CHARLES UNIVERSITY IN PRAGUE

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Michal Topinka

Millennium Development Goal on Education:

An empirical analysis of the determinants of primary school enrolment

Bachelor thesis

Prague 2015

Author: **Michal Topinka**

Supervisor: **doc. Ing. Tomáš Cahlík CSc.**

Academic year: **2014/2015**

Bibliographic note

Topinka, M. (2015). Millennium Development Goal on Education: An empirical analysis of the determinants of primary school enrolment (Bachelor thesis). Charles University in Prague.

Character count: 55 971

Abstract

This bachelor thesis examines the relationship between country-level development indicators and the major indicator of the second Millennium Development Goal on education, the net enrolment ratio in primary education, using econometric analysis of panel data. Given that majority of studies analyzing the determinants of primary school enrolment use data from household surveys containing information about individuals, the study presented in this bachelor thesis investigates, whether variation in primary school net enrolment ratios can be explained using aggregate country-level data on adjusted net enrolment ratios and factors measuring development of particular economies. This study uses data on 192 countries for the period 1999-2012. The findings suggest that data on development are definitely useful in analyzing the determinants of school enrolment, but more detailed data containing information about individuals are needed to assess the causes of different levels of enrolment in individual countries in order to design policies to achieve universal primary education.

Keywords

Millennium Development Goals, MDGs, Education, Net enrolment ratio in primary education.

Abstrakt

Tato bakalářská práce zkoumá vztah mezi informacemi o vyspělosti jednotlivých zemí světa a hlavním indikátorem druhého Rozvojového cíle tisíciletí týkajícího se vzdělávání, kterým je poměr čistého zápisu do primárního stupně vzdělávání, za použití ekonometrické analýzy panelových dat. Vzhledem k tomu, že většina podobných studií využívá data z celostátních průzkumů domácností obsahující informace o jednotlivcích, studie provedená v této bakalářské práci vyšetřuje, jestli lze vysvětlit rozdíly v poměrech čistého zápisu do primárního stupně vzdělávání mezi jednotlivými zeměmi za použití pouze agregovaných dat obsahujících informace o zápisech do primárního školství a rozvojových indikátorech na celostátní úrovni. Tato studie využívá údaje o 192 zemích světa z období 1999-2012. Výsledky výzkumu napovídají, že informace týkající se indikátorů rozvojovosti jsou rozhodně důležité pro analýzu poměru čistého zápisu do primárního stupně školství, nicméně rozsáhlejší data obsahující informace o jednotlivcích státech jsou potřebná, aby bylo možné vytvořit takové politiky, které by napomohly k dosažení universálního primárního školství.

Klíčová slova

Rozvojové cíle tisíciletí, Vzdělávání, Poměr zápisu do primárního stupně vzdělávání.

Declaration of Authorship

I hereby proclaim that I wrote my bachelor thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 31, 2015

Signature

Acknowledgment

I would like to express my sincere gratitude to doc. Ing. Tomáš Cahlík CSc. for his time and valuable suggestions. I would also like to thank my family for their endless support during my undergraduate studies.

Contents

1 Introduction

In September 2000, representatives from 189 United Nations member states committed their nations to a new global development partnership and established a series of time-bound targets, which have become known as the *Millennium Development Goals (MDGs).*

The MDGs symbolize the biggest worldwide development effort in history. It is a set of eight development goals, adopted by most of the world's countries and designated to address not only the needs of the world's poorest, but to generally improve lives of people all around the world. The goals are expected to be achieved by the end of 2015 and aim to *1) Eradicate extreme poverty and hunger; 2) Achieve universal primary education; 3) Promote gender equality and empower women; 4) Reduce child mortality; 5) Improve maternal health; 6) Combat HIV/AIDS, malaria and other diseases; 7) Ensure environmental sustainability* and to *8) Develop a global partnership for development.*

Each MDG is defined by one or more targets, each having specific indicators for measurement of progress. The framework with specific indicators, along with information on how they are supposed to be measured, was developed in 2002. It originally contained 18 targets with 48 indicators and was effective between 2003 and 2007. In 2007 it was revised to include one revised and three new targets, still among the original eight development goals. The current framework is effective since 2007 until the end of the MDGs era in 2015. It contains 21 targets with 60 specific indicators, which are monitored by UN Statistics division and by many other institutions and agencies within and outside the United Nations system.

The progress is reviewed each year in the Millennium Development Goals Report, issued every year in July and available on the official MDGs website. Since Millennium Summit in September 2000, where the MDGs were officially adopted, there has been four major UN summits devoted to MDGs, where the representatives decided on the next steps in achievement of the MDGs depending on current progress.

Although the targets are concrete and specific, monitoring of progress is complicated due to unavailability of reliable data. With the data are missing, estimates are used, if available. This can potentially lead to inaccurate information about current progress and can possibly affect decision-making. The problem of missing data was

identified as a major problem in the last progress report, issued in July 2015. (United Nations, 2015)

There has been several global initiatives to enhance progress in achievement of MDGs. The most important was called The Millennium Project. (Unmillenniumproject.org, 2015) The Millennium Project was an independent advisory body headed by Professor Jeffrey Sachs, composed of worldwide network of development practitioners and experts across an enormous range of countries. Experts from governments, international financial institutions, nongovernmental organizations, academia and private sector conducted extensive researches throughout years 2002 to 2005 and in January 2005 submitted final report, called *Investing in Development: A Practical Plan to Achieve the Millennium Development Goals.* The report proposed a concrete set of actions and specific policies for countries to achieve MDGs. It proposed strategies for developing countries as well as concrete financial plan to support them through official development assistance and domestic resource mobilization.

Another worldwide initiative, The Millennium Campaign, also called End of Poverty 2015, was launched in 2002 and it has been working with UN partners and key global constituencies around the world to promote greater support for the Millennium Development Goals ensuring that they remain priority in the political and public agenda. (Endpoverty2015.org, 2015)

The Millennium Development Goals are considered to be the most successful global initiative in history. Despite the immense progress in most parts of the world, there are still regions who lack significantly behind and where most of the goals will not be fulfilled. (United Nations, 2015)

The main focus of this thesis is on the MDG on achievement of universal primary education. The element of education is widely known to be one of the most essentials. The MDG on education goes along with UNESCO initiative called Education for All (EFA). Education for All contains six specific education goals which were established also in 2000. Both MDGs and EFA have deadline in 2015. (Unesco.org, 2015)

Education is important, not only because it belongs among basic human rights (Un.org, 2015), but also because it is a key to human development. There is evidence that improvement in education will advance improvement in all MDGs. (UNESCO, 2010)

For example, one year of education is expected to rise future income by 10 percent and 171 million people are expected to be lifted out of poverty, provided that all students in low-income countries left schools with basic reading skills, which is equivalent to 12 percent cut in poverty.

Education also promotes gender equality. Achievement of universal primary education would mean that every girl would have the same education as boys and hence equal opportunities and freedom.

Educated mothers are less likely to die while giving birth, their children are more likely to survive first years of their lives and educated mothers have on average less babies, thus they do not suffer from such poverty and can afford to rise their children, provide them with basic vaccination and habits, preventing them from getting various diseases. People with higher levels of educations are also expected to behave in a way that helps to ensure environmental sustainability.

Overall, countries with higher education levels perform better in achievement of all MDGs, which is currently done by the fact that they were not so poor in history and were able to provide current adults with education necessary to achieve better results in all areas. This is very important to the future, because providing current children with education will help countries lift from poverty and achieve all MDGs in the near future.

The MDG on education formally seeks to ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling. It is accompanied by three indicators, namely *Net enrolment ratio in primary education (NER), Proportion of pupils starting grade 1 who reach last grade of primary* and *Literacy rate of 15-24 year-olds.* (Mdgs.un.org, 2015)

The numbers listed for 2015 in the last MDGs Progress Report and in the review below are estimates derived from 2012-2014 values.

The net enrolment rate (NER) in primary education is defined as *the ratio of the number of children of official primary school age who are enrolled in primary education to the total population of children of official primary school age, expressed as a percentage.*

Even though MDGs were established in 2000, most of the targets use year 1990 as a base year, which enables us to compare growth and rate of improvement before

and after establishment of MDGs.

Progress in achievement of universal enrolment has been unstable since 1990. Between 1990 and 2000, the enrolment rates rose from 80 to 83 percent. After establishment of MDGs in 2000 improvement escalated and in 2007 the overall adjusted net enrolment rate reached 90 percent, which indicates significantly better performance in the first half of MDGs period. After 2007, the progress slowed and is expected to reach 91 percent by the end of 2015. The threshold for achievement of universal primary enrolment was set to 97 percent and it has been already met in Northern Africa and Eastern Asia. All regions are close to achieve universal enrolment except for the Sub-Saharan Africa, where the enrolment is expected to reach 80 percent by the end of 2015, which is still an impressive improvement from 52 percent in 1990 and 60 percent in 2000. Moreover, the number of children enrolled in primary education in Sub-Saharan Africa more than doubled between 1990 and 2012.

According to 2012 estimates, 43 percent of current out-of-school children will never go to school. Overall, girls tend to have lower enrolment ratios in all parts of the world, but boys are more likely to leave school early.

Second indicator relating to MDG on education monitors primary school completion rate, defined as *a proportion of pupils enrolled in grade 1 of the primary level of education in a given school year who are expected to reach the last grade of primary school, regardless of repetition.* (Mdgs.un.org, 2015)

Survey data show that in low- and middle-income countries, on the average, the primary school completion rate rose from 70 percent in early 1990s to estimated 84 percent in 2015. (United Nations, 2015) This means that about 100 million children who enroll in primary school are not expected to reach the last grade. Moreover, adolescents from the poorest households are more than five times likely not to graduate from primary school than those from the richest households.

Third and last indicator in second MDG monitors youth literacy rate, defined as *a proportion of the population aged 15–24 years who can both read and write with understanding a short simple statement on everyday life.* (Mdgs.un.org, 2015)

The progress in achievement of this goal has been slow, but steady since 1990. The literacy rates rose, on average, from 83 percent in 1990 to 89 percent in 2010. (United Nations, 2015)

According to projections based on historical trends, youth literacy rate is expected to reach 91 percent in 2015. Young men are expected to reach literacy rate of 93 percent and young women are expected to reach 90 percent, even though they are tend to perform better at school, because more boys are still enrolling at schools than girls.

The gender parity in primary education was achieved in about two-thirds of developing regions, which is about 12 percent increase from 2000. This formally corresponds to the third MDG on gender equality, but it relates to education and so it is stated here.

Despite tremendous progress across world since 2000, the MDG on education nor the EFA goals were reached in most parts of developing world. Strong regional differences remain and more action is needed in post-2015 development agenda to address the needs of specific groups of children, especially girls and children belonging to minorities, engaged in child labor, living in conflict situations and those living with disabilities. It is critical to reflect on and address the root causes of limited progress in youth literacy in some parts of the world. It is also very important to assess if children, who have completed primary school have really obtained basic mathematical and reading skills, which are essential for a human in twenty-first century. (United Nations, 2015)

The MDGs showed that collective action is important and that it can be done, that decent living standard for every human on the planet can be achieved when there would be as much effort as it has been during the last fifteen years. Even though MDGs were not fulfilled in some parts of the world, it is still considered to be a huge success. In fact, world leaders have already established a new set of goals, called *Sustainable Development Goals (SDGs)*, which are expected to be adopted in September 2015. (United Nations Sustainable Development, 2015)

The proposal for SDGs contains seventeen specific targets, ranging from ending poverty to education, health, infrastructure, environment, global partnership and peace. The period for attaining goals remained the same as in MDGs and most of the goals have target dates set to 2030. The set of SDGs is more comprehensive, with each of the goals containing many targets. SDGs set high aspirations and it is obvious that more confident approach was taken when establishing these goals, driven by the success of the MDGs.

The SDGs contain all MDGs in some form, but the goals go far beyond the targets established in MDGs. The SDGs also contain specific goal on education. Its proposal contains ten bullet points aspiring to ensure universal primary and secondary education and a high level of tertiary education and also to improve significantly a quality of education, so that all children and youth can acquire necessary skills needed to provide them with quality lives.

The higher importance is set to environmental issues, which have come to attention of all policy makers all around the world in recent years and definitely need collective agenda to combat various dangers of climate change and depletion of resources.

Last development effort that I would like to mention is called *The Oxford Martin Commission for Future Generations.* The Oxford Martin Commission for Future Generations is an initiative of The Oxford Martin School at the University of Oxford, created by Dr. James Martin in 2005. The Oxford Martin School is a research community of over 300 scholars, working together to harness the opportunities of the 21st century and to address its most pressing challenges.

The Oxford Martin Commission for Future Generations was established in September 2012 by Pascal Lamy, former Director-General of the World Trade Organization. It is a group of 19 leaders from various sectors of society, such as government, academia, business, media and civil society working together to address the growing short-term fixation of modern politics and business. (Oxford Martin School, 2015)

The report analyses the issues and megatrends in the current world, examines lessons from past successes and failures and proposes set of principles to overcome deep political and cultural divides. It also presents practical recommendations for action on critical challenges. Commissioners have been working together with leaders such as Ban Ki-moon, UN Secretary General, Herman van Rompuy, President of the European Council, or Christine Lagarde, Managing Director of International Monetary Fund and continue to engage with governments, non-government organizations and civil society in order to take the recommendations forward.

The key message is for decision-makers to think in a long-term, to adopt long-term policies which will benefit future generations. The Now in the Long Term report highlights MDGs as an example off successful international long-term effort, where almost all national governments incorporated MDGs as central components of their

development agendas and it also encourages world leaders to continue to engage in SDGs.

In this thesis, I would like to further examine MDG on education, particularly the determinants of improvement in the first indicator, net enrolment rate in primary education.

The contribution of this thesis is to determine, if trends in improving primary education can be explained on large scale with only country level on primary net enrolment rates and with data from public accessible datasets using standard econometric methods. Provided that most studies use data from detailed household surveys containing information about number of schools in particular cities, distance to school, education of parents or number of children in households, it would be interesting to see if country level data on standard development indicators can get valuable insights on explaining the determinants of net enrolment rates in primary schools, or if it is really necessary to use deeply detailed data and analyze the situation in each country separately. This would be indicated by inaccurate and possibly misleading results and would mean the need for policy makers to significantly increase the number of surveys in all countries and investing large amount of resources, if trends in education are to be explained on large scale.

2 Literature review

The literature on determinants of school enrolment rates and school completion rates, two major indicators of the second MDG on education, is quite broad. However, majority of studies are targeted to specific countries and use national household surveys containing detailed information about local school system. There is a small number of studies, which tried to assess the problem on larger scale, for example Huisman and Smits (2008), who were estimating the determinants of primary school enrolment in 30 developing countries using data from Demographic and Health surveys program and Pan Arab Project for Family Health.

Huisman and Smits from Radboud University used bivariate and multivariate logistic regression analysis to explain the relationship between school enrolment rates and family background characteristics. As a dependent variable, dummy variable indicating whether or not a child aged 8-11 was enrolled in a school at the time when the household survey was used. The analyses were performed separately for boys and girls.

As independent variables, they used data on parents' employment status and education separately for mothers and fathers. Another variables included household wealth as a proxy for household income, birth order of the child, number of siblings, size of a family and other family-related factors, urbanization and age of interviewed children. The regressions also included information about distance to schools and quality of schools measured by number of teachers and pupil-teacher ratios. Several other independent variables measuring labor, cultural and development statistics were also included. Besides direct effects of the household and district-level variables, also interaction effects between variables at both levels were analyzed.

Socio-economic factors were found to be very important in explaining enrolment rates. Especially parental education and health were found to be very important. Children with parents who attained higher levels of education and who are wealthier had significantly larger chance to be enrolled in primary school. Characteristics of the family structure, such as number of siblings, birth order and size of the family were also found to be important.

Primary enrolment was found to be largely depend on characteristics of the educational facilities in the district where the children live. When children have to

travel longer distances to school, when there are fewer teachers available and when the average class size is larger, both boys and girls are less likely to go to school. Interaction analysis proved that effects of many household-level variables are different when the children live in urban or rural areas with different access to education facilities. The paper concludes with suggestion for possible policies to increase enrolment ratios based on findings presented in a paper.

Richards and Vining (2014) studied the importance of World Bank country development factors and information about national governance in explaining the changes in primary school completion rates, one of the three indicators associated with second MDG, which aspires to achieve universal primary education. Richards and Vining used data on primary school completion, GDP per capita, literacy and governance variables for 66 low income countries between years 2001-2010.

The motivation of this study was to shed a light on determinants of primary school completion in low income countries, where the MDG would not be met for majority of countries and also to determine, if the rate of increase or decrease changed rapidly during the first decade of twenty-first century, the period when MDGs were one of the top world priorities.

Richards and Vining used two sets of regressions. The purpose of the first set of regressions was to examine the relationship between primary school completion rates and GDP per capita, adult literacy rate, defined as literacy rate of those older than 15, government spending per student and data on governance from World Bank Worldwide Governance Indicators (WGI). WGI included information on effectiveness of government spending, governance effectiveness, political stability and government voice and accountability score. Due to expected collinearity among governance variables, each governance variable was used separately, giving four regressions per each set. For the first set of regressions, the whole period of 2001-2010 was used, where binary independent variable was used to indicate whether the observation is from the first or second part of the decade.

Second set of regression aimed to explain the extent to which countries in the second half of the decade improved their average completion rates relative to the first half. The dependent variable for this exercise was a ratio. The numerator was the change in the average completion rate in 2006-2010 relative to 2001-2005. The denominator was defined as 100 percent completion rate minus the average completion

rate for the period 2001-2005. As independent variables, averages for literacy and GDP per capita over years 2001-2005 and averages for governance indicators over a period 2006-2010 were used. As in the first set of regressions, each governance variable was used separately, except for the effectiveness of government spending, which was omitted.

From the first set of regressions, authors concluded that literacy and GDP per capita are important in explaining changes in completion ratios, whereas the governance variables were found only marginally significant. The variable indicating the period from which the observation was obtained was found positive and very significant, implying that a given country's completion rate is expected to be about ten points higher in the second half relative to first half of the decade.

In explaining improvements in completion rates between the first and second half of the decade 2001-2010, neither literacy nor GDP per capita were found to be the key variables, unlike the averages for governance variables, which were found to be very statistically significant. On the other hand, those improvement regressions left much of a variance unexplained, all having low coefficients of determination.

At the end of the paper, authors emphasize possibly the biggest limitation of the MDG on education. The problem is that the MDG on achieving universal primary education is only a quantitative measure and does not measure the quality of schools and real education level obtained by children. Authors demonstrate this thought on India, which is an example of a country that managed to dramatically increase the primary school completion ratios, but experienced large decrease in overall academic results. This problem is also studied in a work performed by Filmer, Hasan and Prichett (2006) from Center for Global Development.

Filmer, Hasan and Prichett argued that reaching the second MDG does not necessarily mean achieving universal primary education. They outlined a concept Millennium Learning Goal, which focuses on a real educational achievement, and used data for seven developing countries - Brazil, Indonesia, Mexico, Thailand, Tunisia, Turkey and Uruguay, to study the importance of reaching the second MDG on achievement of MLG. Filmer, Hasan and Prichett show that MDG and MLG are not closely related, meaning that achievement of MDG on education did not provide children with enough mathematical and reading skills.

They support these finding with various studies, where children who completed

primary schools, when asked, were not able to answer basic mathematical questions or had troubles completing correct word into a sentence. Authors then suggested more practical methods to assess children's skills both in and out of school to monitor real state of educations among children as a basis for achieving the true universal primary education.

Another study on determinants of school enrolment was conducted by Mutangadura and Lamb (2003) from University of North Carolina. In their paper, researchers examine the national characteristics that explain variation in indicators of entry into the first grade and primary school enrolment rates. They used data for 29 countries in Sub-Saharan Africa for the period 1980-1997. Even though the study was conducted using the data from period before establishment of MDGs, the purpose and content of this study is relevant to this thesis and so the review of this work necessarily belongs into this literature review.

Mutangadura and Lamb conducted pooled time series analyses using multivariate regression models to estimate the effect of economic, cultural and demographic variables on two indicators used to measure the children's access to primary education and actual enrolment rates. First dependent variable, called apparent primary intake rate, was defined as a ratio of the number of new entrants to first grade, regardless of age, to the number of all youth who are eligible for first grade. This variable can essentially be defined as more familiar gross enrolment ratio. The second dependent variable was primary net enrolment ratio, defined as the number of age-eligible students that are enrolled in primary school to the total total population of the same age group.

As an independent variables, researchers used GNP per capita, national debt as a percentage of GNP, education expenditure as a percentage of GNP, percentage of urban population, variable indicating number of ethnic groups within a country and variable indicating whether the country is located in West Africa, or South/East/Central Africa. Two models were estimated for each dependent variable, one with the explanatory variables only and one that includes time trend.

When estimating the first regression model with primary intake rate as dependent variable, expenditure on education, percentage of urban population and West Africa region were found to be significant predictors, with first two having positive effect on primary intake rate.

The second model estimating the effects on primary intake rate investigates the impact of adding time trend to the list of variables in the first model. Controlling for the other variables, time trend has a positive effect on rate of entry into first grade, which was increasing about half point annually during the period of study. Among the significant variables from the first regression, only urbanization rate were no longer statistically significant predictor. The models with and without time trend explain 41 and 42 percent of the variance in the observed dependent variable across nations, respectively.

When estimating the effect of independent variables on net enrolment ratio in regression without the time trend, same independent variables as in the first regression plus outstanding governmental debt were found to be significant predictors of predicted variable. When adding the time trend, neither GDP per capita nor the time trend were found to be significant explanatory variables.

Overall, the presented study showed that the major determinants of entry into first grade and primary school enrolment in Sub-Saharan Africa are expenditure on education, urbanization rate and location of the country. GNP per capita and debt as a percentage of GNP were found significant predictors only in one of the two models for each dependent variable. The paper concluded with possible policy implications for governments in Sub-Saharan Africa regarding to finding presented in study conducted by researchers.

There has been several other studies conducted to estimate relevant predictors of primary school enrolment and completion rates, but these were tailored to specific countries mostly using only a few period of data and although they are definitely relevant when predicting specific policies to achieve universal primary education in countries, where the studies were performed, they are not that relevant to this thesis and therefore the reviews of these studies will not be included in this text.

3 Description of data

3.1 Data sources

For this empirical study, I used data from two publicly accessible online data sources. The data on school enrolment were obtained from UNESCO Institute of Statistics (UIS) and the data on development indicators were obtained from World Bank World Development Indicators database (WDI).

The UNESCO Institute of Statistics is the primary source for cross-nationally comparable statistics on education, science and technology, culture, and communication for more than 200 countries and territories. The UIS produces international monitoring indicators based on its annual education survey in more than 200 UN Member States and territories and it is the official source for monitoring education-related MDGs and EFA goals. (Uis.unesco.org, 2015)

The World Bank is a free and open data source and its World Development Indicators database is a collection of development indicators, compiled from officially-recognized international sources. The WDI presents the most current and accurate global development data available, and includes national, regional and global estimates. It contains time series data on variety of country and region development factors from 1960-2014. (Data.worldbank.org, 2015)

3.2 Dependent variable

For my empirical study, I used data on adjusted net primary enrolment rates as my dependent variable. The adjusted NER is defined as *the number of children of official primary school age who are enrolled either in primary or secondary education expressed as a percentage of the total population of children of official primary school age.*.(Mdgs.un.org, 2015)

The adjusted NER provides the most accurate measure of primary school enrolment, because it includes the children who enter primary school early and advance to secondary school before they reach the official upper age limit of primary education. NER below 100 percent provide a measure of the proportion of primary school age children who are not enrolled in primary school and alert policy makers to the need for policies that increase primary school enrolment in order to achieve the goal of universal primary education. (Mdgs.un.org, 2015)

Data on school enrolment are usually recorded by the ministry of education or derived from surveys and censuses. If administrative data are not available, household survey data may be used, although household surveys usually measure self-reported attendance rather than enrolment as reported by schools and might not be comparable between surveys. A serious problem with household survey data is also the inaccurate recording of pupils' ages, depending on the time of the year that the survey is conducted. (Mdgs.un.org, 2015)

The UIS produces time series for adjusted NER based on enrolment data reported by education ministries or national statistical offices through questionnaires sent annually to countries, and United Nations population estimates. The data received by UIS are validated using electronic error detection systems that check for arithmetic errors and inconsistencies and perform trend analysis for implausible results. Queries are taken up with the country representatives reporting the data so that corrections can be made or explanations given to errors and implausible results. (Mdgs.un.org, 2015)

Due to inability of large number of governments to consistently provide enrolment data and due to insufficient number of surveys, UIS is not able to collect or estimate data on NER, which results in large amount of gaps in UIS database.

For the most appropriate measurement, the data should be disaggregated by gender, age, geographic location, social and ethnic groups, and type of school. I use data aggregated by sex, because I do not expect to find important insights when using regression with only country-level data of all variables. The gender separated data are also not available for particular countries, which would lead to less observations available for regression. The gender inequality is likely to be determined by social and ethic factors, which are not observed in this research and therefore I would leave this issue unexamined in this study.

I divided the adjusted NER together with all explanatory variables that are usually expressed as percentages by 100 to convert them to ratios with values from 0 to 1.

3.3 Independent variables

The selection of explanatory variables was, inter alia, influenced by availability of data. For example, I was not able to obtain enough information about quality of school system (e.g. number of schools, number of teachers, etc.). Therefore, the only

data I obtained from UIS are on adjusted net enrolment rate and everything else is take from WDI database. The first independent variable is Gross national income (GNI) per capita converted to U.S. dollars using the World Bank Atlas method, divided by the midyear population. GNI is the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad. (Data.worldbank.org, 2015) The purpose of the Atlas method is to reduce the impact of exchange rate fluctuations in the cross-country comparison of national incomes. (Datahelpdesk.worldbank.org, 2015)

Since I use only country-level data in my analysis, I use GNI per capita to compare income levels across countries. I expect that higher GNI per capita will be generally associated with higher enrolment rates.

For the second independent variable, I chose the level of urbanization. It is calculated as ratio of people living in urban areas expressed as a percentage of total population in individual countries. I expect that countries with higher level of urbanization will have higher level of primary school enrolment, because of better accessibility of schools.

The third explanatory variable is an education expenditure as a proportion of GNI. It refers to the operating expenditures in education, including wages and salaries and excluding capital investments in buildings and equipment. This indicator indicates the importance in government policies in individual countries. The shortcoming in using this indicator is that even when countries invest more of their resources, they may still be very poor and have lower enrolment ratios compared to richer countries, who invest less of their GNI in education. The more appropriate measure would be a logarithm of government expenditure per student expressed in international dollars (An international dollar would buy in the cited country a comparable amount of goods and services a U.S. dollar would buy in the United States.(Datahelpdesk.worldbank.org, 2015)). Unfortunately, such measure is not available. Nevertheless, I expect countries, who invest more resources in education, to have slightly higher enrolment rates.

For the last two explanatory variables, I chose infrastructure variables, namely percentage of population using an improved water source and percentage of population using improved sanitation facilities. Children from countries with low levels of water and sanitation facilities have higher chance of getting serious diseases, which prevents them in enrolling to schools. Girls are especially affected, because fetching water belongs to essential household chores in developing countries, which almost always falls to women and girls, who usually have to stay home to help with work in households. All children need a sanitary and hygienic learning environment, but the lack of sanitation and hygiene facilities in schools has a stronger negative impact on girls than on boys. (UNICEF, 2015) The lower enrolment of girls and boys due to bad sanitation and impact of various diseases associated with bad hygienic environment can significantly reduce NER. I would want to test these assumptions made by UNICEF and I expect both variables to be positively affect the explained variable.

As stated before, all variables, that are originally expressed as percentages (i.e. all except GNI per capita) were divided by 100 to convert them to proportions with values between 0 and 1.

3.4 Summary statistics

To test hypotheses and assumptions presented in the previous section, I created a panel data set with data on country-level data on above mentioned indicators with period from period 1999-2012. The reason for this particular period is that most of the MDGs have 1990 as a base period and progress is measured mostly since 1990. (United Nations, 2015) I chose period from 1999-2012, because there is significantly smaller amount of data in period 1990-1998, compared to 1999-2012. The reason for that is unknown, but may potentially negatively influence the results of regressions and so the data from 1999-2012 are not used in any regression. The amount of data available in period 1999-2012 are almost constant and since 2013 the data are not yet available for some of the selected indicators. The summary of distribution of observations among individual years is presented in Table 1.

To provide better picture of data used in this study, Table 2 summarizes the data obtained on individual observed variables. I separated observations of country-level variables by income groups, to see if the levels of selected indicators are influenced by income levels. Since all of the variables, except maybe for the education expenditure,

are considered to be indicators of development, I expect that higher income will be generally associated with higher levels of these variables.

Table 2 definitely supports the above mentioned hypothesis. It can be seen by looking at mean, minimal and maximal values for individual variables. Low income countries have the lowest mean values in all indicators, except for the quite extreme maximal value for education expenditure, which belongs to Zimbabwe in 1999. When looking at maximal enrolment rates, one can see that there are low income countries who managed to achieve universal enrolment, at least in certain years, indicating that income is not the only factor, which can significantly influence enrolment rates.

Table 2 also proves the consistency of the obtained dataset. Even though the dataset is consistent, it is highly unbalanced. This is due to large amount of missing observations, especially on adjusted NER. This represent a major shortcoming in monitoring MDGs and it was discussed previously in this document. I managed to obtain data on 1499 observations, which represent about one half of possible observations, which would be obtained if I collected data on 226 countries for period of 14 years.

Provided that missing observations are not correlated with current or past values of observed and unobserved factors that change over time, the strongly unbalanced dataset should not cause big problems. The only thing I can expected with certainty is getting less accurate estimates when compared to strongly balanced dataset. However, the consistent and significant estimates could still be obtained, provided that appropriate methods will be used.

Table 1: Data structure

Year	Obs	Percent
1999	104	6.94
2000	109	7.27
2001	103	6.87
2002	109	7.27
2003	108	7.20
2004	108	7.20
2005	119	7.94
2006	111	7.40
2007	114	7.61
2008	105	7.00
2009	107	7.14
2010	101	6.74
2011	100	6.67
2012	101	6.74
Total	1499	100.00

Table 2: Summary statistics 1999-2012

Variable	Income group	Obs	Mean	Std. Dev.	Min	Max
A NER	LI	345	74.869	18.207	25.487	99.596
	LMI	468	90.258	10.825	25.587	99.976
	UMI	298	95.126	4.13	82.387	99.991
	HI	388	97.967	2.133	85.718	99.999
GNI PC	LI	345	457.536	206.15	110	1030
	LMI	468	1959.017	821.573	760	4020
	UMI	298	5725.369	2015.71	2970	12370
	HI	388	32991.52	17074.52	9810	99100
URBAN	LI	345	31.317	12.273	8.246	76.485
	LMI	468	50.523	16.103	12.982	81.964
	UMI	298	60.699	17.699	10.072	94.414
	HI	388	76.129	16	9.092	98.217
EDUC EXPEND	LI	345	3.751	3.26	.85	32.368
	LMI	468	4.127	1.824	.85	9.8
	UMI	298	4.504	1.86	1.7	14
	HI	388	5.073	1.317	1.979	9
SANITATION	LI	345	33.507	24.527	6.4	100
	LMI	468	68.175	22.118	13	100
	UMI	298	85.075	13.421	30	100
	HI	388	98.835	3.252	70.5	100
WATER	LI	345	65.979	15.011	27.1	95.1
	LMI	468	85.488	11.584	39.7	99.7
	UMI	298	94.208	5.079	69.1	100
	HI	388	99.355	1.455	90.8	100

4 Econometric methods

4.1 Econometric models

To study the effect of development indicators on primary school enrolment rates, I estimate the following three econometric models.

$$y_{it} = \beta_0 + \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad t = 1, ..., T \tag{1}$$

$$y_{it} = \alpha + \boldsymbol{x}_{it}\boldsymbol{\beta} + \bar{\boldsymbol{x}}_{\boldsymbol{i}}\boldsymbol{\gamma} + \varepsilon_{it}, \quad t = 1, ..., T \tag{2}$$

$$E(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{w}_{it}, s_{it} = 1) = \Phi\left[\frac{\boldsymbol{x}_{it}\boldsymbol{\beta} + \sum_{r=1}^{T}(\psi_r g_{ir} + g_{ir}\bar{\boldsymbol{x}}_{\boldsymbol{i}}\boldsymbol{\xi}_{\boldsymbol{r}})}{exp\left(\sum_{r=1}^{T} g_{ir}\omega_r\right)}\right] \tag{3}$$

where $g_{ir} = 1[T_i = r], \quad t = 1, ..., T$

To overcome the problem of unbalanced dataset, I decided to employ four methods of estimation and compare results between them. I estimate the first model by the *fixed effects transformation (FE)* and *random effects transformation (RE).*

4.2 FE, RE, CRE

The problem with estimating model (1) by FE or RE is that y is not exactly and *approximately continuous variable.* The adjusted NER is bounded by 0 and 100, or in this case, by 0 and 1. On the other hand, NER can take on any value between 0 and 1 and as stated in Woolridge (2012): "If a strictly positive variable takes on many different values, a special econometric model is rarely necessary." Moreover, both FE and RE methods are very useful when dealing with panel data, especially the fixed effects estimation, because it eliminates the unobserved effects which may influence school enrolments. Unbalanced panel dataset is also not a problem and therefore I decided to use these two methods, because I believe they can still provide valuable insights in my analysis.

Third method, estimating equation (2) is called the *correlated random effects approach (CRE).* This method leads the same estimates as FE, but I decided to use it as well, because it helps to determine which of the two previously mentioned estimation methods is better for particular econometric model. This model suffers from the same shortcoming as the previous two, assuming y to be approximately continuous variable.

All three above mentioned methods are explained in Wooldridge (2010a) and Wooldridge(2012) and their further description will not be provided in this text.

4.3 Fractional heteroskedastic probit

Fractional heteroskedastic probit estimation (FHETPROB) is a method designed to estimate nonlinear models with unobserved heterogeneity when dependent variable is a fraction (i.e. $y_{it} \in \langle 0, 1 \rangle$). This method is based on quasi-maximum likelihood estimation and was firstly proposed in Wooldridge(2010b) and the user-written implementation into statistical software was presented in 2013 also by J. M. Wooldridge. The theoretical specifications of this method are more advanced and can be found in Wooldridge (2010b). Example of practical implementation can be found in Wooldridge (2013) and Bluhm (2013).

This method theoretically appears to be the most suitable for the analysis presented in this thesis, because it allows for unobserved heterogeneity and it accounts for the fact that $y_{it} \in \langle 0, 1 \rangle$. Moreover, it allows for unbalanced panel data. Unfortunately, it is difficult to obtain goodness-of-fit measure, which would enable to compare, for example, between FE and FHETPROB.

4.4 Verification of the assumptions

To estimate equations by $(1), (2)$ by FE, RE and CRE, I shall test validity of the assumptions listed in Wooldridge (2012, p.509-510).

The first assumption of all four estimation methods is the presence of unobserved effect. In this case, it is generally assumed by basic intuition. It is obvious that there exist country specific factors influencing enrolment rates in individual countries. To formally test for presence of unobserved effect I will use the following assumptions:

$$E(c_i|\boldsymbol{x}_{it}) = E(c_i) = 0$$

$$E(c_i^2|\boldsymbol{x}_{it}) = \sigma_c^2$$

$$E(\upsilon_{it}\upsilon_{is}) = \sigma_c^2, \quad for\ all\ t \neq s$$

$$Cov(\upsilon_{it}, \upsilon_{is}) = E(\upsilon_{it}\upsilon_{is}) - E(\upsilon_{it})E(\upsilon_{is})$$

$$E(\upsilon_{it}) = E(\upsilon_{is}) = 0$$

The absence of unobserved effect is equivalent to null hypothesis $H_0 : \sigma_c^2 = 0$. Following above listed assumptions, it is enough to test for AR(1) serial correlation, for which I will use a test proposed in Wooldridge(2010), using implementation derived by Drukker (2003). The test strongly rejects null hypothesis of no first-order autocorrelation, indicating presence of unobserved effect, and also violation of assumption FE.6 in Wooldridge (2012, p.509).

To justify the assumption of random sampling, also needed for all estimation methods, let $\{s_{it} : t-1, ..., T\}$ be a sequence of "selection indicators". Let $s_{it} = 1$ if and only if observation (i,t) is used. The number of time periods available for unit i is $T_i = \sum_{r=1}^{T} s_{ir}$. This is properly viewed as random sampling. (Wooldridge, n.d.) Validity of assumption FE.3 is evident from the nature of obtained data explained in Section 3.

A sufficient condition for consistency of FE on the unbalanced panel is an extension of the usual strict exogeneity assumption:

$$E(u_{it}|\boldsymbol{x}_i, \boldsymbol{s}_i, c_i) = 0, \quad t = 1, ..., T$$

where $\boldsymbol{s}_i = (s_{i1}, ..., s_{i1})$. This condition allows s_{it} to depend on c_i in an unrestricted way, which means that the reason for missing data for certain country can be correlated with country-specific unobserved effects, but not with idiosyncratic error in any time period.

The last important assumption that I need to check is assumption of homoskedasticity (FE.5). I will use Likelihood-ratio test explained in Sanchez (2013) to test for presence of heteroscedasticity across panels. The test strongly indicates the presence of heteroskedasticity, which means that assumption FE.5 is violated.

Fortunately, neither assumption FE.5 nor FE.6 is needed for consistency and unbiasedness of FE estimators. In all regressions, the clustering method of obtaining robust standard errors is used to obtain the best possible results.

The above listed assumptions along with robust statistical inference justify the usage of FE, RE and CRE.

The fractional heteroskedastic probit model estimation requires strict exogenous covariates and ignorable selection. However, it is not advised to model serial correlation. Instead, a robust inference should be used. (Wooldridge, n.d.) No other

assumptions need not to be fulfilled, because the econometric model (3) was specially adjusted to the form that can be estimated and the specific method including adjusting for robust inference is incorporated in user-written program *fhetprob*. More information about the derivation of model (3) and about specifications of *fhetprob* program can be found in Wooldridge (2010b) and Bluhm (2013).

5 Interpretation of the results

5.1 FE and RE regression results

Table 3: adjNER: Fixed Effects and Random Effects, respectively

Variable	**Coef.**	**Std. Err.**	**Coef.**	**Std. Err.**
log($GNIpc$)	-0.005	(0.016)	-0.001	(0.011)
URBAN	0.555*	(0.252)	0.051	(0.066)
EDUexp	0.183	(0.356)	0.207	(0.306)
SANIT	-0.171	(0.199)	0.087	(0.068)
WATER	0.558**	(0.212)	0.433**	(0.141)
2000	0.006	(0.005)	0.007	(0.005)
2001	0.012*	(0.005)	0.015**	(0.005)
2002	0.017**	(0.006)	0.021**	(0.006)
2003	0.020**	(0.007)	0.024**	(0.007)
2004	0.021*	(0.009)	0.026**	(0.008)
2005	0.018	(0.011)	0.023 *	(0.009)
2006	0.024	(0.013)	0.029**	(0.010)
2007	0.026	(0.014)	0.032**	(0.010)
2008	0.029	(0.017)	0.036**	(0.012)
2009	0.028	(0.016)	0.035**	(0.012)
2010	0.030	(0.018)	0.039**	(0.013)
2011	0.028	(0.018)	0.037**	(0.013)
2012	0.030	(0.019)	0.041**	(0.014)
Intercept	0.243	(0.178)	0.398**	(0.083)
Obs.		1499		1499
R^2		0.909		0.505

$\rho < 0.05^*$, $\rho < 0.01^{**}$

Table (3) summarizes regression results from FE and RE estimation. For both FE and RE, the fully robust standard errors are reported in parentheses. The fully robust standard errors were used, because both serial correlation and heteroskedacity were detected, recall section 4.4. The indirect evidence of serial correlation was also indicated by comparing non-robust and robust versions of standard errors, where the latter were significantly larger. Nevertheless, non-robust version of standard errors is not reported, because the conclusions on significance of obtained estimators would not be reliable.

The results from CRE regression are not reported as well, because they provide the same results as FE. The CRE method was used to obtain fully robust regression-based Hausman test to determine, which of the two estimation methods, FE or RE, is more appropriate. The fully robust inference failed to reject the null hypothesis and hence provided a justification for using RE. Interestingly, the traditional Hausman

test, which compares the coefficients on the time-varying explanatory variables and computes a chi-square statistic ruled in favor of FE. However, this test does not enable to use robust standard errors, which is maybe a reason for this confusing results, even though I would normally expect the results to be the same with rejection being stronger in robust case. This indicates that there is either a very strong serial correlation or heteroskedasticity, which makes the traditional Hausman test using non-robust standard errors very inappropriate, or there is another issue with the data or estimation methods. Provided the information I have, I was not able to distinguish between FE and RE using statistical inference.

Now I will provide the discussion of results reported in Table (3) and try to decide between FE and RE using intuition. The most surprising result is that coefficients of $\log(GNIpc)$ are negative for both FE and RE. Using some data processing and intuition, I arrived at following conclusion: *Taking natural logarithm of GNIpc combined with time-demeaning makes data useless to estimate ceteris paribus effect on adjNER.* Taking natural logarithm diminishes differences between poorest and richest countries and time-demeaning is done separately by individual panels, which results in half values being negative and also completely removes differences, because GNI per capita usually increases or decreases slowly and hence the demeaned values would always be close to zero and it does not matter if GDP per capita in particular country is 100$ or 100000$. It implies that an *irrelevant* variable was included in regressions. Including irrelevant variable into regression can potentially have an undesirable effect on variances of estimators.

Therefore, I conducted the same set of three regressions (FE, RE, CRE) and presented the outcome in Table 4. So far, the conclusion is that GNI is not a suitable explanatory variable for regression analyses using wide range of countries. For meaningful analysis of this type with GNI or GDP as explanatory variables would require separation of countries into wide range of income levels. Separating to four levels of income as in Table 2 would still lead to confusing and more importantly, irrelevant results. However, the outcome of Table 2 regarding to GNIpc is very informative and can reveal information about the differences in primary school enrolments between different income groups.

Table 4: adjNER: Fixed Effects and Random Effects, respectively (2)

Variable	**Coef.**	**Std. Err.**	**Coef.**	**Std. Err.**
URBAN	0.548*	(0.245)	0.050	(0.060)
EDUexp	0.197	(0.365)	0.218	(0.312)
SANIT	-0.191	(0.196)	0.081	(0.067)
WATER	0.568**	(0.209)	0.440**	(0.139)
2000	0.006	(0.004)	0.007	(0.005)
2001	0.011*	(0.005)	0.015**	(0.005)
2002	0.017**	(0.006)	0.021**	(0.006)
2003	0.019*	(0.007)	0.024**	(0.007)
2004	0.019*	(0.008)	0.026**	(0.008)
2005	0.017†	(0.009)	0.023**	(0.009)
2006	0.021*	(0.010)	0.029**	(0.009)
2007	0.023*	(0.009)	0.031**	(0.009)
2008	0.026*	(0.011)	0.036**	(0.010)
2009	0.024*	(0.010)	0.035**	(0.009)
2010	0.027*	(0.011)	0.039**	(0.010)
2011	0.024*	(0.011)	0.037**	(0.010)
2012	0.026*	(0.012)	0.041**	(0.011)
Intercept	0.217	(0.152)	0.394**	(0.085)
Obs.		1518		1518
R^2		0.901		0.505

$\rho < 0.05^*$, $\rho < 0.01^{**}$

In further discussion, I will focus on interpreting the coefficients from Table 4, because they are more relevant, even if the differences are small.

Before I discuss signs and magnitudes of specific coefficients, I will firstly clarify the previous discussion about choosing between FE and RE. The situation is the same. The regression-based fully robust Hausman test found coefficients on time-averages to be even less significant and provided even stronger justification for RE. The traditional non-robust Hausman test, on the other hand, still voted in favor of FE estimation. Again, I am not able to say which method is better using statistically inference, even though there is a higher chance of RE being better.

When looking at differences between FE and RE estimators, there are a few important differences. That is consistent with previously mentioned non-robust Hausman test, where the null hypothesis stating that the differences in coefficients are not systematic has been strongly rejected, even though the test estimated FE estimation to be better.

All of the explanatory variables, excluding time dummies, along with explained variable are proportions. This means that the ceteris paribus effect can be interpreted as if it was log-log model (i.e. $\%\triangle y = \beta_1 \%\triangle x$).

For example, if urban population increases relative to total population by 1 percent (i.e. one percent of total population move from rural to urban location), the predicted increase in adjNER is about 0.6 percent for FE and 0.05 percent for RE. Both coefficients have expected signs, but only the FE estimators is significant. The difference in magnitude is huge, more than tenfold, but since RE estimator is not relevant, the further comparison of these two does not make much sense.

Another significant difference is the opposite sign on sanitation and even when neither of two estimators is significant, the RE coefficient has at least an expected sign.

Coefficients on education expenditure are similar in magnitude and also in size of standard errors, but neither is significant on standard significance levels.

Last explanatory variable proved to be the most significant predictor of adjusted net enrolment rates. Both are significant at the 1% level, with FE estimator being slightly higher in magnitude. When percentage of population with access to improved water source increases by 1%, the adjNER is expected to increase by 0.56% and 0.43% for FE and RE, respectively.

The coefficients on time period dummies are significant in both cases, especially in RE case, all coefficients, except for one, are statistically significant at the 1% level. This indicates that trends in improving adjNERs were steadily increasing over the reference period, even if not by significant amount.

The last important difference is in the coefficient of determination. The R-squared for FE is very high, even though I was able to obtain only two significant estimators out of four, when I do not count logarithm of GNIpc from the first two regressions.

To conclude, both FE and RE estimation methods provided similar results, with FE having two significant estimators out of four measured and RE having only one. The RE shows more significant and consistent time trend, but reports significantly lower goodness-of-fit measure.

5.2 Fractional heteroskedastic probit estimation results

Table 6 *(Appendix)* contains regression results with full set of included variables. Table 6 was moved to the *Appendix*, because (1) the coefficients reported in this table are scaled and cannot be interpreted directly and (2) it contains too many coefficients, which were needed for regression to be valid, but are not relevant for our analysis.

In this section, I will focus on interpreting *average marginal effects (AMEs)* of estimated independent variables. Table 5 summarizes these results:

Table 5: Average marginal effects

Variable	**dy/dx**	**Std. Err.**
URBAN	0.371*	(0.149)
EDUexp	0.140	(0.191)
SANIT	-0.072	(0.102)
WATER	0.021	(0.083)

$\rho < 0.05^{*}$, $\rho < 0.01^{**}$

The AMEs reported in this table are partial effects on a mean response, not probability, as in the classic probit case with binary dependent variable. (Wooldridge, 2010) The coefficients can be interpreted in a same way as in FE and RE case.

The only statistically significant effect is the effect of *URBAN* variable. Everything else equal, if urban population increases by 1% relative to total population, adjusted net enrolment rate is expected to increase by 0.4%, which is a little less than it was predicted in FE estimation, where the coefficient on urbanization was also found significant at the same significance level.

In both FE and RE regressions, coefficient on *WATER* was found to be very statistically significant. This conclusion no longer holds here. The estimated effect of increasing the proportion of population with access to improve water source does not seem to have almost any effect on *adjNER* and it is very statistically insignificant.

The other two variables (*EDUexp* and *SANIT*) were also found to be statistically insignificant, similar to FE and RE case.

Overall, this estimation justified the statistical significance of urbanization and provided some justification for FE estimation. It is worth noting that the inference is

again fully robust to violations of underlying assumptions and even if the estimation proved significance of only one variable, the conlusion is likely to be reliable.

To at least partially compare goodness-of-fit measures between FHETPROB, FE and RE, I computed a *pseudo R-squared* measure for FHETPROB. The obtained R-squared is equal to 0.599, which lies comfortably between R-squared measures obtained in FE and RE estimations. However, the comparison must be done carefully.

6 Conclusion

The analysis conducted in this thesis proved that trends in primary school enrolment can be explained on large scale. All regressions obtained at least one statistically significant variable using fully robust statistical inference. This may not seem as such an impressive result, but reader must consider several negative factors influencing the outcome of regressions.

First of all, as noted many times in this thesis, data resources for monitoring MDGs contain large amount of missing data. Therefore, I was not able to obtain sufficient amount of observations for most of the countries, resulting in a highly unbalanced panel dataset.

Moreover, this study used only country-level data. Majority of studies on school enrolment use data from various household surveys, which contain information on individuals. These studies typically contain factors such as distance to school, household income (or at least a suitable proxy), education of parents, or number of siblings. They often use shorter time periods, but have generally more observations. When aggregating the data from household surveys, many information are lost. Therefore I was not able to use data on indicators, which would be expected to better explain the variance in school enrolment rates.

Despite these major difficulties, I was able to obtain convincing results. Even though the first two (three, when adding CRE) are linear models, which are not theoretically designed to fractional dependent variables, they proved to be suitable for the analysis and majority of their fitted values lied comfortably between 0 and 1, or exceeded value 1 (100% net enrolment) only by minimal amount. For RE, only about 5% of predicted fitted values exceeded value 1, with 1.02 being the largest value. For FE, the situation was a little bit worse, where about 30% of fitted values exceeded 1, but only about 6.5% exceeded 1.1 with maximal value being 1.168. Overall, I was not able to conclude with certainty, which of the two methods, FE or RE, was better, because of different results of performed tests, indirect justifications derived from the signs and magnitudes of estimators and results of estimation of fractional probit model.

The usage of fractional heteroskedastic probit model provided a completely different method of estimation. The goal was to compare the results of these three types

of estimation and determine, which model provides the best results. Fractional heteroskedastic method of estimation arrived at slightly different conclusions about the effect of population having access to improved water source, which was very statistically significant in both FE and RE. On the other hand, it confirmed the relevance of urbanization, which was found statistically significant only in FE estimation, thereby provided some justification for FE method.

The analysis proved that country-level regression can lead to reliable results about development indicators, but more detailed analysis would be needed to arrive to such conclusions, which would help policy makers to establish country-specific policies, especially needed for the countries who will not manage to fulfill the MDGs by the end of 2015.

References

[1] Huisman, J., Smits, J. (2009). Effects of Household- and District-Level Factors on Primary School Enrollment in 30 Developing Countries. World Development, 37(1), 179-193. doi:10.1016/j.worlddev.2008.01.007

[2] Mutangadura, G., Lamb, V. (2003). Variations in rates of primary school access and enrolments in sub-Saharan Africa: a pooled cross-country time series analysis. International Journal Of Educational Development, 23(4), 369-380. doi:10.1016/s0738-0593(02)00060-3

[3] Richards, J., Vining, A. (2015). Universal primary education in low-income countries: The contributing role of national governance. International Journal Of Educational Development, 40, 174-182. doi:10.1016/j.ijedudev.2014.09.004

[4] Filmer, D., Hasan, A., Pritchett, L. (2006). A Millennium Learning Goal: Measuring Real Progress in Education. SSRN Electronic Journal. doi:10.2139/ssrn.982968

[5] Wooldridge, J. (2010). Econometric analysis of cross section and panel data. Cambridge, Mass.: MIT Press.

[6] Wooldridge, J. (2012). Introductory econometrics. Mason, Ohio: South-Western Cengage Learning.

[7] Drukker, D. (2003). Testing for serial correlation in linear panel-data models. The Stata Journal, 3(2). Retrieved from http://www.stata-journal.com/article.html?article=st0039

[8] Sanchez, G. (2012). Fitting Panel Data Linear Models in Stata. Presentation.

[9] Bluhm, R. (2013). fhetprob: A fast QMLE Stata routine for fractional probit models with multiplicative heteroskedasticity. Retrieved from http://www.richard-bluhm.com/wp-content/uploads/2013/02/fhetprob.pdf

[10] Wooldridge, J. (2015). Correlated Random Effects Panel Data Models. Presentation, Michigan State University.

[11] Wooldridge, J. (2013). Correlated Random Effects Panel Data Models. Presentation, Michigan State University.

[12] Wooldridge, J. Fractional Response Models with Endogenous Explanatory Variables and Heterogeneity. Presentation, Michigan State University.

[13] Stata.com,. (2015). Stata | FAQ: Testing for panel-level heteroskedasticity and autocorrelation. Retrieved 31 July 2015, from http://www.stata.com/support/faqs/statistics/panel-level-heteroskedasticity-and-autocorrelation/

[14] UNITED NATIONS,. (2015). The Millennium Development Goals Report. New York.

[15] UNESCO,. (2015). EDUCATION FOR ALL: achievements and challenges. Retrieved from http://unesdoc.unesco.org/images/0023/002322/232205e.pdf

[16] UNESCO,. . Education counts: Towards the Millennium Development Goals. Retrieved from http://unesdoc.unesco.org/images/0019/001902/190214e.pdf

[17] Un.org,. (2015). United Nations Millennium Development Goals. Retrieved 31 July 2015, from http://www.un.org/millenniumgoals/

[18] Oxford Martin School University of Oxford,. (2013). Now for the Long Term: The Report of the Oxford Martin Commission for Future Generations.

[19] Mdgs.un.org,. (2015). unstats | Millennium Indicators. Retrieved 31 July 2015, from http://mdgs.un.org/unsd/mdg/

[20] Data.worldbank.org,. (2015). Data | The World Bank. Retrieved 31 July 2015, from http://data.worldbank.org/

[21] Uis.unesco.org,. (2015). UNESCO Institute for Statistics: UNESCO Institute for Statistics. Retrieved 31 July 2015, from http://www.uis.unesco.org/Pages/default.aspx

[22] Unmillenniumproject.org,. (2015). UN Millennium Project | About the MDGs. Retrieved 31 July 2015, from http://www.unmillenniumproject.org/goals/

[23] Endpoverty2015.org,. (2015). End Poverty 2015 | We are the generation that can end poverty. Retrieved 31 July 2015, from http://www.endpoverty2015.org/

List of Tables

Appendix

Table 6: Estimation results : fhetprob (1st part)

Variable	**Coefficient**	**(Std. Err.)**
Equation 1 : adjNER		
URBAN	2.832	(1.222)
EDUexp	1.067	(1.454)
SANIT	-0.549	(0.787)
WATER	0.157	(0.647)
URBANb	-1.990	(1.271)
EDUexpb	1.131	(3.412)
SANITb	2.005	(0.833)
WATERb	0.406	(0.908)
2000	0.014	(0.017)
2001	0.066	(0.028)
2002	0.098	(0.036)
2003	0.118	(0.038)
2004	0.133	(0.043)
2005	0.120	(0.048)
2006	0.159	(0.055)
2007	0.168	(0.058)
2008	0.224	(0.074)
2009	0.218	(0.072)
2010	0.235	(0.080)
2011	0.246	(0.081)
2012	0.236	(0.084)
2000b	-0.103	(1.327)
2001b	-3.279	(2.115)
2002b	-1.485	(5.592)
2003b	-0.817	(0.719)
2004b	-9.146	(3.048)
2005b	-3.703	(3.316)
2006b	-0.389	(6.460)
2007b	2.425	(2.301)
2008b	-2.529	(2.447)
2009b	-7.240	(2.613)
2010b	-4.844	(2.093)
2011b	-4.986	(3.013)
2012b	0.838	(2.700)

Table 7: Estimation results : fhetprob (2nd part)

Variable	Coefficient	(Std. Err.)
n2	-67829.974	(32015.778)
n3	27261.042	(7746.050)
n4	-13888.656	(23273.359)
n5	18109.149	(2731.504)
n6	-0.980	(1.450)
n7	13.295	(5.426)
n8	-12957.350	(1463.801)
n9	-5.526	(2.410)
n10	1.067	(1.191)
n11	1.204	(2.262)
n12	0.540	(0.541)
n13	1465.783	(729.092)
n2URBANb	-41608.665	(58933.664)
n3URBANb	-12973.436	(11223.355)
n4URBANb	-9785.752	(35947.191)
n5URBANb	-23729.167	(12693.077)
n6URBANb	-1.583	(0.766)
n7URBANb	-1.035	(0.533)
n8URBANb	2612.490	(2348.283)
n9URBANb	-3.814	(1.061)
n10URBANb	-1.087	(1.684)
n11URBANb	-0.737	(0.467)
n12URBANb	-1.747	(0.946)
n13URBANb	-711.193	(2030.258)
n2EDUexpb	136007.238	(924557.909)
n3EDUexpb	124904.913	(145842.063)
n4EDUexpb	213890.701	(464427.489)
n5EDUexpb	-129643.798	(34318.878)
n6EDUexpb	-9.028	(3.772)
n7EDUexpb	0.237	(25.207)
n8EDUexpb	184889.258	(53679.923)
n9EDUexpb	25.618	(13.582)
n10EDUexpb	-23.561	(27.475)
n11EDUexpb	-8.372	(21.150)
n12EDUexpb	-26.412	(20.190)
n13EDUexpb	34455.937	(18585.600)
n2SANITb	5723.426	(23088.165)
n3SANITb	53872.652	(9218.823)
n4SANITb	25474.241	(13582.138)
n5SANITb	32241.366	(6909.857)
n6SANITb	-0.106	(0.851)
n7SANITb	4.488	(2.656)

Table 8: Estimation results : fhetprob (3rd part)

Variable	**Coefficient**	**(Std. Err.)**
n8SANITb	-7344.647	(3383.901)
n9SANITb	-4.543	(1.239)
n10SANITb	-1.436	(1.535)
n11SANITb	-0.021	(2.367)
n12SANITb	-1.359	(1.289)
n13SANITb	6724.888	(1185.808)
n2WATERb	146133.569	(107550.950)
n3WATERb	-32257.675	(7591.301)
n4WATERb	66846.397	(32561.033)
n5WATERb	-12933.293	(15001.218)
n6WATERb	1.928	(3.043)
n7WATERb	-17.370	(7.056)
n8WATERb	24266.555	(3477.300)
n9WATERb	10.646	(4.035)
n10WATERb	0.516	(1.458)
n11WATERb	-1.227	(3.404)
n12WATERb	2.959	(3.115)
n13WATERb	-2281.168	(2489.726)
Intercept	1.746	(1.769)
Equation 2 : lnsigma2		
n2	10.314	(0.090)
n3	9.958	(0.046)
n4	10.835	(0.066)
n5	9.559	(0.129)
n6	-0.336	(0.730)
n7	-0.079	(0.573)
n8	8.994	(0.047)
n9	-1.079	(0.329)
n10	-0.542	(0.537)
n11	-0.849	(0.330)
n12	0.146	(1.130)
n13	8.201	(0.050)

N	1514
Log-likelihood	-397.5
pseudo R-squared	0.599