97.33
	(0.34)	(0.54)		-	(0.01)		(0.06)	-			
Morocco	25.53	-	-	62.91	-	-	46.63	-	-	45.02	
	(0.21)			(0.33)			(0.31)				
Palestine	77.88	89.79	0.79	100.00	99.89	-0.01	92.27	99.97	0.51	90.05	96.55
	(0.4)	(0.37)		.	(0.03)		(0.17)	(0.02)			
Sudan	-	48.06	-	-	80.86	-	-	60.69	-		63.20
		(0.74)			(0.62)			(0.67)			
Syria	71.84	-	-	-	-	-	99.14	-	-	85.49	
	(0.19)			.			(0.04)				
Tunisia	67.79	66.75	-0.21	99.13	87.59	-2.31	94.85	98.35	0.70	87.26	84.23
	(0.80)	(0.84)		(0.23)	(0.70)		(0.51)	(0.28)			
Yemen	-	28.04	-	-	45.33	-	-	33.88	-		35.75
		(0.36)			(0.39)			(0.35)			
Average	72.13	73.22		92.26	95.61		90.85	86.34		85.08	85.06
Average*	79.94	84.94		92.26	95.61		96.84	99.35		89.68	93.30

Note: Standard errors in parentheses. Change is percent annual. For actual survey years see Table 8. Average* refers to averages for countries with surveys for both periods.
Source: Authors' calculations.

Table 2.3: Dissimilarity Index in Housing (percent)

	Piped Water			Sanitation			Electricity		
Country	Circa 2000	Circa 2012	Change	Circa 2000	Circa 2010	Change	Circa 2000	Circa 2010	Change
Egypt	7.63	4.22	-0.26	0.31	0.15	-0.01	0.86	0.07	-0.06
	(0.53)	(0.67)		(0.12)	(0.22)		(0.24)	(0.11)	
Iran	4.89	2.15	-0.25	17.19	4.34	-1.07	1.36	0.28	-0.09
	(0.33)	(0.3)		(0.52)	(0.39)		(0.19)	(0.19)	
Iraq	10.02	7.30	-0.45	0.29	0.73	0.07	0.08	0.30	0.04
	(1.01)	(0.47)		(0.19)	(0.14)		(0.15)	(0.08)	
Jordan	0.87	1.45	0.12	0.00	0.01	0.00	0.16	0.00	-0.03
	(0.59)	(0.96)		-	(0.71)		(1.07)	.	
Morocco	40.78	-	-	15.79	-	-	24.59	-	-
	(0.99)			(0.8)			(0.98)		
Palestine	6.50	3.15	-0.21	0.00	0.06	0.00	0.86	0.02	-0.05
	(0.70)	(0.89)		-	(6.69)		(0.29)	(1.46)	
Sudan	-	16.48	-	-	7.08	-	-	14.96	-
		(2.08)			(1.25)			(1.55)	
Syria	7.96	-	-	-	-	-	0.28	-	-
	(0.40)						(0.07)		
Tunisia	14.87	16.18	0.81	0.35	5.25	-0.72	2.04	0.69	-0.03
	(2.14)	(2.39)		(0.49)	(1.57)		(0.99)	(0.62)	
Yemen	24.53	-	-	18.71	-	-	28.18	-	-
	(1.93)			(1.15)			(1.39)		

Note: Standard errors in parentheses. Change is percent annual. For actual survey years see Table 8.
Source: Authors' calculations.

Table 2.4: Coverage Rate in Housing (Percent)

	Piped Water			Sanitation			Electricity		
Country	Circa 2000	Circa 2012	Change	Circa 2000	Circa 2010	Change	Circa 2000	Circa 2010	Change
Egypt	81.36	88.67	0.52	98.56	99.45	0.06	98.54	99.81	0.09
	(0.21)	(0.30)		(0.07)	(0.14)		(0.07)	(0.04)	
Iran	91.39	96.76	0.45	69.35	92.89	1.96	97.89	99.41	0.13
	(0.11)	(0.08)		(0.18)	(0.12)		(0.06)	(0.07)	
Iraq	84.82	87.94	0.52	99.06	98.79	-0.04	99.89	99.24	-0.11
	(0.23)	(0.14)		(0.07)	(0.05)		(0.03)	(0.04)	
Jordan	96.42	93.35	-0.61	100.00	99.99	0.00	99.68	99.00	-0.14
	(0.26)	(0.41)		.	(0.01)		(0.03)	.	
Morocco	43.10	-	-	74.71	-	-	61.79	-	-
	(0.24)			(0.23)			(0.25)		
Palestine	83.30	92.72	0.59	100.00	99.94	0.00	98.11	99.98	0.12
	(0.30)	(0.28)		.	(0.02)		(0.11)	(0.01)	
Sudan	-	55.54	-	-	87.02	-	-	71.37	-
		(0.66)			(0.42)			(0.54)	
Syria	78.05	-	-	-	-	-	99.42	-	-
	(0.15)						(0.03)		
Tunisia	79.63	79.63	-0.65	99.48	92.45	1.49	96.83	99.04	0.08
	(0.53)	(0.54)		(0.14)	(0.43)		(0.32)	(0.16)	
Yemen	37.16	-	-	55.77	-	-	47.17	-	-
	(0.31)			(0.33)			(0.30)		

Note: Standard errors in parentheses. Change is percent annual. For actual survey years see Table 8.
Source: Authors' calculations

2.4.2 Education

The two education outcomes used to construct the education HOI are school attendance for children 10-14 years old and reaching sixth grade on time for13-year old children. We regard attendance as mainly the result of public sector investment; and reaching sixth grade on time as the result of a combination of public investment (in school quality) and parental effort. As a whole, MENA countries have improved according to both indicators (see Table 2.5). The average HOI in education increased from 86% around 2000 to 88% ten years later. The improvement was more pronounced in attendance than reaching sixth grade on time, suggesting that the region has done better in improving access to education than giving all students an equal chance to progress through school.

Again, performance in access to education opportunities is highly variable. Morocco in 2000 and Yemen in 2010 had the lowest HOI values in education (both in school attendance and in reaching sixth grade on time). Interestingly, Sudan, which had the second lowest score in reaching sixth grade on time, had one of the highest HOIs for attendance. Unlike most other countries, Iraq has a higher score in reaching sixth grade on time than in attendance, which implies greater inequality in access to school, and that those who could attend had a more equal chance to reach sixth grade on time. Between 2007 and 2012, the dates of our surveys from Iraq, there is no improvement in HOI for attendance but a slight increase – 2.7 percentage points – in reaching sixth grade on time. Tunisia is the surprising case because of its deterioration in attendance, but it showed improvement in the next category – reaching six grade on time – so its overall education HOI increased by 2.05 percentage points. A possible clue to this finding is in Table 2.6, which shows that Tunisia's D-index for attendance *increased* between 2005 to 2010 (the dates of the two surveys we use). Iran, Palestine, and Tunisia performed well in this period with 10.32, 15.84, and 8.62 percentage points improvement in HOI, respectively.

Comparison of the education HOI with Latin America again favors the MENA region. The average HOI in education in Latin America, as reported in Molinas et al. (2010), was 78% around 2008, compared to 88% in MENA (Table 2.5). There is an equal disparity between the two regions in the average D-index in reaching sixth grade on time: 4.4% in MENA compared to 10.2% in Latin America, indicating a much higher inequality in the latter region.

Table 2.5: Human Opportunity Index for Education (percent)

	School Attendance			Sixth grade On time			HOI_Education	
Country	Circa 2000	Circa 2010	%change	Circa 2000	Circa 2010	%change	Circa 2000	Circa 2010
Egypt	-	91.49 (0.61)	-	80.86 (1.01)	85.91 (1.29)	0.39	80.86	88.70
Iran	89.59 (0.26)	94.87 (0.26)	0.48	78.23 (0.73)	88.55 (0.83)	0.94	88.25	94.24
Iraq	77.63 (1.07)	77.97 (0.46)	0.07	84.42 (1.53)	86.79 (0.98)	0.47	81.03	82.38
Jordan	-	96.93 (0.59)	-	-	96.77 (1.26)	-	-	96.85
Morocco	66.14 (0.55)	-	-	69.86 (1.25)	-	-	68.00	-
Palestine	91.35 (0.69)	97.35 (0.42)	0.40	77.24 (1.78)	93.08 (1.41)	1.06	84.30	95.22
Sudan	-	97.60 (0.59)	-	-	59.80 (3.87)	-	-	83.31
Syria	85.42 (0.42)	-	-	85.55 (0.59)	-	-	85.49	-
Tunisia	90.47 (1.00)	85.96 (1.39)	-0.90	68.14 (3.25)	76.76 (3.71)	1.72	79.31	81.36
Yemen	-	88.28 (0.54)	-	-	46.12 (1.55)	-	-	67.20
Average	83.43	91.42		79.00	80.89		81.03	86.16
Average*	87.26	89.53		79.51	87.23		82.75	88.38

Note: Standard errors in parentheses. Change is percent annual. For actual survey years see Table 8. Average* refers to averages for countries with surveys for both periods.
Source: Authors' calculations.

Table 2.6: Dissimilarity Index in Education (percent)

	School Attendance			Sixth grade on time		
Country	Circa 2000	Circa 2012	%change	Circa 2000	Circa 2010	%change
Egypt	-	2.97	-	5.65	3.63	-0.16
		(1.18)		(1.74)	(2.1)	
Iran	3.73	1.86	-0.17	6.90	3.87	-0.28
	(0.49)	(0.55)		(1.41)	(1.34)	
Iraq	7.06	6.80	-0.05	5.22	2.53	-0.54
	(1.86)	(0.88)		(2.79)	(1.82)	
Jordan	-	1.01	-	-	1.13	-
		(1.2)			(3.1)	
Morocco	11.24	-	-	11.69	-	-
	(1.25)			(2.85)		
Palestine	2.12	0.88	-0.08	5.03	1.59	-0.23
	(1.02)	(0.85)		(3.09)	(2.24)	
Sudan	-	0.95	-	-	12.14	-
		(1.57)			(8.48)	
Syria	4.42	-		3.42	-	-
	(0.49)			(1.02)		
Tunisia	3.23	4.74	0.30	11.41	7.81	-0.72
	(1.85)	(2.81)		(6.40)	(7.14)	
Yemen	3.18	-	-	15.29	-	-
	(0.88)			(4.64)		

Note: Change is percent annual.
Source: Authors' calculations.

Table 2.7: Coverage Rate in Educational Opportunities (percent)

Country	School Attendance			Sixth grade on time		
	Circa 2000	Circa 2012	%change	Circa 2000	Circa 2010	%change
Egypt	-	94.29 (0.40)	-	85.7 (0.74)	89.15 (0.95)	0.27
Iran	93.06 (0.18)	96.67 (0.18)	0.33	84.03 (0.62)	92.11 (0.63)	0.73
Iraq	83.53	83.66	0.03	88.03	89.05	0.20
Jordan	-	97.92 (0.38)	-	-	97.88 (0.76)	-
Morocco	74.52 (0.44)	-	-	79.1 (0.91)	-	-
Palestine	93.34 (0.53)	98.21 (0.27)	0.32	81.33 (1.45)	94.58 (1.06)	0.88
Sudan	-	98.54 (0.35)	-	-	59.8 (3.07)	-
Syria	89.37 (0.20)	-	-	88.57 (0.46)	-	-
Tunisia	93.49 (0.70)	90.24 (1.04)	-0.65	76.92 (2.46)	83.27 (2.75)	1.27
Yemen	91.18 (0.40)	-	-	54.44 (1.39)	-	-

Note: Change is percent annual.
Source: Authors' calculations.

2.4.3 Time decomposition

The purpose of this decomposition is to measure the contribution of the two parts of the HOI, the change in average access (scale effect), $\Delta_{\bar{p}}$, and the change in the distribution of access, Δ_D. The change in HOI between two points in time can be written as (see (de Barros and Molinas Vega 2008)):

$$\Delta = HOI_1 - HOI_0 = \bar{p_1}(1 - D_1) - \bar{p_0}(1 - D_0).$$

Rearranging terms, we get:

$$\Delta = HOI_1 - HOI_0 = \underbrace{(\bar{p_1}(1 - D_0) - \bar{p_0}(1 - D_0))}_{\Delta^{\bar{p}}} + \underbrace{(\bar{p_1}(1 - D_1) - \bar{p_1}(1 - D_0))}_{\Delta^{D}} \tag{2.5}$$

Thus the contribution of change in coverage to change in HOI, $\Delta^{\bar{p}}$, is the change in coverage keeping D at its initial value, and likewise for the contribution of change in the D -index. This procedure is implemented as part of the Stata module hoi.

We perform this decomposition for six counties in our sample for which survey for more than one year is available: Egypt, Iran, Iraq, Jordan, Palestine, and Tunisia. The results are presented in Figures 2.2 and 2.3 for housing and education decomposition, respectively. Our decompositions refer roughly to change during the first decade of the 21st century, when, with the notable exception of Tunisia, all MENA countries improved their overall HOI. In all cases the scale effect dominates the distributional effect, meaning that the main source of improvement in opportunities has been the extension of services, rather than increased equity in their distribution.

First, consider the decomposition of the change in the HOI for housing. In the case of access to electricity, all the countries for which we had more than one year of survey data had very high nearly complete access to begin with. As a result, the change in their HOI is very small and there is not much change to decompose. For piped water, which, except in Jordan and Tunisia, the HOI has improved, the decompositions show that, on average, two-thirds of the improvement in the HOI is due to the scale effect and the remaining one-third is attributable to the distributional effect. In Jordan, which experienced a reduction in HOI, more than 89% of the decline in HOI is due to reduced coverage. In Tunisia, where HOI also declined, the entire reduction is explained by deterioration of equality in distribution. The HOI for sanitation increased in all MENA countries. In Iran, which led the pack by increasing its HOI from 57.4% to 89.0% between 2000 and 2010,[2] 62% of the expansion is estimated to have come from increased coverage. In Tunisia, where the sanitation HOI declined by 11.53 percentage points, the scale effect for this opportunity contributed 60.71% of the decrease.

Next, we consider the decomposition of the change in HOI for education. As noted earlier, except Tunisia, all countries improved their HOI in attendance; about 63% of the improvement can be attributed to increasing in coverage. In Tunisia, where the HOI in attendance deteriorated about 4.5 percentage points, about 30.2% was due to increased inequality in its distribution.

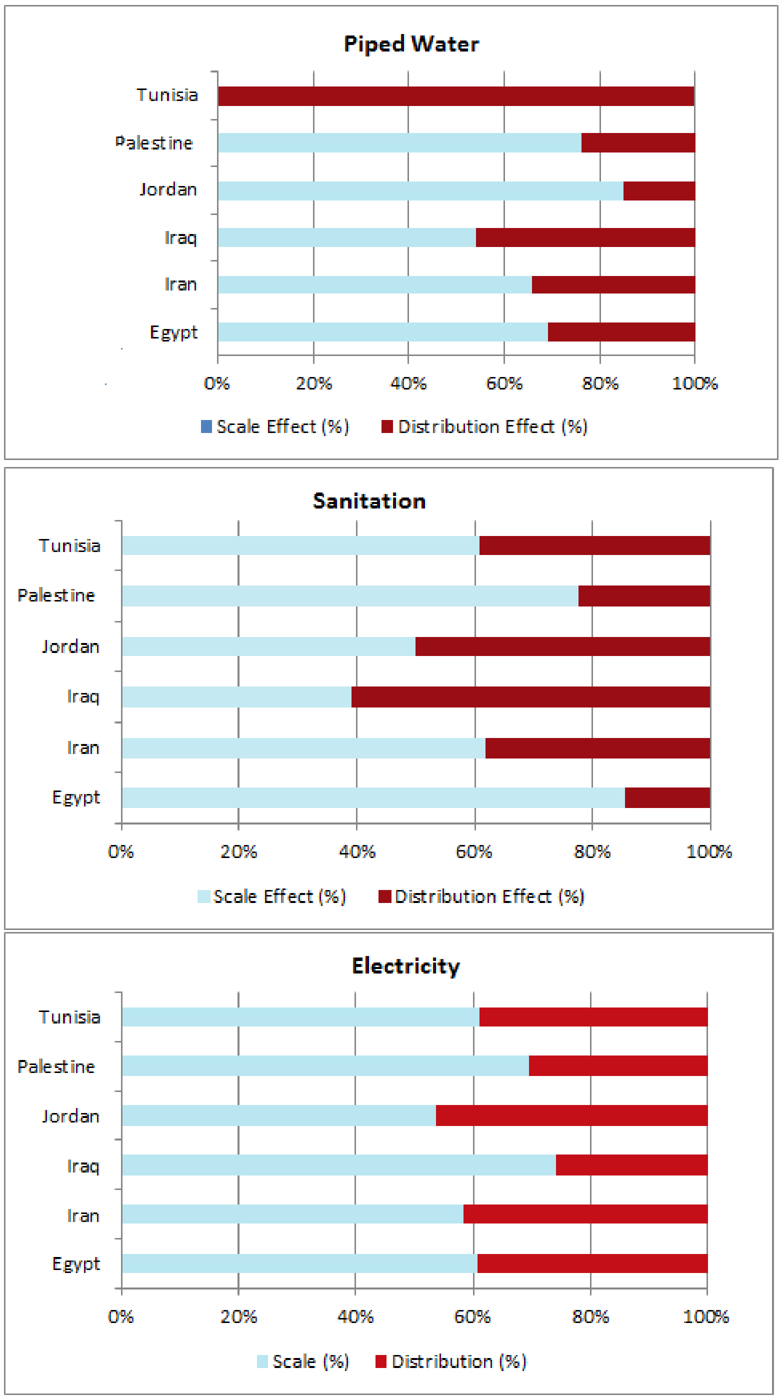
2.4.4 Shapley decomposition of the Dissimilarity Index

Measurement of inequality of opportunity is enriched by information about the contribution of individual (or groups of) circumstances to the index of inequality of opportunity in question. In the case of HOI, this information is in the decomposition of the D-index. The HOI literature has used the Shapley decomposition for this purpose, which we follow here. The idea behind this decomposition is the seminal work of Shapley (1953), who provided a method to quantify the payoff of each player in a cooperative game based on his or her marginal contribution. Shorrocks (2013) used this idea to model a general method to assess the contribution of a set of factors for a statistic such as an inequality index. His solution is based on "calculating the marginal impact of each of the factors as they are eliminated in succession and then averaging these marginal effects over all the possible elimination sequences" (Shorrocks 2013).

The Shapley decomposition results for the D-index using Stata command hoishapley are summarized in Figures 4-6.

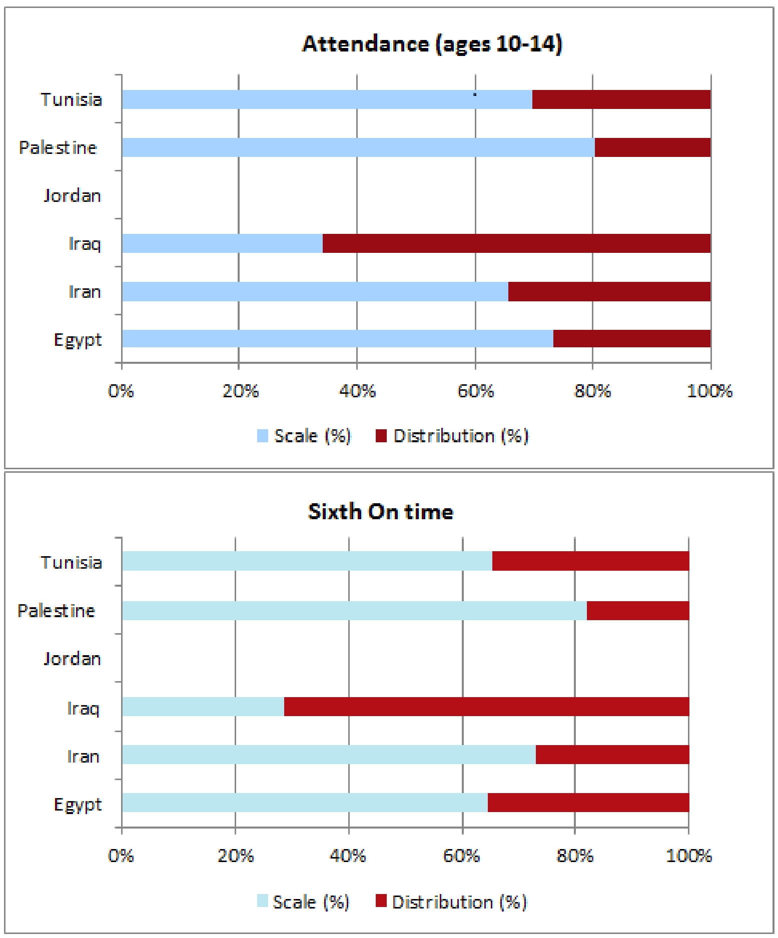
[2]As noted earlier, sanitation in Iran is measured by access to an indoor bathing facility.

Figure 2.2: Time Decomposition in Housing



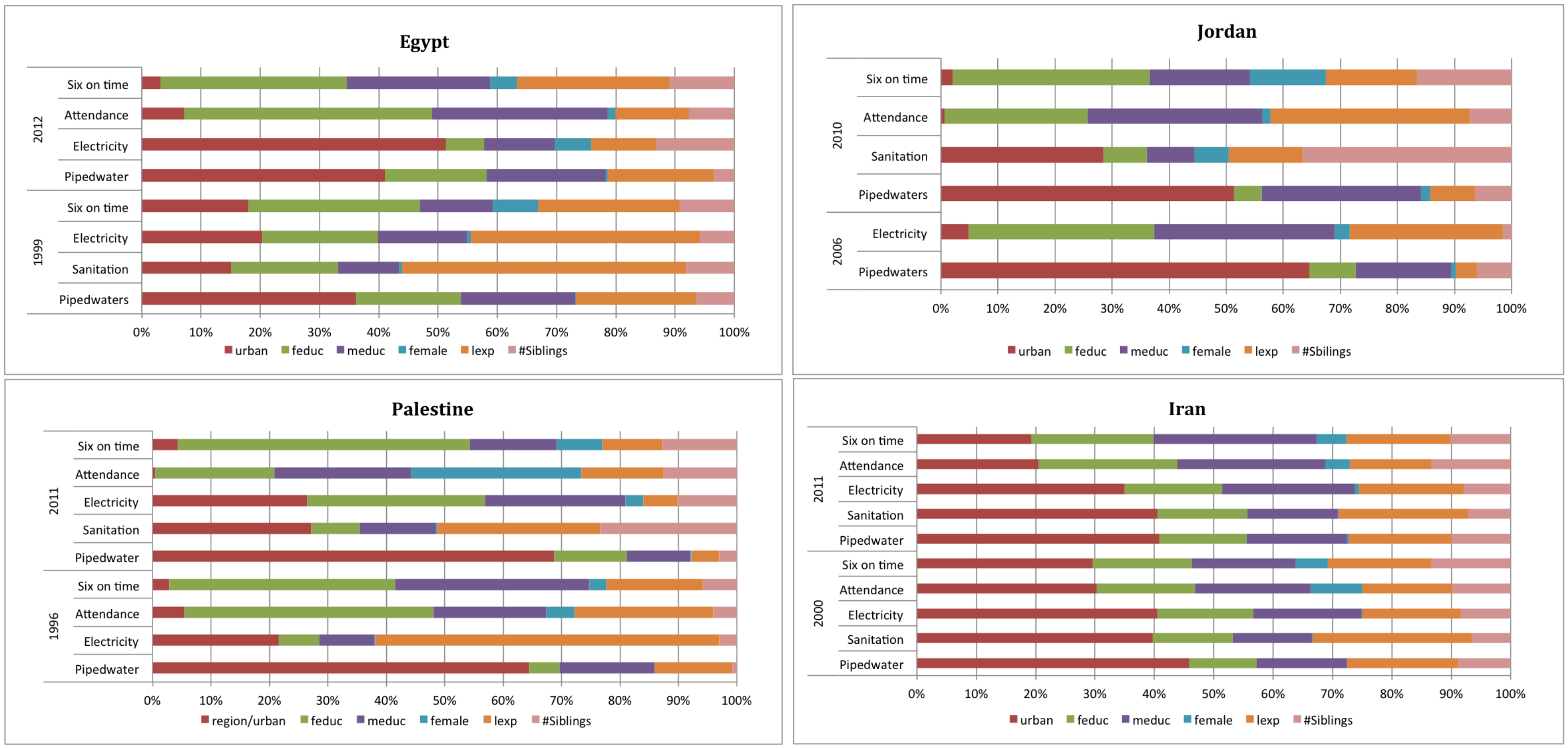
Source: Authors' calculations.

Figure 2.3: Time Decomposition in Education



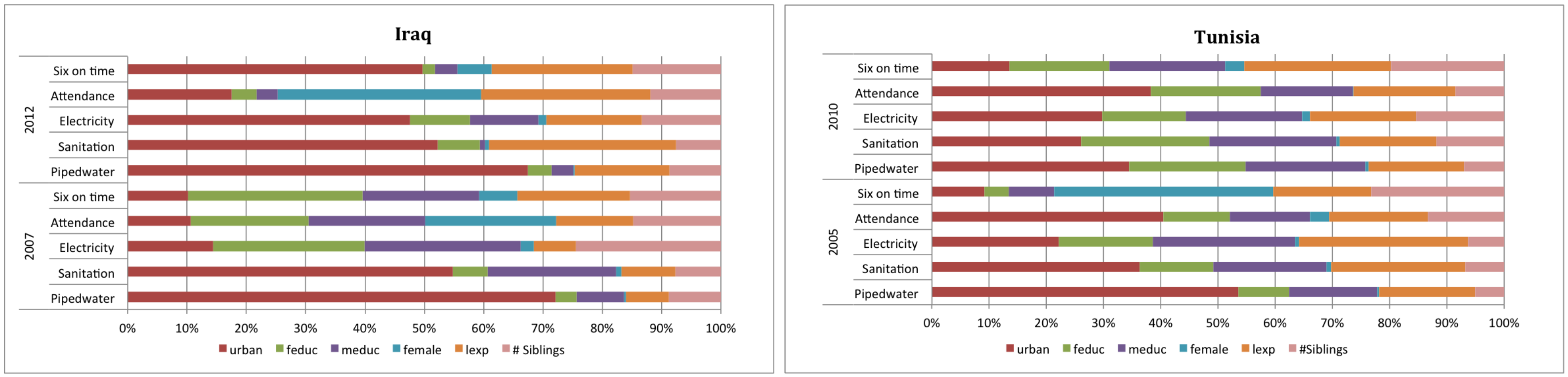
Source: Authors' calculations.

Figure 2.4: Shapley Decomposition for D-index



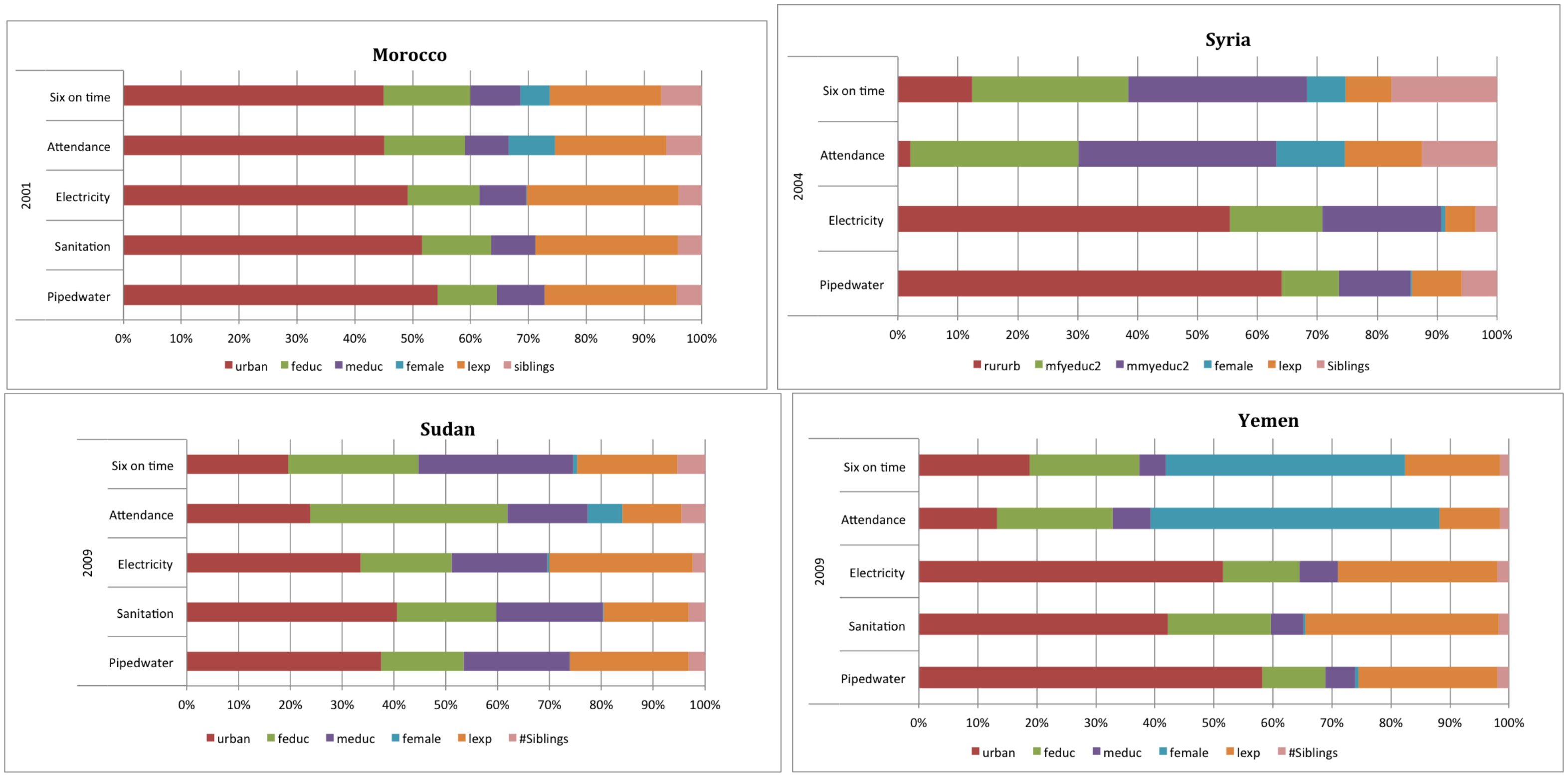
Source: Authors' calculations.

Figure 2.5: Shapley Decomposition for D-index



Source: Authors' calculations.

Figure 2.6: Shapley Decomposition for D-index



Source: Authors' calculations.

For most countries, the circumstance that explains inequality in access to basic services most is location – rural vs. urban – followed by household resources, measured by log expenditure per capita. This is easy to glean from the graphs for piped water, sanitation, and electricity. In the case of piped water, except in Egypt, Iran, and the Sudan, more than half of the inequality in access is explained by location. A high share of location, such as 72.14% in Iraq in 2007, implies inequitable government policies in provision of piped water, and a low share, such as 34.51% in Tunisia in 2010, signals more equitable policies, which is at odds with common wisdom about regional inequalities in Tunisia.[3]. The number of siblings matters in most cases, especially in Jordan, but its importance may be because of its correlation with fertility and hence household income and education rather than greater competition for resources among children. The gender of the child does not play an important role in equity in access to these services, indicating that boys and girls have equal access, most likely not an indicator of gender equity but a product of the fact basic infrastructure is supplied to the household unit, not individual members.

In contrast to basic services, education opportunities in most countries are most affected by parent education and income. In Egypt, Jordan, and Palestine, for reaching sixth grade on time, the latter account for more than 90% of the inequality as indicated by the D-index. In Syria, too, geography matters much less: attendance is nearly entirely accounted for by household characteristics, and for reaching sixth grade on time close to 90%. Gender is important for attendance in the two poorest countries in our sample, Iraq, and Yemen, suggesting that when resources are scarce it is girl education that suffers. In Yemen, about 50% of inequality is explained by gender and 22% in Iraq. According to these estimates, between 2005 and 2010, Tunisia managed to substantially reduce the contribution of gender to inequality in reaching sixth grade on time. The complementary roles of governments and parents in reducing inequality of opportunity of children is evident in the divergent decomposition results we obtained for basic services vs. education. Governments make the critical decisions to supply basic services whereas parents make the important decisions for schooling.

[3]See Robert Joyce, The Regional Inequality Behind Tunisias Revolution," *The Atlantic*, December 17, 2013; and the World Bank, *The Unfinished Revolution: Bringing opportunity, good jobs and greater wealth to all Tunisians, 2014.*

2.5 Conclusion

Populism in the Middle East has been an integral part of the authoritarian bargain (Desai, Olofsgård, and Yousef 2009). Since independence, most MENA countries have been ruled by autocratic leaders who promised to deliver education and basic services to their citizens. The failures of autocratic rule and the attendant the We presented estimates of the Human Opportunity Index for 10 MENA countries, for several at more than one point in time. The HOI measures the extent to which the basic opportunities that matter for a child to succeed are available and are provided equitably. It measures not only the average rates of access to these services but also the equity with which they are supplied. As such, the HOI is not only a useful measure of a country's progress in human development, in the case of MENA countries it provides evidence on the degree to which the region's authoritarian leaders have fulfilled their end of the bargain.

We began this paper by posing a MENA inequality puzzle: widespread feelings of social injustice in the region despite low estimates of inequality of outcomes observed from cross-section survey data. We also noted that recent evidence on inequality of opportunity – which is, roughly speaking, the part of the inequality of an outcome that is due to circumstances beyond individual control – helps resolve the puzzle by pointing out that the main game for upward social mobility – especially in education – is unfair. A large part of the inequality in early childhood development and educational achievement is accounted for by circumstances (El-Kogali and Krafft 2015; Salehi-Isfahani, Hassine, and Assaad 2014). In contrast, the evidence on inequality in access to basic opportunities that we present in this paper does not indicate particularly high inequality of opportunity. We find that the overall index of human opportunities has improved over time. We also find that the HOI for nearly all basic opportunities that we consider are higher than in Latin America, a region with roughly the same level of income and economic development. One way to view these findings is that they do not help resolve the puzzle. Broadly speaking, the authoritarian bargain has been for nothing. In Syria, a country rules for decades by a textbook case of a populist, authoritarian regime, has a higher HOI in housing than a typical country in Latin America, and considerable greater equity in the delivery of basic services that than Morocco, a MENA country with similar level of income but less harsh authoritarian rule.

One final thought: the rising and relatively high levels of HOI in certain MENA countries should not be understood to mean that the citizens of these authoritarian countries should be content with their political and economic systems. Indeed, the main reason why people at the bottom rungs of the economic ladder might be willing to give up their political liberties in exchange for economic improvement is to escape poverty. If the bargain did work for them and broad sections of the society experienced improvements, but there is still widespread discontent, it could well be the time they want their democratic rights back.

Chapter 3

Inequality of Education Opportunity: An Analysis of its Determinants

3.1 Introduction

Identifying barriers that preclude education attainment are some of the greatest concerns in the field of labor economics. Knowing about which circumstances reduce the student achievement is important for a policy design. Inequality of opportunity (IOP) is a new way of looking at justice and equity that tries to demonstrate the effect of various circumstances such as gender, ethnicity, and family background on the outcome. From the policy point of view, it is important to find which circumstances affect more an individual outcome. As a result, a good policy may alleviate the effects of those circumstances and equalize the opportunities among people.

In this paper, I compute an index for measuring inequality of educational opportunity. I choose education attainment as an outcome since several studies show the effect of early childhood education on the future life career. Education produces substantial returns to individuals in terms of earnings as well as economic growth (Wößmann 2008 or WorldBank 2005). Measuring inequality of opportunity in education for early childhood education and secondary level can illuminate the effects of policies in educational systems around the world. In this study, I first measure the size of IOP between fourth and eights graders and look at the relevant hypothesis to explain the difference between grade four and eight.

The initial results show that IOP increases in mathematics and science for most countries in grade 8 compare to grade 4. Several reasons may explain this results. As some researchers proposed tracking may explain part of these gaps (For instance see: Hanushek et al. (2006), Brunello and Checchi (2007)). Here I look at the correlation between the index of IOP and economic characteristics such as the Gini index (which explains the level of income inequality in an economy) and per pupil expenditures.

The analysis proceeds as follows: in section 3.2 I review the relevant literature, section 3.3 investigates part of the methodology that is used in this research. Section 3.4 and 3.5 present the data and results, and section 3.6 is concluding remarks.

3.2 Literature Review

Inequality and justice are one of the controversial issues both in political philosophy and economics. Egalitarianism in utilitarian's ethics means equality of utility. This definition of egalitarianism ignores the role of person's choices and preferences. Utilitarianism was the dominant theory in the distributive justice literature before John Rawls's famous book "A Theory of Justice"' in 1971 (Rawls 1999). Rawls (1999) tried to formulate a new approach to egalitarianism which inserted personal responsibility as an essential factor for qualifying degree of equality in a desirable and ethical way (John Roemer 2013). The objective of utilitarian is to maximize the sum of utilities for each individual in society. Rawls argues that justice is beyond the society welfare and it should focus on a set of "primary social goods." He suggested "rights, liberties and opportunities, income and wealth" as primary social goods.

The main philosophical contributions to the new approach towards justice and equality after Rawls theory were from (Sen 1980), (Dworkin 1981a; Dworkin 1981b), (Arneson 1989) and (Cohen 1989). For instance, Dworkin proposed the idea that people should be responsible for the choices they make, but they should be compensated for the random distribution of resources in society. In his famous papers, he considered two general theories of distributional equality. He said, "equality of welfare holds that distributional plan treats people as equals when it transfers resources among people until no further transfer would leave them more equal in welfare. On the other hand, equality of resources holds that it treats them as equals when it distributes so that no further transfer would leave their shares of the total resources more equal." His theory of marking off inequalities by their origins contributes the theoretical foundation for research on equality of opportunity.

The concept of equality of Opportunity (EOP) focuses that society should do to "level the playing field" among individuals who compete for positions. The other conception for EOP can be named as a "nondiscrimination principle". As (Roemer 1998) noted, the nondiscrimination principle means that in a competition for positions in society, all eligible individuals who possess the relevant characteristic should be included in the pool of candidates and they should be judged based on those characteristic and not based on attributes such as religion, gender, and ethnicity.

The literature in the field of equality of opportunity has been expanded during the past two decades. Some contributions have addressed policy implications of EOP, the problem of designing fair allocation rules with respect to EOP and measuring of IOP in different countries. In this study, I focus mostly on measuring the IOP and public policies, which have effects on the IOP.

3.2.1 Measuring of Inequality of Opportunity

One of the growing empirical literature in the field of EOP is measuring the degree of inequality of opportunity for education attainment, income or health care. The philosophical interpretation of equality as an equality of opportunity is applied in several empirical studies recently (see, for example, Bourguignon et al. 2007, Lefranc et al. 2009, Checchi and Peragine 2010 and Salehi-Isfahani et al. 2012)

(Bourguignon, Ferreira, and Menéndez 2007) quantified the role of IOP in generating inequality in current earnings of people in Brazil. They estimated the impact of circumstances both directly and indirectly on earnings. Their results showed that circumstances such as parental education, father's occupation, race and place of birth explain more than 20% of total earnings inequality within gender/cohort groups in Brazil in 1996. Bourguignon & Ferreira also showed that among various circumstances, family background is the most important circumstances affecting on a person's opportunities. It makes sense since a kid whose parents have university degree can be better educated at home and his/her human capital is much higher than a kid who comes from an illiterate family. In addition, wealthy parents can prepare better facilities for their off springs. This fact is considered in several papers such as (**?**) & (Schütz, Ursprung, and Wößmann 2008). Bourguignon & Ferreira estimated that more about 55%-75% of the total effect of circumstances can be attributed to parental education alone.

(Ferreira and Gignoux 2011) constructed an index of IOP which captures between group inequality when groups are specified on the basis of circumstances. The measurement index of IOP in their study is revealed that IOP shares ranging from 25% to 50% in the six Latin-American countries of total consumption inequality. Later on, (Ferreira and Gignoux 2013) showed that the simple R^2 of OLS equation of outcome on circumstances can be substituted in the ex-ante index of IOP and measured the level of it between countries. In this study, I applied both decomposition method and R^2 for measuring IOP among participating countries to investigate the trend of IOP around the world.

3.2.2 Equality Opportunity & Educational Policies

Other than measuring IOP in education, finding the effect of policies on equality of opportunity is investigated by several researchers. (Hanushek et al. 2006) show that early tracking increases inequality in achievement. They use a difference-in-difference method to address the effect of early tracking on the distribution of outcomes. Some of the literature has suggested that there is a channel for increasing inequality which supports the effects of family background. In fact, if early achievement is associated with the difference in family background, track placements will be related to family background. For instance, (Schnepf 2003) showed that family background is a driving force in setting track systems.

(Schütz, Ursprung, and Wößmann 2008) provide a comparable measure for 54 countries and show how strongly children educational performance is related to family background and also they identify the impact of education policies on EOP by using the cross-country variation in education policies and its interaction with a family background at the student level. Their results show that the variation between family background effects the student performance which is related to certain systematic features of the countries' education systems. Their specification indicates the significance of extensive early childhood education and late tracking. (Brunello and Checchi 2007) investigate the interaction between family background and secondary school tracking and its effect on human capital accumulation. They confirm that school tracking reinforces family background impacts through educational attainment while school tracking does not have any effect on literacy and on-the-job training. Brunello & Checchi concluded that school tracking has an ambiguous effect in the sample they used. They deduced that reducing the extent of student tracking may be good for social mobility in educational attainment, but it may increase social exclusion for people come from disadvantaged families. They also explained the effect of tracking in a theoretical framework. Taking advantage of (Benabou 1996), they considered a simple model to investigate the relation between tracking and peer effects. Advocates of tracking emphasize to the efficiency gains that this policy is made. In principle, efficiency is higher in a homogeneous class and teachers can teach better in a class where students have same ability level. On the other hand, some scholars discussed that lower ability students will worse off in a tracked system. As a result, we can say that tracking related to the trade-off between efficiency and equality. In their model, they show that if ability types are substitutes, a comprehensive educational system is more efficient whereas tracking system is more efficient when abilities are complements ((Betts 2011))

3.3 Methodology

In this study, I will investigate two different issues. First, I will measure Inequality of Opportunity (IOP) in education achievement for 20 countries around the world both for the fourth and eighth grade, and second I will investigate the relevant policies affected the IOP between grade 4 and 8. The ambitious question is what explains the level of IOP in the countries I considered. Answers to this question have clear value in policy implication of inequality of educational opportunity. In this paper, I pursue graphical exploring the correlates of IOP in the considered countries.

3.3.1 Measuring of IOP

The idea behind the concept of equality of opportunity is to level playing field for individuals. Equality of Opportunity implies that circumstances beyond the control of an individual should not influence his/her outcome. The question is how much of that inequality can be spelled out by circumstances such as family background, gender or ethnicity and how much are due to efforts or luck.

There are two main approaches for measuring IOP for an outcome such as education achievement. Both of them start by searching for agreement on the set of individual characteristics which are beyond the individual's control. In the literature, these variables are called as circumstances. C is the vector of circumstances in the model. Based on C the society can be partitioned into different groups with identical circumstances. Each of these groups is called as a type. If we show the whole society as S and the set of type as Γ where $\Gamma = \{T_1, T_2, ...T_k\}$ such that $T_1 \cup T_2 \cup ... \cup T_k = S$ and $T_i \cap T_j = \emptyset, \forall i, j$ and $C_i = C_j, \forall i, j | i \in T_k \& j \in T_k, \forall k$. Given this partition, we can define two approaches for measuring IOP. The one is called the ex-ante approach and the second is the ex-post approach. In the ex-ante approach, the outcome is evaluated for each type and equality of opportunity is obtained when there is perfect equality on evaluated values for all types. In practice, the researchers have used the mean outcome (such as mean income or mean education achievement) of the type as an estimate of the value of the opportunity set. The differences between types show the inequality of opportunity. The ex-ante approach measures between-type inequality. On the other hand, one can measure IOP through ex-post approach proposed by (**?**). In this approach, EOP achieves only when persons with the same efforts obtain the same outcome regardless of their circumstances. IOP can be captured by the sum of inequality within groups characterized by the same degree of effort. (Ferreira and Gignoux 2013)

In this study, First, I measure IOP with the ex-ante approach context by using decomposition method and then I check my results by using R^2 as (Ferreira and Gignoux 2013) proposed.

Ex-ante Approach: Decomposition Method

As (Checchi and Peragine 2010) defined, "there is EOP if all the types have the same mean income. Inequality of opportunity decreases if inequality between the types income decreases." One can formulate this definition as:

$$f\,(y|C) = f(y), \tag{3.1}$$

where y is the outcome of interest (here education attainment) and C are relevant circumstances ((Salehi-Isfahani, Hassine, and Assaad 2012)). In the ex-ante approach, one should decompose the effect of efforts and luck on outcome from effect of circumstances. (Checchi and Peragine 2010) proposed to apply a smoothing transformation to remove the residual inequality. The smoothed distribution $\left\{\mu_t^k\right\}$ is derived from y distribution where each individual outcome in type k is replaced by the group-specific mean, $\mu^k(y)$ which means that outcome of individuals in each type is replaced by the mean of outcomes in that type. More formally for N individuals and K types we have ((Ferreira and Gignoux 2011)):

$$\left\{\mu_i^k\right\} = (\mu_1^1, ...\mu_{n_1}^1; ....; \mu_i^K, ..., \mu_N^K) \text{ with } \mu_g^k = ...\mu_i^k = ...\mu_h^k \,\forall k, g = 1 + \sum_{l=1}^{k-1} n_l,\;\; h = \sum_{l=1}^{k} n_l \tag{3.2}$$

Then, the measure of inequality of opportunity will map joint distribution of outcomes and circumstances in each type to a positive real number. As a result the relative version of this index can be defined over 0 and 1 interval.

$$\theta_r = \frac{I(\left\{\mu_i^k\right\})}{I(y)}, \tag{3.3}$$

where I() is any inequality index that satisfies symmetry, the transfer principle, scale invariance, population replication, and additive decomposability. (Foster 1985) showed that a positive multiple of a member of the "Generalized Entropy" class satisfies four basic properties and additive decomposability. In this paper, I use the parametric approach which is used by (Salehi-Isfahani, Hassine, and Assaad 2012) to measure ex-ante IOP. The brief formulation of their paper is as followed:

$$y_i = C_i\beta + \epsilon_i \tag{3.4}$$

Where C is vector of circumstances in addition to constant and ϵ_i shows all unobserved factors.

$$\tilde{z}_i = C_i\hat{\beta} \tag{3.5}$$

Running a simple OLS model, the fitted value of y_i represents all effects of circumstances on outcome. The relative inequality estimated by fitted value over total inequality will be an indicator of IOP.

$$\theta_d = \frac{I(\tilde{z}_i)}{I(y_i)} \tag{3.6}$$

Based on ex-ante approach, we need to measure inequality due to inequality in efforts. The difference between total inequality and inequality due to efforts obtains IOP. By dividing this difference to total inequality we can derive an index to measure IOP. One can say y_i is sum of fitted value and residuals, more formally, $y_i = \tilde{z}_i + \hat{\epsilon}_i$. We can say $\hat{\epsilon}_i$ measures the contribution of effort and luck to outcomes. With helping a synthetic distribution where circumstances are fixed to be equal in their mean values, we can compute an index of IOP.

$$\tilde{y}_i = \bar{C}\hat{\beta} + \hat{\epsilon}_i \tag{3.7}$$

So

$$\theta_r = 1 - \frac{I(\tilde{y}_i)}{I(y_i)} \tag{3.8}$$

In this research I use GE(2) as an index of inequality and report the results for θ_r only since the results of indirect method are more reliable compared to direct method.

Ex-ante Approach: Using R^2

(Ferreira and Gignoux 2013) proposed a new method for measuring IOP which is adapted with the ex-ante approach. By using equation 3.4 we can derive $\tilde{z}$ which denotes the vector of predicted education achievement from regression. Under the assumption of the linear relationship between circumstances and outcomes, $\tilde{z}$ is equivalent to the "smoothed distribution." The authors argued in their paper that simple variance can be used as inequality index so by adapting variance into equation 3.6 we will have:

$$\hat{\theta}_{IOP} = \frac{Var(C_i'\hat{\beta})}{Var(y_i)} \tag{3.9}$$

This index has some interesting features. In fact, it is R^2 of an OLS regression of student's test score on the vector of circumstances. Running equation 3.4 and computing its R^2 will give us a measurement of IOP in education. In addition to its simplicity, this index is a parametric estimation to the lower bound on the share of overall inequality in education achievement. A third feature of equation 3.9 is that we can use more information on circumstances compared to previous studies, which investigates the effect of a smaller set of circumstances. So our estimation of IOP will be closer to actual value of IOP in the society.

3.4 Data

For measuring inequality of opportunity, I utilize data from the fifth rounds of Trends In International Mathematics and Science Study (TIMSS) database at 2011. TIMSS measures trends in mathematics and science achievement both for the fourth and eighth grade in participating countries in the world. In this database, there is huge and rich socioeconomic information of students around the world. Background information provides materials for doing comparative studies within, and between, countries in the context of different educational systems, school organizational approaches, and instructional practices (Foy 2013). Among 63 countries and 14 benchmarking participants, in this study, I choose those countries which participated both in fourth and eighth grade and have information about parent educations. The fifth round (i.e. 2011) is the first round that asks questions about parent education from students in the fourth grade. Having information on parent education both in the fourth and eighth grade will facilitate using the difference-in-differences method for addressing the effect of tracking. In addition, having information for parental education will help me for comparing IOP between students in primary and secondary schools.

Twenty countries have participated both in fourth and eighth grade and have information on parental education for fourth grade. The countries are reported at table 3.1. TIMSS tests are not uniform for all students across schools and countries. Item Response Theory (IRT) will standardize and scale ex-post scores to make them comparable. Scores generated by IRT are mapped onto an international achievement scale with the mean of 500 and the standard deviation of 100 in 1995. As a result, the scores are comparable across countries and over time. TIMSS reports test scores as five "plausible values" both for mathematics and science.

For measuring IOP, I use from average values of these scores. TIMSS employs a random selection of schools in the first stage and then chooses randomly one or two classes in each school in the second stage. For my purpose, I construct two databases, one for fourth grade and the other for eighth grade, combining students test scores in mathematics and science, family background, student-specific characteristics and resources of schools. Most variables in the sample are categorical and qualitative variables which were transformed into dummy variables. For schooling resources, TIMSS data have detailed information about teacher characteristics both for math and science. Due to endogeneity problem, in this study, I use these data at community level (Salehi-Isfahani, Hassine, and Assaad 2012).

In this study, I choose parent education, the number of books at home, access to the internet, computer, gender, ethnicity, the place where a student lives, and teacher characteristics as circumstances.

Table 3.1 & 3.2 shows summary statistics of chosen countries for fourth grade and eighth grade students respectively.

Figures 3.1,3.2, 3.3, and 3.4 show the conditional density of average test scores in mathematics and science for different attributes including gender, mother education, school type, and community size. The graphs for other countries is not reported here for saving space but it is available under the request. As these graphs show, gender is not an issue anymore even in Iran and Qatar where they are traditional countries. In contrast to gender, mother education has a significant effect on the distribution of test scores both in mathematics and science and especially in Iran. These kernel densities illuminate the effect of different circumstances on the distribution of average test scores which lead us to choose appropriate circumstances for measurement of inequality of opportunity.

Sample size ranges from 3000 to more than 14000 for students in fourth grade and it ranges from 3800 to 14000 in eighth grade. In fourth grade in 2011, Morocco, Oman, and UAE are among countries with a large sample while Norway, Hong Kong and Qatar have the smallest samples. In eighth grade, Oman and UAE participate with more than 9000 students while Norway and Italy have samples with less than 4000 students.

Table 3.1: Summary Statistics-Fourth Grade

Fourth Grade Students-2011									
		Mathematics				Science			
Country	Obs	Mean	SD	Min	Max	Mean	SD	Min	Max
Australia	6146	513.2	82.7	192.9	745.8	514.1	76.7	209.5	733.4
Finland	4638	545.6	63.6	272.4	736.3	569.2	60.8	289.9	772.9
Georgia	4799	451.7	81.4	159.5	730.5	456.6	78.6	189.4	711.9
Hong Kong	3957	602.3	61.2	286.8	767.6	535.9	67.5	170.6	715.5
Hungary	5204	521.2	83.6	196.1	742.8	539.7	79.6	184.3	748.5
Iran	5760	435.3	87.7	124.7	720.1	458.2	93.9	132.3	739.9
Italy	4200	506.1	69.1	256.5	703.8	521.9	70.3	246.1	740.1
Lithuania	4688	539.4	68.9	259	739.2	520.3	61.8	268.3	703.1
Morocco	7841	356.6	90.9	113.8	680.8	267.8	113.6	24.9	677
Norway	3121	496.7	64.1	259.6	698.8	495.9	58.6	264.2	691.4
Oman	10411	383.4	97.9	88.1	717.1	374.1	116.9	42.9	748.8
Qatar	4117	407.9	99.9	150.2	712.8	390.8	119.2	66.9	729.2
Romania	4673	496.2	97.6	151.2	779.3	521.2	97.8	156.3	774.7
Russia	4467	545.4	69.3	314.8	782.4	555.6	66.6	316.5	754.2
S. Arabia	4515	406.2	93.8	125.5	815.1	425.1	99.6	98.1	819.1
Singapore	6368	603.8	74.9	313.2	787.2	580.9	83.2	264.5	797.7
Slovenia	4492	510.4	65.1	223.8	697.7	517.4	71.1	179.4	728.1
Sweden	4663	500.3	63.9	281.8	697.8	527.8	72.1	274.9	743.4
Taiwan	4284	593.9	68.1	331.6	768.2	554.5	67.9	300.9	750
UAE	14720	435.8	95.3	154.3	779.5	428.4	107.2	107.2	827.8
Dubai	6151	455.5	100.2	178.6	766	446.9	111.4	148.8	761.9

Table 3.2: Summary Statistics-Eight Grade

Eighth Grade Students-2011									
		Mathematics				Science			
Country	Obs	Mean	SD	Min	Max	Mean	SD	Min	Max
Australia	7556	496.9	83.4	222.1	759.9	513.3	82.1	219	776.6
Finland	4266	514.2	62.8	276.5	705.5	552.1	62.1	313.3	758.7
Georgia	4563	437.5	100.8	142.5	783.7	423.7	83	141.8	681.2
Hong Kong	4015	587.4	81.8	243.7	789.8	536.4	71.3	195.6	736.5
Hungary	5178	513.5	82.7	175.8	754.7	529.9	74.8	161.5	740.1
Iran	6029	419.1	90.5	138.9	767.2	478.2	85.8	215.2	767.9
Italy	3979	498.7	69.5	266.5	721.3	501.7	71.5	204.2	748.7
Lithuania	4747	508.9	74.7	213.8	754.7	519.3	71	158.9	742.5
Morocco	8986	376.9	81.8	127.9	666.5	381.3	79.9	107.6	681.3
Norway	3862	476.5	61.8	203.2	647.1	496.1	69.1	211.7	686.8
Oman	9542	370.4	105.1	78.6	712.5	420.5	107.7	88.4	728.3
Qatar	4422	416.8	104.1	81.1	736.1	427.5	112.9	47.4	759.9
Romania	5523	469.3	100.5	176.1	779.1	472.1	81.9	158.7	728.5
Russia	4893	542.6	77.8	308.7	804	544.9	72.2	266.3	803.5
S. Arabia	4344	393.3	86.7	123.1	658.9	436.1	76.5	197.5	673.7
Singapore	5927	607.5	81.6	327.7	799.6	585.9	93.7	273.8	818.4
Slovenia	4415	504.7	67.8	282.2	722.5	542.2	72.1	279.8	757.1
Sweden	5573	483.4	64.7	241.2	702.4	507.7	77.2	186.2	756
Taiwan	5042	613	102.6	206.5	872.9	566.4	80.1	195.2	779.5
UAE	14089	453.1	86.1	178.3	767.1	460.5	94.2	156.7	819.9
Dubai	5571	464.2	91.8	181.3	767.1	469.8	101.2	161.6	819.9

As I said before, TIMSS provides information about the students' family, community characteristics, and school quality. Family background variables are parents' education, the number of books at home, access to the computer and the internet. All of these variables are recorded at categorical variables. For parents' education, the category consists of eight levels which are based on ISCED level defined by the UNESCO. For instance in Oman for students in fourth grade, about 34.5% of Omani students have mothers with primary or no schooling education whereas only 13.12% of mothers have a university degree. These number is changed for eighth-grade students where 23.75% mothers have education lower than primary school and about 10.67% of them have a university degree. On the other side of the distribution, Norwegian students' mothers are more educated compare to their counterpart in Oman. Only 0.97% and 0.78% of fourth and eighth-grade students have mothers with education level less than primary school while 52.95% and 32.34% of fourth and eighth-grade students' mothers have a university degree. It is interesting that the average scores of mathematics for fourth-grade students in Oman and Norway are 383.36 and 496.68 respectively. One can say that some parts of the dispersion between school achievement of Omani & Norwegian students can be explained by mother education level.

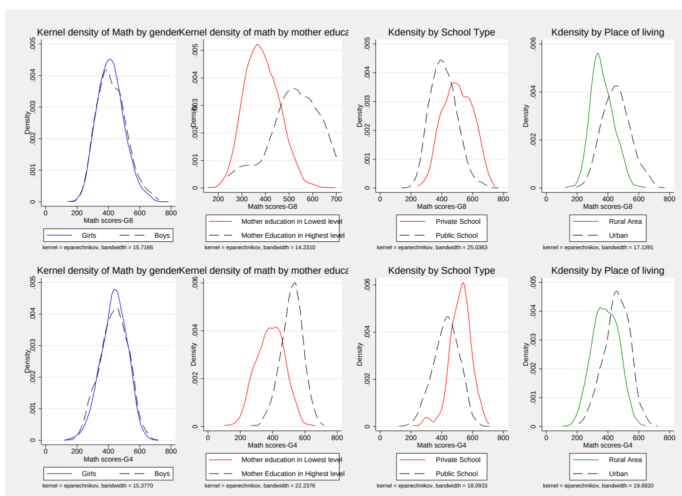


Figure 3.1: Conditional Densities in Mathematics in Iran by key circumstances

TIMSS data include a variable indicating if the test was taken in the language spoken at home. This variable is a categorical variable with four categories, namely: always, almost always, sometimes, never. I use this variable as an indicator of minority status or ethnicity. I also measure IOP among children who are the resident of the country and not immigrants.

A number of books at home are another variable that can be used as a proxy of the home

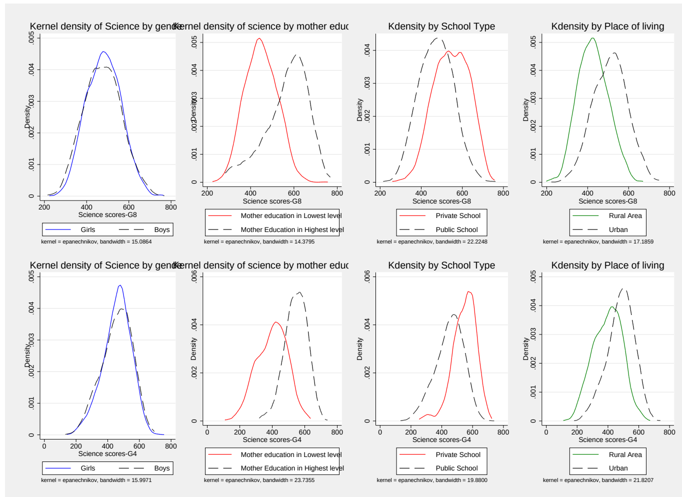


Figure 3.2: Conditional Densities in Science in Iran by key circumstances

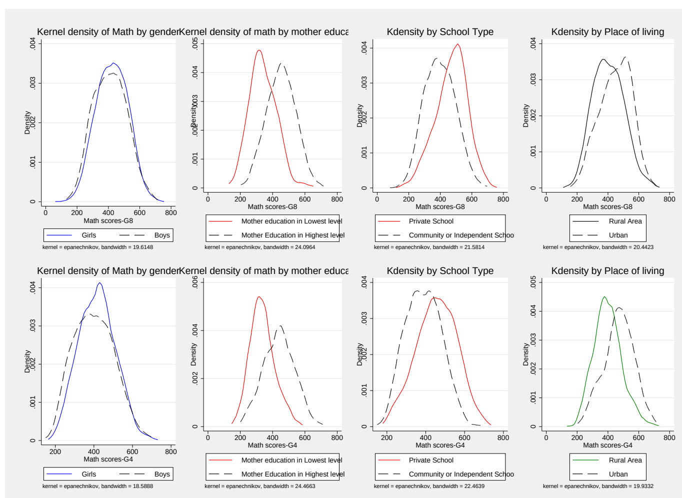


Figure 3.3: Conditional Densities in Mathematics in Qatar by key circumstances

environment for education. This variable classifies a number of books in five categories: 0-10 books, 11-25 books, 26-100 books, 101-200 books, and more than 200 books. For students in grade eight, Morocco, Oman, Saudi Arabia, and Iran rank poorly according to this measure, while Australia, Norway, Georgia, and Hungary are well and more than 40% of students live in houses with more than 100 books in them.

I also use variables for community and school characteristics. These variables include community size, class size, teacher education, teacher age, teacher experience, as well as the percentage of students in school come from disadvantaged families or affluent ones. These

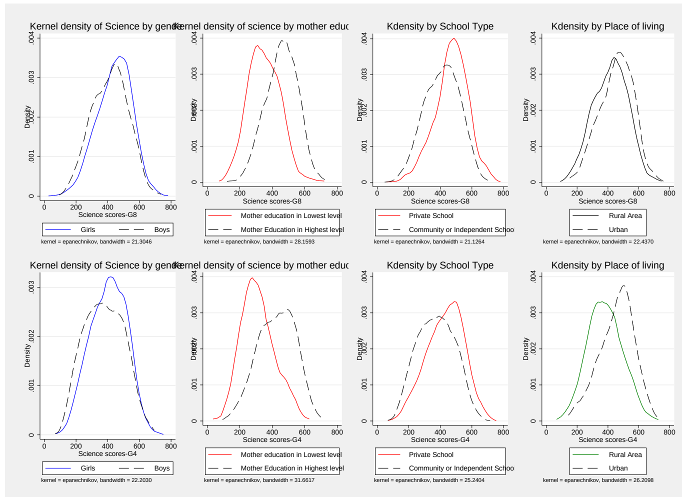


Figure 3.4: Conditional Densities in Science in Qatar by key circumstances

variables have been shown to affect student performance ((Wößmann 2003)). As I said before, due to endogeneity problem I use average of school-level variables computed at the sampling zone level.

TIMSS data includes students who have entered and remained in school until grade 8. It may be plausible that students who come from lower socio-economic background leave school earlier. Based on the UNESCO educational database, grade eight is part of the compulsory education in all of the countries in the sample and based on the table 3.3, the enrollment rate is quite high in all of the countries and both gender in the sample. So selection is not a serious issue for the comparison across countries. In addition, like most samples, there are missing observations. Missing values for family background variables are about 13.5% and 2.7% on average for fourth and eighth grade respectively. I follow the procedure that (Salehi-Isfahani, Hassine, and Assaad 2012) where they mark the missing values of independent variables with dummy variables.

Table 3.3: Net Enrollment Rates, Primary and Secondary Schools by Gender and School Type-2011

Country Name	Primary, female	Primary, male	Primary,private (% of total primary)	Secondary, female	Secondary, male	Secondary,private (% of total secondary)
Australia	96.78	96.03	31.08	86.09	85.16	35.24
Finland	98.18	98.16	1.60	93.25	92.69	9.36
Georgia	98.99	97.79	9.51	80.34	84.43	6.41
Hong Kong	89.64	90.89	16.69	77.87	78.32	16.73
Hungary	91.96	93.08	9.20	91.70	92.33	13.36
Iran,	96.24	98.11	7.33	77.40	81.28	11.88
Italy	96.66	97.76	6.93	92.43	91.38	8.52
Lithuania	95.20	97.15	1.06	98.11	98.18	1.31
Morocco	96.04	96.79	11.77	-	-	4.75
Norway	99.35	99.20	2.36	95.86	95.24	8.17
Oman	96.49	95.53	17.23	86.33	89.72	6.98
Qatar	89.90	94.27	56.67	100.00	90.56	39.62
Romania	87.38	88.39	0.34	80.70	79.35	1.82
Russia	93.42	93.43	0.64	-	-	0.71
S.Arabia	93.19	90.28	9.70	–	–	12.61
Singapore	–	–	–	–	–	6.37
Slovenia	97.48	96.85	0.44	93.62	92.86	1.49
Sweden	99.23	99.70	10.15	91.55	91.44	19.39
UAE	85.60	92.03	71.72	79.14	72.50	57.62
Bahrain	95.88	97.05	32.46	91.96	94.51	21.81
Lebanon	87.71	94.90	73.66	69.72	69.52	61.26
Palestinian	86.36	86.97	12.23	84.79	76.99	5.69
Tunisia	95.57	96.84	2.43	–	–	4.77
Turkey	86.36	86.97	0.75	81.95	87.34	–
Jordan	96.85	98.23	33.62	89.37	86.49	18.68

Note: For some countries the nearest available year is chosen.
Source: UNESCO education database and WDI.

3.5 Results

3.5.1 Decomposition and R^2 results

Here, I present estimates of IOP using both parametric decomposition and R^2 of OLS regression for the fourth grade of 20 countries and eighth-grade students of 26 countries at 2011. As I mentioned above, I use gender, ethnic background (indicated by home language), parental education, the number of books at home, access to computer and the internet, place of living as well as characteristics of community such as teacher experience, teacher education, economic status of classmates and average class size. Tables 3.4 and 3.5 offers a summary of results for fourth and eighth grade.

Table 3.4: Parametric estimates of inequality of educational opportunity for fourth grade-2011

Country	Mathematics	R^2	Science	R^2
Australia	0.388	0.27	0.424	0.29
Finland	0.192	0.20	0.263	0.24
Georgia	0.217	0.22	0.103	0.22
Hong Kong	0.360	0.21	0.363	0.19
Hungary	NA	0.41	0.200	0.43
Iran	0.349	0.32	0.360	0.35
Italy	0.212	0.14	0.229	0.18
Lithuania	0.276	0.26	0.085	0.29
Morocco	0.272	0.26	0.276	0.23
Norway	0.225	0.19	0.252	0.24
Oman	0.247	0.19	0.131	0.19
Qatar	0.463	0.40	0.455	0.35
Romania	0.369	0.28	0.434	0.37
Russia	0.214	0.20	0.199	0.19
S. Arabia	0.185	0.14	0.175	0.21
Singapore	0.350	0.28	0.397	0.34
Slovenia	0.248	0.23	0.241	0.25
Sweden	0.320	0.21	0.389	0.24
Taiwan	0.274	0.23	0.277	0.25
UAE	0.313	0.25	0.304	0.25
Dubai	0.424	0.37	0.356	0.37

Figure 3.5 & 3.6 show the trend of IOP in math and science between fourth and eighth

grades. As it is obvious, IOP both in mathematics and science increases for most countries from grade 4 to 8. It makes sense because as time passes, parents have obtained more information about the ability level of their progeny. In addition, in most of the countries, tracking systems will start either at grade 9 or 10. My hypothesis is as time passes, parents invest more in their children to increase their chances to stay at the academic level. Parents who have access to better opportunities will invest more and as a result, the role of family background on students' outcome increases which means the inequality of opportunity rises

Table 3.5: Parametric estimates of inequality of educational opportunity for eighth grade-2011

Country	Mathematics	R^2	Science	R^2
Australia	0.327	0.360	0.387	0.370
Finland	0.190	0.160	0.246	0.210
Georgia	0.320	0.290	0.327	0.270
Hong Kong	0.465	0.460	0.407	0.390
Hungary	0.4042	0.430	0.4105	0.440
Iran	0.384	0.370	0.359	0.350
Italy	0.248	0.250	0.296	0.300
Lithuania	0.307	0.310	0.294	0.300
Morocco	0.283	0.230	0.242	0.190
Norway	0.280	0.240	0.293	0.240
Oman	0.346	0.310	0.404	0.350
Qatar	0.336	0.450	0.381	0.430
Romania	0.415	0.360	0.366	0.330
Russia	0.211	0.190	0.237	0.220
S. Arabia	0.233	0.210	0.272	0.230
Singapore	0.264	0.260	0.331	0.330
Slovenia	0.240	0.250	0.256	0.260
Sweden	0.225	0.200	0.291	0.250
Taiwan	0.339	0.320	0.360	0.340
UAE	0.241	0.210	0.261	0.220
Dubai	0.332	0.300	0.352	0.330
Bahrein	0.402	0.430	0.413	0.415
Lebanon	0.348	0.330	0.347	0.350
Palestinian	0.243	0.220	0.228	0.240
Tunisia	0.228	0.275	0.230	0.220
Turkey	0.301	0.310	0.318	0.310
Jordan	0.284	0.230	0.360	0.260

between fourth and eighth grade.

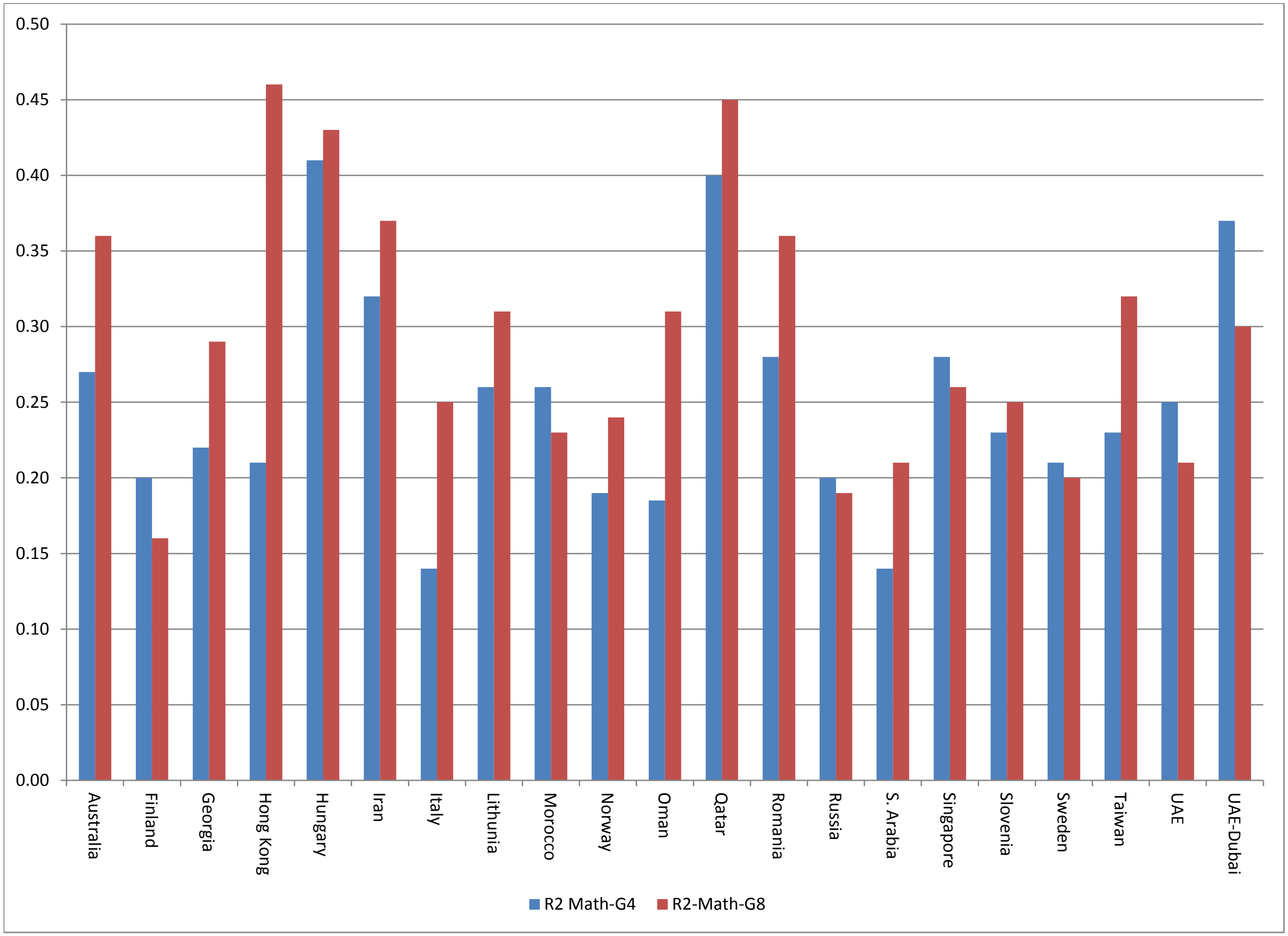


Figure 3.5: IOP in Math between fourth and eighth grades

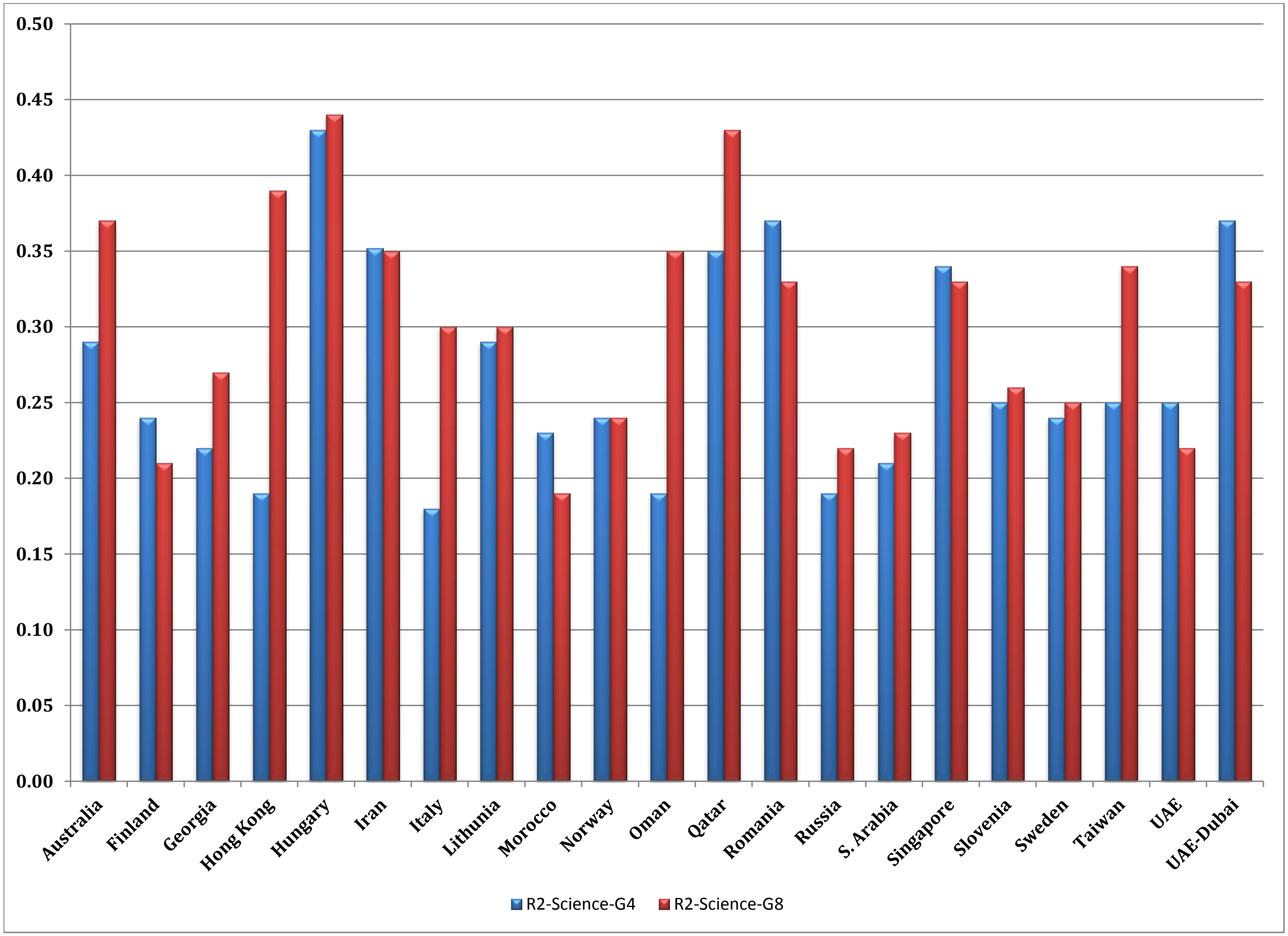


Figure 3.6: IOP in Science between fourth and eighth grades

3.5.2 Inequality of Opportunity and Country Characteristics

The wide ranges of IOP between countries and within grades are interesting to ask what explains the level of IOP in these countries and why IOP changes from grade 4 to 8. The answers to these questions will be helpful from policy designing point of view. Here I investigate several variables which may have some relationship with IOP index. Among them, I focused on per pupil expenditures, the Gini index, GDP and GDP per capita, and economic growth.

First, I want to explore the correlation of expenditures in education and IOP. Expenditures per pupil as a percentage of GDP for primary and secondary education is the highest in Slovenia and Finland respectively. In Slovenia, the expenditures per pupil in primary education is 31.13% of GDP per capita while UAE spends only 5.91% of GDP per capita in primary education for each student. On the other hand, for secondary education, Finland spends about 36.53% of GDP per capita for each pupil while Lebanon's expenditure is only 3.73% of GDP per capita for each student. The median of this characteristic is 18.17% and 19.88% of GDP per capita for primary and secondary education respectively. Table 10 in the appendix C shows the details of educational expenditures in the countries of sample. In addition to expenditures per pupil, we have public spending on education as a percentage of GDP and government expenditures. The correlation of IOP in mathematics at primary school with those variables are -0.336, -0.378, and 0.056 respectively.

Figures 3.7 and 3.8 show this fact for primary and secondary expenditures for mathematics in fourth and eighth grade. For saving space I report the trend for science in the appendix C. Figures 17 and 18 show this trend for science.

As it is obvious from the figures, there is a negative correlation between IOP and per pupil expenditures both in primary and secondary schools. The relation is more negative at secondary schools too. The correlation of per pupil expenditures in secondary level and IOP in math and science in eighth grade is -0.48 and -0.50 respectively.

In fact, the philosophy of public school and public expenditures on education is equalizing opportunities among students. As (Salehi-Isfahani, Hassine, and Assaad 2012) noted, "The level of education expenditures relative to GDP per capita is a better indicator of the relative strength of public vs. private spending on education, and therefore a better candidate as a determinant of equality of opportunity." Next, I look at the correlation between inequality of income measured by Gini index and IOP in education achievement. I expect a greater inequality in income coincides with higher IOP in education, however, this conjecture is weakly supported by the data I have. Table 11 and figure 19 present data and graphical correlations for science in appendix C. Figure 3.9 shows the relation for mathematics. The correlation between IOP in education and Gini index is about 0.27 and 0.16 for mathematics and science in grade eight. There is also a positive correlation between differences in IOP in fourth and eighth grade and Gini index. One can say that inequality of income may explain part of the gap between students in fourth and eighth grade. My conjecture is the deeper and higher inequality in a society tends to have a larger gap between students in different ages. The correlation between Gini index and IOP is strong in mathematics regardless of the education grade, however, it is weak in science for fourth graders. Overall, the relationship between IOP and Gini is stronger in mathematics than science.

Finally, I explore the correlation average economic growth and GDP per capita, as two proxies for level of development, and IOP. There is a literature which emphasizes on development and equity in one hand and economic growth and equality on the other hand. The World bank development report in 2006 is a good example of studies which investigate the relationship between development and equality. (Marrero and Rodríguez 2013) show that there is a negative relationship between inequality of opportunity and growth while there is a positive relation between inequality of effort and growth in the USA at 1970,80, and 90. In addition, Kuznet curve said that economic growth accompanies with inequality. Inequality of opportunity is one aspect of inequality. Since most countries in my sample are either developing countries or the Middle Eastern countries which are in their early stage of development, my conjecture is that greater growth in GDP accompanies with greater IOP. For investigating this hypothesis, I used the average growth rate of GDP for each country from 1991 to 2011 (21 years). "Qatar" and "Palestinian" have data for 11 years. The correlation between economic growth and IOP in mathematics is 0.16 both for grade four and eight. It is almost no relationship between economic growth and IOP in science.

Another variable of interest is looking at GDP per capita. I use GDP per capita (PPP, 2011 constant international \$). Excluding "Qatar" as an outlier remains this relation positive but the slope is smaller for fourth graders in mathematics and science. This relationship is robust for eight graders in science as well while there is almost no relationship in mathematics for eight graders. Figure 3.12 presents the relationship between IOP and GDP per capita in mathematics. Figure 22 presents the same relationship in science in the appendix C.

3.6 Conclusion

Inquality of opportunity is a new concept after John Rawls work that justifies the meaning of justice in a practical way. Measuring IOP for an outcome is the first step for doing research in this field. Using TIMSS data for 2011, I derive the lower bound of IOP in education for a sample contains twenty countries both for students in primary and secondary schools. The results both using variance decomposition method and R^2 show the expansion of IOP in most countries of my dataset. The index of IOP is lower on average for OECD countries while it is higher for Middle Eastern countries. Given the results of the Human Opportunity Index and D-index in chapter 1, we can say that MENA governments are successful in providing of basic services such as piped water or access to education, however, they are not successful in provision of higher level services such as education achievement. In other words, they provide successfully education but they could not provide a quality of education leveled for all.

The preliminary investigations show that there is a negative correlation between per pupil expenditure and level of IOP. The relationship between Gini index and IOP is positive suggesting there is more likely to have higher level of inequality of opportunity when there is initially high income inequality existed in society.

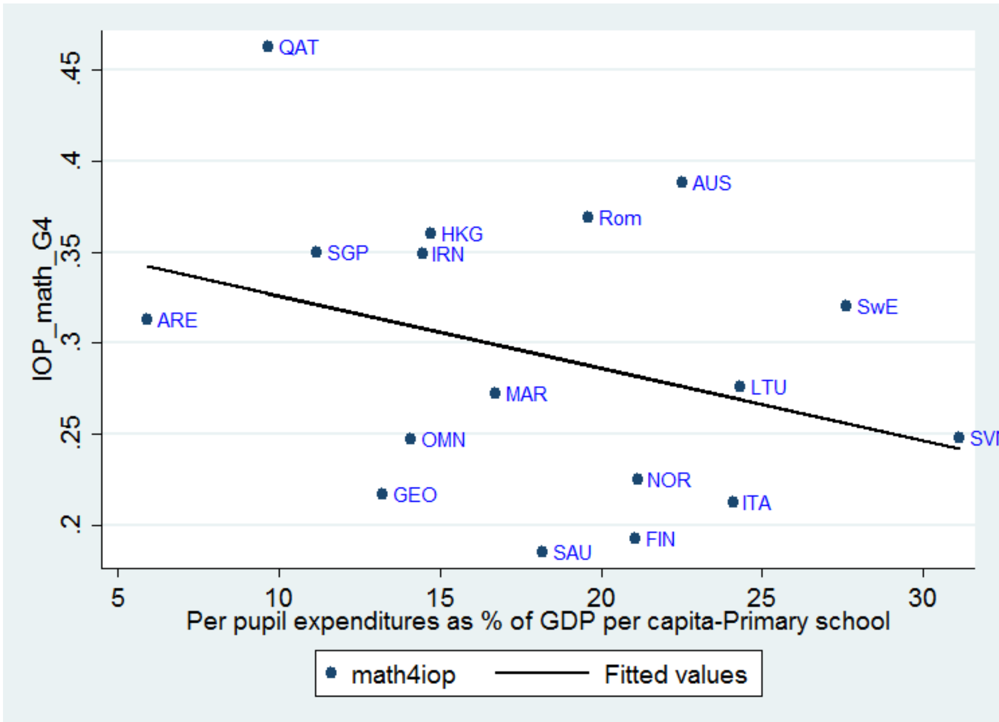


Figure 3.7: IOP in mathematics in grade 4 and per pupil expenditures in primary education

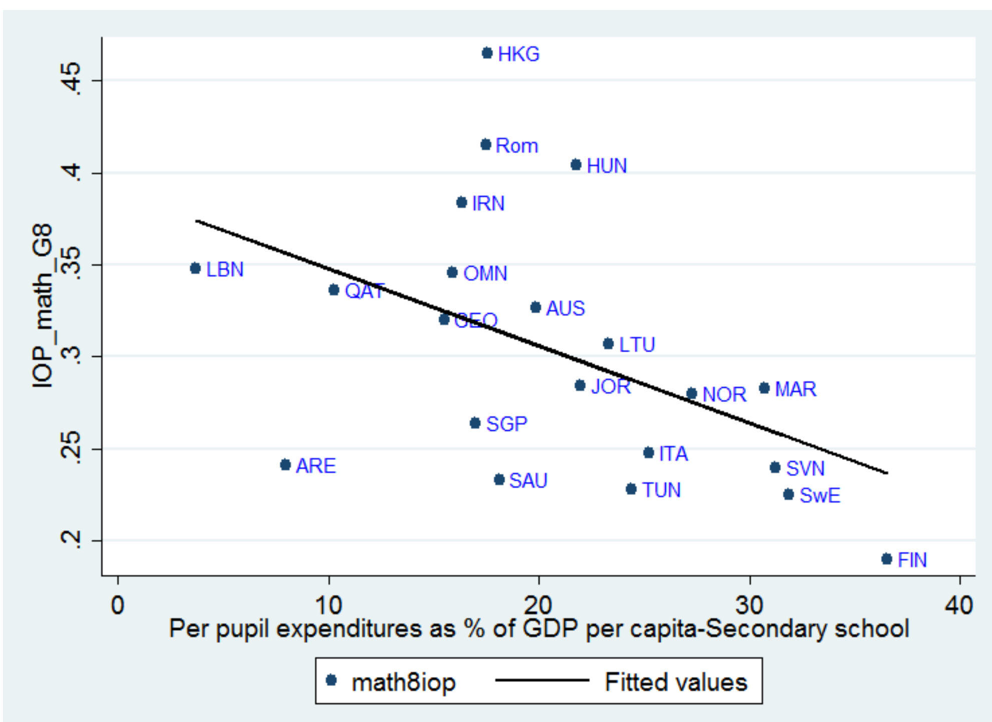


Figure 3.8: IOP in Math in grade 8 and per pupil expenditures in secondary education

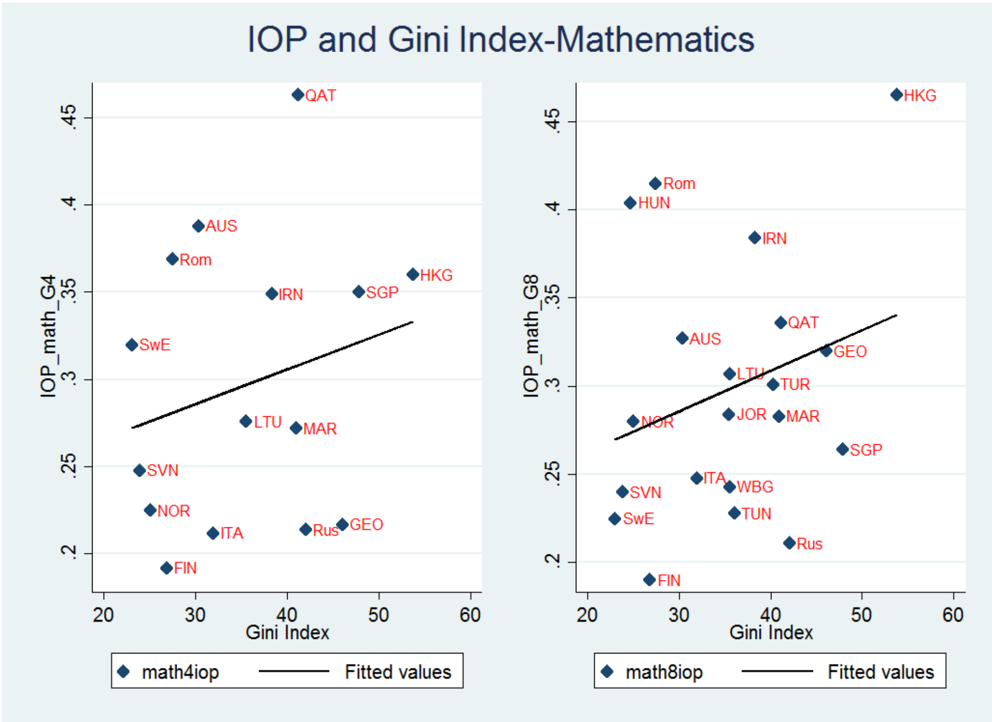


Figure 3.9: IOP and Gini Index-Mathematics

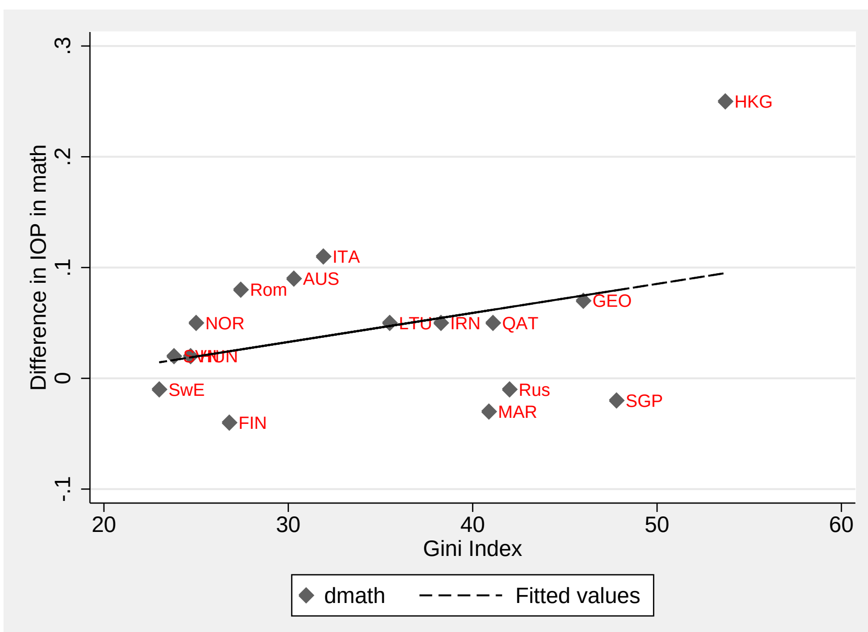


Figure 3.10: Difference in IOP in math between grade 4 and 8 and Gini Index

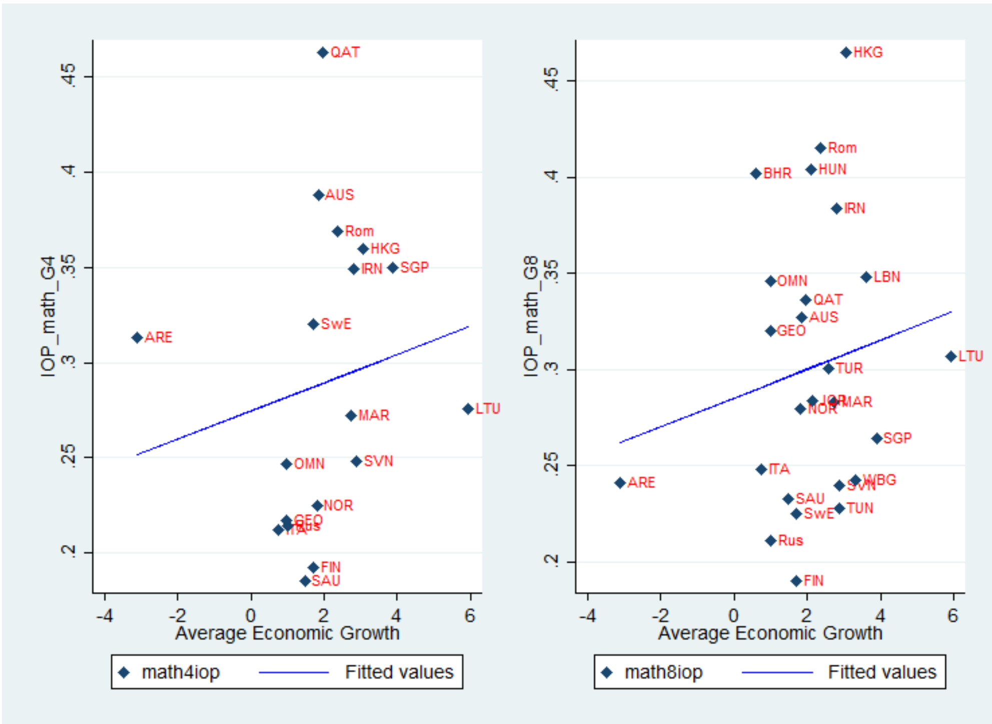


Figure 3.11: IOP and Average Economic Growth-Mathematics

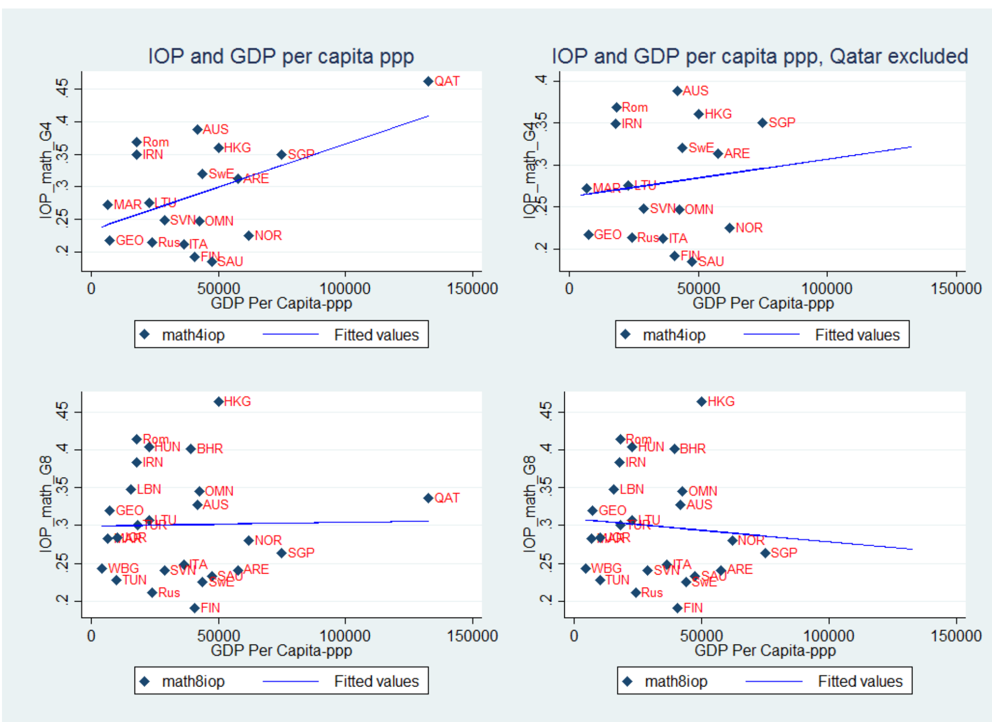


Figure 3.12: IOP and GDP per capita-Mathematics

Bibliography

Alderman, H., M. Babita, G. Demombynes, N. Makhatha, and B. Özler (2002). How low can you go? combining census and survey data for mapping poverty in south africa. *Journal of African Economies 11*(2), 169–200.

Angrist, J. D. and A. B. Keueger (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics 106*(4), 979–1014.

Aran, M. A. and L. Ersado (2013). Inequality of opportunity in access to basic services among egyptian children.

Arneson, R. J. (1989). Equality and equal opportunity for welfare. *Philosophical studies 56*(1), 77–93.

Assaad, R., C. Krafft, J. Roemer, and D. Salehi-Isfahani (2016). Inequality of opportunity in income and consumption: the middle east and north africa region in comparative perspective. In *Economic Research Forum Working Paper Series*, Number 1003.

Assaad, R. and M. Saleh (2016). Does improved local supply of schooling enhance intergenerational mobility in education? evidence from jordan. *The World Bank Economic Review*, lhw041.

Assaad, R., D. Salehi-Isfahani, and R. Hendy (2013). Inequality of opportunity in educational attainment in the Middle East and North Africa: evidence form survey data. ERF working paper.

Assaad, R., D. Salehi-Isfahani, and R. Hendy (2014). Inequality of opportunity in educational attainment in middle east and north africa: evidence from household surveys.

Becker, G. S. (1975). Front matter, human capital: a theoretical and empirical analysis, with special reference to education. In *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition*, pp. 22–0. NBER.

Bedi, T., A. Coudouel, and K. Simler (2007). *More than a pretty picture: using poverty maps to design better policies and interventions*. World Bank Publications.

Belhaj-Hassine, N. (2012). Inequality of opportunity in Egypt. *World Bank Economic Review* (2), 265–295.

Belhaj-Hassine, N. (2015). Economic inequality in the arab region. *World Development 66*, 532–556.

Benabou, R. (1996). Equity and efficiency in human capital investment: the local connection. *The Review of Economic Studies 63*(2), 237–264.

Betts, J. (2011). The economics of tracking in education. *Handbook of the Economics of Education 3*(341-381), 4.

Bibi, S. and M. K. Nabli (2009). Income inequality in the arab region: Data and measurement, patterns and trends. *Middle East Development Journal 1*(02), 275–314.

Bourguignon, F., F. H. Ferreira, and M. Menéndez (2007). Inequality of opportunity in brazil. *Review of Income and Wealth 53*(4), 585–618.

Brunello, G. and D. Checchi (2007). Does school tracking affect equality of opportunity? new international evidence. *Economic Policy 22*(52), 781–861.

Butcher, K. F. and A. Case (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics 109*(3), 531–563.

Cammett, M., I. Diwan, A. Richards, and J. Waterbury (2015). *A political economy of the Middle East*. Westview Press.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research.

Card, D. (1994). Earnings, schooling, and ability revisited. Technical report, National Bureau of Economic Research.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics 3*, 1801–1863.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*(5), 1127–1160.

Checchi, D. and V. Peragine (2010). Inequality of opportunity in italy. *The Journal of Economic Inequality 8*(4), 429–450.

Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics*, 906–944.

Dabalen, A., A. Narayan, J. Saavedra-Chanduvi, and A. H. Suarez (2014). *Do African Children Have an Equal Chance?: A Human Opportunity Report for Sub-Saharan Africa.* World Bank Publications.

de Barros, P. and S. Molinas Vega (2008). Measuring inequality of opportunities for children.

de Barros, R. P. (2009). *Measuring inequality of opportunities in Latin America and the Caribbean.* World Bank Publications.

Desai, R. M., A. Olofsgård, and T. M. Yousef (2009). The logic of authoritarian bargains. *Economics & Politics 21*(1), 93–125.

Duflo, E. (2000). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. Technical report, National Bureau of Economic Research.

Duflo, E. (2004). The medium run effects of educational expansion: Evidence from a large school construction program in indonesia. *Journal of Development Economics 74*(1), 163–197.

Duflo, E. et al. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review 91*(4), 795–813.

Dworkin, R. (1981a). What is equality? part 1: Equality of welfare. *Philosophy & Public Affairs*, 185–246.

Dworkin, R. (1981b). What is equality? part 2: Equality of resources. *Philosophy & Public Affairs*, 283–345.

El-Kogali, S. and C. Krafft (2015). Expanding opportunities for the next generation: Early childhood development in the middle east and north africa. Technical report, Washington, DC: World Bank.

Elbers, C., T. Fujii, P. Lanjouw, B. Özler, and W. Yin (2007). Poverty alleviation through geographic targeting: How much does disaggregation help? *Journal of Development Economics 83*(1), 198–213.

Elbers, C., J. O. Lanjouw, and P. Lanjouw (2003). Micro–level estimation of poverty and inequality. *Econometrica 71*(1), 355–364.

Ferreira, F. H. and J. Gignoux (2011). The measurement of inequality of opportunity: Theory and an application to latin america. *Review of Income and Wealth 57*(4), 622–657.

Ferreira, F. H. and J. Gignoux (2013). The measurement of educational inequality: Achievement and opportunity. *The World Bank Economic Review*, lht004.

Fleurbaey, M. and V. Peragine (2013). Ex ante versus ex post equality of opportunity. *Economica 80*(317), 118–130.

Foster, J. E. (1985). Inequality measurement. *Fair allocation 33*, 31–68.

Foy, P. (2013). Timss and pirls 2011 user guide for the fourth grade combined international database.

Greene, W. (2002). The bias of the fixed effects estimator in nonlinear models.

Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society*, 1–22.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational evaluation and policy analysis 19*(2), 141–164.

Hanushek, E. A. et al. (2006). Does educational tracking affect performance and inequality? differences-in-differences evidence across countries*. *The Economic Journal 116*(510), C63–C76.

Hentschel, J., J. O. Lanjouw, P. Lanjouw, and J. Poggi (2000). Combining census and survey data to trace the spatial dimensions of poverty: A case study of ecuador. *The World Bank Economic Review 14*(1), 147–165.

Hindriks, J., M. Verschelde, G. Rayp, and K. Schoors (2010). School tracking, social segregation and educational opportunity: evidence from belgium. *CORE DP n 2010 8*.

Ianchovichina, E., L. Mottaghi, and S. Devarajan (2015). Inequality, uprisings, and conflict in the arab world. *Washington, DC: World Bank*.

John Roemer, A. T. (2013, Octorber). Equality of opportunity.

Kanbur, R. and A. Wagstaff (2014). How useful is inequality of opportunity as a policy construct? Technical report, World Bank Policy Research Working Paper 6980.

Lanjouw, P. F., J. O. Lanjouw, and C. Elbers (2002). Micro-level estimation of welfare.

Lefranc, A., N. Pistolesi, and A. Trannoy (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in france. *Journal of Public Economics 93*(11), 1189–1207.

Maluccio, J. (1998). Endogeneity of schooling in the wage function: Evidence from the rural philippines.

Marrero, G. A. and J. G. Rodríguez (2013). Inequality of opportunity and growth. *Journal of Development Economics 104*, 107–122.

Mincer, J. (1974). Schooling, experience, and earnings. human behavior & social institutions no. 2.

Molinas, J. R., R. Paes de Barros, J. Saavedra, M. Giugale, L. Cord, C. Pessino, and A. Hassan (2010). Do our children have a chance? the 2010 human opportunity report for latin america and the caribbean. *World Bank, Washington, DC*.

Neilson, C. A. and S. D. Zimmerman (2014). The effect of school construction on test scores, school enrollment, and home prices. *Journal of Public Economics 120*, 18–31.

Nishimura, M., T. Yamano, and Y. Sasaoka (2008). Impacts of the universal primary education policy on educational attainment and private costs in rural uganda. *international Journal of Educational development 28*(2), 161–175.

Pischke, J.-S. and A. Manning (2006). comprehensive versus selective schooling in england in wales: What do we know? Technical report, National bureau of economic research.

Rawls, J. (1999). *A theory of justice*. Harvard university press.

Roemer, J. E. (1998). *Equality of opportunity*. Cambridge Univ Press.

Roemer, J. E. (2013). Economic development as opportunity equalization. *The World Bank Economic Review*, lht023.

Salehi-Isfahani, D., N. B. Hassine, and R. Assaad (2012). Equality of opportunity in educational achievement in the middle east and north africa. *The Journal of Economic Inequality*, 1–27.

Salehi-Isfahani, D., N. B. Hassine, and R. Assaad (2014). Equality of opportunity in educational achievement in the middle east and north africa. *Journal of Economic Inequality 12*(4), 489.

Salehi-Isfahani, D., I. Tunali, and R. Assaad (2009). A comparative study of returns to education of urban men in egypt, iran, and turkey. *Middle East Development Journal 1*(2), 145–187.

Schnepf, S. V. (2003). Inequalities in secondary school attendance in germany.

Schultz, T. P. (2004). School subsidies for the poor: evaluating the mexican progresa poverty program. *Journal of development Economics 74*(1), 199–250.

Schütz, G., H. W. Ursprung, and L. Wößmann (2008). Education policy and equality of opportunity. *Kyklos 61*(2), 279–308.

Sen, A. (1976). Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, 219–231.

Sen, A. (1980). Equality of what? *The Tanner lectures on human values 1*, 353–369.

Shapley, L. S. (1953). Stochastic games. *Proceedings of the national academy of sciences 39*(10), 1095–1100.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the shapley value. *Journal of Economic Inequality*, 1–28.

Son, H. H. (2013). Inequality of human opportunities in developing Asia. *Asian Development Review*.

Tarozzi, A. and A. Deaton (2009). Using census and survey data to estimate poverty and inequality for small areas. *The review of economics and statistics 91*(4), 773–792.

Trostel, P. A. (2005). Nonlinearity in the return to education. *Journal of Applied Economics 8*(1), 191.

Vélez, C. E., S. Al-Shawarby, and H. El-Laithy (2012). Equality of opportunity for children in egypt, 2000-2009: achievements and challenges.

Waldinger, F. (2006). Does tracking affect the importance of family background on students test scores? *Unpublished manuscript, LSE*.

WorldBank (2005). *World Development Report 2006: Equity and Development*. Oxford University Press, Incorporated.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. *Oxford Bulletin of Economics and Statistics 65*(2), 117–170.

Wößmann, L. (2008). Efficiency and equity of european education and training policies. *International Tax and Public Finance 15*(2), 199–230.

Appendices

Appendix A

Table 6: Probability of Finishing High School Condition on School Availability

	Dependent Variable: Finish High School (Whole Sample, 25=Age=60) [Logit Model]								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6	leduc6
qwealth2	0.505***	0.543***	0.563***	0.517***	0.576***	0.612***	0.510***	0.530***	0.549***
	(0.031)	(0.032)	(0.031)	(0.044)	(0.045)	(0.046)	(0.030)	(0.031)	(0.030)
qwealth3	0.764***	0.839***	0.894***	0.798***	0.908***	0.974***	0.754***	0.799***	0.856***
	(0.037)	(0.038)	(0.040)	(0.049)	(0.051)	(0.056)	(0.037)	(0.038)	(0.037)
qwealth4	1.066***	1.219***	1.273***	1.074***	1.303***	1.366***	1.087***	1.196***	1.255***
	(0.037)	(0.039)	(0.038)	(0.052)	(0.056)	(0.058)	(0.035)	(0.036)	(0.034)
qwealth5	2.086***	2.355***	2.414***	2.048***	2.470***	2.530***	2.176***	2.379***	2.457***
	(0.050)	(0.048)	(0.034)	(0.061)	(0.059)	(0.054)	(0.046)	(0.043)	(0.034)
rural	-0.872***	-0.889***	-0.902***	-1.084***	-1.122***	-1.133***	-0.737***	-0.743***	-0.762***
	(0.046)	(0.049)	(0.034)	(0.065)	(0.069)	(0.044)	(0.037)	(0.039)	(0.031)
hschool	1.089***	0.551***	0.576***	1.755***	0.971***	0.943***	0.783***	0.397***	0.461***
	(0.061)	(0.093)	(0.067)	(0.107)	(0.154)	(0.110)	(0.051)	(0.074)	(0.061)
Cohort FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE (province)	No	No	Yes	No	No	Yes	No	No	Yes
Sample Size	Whole	Whole	Whole	Female Only	Female Only	Female Only	Male Only	Male Only	Male Only
Observations	536,655	536,655	536,655	267,901	267,901	267,901	268,754	268,754	268,754

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Census 2006

Table 7: Probability of Ever Attend to Lower Secondary School Condition on School Availability

	Dependent Variable: Ever Attend Secondary School (Whole Sample, 25=Age=60)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5	leduc5
qwealth2	0.352***	0.467***	0.449***	0.429***	0.550***	0.533***	0.315***	0.430***	0.414***
	(0.025)	(0.025)	(0.023)	(0.035)	(0.036)	(0.033)	(0.028)	(0.028)	(0.027)
qwealth3	0.355***	0.551***	0.569***	0.449***	0.640***	0.653***	0.304***	0.510***	0.536***
	(0.027)	(0.026)	(0.025)	(0.039)	(0.040)	(0.038)	(0.029)	(0.029)	(0.028)
qwealth4	0.337***	0.678***	0.695***	0.401***	0.759***	0.767***	0.315***	0.653***	0.683***
	(0.028)	(0.028)	(0.027)	(0.040)	(0.042)	(0.040)	(0.030)	(0.030)	(0.030)
qwealth5	0.164***	0.655***	0.652***	0.270***	0.812***	0.791***	0.110**	0.581***	0.596***
	(0.051)	(0.051)	(0.056)	(0.061)	(0.071)	(0.081)	(0.051)	(0.046)	(0.047)
rural	-0.351***	-0.381***	-0.379***	-0.552***	-0.578***	-0.578***	-0.209***	-0.245***	-0.241***
	(0.023)	(0.024)	(0.023)	(0.032)	(0.033)	(0.033)	(0.024)	(0.024)	(0.024)
hschool	1.717***	0.331***	0.393***	2.221***	0.597***	0.639***	1.454***	0.219***	0.297***
	(0.074)	(0.061)	(0.055)	(0.122)	(0.103)	(0.095)	(0.064)	(0.056)	(0.052)
Cohort FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Location FE (province)	No	No	Yes	No	No	Yes	No	No	Yes
Sample Size	Whole	Whole	Whole	Female Only	Female Only	Female Only	Male Only	Male Only	Male Only
Observations	536,655	536,655	536,655	267,901	267,901	267,901	268,754	268,754	268,754

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Census 2006

Appendix B

Table 8: List of countries and surveys used in this study

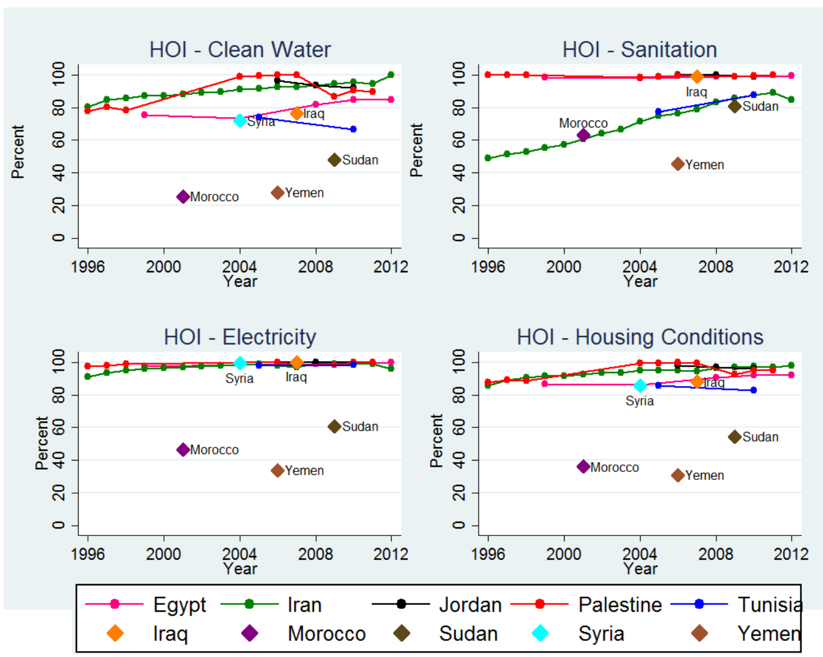
Country	Circa 2000	Circa 2010	Survey
Egypt	1999	2012	Household Income, Expenditure, and Consumption Survey, (HIECS)
Iran	2000	2011	Household Expenditure and Income Survey, (HEIS)
Iraq	2007	2012	HEIS (2007), Household Socio-Economic Survey (IHSES) (2012)
Jordan	2006	2010	Household Expenditure and Income Survey, (HEIS)
Morocco	2001	-	Household Expenditure and Income Survey, (HEIS)
Palestine	1996	2011	Palestine - Expenditure and Consumption Survey, (PECS)
Sudan	-	2009	National Baseline Household Survey (NBHS)
Syria	2004	-	Household Expenditure and Income Survey, (HEIS)
Tunisia	2005	2010	National Survey on Household Budget, Consumption and Standard of Living, (EBCNV)
Yemen	-	2006	Household Expenditure and Income Survey, (HEIS)

Source: ERF Data Portal, Statistical Center of Iran, World Bank.

Table 9: Sample sizes

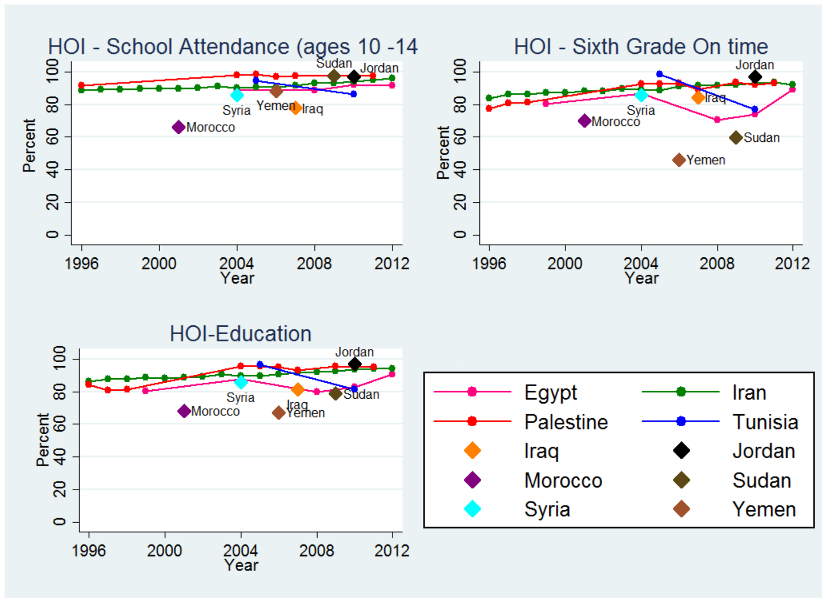
Country	Circa 2000	#Observation	Circa 2010	#Observation
Egypt	1999	40,247	2012	11,273
Iran	2000	50,634	2011	37,730
Iraq	2007	45,574	2012	66,441
Jordan	2006	6,782	2010	5,809
Morocco	2001	30,919	-	-
Palestine	1996	13,745	2011	10,796
Sudan	-		2009	20,746
Syria	2004	77,009	-	-
Tunisia	2005	16,252	2010	13,044
Yemen	-		2006	47,763

Figure 13: Human Opportunity Index in Housing over time



Source: Authors' calculations.

Figure 14: Human Opportunity Index in Education over time

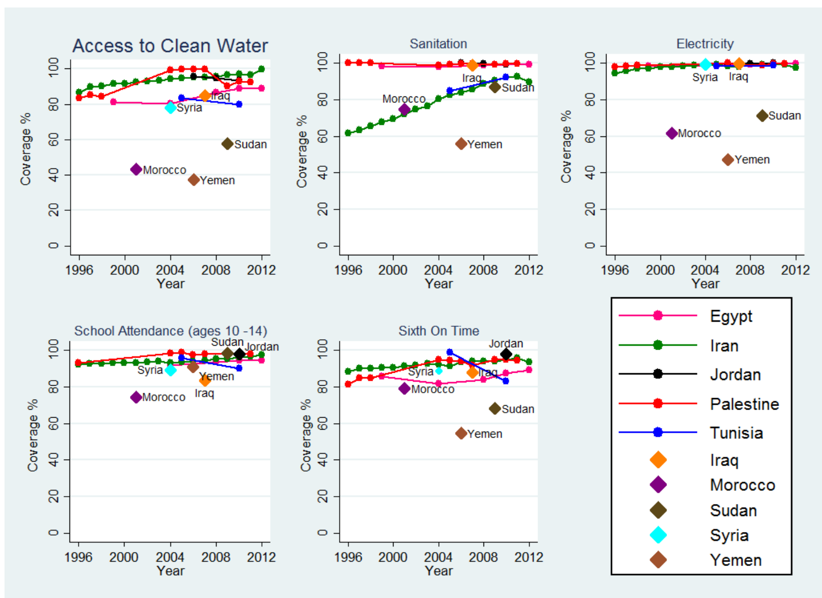


Source: Authors' calculations.

Figure 15: Equality of Opportunity index (1-D) over time

Source: Authors' calculations.

Figure 16: Coverage rate over time



Source: Authors' calculations.

Appendix C

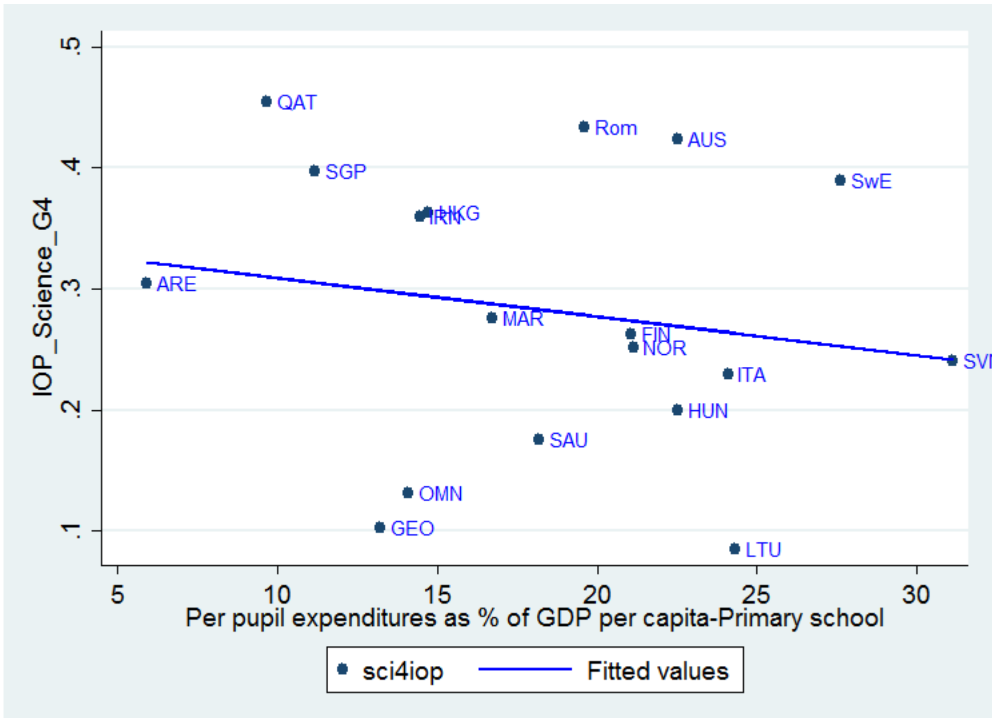


Figure 17: IOP in Science in grade 4 and per pupil expenditures in primary education

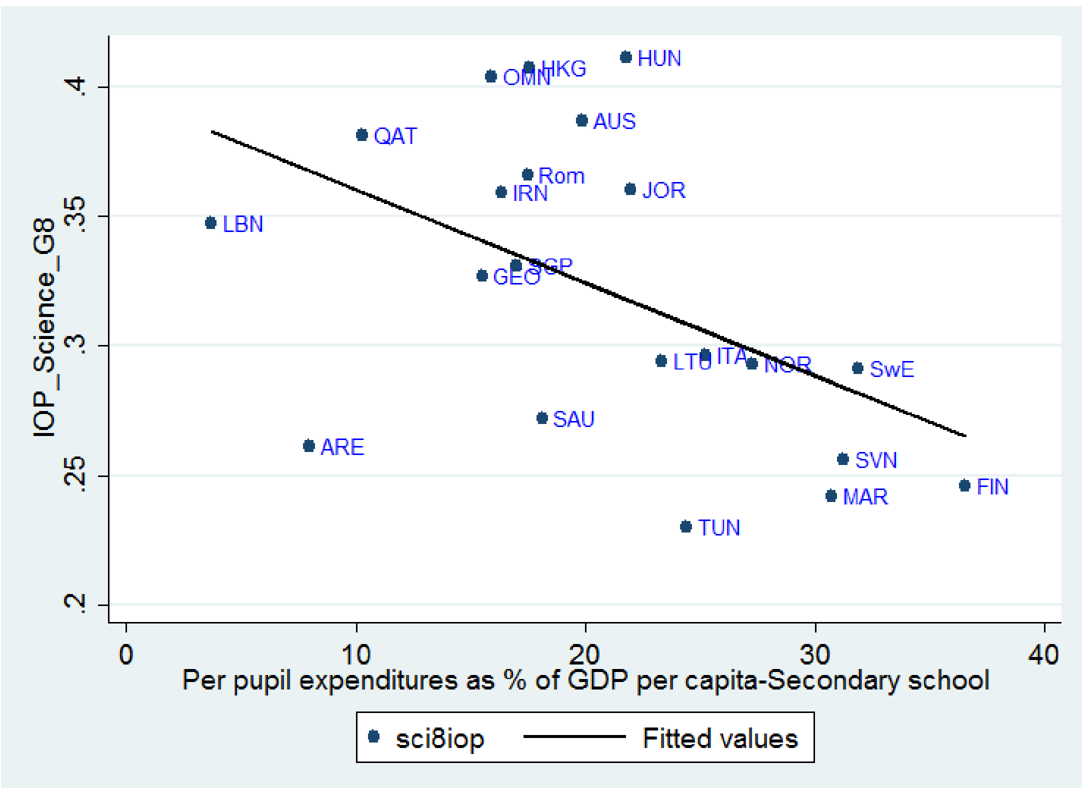


Figure 18: IOP in Science in grade 8 and per pupil expenditures in secondary education

Table 10: Expenditures in Education-2011

Country Name	Expenditure per student, primary (% of GDP per capita)	Expenditure per student, secondary (% of GDP per capita)	Public spending on education, total (% of GDP)	Public spending on education, total (% of government expenditure)
Australia	22.52	19.88	5.59	14.33
Finland	21.06	36.53	6.85	12.27
Georgia	13.23	15.49	2.70	9.27
Hong Kong	14.72	17.55	3.42	17.43
Hungary	22.52	21.81	4.90	9.80
Iran	14.45	16.37	4.05	15.39
Italy	24.09	25.26	4.50	8.93
Lithuania	24.31	23.30	5.42	12.87
Morocco	16.73	30.73	5.38	–
Norway	21.13	27.28	6.87	15.31
Oman	14.11	15.89	4.32	–
Qatar	9.69	10.29	2.45	7.36
Romania	19.61	17.47	4.24	–
Russia	–	–	4.10	11.96
S.Arabia	18.17	18.11	5.14	17.74
Singapore	11.17	17.00	3.14	20.52
Slovenia	31.13	31.26	5.69	12.08
Sweden	27.61	31.87	6.98	13.35
UAE	5.91	7.98	1.11	–
Bahrain	–	–	2.58	8.87
Lebanon	7.75	3.73	1.65	5.73
Palestinian	–	–	–	–
Tunisia	17.34	24.42	6.17	17.34
Turkey	–	–	2.86	8.55
Jordan	22.50	21.97	4.95	14.35

Note: For some countries the nearest available year is chosen.
Source: WDI.

Table 11: Gini Index

Country Name	GINI index	Country Name	GINI index
Australia	30.3	Russia	42
Finland	26.8	S.Arabia	–
Georgia	46	Singapore	47.8
Hong Kong	53.7	Slovenia	23.8
Hungary	24.7	Sweden	23
Iran	38.28	UAE	–
Italy	31.9	Bahrain	–
Lithuania	35.5	Lebanon	–
Morocco	40.88	Palestinian	35.5
Norway	25	Tunisia	36.06
Oman	–	Turkey	40.2
Qatar	41.1	Jordan	35.43
Romania	27.42		

Note: For some countries the nearest available year is chosen.
Source: WDI and
https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html

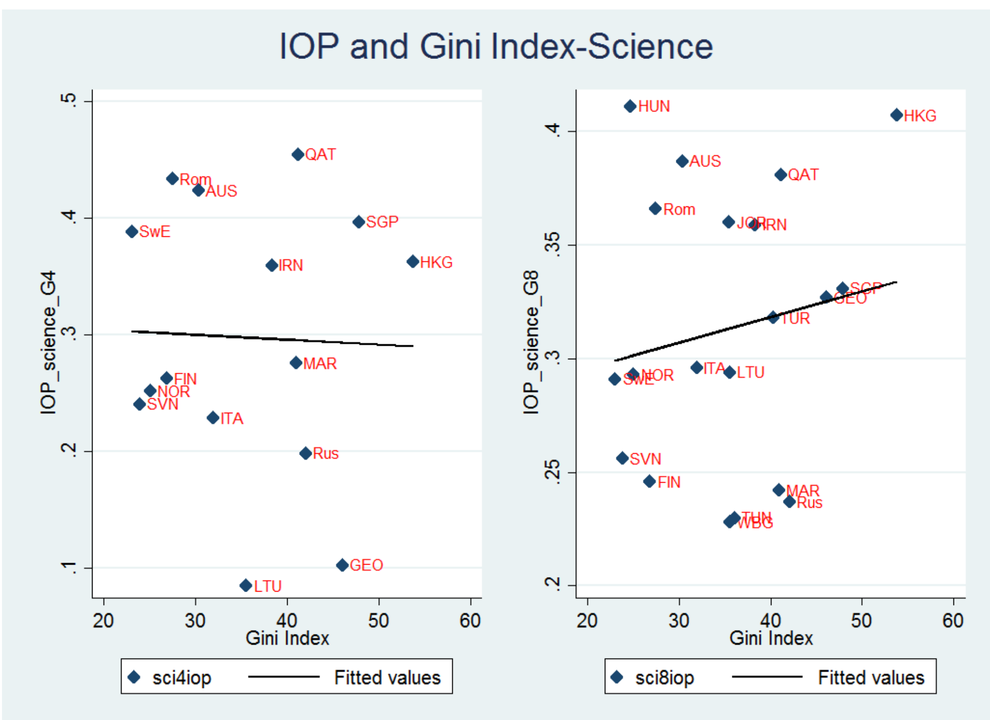


Figure 19: IOP and Gini Index-Science

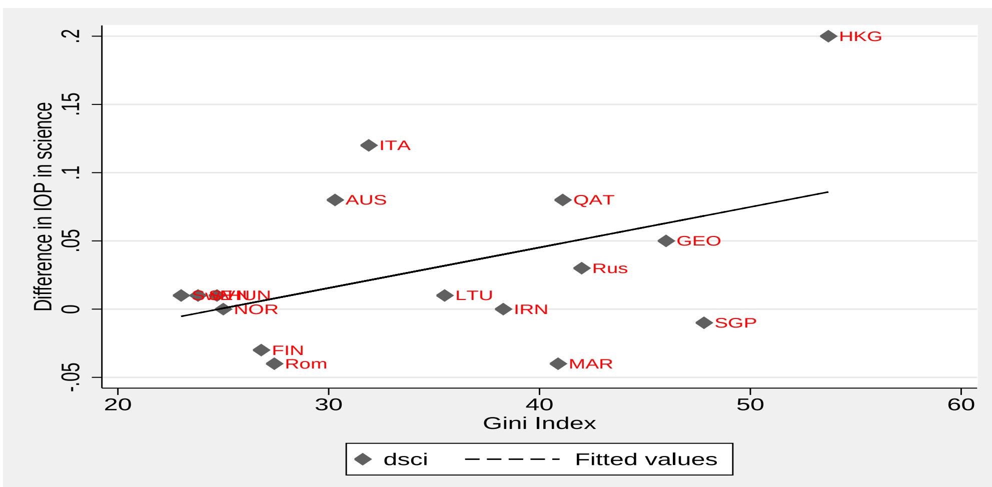


Figure 20: Difference in IOP in science between grade 4 and 8 and Gini Index

Table 12: National Balance

Country Name	GDP (PPP Constant 2011 International $)	GDP per capita (PPP, Constant 2011 International $)	Average Economic Growth (from 1991-2011)
Australia	9.330E+11	41763.119	1.834
Finland	2.192E+11	40683.528	1.713
Georgia	2.835E+10	7315.091	0.986
Hong Kong	3.542E+11	50085.959	3.051
Hungary	2.266E+11	22729.184	2.116
Iran, Islamic Rep.	1.350E+12	17949.244	2.796
Italy	2.158E+12	36347.343	0.757
Lithuania	6.918E+10	22845.336	5.924
Morocco	2.230E+11	6746.899	2.728
Norway	3.074E+11	62060.959	1.799
Oman	1.364E+11	42479.201	0.985
Qatar	2.525E+11	132514.495	1.969
Romania	3.646E+11	18095.013	2.372
Russian Federation	3.442E+12	24074.365	1.019
Saudi Arabia	1.367E+12	47474.043	1.467
Singapore	3.883E+11	74910.183	3.887
Slovenia	5.913E+10	28804.702	2.887
Sweden	4.135E+11	43755.060	1.704
UAE	5.031E+11	57594.127	-3.095
Bahrain	5.134E+10	39311.325	0.592
Lebanon	7.201E+10	15683.583	3.618
Palestine	1.711E+10	4356.342	3.315
Tunisia	1.089E+11	10123.120	2.869
Turkey	1.343E+12	18269.838	2.582
Jordan	6.980E+10	10324.447	2.158

Note: For some countries the nearest available year is chosen.
Source: WDI

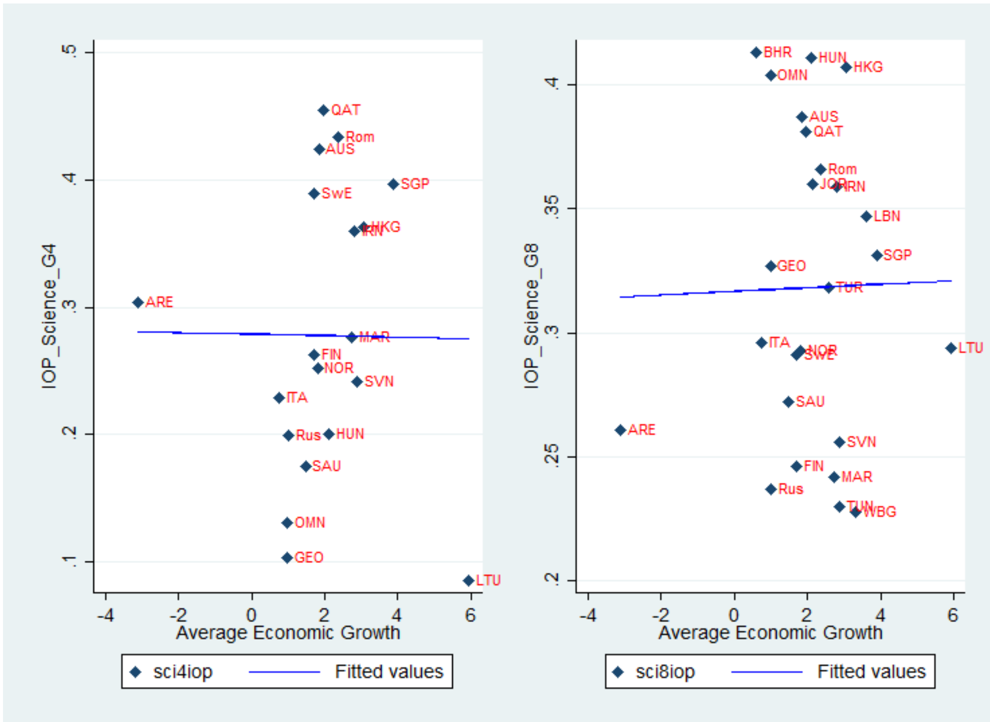


Figure 21: IOP and Average Economic Growth-Science

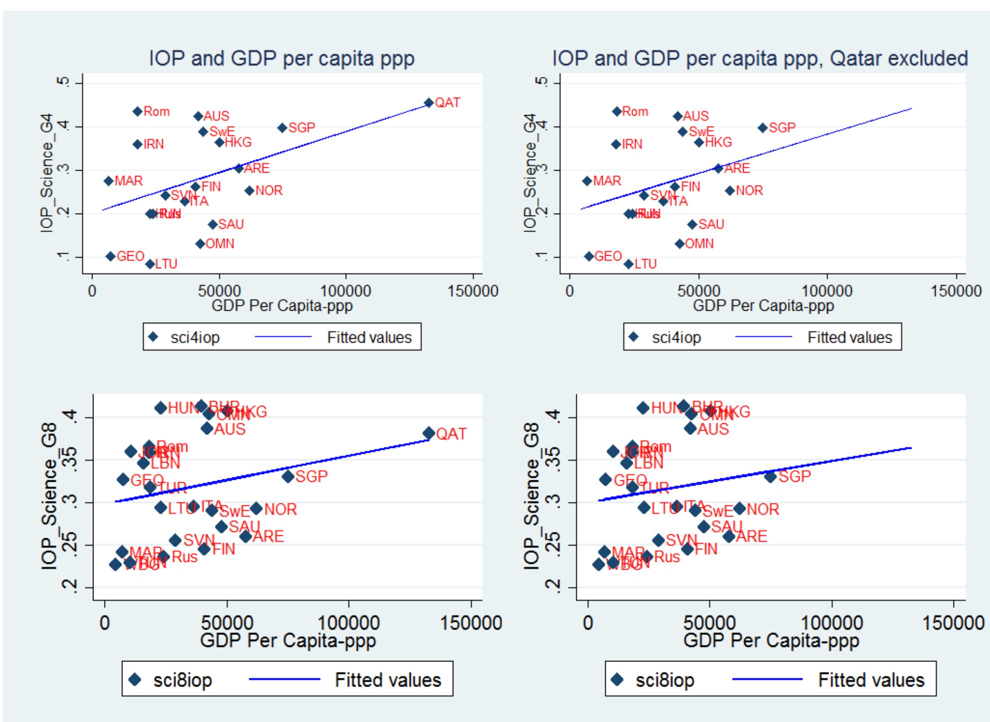


Figure 22: IOP and GDP per capita-Science