



HAL
open science

Human capital, public debt and long-term economic growth

Maya Murched

► **To cite this version:**

Maya Murched. Human capital, public debt and long-term economic growth. Economics and Finance. Université d'Angers, 2016. English. NNT : 2016ANGE0003 . tel-01461620

HAL Id: tel-01461620

<https://theses.hal.science/tel-01461620>

Submitted on 15 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de Doctorat

Maya MURCHED

*Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université d'Angers
sous le label de L'Université Nantes Angers Le Mans*

École doctorale : DEGEST (Droit, Economie, Gestion, Environnement, Société et Territoires)

Discipline : Sciences Economiques

Unité de recherche : UMR GRANEM, Université d'Angers, Agrocampus Ouest centre d'Angers

Soutenue le **15/01/2016**
Thèse N° :

Capital Humain, Dette Publique et Croissance Economique à Long Terme

JURY

Rapporteurs : **Fredj JAWADI**, Maître de conférences HDR, Université d'Evry Val d'Essonne
Shyama RAMANI, Professeure, United Nations University at Maastricht

Examineur : **Muriel TRAVERS**, Maître de conférences HDR, Université d'Angers (Présidente du jury)

Directeur de Thèse : **Walid OUESLATI**, Professeur, AGROCAMBUS OUEST centre d'Angers

Co-directeur de Thèse : **Damien ROUSSELIERE**, Professeur, AGROCAMBUS OUEST centre d'Angers

Thèse de Doctorat

Maya MURCHED

*Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université d'Angers
sous le label de L'Université Nantes Angers Le Mans*

École doctorale : DEGEST (Droit, Economie, Gestion, Environnement, Société et Territoires)

Discipline : Sciences Economiques

Unité de recherche : UMR GRANEM, Université d'Angers, Agrocampus Ouest centre d'Angers

Soutenu le

Thèse N° :

Capital Humain, Dette Publique et Croissance Economique à Long Terme

JURY

Rapporteurs : **Fredj JAWADI**, Maître de conférences HDR, Université d'Evry Val d'Essonne
Shyama RAMANI, Professeure, United Nations University at Maastricht

Examineur : **Muriel TRAVERS**, Maître de conférences HDR, Université d'Angers (Présidente du jury)

Directeur de Thèse : **Walid OUESLATI**, Professeur, AGROCAMPUS OUEST centre d'Angers

Co-directeur de Thèse : **Damien ROUSSELIERE**, Professeur, AGROCAMPUS OUEST centre d'Angers

Preface

This thesis was carried out in GRANEM, AGROCAMPUS OUEST centre d'Angers, University of Angers and financed by the Syrian Government.

I would like at the end of this work to address special thanks to my supervisors, Prof. Walid Oueslati and Damien Rousselière for the patient guidance and encouragement. Their guidance helped me in all the time of research and writing of this thesis.

I sincerely thank all the members of the jury, in particular, Professor Muriel Travers for agreeing to chair the jury of my thesis, Professor Shyama Ramani and Professor Fredj Jawadi for generously agreeing to review my thesis. I also thank Mr Mouezz Fodah and Mr Pierre-André Jouvét for agreeing to be a member of thesis monitoring committee for the four years of my thesis preparation, and for their careful reports and for all exchanges, discussions and useful interactions we had during the course of my thesis and its validation. I would like to thank also Madame Gemma Davis, research assistant for her editorial assistance.

My first and infinite thanks to the soul of my father Ahmad, who could not resist his enormous illness to be with me this day. I would thank him there, and wish him all the peace. I would like to thank also infinitely my mother Mouna Youssef to her courage and support.

I must express my gratitude to Wissam, my scintillating husband, for his continued support and encouragement despite the thousands of kilometers that separate us. He was always there cheering me up and stood by me through the good and bad times. Special thanks to my little sweet princess Mia for her presence in my life.

I can not forget the enormous help, support and love I had and still receive from my gentile friend Nathalie and her husband Benoit Chevillard.

I wish to warmly thank my big family and friends : my parents in law Oyma and Adnan Chahin, my sisters Maha and Haneen Murched, my sisters in law Zainab and

Baraa Chahin, my brothers in law Rami Kazak and Somer Chahin, my grand mother Ghania, my respected aunt Ramia Youssef and her husband Bassel Mahmmoud, my friends Houssam,Hiba, Aida, Eiman,... and Rabab for their persistent support.

Last but not the least, I would like to thank my all my colleagues, friends and members of the lab, especially Thomas Coisnon, Marouen Mbarek, Marjorie Tendero, Amal Jemai for their assistance and support.

Special thanks go to for their unlimited support during and after this thesis.

My great thanks to syrian government for offering me the opportunity to be here and prepare this thesis, to my homeland **SYRIA** God bless her.

Abstract

Economic growth and its driving forces have been the main topic preoccupying economic researchers since long time in macroeconomic branch. Public investment in human capital through educational system plays an ultimate role in boosting economic growth and development, this role has taken a place since the innovative dawn of endogenous growth theory. The focus and efforts of investing in human capital could be destabilized by the global and recent return of sovereign debt crisis in several countries, which continues its rise since the early 2007, and the after-crisis necessary adjusting policies. Getting back wise policies should be composed of mixture of growth fostering activities, including the investment in human capital, austerity and forbearance. The main purpose of this thesis is to provide new empirical inferences on debt-growth relationship and its interaction with human capital formation. Estimates are carried on a recent and complete data set that spans over 22 years and involves 76 countries worldwide. The range of invested variables encompasses many macroeconomic aggregates such as : GDP annual growth rates, public debt to GDP ratio, and public education expenditure to GDP ratio, average schooling years, inflation rate, and others. Using a superior estimation semi-parametric technic which accounts for some data issues, the empirical results suggest a heterogeneous impact of public debt and education expenditures levels on GDP growth rates. Henceforth, the use of education expenditure in the whole sample is inefficient, where policy makers should adjust and well manage the function of these expenditure in line with the public efforts to reduce debt high levels and rise economic growth. We also show that the use of textual analysis tools in economic studies, such in growth literature, offers a rapid and total lecture of the hidden research trends embodied in the huge empirical and theoretical literature of economic growth.

Keywords : Economic growth, human capital, public debt, education expenditure, efficiency, multiple imputation, quantile regression, panel data

Résumé

La croissance économique et ses moteurs représentent le principal sujet préoccupant les chercheurs en macroéconomie depuis longtemps. Investir en capital humain à travers le système éducatif joue un rôle important pour stimuler la croissance et le développement économique, cette accentuation a pris place depuis la naissance innovante de la théorie de la croissance endogène. L'attention et les efforts dévoués à l'investissement dans le capital humain peuvent être déstabilisés par le retour global et récent de la crise de la dette souveraine dans plusieurs pays, dette qui poursuit son ascension depuis 2007, et les politiques d'ajustement nécessaires d'après-crise. Des judicieuses politiques de redressement devraient être composées d'un mélange des activités encourageant la croissance économique, y compris l'investissement dans le capital humain, l'austérité et le long terme. L'objectif principal de cette thèse est de fournir des nouvelles évidences empiriques sur la relation dette-croissance économique et leurs externalités sur la formation de capital humain, les estimations sont réalisées sur un jeu de données récent et complet couvrant 22 années et 76 countries dans le monde. L'ensemble des variables utilisées englobe de nombreux agrégats macroéconomiques tel que : taux de croissance annuel du PIB, la dette publique en % de PIB, les dépenses publiques d'éducation en % de PIB, le moyen d'année de scolarité, le taux d'inflation, et d'autres. En utilisant une technique d'estimation semi-paramétrique appropriée qui offre des solutions pour de nombreux problèmes concernant les données, les résultats empiriques suggèrent un impact négatif et hétérogène de dette et des dépenses d'éducation publiques sur la croissance du PIB. Là où, l'utilisation des dépenses d'éducation dans l'ensemble de l'échantillon est inefficace, les décideurs politiques devraient ajuster et bien gérer la fonction de ces dépenses en même temps de viser des efforts publics pour réduire les niveaux élevés d'endettement et d'augmenter la croissance économique. Nous montrons également que l'utilisation des outils d'analyse textuelle en économie, offre une lecture rapide et globale des courants de recherche contenus dans la littérature empirique et théorique de la croissance économique.

Mots clés : Croissance économique, capital humain, dette publique, dépenses d'éduca-

tion, efficacité, imputation multiple, regression par quantile.

Contents

Preface	3
Abstract	5
Résumé	7
Liste des Acronymes	17
1 General Introduction	19
1.1 Problem statement	25
1.1.1 Historical background: debt and crisis	25
1.1.2 Stylized facts: debt and education expenditure	29
1.1.3 Growth trajectory heterogeneity	31
1.1.4 Growth determinants uncertainty	32
1.1.5 Thesis research questions	32
1.2 Methodology	33
1.2.1 Model uncertainty: textual analysis & variables selection	33
1.2.2 Empirical growth framework	39
1.2.3 Estimation method: quantile regression	40
1.3 Data	47
1.3.1 Textual data set	48
1.3.2 Quantitative data set and descriptive statistics	49
1.3.3 Missing data issue	52
1.3.4 How to deal with the problem of missing data?	54
1.4 Thesis outlines	63
2 Identifying Economic Growth Schools: The Case of Human Capital	71
2.1 Introduction	72
2.2 Economic growth research streams	74

2.3	Methodology: textual analysis tools	77
2.3.1	Collecting data procedure and analysis factors	79
2.3.2	Corpus data statistics	81
2.3.3	Primary corpus description	82
2.4	Results and Discussion	84
2.4.1	Human capital context in growth literature	93
2.5	Conclusion and limitations	94
3	Estimating the heterogeneous effect of Education Expenditure and Public Debt on Economic Growth: A quantile regression with multiple imputation approach	99
3.1	Introduction	100
3.2	Debt-growth: endogeneity issue & proposed solutions	103
3.3	Data	104
3.3.1	Data sources	104
3.3.2	Missing data: multiple imputation solution	106
3.4	Empirical strategy: model & estimation methods	107
3.4.1	Baseline model	107
3.4.2	Econometric approach: the semi-parametric method of QR	109
3.5	Results and discussion	111
3.5.1	Robustness tests	116
3.6	Conclusion	123
4	Growth and Education Expenditure Efficiency: A Two Stage Longitudinal Quantile Regression Approach	125
4.1	Introduction	126
4.2	Methodological strategy	129
4.2.1	Limits of traditional efficiency evaluation methods	130
4.2.2	Quantile regression approach to estimate economy efficiency	132
4.3	Data	134
4.4	Efficiency estimation	135
4.4.1	First stage: QR estimated efficiency score	135
4.4.2	Second stage analysis: Bayesian model averaging	138
4.5	Conclusion and policy implications	146
5	Summary and General Discussion	149
5.1	Limitations and future perspectives	155
	References	157

Appendices	177
A Introduction Chapter Appendix	177
A.1 IRaMuteQ Description	177
A.2 R-environment	178
A.3 R packages for multiple imputation	179
A.4 Imputation Codes	180
A.4.1 Imputation code, m=100	180
A.4.2 Imputation code, m=5	183
A.5 Introduction Chapter Figures	188
B Chapter 2 Appendix	217
B.1 Chapter 2 Tables	217
B.2 Chapter 2 Figures	221
C Chapter 3 Appendix	225
C.1 Chapter 3 Tables	225
D Chapter 4 Appendix	233
D.1 Chapter 4 Tables	233

List of Figures

1.1	Gross Central Government Debt as a Percentage of GDP	28
1.2	GDP growth rate, and public debt as % of GDP average	30
1.3	GDP growth rate, and public education expenditure as % of GDP	30
1.4	Multiple Imputation publications (log) between 1977 and 2010	57
1.5	MI schematic (Honaker et al., 2011)	60
1.6	Densities Diagnostic I, m=100	64
1.7	Densities Diagnostic II, m=100	65
1.8	Densities Diagnostic III, m=100	66
1.9	Overimputation Diagnostic, m=100	67
2.1	Correspondence analysis of active forms, and factors	85
2.2	Hierarchical Descending Classes:Classification Dendrogram	87
2.3	Publishing year's χ^2 by class	88
2.4	Publishing journal's χ^2 by class	89
2.5	Citation number's χ^2 by class	90
2.6	Human Capital Words Graph	94
3.1	Comparison of parameters estimates I, PFE-QR, TFE and SFA	120
3.2	Comparison of parameters estimates II, PFE-QR, TFE and SFA	121
4.1	Rural population in % of total population effect on efficiency score	142
4.2	Student per teacher effect on efficiency score	143
4.3	Countries fixed effects	144
A.1	OECD and Non-OECD growth rate	188
A.2	OECD and Non-OECD growth rate II	189
A.3	Densities Diagnostic IV, m=100	190
A.4	Densities Diagnostic V, m=100	191
A.5	Densities Diagnostic VI, m=100	192
A.6	Densities Diagnostic VII, m=100	193

A.7	Densities Diagnostic VIII, m=100	194
A.8	Densities Diagnostic IX, m=100	195
A.9	Densities Diagnostic X, m=100	196
A.10	Overimputation Diagnostic II, m=100	197
A.11	Overimputation Diagnostic III, m=100	198
A.12	Overimputation Diagnostic IV, m=100	199
A.13	Overimputation Diagnostic V, m=100	200
A.14	Densities Diagnostic I, m=5	201
A.15	Densities Diagnostic II, m=5	202
A.16	Densities Diagnostic III, m=5	203
A.17	Densities Diagnostic IV, m=5	204
A.18	Densities Diagnostic V, m=5	205
A.19	Densities Diagnostic VI, m=5	206
A.20	Densities Diagnostic VII, m=5	207
A.21	Densities Diagnostic VIII, m=5	208
A.22	Densities Diagnostic IX, m=5	209
A.23	Densities Diagnostic X, m=5	210
A.24	Densities Diagnostic XI, m=5	211
A.25	Densities Diagnostic XII, m=5	212
A.26	Densities Diagnostic XIII, m=5	213
A.27	Overimputation Diagnostic I, m=5	213
A.28	Overimputation Diagnostic II, m=5	214
A.29	Overimputation Diagnostic III, m=5	214
A.30	Overimputation Diagnostic IV, m=5	215
A.31	Overimputation Diagnostic V, m=5	215
A.32	Overimputation Diagnostic VI, m=5	216
A.33	Overimputation Diagnostic VII, m=5	216
B.1	Class1: Graph of words: class 1	221
B.2	Class2: Graph of words: class 2	222
B.3	Class3: Graph of words: class 3	222
B.4	Class4: Graph of words: class 4	223
B.5	Class5: Graph of words: class 5	223
B.6	Class6: Graph of words: class 6	224

List of Tables

1.1	Growth determinants in growth literature	36
1.2	Articles' Summary	49
1.3	Variables' sources	50
1.4	Descriptive Statistics	51
1.5	Missing data per year, variable and country	53
2.1	Journal Ranking by Impact Factor	80
2.2	Time's Division	81
2.3	Articles' Summary	82
2.4	Journals order by article citation	82
2.5	Corpus Description	83
2.6	Active Forms Frequency	84
2.7	Classes Profiles	86
2.8	Classes' labels	92
3.1	Variables selection	105
3.2	Penalized quantile regression with fixed effects estimates (Benchmark)	112
3.3	Penalized quantile regression with fixed effects estimates (Model C)	113
3.4	Standardized Coefficients of some variables	116
3.5	Model Specification Tests	117
3.6	SFA & TFE Panel data model estimation	118
3.7	Quantile regression with correlated random effects, Model C	122
4.1	Efficiency Estimation	136
4.2	Public Education Expenditure fraction and coefficients	137
4.3	BMA estimates	141
5.1	Thesis contributions	151
B.1	Articles' authors & titles I	218

B.2	Articles' authors & titles II	219
B.3	Articles' List	220
B.4	Journals abbreviation	221
C.1	Country classification	226
C.2	Variables Description I	227
C.3	Variables Description II	228
C.4	Penalized quantile regression with fixed effect estimates, Model A	229
C.5	Penalized quantile regression with fixed effect estimates, Model B	229
C.6	Panel data model estimation: TFE	230
C.7	Quantile regression with correlated random effects, Model C	232
D.1	Countries by quantiles	234
D.2	Education and health expenditure in % of GDP	236
D.3	Quantile regression estimation	236
D.4	Some variables averages by quantiles	237
D.5	BMA estimates (Imputation 1)	239
D.6	BMA estimates (Imputation 2)	241
D.7	BMA estimates (Imputation 3)	244
D.8	BMA estimates (Imputation 4)	246
D.9	BMA estimates (Imputation 5)	248
D.10	Student per teacher ratio & Rural population levels by development level	248

List of Acronyms

APA	-	A merica, P sychological A ssociation
BACE	-	B ayesian A veraging of C lasical E stimates
BMA	-	B ayesian M odel A veraging
BOCF	-	B aseline O bservation C arried F orward
CA	-	C orrespondence A nalysis
CPU	-	C entral P rocessing U nits
CRE-QR	-	Q uantile R egression with C orrelated R andom E rrors
DEA	-	D ata E nvelopment A nalysis
EBA	-	E xtrême B ound A nalysis (E B A)
EMB	-	E xpectation- M aximization with B ootstrapping
Emis	-	E xpectation- M aximization importance sampling
EU	-	E uropean U nion
FCS	-	F ully C onditional S pecification
FDH	-	F ree D isposal H ull
FE	-	F ixed E ffects
FIML	-	F ull I nformation M aximum L ikelihood
FTDA	-	F unctional T extual D ata A nalysis
GDP	-	G ross D omestic P roduct
GETS	-	G eneral T o S pecific modelling
GMM	-	G eneralized M ethod of M oments
HDC	-	H ierarchical D escending C lassification
HDI	-	H uman D evelopment I ndex
IP	-	I mputation P osterior
LR	-	L ikelihood R atio
LS	-	L istwise D eletion
LOWESS	-	L Ocally W eighted S catter plot S moothy
LOCF	-	L ast O bservation C arried F orward
LA	-	L exicographic A nalysis

List of Abbreviations

MAR	-	Missing At Random
MBCC	-	Model-Based Curve Clustering
MCAR	-	Missing Completely At Random
MCMC	-	Markov Chain- Monte Carlo
MI	-	Multiple Imputation
MICE	-	Multivariate Imputation by Chained Equations
MNAR	-	Missing Not At Random
MS	-	Mean Substitution/mean imputation
NEG	-	New Economic Geography theory
OLS	-	Ordinary Least Squares
OECD	-	Organisation for Economic Co-operation and Development
PFE-QR	-	Quantile Regression with Penalized Fixed Effects
PWT	-	Penn World Table
QR	-	Quantile Regression
RE	-	Relative Efficiency
SFA	-	Stochastic Frontier Analysis
SGP	-	Stability Growth Pact
TA	-	Textual Analysis
TFE	-	Time Fixed Effects
UNDP	-	United Nation Development Program
VIF	-	Variable Inflation Factor
WEO	-	World Economic Outlook database
WDI	-	World Development Indicators
WGI	-	World Governance Indicators

Chapter 1

General Introduction

Nations economic growth performance is the core subject that has interested economic researchers around the globe for many years. Through this subject, researchers have insisted on the important role that human capital plays in developed as well as in developing markets. The recognition of this role came at the end of the last century, henceforth human capital is considered as a leading force of sustainable per capita growth in endogenous, and neoclassical growth theories (Lucas Jr, 1988), (Uzawa, 1965).

Many theoretical growth models were proposed before the development of the concept of *Human Capital* in growth context. The most important work was that of Solow (1956) which is based on the basic growth function edited by Cobb and Douglas (1928) where economic growth motors are exogenous: *saving rate, physical capital depreciation rate and population growth rate*. Solow neoclassical growth model represents the starting point of conventional economic growth theorization. Model assumptions were: constant returns to scale, marginal productivity of capital, exogenous technical progress, a substitutable capital and labor.

Solow's model was and still is used to study differences in cross-countries growth rates which, according to him, come from two channels: either from differences in *total factor productivity (TFP)* or from the different levels of *capital per worker*. According to this, a country could be on a sustainable growth path only if it invests in research&development and education. Lucas's Lucas Jr (1988) contributions add to economic literature enrichment, while trying to find suitable mechanics for studying economic development, he states that there are two kinds of capital in growth production: the first is *physical capital* in form of buildings, machines, etc., which is accumulated and utilized under a familiar neoclassical technology, while the second is the *human capital* in form of skilled and educated workers that enhance productivity. Lucas's con-

tribution highlights the importance of education and human capital skills on worker outputs. Only two years before Lucas's contributions, [Romer \(1986\)](#) argues that increasing growth rates are feasible since knowledge, that has increasing marginal productivity, is one of two state variable in a growth model with endogenous technology. Both, Romer's and Lucas's contributions have consolidated the birth of endogenous economic growth theory which supposes an increasing return to capital once new accumulation factors are included such as knowledge and innovation which are embodied in Human capital.

Later theoretical development have been empirically tested by some economists such as [Mankiw et al. \(1992b\)](#) whose results emphasize the role of human capital in economic growth under the form of educational attainment, and confirm that Solow's human capital augmented model provides an excellent description of cross-country data and that education policies differences explain cross-countries growth differences. In more recent growth accounting prospect, [Mäkeläinen \(2010\)](#) argues that increasing education level for labor force rises worker supply which in turn increases outputs levels.

Thus, endogenous growth theory amongst others demonstrates that human capital is a very important input in the economy production process. Any investment in human capital, either made by the individual through schooling or by the firms on the job training, would have positive results on growth rates, this represents the typical feature within the endogenous growth literature ([Zotteri, 2002](#)). Denoted firstly by [Schultz \(1988\)](#), that human capital creates economies of scale, which implies that further investment to increase human capital accumulation will lead to enhanced economic growth rates. Interested in studying human capital, Schultz states that an important source to produce and improve human capital quality, amongst others, is the educational system where education provides more flexible skills, highly educated workers that are more able to produce and innovate using the actual technological level, ([Zotteri, 2002](#)).

Thus, increasing human capital level would affect productivity growth in two ways: firstly, highly educated individuals work to increase the stock of knowledge through the development of new process and technologies. Secondly, education affects economic growth through the diffusion and transmission of knowledge, an educated worker helps his colleague to be more productive by creating learning opportunities via social and professional interaction ([Mäkeläinen, 2010](#)). Moreover, externalities of investment in human capital can be a helpful factor in explaining the differences between countries in per capita income. Human capital has many interesting characteristics, [Crawford \(1991\)](#) identifies some of them as follows:

-
1. Human capital is expandable and self generating which is linked to the possibility that the stock of knowledge increases individual human capital.
 2. The transportable and shareable characteristics mean that the holder of knowledge shares it and distributes it with others.

Building human capital endures all lifetime through different procedures. Human capital theory developed by [Schultz \(1961\)](#) suggests two forms of human capital formation: the first form is the general formation, where humans are usually qualified by diploma and do not have a specialized qualification. While the second, is the specific formation where individuals have a specialized experience, usually made during work time, and their professional experience is measured by age. A general formation is acquired in the educational system through the compulsory education path, whereas more specific formations start after school time where individuals have more possibilities. The formation process needs some input: time sacrifice by individual, physical capital represented by equipment and endowments necessary to compile the formation mission, human capital it self and finally the necessary funds to finance the construction of physical capital and a guarantee formation process function. Funds come from two sources either public or private agents or both. Models like in [Uzawa \(1965\)](#) , [Lucas Jr \(1988\)](#) and [Rebelo \(1991\)](#) do not consider public spending as an input to format human capital. In counterparts, for others like [Beauchemin \(2001\)](#) and [Blankenau and Simpson \(2004\)](#) public agent participates in human capital formation by allocating the necessary public funds, implying that education is an investment field through which governments gain access to control public objectives in human capital formation and accumulation.

Education and health expenditures represent a high portion of public expenditure ([Afonso and St Aubyn, 2004](#)). Therefore, the focus on estimating their efficiency has increased. For example [Gupta et al. \(2002\)](#) use cross sectional data for (50) countries, while [Han and Miao \(2010\)](#) analyze local health expenditure efficiencies using the two stage framework of a **DEA**-Tobit model based on the panel data for (31) provinces in China from 1997 to 2007. In another study of [Gupta and Verhoeven \(2001\)](#), authors assessed government spending on education and health in (37) countries in Africa from 1984 to 1995. This study indicates that different countries experience different efficiency levels depending on regional factors. In order to estimate expenditure efficiency, several measuring procedures have been examined in econometric literature. In the work [Afonso and St Aubyn \(2004\)](#), authors have measured the efficiency in education and health sectors in **OECD** countries using **DEA**. In another work [Herrera and Pang \(2005\)](#), authors have also examined the efficiency of government spending in health education sectors in developing countries, as did [Prasetyo and Zuhdi \(2013\)](#) in assessing the government expenditure efficiency towards the human development using **HDI**.

To finance education expenditures governments may use either tax revenue or public deficit or both of them. Being one of infrastructure expenditures, education expenditure, financing them through public deficit will not have a distortion impact on economic environment, since future formed individuals will pay the outstanding debt and increase the long-term growth rates, (Turnovsky and Fisher, 1995). Many reasons can push governments in advanced as well as advancing countries to increase the stock of its debt to finance public investment or any kind of activities that contribute to a broader economic and socially useful purposes such as education and health. Some of these reasons might be either a weak national income, an unusual events such war and natural disasters (Daniel et al., 2003), government desire to avoid market distortion created by raising taxes level, (Barro, 1979), or for the controversial Keynesian countercyclical fiscal policy characterized by a recourse to deficit financing, named also short-run fiscal activism, (Dwyer, 2011)¹.

Amongst the other financial instruments, the use of public debt in financing economic activities has received much attention, lately renewed with the return of financial crisis all around the world. Debt-growth correlation has been explored theoretically as well as empirically. From the theoretical perspective, debt-growth correlation is rather characterized negatively (Greiner, 2013). In an infinite individual life economic growth model, Turnovsky (2000) states that using public deficit to finance public spending have an increasing impact on the long-run growth rate. This increasing impact comes from the fact that public deficit does not have any distortion impact in the model, where individual life is infinite. Once the government has stuck to their budgetary rules, growth rates created by the use of public deficit in financing public investment becomes smaller (Greiner and Semmler, 2000). Coming back to reality with finite life individuals, an endogenous growth model with productive public spending and limited public debt has been developed by Futagami et al. (2008). The authors found that when public debt to GDP ratio is the smaller, the balanced growth rate is the higher. With public capital and public debt Greiner (2008a) presents another endogenous growth model, where government finances productive and unproductive public spending through income taxation and public deficit, and assumes that budget surplus to GDP is a positive function of debt to GDP². Greiner's main insight was that, when debt to GDP is zero the balanced growth rate is higher compared to a situation with positive debt ratio. The same negative debt-growth correlation has been confirmed in a recent contribution of

1. Where government uses further budget deficit, cuts taxes and increases spending volume during recession. The growth in public expenditure is sufficient to maintain full employment and debt is subject to sustainability constraint.

2. This assumption is a necessary condition for the inter-temporal budget constraint of the government to be fulfilled.

[Greiner \(2011\)](#), where he presents an endogenous growth model without productive public spending, with an elastic labor supply. For this model high public debt ratio reduces economic growth which is due to the fact that higher public debt leads to a lower shadow price of savings which reduces labor supply and investment. In the case of wage rigidities and unemployment, if the deficit is used for productive public investment a higher debt can lead to a higher growth and less unemployment, proven in an empirical example by [Greiner and Flaschel \(2010\)](#). An exploration work of [Greiner \(2013\)](#) concerned with the validity of [Checherita-W et al. \(2012\)](#), results which denote that debt and growth, have an U-inverted relationship. Greiner argues that once the deficit policy is allowed to vary, the higher the growth rate is, the smaller public deficit and public debt are.

In a specific case [Greiner \(2008b\)](#) develops an endogenous growth model with human capital, where the government finances education spending. Greiner states that using public deficit to finance public investment including education sector could increase growth, since primary surplus is a positive function of public debt and fiscal policy is neither loose nor strict. By assuming public debt sustainability, governments can guarantee a sustainable long-run growth path, if sustainable debt is not a feasible fiscal policy should establish a consolidation policy that is based on austerity principles. Where public expenditure, including education expenditure, should be reduced in order to reduce deficit amounts as a response to high debt levels. Beside that, policy makers should put fiscal policy question of efficiency to know whether public expenditure were efficient or not and how to rise their efficiency in a way to avoid any adverse impact of volume reduction.

From the empirical side, [Ferreira \(2009\)](#) and [Woo and Kumar \(2010\)](#) provide evidence of this negative correlation. Studying the OECD members ([Ferreira, 2009](#)), confirms that debt-growth relationship is negative and bi-directional along with the presence of bi-directional causality, he points out heterogeneity across OECD countries. While further empirical studies provide evidence on the non-linearity of debt-growth relationship such as that of [Reinhart and Rogoff \(2010b\)](#). Where, in a sample of (44) country spanning over (200) years, [Reinhart and Rogoff \(2010b\)](#) have found a varying effect of public debt on economic growth, where debt accumulation impact on GDP growth is weak until debt to GDP did not yet attain the level of (90%) which is equal for advanced as well as emerging economies. At the same threshold [Woo and Kumar \(2010\)](#) reveals that debt impact becomes non-linear after this level (%90), even more the negative effect is weaker in advanced countries, [Baum et al. \(2013\)](#) also shows that there is a negative correlation between public debt and economic growth in advanced and emerging countries. Whereas, [Cohen \(1993\)](#) argues that before certain levels debt

impact on growth is positive.

It is also proven that debt use could become dangerous in case of mismanagement, for instance [Cohen \(1997\)](#) has found that external debt can be very dangerous for countries if not managed properly. External debt can also contribute in a reduction of investors incentives³ then reduce capital accumulation together with putting more charges on the future generation in the phenomena of *debt overhang* defined by [Krugman \(1988\)](#). A large fraction of external debt will decrease the productivity to innovate as foreign investors would benefit more, ([Poirson et al., 2004](#)). Further, rising debt undermines confidence and threatens economic recovery implying fiscal vulnerability. For example general government debt in the G20 advanced economies raised from (78%) of GDP in 2007 to (97%) of GDP in 2009 and it is supposed to raise till (115%) of GDP in 2015⁴. The dangerousness of rising debt levels together with the last financial crisis in many countries with a contagion risk, impose the concern of public debt recurrence and management as a crucial topic to be studied in order to find suitable solutions that enable governments to avoid the debt spiraling out of control.

Contrary to previous studies, and instrumenting debt with a variable that capture valuation effects, [Panizza and Presbitero \(2014\)](#) did not find any evidence that high levels of debt hurt future growth in advanced economies, at the same time they did not claim that countries can sustain any level of debt. As a consequent, they think that this relationship should not be used as an argument in support of fiscal consolidation, rather they consider that the establishment of restrictive fiscal policies in order to avoid sudden shocks may reduce growth especially if implemented during recession circumstances as argued in [Perotti \(2012\)](#). Because only at this point, debt high levels will lead to contractionary policies, that is why it would be more fruitful to apply such policies out crisis time. Furthermore, [Pescatori et al. \(2014\)](#) indicate that there is no clear debt threshold that severely impairs medium term growth, only at the shortest-term horizons.

To sum up this brief literature review, one can observe the importance of each of debt-growth and growth-human capital relationships, thereby debt-human capital one. Each link is worth further investigation and exploration from theoretical as well as empirical perspective, that all the aforementioned studies sometimes show contradicting results. Especially in what concerns debt-growth relationship and the potential sub-

3. Investors perceives the high levels of debt as a future increase in taxes level, that will imply a decrease in the expectation of returns on investment as a consequent the net amount of investment in the economy will decrease.

4. Source: International Monetary Funds, Moving public debt onto a sustainable path.<http://www.imf.org/external/pubs/ft/survey/so/2010/pol090110a.htm>.

sequences on growth determinants and economy functioning. That is why one of the main exploration points in this thesis is to figure out whether controlling for public debt in an empirical growth framework will have any consequence on the relationship relating GDP growth rates with human capital factor. A human capital factor which is approximated by education expenditure and the stock of schooling years, whilst at the same time we look for the presence of public debt impact heterogeneity through a time-series cross sectional data base. These subjects and more have been investigated and analyzed through the three chapters of this thesis. The last and renewable debt crisis all over the World raises the interrogation about the relationship between public debt and economic growth, its nature and consequences on another economic aggregates amongst others the most ultimate growth determinants. This question has been explored theoretically as well as empirically, in this thesis in turn we try to extract an empirical evidence using the appropriate econometric tools in order to prepare the solid baseline for a latter theoretical development.

1.1 Problem statement

From macroeconomic prospective nations economic growth was and is still the main preoccupation of researchers, more precisely with the raising debt levels after the remarkable return of financial and banking crisis that started recently from 2007 in US and extended then to the European Union (EU) and the rest of the world. This meaningful return has resurrected the interest of developing new growth models that accounts for debt levels and their effects on economic growth process, in general, and on growth driving forces represented particularly by human capital as it has been already done lately by [Greiner \(2008b\)](#). Taking a historical view on both of debt and human capital proxies, all with accounting for growth literature and empirics famous issues would be useful to the precision of research questions in this work of thesis.

1.1.1 Historical background: debt and crisis

Economic and financial crisis reduce the amount of public resources which affect directly and negatively public finances. For instance Asiatic bank crisis between 1997 – 1998 was followed by banking system rescue plan, this plan had an important cost about (16%) of GDP in **Thailand** and (50%) in **Indonesia**. Moreover, currency crisis weighed down the weight of public debt in foreign currency. Whatever the source of crisis having an amplifying impact on public debt, public finance and public debt

should assure their sustainability to face crisis, (Bachellerie and Couillault, 2005). From the historical perspective, in the last decades many countries such as those of East Asian, central Asia and Latin American and European members have known some financial crisis that contributed in rising debt levels and worsening fiscal imbalances. Many of these financial crisis waves are, in first place, the result of financial sector liberalization, then the periods of huge financial sector growth and development which are usually accompanied by rising private indebtedness, (Reinhart and Rogoff, 2013). Since the World War II, Reinhart et al. (2012) have identified (26) episodes of excessive debt in advanced countries where debt levels had exceeded the threshold of (90%) of GDP, the official public debt represents only one piece of the larger debt overhang. The small number of cases where this crisis lasted less than (5) years, where usually the average duration of overhang episodes is (23) years.

Here, we present a brief historical background of some sampled countries where the source of the crisis are different but the final consequence was a rising public debt:

1. The oldest Asian crisis started in **India** at the second half part of 1980. Where, India external indebtedness has doubled from (35) billion dollars at the end of 1984/85 to (69) billion dollar by the end of 1990/91 and also did medium and long term commercial debt, making India more vulnerable to external, (Cerra and Saxena, 2002). The changes in the Indian economic environment (the raised current account deficit, the gradual process of trade liberation, investment and financial market) were the source of debt rising. At the same time, while the Indian petroleum consumption was still growing causing high value of petroleum importation, petroleum production was slowing. Moreover, Indian reaction policies⁵ made India more dependent on high cost short maturity financing and heightened sensitivity shifts in creditor confidence.
2. After the 1970 – 1995 Asian-5 economic growth saga⁶, a sizeable current account deficits followed by a collapse of national currencies had caused the Asian-5 economies **Indonesia, Malaysia, Philippines, South Korea and Thailand**⁷ financial crisis in the 1990s. The main reason behind this crisis was the reversal influx of capital coming from the self-fulfilling panic, this creditor panic has made Asian-5 economies more vulnerable to crisis.
3. **Russian financial crisis** of 1998, was the most severe crisis between 1997 – 2001, where many other countries were living turbulent crisis from East Asia, Turkey

5. Indian policy makers have relied on borrowing at commercial terms, and the remittance of nonresident workers.

6. Economic growth rate of the (5) economies has doubles those of OECD, life expectancy has raised by (20%) and adult literacy also did by (25%).

7. All four countries are in the sample except South Korea.

and Argentina. Russian real interest rate and real appreciation rate were remarkably high which made it difficult for companies to get profits. Beside that privatization did not yield the expected effect in term of corporate management and governance. All these effects have represented clear signs of the Russian financial crisis, despite the fact that foreign investors continued to finance Russian government debt⁸, (Pinto and Ulatov, 2010). The Russian crisis spread to a number of international investors also some financial institutions saw heavy lost, but thanks to the intervention of central banks over the wide world this crisis did not spread and lead to a financial bankruptcy, (Constâncio, 2012).

4. **Latin American crisis**, after the 1980s crisis the Mexican economy was getting back up, when Mexican peso currency overvaluation contributed in ballooning current account deficit. In addition to rebellion activities which launched fears of political instability and set off brief financial panic. The latter raised interest rates and lowered Peso's value, more details about crisis circumstances are available in the work of Whitt Jr (1996). To which, a widening of current payment deficit of (9.8) in 1993 to (7) in 1994, a new confidence crisis in November, and peso floating in (22) December have been added resulting in increased debt until (18) point of GDP, (Bachelier and Couillault, 2005). In **Brazil** 1998, weak economic growth, current payment (-4.3), public deficit (-7.9), capital outflows, rising interest rates, floating peso in 1999, all these aspects have raised public debt by (10) point of GDP in 2002, (Bachelier and Couillault, 2005).
5. The latest and famous worldwide financial and economic crisis started in August 2007, then the banking system crisis peaked in 2008 with the bankruptcy of Lehman Brothers and the support offered to the financial system. Since 2010, this crisis has been transformed into sovereign debt crisis, then due to contagion phenomenon this financial instability been spread⁹ to cover further countries in the European Union, (Constâncio, 2012). For example **Italy** has known series of high level debt¹⁰ over the years between 1980 – 2010. Italian high fiscal deficit (12%) of GDP in the last (22) years and the economic shrinking (-5.2%) were the evidence of debt rising. In Portugal, debt raised in 2007 as a result of the worsened Portuguese trade balance, low growth and high fiscal deficit. **Greece** in 2009 has started its own crisis with large fiscal deficit (12.7%) of GDP and persistent high inflation rates. Then, it has been nourished by the corruption installed latter. In

8. This strange behavior was known as *moral hazard*, where the main objective of investors is to get their part of cake, that's why they have invested their money in developing countries the less performing contrary to the prediction of the neoclassical partisan (Gourinchas and Jeanne, 2013).

9. Financial instability does not spread if there is no initial shock.

10. Relationship between debt and economic growth in this country has been showed negative of (-0.64) from 1950 – 2010 as found by a study realized in 2011 by Balassone et al. (2011).

Portugal also, Portuguese exportation has known an increased competition from those coming from China also with those of the European Union. All these factors strengthened its deficit.

During the period between 1990 – 2011 some countries could keep their debt levels stable in 2002 such as; Mexico. Other countries like: Brazil, Ukraine, Thailand and Colombia have increasing debt levels, India, Indonesia and Philippines have witnessed a substantial debt increasing while Argentina, Jamaica, Uruguay, and Pakistan have attained the threshold of (100). In counterpart, Chile, Russia and Hungary experienced an impressive debt reduction, as Jordan, Ecuador, Bulgaria, Malaysia and Morocco, they could carry out a debt reduction relative to 1990 levels, (Budina and Fiess, 2005).

Figure (1.1) shows the evolution of central government public debt of advanced and emerging economies since 1900 up to 2011, as calculated by Reinhart et al. (2012). As shown in the figure, in recent years advanced economies are more touched by debt crisis than emerging ones.

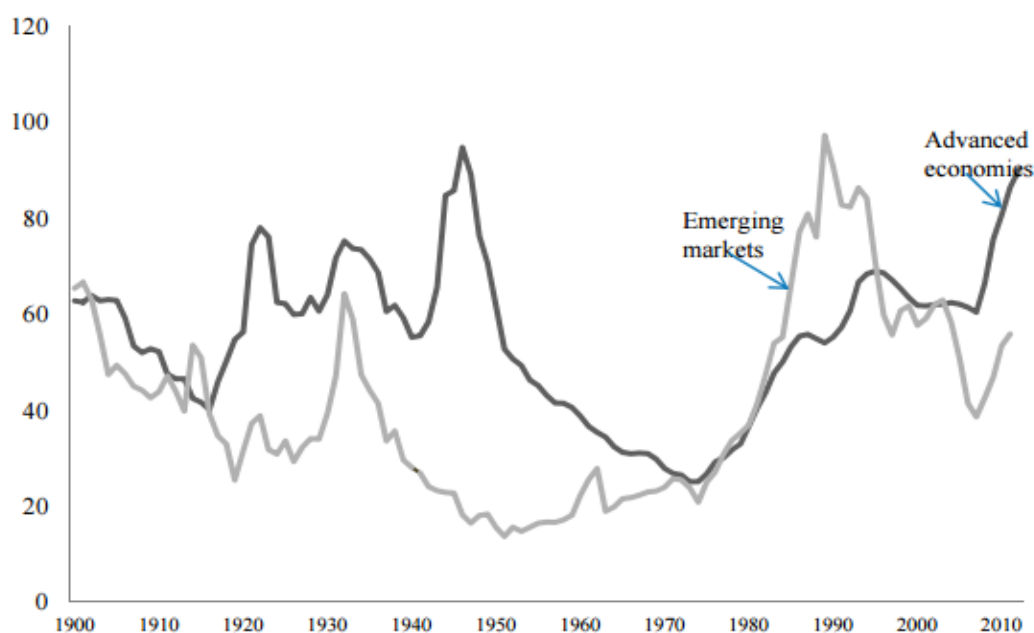


Figure 1.1: Gross Central Government Debt as a Percentage of GDP: Advanced and Emerging Market Economies, 1900 – 2011, ((Reinhart et al., 2012), and sources cited therein).

As a consequent of the different sources of fiscal imbalances, policy makers have to face debt rising levels by establishing the appropriate fiscal policies that consolidate the fiscal position and reduce debt levels. Fiscal adjustment-austerity is one of five ways to reduce large debt to GDP: *economic growth, explicit default or restructuring and a steady*

dose of financial repression accompanied by a steady dose of inflation, (Reinhart and Rogoff, 2013). The public expenditure cut off resulted because of the austerity adjustment to reduce deficit amounts could be made inter alia at the expense of human capital formation, motivated Greiner (2008b) in order to invest in a theoretical model to identify when debt use can have negative impact on human capital formation. He built an endogenous growth model with debt and human capital formation to figure out how could debt use influence human capital formation. Greiner suggests that the use of public debt in financing public activities, implying human capital formation, would be fruitful if debt is assumed to be sustainable and debt managing is neither loose nor strict.

1.1.2 Stylized facts: debt and education expenditure

Public finances in many countries, including countries of our sample, have known a rising trend in the last few years. Stated in 1990 up to now, figure (1.2) depicts both of debt to GDP levels and GDP annual growth rates for (76) countries.

Figure (1.2) reflects a rising trend of public debt that starts after 2007 with the birth of the last world financial crisis in USA, where at the same time annual growth rates have a slowdown phase. Debt-GDP growth rate link shows an inverse association and debt average is the lowest since the 1990s. After that GDP growth rates decreases sharply while the debt level is back on the increase. According to the data plotted in figure (1.2), we conclude that the potential adverse link of debt-growth relationship is worth to be estimated again.

Equally for the fraction of education expenditure in GDP shown in figure (1.3). The average of education expenditure (EDUC) and schooling years (SCH) both have an increasing trend over study period even when GDP growth rate decreased at the end of 2000s. The level of educational spending depends on the overall education system which differs among countries, where as in some countries the public sector provides services to education system this is the case in some countries more than others.

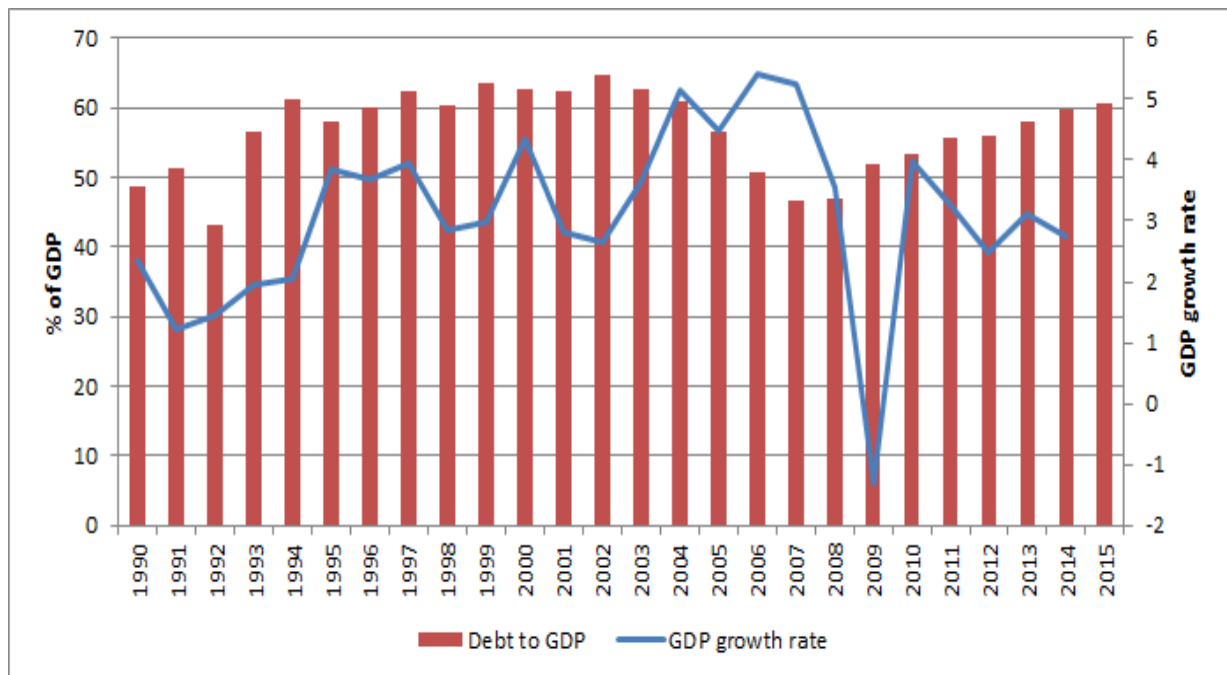


Figure 1.2: GDP growth rate, and public debt as % of GDP average for the whole sample (1990 – 2015), WEO and the World Bank databases.

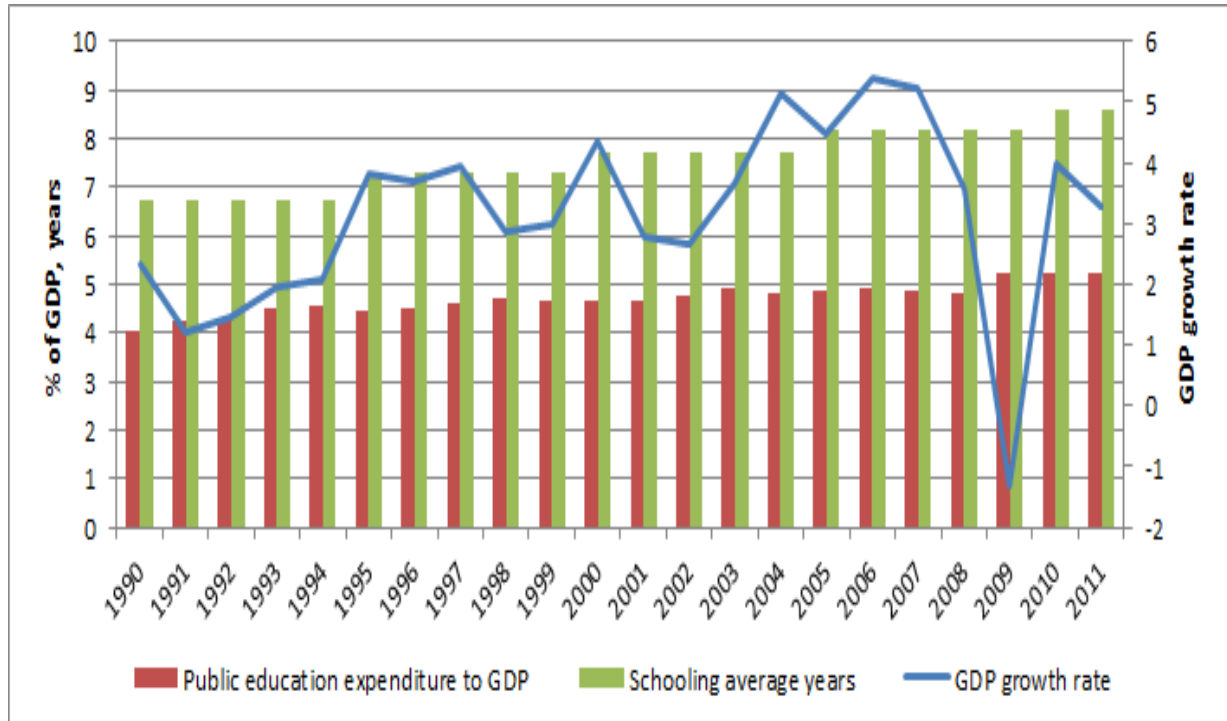


Figure 1.3: GDP growth rate, and public education expenditure as % of GDP (1990 – 2011)

For a closer view we split the sample into two groups according to their belonging to OECD organization¹¹. Each country group has a specific characteristics: *government revenue, trade openness, the quality of domestic institution and the political system* which might be considered as determinant for over borrowing, (Daniel et al., 2003). For example; in non-OECD group, where countries are developing economies, public revenues are lower and more volatile compared to those of OECD members. Looking at debt raising trend, graphs (A.1) and (A.2) in the appendix A show that since 2007 OECD members have accumulated huge amounts of debt to GDP compared to non-OECD members. This fact is due to the famous and last financial crisis that has worsened these countries' fiscal imbalances. At same period also, mean GDP growth rate of OECD members over (22) years equals (2.439), which is low compared to that of non-OECD sampled countries whose mean growth rate is (3.563). While figure (A.2) shows that, even though the difference in public education between the OECD and non-OECD members is not bigger than (1%) of GDP, but the average of schooling years in OECD members is bigger by (3.309) years of non-OECD average schooling years.

1.1.3 Growth trajectory heterogeneity

The question of differences in growth realization has recently been at the core of researcher's subjects. That some countries produce more than others is the result of different growth trajectory, where the traditional neoclassical assumption, which says that all countries have the same production technology and growth inputs are efficiently used, no longer holds as confirmed by Bos et al. (2010) while answering the question of: *do all countries grow alike?* Using a *stochastic frontier model* augmented with *latent class structure* on (77) country over (30) years period, Bos et al. (2010) could identify (3) growth regimes *A,B,C*, with three different sets of growth determinants. Group *A* gathers many Asian countries representing the *emerging regime*, their growth is driven primarily from factors accumulation and technical change, where financial development through domestic saving and foreign capital facilitate the accumulation of factors. United States and European countries belong to mature regime *B*, the increase in capital stock motivated by the high stock of human capital is the underlying force of the highest output per capita. Finally, African countries belong to developing regime *C*, output per capita growth is driven from technical change, factor accumulation {labor, capital} and efficiency change.

11. Founded in 1961 to stimulate growth and world trade, it contains 34 countries that pretend themselves as committed to democracy and market economy, providing a platform to compare policy experiences, seeking answers to common problems, identify good practices and coordinate domestic and international policies of its members.

Consequently, growth experiences and force driving are not unique, for example; the differences in human capital endowments expressed by workers heterogeneity can change the growth rate of the economy, (Zotteri, 2002).

1.1.4 Growth determinants uncertainty

Beside the last issue, another controversial one adds further difficulty to growth empirics. Henceforth, economists' efforts and attention have been shifted from theory to empirical investigation and vice versa, in both cases results are not fully satisfactory. On the one hand because many growth theories have been born, Durlauf et al. (2005) has identified the existence of (43) distinct growth theory with (145) significant growth factors. Those growth theories and factors are compatible with one another, where the validity of one theory does not imply the falsity of another. This constat renders difficult researchers intention to study growth empirics, particularly when available data size is restrained, for instance Durlauf et al. (2005) argues that there exist limits in drawing conclusion from growth data, those limits concern model uncertainty and the general weakness of available data relative to the sorts of question for which they are employed. Recently Moral-Benito (2014), based on Durlauf et al. (2005) growth survey and other economists works, concludes that once a researcher account for model uncertainty and weakly exogenous regressors issue, empirical findings suggest that economic growth does not appear to be robustly related to the determinants proposed in growth literatures.

Henceforth, together with growth models and determinants uncertainty, economists should keep in consideration, in following efforts, that nations may have a different growth trajectory with different growth enhancing forces which should be translated in appropriate manipulations and econometric estimation tools with the goal of carrying out a conclusive work that add valuable information to present growth empirics.

1.1.5 Thesis research questions

Since the public actor is the principal provider of many services, and under high indebtedness circumstances in many economies nowadays, pressure rises on policy makers to well managed fiscal policy in order to face public needs to fund all in assuming finance sustainability. Otherwise, reducing the needs of public finance can be made either by reducing public spending or by rising expenditure efficiency (increasing output with the same spending amounts, or reducing spending while improving

output quantity and quality) as an ultimate. Moreover, in the light of Greiner theoretical framework, an empirical application on realistic data of growth model gathering debt and human capital formation would be interesting.

Thereby, the core of this thesis is constructed in an attempt to answer the following questions:

1. *given growth determinants and theories uncertainty, in an attempt, we aim to identify through the enormous register of growth literature dated from 1990 to 2011 the presence economic growth engine force, inter alia, human capital and its context.*
2. *Then, in light of Bos et al. (2010) findings about the multiplicity of growth trajectory and debt rising levels, we try to identify whether public debt and human capital have a heterogeneous impact on different growth perspectives? Does ignoring debt alters the impact of human capital proxies on economic growth?*
3. *How could education expenditure contribute efficiently in economy production and functioning? Which factors do matter to improve economy efficiency?*

1.2 Methodology

In this section of thesis, we present the methodological path including all the econometric and analyzing technics we use here in aim to answer thesis core questions.

1.2.1 Model uncertainty: textual analysis & variables selection

Growth research stream concerned with growth factors that explain growth differences across countries have been interestingly evolved. Such evolution has resulted in a huge number of studies, especially that researchers have to account for the important challenge of model uncertainty.

Textual analysis

For a long time, a consensus about empirical growth framework does not exist which has resulted in a high number of empirical papers, each of them with its specific growth factors. The number of proposed growth factors attains (143) variables, where no study can run them all at the same time because of countries limited number and the limitation of their data availability. Rather researchers tend to select a subset of explanatory regressors to provide evidence of one or two interest variables, but this

approach does not receive a global consensus because it is sensitive to the inclusion and/or exclusion of variables. Moreover different forms of the model may offer a reasonable representation for data but different conclusions about what causes economic growth thereby the choice of one model will be misleading, (Ulasan, 2011).

Thereafter model uncertainty is a continuous issue, testing the robustness of coefficients estimates for a large number of policy indicator Leamer (1983) has proposed the sensitivity tool of *Extreme Bound Analysis (EBA)* which is concerned with quantifying fragile and unreliable estimated coefficients. After them, Levine and Renelt (1992) paper is the first to take into account model uncertainty in the empirical cross-country growth literature, by applying a modified version of EBA where considering one variable as robust depends whether or not this variable changes sign and statistical significance over all possible models. In short, besides being criticized as too extreme, EBA fails in providing a satisfactory solution to the problem of identifying the true determinants of growth, (Ulasan, 2011). Another recently appeared approach based on *rich regression* that involves all variables that have information on dependent variables then represented by a parsimonious regression that is so called *specific regression*. This approach known as *general-to-specific modeling (GETS)* and applied by Hendry and Krolzig (2004) and others, has been sometimes referred to as **sophisticated data mining** as pointed out by Hendry (1995), that because there can be several simplification paths from the general model and there is no guarantee that a particular simplification path leads to the true model.

(Sala-I-Martin et al., 2004) applies an alternative sensitivity approach to determine the importance of variables in cross-country growth regression, so called *Bayesian averaging of classical estimates (BACE)*. This approach combines a Bayesian concept which is the averaging across models with classical OLS estimation which comes from the assumption of diffuse prior. BACE does not require any specification but one simple prior hyper-parameter. As a result, the author's main finding was, in contrast to Levine and Renelt (1992), that a good number of variables have robust partial correlation with long-run growth. Surveying cross-country growth empirics, Ulasan (2011) argues that these studies results have been controversial in terms of robustness, and introducing new statistical tools and better variables proxies will make cross-country growth regressions more informative.

In this thesis, we do not seek to resolve growth empirics uncertainty, rather we prefer to select our framework elements from works that have been interested in identifying the most important growth regressors. Then we devote a particular attention to identify the context of human capital in growth literature. To address this objective, in

Chapter 2, we analyze a qualitative data set *corpus*, where research units are text segments issued from specialized growth empirics papers. This corpus will be subject to quantitative analysis tools that trace the occurrence and distribution of text keywords with the aim of extracting hidden meaning in discourse and latent temporal patterns. The textual analysis approach is the fruit of the match between several disciplines like: discourse analysis, statistics, linguistics, informatics and survey treatment (Lebart and Salem, 1994). In summary, textual analysis offers a rapid and efficient solution that describes content, structure and function of the messages contained in texts.

Application field of this approach can be either: open survey responses, interviews or composed corpus. Communication researchers have used textual analysis to describe and interpret the characteristics of a recorder or visual message. In more recent studies like in Saviotti et al. (2005), authors have used lexicographic analysis (LA) to map changes in the knowledge base of three firms overtime. Also it has been used to study the historic statistics by Trevisani and Tuzzi (2014). Computerized programs used to carry out this kind of analysis have been developed over the past decades. In this thesis we use *IRaMuteQ* a textual statistical analysis program, based on R statistic software, see (A).

Variables selection

As we have mentioned earlier, in this thesis we do not seek to identify robust growth determinants as in Levine and Renelt (1992), that is why we do not suggest a specific econometric approach close to what had been experienced before like EBA, BMA or GETS. We simply select an explanatory variables model based on empirical growth papers that were concerned with exploring growth determinants, by choosing the factors that have proven to have a significant impact on growth. As a result, the annual GDP growth rate, measured at market prices based on constant local currency, is assumed to depend on over (28) growth determinants.

From macroeconomic perspective there are many variables that are supposed to participate in growth production, that is why we extend covariates set to include this high number of growth determinants in an attempt to be as close as possible to reality. At the same time, there are further factors that we ignore in our analysis for many reasons. One of them is lack of data such for the following variables: (expenditure per educational level) and (enrollment rate per level). Also, we had to limit the number of explanatory variables in order to have more gain in terms of freedom degree. The subsequent list (1.1) demonstrates growth determinants, that we selected in this thesis,

and some of empirical papers motivating the addition of this predictors:

Variable	Studies
Deficit	(Levine and Renelt, 1992)
Public debt	(Greiner and Fincke, 2009)
Public spending (education, health and military)	(Aizenman and Glick, 2006), (Levine and Renelt, 1992), (Ciccone and Jarociński, 2010)
Initial GDP level	(Barro, 1991), (Durlauf et al., 2005)
Human capital measures	(Barro, 1991)
Life expectancy	(Sala-I-Martin et al., 2004), (Barro and Lee, 1994a)
Investment goods prices	(Agarwala, 1983) (Sala-I-Martin et al., 2004), (Ciccone and Jarociński, 2010)
Average years of schooling	(Sala-I-Martin et al., 2004), (Barro, 1991)
Inflation rate	(Kormendi and Meguire, 1985), (Levine and Renelt, 1992)
Purchasing power parity (PPP)	(Barro, 1991)
Saving and population growth rates	(Kormendi and Meguire, 1985), neoclassical growth theory (Solow, 1956)
Openness to international trade (exports, imports in % of GDP)	(Helliwell, 1996), (Tyler, 1981), and (Easterly and Rebelo, 1993), (Ciccone and Jarociński, 2010)
Gross fixed capital formation in % of GDP	(Plosser, 1992), (Levine and Renelt, 1992), (Sala-i Martin, 1997)
Foreign direct investment (FDI)	(Moudatsou, 2003)
Total land area	(Barro and Lee, 1994b)
Labor force	(Blomstrom et al., 1993), The standard neoclassical growth theory
Current account balance	(Sinha et al., 2011)
The initial level of GDP in 1990 and population number in 1990	(Durlauf et al., 2005), (Ciccone and Jarociński, 2010)
The initial level of public debt in 1990	(Woo and Kumar, 2010)
Political stability measure	(Barro, 1991), (Aisen and Veiga, 2013), (Mello and Perrelli, 2003)
Government effectiveness	(Helliwell and Putnam, 1995), (Grigoli, 2014), (Mandl et al., 2008)
Level of adults literacy	(Pang and Herrera, 2005)
Rural population	(Grigoli, 2014), (Jayasuriya and Wodon, 2003)
Income inequality indicator, GINI	(Pang and Herrera, 2005)
Student-to-teacher ratio at primary level	(Grigoli, 2014)
Regional dummies	(Ciccone and Jarociński, 2010)

Table 1.1: Growth determinants in growth literature

Since public debt is at the core of the investigation of this thesis, the ratio of **public debt to GDP** represents the main interest variables in our framework. **Interest rate** and **primary surplus** together with being an elementary factor that have a direct effect on debt accumulation, interest rate controls for international macroeconomic environment. The following mechanism illustrated by [Greiner and Fincke \(2009\)](#) show the link relating them to debt:

$$\dot{B}_t = r_t B_t - S_t$$

Where B_t is debt at time t , r_t real interest rate, and S_t is a primary surplus exclusive from interest payments on public debt. While, budget surplus (deficit) has been considered as an approximation of fiscal policy that is supposed to have a negative influence on growth rates as stated by [Levine and Renelt \(1992\)](#). Fiscal policy has also been approximated by government expenditures in total sum or separately such as military spending ([Aizenman and Glick, 2006](#)), which are expected to have a negative influence.

In [Levine and Renelt \(1992\)](#) sensitivity analysis, authors examine whether the conclusions of a high number of growth studies are robust to small changes in the conditioning set. Authors find that almost all results are fragile. In 1991, same authors have already shown (41) growth papers survey that: (33) include investment share, (29) include annual population growth rate, (18) include a measure of initial income which holds a negative link to growth as illustrated by [Barro \(1991\)](#), finally (13) include a human capital measure as an ultimate element to catch up for poor countries.

Then [Sala-I-Martin et al. \(2004\)](#), in a cross-sectional growth regression, using BACE have found evidence that variables of **investment goods prices**, **life expectancy** and **average years of schooling** have a huge impact on growth. Investment price has a high probability to be in economic growth model with negative impact, because as argued in [Agarwala \(1983\)](#) any distortions of market prices have a negative impact on economic growth. Schooling years and life expectancy are positively related to growth per capita. Life expectancy as a growth determinant measures the effect of many another factors such as: nutrition, health care security social, and literacy rate. To measure human capital we have added more than one proxy: **education expenditure**, **labor force**, **health expenditure**, life expectancy, and schooling years. That is because human capital involves many skills, competencies and offers many services that could not be measured simply. These proxies do not ignore the different human capital compositions, neglected for a long time from growth models. Thus accounting for those different possible compositions is important in determining the growth rate of the economy, and in particular the complementarities between the different types of human capital investment are important, ([Zotteri, 2002](#)).

As one would expect, intuitively, investment matters for economic growth, that is because it has two ways to influence the function of the economy; either by increasing the physical capital stock in domestic economy directly (Plosser, 1992), or indirectly by promoting the technology (Levine and Renelt, 1992). The intuition of the importance of investment to growth has been proven to be true in many empirical studies as Levine and Renelt (1992) who argues their robustness. In what concerns **foreign direct investment (FDI)**, Moudatsou (2003) empirical results show a positive effect of FDI on the European Union EU growth rates both directly and indirectly. Not to forget also that **FDIs** in developing countries have a crucial influence on their growth rates, mainly because these investments can replace the lack of national investment activities in developing economies, and facilitating technological and knowledge transfer contributing thereafter in improving human capital formation and Total Factor Productivity (TFP).

In a study of Kormendi and Meguire (1985) involving (47) countries' post-war data, authors find evidence that inflation reduce economic activities thereby economic growth, also confirmed by Levine and Renelt (1992) that inflation rate acts negatively on growth. **Purchasing power parity (PPP)**, with broad price indices that are systematically related to the level of economic development and perhaps to the presence of natural resources, controls for standard living conditions and measures market distortion. According to Barro (1991) PPP is expected to be positively related to economic growth rates.

Since Adam Smith, **openness to international trade** has been argued to enhance and encourage productivity and specialization¹². Latterly it has been stated as a key factor to explain differences in growth rates among countries such between Asian economies as proven by Helliwell (1996). An older study of (55) developing countries between 1960 – 1977 Tyler (1981) found a positive significant association between exports, and growth. That in developing countries, which they represent (55.26%) of our sample, fiscal policy sources are based on international trade taxes contrary to developed countries where fiscal structure feeds from income taxes, (Easterly and Rebelo, 1993).

The battery of explanatory variables we use in this thesis, contains some time fixed control variables. Those invariant variables capture any systematic differences in the level of dependent variable that exist across regression and could not be captured by independent variables. Some of these fixed variables are for the aim of capturing any evidence of income convergence, that is why we keep some variables measured as close as possible to the beginning of the sample period in 1990. The most famous time invariant variable are **the initial level of GDP in 1990 and population number in 1990** also used by Durlauf et al. (2005). While **the initial level of public debt in 1990** has

12. Specialization would not be profitable in small markets.

been added in order to participate in handling the endogeneity problem between debt and growth. [Bos et al. \(2010\)](#) empirical outcomes show that countries in the same geographical zones may have identical endowments that are useful to economy growth such as natural sources. As a consequent geography matters for economic growth, also confirmed earlier by [Barro \(1991\)](#). Accordingly to this evidence, we add a **regional dummies** to the model.

Later in Chapter 4 of this thesis, further 5 indicators are added in order to tackle the issue of education expenditure efficiency: rural population, where this fraction has a meaningful signification as it may represent the fraction of human capital devoted to land work in agriculture and livestock, that together with agricultural land they offer an approximation of agricultural sector. In second place, it reflects economy dispersion thereby influences public expenditure efficiency and repartition between economic activities. For that we consider rural population sufficiently interesting to be added to model as one of growth determinants. We also add a measure of the number of **students per teacher in primary, adult literacy, OECD countries dummy, GINI index, government effectiveness, and political stability**. The inclusion of political stability measure has been motivated by the finding of [Barro \(1991\)](#), that political instability reduce the per capita growth rates. Government effectiveness measures the perception of public services, the quality of civil services, the quality of policy formulation and implementation and finally the credibility of the government's commitment to such policies, is recognized to facilitate and strengthen the convergence of per capita income among regions as experienced in Italy between 1960 – 1970 by [Helliwell and Putnam \(1995\)](#).

The majority of variables are presented in ratio of GDP, percentage, level or currency dollar, more details about selected variables sources, abbreviations, source note and units are available in table (C.2) in the appendix C.

1.2.2 Empirical growth framework

In the rest of this thesis, mainly Chapters 3 and 4, we adopt a quantitative methodological techniques that will be deployed in a framework of international data. At the beginning, it is worth mentioning that contrary to research by [Reinhart and Rogoff \(2010b\)](#) and closer to growth empirics and interactions, the debt threshold issue is out of the scope of this thesis.

Thus our focus is put on the long-term relationship between education expenditure, public debt and growth. To this objective, we retain a linear empirical growth model based on the augmented Solow model where human capital participation in economic

growth starts to attract more attention. As a baseline, this model relates the annual growth rate of gross domestic product to a set of explanatory variables including human capital as a principal economic growth determinant. To approximate human capital we retain mainly: the stock of human capital and the number of education schooling years, gross fixed capital formation represents the factor of physical capital. We augment the basic model by adding the part of debt to GDP. Later on, other sets of economic growth determinants are included to capture more output variations.

For a country i observed at a time t , growth model form is the following:

$$Y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + \lambda S_{it} + \epsilon_{it}, \quad (1.1)$$

where:

- X is the vector of interest variables, those are: debt, education expenditure, schooling years, and gross fixed capital formation,
- Z is the vector of 20 control variables approximating economic environment,
- S is the vector of 4 time invariant variables, and finally,
- ϵ is the error term.

1.2.3 Estimation method: quantile regression

Why should-we use QR?

There are multiple sources of endogeneity bias in growth empirical studies such as: right hand side variables measured with error, the presence of unobservable time-invariant country specific characteristics, adding to them debt-growth reverse causation, that when low growth rates lead to borrow rising by that debt burdens. Scholars have worked hard to address this problem. *Internal instruments* such as lagged explanatory variables have been used as in the method of generalized moments (GMM) (Bond, 2002), although this technique is not suitable for macroeconomic data sets with relatively small numbers of cross-sectional units. Also GMM could suffer from instrument weakness as it supposes instruments to be strong without testing them, (Bun and Windmeijer, 2010). Authors in Woo and Kumar (2010) suggest that the system GMM estimates are similar to those of ordinary least squares (OLS), Panizza et al. (2013) argue that there are two possible interpretations to this result, either debt is not endogenous, or the system GMM does not solve the problem. External instrument is the other alternative solution that is rarely found, because it should satisfy a strict condition to be a strong instrument. Another solution has been considered in growth models by Caselli

et al. (1996) is the use of the first-differenced GMM estimator applied to the dynamic panel data models. In growth context, this method suffers from large finite sample biases because of the weak instrument problem. Also, Woo and Kumar (2010) suggests that the use of initial government debt to examine the impact on subsequent growth may avoid reverse causality. But when endogeneity source is a third omitted factor that affects both of debt and growth at the same time, handling endogeneity becomes more complicated. Using lagged variables to resolve the endogeneity issue could be misleading as discussed by Panizza et al. (2013) because it just mitigates the problem.

That is why we apply an alternative method that surpasses endogeneity without looking for strong instruments. This method is an extension of the standard *quantile regression* (QR) approach, introduced firstly by Koenker and Bassett Jr (1978), and then applied by a number of economists to the study of economic growth (Mello and Perrelli, 2003; Andini and Andini, 2014). The function of QR method as described by Buhai (2004) is based on minimizing asymmetrically weighted absolute residuals and intended to estimate conditional median function and a full range of other conditional quantile functions. The main motivation laying beyond the use of QR is that the stochastic relationship between random variables can be portrayed much better and with much more accuracy than the simple mean regression. Another interesting motivation, that QR is well suited to focus on particular segments of the analyzed conditional distribution, or on upper or lower quantile reference curves as a function of several covariates of interest, without having to impose any strict parametric assumptions (Buhai, 2004).

QR as a statistical procedure intended to estimate conditional quantile function has recently received increasing attention. It has been widely employed with labor and educational economics to study wage determinants, discrimination effects, transition, etc., also used in micro-demand analysis and in empirical finance.

For a quantile τ with $\tau \in [0, 1]$, Quantile regression is a linear conditional quantile function, given that Y is a random variable (Bernini et al., 2004):

$$Q_\tau = (Y|X = \tilde{X}) = \tilde{X}\beta_\tau, \quad (1.2)$$

where

$$\beta_\tau = \arg \min_{\beta} \sum_{n=1}^N |\tau - I_{(Y_n - \tilde{X}_n\beta < 0)}| (Y_n - \tilde{X}_n\beta). \quad (1.3)$$

τ is a quantile estimator and measures the relationship among a vector of inputs X_n and the output variable Y on the τ quantile of conditional distribution $Y_{(\tau)}|X$.

Further interesting features offered by QR are:

1. Instead of running mean estimation methods, where regression line passes through the mean of the data, QR estimation method allows a whole set of regression lines to be produced that pass through different quantiles of the data giving thereafter a much broader picture, (Kumbhakar et al., 2015). Buhai (2004) argues that QR complements the established mean regression methods by adding more flexibility in the estimation and more robustness particularly in non-Gaussian distribution settings. Henceforth, the response variable is considered as a conditional function on a set of covariates (Bernini et al., 2004), where it is easier to capture how explanatory variables will influence location, scale and shape of the response variable distribution unlike conventional techniques which usually investigated location-shift paradigms (Alexander and Lamarche, 2011).
2. Quantile regression estimation allow the control of unobserved heterogeneity of data and variables. Thereby the feature of considering heterogeneity by QR makes it very attractive, given the complex nature of real data. For example estimating debt-growth relationship would be biased if the researcher did not account for heterogeneity across countries, consequently QR could be the appropriate estimation method to apply on our case.
3. In case of heteroscedasticity, the conditional quantile regression functions are not merely shifted vertically but can also have varying parameters for differing values of τ the quantile of Y . In this case QR is especially useful, (Behr, 2010).
4. By contrast to mean estimation methods, mainly OLS, which are sensitive to outliers¹³, QR approach overcomes the outliers sensitivity, it is more robust facing them than do OLS, (Behr, 2010), more particularly in growth regressions where some of the large effects on growth of a policy variables in the earlier empirical studies were often caused by outliers that represent extremely bad policies, (Easterby, 2005). Even though, QR addresses many data issues that may complicate and bias econometric study, QR is not geared toward econometric rigor, instead the aim of such approach is to gain reliable inferences and further estimation flexibility, (Kumbhakar et al., 2015).
5. Moreover, this statistical technique is also used to estimate efficiency of frontier production function. The appropriateness of this application is shown when the conditional distribution does not have a standard shape, such being an asymmetric or truncated one (Behr, 2010). Also it is well suited for efficiency estimation when there is considerable heterogeneity in data. Using QR in efficiency studies allows the interpretation of different efficiency levels over the spectrum of the

13. Observations with unusually large errors or influential observations with unusual values of explanatory variables (Woo and Kumar, 2010).

production function together with identifying which factors are responsible for such efficiency level, in the case of multiple inputs, (Bernini et al., 2004).

Penalized fixed effects or correlated random effects?

In panel data models, constant is usually expected to capture variability sources e.x. individual specific effects, or unobserved heterogeneity, that other covariates parameters do not control for. That is what α parameters control for, but once the number of parameters to estimate is high, that can significantly inflate the variability of estimates of other covariates effects, that is why Koenker (2004) has proposed a regularization of these individual effects by shrinking them toward a common value which can help to mitigate the inflation effect. The shrinkage step improves the performance of individual fixed effects and β . This model developed by Koenker (2004) is known under *quantile regression with penalized fixed effects model (PFE-QR)*, where penalty shrinkage ℓ_1 maintains the linear programming form of the problem and preserves the sparsity of the resulting design matrix.

A simple quantile regression model with fixed effects takes the following mathematical form:

$$Q_{y_{ij}}(\tau \setminus x_{ij}) = \alpha_i + x_{ij}^\top \beta(\tau) \quad j = 1, \dots, m, \quad i = 1, \dots, n \quad (1.4)$$

α_i is intended to capture some individual specific source of variability, solving this model requires shrinking $\hat{\alpha}$'s to a common value with further conditions including symmetry of the τ_k 's and w_k 's to be specified when individuals number is large relative to the m_i 's shrinkage may be advantageous in controlling the variability introduced by the large number of estimated α parameters, (Koenker, 2004). For the quantile loss function, it is convenient to consider the ℓ_1 penalty:

$$P_{(a)} = \sum_{i=1}^n |\alpha_i|, \quad (1.5)$$

Model is solved when :

$$\min_{(\alpha, \beta)} \sum_{k=1}^q \sum_{j=1}^n \sum_{i=1}^{m_i} w_k \rho_{\tau k}(y_{ij} - \alpha_i - X_{ij}^\top \beta(\tau k)) + \lambda \sum_{i=1}^n |\alpha_i|, \quad (1.6)$$

thereby model is solved once $\lambda \rightarrow \infty$ and $\hat{\alpha} \rightarrow 0$

Briefly, PFE-QR is based on minimizing a weighted sum of K ordinary quantile re-

gression objective functions corresponding to a selection of K values of α , with user specified weights (Koenker, 2004). The slope coefficients of this objective function are α dependent, whereas coefficients corresponding to the fixed effects are assumed to be independent of α . The vector of fixed-effects coefficients is penalized by an ℓ_1 (lasso)¹⁴ penalty term with associated penalty parameter (λ), thereby shrinking these coefficients toward zero. The implementation of FE estimation model is based on the strict heterogeneity assumption of regressors. As FE cannot handle the scale effect (variant effect over quantiles), Bache et al. (2013) shows that this procedure is less efficient than PFE and CRE, even more biased.

Not forgetting that individuals in panel data have a specific unobserved characteristics. Those characteristics may influence dependent variable and be correlated with another model predictors and impact, which makes the isolation of causal effect between covariates and dependent variable difficult. To avoid this situation Chamberlain (1982) has developed a *correlated random effects model (CRE)* that considers the unobserved terms as projections on the observables plus a disturbance. CRE considers endogeneity as an unobserved heterogeneity problem without imposing any strict conditions or requiring any instruments. Thanks to Abrevaya and Dahl (2008), the correlated random effects model of Chamberlain (1982) has been extended to QR framework, yielding an interesting intersection between panel data and quantile regression method (QR-CRE). Abrevaya and Dahl in 2008, have used this model to estimate the effect of birth inputs on birthweight. Henceforth, there is no need to apply more sophisticated methods to deal with endogeneity problem such instrumental variables (Bache et al., 2013). This intersection enables researchers to control for unobserved constant heterogeneity via fixed effects or correlated random errors model. A standard linear panel-data model, has the next form:

$$Y_{mb} = X'_{mb}\beta + c_m + u_{mb} \quad b = 1, 2, \quad m = 1, \dots, M \quad (1.7)$$

where:

- m denotes individuals,
- b denotes birth number, in general framework it represents time
- c indicates the unobservable individual effect,
- u denotes the disturbance.

In Chamberlain (1982) correlated random error model considers the unobservable

14. Least absolute shrinkage and selection operation LASSO, is a regression method that involves penalizing the absolute size of the regression coefficient. This method is convenient when dealing with high correlated predictors (Hastie et al., 2005).

individual characteristics c_m as a linear projection on covariates plus the disturbance:

$$c_m = \psi + \dot{X}_{m1}\lambda_1 + \dot{X}_{m2}\lambda_2 + v_m \quad (1.8)$$

where ψ is a scalar and v_m is the disturbance that by default is not correlated with covariates, thereby after combining equations 1.7 and 1.8 the meaning of β is given by:

$$\begin{aligned} \beta &= \frac{\partial E(Y_{m1}|X_m)}{\partial X_{m1}} - \frac{\partial E(Y_{m2}|X_m)}{\partial X_{m1}} \\ &= \frac{\partial E(Y_{m2}|X_m)}{\partial X_{m2}} - \frac{\partial E(Y_{m1}|X_m)}{\partial X_{m2}} \end{aligned} \quad (1.9)$$

Then, β shows how much X_{m1} affects $E(Y_{m1}|X_m)$, and beyond the effect of c_m . In conditional quantile spirit, the last equation cannot hold the same way since quantiles are not linear operators, as a result we obtain two separate equations to solve:

$$\frac{\partial Q_\tau(Y_{m1}|X_m)}{\partial X_{m1}} - \frac{\partial Q_\tau(Y_{m2}|X_m)}{\partial X_{m1}} \quad (1.10)$$

and

$$\frac{\partial Q_\tau(Y_{m2}|X_m)}{\partial X_{m2}} - \frac{\partial Q_\tau(Y_{m1}|X_m)}{\partial X_{m2}} \quad (1.11)$$

Under error disturbance independence of X_m , the next two equations parameters can be solved by a linear quantile regression (Koenker and Bassett Jr, 1978) :

$$Q_\tau(Y_{m1}|X_m) = \phi_\tau^1 + \dot{X}_{m1}\theta_\tau^1 + \dot{X}_{m2}\lambda_\tau^2 \quad (1.12)$$

$$Q_\tau(Y_{m2}|X_m) = \phi_\tau^2 + \dot{X}_{m2}\theta_\tau^2 + \dot{X}_{m1}\lambda_\tau^1 \quad (1.13)$$

Thus we opt to implement PFE-QR to evaluate public debt impact on growth. To control for possible individual heterogeneity more frequent in panel data, then for endogeneity issue we repeat estimates with a CRE-QR estimator.

Bayesian model averaging

In Chapter 4, we firstly use the QR approach to estimate the frontier production function by setting $\tau = 0.95$. Members as those having the most efficient production function, then by comparing the residual individual production outputs to the most efficient quantile we obtain their efficiency score. Secondly, we use the Bayesian Model

Averaging (BMA) for policy implications. In the mid of 1990s BMA has come to a prominence in statistics, then it has been expanded to further fields such as; economics, biology and others. The use of BMA is useful under three cases that have been pointed out by [Montgomery and Nyhan \(2010\)](#):

1. The first case is when there is uncertainty over control variables in a model, here BMA can test the robustness of variables estimates more systematically than is possible under a frequentist approach. Therefore, BMA can check the robustness to show whether inferences are sensitive to plausible variations in model specification.
2. It would be valuable to use BMA to estimate the effect of a large number of possible predictors of a substantively importance for dependent variable.
3. Finally, researcher can use BMA to assess the evidence in favor of two or more competing measures of the same theoretical concept.

The important use of this method as stated by [Hoeting et al. \(1999\)](#) comes from the fact that ignoring model uncertainty, as with most standard statistical practices could lead to overconfident inferences and more risky decisions based on this inferences, while BMA represents a coherent mechanism accounts for this uncertainty. BMA is not concerned with offering solutions for estimation problems such endogeneity, and causal inferences. Also, as all statistical models BMA cannot defeat unscrupulous researchers, where one could use BMA to identify a model specification that maximizes fit to the data and then present his choice as the result of the theory. BMA principal idea is to account for model uncertainty in statistical inferences such that the true model is considered as an unobservable random variable, consequently over an interest variable BMA considers model uncertainty.

BMA basic mechanism as demonstrated by [Zeugner \(2011\)](#) has the following linear model structure:

$$y = \alpha_\gamma + X_\gamma \beta_\gamma + \epsilon, \epsilon \sim N(0, \sigma^2 I) \quad (1.14)$$

where:

- X a vector of k potential explanatory variable,
- α_γ is a constant,
- β_γ is the coefficients,
- ϵ a normal IID with variance σ^2 .

BMA estimates all possible combination of X models, then it constructs a weighted average over all of them. Model weights stem from the posterior model probability

(PMP) that arises from Bayes' theorem:

$$\begin{aligned} P(M_\gamma|y, X) &= \frac{P(y|M_\gamma, X)P(M_\gamma)}{P(y|X)} \\ &= \frac{P(y|M_\gamma, X)P(M_\gamma)}{\sum_{s=1}^{2^k} P(y|M_s, X)P(M_s)} \end{aligned} \quad (1.15)$$

$P(y|X)$ indicates the integrated likelihood which is constant over all models and is thus simply a multiplicative term, $P(M_\gamma|y, X)$ represents PMP¹⁵, $P(y|M_\gamma, X)$ represents the probability of the data given the model M_γ , and $P(M_\gamma)$ denotes prior model probability. The model weighted posterior distribution for any statistic θ ¹⁶:

$$P(\theta|y, X) = \sum_{\gamma=1}^{2^k} P(\theta|M_\gamma, y, X)P(M_\gamma|X, y) \quad (1.16)$$

BMA computes the *posterior inclusion probability (PIP)*, which indicates whether examined variables are included in the regression, PIP represents an indicator for the weighted average goodness of it of models containing a particular variable relative to models not containing that variable. Consequently, variables with low PIP are rejected from model potential regressors.

1.3 Data

To perform the analysis in this thesis, we had to construct two kinds of data sets in order to answer research questions. A broad range of data sources have been solicited to begin with in order to construct a corpora of related economic growth publications, secondly to construct a base with many macroeconomic indicators from available international databases. The second set has the known cross-sectional time-series (*panel data*) form, where each cross-sectional unit has been observed over certain periods. This kind of data structure is very useful in empirical macroeconomic studies for several reasons. First off, that a panel structure is well suited to study dynamics of changes by studying the repeated cross section observations over time (Gujarati, 2003). It also provides a richer and more informative exploitable data set, therefore, analysts can account for both temporal and individual dimensions of data, which means there is no

15. Model prior probability reflecting prior belief has to be estimated by researcher. Setting a uniform prior probability for each model is a popular choice to represent the lack of prior knowledge. The literature standard is to use a bayesian regression linear model with specific prior structure called Zellners g prior.

16. θ could be the coefficients β

gaps in the spatial dimension as in the case of a time series, or missing time dimension like in cross-sectional analysis. Thereby, increasing efficiency and degree of freedom. This leads to more reliable results for generalization since the limited number of countries available is compensated by the addition of a within-country variation to multiply the number of observations. Panel data, also, solves empirical estimates inconsistency resulting from omitting a country specific effect, leading to a misspecification of the underlying dynamic structure, (Moral-Benito, 2014). Moreover, using panel structure alleviates data multicollinearity through individual heterogeneity, also avoids correlation between disturbance and some explanatory variables, (Piotte, 2011).

Without neglecting that panel data has also some inconvenience related to the high individual dimension that should be considered when treating data. Firstly, contrary to cross section data structure in panel data the possibility of having outliers and missing data is higher. Most importantly, estimation method have to account for data heterogeneity, since there is a high number of parameters to be estimated, otherwise estimates will be biased.

1.3.1 Textual data set

Proceeding to applying the necessary empirical instrument in an attempt to answer the first question in this thesis, starts with the construction of textual data set. To prepare this set we have gathered summaries of (65) papers where the economic growth topic and its determinants are present. Since impact factor is a proxy of the relative importance of a journal, article selection has been restrained to those published in journals with an important 2013 two year impact factor, covering the period from 1990 to 2014. Defining the number of selected journals and papers has been subject to certain consideration that we detail in Chapter 2. The main source of concerned papers was online database namely *BusinessSourceComplete*.

Collection efforts have yielded a corpus of (65) texts published in (15) journals, those texts contain (8476) words with (130.40) words by text in average. Words are classed in (1307) active forms, with (771) hapax¹⁷. An active form may be one of the following: verbs, adjectives, adverbs or nouns.

Table (1.2) illustrates the number of collected articles per: journal, year, and citation range.

17. Hapax is a word that occurs only once within a context, either in the written record of an entire language, in the works of an author, or in a single text.

Publishing Journal	Number	Period	Number	Citation Range	Number
J-Dev-Eco	11	1990-1995	10	2-45	17
J-Macr	9	1996-2000	6	46-605	31
J-Eco-Lit	8	2001-2005	19	606-12309	17
J-Eco-Gr	7	2006-2010	18		
A-Eco-Rev	7	2011-2014	12		
W-B-Eco-Rev	4				
Ecol-Eco	4				
Q-J-Eco	4				
Eco-J	2				
Int-Eco-Rev	2				
J-Mon-Eco	2				
J-Ect	1				
J-Eco-Dev	1				
J-Int-M-F	1				
J-Pol-Eco	1				
Macr	1				
Total	65		65		65

Table 1.2: Articles' Summary

1.3.2 Quantitative data set and descriptive statistics

To estimate the growth model we had already exposed in this thesis, we looked for the available web data bases like the *World Bank data*, *World Economic Outlook (WEO)*, *World Governance Indicators (WGI)* and others, to construct a dataset of (76) countries all around the globe, (11) from Asia, (20) from America (19) from Europe (15) from Africa and Australia, over (22) years 1990-2011 that accounts for (1672) observations per variable. Countries included in the dataset belong to all economic development levels where: (42) developing countries, (27) developed countries and (7) countries in transition. Accordingly there is no place to selection bias and one may expect that results can be totally generalizable to another extern cross-sectional contexts. Countries and their classification are displayed in table C.1 in the appendix C.

At data collection, data was not available for a much longer time span especially for the *dependent variable*, then for *public debt*¹⁸, *education expenditure* and other variables that is why data concerns only (76) countries and temporal dimension is limited to the upper boundary of 2011. Even after ignoring countries, and years with high missingness values, we obtain an unbalanced panel set. Total variables invested in this thesis is

18. The different government regimes that govern in countries are one of the underlying reasons behind the lack of some data, such in public debt data.

made up of (38) dependent and explanatory variables. Variables names, abbreviations and sources description all are offered in table (C.2) in the appendix C. The next table (1.3) shows in descending order variables effective per source:

Total	Source
23	World bank open data
7	World Economic outlook database 2012/FMI/
5	Penn world table 7.1
2	Worldwide Governance Indicators
1	Barro et Lee (2000)

Table 1.3: Variables' sources

Five of the total variables number have not been used in the estimation model, those are; *CONS, INV, XRAT, OPEN, EXPENS*. Chapter 3 estimation model contains (28) predictors *EDUC, DEBT, SCH, GFCF, POP, LIFE, UNEMP, INF, FDI, LABOR, PPP, CACNT, PINV, SOLD, SAVE, HEXP, MEXP, EXPO, IMPO, INTEREST, REVENU, LAND, AGR, ASI, GDP90, POP90, DEBT90*¹⁹, where (3) of them are considered as time fixed control variables, plus one dummy regional variable *ASI*²⁰.

In the Chapter 4, first stage estimation model contains (24) explanatory variables, then another (6) independent variables; *RPOP, STU-TEACH, GINI, POLIST, GOVEFF, LITTER* coupled with *OECD* dummy variable have been added in the second stage estimation.

A glance at table (1.4) shows descriptive statistics of original and complete collected data.

19. Please see code variables in table (C.2) in the appendix (C).

20. We could not include further dummy regional variables, as they have stopped the function of estimation code. As a consequent, we include only one dummy variable in order to detect the presence of regional effect in the set.

	Incomplete data						Imputed data			
	Variable code	Mean	Min	Max	STD.DEV	Missingness	Mean	Min	Max	STD.DEV
Annual economic growth rate	GDP	3.16	-50.25	35.22	4.643	0	3.160	-50.250	35.220	4.643
Public debt part in GDP	DEBT	57.62	0.00	235.60	33.236	374	57.961	0.364	235.600	33.802
Education public expenditure portion in GDP	EDUC	4.67	0.95	16.06	1.797	453	4.703	0.312	16.060	1.797
Average years of schooling for persons over 25 years	SCH	7.59	1.75	13.20	2.619	0	7.585	1.750	13.200	2.619
Gross fixed capital formation portion in GDP	GFCF	21.32	5.20	76.69	6.091	13	21.336	5.200	76.690	6.096
Health public expenditure portion in GDP	HEXP	6.92	1.52	17.86	2.406	380	6.799	1.061	17.855	2.398
Public military expenditure portion in GDP	MEXP	2.19	0.05	19.09	1.611	133	2.152	0.050	19.090	1.605
Annual population growth rate	POP	1.24	-7.53	11.18	1.175	0	1.238	-7.530	11.180	1.175
Rural population part of all population	RPOP	38.66	0.000	94.58	22.009	0	38.659	0.000	94.584	22.009
General government revenue portion in GDP	REVENU	31.08	0.00	67.05	12.699	182	30.577	0.000	67.050	12.658
Export of goods and services portion in GDP	EXPO	39.55	5.15	241.40	28.955	20	39.590	5.150	241.400	28.865
Import of goods and services portion in GDP	IMPO	43.19	4.63	219.07	27.710	20	43.212	4.630	219.070	27.609
Real interest rate	INTEREST	7.69	-97.62	97.47	12.702	264	7.260	-223.275	97.470	14.657
Labor force participation rate	LABOR	62.19	38.90	89.10	8.725	76	62.213	38.900	89.100	8.730
Inflation rate	INF	29.68	-9.62	7481.66	264.07	20	47.553	-146.555	13596.612	443.765
Life expectancy at birth	LIFE	70.40	26.82	82.93	9.053	40	70.554	26.820	83.599	9.047
Price Level of Investment	PINV	63.50	6.98	374.82	25.85	87	63.758	-61.345	374.820	25.885
Primary sold portion in GDP	SOLD	-1.17	-34.24	20.49	4.503	609	-1.299	-34.240	20.490	4.341
Gross national savings portion in GDP	SAVE	20.71	-27.80	57.51	8.482	18	20.790	-27.800	57.510	8.556
Unemployment rate	UNEMP	8.77	0.30	9.30	5.146	19	8.874	-35.019	39.300	5.509
Current account balance portion in GDP	CACNT	-1.64	-41.09	39.58	7.404	23	-1.587	-41.090	39.580	7.421
Foreign direct investment	FDI	9076E + 6	-8466E + 7	3.401E + 11	2.9161E + 10	49	9047935857	-84662791822	3.40065E + 11	29047264633
Purchasing power parity	PPP	85.81	0.00	4266.32	324.367	25	86.035	-388.351	4266.320	323.054
Country land area	LAND	1075202.00	320.00	16389950.00	2614693.034	0	1075201.884	320.000	16389950.000	2614693.034
Agricultural land	AGR	40.70	1.00	85.46	21.182	34	40.701	0.524	85.465	21.125
Student per teacher ratio	STU-TEACH	24.28	9.266	69.286	11.113	501	23.919	4.109	69.286	10.754
Literacy rate	LITER	84.618	27.245	99.784	15.784	1461	85.18	24.42	100.00	14.856
Policy stability	POLIST	0.005	-2.812	1.668	0.939	684	-0.051	-2.812	2.042	0.963
Government effectiveness	GOVEFF	0.388	-1.339	2.43	0.976	684	0.309	-3.784	2.43	1.0
GINI index	GINI	38.659	19.49	94.584	10.351	1195	40.759	8.391	71.594	11.099

Table 1.4: Descriptive Statistics

1.3.3 Missing data issue

Coming back to descriptive statistics in table (1.4), missing values are scattered throughout all data cases and variables where many explanatory variables (except: time fixed, average schooling years, land, rural population and population growth rate variables) have incomplete observations. The most important missingness fraction is that of public education spending with effective of (27.09%) of the total variable, also public debt series has an important missingness portion of (22.37%). The high fraction of missing debt data might be explained by the application of strict policies concerning the circulation of financial information. GFCF (gross fixed capital formation) has the lowest missingness percentage of (0.78%). Missingness fractions differ from variable to another, as well as from country to another, where data set countries have neither the same missingness portion nor the same likelihood to missing observation, for example Netherland has only (2.372%) of total observations missing while (16.601%) of missingness for Ivory Coast. Since not all countries have the same probability to missing data, thereby assuming data *Missing Completely At Random MCAR*²¹ cannot be true (Abayomi et al., 2008). In general the condition of MCAR is hard to fulfil, that is why we relax it by assuming our data *missed at random (MAR)* where Y_{mis} missing values depend only on Y_{obs} observed ones. Beside MCAR and MAR mechanisms, there is a third pattern where missing observations Y_{mis} may depend on extern elements that had not been measured by the researcher, in such case data $Y = (Y_{obs} + Y_{mis})$ is said to be missing not at random MNAR, (Wayman, 2003). Tabachnick et al. (2001) states that the pattern of missing data is more important than the amount missing, that because randomly missing values through data matrices poses less problem and does not affect the generalizability of results as does missing non randomly data.

Identifying missingness pattern is a crucial need for the pursuit of a remedy. Table (1.5) shows how missing data has been scattered randomly through variables, years and countries:

Missing data is such a common issue in emerging as well as in developed countries as shown in table (1.5), particularly for debt data in emerging countries as confirmed by Daniel et al. (2003) whose state that the availability and coverage of public debt data vary considerably between countries²² and there is no single source to obtain data.

Until now there is no known threshold to accept missing data fraction, for instance Schafer (1999) considers that a $\leq 5\%$ missing fraction is inconsequential, while Bennett

21. Under MCAR mechanism all units have to have the same missingness probability.

22. Authors of this study, analysis emerging market economies public debt.

Year	Missing fraction	Variable name	Missing fraction	Country	Missing fraction	Country	Missing fraction
2004	3.204	GDP	0.00	Netherlands	2.372	Brazil	7.510
2003	3.375	DEBT90	0.00	Paraguay	2.569	El Salvador	7.510
2008	3.432	RPOP	0.00	UK	2.569	Lesotho	7.510
2002	3.490	POP90	0.00	Italy	2.767	Egypt	7.905
2006	3.490	SCH	0.00	France	3.755	Guatemala	7.905
2005	3.547	POP	0.00	Malaysia	3.755	Romania	7.905
2007	3.661	GDP90	0.00	Australia	3.953	South Africa	7.905
2000	3.947	LAND	0.00	Canada	3.953	Iceland	8.103
2001	4.005	GFCF	0.778	Finland	3.953	Belgium	8.498
2009	4.577	SAVE	1.077	Jordan	3.953	Israel	8.498
1999	5.435	EXPO	1.196	Hungary	4.150	Kenya	8.696
2010	6.293	IMPO	1.196	Philippines	4.150	Bolivia	9.091
1998	6.465	INF	1.196	Bulgaria	4.348	Argentina	9.289
1996	6.522	CACNT	1.376	Norway	4.348	Belize	9.289
1997	7.037	PPP	1.495	Peru	4.348	Dominican Rep	9.289
1995	7.208	AGR	2.034	Sweden	4.348	Latvia	9.289
1994	12.757	LIFE	2.392	Denmark	4.545	Algeria	9.486
1993	13.501	FDI	2.931	Mexico	4.743	Botswana	9.684
1992	15.103	LABOR	4.546	Singapore	4.743	Mauritius	9.881
1991	17.849	PINV	5.203	Colombia	4.941	Ecuador	10.474
1990	17.963	MEXP	7.955	Spain	4.941	Malta	10.474
2011	20.995	REVENU	10.885	Uruguay	4.941	Moldova	10.870
		UNEMP	13.098	India	5.534	Zambia	11.265
		INTEREST	15.79	USA	5.534	Slovak Rep	11.462
		DEBT	22.368	Greece	5.929	Albania	11.858
		HEXP	22.727	Pakistan	5.929	Honduras	12.055
		EDUC	27.093	Portugal	5.929	Jamaica	12.253
		SOLD	36.423	Chile	6.126	Nepal	12.648
				Indonesia	6.126	Ukraine	12.846
				Germany	6.324	Uganda	13.043
				Nicaragua	6.324	Costa Rica	14.032
				Thailand	6.324	Kyrgyzstan	14.032
				Japan	6.522	Luxembourg	14.229
				Morocco	6.522	Russia	14.427
				Austria	6.719	Rwanda	14.427
				Panama	6.917	Trinidad and Tobago	15.217
				Venezuela	7.917	Benin	16.206
				Cyprus	7.115	Ivory Coast	16.601

Table 1.5: Missing data per year, variable and country

(2001) states once the amount of missing is $\geq 10\%$ that will probably imply a biased statistical analysis.

1.3.4 How to deal with the problem of missing data?

The issue of incomplete datasets is a common encumbrance in the world of empirical economic studies. Through surveying quantitative empirical studies in (11) educational and psychological journals between 1998 to 2004 by Peng et al. (2006), (48%) of these studies have been identified with missing data, against (16%) that cannot be determined. Allison (2001) states that : "Sooner or later any one who does statistical analysis runs into problems with missing data". Also does Enders (2003), he asserts that a missing rate between (15%) to (20%) is common in educational and psychological studies. This phenomena hampers scientists' intentions of getting robust results for three important related points as stated by Horton and Lipsitz (2001); the first one is the loss of efficiency, the second is the complication in data handling and analysis, where most statistical procedures are designed for a complete data set (Schafer and Graham, 2002), and bias due to the differences between the observed and unobserved data. Missing data imposes technical difficulties, besides being a sign of sloppy research that is why many scientists prefer to delete them, also remarked by Orchard et al. (1972): " Obviously the best way to treat missing data is not to have them".

Inspired by Van Buuren (2012), we provide a brief survey of missing data solutions:

1. The default procedure to deal with missing values is by dropping them, keeping only observed ones, this method is called *listwise deletion (LD)*. The implementation of this method is good when missing data assumes MCAR pattern and only a few cases have missing values, in such case LD gives unbiased and significance results for a reduced database. LD procedure has many drawbacks the main one is that although the abundance of omitting information in real life application, deleting missing data will lead to information loss which in turn decreases the statistical power and increases standard errors, (Peng et al., 2006). For example in political sciences, King et al. (2001) has estimated that incomplete records are over (50%), while in some studies that percentage exceeds (90%). This issue becomes more problematic when the hypothesis of MCAR cannot be held.
2. Another solution to handle missing values consists of estimating these values, then use the new values. Estimation can be made through different methods, (Tabachnick et al., 2001): *prior knowledge, inserting mean values, using regression, Expectation Maximization (EM) and multiple input*, or even: *Pairwise Deletion PD; stochastic regression imputation; Last Observation Carried Forward (LOCF) and Base-*

line Observation Carried Forward (BOCF); and indicator method. More details about these methods are available in [Van Buuren \(2012\)](#). Quick, easy and less commonly used nowadays is the *mean imputation or estimation*, where means are calculated from observed data and used to replace missing cases. Although this may be handy to mean imputed data, this method has serious disadvantages like: disturbing the relation between variables, underestimating the variance, and biasing any estimate other than the mean. Similarly to LD, the application of mean imputation needs to assure that data is completely missing at random, otherwise the mean estimation will be biased.

3. According to [Dong and Peng \(2013\)](#) principal methods that take into consideration missing data conditions, such as *Multiple Imputation (MI)*; *the full information maximum likelihood (FIML)* and EM, provide better estimates than LD or PD. [Van Buuren \(2012\)](#) states that multiple imputation MI method is the best one given all other imputation methods disadvantages. A common point between MI and all others methods, is that all of them recognize the independence between the imputation step and the analyzing one.

The alternative approach to complete missing data: multiple imputation

The MI seminal work of [Rubin \(1976\)](#), is nowadays universally accepted. Rubin has used MI in a public use data setting, then it has been extended to more general purposes. He observed that a single imputation could not be true that why imputation should be made multiply, he relates observed and unobserved data in the same model that reflects the uncertainty resulting from missing. After all, further attractive points make the procedure of MI as the most appropriate solution for complete data statistical analysis. To sum up, MI is not a magical solution but at least is a modern statistical approach that has so many important advantages, some of these advantages are listed below as presented by [Wayman \(2003\)](#) and other researchers:

1. Wayman states that MI restores natural variability in the missing data beside accounting for data uncertainty caused by estimating unobserved values. Since imputing missing is based on the observed values that are correlated with the original missing and the causes of missingness, that maintains data original variability, and uncertainty is assured through the imputation of different versions of each missed one.
2. MI procedure is computationally simpler to use than other methods such as maximum likelihood estimation, implying by the way a good balance between results

quality and the ease of use. And by assuming different imputation, it contributes in eliminating the possible dependence between the different conducted imputations, and removing the overconfidence of the unique imputation ([Honaker and King, 2010](#)).

3. The implementation of MI does require a familiar interaction for many researchers users, thereby analyzing complete imputed sets can be done by any method or software package.
4. MI produced estimates are almost more representative of the population than the more popular methods of handling missing data like listwise deletion LD and mean substitution MS. Contrarily to LD, MI improves estimation efficiency, reduces its bias, and avoids variance and covariance bias that results in mean imputation, ([Van Buuren, 2012](#)).
5. MI codes data uncertainty through data itself, it also separates the solution of missing values problems from the solution of the complete data problems. This features make of MI a simple, elegant and powerful imputation method, ([Van Buuren, 2012](#)).
6. MI is much appropriate for standard statistics because it maintains the rectangular form of data set, thus statistical analysis are applicable on each complete data set without making data up. The attentive implementation of this procedure can minimize the importance of missingness problem, ([King et al., 2001](#)).
7. No more tradeoff between the temporal and individual dimension of the dataset, therefore advanced and developing countries can be in the same dataset regardless of their missingness portion. [Castellacci and Natera \(2011\)](#) indicates that MI allows to invest all data and available statistical information in a more efficient way, coupled with the possibility to extend time period study enabling thereby a dynamic analysis. So, researchers are capable of realizing imputation for large databases with even (240) variables, as proved by [Honaker and King \(2010\)](#).
8. Standard error SE of MI is bigger than those of other single imputation methods. That MI pooled SE has, in addition to the inherent uncertainty of any estimation method, the uncertainty that is related to missing data. Therefore, MI minimizes the bias in the SE of a parameter estimates resulted in single imputation method, ([Dong and Peng, 2013](#)).

Due to all the above mentioned features, MI has been favored in the publication of [Castellacci and Natera \(2011\)](#) where thanks to MI, they have a less biased and more representative view of the relevance of national system for development. Consequently, they conclude that MI provides great advantages vis-a-vis all other commonly adopted

ad hoc methods to deal with missing data problem. That is why the use of MI has risen notably since Rubin's seminal work, graph (1.4) which clearly displays the evolution of publications about MI.

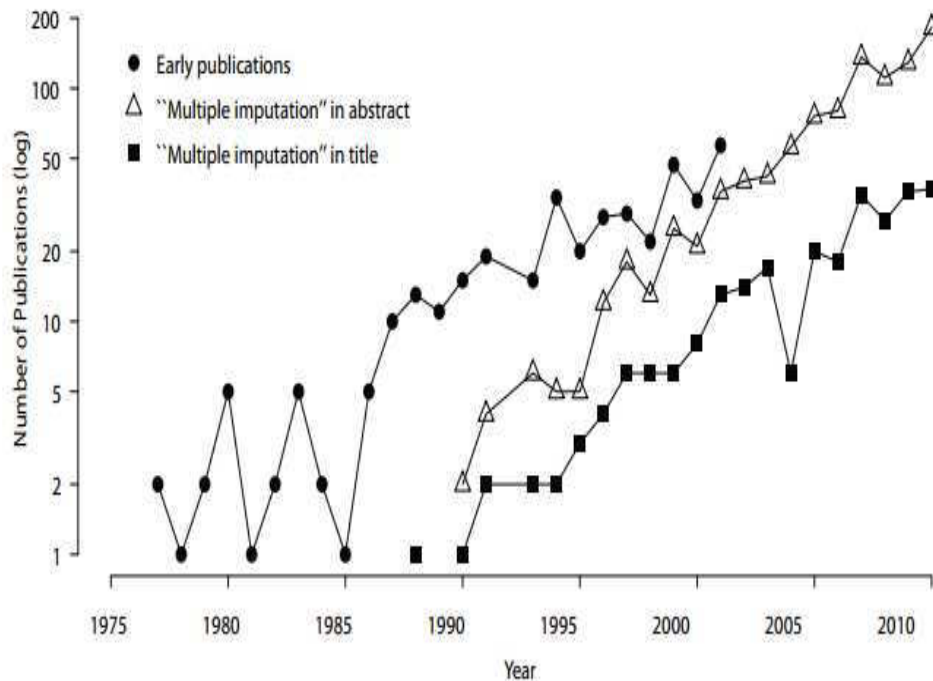


Figure 1.4: Multiple Imputation publications (log) between 1977 and 2010 according to three counting methods (Van Buuren, 2012)

Seen the important missing part in interest variables and our desire to have a rectangular complete database that enables us to carry out the statistical analysis with more confidence, we opted to experiment this imputation method in this thesis offering an empirical application on international data.

The mechanism of multiple imputation: expectation maximization bootstrap algorithm

MI is a predictive procedure that draws imputation from the distribution of observed values, then imputes each missing case m times. In the m complete data sets, units have the same observed values with m different imputed values. Since an appropriate imputation model is a determinant key to the effectiveness of MI, the imputation model should be general enough to capture data structure (Dong and Peng, 2013), as a

result the imputation model should not have more restriction than the analysis model. Denoted early by [Meng \(1994\)](#), he states that if the imputation model omits important relations that are needed in the analysis model, that would generate undesirable situation so called *uncongenial*. Consequently, the imputation model should contain useful variables for imputation²³

The function of MI is based on the Expectation-Maximization Bootstrap algorithm (EMB), EMB regroups the EM algorithm that serves to compute the posterior and the bootstrap approach to draw inferences from the posterior. For each draw, data is bootstrapped to simulate uncertainty, the EM algorithm finds the mode of the bootstrapped data, ([Honaker et al., 2011](#)). This algorithm developed by [Honaker and King \(2010\)](#) uses the Bayesian analysis in imputation procedure as the bootstrap algorithm in order to extract the new parameters for imputed data. The R package *Amelia* takes m bootstrap samples, then apply the EMB algorithm on each sample aiming to find correct estimates for variables' means and covariances under the assumption of multivariate normal distribution.

This algorithm, which is fast and easy to use, replaces two other algorithms that have been used before in the imputation operation: imputation posterior-**IP** and Expectation Maximization importance sampling-**EMis**²⁴. Equally important advantage of **EMB** is called **embarrassingly parallel**, this characteristic allows to divide the operation of imputation between many **CPU** units in parallel without any intra dependence, allowing finishing the imputation job even before **IP** and **EMis** start to run.

Thus, MI function goes through two steps, as follows:

- (i) Each missed value will be simulated independently using the conditional distribution of D_{mis} , given D_{obs} :

$$p(D_{mis}|D_{obs}, \theta_t),$$

where θ is the estimation of a current parameter, and t is the number of iteration.

- (ii) Simulation of the average and covariance matrix θ_{t+1} of the posterior population using the complete estimation of the sample.

These two steps will be repeated iteratively so that MI becomes credible ([Schafer, 1997](#)):

23. [Schafer \(1997\)](#) indicates three types of variables to be included in MI model. Those are:

- (i) variables of theoretical interest.
- (ii) variables that are associated the missing mechanism.
- (iii) designate variables that are correlated with incomplete variables.

24. The problem of **IP** is that, its application needs experts, and needs time to run, whereas the second algorithm which is based on Markov Chain- Monte Carlo is faster than the first one and needs less experience to be applied.

$$p(\theta|Y_{obs}, Yt + 1_{obs}),$$

creating the following Markov Chain:

$$(Y(1)_{mis}, \theta(1)), (Y(2)_{mis}, \theta(1)), \dots$$

which converges to

$$p(Y_{mis}, \theta|Y_{obs}).$$

The last step after the implementation of EMB algorithm is the combination of parameter estimates \bar{q} over the m imputed data set, it is easily done through computing the average of the m separate parameter:

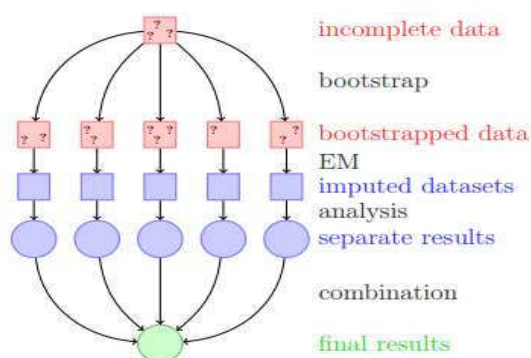
$$\bar{q} = 1/m \sum_{j=1}^m q_j \quad j = 1, 2, \dots, m$$

and the variance of the parameter is the average of the variances from *within* m completed data set plus the sample variance of parameter estimate *across* the data set:

$$SE(q)^2 = 1/m \sum_{j=1}^m SE(q_j)^2 + S_q^2(1 + 1/m),$$

S_q^2 denotes sample variance across the m parameter estimates.

To sum up, MI whole process is simply shown in the next schematic (1.5):



A schematic of multiple imputation approach with the EM algorithm.

Figure 1.5: MI schematic (Honaker et al., 2011)

Executing the imputation step requires the satisfaction of two necessary conditions, the first was that missing data should be MAR, while the second and important one is that data should be multivariate normal:

$$D \sim N_k(\mu, \Sigma)$$

Assuming data to follow a multivariate normal distribution means that variables and all linear combinations between variables should be normally distributed. This assumption consists of a crude approximation of data real distribution, thus when data does not follow a multivariate normal distribution then one can transform them to make this assumption more plausible. Schafer (1997) states that even though the definition of a large sample or the low missingness rate is not specified, but the assumption of multivariate normal distribution is robust when the sample size is large and the fraction of missing data is low. However, Tabachnick et al. (2001) argues that with sufficiently large sample size, the Central Limit Theorem reassures that, sampling distributions of means are normally distributed regardless of the distribution of variables, which corresponds to our case where ($T = 22, N = 76$). As a consequence, we consider that the multivariate normal assumption holds²⁵.

Another important element to assure imputation effectiveness is the choice of imputation's number which depends on missing data part. Rubin (2004) confirms that a

²⁵. Transformation of many types of variables can often makes normality assumption more plausible, Amelia package allows the implementation of such transformation.

multiple imputation of (5) times guarantees an efficiency level up to (95%) for missingness fraction of (50%). To remove noise from estimations [Schafer and Graham \(2002\)](#) recommends a more restrictive imputations number of (20). Seen the importance of m imputation number, [Van Buuren \(2012\)](#) considers setting m to a higher number between (20) to (100) would be more useful. In first MI application we had set $m = 100$ to impute thesis quantitative data set, in second application we set $m = 1$, then we compare imputation through diagnostic instruments in order to capture the effect of varying m on imputations validity.

Data set in Chapter 3 has been imputed with $m = 100$ times. We choose to set $m = 100$ sufficiently high, in order to raise the likelihood of approximating missed information as much as possible. While the same data set with supplement variables, mainly (5) new variables with high portion of missing values, will be imputed this time for (5) times in order to explore the sensitivity of results to imputation number.

The imputation model contains as much information as the analysis model, that is why we have added 5 more variables to imputation model those are: public consumption (*CON*), investment part in **GDP** (*INV*), public expenditure in **GDP** (*EXP*), openness to trade (*OPEN*) and exchange rate (*EXCHANG*). Doing so, the prediction power of the model is increased and the assumption of random missed data is guaranteed. We also impose a logical condition on (12) regressors. Those regressors are: *DEBT*, *EDUC*, *HEXP*, *REVENU*, *MEXP*, *EXPO*, *IMPO*, *GFCF*, *PINV*, *UNEMP*, *PPP*, *AGR*. The main objective of these conditions is to ensure that the new imputed values will not go over each of the variables limits (*the minimum and maximum observed value*). For a more realistic imputation we also have employed the lagged and future values of all regressors. Health expenditure variable had missing values for all individuals between 1990 – 1994, these missed values have been replaced by the value of $1995 \pm 2 * StandardDeviation$. Imputation code is available in the Appendix A.

Imputation diagnostic for $m = 5, 100$

Verifying impute validity does not mean hypothesis testing, since the plausibility check uses external information or speculation together with the fact that imputes are unknown. Rather, diagnostic consists of checking whether the imputation model fits observed data by controlling for differences between observed and missing data. Thereby, a diagnostic can be made either by a comparison with external knowledge, or internal related to observations and model ([Abayomi et al., 2008](#)). Internal tests can be performed on the imputation model itself, that graphical figures represent an external

diagnostic tool in sense of observed data.

Many diagnostic options are possible to verify imputes validity and plausibility, for example:

1. The simplest diagnostic visual tool is to compare variables distribution before and after imputation, to check whether the mean prediction of the imputation falls within known bounds when such bounds exist (Honaker et al., 2011). MCAR hypothesis can yield some differences in imputes' distribution, such differences are explicable by other variables in the set. But if the distribution graph is strange or very far from the observed one, imputation could be incredible.
2. Because of the nature of the missing data mechanism, and since these missing values do not exist to enable one to compare after imputation data with the missing, Honaker et al. (2011) have developed another technique (*overimputation*) to judge the fit of the imputation model. Overimputing data considers sequentially all observed values of each variable as unknown to be imputed for hundreds of times in order to construct a confidence interval of what the imputed values would have been, had any of the observed data been missing. Hence, by checking how many confidence intervals cover the $y = x$ line, an analyst can tell if imputation model was a perfect predictor of true values. The application of overimputation technique is possible by Amelia package as presented in Honaker et al. (2011).
3. Bivariate scatter plot, this plot compares the internal consistency of the observed and imputed data to detect whether there is unusual differences in the variables distribution before and after the imputation, as suggested by Abayomi et al. (2008).
4. Further tools are available by Amelia like: over-dispersed starting values, and time-series (Honaker et al., 2011).

To inspect imputations plausibility for $m = 100$ and $m = 5$, in this thesis we have chosen to apply two main diagnostic tests: *overimputation*, and *density graphes*. First off, we compare distribution of imputed versus observed values in order to capture any kind of after imputation distribution anomaly given the distribution of observed data. Then, we overimpute observed values to further verification. For m set to (100), densities graphs for almost all variables after imputation overlay those observed values, in figures (1.6, 1.7) in red represent the after imputation distribution line, while the black represent true observed data. Densities diagnostics have identified 3 categories of goodness of imputation model fit between imputes and real observed values. The first category, involves variables with a perfect imputation distribution: *EXP, IMP, DEBT, EDUC, HEXP, INTEREST, LABOR, PINV, REVENU, LIFE, MEXP, SOLD, SAVE, UNEMP*. The second contains 4 variables: *FDI, CACNT, AGR, GFCF* for

which the after imputation densities are slightly different. Finally, the last category has two variables *PPP*, *INF* for which the imputation model did not succeed completely to fit unobserved data distribution. Figures (1.6, 1.7) and 1.8) show densities for only (6) variables, two from each category, whilst the rest of the variables graphs are plotted in the appendix A.

Line color in overimputation graphs refers to the fraction of missing observation in the pattern of missingness for that observation²⁶, whereas dots represent mean imputation of each observation. The confidence interval width, for almost all variables, does not reflect a great amount of missing information for variables missingness patterns. In overimputed variables, the percentage of observations that have fewer observed covariates and higher variance across the imputed values is low. Therefore, since a high portion of observed data falls in their range, where their confidence intervals cover the $x = y$ line of perfect agreement, thus *imputation model has well imputed observed values, by consequent missing values*. Unlike densities graphs, in overimputation images (see 1.9) only a few values are dropped out of the $x = y$ line, especially for *FDI*, *INF*.

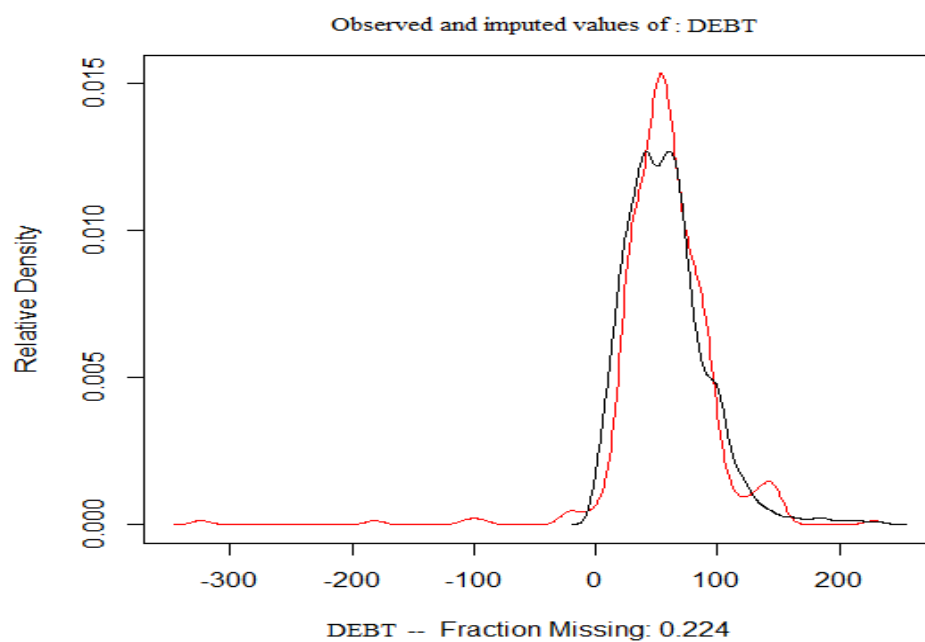
Identically, for $m = 5$, distributions diagnostic and overimputation figures are plotted in the appendix A. Looking at them, we observe that they are very close to $m = 100$ diagnostic graphs, accordingly we can say that in both cases $m = 5$ or 100 MI model has been successfully made to yield such valid imputations. Henceforth we dispose a complete panel dataset to follow estimations in this thesis.

1.4 Thesis outlines

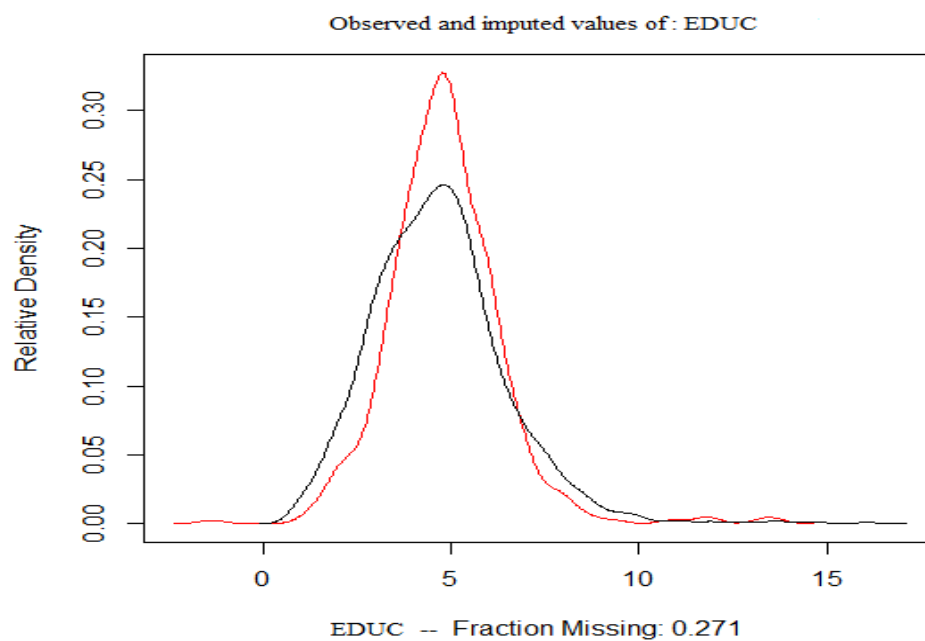
As we mentioned earlier, given growth model uncertainty and the abundance of theoretical and empirical growth literature, adopting a growth model and identifying its determinants is a difficult task. To this respect in Chapter 2, we aim to identify, through an important register of economic growth articles, the presence of streams of growth determinants, and capture their evolution through time. Our attention is mainly put on human capital for the following important reasons:

- (i) Recent studies consider human capital as the most important element to sustain growth since the birth of the human capital theory in the studies of Schultz (1961).
- (ii) The role of human capital in the growth production function has been in discussion for a long time between economists: should they consider human capital stock or its accumulation? This question has been the focus of Lucas Jr (1988) and Change (1990).

26. In graph legend, the fraction of missing data and lines colors are shown.

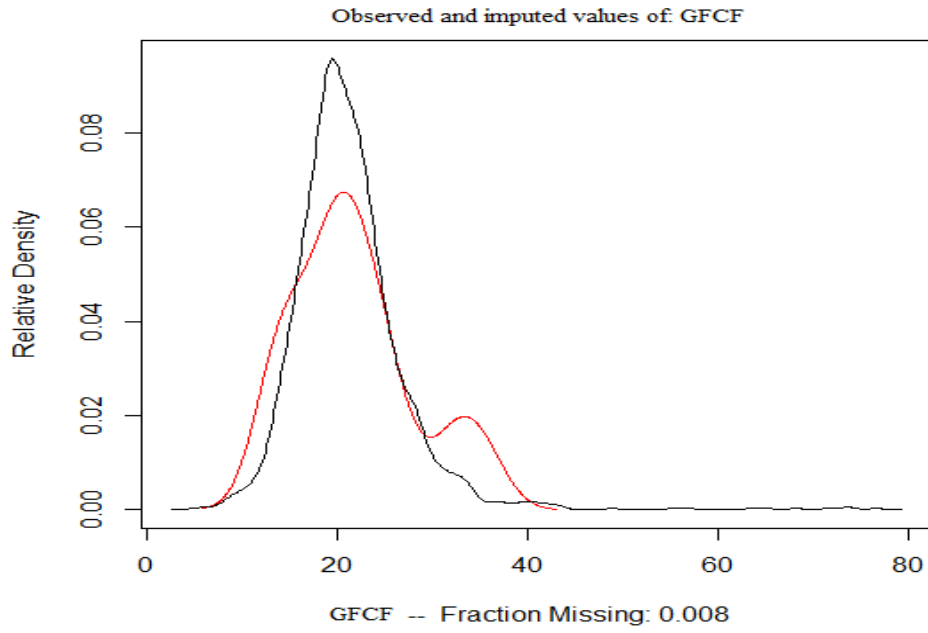


(a)

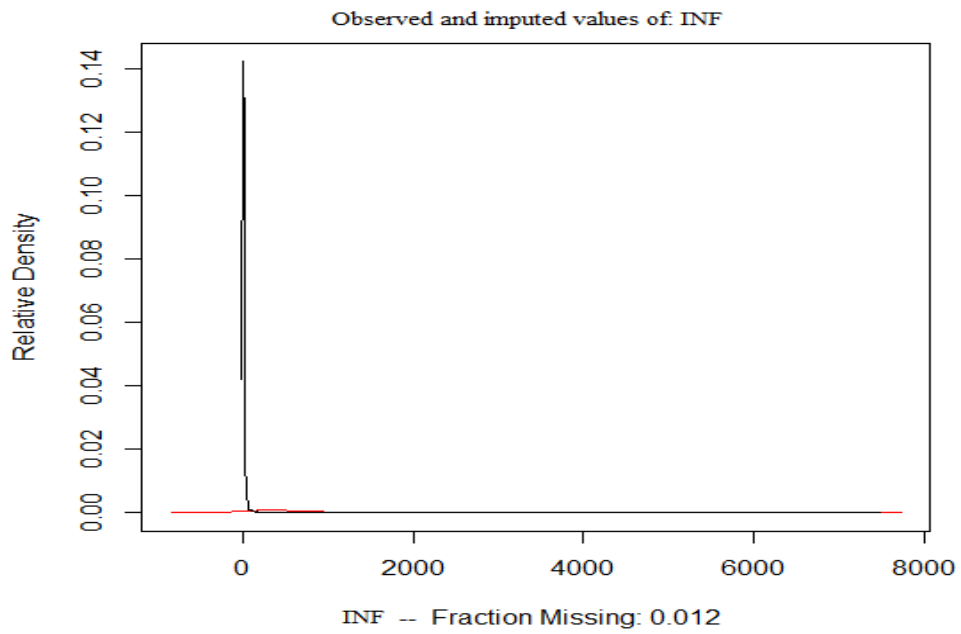


(b)

Figure 1.6: Densities Diagnostic I, $m=100$

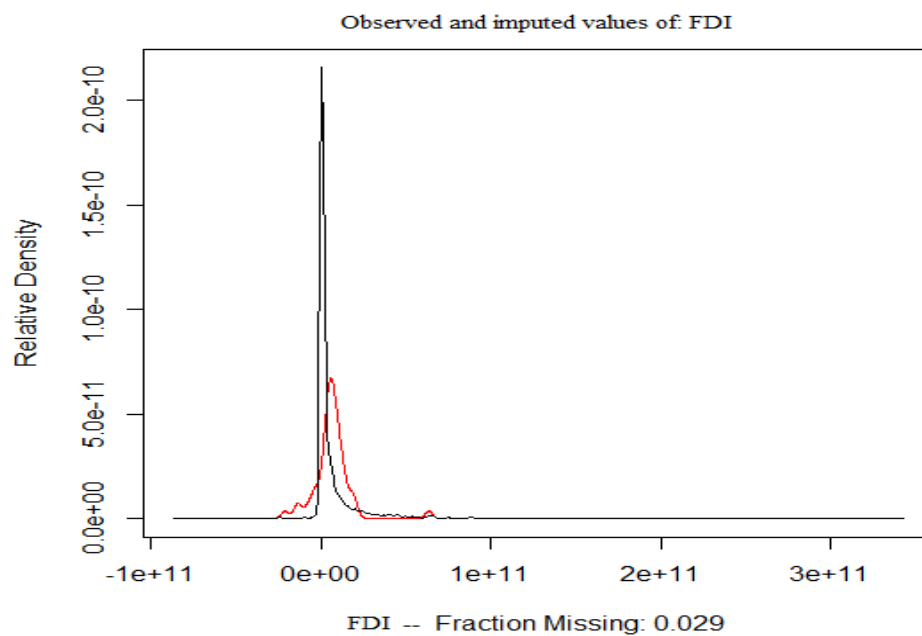


(a)

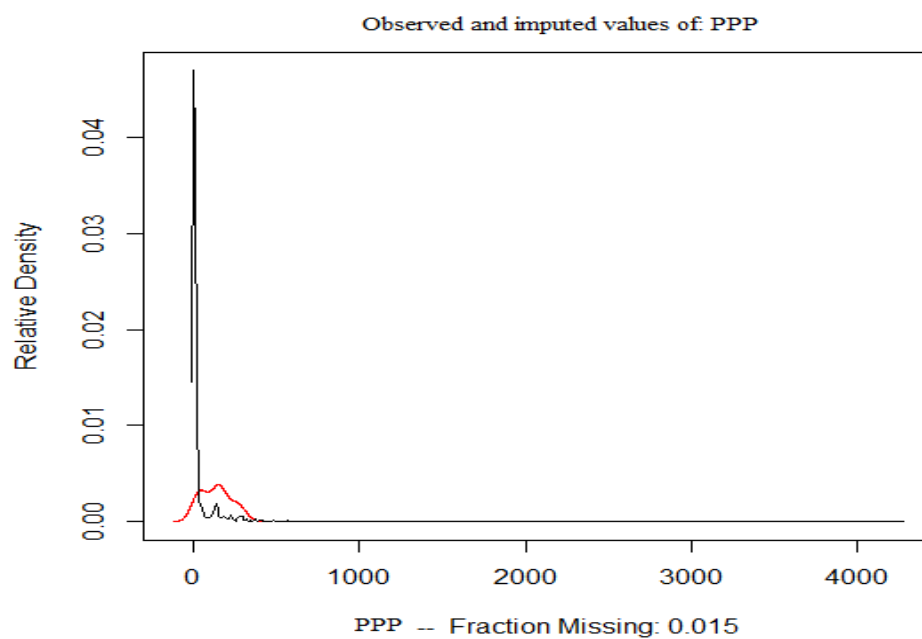


(b)

Figure 1.7: Densities Diagnostic II, $m=100$



(a)



(b)

Figure 1.8: Densities Diagnostic III, $m=100$

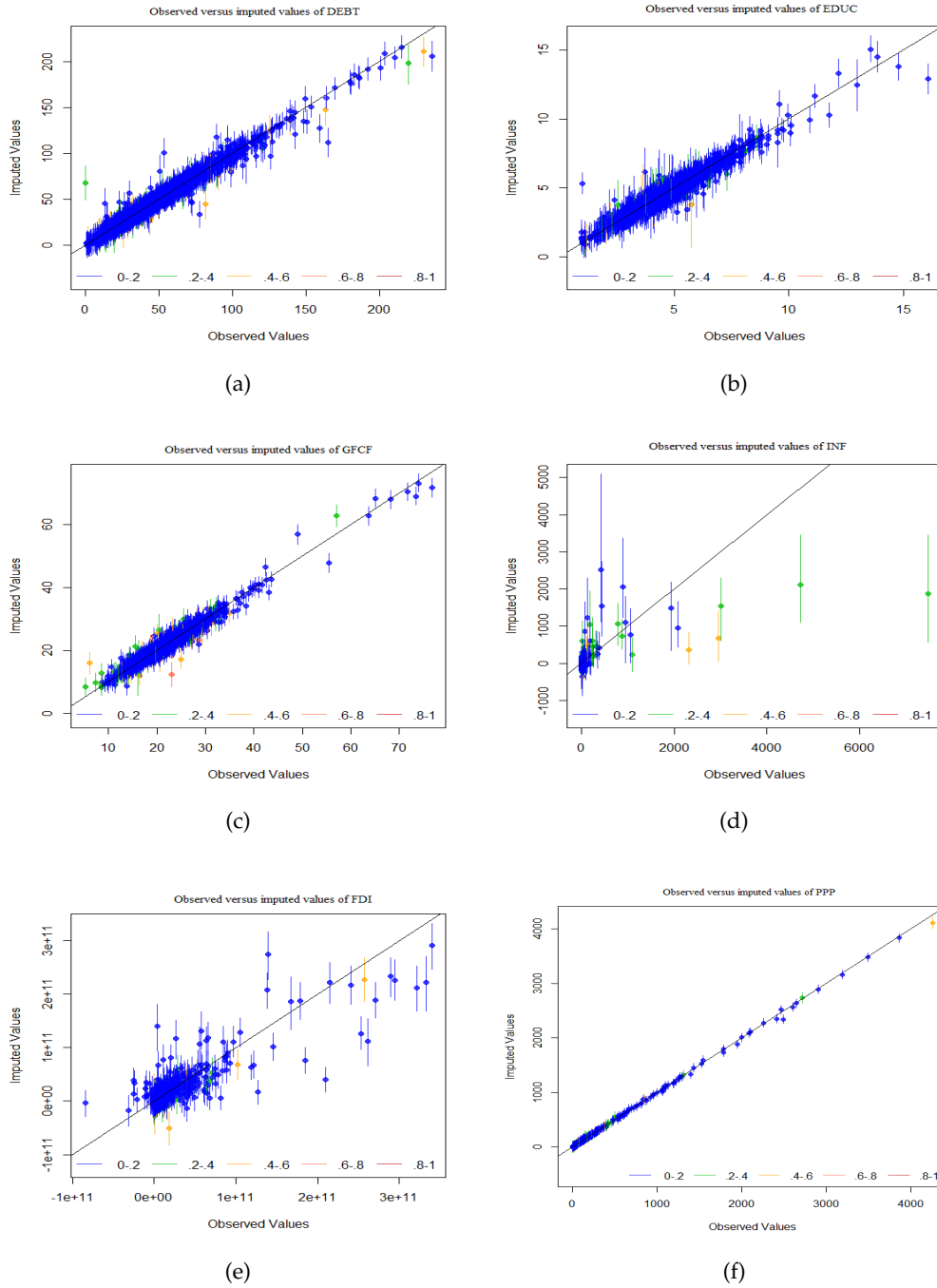


Figure 1.9: Overimputation Diagnostic, $m=100$

(iii) There is no real consensus on how human capital could be measured and represented. For instance, the conventional standard to measure human capital stock that is categorized into three parts: Output-, Cost-, and Income-based approach, OECD measure of human capital where the constitution of human capital measure is based on three factors: investment, quality adjustment, and result of education, and finally the index of human development considered by United Nation Development Program (UNDP) (Kwon, 2009).

In this respect and due to the complexity of this task, we use a specific analyzing approach destined to deal with qualitative data that is based on textual statistics so called the *textual analysis*, to be implemented on textual corpus that we have constructed from a number of scientific papers abstracts. This approach is the result of the link between different disciplines: discourse analysis, statistics, linguistic informatics and survey treatment, (Lebart and Salem, 1994). Employing textual analysis in research studies is recent, and goes back a dozen of years. Then its use has become more common, Saviotti et al. (2005) used the textual analysis to map the changes in knowledge base of three firms overtime. Also, in 2014 it has been used to study the historic of statistics through articles' titles (Trevisani and Tuzzi, 2014).

The literature review demonstrated a link between public debt and economic growth. Governments frequent use of debt to finance objectives imposes heavy burdens on developed and developing economies. This impact could vary according to countries development level and growth prospects. Moreover, the economic and financial crisis have interestingly contributed in debt build up, that is why studying debt is so relevant. Thus our objectives in Chapter 3 are of a purely empirical nature, where firstly, we seek to evaluate the long-run relating between both of public debt and public education expenditure with GDP growth rates. Then secondly, we detect whether the last relationships are heterogeneous through the different countries and growth rates.

Our latter objective is of main interest for many reasons. The first suggested by Bos et al. (2010) states that countries do not all grow alike which implies that growth relationships with macroeconomic indicators could also be different. While the second reason comes from the fact that countries' capacities to tolerate high levels of debt are variant depending on a number of characteristics that some of them may be unobservable (*unobserved heterogeneity*). Some of these factors could be due to the crisis, macro and institutional structures (Reinhart et al., 2003). Thereby considering common parameters for all countries will bias results as they ignore this heterogeneity. To this objective the thesis invests in a linear model inspired from growth theory and augmented by the ratio of debt to GDP. Thirdly, we also have been inspired by Greiner (2008b) who had already analyzed an endogenous economic growth model with human capital for-

mation and public debt. Greiner has found that using public debt could have a negative reaction on human capital formation if debt levels have not been controlled. Neither strict nor a loose fiscal policy and debt management should be conducted to assume the good investment in human capital through education expenditure. It will be sufficient to guarantee the primary surplus in a positive function of debt. In line with Greiner study, we try to estimate empirically how debt could change human capital formation relationship with growth through human capital proxies mainly education expenditure and average years of schooling.

Before the model estimation, due to the fact that long data series are not widely available to estimate relationships under question, we have limited the size and length of data to only (76) countries over (22) years. Given the number and the importance of missing cases, we preferred to impute data using the *Multiple Imputation* approach in order to have a rectangular matrix. Then in order to have robust inferences and to overcome the problem of growth model linearity, we estimate the model using the flexible method of *Quantile Regression*. QR allows to estimate the impact of explanatory variables on the conditional function distribution of the dependent variable. It also provides the appropriate estimator that respects the longitudinal nature of data (*Quantile Regression with Penalized fixed effects*) and handles the common growth models issue of endogeneity (*Quantile Regression with Correlated Random Effects*).

Public means are mostly issued through taxes, these taxes have a distortion impact that may constraint economic growth. For this reason, the use of public expenditure has to improve the long-term growth perspective with respecting other public targets like equity, stabilization and sustainability. Investing in human capital formation through the education system is the necessary activity to sustain growth then assuming all public targets. Thereby, when public funds are devoted to economic activities, such as education expenditure, economic conditions and public objectives should be considered. Consequently public spending efficiency and effectiveness is improved, which not only helps maintain fiscal discipline, but also is primordial in promoting the structural reform agenda and alleviating budget constraints as it allows outcomes at the same level of inputs to be achieved. Because lowering public funds efficiency is translated by: less effective capital, lower GDP, lower fiscal revenues and potentially unsustainable debt; the government as economic leader, should control their use as the good value of money is a necessity regarding the continuous pressure on public expenditure, especially in time of a debt crisis. For this purpose, highlighting the concept of efficiency of public performance becomes elementary. Simply, efficiency represents the input-output relationship evoked earlier in [Farrell \(1957\)](#).

In a continuous perspective to Chapter 3, in Chapter 4 we are interested in estimating the contribution of education expenditure in whole economy technical efficiency, and in how to improve their efficiency. Techniques measuring efficiency have been evolved since the first investigation of Farrell (1957).

Contrary to the classic parametric and non-parametric efficiency estimation methods, we employ a semi-parametric method through a *composed approach*. In first step, we estimate the frontier production function, where education expenditure is considered one input amongst others. Which it is worth to acknowledging that the common efficiency estimation methods usually do not allow for more than one or two inputs, in contrast to the quantile regression approach that we use in this chapter. Then, in second step we continue using the *Bayesian Model Averaging* approach to regress first step efficiency scores on a new set of environmental, and educational system variables, aiming at identify efficiency enhancing determinants. Equally to Chapter 3, and since it concerns efficiency estimation, the data set used in this chapter has been imputed depending on missingness fraction in order to obtain more reliable results.

Chapter 2

Identifying Economic Growth Schools: The Case of Human Capital

Contents

2.1 Introduction	72
2.2 Economic growth research streams	74
2.3 Methodology: textual analysis tools	77
2.4 Results and Discussion	84
2.5 Conclusion and limitations	94

Economic growth has preoccupied researchers concerned with offering a more comprehensive view of growth production formulation and sources. These efforts has resulted in a huge amounts of growth literature that are not only non conclusive but also complicate empirical studies researches. That is why in an attempt to overcome this difficulty, we deploy the textual analysis technic, that has been implemented for long time in social studies and recently in economic literature to analyze the evolution of growth research streams focusing on growth determinants, mainly *human capital*. This technic is based on extracting quantitative results and hidden messages through the analysis of an textual corpora. The application of *textual analysis* on a corpus, whose content is extracted from economic growth specialized papers published in journals of important impact factor, has enable us to identify the main themes associated to economic growth in these papers thanks to the creation of classes (with further graphical representation of each class' words). Empirical results have revealed the presence of six different clusters of growth contexts and determinants. We highlight the place of human capital in some of these clusters.

2.1 Introduction

Economic growth topic had mainly known two streams of studies through time: *theoretical and empirical*. Scholars main focus in theoretical studies was to identify the deriving forces of economic growth in order to have the first understanding on the motors of economic growth. While in the empirical branch the objective is to understand the underlying reasons of growth effective differences. In the under theoretical lines, many economists of different economic doctrines had devoted a huge amounts of time and effort to develop economic growth models within different mathematical shapes. One of the most known theoretical doctrines is the *neoclassical* whose main contribution is the exogenous growth theory, where growth factors are determined out of model or "exogenously". Such as the earlier Cobb-Douglas model (Cobb and Douglas, 1928), where growth is the result of the contribution of three principal rates: saving rate, physical capital depreciation rate and population growth rate. Followed by Solow (1956), where the author offers a new version of growth model that explains differences between countries through three factors: technical changes, scientific knowledge and creative human capital. Contributions of Romer (1986) and Lucas Jr (1988) have consolidated the birth of the endogenous growth theory, since new endogenous factors are considered such as knowledge and innovation, implying an increasing return to scale. Thereby a self-sustainable growth can be derived from: new knowledge (Change, 1990; Grossman and Helpman, 1991), innovation (Howitt et al., 1992), and public infrastructure (Barro and Sala-i Martin, 1990). As a result, together with the basic growth factors (physical capital, and labor), human capital is the key determinant of growth as it is the source of technical change, innovation and research & development works.

Another economic growth model is the cumulative causation approach¹ referred by Myrdal (1957) and Kaldor (1970) where the latter attempts to explain the reason lying behind the slow rate of growth in the UK. This theory has some similarities with the endogenous growth one.

New economic geography theory (NEG) suggests that economic activities tend to agglomerate in a specific region and choose a location with a large local demand resulting in a self-reinforcing process. The location of economic activities and its externalities on growth is the main concern behind this theory (Petraikos et al., 2007). Further theoretical and empirical approaches emphasis on the significant contribution of some non-economic factors, those are discussed in section 2.2.

1. Cumulative causation is qualified as the process of interaction between increase of demand induced by increases in supply and increase in supply in turn is generated in response to increases in demand.

Empirically, more attention and techniques have been deployed to estimate theory and explain cross-country growth differences. A large number of empirical studies have been carried out on real life data (Barro and Sala-i Martin, 1990; Barro, 1991; Sala-i Martin, 1997), and (Sala-I-Martin et al., 2004). All of these studies and many others do not agree on a unique set of growth determinants, even though having the same economic growth determinants does not necessarily imply necessarily that all countries have the same growth process. According to Bos et al. (2010), each country or group of countries has its specific process where some determinants are more important than others. The contribution and importance of those determinants are variable and depend on the country's situation. One can find the empirical evidence in Petrakos et al. (2007), where economic growth determinants do not all share the same influence on economic growth from one country to another, even from one region to another.

Carrying out a macroeconomic study requires a growth model, set up using a battery of growth factors. In the light of the growth theories' uncertainty and ambiguity, this study proposes the identification of streams of growth determinants and their evolution through time, with the use of an extensive bibliography of economic growth articles. Our attention is mainly focused on human capital due to the following reasons:

- (i) Recent studies consider human capital as the most important element to sustain growth since the dawn of human capital theory following the studies of Schultz (1961).
- (ii) The role of human capital in growth production function has been in discussion for a long time between economists: should they consider human capital stock or its accumulation? This question has been the focus of Lucas Jr (1988) and Change (1990).
- (iii) There is no real consensus on how human capital could be measured, the conventional measurement, or the index of human development (Kwon, 2009).

To this respect, we use a novel approach that is based on textual statistics, so called *Textual Analysis (TA)*. The birth of this approach goes back many years through the match between several disciplines like; discourse analysis, statistics, linguistic informatics and survey treatment (Lebart and Salem, 1994).

Using *functional textual data analysis (FTDA)*² and *model based curve clustering (MBCC)*, in 2014, Trevisani and Tuzzi (2014)'s study explored the opportunity of learning about the evolution of concepts, methods and application through studying the evolution of keywords in titles of paper published in the Journal of American Statistical Association (JASA) and its predecessors between 1888 and 2012. In this study also, corpora texts

2. FTDA represents and compare the individual trajectories of each keyword resorting to a wavelet based-decomposition of the signal-trajectories, and a functional mixed model to clustering them.

are arranged chronologically, in order to retrieve information about the evolution of a language. This *Lexicographic analysis (LA)* was used by [Saviotti et al. \(2005\)](#) to map the changes in a knowledge base of three firms over time. Each time the application of this approach was possible thanks to a computerized program created for this purpose. According to [De Looze et al. \(1999\)](#) LA is originated in linguistic engineering³, this analysis was applied to all text in order to identify themes and keywords.

This analysis approach has two main applications:

- (i) The stylometry: whose goal is to compare texts based on a quantitative study of vocabulary, words distribution, etc.
- (ii) Contents analysis: which is the most important application where the sense of the text is sought more than its form. This application known for its efficiency, dates back to 1995. The application of this method can be done either on responses of open questionnaires found in investigation, or on a corpus composed of different texts.

Moreover, textual analysis allows the detection of the latent temporal patterns in subjects' keywords.

This chapter is organized as follows. In section (2.2), we present a brief overview on the main research streams of economic growth. The methodology of research adopted in this chapter is presented in section (2.3). Results are provided in section (2.4). In section (2.5), we expose the conclusion and main limitations.

2.2 Economic growth research streams

Until today economic growth process was still inadequately conceptualized and poorly understood due to the lack of a generalized theory explaining the function of economic growth. This lack of a unique theory has led to a multiplicity of theoretical bases adopted in applied studies, resulting in unreliable findings. But this handicap did not hamper researchers from investigating in this field and trying to point out the role of various factors in determining economic growth process.

Through the wide range of applied studies specialized in economic growth factors, some factors have gained more emphasis than others. The work of [Barro \(1991\)](#) constitutes the seminal empirical application on economic growth literature. It has been followed by a series of similar works ([Sala-i Martin, 1997](#)), ([Doppelhofer et al., 2000](#)),

3. Linguistic engineering means a constructed dictionary containing relevant keywords. The relationship of different documents can then be studied by measuring the frequency of co-occurrence of keywords between different patents or different technological classes.

to name a few. In [Barro \(1991\)](#) authors examine (33) variables in a cross-section of (98) countries. Their main finding was that per capita growth rate is related to the initial level of human capital, where there is no strong relationship between economic growth per capita rate and the initial level of income, which contradicts neoclassic convergence condition.

Here we present briefly some of growth regressors that have been empirically analyzed:

1. Investment rate: which is a factor that has the agreement of neoclassical and endogenous growth theory. Many studies have tried to investigate the relationship between economic growth and investment rate: ([Sala-i Martin, 1997](#)), ([Barro and Sala-i Martin, 1995](#)), ([Mankiw et al., 1992a](#)) and ([Levine and Renelt, 1992](#)) to mention a few.
2. Innovation and R&D activities increase productivity and growth. This role has been conducted by various endogenous growth models, ([Ulku, 2004](#)).
3. Openness to trade: the importance of this factor comes from its role of enabling technology transfer between countries, also the exploitation of comparative advantage, ([Dollar, 1992](#)).
4. Foreign direct investment FDI as openness trade: this factor plays a crucial role in technology transfer and economic growth. Applied studies about the FDI's relationship with economic growth have provided relatively consistent findings affirming a significant positive link between them, ([Lensink and Morrissey, 2006](#)).
5. Political environment: examining the relationship between economic growth and political factors has been carried out by sole scholars who made it clear that the political environment plays an important role in the economic growth path, ([Lensink et al., 1999](#)), ([Lensink, 2001](#)).
6. Social-cultural factors have recently found their way into economic growth research. A wide range of empirical studies, ([Granato et al., 1996](#)), ([Zak and Knack, 2001](#)), ([Barro and McCleary, 2003](#)), confirm both positive and negative contribution of some social factors in economic growth. While other studies find indirect and unclear contribution for some factors such as trusting economies, ethnic diversity, religion beliefs, languages and social/ethnic conflicts.
7. Institutional structures impact not only economic growth itself but also other growth determinants. That is why [Easterly and Easterly \(2001\)](#) state that none of the traditional factors would have an impact on economic performance if a stable and trustworthy institutional environment had not been developed. Some of institutional structure components are: property rights, and regulatory institutions, where the most used ones are: the corruption, property rights and the role

of law and bureaucratic quality, (Knack and Keefer, 1995).

8. Numerous geographic variables are also examined, where some of them have proved to have a direct effect on growth. Examples of such variables: natural resources, climate, and topography (Armstrong and Read, 2004), (Masters and McMillan, 2001), where no significant impact is found of geography on growth after controlling for institution (Easterly and Levine, 2003).
9. Population growth rate is the most explored demographic factor in both theoretical and applied economic growth studies. Another investigated factor is migration and age by Kelley and Schmidt (1995). However more demographic aspects have not been explored yet. Population structure could influence economic growth through its impact on some determinants of growth such as investment and saving rate.
10. Macroeconomic conditions and economic policies: these two factors have been the center of interest for scholars since determining the framework within which economic growth takes place, (Barro, 1991), (Easterly and Rebelo, 1993), (Barro and Sala-i Martin, 1995). A stable macroeconomic environment is a necessary condition favoring economic growth, some macroeconomic factors were revealed to be more important for growth than others such as inflation, fiscal policy, budget deficit and tax burdens (Petraikos et al., 2007). Similarly for economic policies, they influence economic aspects through investment in human capital and infrastructure.
11. Workers' skills and acquisition through education and training system represent human capital. Human capital is the engine of growth in the endogenous economic growth theory, measured in different ways, human capital contributions have been confirmed by Barro and Sala-i Martin (1995), Barro (1991), and Mankiw et al. (1992a).

Applied contributions have been facilitated thanks to the remarkable development in databases as larger and richer bases are now available. Nonetheless, this evolution has included further difficulties in estimating growth determinants, for instance, many of the 1975 – 1996 growth determinants according to World Bank income data turn out to be irrelevant once Penn World Table (PWT) data are used instead, as World Bank adjusts for purchasing power parity using a slightly different methodology. Not only but also each revision of 1960 – 1996 PWT income data brings substantial changes in terms of growth factors, (Ciccone and Jarociński, 2010).

Advanced statistical techniques have been put into service through innovative programs that are capable of handling huge datasets by applying improved econometric techniques. Now, scientists possess more tools and resources in order to explore with

higher precision and confidence all potential economic growth factors across available data. Multiple estimation methods have been explored through in the past, for example: [Levine and Renelt \(1992\)](#) use the *Extreme Bounds Analysis (EBA)* in order to conclude the robustness of growth determinants. To the same end [Sala-i Martin \(1997\)](#) constructs a weighted average of OLS coefficients. Later, to determine growth determinants [Sala-I-Martin et al. \(2004\)](#) employ the *Bayesian Averaging of Classical Estimates*. For the same objective, [Ley and Steel \(2009\)](#), and [Moral-Benito \(2012\)](#) apply the *Bayesian Averaging Model*. This variety of estimation approaches has led to a variety of results which sometimes worsens and complicates the situation. For all of the aforementioned factors, results are not conclusive. Our goal in this chapter is to look through a defined number of published articles for the most present growth determinants in the last years that are worth shedding light on in any subsequent empirical study.

2.3 Methodology: textual analysis tools

The high attention and efforts scarified by scholars on economic growth and its determinants have led to produce a huge number of empirical and theoretical papers. Therefore, since a text represents a multidimensional world with many independent entrance, and data is not only stored in structured databases, it becomes crucial to extract the current events and conclusions from the large compilation of texts. Then it is much richer than any natural science because it contains an infinite number of non-material dimensions, ([Popescu et al., 2009](#)). Thus, textual analysis comes to offer a rapid and efficient solution that describes the content, structure and function of the messages contained in texts. Textual statistics consists of seeing the document through a prism of numbers and figures, producing information on the frequency counts of words ([MacMurray and Shen, 2010](#)). Thereby, textual analysis uses qualitative information stocked in texts to extract quantitative conclusion. This tool has been tested by [MacMurray and Shen \(2010\)](#) to detect business and economic statistically significant events in the corpus of the New York Times. Textual analysis is the method used by communication research to describe and interpret the characteristics of a record or visual message. Lexical analysis has been used firstly for statistical goals, like quantitative study of text vocabulary, words distribution form [Guérin-Pace \(1997\)](#). There are four approaches to textual analysis: rhetorical criticism, content analysis, interaction analysis, and performance studies. In this chapter we use the content analysis to capture and interpret the occurrences of specific messages and message characteristics embedded in texts, ([Frey et al., 2000](#)).

The well-known tools in textual analysis are: *classification* and the *correspondence analysis(CA)* of the lexical contingency tables. Correspondence analysis has been firstly used in linguistic data to analyze the sequence of words, (Reinert, 1993). CA is an explorative and model-free approach providing a general graphic overview of the structure of the association among keywords, years and keywords as in Trevisani and Tuzzi (2014) study, where the occurrences of keywords are transformed into coordinates on the multidimensional Cartesian system, (Murtagh, 2005).

Information provided by factorial analysis are completed by the global vision of the classification tool. The *Hierarchical Descending Classification (HDC)* proposed by textual analysis, was developed by Reinert (1987). This classification aims to gather homogeneous individuals/texts in groups known as classes as well as separate classes with contrasting profiles. Each class is represented by pertinent textual occurrences. HDC Classification enables the analyst to have a clear visualization of co-occurrence relationship between every class of words. HDC has three forms (Baril and Garnier, 2013):

- 1. Text simple classification: it allows the consolidation of similar texts, without changing the entirety.
- 2. Text segments simple classification: where classification is applied on text segments.
- 3. Dual classification on the combinations of text segments: here, the classification is carried on two tables where the lines are no longer text segments rather the combination of text segments. The same treatment is done twice after changing the number of active forms by the combinations of text segments. Since, the choice of analysis units is a delicate step and one that depends on the analyst objective, who should also have prior information about his texts which is rarely acquired, (Reinert, 1995). Repeating analysis twice on texts units before and after regroupment in order to capture how the choice of analyzing units could influence results, represents in some way a robustness analysis. Notably, for this reason we went to implement this form of classification in this chapter.

Over the past decades, text analysis tools have seen a computerized development, where new programs have been made to this end. In general, those programs are categorized in three types as suggested by (Bauer (2003), and Corman et al. (2002)):

- (i) *Positioning*, where keywords position in the context matters,
- (ii) *inferential*, where significant meaning is assigned to keywords in outside of their context,
- (iii) *representational*, that focuses on the most important words in the text.

Alceste developed by Max Reinert, is one of most popular textual analyzing software, *IRaMuteQ* also is a software that allows texts analyzing freely based on R statistic soft-

ware.

2.3.1 Collecting data procedure and analysis factors

Collecting steps

This subsection is divided in two parts. In the first one we describe how we collect data. In the second, we present corpus characteristics as shown by IRaMuteQ.

We have mentioned earlier that our interest is growth determinants, particularly human capital. That is why we build up a corpus, that its components are the abstracts of articles whose subjects are linked to economic growth and its factors. The data collecting approach is defined by the following steps:

First Step: we looked for articles having certain keywords in their abstracts such as: "economic growth" and "determinant", while others are more general like: "economic", "countries", "determinants". Since many studies compare economic growth across countries or look for economic growth origins for groups of countries, we look also for *origin and countries* in keywords. Thus the research procedure has the following order:

1. At the top of the most interesting papers are those who have each of the following in their abstract: *economic growth, determinants* in their abstracts.
2. Then: *economic, countries, determinants, but not economic growth*.
3. In the third place abstract with words: *Growth, countries, origins, but not economic growth or determinants*.
4. Followed by: *economic growth, origins, but not determinants*.
5. Then we added some of the abstracts from articles about growth determinants that we used as references within this thesis.
6. Finally, we extended data set by adding some articles cited in [Bos et al. \(2010\)](#), article titled: *Do all countries grow alike?*. The reason behind choosing this article is that Bos studies the differences of growth process and its factors in a cross countries set.

Pooling abstracts of all previous results in one text, gives the actual corpus to analyze⁴.

Journal's impact factor is a proxy of the relative importance of a journal through its field. Thereby, **the second step** concerns publishing journals, where we restrain selecting papers to those we could obtain through their 2013 2-year journal impact factor

4. We found almost all articles in *Business Source Complete* database. <https://www.ebscohost.com/academic/business-source-complete>

(2-*JIF*). The 2-year impact factor published by Thomson Reuters in the Journal Citation Reports (JCR) is defined as the average number of references to each journal in a current year with respect to 'citable items' published in that journal during the two preceding years (Garfield, 1972).

Classed in ascending order, table (2.1) contains 2013 impact factors of 15 journals for which we have could obtain their impact factors.

Journal Name	2013
Journal of Macroeconomics	0.621
Journal of International Money and Finance	1.09
World Bank Economic Review	1.095
International Economic Review	1.415
Journal of Econometrics	1.533
Journal of Monetary Economics	2.065
Journal of Development Economics	2.411
Ecological Economics	2.517
The Economic Journal	2.59
American Economic Journal: Macroeconomics	2.887
Journal of Economic Growth	3.125
American Economic Review	3.31
Journal Of Political Economy	3.617
Quarterly Journal of Economics	5.966
Journal of Economic Literature	7.56

Table 2.1: Journal Ranking by Impact Factor

The journal with the highest citation number in the two years preceding 2013 with effective of (7.56) times is the *Journal of Economic Literature*, this journal's topics are Economics, Econometrics and Finance. While the lowest cited articles in the sample we have collected, have been published in *Journal of Macroeconomics* which has an impact factor of (0.621) classed the last in table (2.1). The interest field of (15) selected journals are variant, some of them have more specified items such as: *Journal Of Political Economy*, *Journal of International Money and Finance*, *Journal of Econometrics*, while others are more general with widen topics like: *The Economic Journal*, *Journal of Economic Literature*, *Quarterly Journal of Economics* . Having these journals in the study adds more credibility to final results that will be discussed in section 2.4. The choice of 2013 impact factor is due to the difficulties of obtaining the whole list of economic journal ranking by impact factor on the web, because of the multiplicity of published journal classification terms. At the end, we could build a corpus of (65) texts, where articles' authors and titles are listed in the Appendix B, tables (B.1,B.3). **The third** and last key criterion that we have respected in constructing the corpus is when the articles were published, the chosen

articles cover the period between 1990 and 2014.

Analysis factors

Thereafter, the corpus analysis would be conditioned using the three following variables: publishing journal, period slice, and citation quartiles. In table (2.1), we have already exposed publishing journals, abbreviations are shown in table (B.4) in the appendix B. In order to facilitate result interpretation, publishing periods have been categorized in (5) subperiods each, except the last group, of five years, see table (2.2).

Periods	Covered period
1	1990 – 1995
2	1996 – 2000
3	2001 – 2005
4	2006 – 2010
5	2011 – 2014

Table 2.2: Time's Division

For citation number, we compute the average of each paper per year between 1990 – 2014, then we split them into four quartiles. The first quartile contains the (25%) less cited articles in the whole set, the last one has the (25%) most cited articles.

2.3.2 Corpus data statistics

First of all we present some general statistics about corpus texts. Table (2.3) illustrates the number of collected articles per: journal, year, and citation range. Further information like article's title, citation number and author are showed in tables (B.1, B.3) in the Appendix B.

The majority of selected articles are published in the between 2000 and 2014, whereas only (24.62%) of papers belong to an earlier period, which reflects that corpus content is recent and has a great influence depending on the criteria of the impact factor. Table (2.4) shows the (8) journals containing the most cited published papers, in the rang of (24.25 – 492.36), through study period.

Publishing Journal	Number	Period	Number	Citation Range	Number
J-Dev-Eco	11	1990-1995	10	0-1.8	17
J-Macr	9	1996-2000	6	1.9-5.64	16
J-Eco-Lit	8	2001-2005	19	5.65-24.24	16
J-Eco-Gr	7	2006-2010	18	24.25-492.36	16
A-Eco-Rev	7	2011-2014	12		
W-B-Eco-Rev	4				
Ecol-Eco	4				
Q-J-Eco	4				
Eco-J	2				
Int-Eco-Rev	2				
J-Mon-Eco	2				
J-Ect	1				
J-Eco-Dev	1				
J-Int-M-F	1				
J-Pol-Eco	1				
Macr	1				
Total	65		65		65

Table 2.3: Articles' Summary

Journal name	Papers number
Quarterly Journal of Economics	4
Journal of Economic Literature	4
American Economic Review	3
Journal of Economic Growth	2
Journal of Monetary Economics	1
Journal of Political Economy	1
Economic Journal	1
Journal of Development Economics	1

Table 2.4: Journals order by article citation

At the head of table (2.4), we find two journals: *Quarterly Journal of Economics*, and *Journal of Economic Literature* highly noted in 2013, these two journals study field of economics is wide, in contrast to others such *Journal of Monetary Economics* or even *Journal of Economic Growth*.

2.3.3 Primary corpus description

The corpus is made of (65) texts, that are published in the (15) journals showed earlier in table (2.1). Barro's (1991) article titled *Economic growth in a cross section of*

countries and published in *Quarterly Journal of Economics* has the highest citation average per year of (492.36). While the lowest journal in terms of citation average, (0.08), was that published lately in 2014 by Pops Alexander in the *Journal of Development Economics*, under the following title: *Credit constraints, equity market liberalization, and growth rate asymmetry*.

After some text manipulation, loading corpus to *IRaMuteQ* will provide further detailed description on lexical texts contents statistics, available in table (2.5):

Texts number	65
Segments number	238
Occurrences number	8476
occurrence mean by text	130.40
Lemmas Number	1602
Active forms number	1307
Supplement forms number	295
Forms mean by segment	35.61
Hapax number	771

Table 2.5: Corpus Description

From table (2.5), we observe that the (65) texts have (238) segments, (8476) words with on average (130.40) words per text. Occurrences are arranged in (1602) forms, (1307) are active and (295) are supplement. (771) words are characterized as hapax, which signifies that each word has a unique occurrence all over the corpus. The number of hapax increases with text length, also sometimes used as a measure of vocabulary richness, (Popescu et al., 2009). Out of all forms, the active form carries more sense than others also it transmits certain messages that authors want to pass to reader. Due to these forms, textual analysis can follow topic evolution and relative importance. Active forms can be either verbs, adjectives, adverbs, or nouns, table (2.6) shows some of them according to their appearance in the corpus.

Word	Frequency	Form	Word	Frequency	Form
growth	215	noun	model	44	verb
country	96	noun	increase	25	verb
determinant	51	noun	show	19	verb
development	46	noun	develop	18	verb
rate	36	noun	find	17	verb
effect	36	noun	relate	16	verb
income	34	noun	control	16	verb
paper	30	noun	examine	13	verb
economy	28	noun	suggest	12	verb
literature	26	noun	analyze	12	verb
long	19	adj			
cross	19	adj			
high	16	adj			
large	15	adj			
recent	12	adj			
positive	12	adj			
strong	10	adj			
robust	10	adj			
numb	9	adj			
low	9	adj			

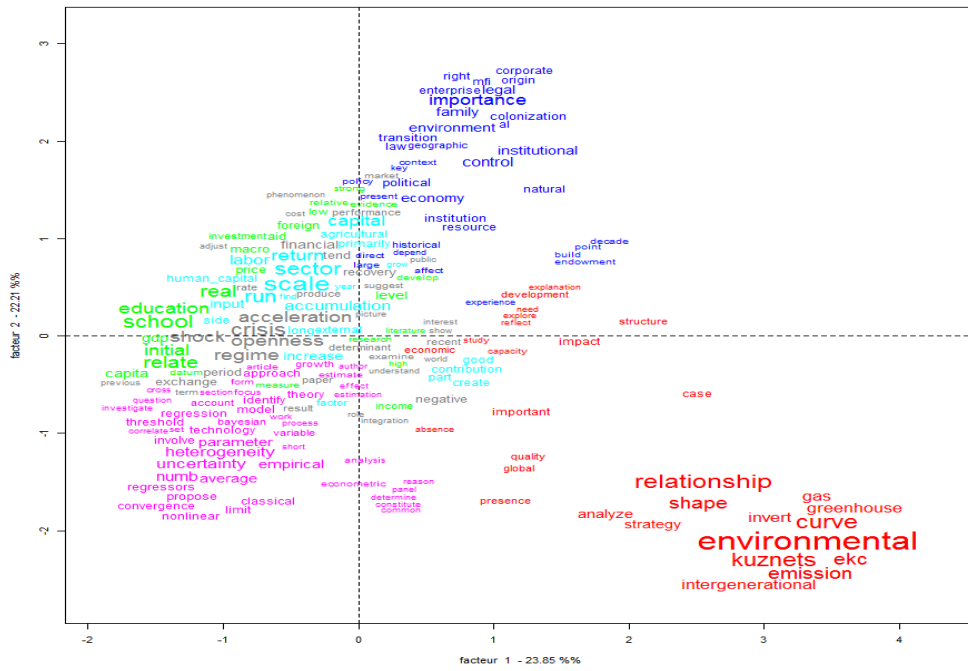
Table 2.6: Active Forms Frequency

At the top of the most cited nouns are the key words: *growth*, *country*, and *determinant*. In verbs, *increase*, *develop*, *relate*, *control*, and *long*, *cross*, *high* as adjectives. From these forms, one can see that articles' main subjects were turned towards countries growth determinants, and how to increase growth rates in interaction with the different aggregates cross country.

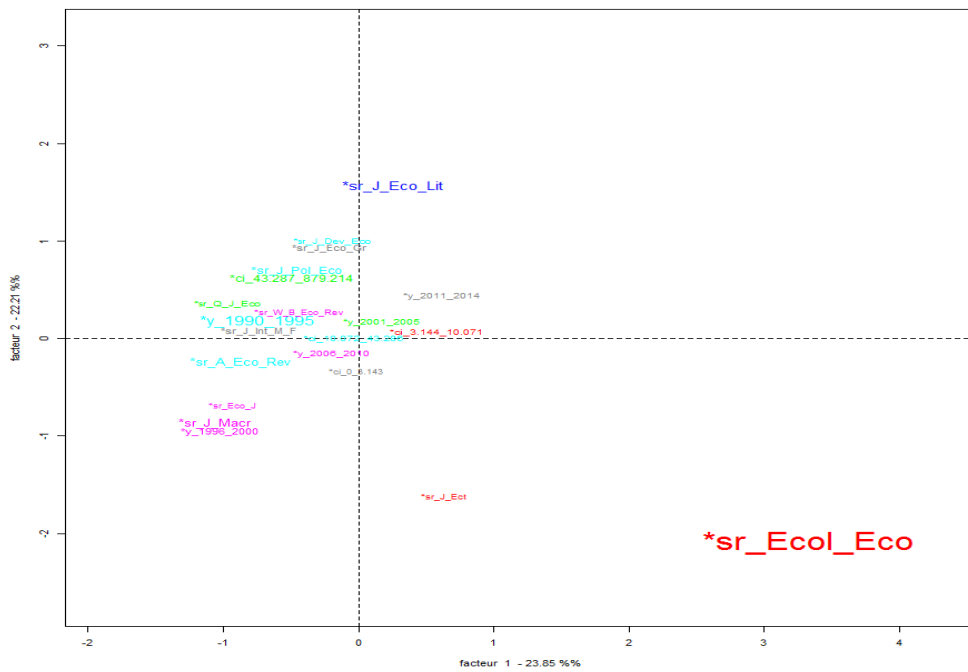
2.4 Results and Discussion

The main outputs consist of correspondence analysis figures and classes profiles, presuming content topics and characteristics. CA figure (2.1), represents the relationship between words on a map, where word mass represents the weight with the relative frequencies accorded to this word, while its place reflects the coordinate position. Distance separating words is the geometric representation of their own chi^2 values. Word colors express their homogeneity and belonging to the same group, namely *class*, thereafter all words of same color represent the visualized profile of each class.

In this figure (2.1), (23.85%) of inertia is in the first dimension and (22.21%) in the



(a)



(b)

Figure 2.1: Correspondence analysis of active forms, and factors

second, their sum means that the map shows only (46.06%) of the strongest patterns in the data. Further information is offered through the classification results, which show that only (123) text segments from a total of (238), representing (51.68%) of the whole corpus, have been analyzed. Reinert HDC dual classification has yielded (6) classes, each class owns a specific profile. Where each profile is overrepresented by a group of active forms, that identifies the major topic discussed in each class. The categorization of corpora texts in homogenous classes, where each class has its own overrepresented keywords⁵ is depicted in figure (2.2).

Each branch of the dendrogram has three classes, the left branch is slightly bigger than the right one where (50.4%) of forms are gathered in classes 1, 4 and 5, while classes 2, 3 and 6 composing the right branch have the rest of the forms. Classes in the same branch are supposed to have homogenous profiles or even common characteristics. Further information concerning classes profiles are shown in table (2.7).

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Publishing year	2011-2014 2001-2005	2011-2014 2001-2005	2001-2005 1990-1995	1996-2000	2001-2005 2011-2014 2005-2010	2005-2010 1996-2000
Citation range	1.9-5.64 0-1.8	0-1.8	24.25-492.36	5.65-24.24	1.9-5.64 24.24-492.36	0-1.8 1.9-5.64
Publishing journal	Ecol-Eco J-Ect	J-Eco-Gr J-Int- M-F J-Macr	J-Eco-Gr Q-J-Eco A-Eco-Rev J-Eco-Lit	Q-J-Eco A-Eco-Rev J-Pol-Eco J-Dev-Eco Eco-J	J-Eco-Gr J-Dev-Eco J-Eco-Lit	J-Ect J-Macr A-Eco-Rev Eco-J

Table 2.7: Classes Profiles

Table (2.7) of classes profiles show the characteristics of each class' texts in term of publishing year, publishing journal and citation category. The same information is also available in figures (2.3, 2.4 and 2.5), where variables modalities are associated with the different classes according to their χ^2 value.

5. For further representing keywords for each class, (6) graphes of words (B.1, B.2, B.3, B.4, B.5, B.6) are available in the appendix B.

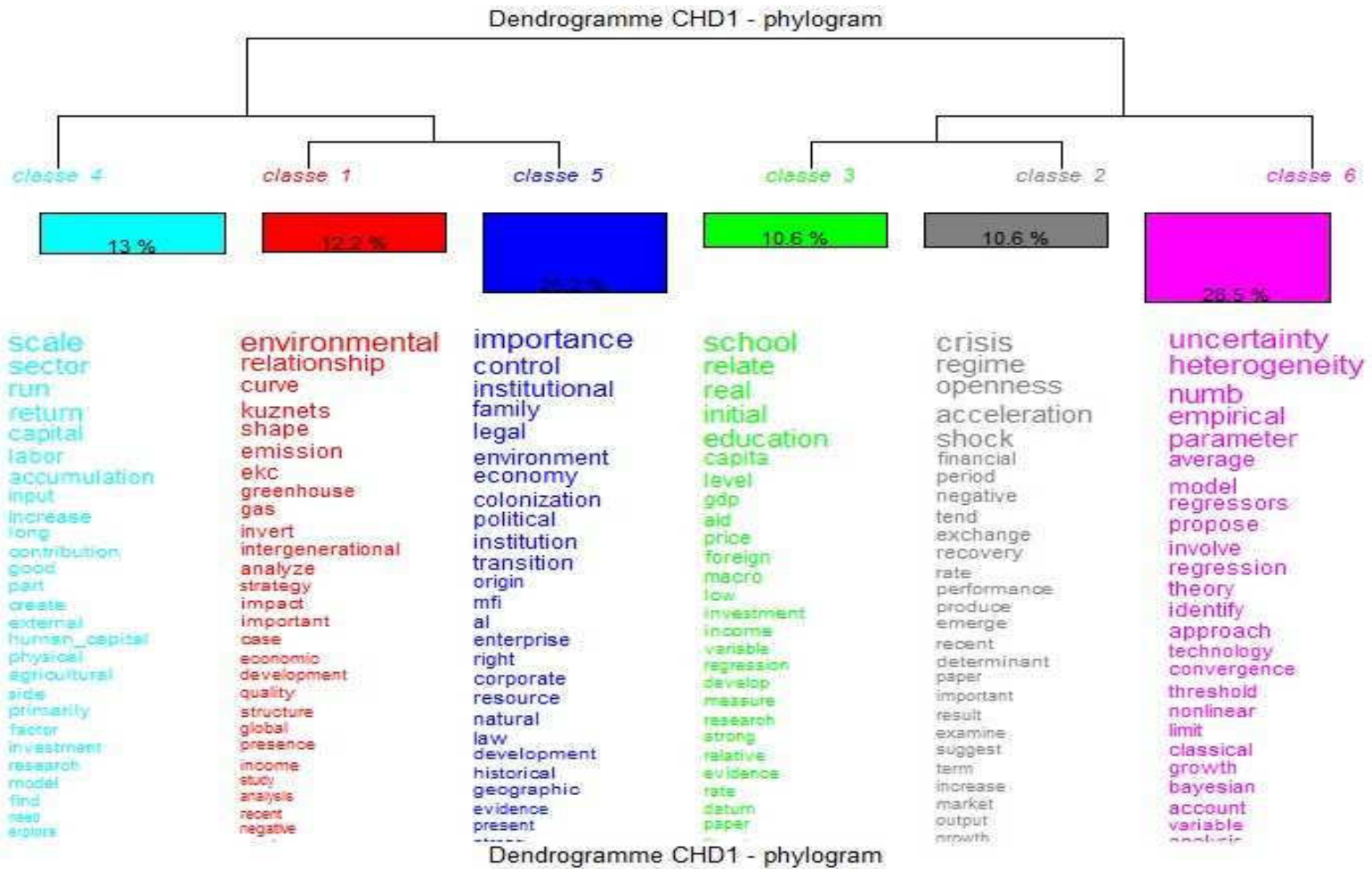


Figure 2.2: Hierarchical Descending Classes:Classification Dendrogram

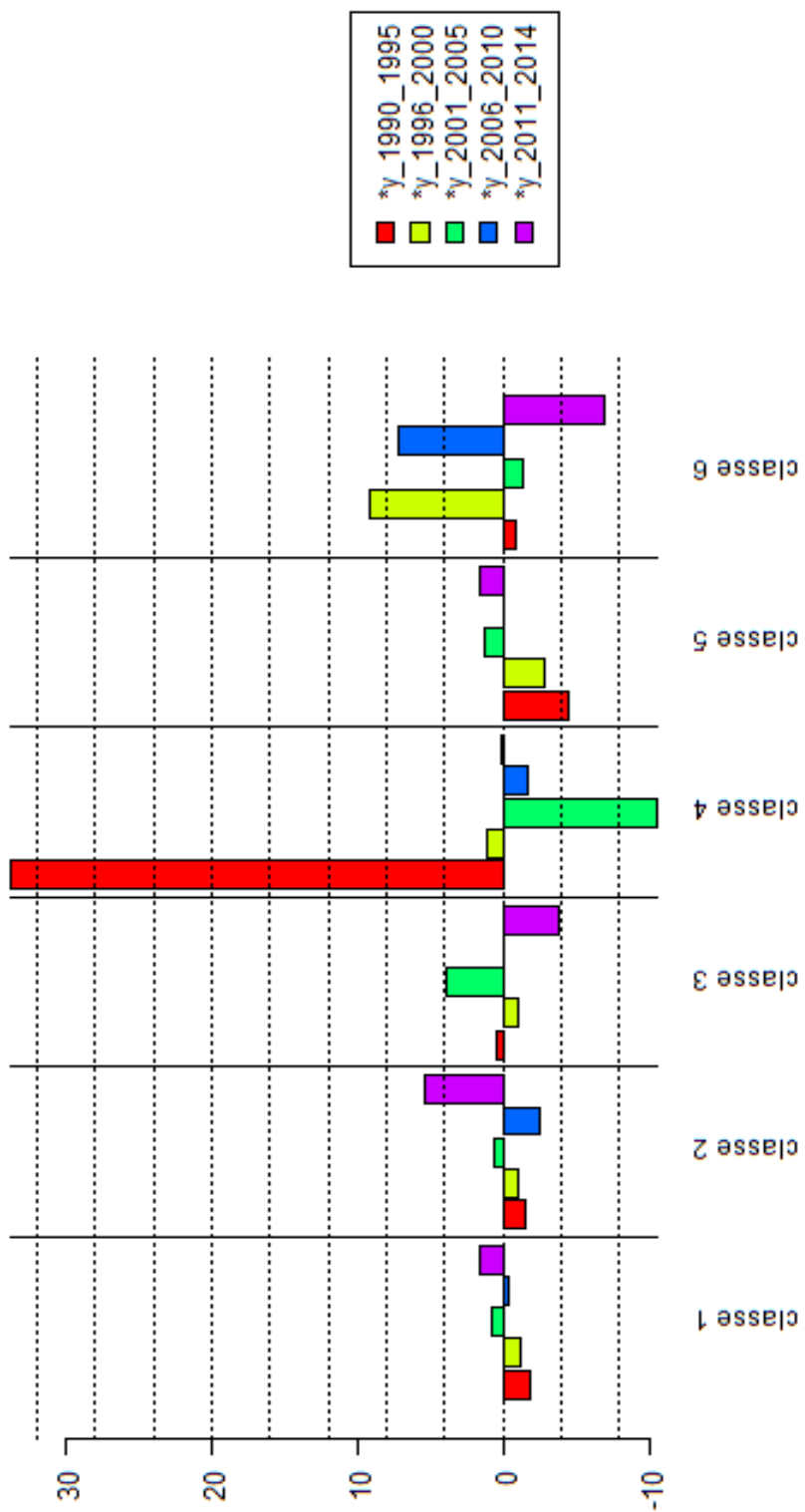


Figure 2.3: Publishing year's χ^2 by class

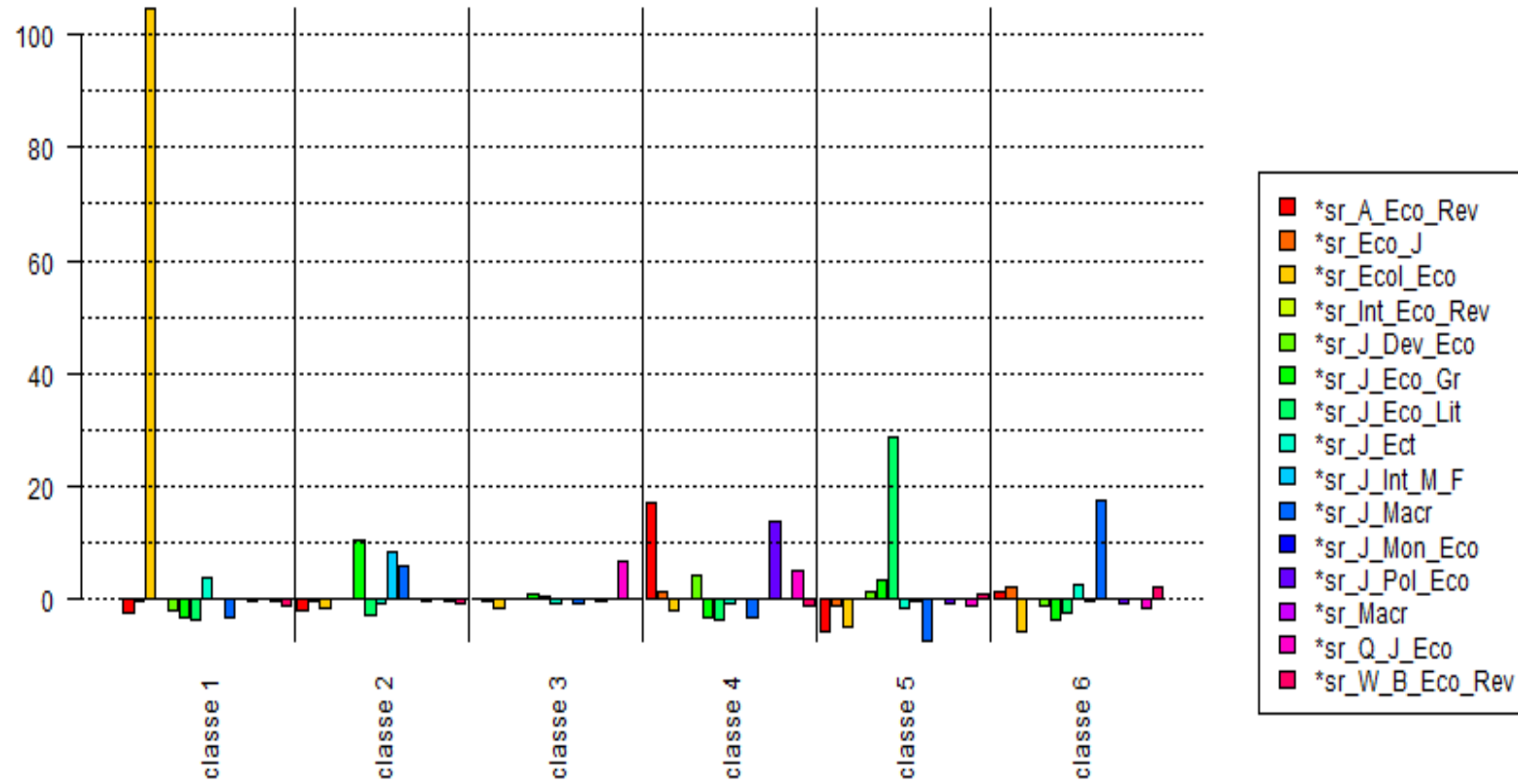


Figure 2.4: Publishing journal's χ^2 by class

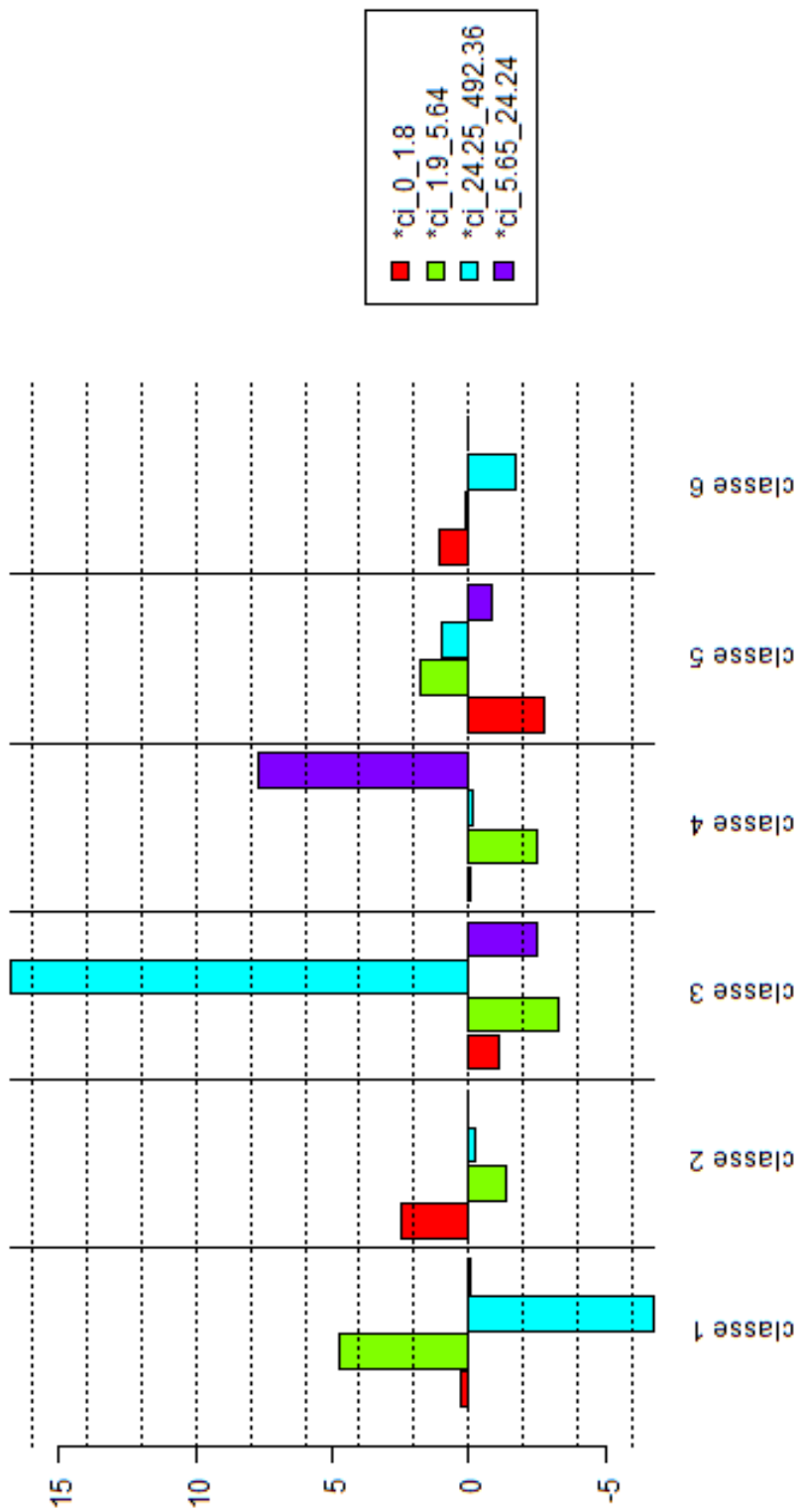


Figure 2.5: Citation number's χ^2 by class

Coming back to figure (2.2), the (6) classes are split in two distinct branches, each branch contains homogenous classes that have similar characteristics. The classification dendrogram in figure (2.2) represents a tree with two branches, each branch also has sub-branches. The classes' profiles are composed of forms of high frequency.

Here we describe the dendrogram shown in figure (2.2) from left to right.

- (i) The classification dendrogram has two branches, each branch is composed of two sub branches. The left sub branch represents class 4 which embodies (13%) of segments. The main inspiration of its overrepresented words can be how to increase economic growth and which factors to use, inspired from words like: *return, scale, increase, accumulation, contribution, labor, human capital and factor*. These keywords characterize the classical economic growth function and factors, consistent with controversial topics in the second half of ninetens between 1996 – 2000 published in journals such as: *economic journal, journal of development economic and quarterly journal of economics*. χ^2 measuring active forms depending on this class between (42.18 – 2.17) for this class.

The right side of this branch has an under sub branch with two classes which are class 1 and 5, supposedly closer to each other than to class 4 since all of these 3 classes are in the same branch. These two classes contain respectively (12.2%) and (25.2%) of total segments.

At the head of class 1 profile, we detect the presence of environmental terms like : *Kuznets curve, gas, emission, and greenhouse*. That corresponds to the fact that text segments analyzed in this class come mainly from the journal of Ecological Economics, where environmental and sustainable development subjects are at the head of discussion topics in the past few years. This has been confirmed from table (2.7) where all texts belong firstly to the period between 2011 – 2014 then 2001 – 2005. While other words like *analyze, analysis, and negative* concern econometrics side make reference to Econometrics Journal. Each text in this class is cited on average between (46 – 605). χ^2 reflecting active forms depending on this class are between (61.61 – 2.63).

The institutional aspects governing economic activities: *institutional, legal, and political* are more present in class 5. Texts in this class are extracted particularly from after 2000-to-2014 published works in classical economic journals: *journal of economic literature, journal of development economics*.

- (ii) The dendrogram's right branch also contains two sides, the right sub-side represents the sixth class that owns (28.5%) of analyzed segments making this class biggest. This class keywords such as *empirical, heterogeneity, parameter, nonlinear, bayesian and regression* reflect the wide range of empirical approaches invested in

macroeconomic subjects. Texts from the sixth class are mainly issued from *Journal of Econometrics*, *The Economic Journal* and *Journal of Macroeconomics*. The left side of this branch contains two classes 2 and 3 that identically have (10.6%) of all the classed segments. Subjects in this branch as shown are centered about economic performance and activities. More specifically, class 2 is concerned with economic performance and recovery during crisis and potential financial shocks *crisis, shocks, and financial*, that is justified by the intention of publishing journal *Journal of International Money and Finance*. Another indicator is that class 2 articles' are published between 2011 and 2014 which corresponds to the last financial crisis starting in 2008 that the world still going through.

While in the third class, economic performance is shown in function of factors like *school, education, price and foreign*, where schools are responsible of general human capital formation through education. Texts in class 3 are mainly issued from the *Quarterly Journal of Economics*, which also has the highest citation number making this class more valuable than others. Both classes treat macroeconomic subjects which can be shown by lexical terms like : *determinant, exchange, output, growth and market*. χ^2 for classes 2, 3 and 6 of this branch take respectively the following values (34.98 – 2.49) ,(34.98 – 2.24) and (21.51 – 2.21).

Classes in the left branch cover a variety of development economic subjects as *Journal of Development Economics* is highly present. Whereas, subjects dealing with economic growth and macroeconomic issues are more negotiated in classes of right dendrogram branch. Texts of this branch come from : *Journal of Economic Growth*, *Journal of Development Economics*, *Quarterly Journal of Economics*. Class 6 is the largest class that corresponds to the empirical nature of published papers where many estimation approaches can be implemented from a robustness aspect.

In table (2.8), we tried to give each class a label in function of its subject and characterization:

Class	Label
First Class	Environment and growth
Second Class	Financial crisis and growth
Third Class	Schools education and economic growth
Fourth Class	Economic growth factors, improvement
Fifth Class	Economic growth and institution
Sixth Class	Empirical approaches, econometrics

Table 2.8: Classes' labels

Thus, in line with [Trevisani and Tuzzi \(2014\)](#) we carried out a textual analysis study

on a corpora of abstracts. Our main study tools were CA and hierarchical descending analysis, whereas in Trevisani's study, they used CA with model-based clustering on titles corpora, (13) clusters have been found with one garbage cluster. We have obtained (6) classes identifying the different research streams between 1990 – 2014, the sixth class being the biggest and representing the various econometric tools and approaches that have been explored through all the corpus.

2.4.1 Human capital context in growth literature

The renewable subject of economic growth factors has resulted in a new stream of growth theories namely endogenous growth, where growth key engine is human capital *stock, accumulation*. The seminal work concerning human capital role in economic growth is due to works of endogenous growth theory (Change, 1990). Following the trace of Romer at the beginning of 90's, other scholars have also been interested in studying and especially measuring human capital (Barro, 1991) and (Barro and Lee, 1993). Among other researchers, Barro was trying to find credible measures of human capital factor, while Hénin and Ralle (1993) in their contribution of new growth theories studies, focus on human capital characteristics. Then in the twentieth century human capital studies have known further evolution.

Since we aim to position human capital in the mass of economic growth papers, textual analysis graphs have shown us that human capital trace has been identified in two classes. Where human capital χ^2 of belonging to classes, denotes that class 3 and 4 lexicons have contained human capital in some way. For classes 4 and 3, χ^2 takes respectively (7.82) and (2).

In class 3 overrepresented keywords, there are words like: *school, education* which makes reference to human capital general formation that starts through education in schools. The accompanied context also contains words such as: *investment, level, laws, initial condition, income growth, etc.*, these words make reference to the link between investing in human capital formation and economic growth levels and initial conditions. Whereas in class 4, human capital exists amongst other growth determinants such as: *labor, accumulation, capital, physical, investments, etc.*, which might correspond to the period 1996 – 2000 key topics where representing texts that have been published. Henceforth, extracting the relationships relating human capital to another corpora words is disposable thanks to textual analysis specialized computerizing programs for instance; IRaMuteQ, figure (2.6) exposes co-occurrence graph of *human capital*:



Figure 2.6: Human Capital Words Graph

As figure (2.6) shows human capital is related to concepts like accumulation and investment, these concepts have long since been evoked in economic growth literature debates around the role of human capital, stock or accumulation. Another liaison with the word of *physical*, *primarily* which is based on the elementary role, in neoclassical growth theories, played by both of human and physical capital in growth production. Moreover, this link could emerge from the fact that investing in human capital needs an identified quantity of physical equipment. *Agriculture* evokes the strong relationship relating human capital to agricultural sector, as one of its important inputs.

Out of these 2 classes, class 1 was where we found references to durable economic growth and environmental aspects in words like *environmental*, *kuznets curve*, *development*, and *impact*, there is no trace of human capital or any of its possible proxies although investing in human capital is crucial in order to sustain growth with respect to the environment. This paradoxical absence of human capital merits more research in the future in order to be clarified.

2.5 Conclusion and limitations

Economic growth has preoccupied researchers for many years with the aim of offering a more comprehensive view of growth production formulation and sources. These efforts have resulted in a huge amount of growth literature that is not only non conclu-

sive but also composed of many complicated empirical research studies. There is a lack of a common consensus in order to offer a universal set of growth determinants each time the growth subject is studied. That is why in an attempt to overcome this difficulty, we deploy the textual analysis technique, that has been implemented for a long time in social studies, and recently in economic literature, to analyze the evolution of growth research streams focusing on growth determinants, mainly *human capital*. This technique is based on extracting quantitative results and hidden messages through the analysis of textual corpora. The application of *IRaMuteQ* on a corpus, whose content is extracted from economic growth specialized papers with a considerable impact factor, has enable us to identify the main themes associated with economic growth in these papers thanks to the creation of classes and then a further graphical representation of each class' words. Empirical results have revealed the presence of six different growth contexts and determinants, where human capital is the most represented determinant.

Through research work in this chapter, we take note of some limits during this study work that may have an influential impact on results:

- The first limitation, concerns articles publishing time. We had restrained the collection procedure to papers published from 1990 for two reasons. The first one, because empirical analysis that has been carried out in the following chapters is made on data going from 1990 to 2011. While the second, is related to the fact that extending collection time would imply a huge number of research papers thereby an enormous time and effort to realize this task. Without forgetting, that the latest years have known a rapid and wide evolution in research themes, especially in growth theories *the endogenous growth theory*.
- In order to have the most credible results in the last years, we had to limit research procedures to papers published in journals with an important impact factor regardless of the paper's citation number. That may eliminate some valuable articles published in other weakly noted or new journals. Where, each year many new journals start publishing with emerged and interesting topics.
- Our corpus has been built up by gathering the abstracts of (65) scientific articles. In general, abstracts could be built in a monotonic way, or even they might ignore some interesting information which would represent a potential bias source.
- In textual analysis, an important part of data analysis depends on the method employed by the researcher to read quantitative results of qualitative data. Consequently, final conclusion of such an analysis could be over or underestimated in relation to the researcher's expertise and knowledge.
- The computerized textual analyzing program investigated in this chapter is based on a corpus of texts in the same language, which lowers the potential number of

targeted papers and possibly even journals.

Then and most interestingly, our main findings based on abstracts analyzing, are summarized in the next points:

- Textual analysis approach is not very common in economic growth studies as we have remarked, or at least as far as we have observed. To this mark and regardless of the results, our work could be an original contribution at this point together representing a cornerstone for further studies. Using IRaMuteQ software to implement textual analysis provides corpus characterization, enabling simpler and a logical read of the corpus contents, principally in large text corpora case. IRaMuteQ also provides images of keywords clusters that show the synthetic content, and themes evolution by periods of five years. Seen the precious functionalities of using textual analysis in lexical studies, we also stress the importance, time saving, and the easy use of this method in analysis, especially in cases where study subject has a wealthy register like ours.
- The main output was the identification of six classes of growth interactions. Each class has its own characteristics, regrouping similar research themes.
- An interesting presence of human capital has been identified in (2) classes through the (123) classed texts segments. Compared to other growth factors, this reflects the ascending weight associated with human capital in growth literature since the early 1990's, in spite of the uncertainty characterizing the different growth determinants. Human capital is sufficiently important to be present in two growth streams as results confirm. To this respect, in any subsequent growth study human capital should be considered as a primarily growth factor.
- A paradoxical absence of human capital source of sustainable growth ([Change, 1990](#); [Lucas Jr, 1988](#)) in the class 1 keywords concerning durable growth and environment. This constatation is worth a more in-depth and careful study.

Textual data analysis is a prominent instrument to explore an economic discipline as it used to be in other disciplines, such in sociology. That because, this exploration tool enables the analyst to go through texts and further identify the creation and emergence of new subjects. For future research, it would be interesting to invest in a more representative corpus where the text collection is extended to other temporal dimensions, publishing journals, or even to resume. Focusing on co-citation in scientific papers, another interesting future study idea could be explored using the textual analysis tool. It would be to inspect the existence of a grid effect among economic article authors, the same aspect has been explored early in corporate strategy and industrial organization by [Boissin et al. \(2000\)](#). That because a network design a mapping of the most cited authors in a specific research field. Therefore, the co-citation concept refers to a statisti-

cal approach that allows to locate beside the most prominent works, authors of a specific research field, the intensity of association and the ramification entre researchers, ([Boissin et al., 2000](#))

Chapter 3

Estimating the heterogeneous effect of Education Expenditure and Public Debt on Economic Growth: A quantile regression with multiple imputation approach

Contents

3.1 Introduction	100
3.2 Debt-growth: endogeneity issue & proposed solutions	103
3.3 Data	104
3.4 Empirical strategy: model & estimation methods	107
3.5 Results and discussion	111
3.6 Conclusion	123

Building on a growth framework with the human capital accumulation, in this chapter we make an empirical contribution to the debate about the growth-debt and human capital formation relationship. The main contributions of this chapter to economic growth literature are twofold. Firstly, we implement a multiple imputation method using an Expectation Maximization Bootstrapped algorithm in order to complete unbalanced data set of (76) countries covering all continents over (22) years, improving data quality and inferences validity. Multiple imputation has been successfully done, resulting in improved data and generalizable results. Secondly, using a suitable method

to control for cross-country varying coefficient through linear model and endogeneity, we apply two different quantile regression estimates. Results suggest a negative link between public education expenditure and debt ratios with economic growth rate. Estimates present evidence that countries in the bottom tail of the conditional growth distribution have the most influential negative and significant impact of both public debt and education expenditure to GDP. Inversely, schooling average years have a positive and significant effect on GDP growth rates, countries at the upper tail hold the most influential effect. These results are proved through different estimation methods.

3.1 Introduction

Endogenous growth theory holds that economic growth is mainly the result of endogenous factors. For instance, it has been proven that the explication power of the two classical determinants of growth rates, namely: (i) physical capital accumulation (K) and (ii) labor force (L), represent only (50%). The rest unexplained portion has encouraged many economists to focus on other determinants (Schultz, 1961; Denison, 1962). In particular, these economists have found that human capital and the role of education in accumulating this capital are influential factors for growth. Although, the human capital role in economic growth has been generously analyzed since the early works of Adam Smith and Alfred Marshall.

There was a disagreement between economic scholar on the manner that human capital contributes in creating and enhancing economic growth. The first stream motivated by Lucas Jr (1988) and earlier Uzawa (1965) considers the changes in human capital stock as the main engine to improve growth. While the second stream represented by (Change, 1990) and others considers that the level of human capital is the key in economic development. According to the last stream, policy makers have to promote and support human capital accumulation in order to provoke growth (Temple, 2002).

To summarize these viewpoints, some research, mainly Lucas Jr (1988) and Change (1990), demonstrate that human capital is the most important input to sustain growth in production process. That is human capital has many channels able to influence economic growth. Firstly, human capital helps in creating economies of scale leading, thereby, to high output amounts with each further input unit as stated by Schultz (1988). Thus more investment in human capital accumulation will increase output quantity. Secondly, the human capital factor is a helpful element in explaining the differences between countries in per capita income. Thirdly, human capital can also promote economic growth production through its effects on the expansion of research and develop-

ment activities, being the source of any technological progresses responsible of raising other inputs' productivity (Jones, 1998).

The compulsory path for the creation and development of human capital, known also as general formation of human capital, is made through education. Thus education becomes the most important tool enabling governments to gain access in order to control their objectives in human capital formation and accumulation. To this aim, governments have to provide an educational system with sufficient funds in order to undertake its objectives, using any of fiscal policy instruments. As a consequence, almost all governments have this ultimate commitment to fulfil. Many reasons underlie the government's willingness to use debt as a financing source in developing as well as advanced countries, like for example: being in a declining national income period, war, a country's specific institutional characteristics, or even the government's desire to avoid the risk of sudden financial shocks. To this intention they contract further debt instead of creating market distortion by raising taxes levels (Barro, 1979).

Since using debt to finance public activities is risky, the subject have been the concern of many economists, (Reinhart et al., 2003; Reinhart and Rogoff, 2010b; Woo and Kumar, 2010; Checherita-W and Rother, 2012; Panizza and Presbitero, 2014; Herndon et al., 2014). Globally, authors have discussed and analyzed the different economic externalities associated with the use of debt in an attempt to establish a sustainable debt threshold. Because debt levels have continued the raising chain again in emerging as well as advanced markets especially since the last financial crisis in 2008, debt studies have enchainned research work this time by studying how debt may influence growth through its impact on human capital as Greiner did in Greiner (2008b). In an endogenous economic growth model with debt and human capital formed by education, where governments have to increase surplus as a response to rising debt, Greiner derives some implications. His main outcome was that the government should stabilize debt being neither too strict nor too weak, and invest sufficiently in human capital formation to make a sustainable growth possible.

In line with Greiner (2008b), by estimating an empirical growth model with cross-section time series panel data we explore the relationship between human capital and economic growth under indebtedness situation. Due to the composed nature of human capital as a growth factor it does not have a unique measure that is why there are several measures developed by international organizations and economists. These measures contain a variety of human capital proxies such as educational gauges, composite enrolment rate, average years of schooling, rate of academic achievement, education expenditure, literacy, health expenditure, etc. All these measures represent the different

compositions of human capital structure. As we aim to estimate the basic human capital formation, we approximate human capital by a set of variables (as it is done in the calculation of *HDI* (Human Development Index)). At first place comes each of public education expenditure and average years of schooling for person aged over (25) years, while the rest of proxies are for robustness respect : life expectancy, health expenditure and labor force.

Our main purpose in this chapter is empirical. Where using an empirical framework, we focus on the impact of model covariates on the whole conditional growth distribution. The growth model is augmented to contain human capital and public debt as a potential economic growth regressors. We contribute to the existing research in two points :

- (i) Estimates are carried on a large and balanced panel data set of (76) countries including countries with different development levels. It is common that observation for many variables, particularly for *debt*, are not available even in international database. Although in most debt studies authors either did not refer to this issue, or simply they restrain the sample to just observed values consequently limiting the generalization of results (Hineline, 2008). Ignoring missing data might bias results since observed ones could be different from the observed as denoted by Barnard and Meng (1999), or even results will lose efficiency (Horton and Lipsitz, 2001). Thanks to the *multiple imputation* method (Honaker and King, 2010), we overcome the issue of missing data and obtain a complete observation of 22 years of debt and other variables. Thereafter, each estimate accounts for missing values and their uncertainty. Thereby our contribution constitutes one of the rare growth studies handling missingness problems and estimating complete data sets.
- (ii) We build on a theoretical growth model where debt and human capital are potential growth regressors inspired from Greiner (2008b). To estimate model we use the approach of *Quantile Regression (QR)* which allows debt to react differently across quantiles then we can detect any parameters' heterogeneity along the conditional growth distribution. Further, we implement quantile regression approach to address some endogeneity problems, using *Penalized fixed effects (PFE)* (Bache et al., 2013; Abrevaya and Dahl, 2008). The use of this method in the empirical growth literature has been inspired by the works of Andini and Andini (2014) and Mello and Perrelli (2003). To our knowledge, this is the first attempt to study the heterogeneous effect of debt and human capital together on growth.

The estimation model does not have any prior assumptions such as the condition of sustainable public debt imposed in Greiner (2008b). The rest of this chapter is orga-

nized as follows: in Section (3.2) we discuss how we deal with the endogeneity issue. Section (3.3) provides the data set investigated and the multiple imputation method used. In Section (3.4), we present our model and main strategies for the methods of estimation and analysis of the obtained data. We present and discuss results in Section (3.5). Concluding remarks are given in Section (3.6).

3.2 Debt-growth: endogeneity issue & proposed solutions

The issue of regressors endogeneity is well known for economists studying growth empirics, particularly known in the debt-growth relationship, many sources are responsible for this problem as growth regressors are typically endogenous, measured with error and correlated with unobservable time-invariant country-specific effects influencing growth rates. However, in the presence of unobserved individuals' heterogeneity, it is difficult to isolate the causal effects of the various economic growth determinants.

Existing literature has already tried to address this endogeneity through different tools. Some of them like [Cecchetti et al. \(2011\)](#) have used lagged variables, while others like [Woo and Kumar \(2010\)](#) use internal instruments with *Generalized Method of Moments* (GMM) estimations developed by [Blundell and Bond \(1998\)](#), or even by instrumenting debt to GDP ratio by the average debt of partners countries in [Checherita-W and Rother \(2012\)](#). The previous approaches are not fully sufficient to resolve endogeneity, for example since debt series are persistent the use of lagged variables is problematic, which limits their validity as well as the internal instrument with the standard GMM, also the use of lagged variables is not evident to resolve endogeneity since it could come from a third variable that influences growth and debt at the same time. Moreover, with a small cross-sectional macroeconomic data sets, system GMM estimators are poorly suited ([Bond, 2002](#)). The resort to instrumental variables, as in [Harding and Lamarche \(2009\)](#), is also problematic and doubtful. Since it requires a careful selection under the strict exogeneity condition, it is highly exigent and rarely well applied. [Panizza and Presbitero \(2014\)](#) states also that the use of instrument variables is never perfect since it involves a trade off between efficiency and consistency. Treatment approach applied by [Checherita-W and Rother \(2012\)](#) is also problematic with the spillover effect of financial crisis and global shocks, ([Panizza and Presbitero, 2014](#)). Using panel data econometrics with unit specific fixed effects was always the most prominent way to overcome endogeneity, since this consideration accounts for time-invariant heterogeneity and attenuates the bias generated by omitted variables, beside reducing the deviation from the mean and amplifying the consequences of measurement error, ([Panizza and Presbitero,](#)

2014).

An alternative approach offered by [Mundlak \(1978\)](#) and improved by [Chamberlain \(1982\)](#) is the *Correlated Random Effects (CRE)*, that has been lately extended by [Abrevaya and Dahl \(2008\)](#) to a QR framework resulting a *Quantile Regression with Correlated Random Effects (CRE-QR)*. CRE-QR considers the unobservable term as a linear projection on the observable term plus the disturbance, which is not correlated with the observable terms. CRE-QR considers endogeneity as an unobserved heterogeneity problem without imposing any strict conditions or requiring any instruments, offering more robust results than those of PFE results and avoiding the implementation of sophisticated methods ([Bache et al., 2013](#)). The CRE estimator does not estimate the fixed effects but controls for the time invariant dependence between the fixed effects and a set of covariates x_i by linearly including time invariant CRE transformations of possibly endogenous time varying variables. In this work, we feel that it is to implement this approach in order to face debt-growth potential endogeneity.

3.3 Data

In this section, we present some description about data source, model variables, then we tackle data problems and solution.

3.3.1 Data sources

The data on which we carry estimation in this chapter is an unbalanced panel data containing information on (76) countries over the 1990 – 2011 period. Data on dependent variable and other covariates has been collected mostly from the World Bank data set and the other international data sets available on web. Annual average growth rate, measured at market prices and based on local currency, is the dependent variable on which we regress a set of (32) explanatory variables. Further information with statistics concerning the (32) variables used in this study is shown in tables (C.2 in Appendix (C) and (2.3) in Chapter (2) , also country list is available in table (C.1) in Appendix (C).

Due to the well-known problem related to the abundance of economic growth theories and the absence of a clear guidance to select growth factors, model selection becomes difficult and should be made cautiously in order to avoid the over-representing issue. There exists many growth models with different factor combinations, where each of these models has a probability to be true. Empirical literature has widely discussed

model uncertainty issue (Durlauf, 2001). For example, Durlauf et al. (2005) has found (43) theories of economic growth with (145) significant growth regressors. Each theory was at least significant in one study while the determinants of growth may be the same in all of these theories with differences in their importance degree. For this reason, Ley and Steel (2007) stresses the importance of paying more attention to the starting assumptions, whereas Ciccone and Jarociński (2010) sheds light on the differences in sample construction and its implication on regressors choice, i.e., the set of individuals included in data samples and the time dimension of each sample.

Thus to select model regressors we relied on the inferences of some economists that have, in their turn, been interested in addressing uncertainty issue. For instance, Levine and Renelt (1992) carried out a sensitivity analysis to distinguish valid from fragile inference on many growth empirical studies results. Or even Sala-I-Martin et al. (2004) by applying model averaging approach. Depending on these studies amongst others we have established a set of explanatory variables including (28) regressors, with further dummy regional variables. The inclusion of different regressors to growth regressions has been suggested by Solow in order to add predictable and supplement heterogeneity in the steady state.

This set of regressors is split into three categories: *interest*, *control* and *time-invariant* variables. Table (3.1) makes reference to empirical and theoretical growth papers by which we have been motivated to select regressors.

Variable category	Name	Source
Interest variables	public debt, education expenditure	(Greiner, 2008b)
	Gross fixed capital formation as investment proxy	(Levine and Renelt, 1992)
	Schooling average years	(Barro, 1991)
Control variables	Inflation rate, deficit, health expenditure	(Levine and Renelt, 1992)
	Military expenditure	(Aizenman and Glick, 2006)
	Life expectance, investment price goods	(Sala-I-Martin et al., 2004)
	Purchasing power parity	(Barro, 1991)
	Population growth rate, saving rate	(Kormendi and Meguire, 1985), (Solow, 1956)
	Exportation and importation as openness proxies	(Krueger, 1998)
	Land area	(Barro and Lee, 1994a)
Foreign direct investment	(Moudatsou, 2003)	
Time invariant	Initial income	(Barro et al., 1991)
	Dummy regional variables	(Barro, 1991)

Table 3.1: Variables selection

To table (3.1) we add further variables to control for different aspects between countries such as: agricultural land as % of total land area, rural population as % of total population, and current account balance, these variables have been used in other empirical studies, see for example (Sahin et al., 2014).

3.3.2 Missing data: multiple imputation solution

As mentioned above, the issue of incomplete data sets is a common encumbrance in the world of empirical economic studies which hampers scientists in their quest for obtaining unbiased results. Proceeding to casewise/listwise deletion limits the generalization of results. In public debt observations the portion of missing values is of (22.37%) and (27.09%) in public education expenditure.

Solutions exist to replace or delete missing cases, but since the seminal work of Rubin (1976), *multiple imputation (MI)* has become the most popular method in the last thirty years. Multiple imputation consists of imputing m times each missing value, then it fills every missing cell by one drawn from a number of possible values. The new value is supposed to be close to the expected real value. In Honaker et al. (2011), authors suggest to set $m = 5$, when the missingness portion is negligible. Imputation results are more robust than alternative methods to sparse data (King et al., 2001).

Since the MI procedure is based on *Expectation-Maximization with Bootstrapping (EMB)* algorithm, that generates estimates of missing data using the observed ones, it gives a good approximation to the optimal posterior distribution and valid inferences even with a very small m . To impute missing fraction, we use *Amelia II* which has been written by the simple and powerful EMB algorithm, as a consequent Amelia II never crashes, also it is faster than alternatives. Further Amelia II has many features to give more accurate and valid imputations for time-sires cross-sectional data, like allowing to include prior information that fosters prediction power of imputation code (Honaker et al., 2011).

D denotes the data matrix assumed to be multivariate normal¹, that is composed of missing and observed values $D = \{D^{obs}, D^{mis}\}$. The missing value \tilde{X}_{ij}^{mis} for the observation i and the variable X_j are imputed from a linear regression:

$$\tilde{X}_{ij}^{mis} = X_{i,-j} \cdot \tilde{\beta} + \tilde{\epsilon}_i$$

with β denoting the regression coefficients calculated deterministically from μ and Σ .

1. $D \leftrightarrow N(\mu, \Sigma)$ with a mean μ and a variance Σ

A crucial element that the imputation model should assume in order to improve imputes quality, is that it should contain as many variables as estimation model, (Meng, 1994). To complete missing values in our data we used a multiple imputation package *Amelia II* available in *R* program (Honaker and King, 2010), where we set $m = 100$. Imputation process has successfully created new values, henceforth we can run more confident estimations of inferences.

Although this method is easy to use in theory, in practice it has some drawbacks. It requires computational algorithms that can take many hours or even days to run and cannot be fully automated. As these algorithms rely on concepts of stochastic convergence, it requires an expert judgment to know at which moment of iteration program should be stopped (King et al., 2001).

Imputation code, and diagnostics figures are available in the Appendices (A).

3.4 Empirical strategy: model & estimation methods

3.4.1 Baseline model

We set up a simple, empirical growth model where human capital is one of growth engines. The baseline specification assumes a linear relationship between the annual growth rate and a set of explanatory variables. Model main variables are: human capital, a proxy of physical capital, and public debt. Debt specifically consists of general government gross debt. Human capital is approximated by education expenditure and education school years for above (25) year aged person to which we add gross fixed capital formation as a measure of gross net investment. Physical capital measured through the gross fixed capital formation. For a country i observed at time t , growth model has the following mathematical form:

$$Y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + \lambda S_{it} + \epsilon_{it},$$

where X is the vector of interest variables, Z vector of control variables approximating economic environment, S vector of time invariant variables and ϵ is the error term.

In order to capture output variations control, two more sets of regressors to include to estimation model successively, consequently model has the following structures:

1. Benchmark model : Benchmark model controls for interest variables without debt:

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it}, \quad (3.1)$$

$$GDP_{it} = \alpha + \beta_1 EDUC_{it} + \beta_2 SCH_{it} + \beta_3 GFCF_{it} + \epsilon_{it}, \quad (3.2)$$

where X contains three regressors: education expenditure $EDUC$, the education stock of years SCH , gross fixed capital formation $GFCF$.

2. **Model A:** In model A we add public debt as a % of GDP $DEBT$ to Benchmark model:

$$GDP_{it} = \alpha + \beta_1 EDUC_{it} + \beta_2 SCH_{it} + \beta_3 GFCF_{it} + \beta_4 DEBT_{it} + \epsilon_{it}. \quad (3.3)$$

3. **Model B:** Here, to model B we introduce the first set of control covariates aiming to capture economic environment:

$$Y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + \epsilon_{it}, \quad (3.4)$$

$$\begin{aligned} GDP_{it} = & \alpha + \beta_1 EDUC_{it} + \beta_2 SCH_{it} + \beta_3 GFCF_{it} + \beta_4 DEBT_{it} + \beta_5 INTEREST_{it} \\ & + \beta_6 PINV_{it} + \beta_7 MEXP_{it} + \beta_8 LABOR_{it} + \beta_9 UNEMP_{it} + \beta_{10} INF_{it} \\ & + \beta_{11} CACNT_{it} + \beta_{12} SOLD_{it} + \beta_{13} LAND_{it} + \beta_{14} AGR_{it} + \beta_{15} POP_{it} \\ & + \beta_{16} RPOP_{it} + \beta_{17} LIFE_{it} + \beta_{18} HEXP_{it} + \beta_{19} FDI_{it} + \beta_{20} IMPO_{it} \\ & + \beta_{21} EXPO_{it} + \beta_{22} REVENU_{it} + \beta_{23} SAVE_{it} + \beta_{24} PPP_{it} + \epsilon_{it}. \end{aligned} \quad (3.5)$$

where Z holds some control variables those are as shown in equation: interest rate ($INTEREST$), investment goods prices ($PINV$), military expenditure ($MEXP$), labor force ($LABOR$), unemployment rate ($UNEMP$), inflation rate (INF), current account balance ($CACNT$), primary deficit ($SOLD$), land area ($LAND$), agricultural area (AGR), population growth rate (POP), rural population percentage in total population ($RPOP$), life expectancy ($LIFE$), health expenditure ($HEXP$), foreign direct investment (FDI), the fraction of importation and exportation in GDP ($IMPO$, $EXPO$), general government revenue in GDP ($REVENU$), gross national saving in GDP ($SAVE$), parity power of purchasing and finally the part of saving in GDP (PPP)².

4. **Model C:** Model C represents extended version of standard model where we account for all regressors including time-invariant ones:

2. Variables are not in \log form, since some of them have a negative values or zero.

3.4. Empirical strategy: model & estimation methods

$$Y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + \lambda S_{it} + \epsilon_{it}, \quad (3.6)$$

$$\begin{aligned} GDP_{it} = & \alpha + \beta_1 EDUC_{it} + \beta_2 SCH_{it} + \beta_3 GFCF_{it} + \beta_4 DEBT_{it} \\ & + \beta_5 INTEREST_{it} + \beta_6 PINV_{it} + \beta_7 MEXP_{it} + \beta_8 LABOR_{it} + \beta_9 UNEMP_{it} \\ & + \beta_{10} INF_{it} + \beta_{11} CACNT_{it} + \beta_{12} SOLD_{it} + \beta_{13} LAND_{it} + \beta_{14} AGR_{it} \\ & + \beta_{15} POP_{it} + \beta_{16} RPOP_{it} + \beta_{17} LIFE_{it} + \beta_{18} HEXP_{it} + \beta_{19} FDI_{it} \\ & + \beta_{20} IMPO_{it} + \beta_{21} EXPO_{it} + \beta_{22} REVENU_{it} + \beta_{23} SAVE_{it} + \beta_{24} PPP_{it} \\ & + \beta_{25} GDP90_i + \beta_{26} POP90_i + \beta_{27} DEBT90_i + \beta_{28} AFR_i + \beta_{29} ASI_i \\ & + \beta_{30} AMR_i + \epsilon_{it}. \end{aligned} \quad (3.7)$$

S is composed of: GDP level in 1990, population volume in 1990, debt initial level in 1990, and three dummy regional variables (with Europe as the reference continent)³.

Estimates for model A and B are available in the Appendix (C).

3.4.2 Econometric approach: the semi-parametric method of quantile regression

Estimating the growth process for a group of different countries by implementing a common linear model is henceforth possible since estimation approach accounts for heterogeneous effects such as the semi-parametric methods of *Quantile Regression (QR)* where covariates parameters can vary across countries and response variable quantiles (Durlauf et al., 2005). Thereby, through linear model estimation method can account for nonlinearity in covariates relationship. In this context, QR allows to assess whether the impact of certain regressor on dependent variable is stronger (or weaker) when dependent variable is unusually high (or low). Thereafter as argued by Jawadi and Sousa (2014), this approach reveal a more complex and a richer dynamics of the relationship between model variables, than do the linear models.

In the empirical growth literature, different approaches have been explored. Since our focus is no longer to estimate covariates impact on dependent variable mean, we implement the quantile regression approach that enables us to identify the effect of explanatory variables on dependent variable at different points in the distribution.

3. During estimation, we due to remove some regional variables as they hamper the estimation process.

Thereby estimates parameters for a given quantile give the marginal change in dependent variable with a marginal change in one independent variable for countries of that quantile, (Mello and Perrelli, 2003).

Introduced by Koenker and Bassett Jr (1978), QR estimation method shapes the relationship between the explanatory variables and the conditional distribution of dependent variable. Application of QR has been started firstly in cross section data, then QR use has been extended to both time-series and panel data.

For a quantile α with $\alpha \in [0, 1]$, Quantile regression is a linear conditional quantile function:

$$Q_\alpha = (Y|X = \tilde{X}) = \tilde{X}\beta_\alpha, \quad (3.8)$$

where

$$\beta_\alpha = \arg \min_{\beta} \sum_{n=1}^N |\alpha - I_{(Y_n - X_n\beta < 0)}| (Y_n - X_n\beta). \quad (3.9)$$

Already applied to economic growth studies by Mello and Perrelli (2003); Ram (2008); Andini and Andini (2014), this estimation method offers many interesting technical advantages. QR is well-suited when the data has considerable heterogeneity at an individual level. Moreover, together with capturing the heterogeneity, QR assess how policy variables affects country according to their position along the conditional growth distribution, (Mello and Perrelli, 2003). Contrary to *the ordinary least square (OLS)*, this semi-parametric method does not magnify the impact of outliers on dependent variable, consequently it is robust against outliers (Mello and Perrelli, 2003). Further, Koenker and Xiao (2002) states that the quantile estimation is more appropriate than estimation of the conditional mean, that QR allows policy variables to affect conditional dependent variable distribution by different ways, either by affecting skewness or dispersion. Moreover, this method handles conditional distributions that does not have a standard shape, such as asymmetric or truncated. The use of QR has been extended to deal with different issues in many applications thanks to development led by many authors like in Koenker (2005); Chernozhukov and Hansen (2008); Bernini et al. (2004); Harding and Lamarche (2009) and Canay (2011). One disadvantage of QR revealed by Alexander and Lamarche (2011) is that the inclusion of a lagged variable into model variables is not straightforward.

Since growth data in general, including our case, is cross-sectional time series said longitudinal, then if each individual observation is allowed for specific location shift effect then the parametric dimension of the resulting estimation problem can be quite large. This concern has incited Koenker (2004) to develop specific version of the quan-

tile regression estimator to fit and respect the longitudinal nature allowing at the same time covariates to play a more flexible role. It is the *Penalized Quantile Regression with Fixed Effects (PFE-QR)*, which Koenker states as the most flexible and robust approach that is forged outside the Gaussian random effect framework. PFE approach performs better than other models when there are a high number of predictors, and when predictors are correlated together or with dependent variable (Lazaridis, 2014).

PFE-QR is based on minimizing a weighted sum of K ordinary quantile regression objective functions corresponding to a selection of K values of τ the quantile of Y , with user specified weights (Koenker, 2004). The slope coefficients of this objective function are τ dependent, whereas coefficients corresponding to the fixed effects are assumed to be independent of τ . The vector of fixed-effect coefficients is penalized by an l1 (lasso⁴) penalty term with associated penalty parameter λ , thereby these coefficients shrink toward zero.

Abrevaya and Dahl (2008) have developed another QR extension that handles endogeneity problem generated from the dependence between the time-invariant specific countries effects and one or more of model regressors. This technique includes a *correlated random effect* transformation of possibly endogenous time-varying variables, see also (Wooldridge, 2010). Henceforth, there is no need to look for robust instruments.

As we discussed early here, we also use the alternative approach of quantile regression with correlated random effects to address the well-recognized issue of debt-growth endogeneity. Thereby, we hold to implement both of PFE-QR and CRE-QR since each of them handles a specific data issue, then we compare results in respect to verify estimates robustness.

3.5 Results and discussion

In this section, we combine the advantages of QR approach and its extension with those of panel data in order to obtain a robust estimates. First of all, we sue variance inflation factor (VIF) to evaluate multicollinearity. The mean of VIF of all regressors equals ($8.543 < 10$) thus we conclude that we can neglect this issue. Tables (3.2) and (3.3) contain growth factors estimates reflecting their impact on the conditional growth distribution:

4. Least absolute shrinkage and selection operation LASSO, is a regression method that involves penalizing the absolute size of the regression coefficient. This method is convenient when dealing with high correlated predictors (Hastie et al., 2005).

Chapter 3. Estimating the heterogeneous effect of Education Expenditure and Public Debt on Economic Growth: A quantile regression with multiple imputation approach

Quantiles	τ	0.25	0.5	0.75	0.95
Benchmark Model	INTERCEPT	2.245 ^(**)	2.633 ^(***)	3.815 ^(***)	9.705 ^(***)
	EDUC	-0.342 ^(***)	-0.263 ^(***)	-0.199 ^(***)	-0.418 ^(***)
	SCH	-0.143 ^(*)	-0.137 ^(**)	-0.144 ^(.)	-0.141 ^(.)
	DEBT				
	GFCF	0.123 ^(**)	0.147 ^(***)	0.157 ^(***)	0.084 ^(*)

Significance code: ./*/**/***, at 1%/5%/10%/0.1%, respectively.

Table 3.2: Penalized quantile regression with fixed effects estimates, Benchmark Model

From table (3.2) representing Benchmark PFE-QR estimated parameters, we find that both of education expenditure as a % of GDP and average years of schooling have a negative impact on growth rates. Countries at the upper tail of growth distribution have the worst negative influence of education expenditure on growth, while countries at the middle in quantiles $\tau = 0.50, 0.75$ education expenditure influence is mitigated, all coefficient are statistically different from zero. Surprisingly, schooling years have a negative and significant estimates.

Estimating the role of human capital on economic growth only through education can be misleading because it overvalues human capital contribution and underestimates the role of total factor productivity (TFP) growth (De la Escosura and Rosés, 2010). That is why we use a battery of measures mainly: schooling average years and education expenditure, while the rest of measures: health expenditure and life expectancy labor force represent work experience embodied in human capital, are for robustness aim.

Quantiles	τ	Before MI				After MI			
		0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
Interest Variables	INTERCEPT	20.133(**)	15.001(**)	16.997(*)	15.847	8.309(*)	7.887(**)	7.581(·)	16.248
	EDUC	-0.586(*)	-0.458(*)	-0.032	0.335	-0.377(**)	-0.247(*)	-0.216(**)	-0.271
	SCH	0.033	0.026	0.023	0.313	0.120(·)	0.179(*)	0.184(*)	0.555(**)
	DEBT	-0.014	-0.012	0.007	0.016	-0.024(***)	-0.019(***)	-0.019(***)	-0.01
	GFCF	-0.419(**)	-0.208(·)	-0.142	0.172	-0.010	-0.054	-0.018	-0.037
Control Variables	CACNT	-0.459(***)	-0.233	-0.200	0.019	-0.064(*)	-0.116(*)	-0.110	-0.082
	UNEMP	-0.027	-0.023	-0.037	0.077	-0.048(·)	-0.023	-0.001	0.036
	SOLD	-0.005	-0.060	-0.013	0.089	0.142(***)	0.070(**)	0.082	0.079
	MEXP	0.164	0.18327	0.344(*)	0.967(**)	-0.240	0.113	0.175(*)	0.086
	HEXP	-0.49(·)	-0.408(**)	-0.387(*)	-0.123	-0.092	-0.131	-0.112	-0.231
	SAVE	0.496(***)	0.283(*)	0.282(*)	0.016	0.072(·)	0.124(*)	0.129(*)	0.169(·)
	LIFE	-0.167(·)	-0.117(*)	-0.152(*)	-0.107	-0.040	-0.049	-0.053	-0.153
	EXPO	-0.088(**)	-0.071(*)	-0.078(*)	-0.02	-0.024	-0.016	-0.013	-0.031
	FDI	0.00000	0.00000	0.00000	0.00000	$8.06e-12$ (*)	$1.04e-12$	$-3.58e-12$	$-9.148e-12$
	INTEREST	-0.073(*)	-0.049	-0.019	0.016	-0.016	-0.008	$-3.082e-6$	$-2.55e-4$
	IMPO	0.074(*)	0.084(**)	0.094(*)	0.044	0.030	0.029	0.022	0.033
	INF	-0.278(***)	-0.135(**)	-0.101(**)	-0.026	-0.01(***)	-0.004(*)	-0.003(***)	-0.002(·)
	LABOR	0.026	0.048	0.056	-0.028	-0.043	-0.006	0.019	0.050
	PINV	-0.021	-0.012	-0.01	-0.001	-0.014(**)	-0.010	-0.013	-0.007
	POP	-0.289	-0.162	-0.717(·)	-0.823	0.717(**)	0.344	0.219	0.460
	RPOP	-0.025	-0.024	-0.035	-0.022	0.002	-0.006	-0.007	-0.033
	PPP	-0.00001	-0.001	-0.0004	0.004	$3.181e-4$	$-8.213e-6$	$-4.583e-4$	-0.002
	REVENU	0.072(·)	0.024	-0.051	-0.194(**)	0.036(·)	-0.001	-0.021	-0.101(*)
LAND	0.00000	0.00000	0.00000	0.00000	$-4.515e-8$	$-55.340e-8$	$-5.705e-8$	$-6.505e-8$	
AGR	-0.011	-0.016	-0.021	-0.017	-0.001	-0.011	-0.011	-0.002	
Time Invariant Variables	ASI	-0.807	-1.198	-1.119	-1.401	0.981(·)	0.113	-0.217	-1.033
	DEBT90	0.006	0.001	-0.009	-0.029(*)	0.008	0.003	0.007	0.006
	GDP90	0.00000	0.00000	0.00000	0.00000(·)	$2.389e-15$	$-3.544e-16$	$-3.901e-15$	$-6.934e-15$
	POP90	0.00000	0.00000	0.00000	0.00000	$2.676e-09$ (·)	$2.948e-09$ (·)	$2.141e-09$	$1.674e-10$

Significance code: ./*/**/***, at 1%/5%/10%/0.1%, respectively.

Table 3.3: Penalized quantile regression with fixed effects estimates (Model C)

In table (3.3) after controlling for public debt in model, schooling average years coefficients become significantly positive through the complete growth distribution. The marginal effect of schooling years is higher in countries of the higher growth rates, where one supplement year will increase growth rates in $\tau = 0.95$ countries of (0.56%) of GDP against only (0.12%) in countries at the bottom quantile. The negative effect of education expenditure on growth does not change sign but have lower impact and loss significance at the highest quantile. Again countries with lower growth rates have the highest adverse education expenditure impact, where GDP growth rate decreases by (0.38%) with each further education expenditure increase in GDP. The highest adverse impact of education expenditure is in countries of the lower conditional growth distribution, where education expenditure accounts for (5.03%) of GDP at the same quantile.

This finding is not in line with studies about education expenditure such as [Baldacci et al. \(2008\)](#) where authors state that education and health expenditure have indirect positive and significant impact on growth. And [Lee and Francisco \(2012\)](#), where authors state that education expenditure plays a positive role in improving the accumulation of human capital in a long and short run, but this contribution is so sensitive to good governance. Whereas [Blankenau and Simpson \(2004\)](#), argue that the positive impact of education expenditure could be neglected once another growth determinant is negatively affected by adjustments. Also [Greiner \(2008b\)](#) stresses that the adjustment following debt use could act negatively and hamper human capital formation through education. Thereby, beside the Greiner's explanation, the negative coefficient of education expenditure could be explained either by the bad governance and overinvesting in education expenditure at the expense of another economic activities, misallocation or the mismanagement of these funds. Or simply, this can be explained by the fact that investment in education leads to influence economic growth after some time lag which should be taken into consideration by economists and policy makers as noticed by [Chandra \(2010\)](#).

For the rest of human capital proxies: *health expenditure, life expectancy, and labor force*, coefficients indicate non statistically different from zero a negative and heterogeneous impact of life expectancy and health expenditure⁵. Labor force estimate, before and after imputation, are not different from zero. Regional variable supposed to matter for economic growth in [Barro \(1991\)](#), here Asia dummy variable has a negative and significant impact at $P - value = 0.05$, which denotes that Asian countries have a specific effect that is associated with lower growth rates compared to all other countries in the

5. The main proxies of human capital, in this chapter, public education expenditures and the output of these expenditures represented by the average years of schooling. Those two variables are significant in estimates after imputation.

sample, whose estimates are not statistically significant.

Coming back to public debt, theoretical studies point out a negative relationship between GDP per capital growth rate and the public debt ratio to GDP. While some empirical research find evidence of a threshold effect of debt, for instance [Baum et al. \(2013\)](#); [Chang and Chiang \(2009\)](#) prove that it promotes economic growth in the short run when its level is bounded to strict limits that respect each country economic conditions like in the European Union. Also illustrated in a sample of advanced economies by [Reinhart and Rogoff \(2010a\)](#) and [Reinhart and Rogoff \(2010b\)](#), that there is evidence of a negative link between growth and debt when debt-to-GDP is higher than (90%).

Far from all these studies, debt estimates show a negative correlation between debt levels and growth rate at all growth levels, more notably in countries of the tail growth distribution where increasing debt by (1) percentage to GDP will lead to reduce growth rate by (0.02) percentage. All debt coefficients are statistically significant except that of the upper conditional growth distribution. Identically to education expenditure parameters, the highest public debt influence is in countries located at the lower conditional growth distribution that correspond to the fact that average public debt to GDP of this quantile is the highest (71.94) to GDP.

Testing for conditional income convergence through the inclusion of initial income regressor, shows positive estimates in all countries including less developed ones. According to [Mello and Perrelli \(2003\)](#) this result is explained by the fact that in large cross-section samples, countries do not have necessarily the same technological parameters.

The rest of the variables take their expected sign for example, price level of investment sign is negative in consistence with [Sala-I-Martin et al. \(2004\)](#), saving rate and population growth rate sign is positive confirming the evidence proposed by [Kormendi and Meguire \(1985\)](#), the negative inflation rate sign as in [Levine and Renelt \(1992\)](#), interest rate has a negative sign. For deficit, we find that it holds a positive and significant impact on growth, contrary to [Levine and Renelt \(1992\)](#) findings.

Controlling for the reverse causality between economic growth rate and debt by the addition of the initial level of public debt shows a positive contribution in all quantiles, especially in countries in the first quantile whose debt average is the highest compared to the rest quantiles.

Since growth regressors are of different nature and units, that implies comparing their impacts on growth should be made through standardized coefficient rather metrics. Table (3.4) presents standardized coefficients.

Quantile	0.25	0.5	0.75	0.95
EDUC	-0.146 ^(**)	-0.095 ^(*)	-0.083 ^(**)	-0.104
SCH	0.068 ^(.)	0.101 ^(*)	0.104 ^(*)	0.313 ^(**)
DEBT	-0.221 ^(***)	-0.164 ^(***)	-0.170 ^(***)	-0.09
HEXP	-0.047	-0.067	-0.057	-0.118
SOLD	0.151 ^(***)	0.075 ^(**)	0.087	0.084

Table 3.4: Standardized Coefficients of some variables

Henceforth, one standard deviation⁶ increase in public debt to GDP will reduce (0.221) to (0.09) standard deviation of GDP annual growth rates, and public education expenditure with (0.146) to (0.104) decrease in dependent variable standard deviation. Whereas an increase by one standard deviation of schooling years rises dependent variable by (0.068) to (0.313) standard deviation. After remark, standardized estimates show that both education and schooling effects are lower than the metric coefficient. In contrast, debt standardized coefficient is more influent.

3.5.1 Robustness tests

Checking the robustness of our results is a very important objective that we try to fulfill through the following robustness tests. Robustness tests applied here are split into two groups, on first hand we evaluate results robustness in relation to before imputation data. On the other hand, assessment is related to estimation method.

- Regarding the data missingness issue, we estimate data before imputation in a way to evaluate whether the imputation procedure has improved data by further inducing information that has been hidden with the missing fraction. Table (3.3) displays both estimates of complete and incomplete data⁷. Regression estimates on the right hand side in table (3.3) compared to left hand side completed data, shows that in general coefficients of interest variables as well as of control ones are the same at least in the bottom conditional growth quantiles $\tau = 0.25, 0.75$. Whereas, at the upper level of growth rates, regressor coefficients change sign and loss significance. For instance, that is the case for each of: education expenditure, schooling years, net investments, and deficit. Moreover imputation diagnostics denote that the imputation procedure has successfully replaced the missingness data by new ones close to the real unobservables, and after imputation interest

6. Standard deviations of following variables EDUC, SCH, DEBT, HEXP, SOLD, respectively are:(1.79, 2.62, 41.19, 2.37, 4.95).

7. comparison is made only for model C

variables estimates gain significance. We conclude that imputation has been a crucial step that we did not neglect, it has improved data quality and contributed in obtaining reliable estimates.

- Mean growth regression estimates: the first estimator is addressed to panel data, it controls for heterogeneity bias that might result from correlation between the country specific fixed effects and regressors affecting model consistency. The second model accounts for production function inefficiency, this approach is the stochastic analysis estimator SFA. Estimating a panel data goes through a battery of tests concerned to find the ad hoc model that fits data nature, tests and results are presented in table (3.5):

Model	Lagrange Multiplier1	Lagrange Multiplier2	Hausman Test
Benchmark	<2.2e-16	<2.2e-16	0.003
A	<2.2e-16	<2.2e-16	2.357e-06
B	<2.2e-16	1.03e-06	0.012
C	<2.2e-16	8.52e-06	

Lagrange Multiplier1 is used to detect whether data present a panel structure or not.
 Lagrange Multiplier2 tells whether data has individual fixed effects or temporal fixed effects.
 Hausman Test allows to know whether model effects are random or fixed.

Table 3.5: Model Specification Tests

Columns in table 3.5 contain identification tests $P - value$: Lagrange multiplier, and Hausman test (Hausman, 1978). According to first columns $P - value$, data represents a *panel structure* implying that a pooled cross sectional estimation is rarely a good idea when individuals have unobserved effects that may correlate with included variables (Bache et al., 2013). Second column $P - value$ tells that data set have been characterized with *temporal fixed effects (TFE)*. SFA and TFE estimation for benchmark and C model are reported in table (3.6):

Chapter 3. Estimating the heterogeneous effect of Education Expenditure and Public Debt on Economic Growth: A quantile regression with multiple imputation approach

Model	Benchmark		C		
	TFE	SFA	TFE	SFA	
Interest Variables	INTERCEPT	7.168 ^(***)		8.158 ^(***)	
	EDUC	-0.314 ^(***)	-0.280 ^(***)	-0.46 ^(***)	-0.43
	SCH	-0.243 ^(***)	-0.188 ^(***)	0.099 ^(*)	0.187
	DEBT			-0.025 ^(***)	-0.025
	GFCF	0.156 ^(***)	0.138 ^(***)	0.108 ^(**)	0.08
Control Variables	CACNT		0.021	0.023	
	UNEMP		-0.028 ^(*)	-0.021	
	SOLD		0.152 ^(***)	0.185	
	MEXP		-0.037	-0.156	
	HEXP		-0.077	-0.097	
	SAVE		-0.019	-0.001	
	LIFE		-0.03	-0.03	
	EXPO		-0.022	-0.032	
	FDI		-1.86e - 13	4.65e - 12	
	INTEREST		0.005	0.002	
	IMPO		0.02	0.030	
	INF		-0.033 ^(***)	-0.004	
	LABOR		-0.043 ^(**)	-0.044	
	PINV		-0.005	-0.008	
	POP		1.047 ^(***)	1.030	
	RPOP		0.002	0.001	
	PPP		-0.000	-2.33e - 05	
REVENU		0.026	0.019		
LAND		-9.32e08 ^(.)	-9.96e - 08		
AGR		-0.009	-0.009		
Time Invariant Variables	AFR		0.076	-0.353 ^(**)	
	AMR		0.482 ^(.)	0.568 ^(**)	
	ASI		-0.067	0.011 ^(**)	
	DEBT90		0.007 ^(.)	0.009	
	GDP90		2.38e - 15	1.92e - 15	
	POP90		2.93e - 09 ^(*)	3.23e - 09	
R	0.083		0.274		
Adj-R	0.082		0.266		

Significance code: ./*/**/***, at 1%/5%/10%/0.1%, respectively.

Table 3.6: SFA & TFE Panel data model estimation

Estimates of TFE and SFA on the mean of the conditional growth distribution from one side are similar to each other. Parameters estimates are in line with PFE-QR estimates for both of Benchmark and model C. But SFA estimates statistically do not differ from zero, and the TFE⁸ estimation method does not allow for varying parameters. Moreover it does not have the interesting advantages and the capacity to handle data issue as does the PFE with quantile regression. From all these facts, our estimation is better and results are robust. That has also been visually shown in figures (3.1, 3.2), PFE with quantile regression estimators allow

8. TFE estimation for models Benchmark, A and B are disposable in the appendix (C)

to capture the different quantities of explanatory variables at each growth level which is not the case in mean distribution methods.

Observing graphs in figures (3.1, 3.2) confirms that variables do not act identically in all countries as shown through standard mean estimation methods TFE,SFA, for example, contrary to mean estimation method the method of quantile regression allows us to detect a positive rising impact of schooling years in countries at the top of growth conditional distribution.

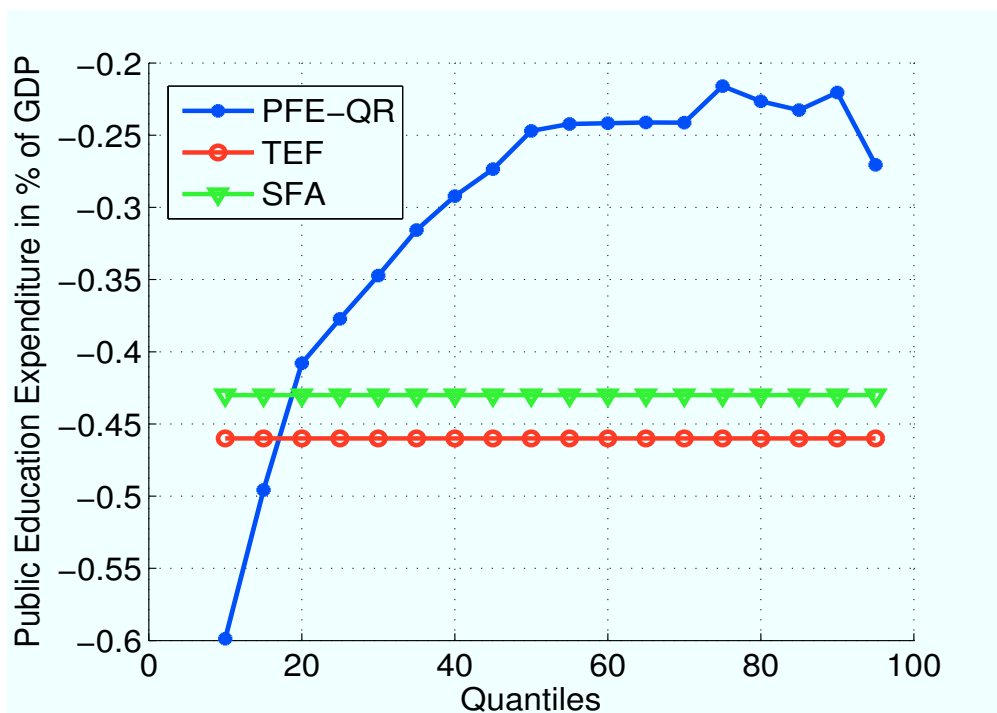
- Consistent with [Abrevaya and Dahl \(2008\)](#), we run CRE to control for unobservable heterogeneity that may alter other growth determinants impact. Estimates displayed in table (3.7), show some differences with PFE-QR results. Firstly, education expenditure and debt parameters have not greatly changed, they maintain their sign and direction and still significantly in the same range. While, not only schooling average years had no significant impact on economic growth, contrary to PFE-QR results, but also it's positive impact is attenuated. Equally as in PFE-QR, debt and education expenditure parameters loss significance in countries at the upper growth distribution.

Unlike [Panizza and Presbitero \(2014\)](#) who claim no evidence for a negative effect of debt on growth, and identically to many papers such as [Saint-Paul \(1992\)](#) and [Schclarek \(2004\)](#), mean and conditional growth estimates confirm that debt as a ratio of GDP hold a negative and heterogeneous impact on GDP growth rates through all countries, implying that debt impact is robust to different estimation methods even after controlling for debt-growth endogeneity issue, unless for the upper tail distribution. With heterogeneous and sizable effect, in countries with the highest average of debt to GDP ratio (71.941)⁹ and the lowest annual growth rates through (22), the impact of (10) percentage increase in debt ratio will decrease significantly growth rates by (0.24) per year.

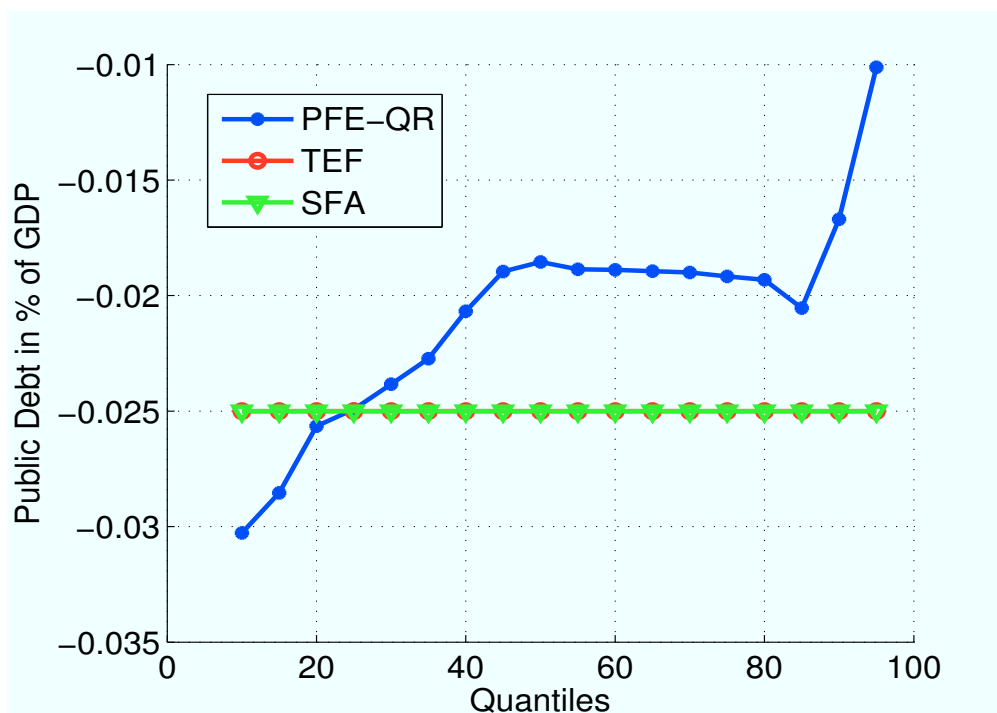
Human capital measured by education expenditure and the stock of schooling years for person aged above (25) year, has two different impacts. First, that of education expenditure effect, which is equally to debt impact is negatively and significantly associated with growth rates. The highest negative influence on subsequent economic growth is in countries at the bottom of growth conditional distribution. The stock of schooling years has a negative relationship with growth if the model does not account for debt to GDP ratios, while after including this indicator to model the impact of lagged education expenditure embodied in the average of schooling years creates a positive influence on growth rates that are statistically different from zero. Thereby, for countries¹⁰ with low

9. See table (D.4) in the appendix (D), for further information about more quantiles.

10. These countries average of schooling years is about (6.681) years and GDP growth rate average is the highest in the sample, $\tau = 0.95$

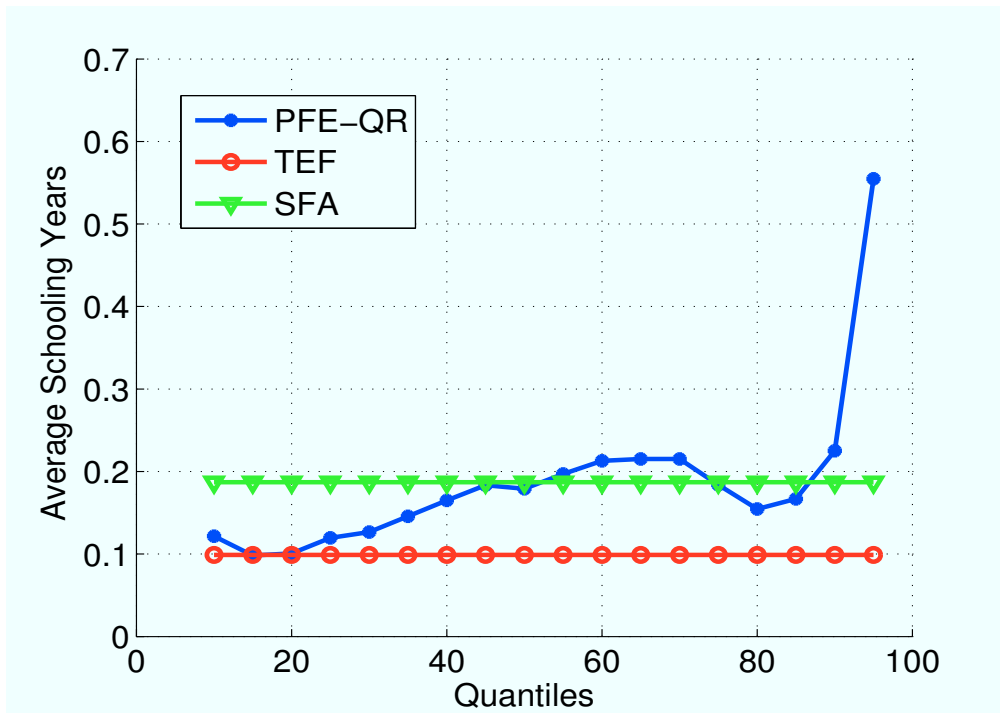


(a)

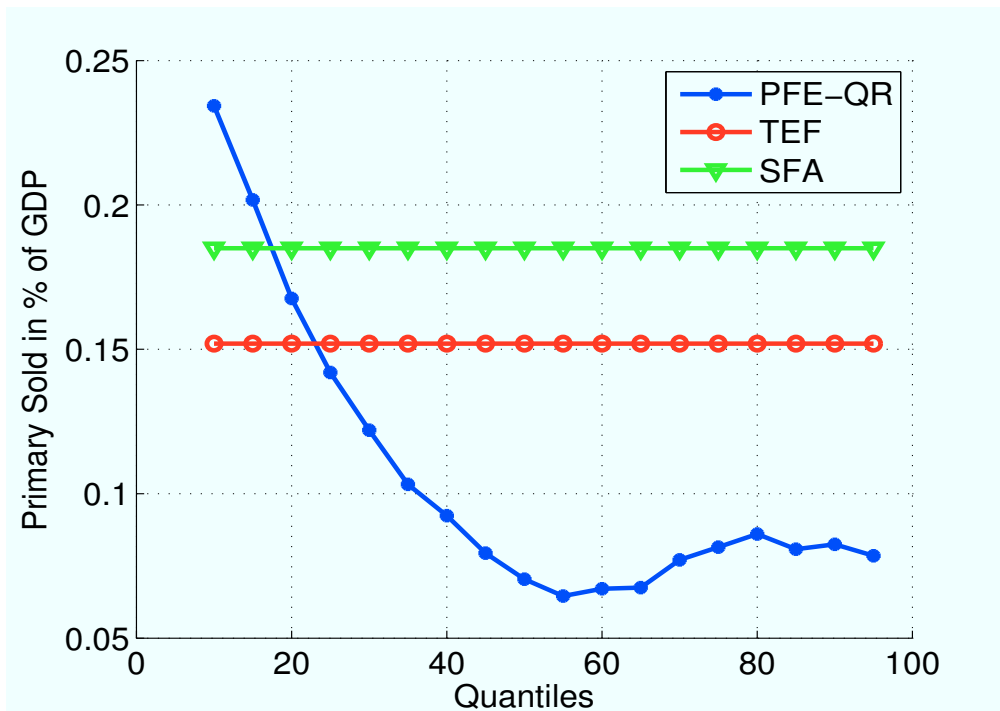


(b)

Figure 3.1: Comparison of parameters estimates I, PFE-QR, TFE and SFA



(a)



(b)

Figure 3.2: Comparison of parameters estimates II, PFE-QR, TFE and SFA

Chapter 3. Estimating the heterogeneous effect of Education Expenditure and Public Debt on Economic Growth: A quantile regression with multiple imputation approach

Quantiles	τ	0.25	0.5	0.75	0.95
Interest Variables	INTERCEPT	15.165 ^(*)	15.097 ^(**)	14.146 ^(**)	16.976
	EDUC	-0.284 ^(.)	-0.244 ^(**)	-0.215 ^(.)	-0.194
	SCH	0.167	0.128	0.153	0.155
	DEBT	-0.028 ^(**)	-0.014 ^(**)	-0.01 ^(*)	-0.009
	GFCF	-0.044	-0.038	0.023	0.031
Control Variables	CACNT	-0.121	-0.123 ^(*)	-0.093	-0.101 ^(.)
	UNEMP	-0.013	0.005	0.011	0.032
	SOLD	0.154 ^(**)	0.102 ^(***)	0.085 ^(**)	0.076 ^(*)
	MEXP	-0.322	-0.058	0.103	0.183
	HEXP	-0.135	-0.152	-0.155 ^(*)	-0.135
	SAVE	0.148 ^(*)	0.147 ^(**)	0.13 ^(*)	0.131 ^(*)
	LIFE	-0.081	-0.102 ^(*)	-0.108 ^(.)	-0.134
	EXPO	-0.034	-0.026	-0.022	0.010
	FDI	4.81e - 12	3.71e - 12	3.16e - 12 ^(*)	1.1e - 12
	INTEREST	-0.018	-0.02	-0.015	-0.005
	IMPO	0.046	0.052 ^(*)	0.046	0.014
	INF	-0.008 ^(**)	-0.004 ^(**)	-0.003 ^(*)	-0.002 ^(**)
	LABOR	-0.138 ^(.)	-0.089 ^(.)	-0.042	-0.030
	PINV	-0.015 ^(**)	-0.01	-0.010	-0.001
	POP	0.677 ^(.)	0.248	0.072	0.056
	RPOP	-0.001	-0.003	-0.009	-8.53e - 05
	PPP	0.001	0.001	0.000	0.000
	REVENU	0.027 ^(.)	0.022 ^(.)	-0.001	-0.03
	LAND	1.48e - 07 ^(.)	1.2e - 07	3.47e - 08	-2.28e - 08
AGR	-0.008	-0.016	-0.019	-0.032	
Time Invariant Variables	ASI	0.353	0.272	-0.311	-0.46
	DEBT90	-0.002	-0.006	0.002	0.011
	GDP90	-1.94e - 16	-1.93e - 15	-5.10e - 15	-3.6e - 15
	POP90	1.19e - 09	2.25e - 09	2.83e - 09	1.64e - 09

Significance code: ./*/**/***, at 1%/5%/10%/0.1%, respectively.

Correlated random effects coefficients available in appendix (D), are not significant.

Table 3.7: Quantile regression with correlated random effects, Model C

schooling years average and high growth rates each supplement year of schooling is associated with an increase of (0.555) in growth rates percentage per year against only (0.120) for countries having the lowest average of annual growth rates.

Once we control for endogeneity using the quantile regression model with correlated random effects, the positive impact of schooling years is no longer different from zero, whereas debt and education expenditure ratios to GDP maintain a negative and significant impact through all quantiles except for the top quantile $\tau = 0.95$ also showed in PFE-QR estimates. Since, the correlated random effects, reported in table (3.7) in the appendix, are not statistically different from zero, as a consequence there is no endogeneity generated by the dependence between time invariant effects and the regressors, thereby endogeneity issue at this stage can be neglected. Thus, we hold the estimates of PFE-QR.

3.6 Conclusion

Given the continuous increase in public debt levels all over the world, particularly after the last financial crisis. And the important attention devoted to study human capital role in growth literature, mainly since the seminal work of Lucas Jr (1988) and Change (1990). It became essentially to evaluate the interaction and dual effect of both (debt, and human capital) on subsequent growth rates. Developing a theoretical model, Greiner (2008b) has alluded to the adverse impact of rising debt level followed by strict adjusting policies on education expenditure then on human capital formation.

In this chapter we propose an empirical contribution to the debate about growth-debt and human capital formation, by presenting our empirical evidences based on a large data set. Data quality has been improved thanks to the most appropriate instrument *multiple imputation* that has been used to complete data missingness. We provide empirical evidences carried out on a panel of (76) developing and advanced countries all over the world for a period that covers 1990 to 2011. The sample was suffering from data lack mainly for debt and education expenditure variables, it was then balanced after the implementation of a customized multiple imputation code that fits observed as well as missing values. Methodologically this work is built on a wide strand of economic growth determinants, where a range of econometric techniques have been run to analyze data. Mainly the approach of *quantile regression for longitudinal data* that helps in exploring the existence of nonlinear or heterogeneous relationships between interest variables and the response one, together with the standard mean estimation methods such as the typical panel data models with time and individual fixed effects, or even the

stochastic frontier analysis approach. In this work, we do not treat the issue of reverse causality between debt-growth. Moreover, we pay a specific attention for another recognizable issues, particularly endogeneity, that could influence results. To this respect, we deploy the new and superior method of *Quantile Regression with Correlated Random Effects* developed by [Abrevaya and Dahl \(2008\)](#) for addressing endogeneity problem.

Thereby, controlling for a battery of growth determinants, main empirical estimates suggest an adverse and heterogeneous cross-country impact of debt on GDP growth. On average an increase of (10) percentage of debt to GDP ratio, is followed by a slow-down in annual GDP growth rates between (0.1 – 0.24) per year for sample entities. Countries with the lowest annual growth rates and highest debt levels show evidence of a non-linear effect on growth rates with a higher negative and significant impact of further debt use, against a non significant impact in countries at the upper conditional growth distribution. This result holds in all estimation method, even after controlling for endogeneity. At the same time, although we do not control for debt-growth reverse causality, estimates show that debt is a significant and robust growth determinant.

Education expenditure as a measure of human capital has a negative influence on growth rates, which supposes a need to optimize their amounts especially in the light of debt levels during and before financial crisis. Ignoring debt ratios from estimation model leads to a negative association between growth rates and the stock of schooling years, which does not hold once we add debt to model, thereafter a positive and significant impact of any additional schooling year on growth rates mainly for countries at the upper conditional growth distribution. The control for endogeneity problem using the quantile regression approach motivated by the correlated random effects and developed by [Abrevaya and Dahl \(2008\)](#), has been rejected since the correlated random effects were not significant.

These results imply that countries in the bottom quantiles should establish a public debt consolidation policies in order to attenuate the negative debt contribution in economic process, as well as improve the contribution of education expenditure in economy functioning. Or even, more in-depth studies have to be made on the utility and efficiency of these expenditures in order to rationalize policy implication in this respect. Moreover, since any additional schooling years increases growth rate, education expenditure should be assigned in a way to increase schooling years.

Chapter 4

Growth and Education Expenditure Efficiency: A Two Stage Longitudinal Quantile Regression Approach

Contents

4.1 Introduction	126
4.2 Methodological strategy	129
4.3 Data	134
4.4 Efficiency estimation	135
4.5 Conclusion and policy implications	146

In this chapter we attempt to assess the effect of public education expenditure on the economy efficiency level of (76) countries. Efficiency estimation is carried out in two stages. In the first stage, using a flexible and robust semi-parametric method namely the quantile regression, we estimate a linear growth model. The use of this method in the field of efficiency estimation has recently been revealed robust and advantageous (Behr, 2010). Estimates suggest that public education expenditure has a linear direct adverse effect on economy efficiency in all countries. Then the second stage is accomplished using a bayesian averaging model with fixed effects estimates. This model suggests that facilitating rural population school attending and rising teachers ratio per student would improve economy efficiency level. Surprisingly, literacy rate does not have a potential role in improving efficiency.

4.1 Introduction

A knowledge-based economy, technological changes and other challenges, have been the reason behind countries race to have a competitive advantage. Especially, after the recognition of human capital central role in economic growth theories, endogenous growth theory, made by (Lucas Jr, 1988), and (Change, 1990). Henceforth, human capital is considered as a crucial growth determinant to boost economic development. And any country success depends, mainly on individuals with high competence, skills and knowledge level.

The investment in human capital, *human capital formation*, starts early at the school, and other education system institutions. This formation represents the cornerstone in preparing human capital to continue the long and rich formation and skill acquisition path. Sleezer et al. (2004) state that learning is an important component to obtain much knowledge and skills through lots of acquisition ways including relationship between the individual and the others. However, investing in human capital is not only an individual need, hence it should be considered as a collective investment in the society's future. To this respect, such investment should be supported by offering the necessary inputs such as: schools, teachers, education endowments and, more importantly the *elementary function funds*, because as stressed by Ramani (2014), if states funds are invested into science and technology parks-innovation, necessary for sustain economic growth, will come to life.

Education systems all around the globe may receive finance from three sources: either from public sector, private sector or international community. Many incentives motivate the investment in the educational system at the public as well as individual level. Education does not pay off only at the individual level by substantial returns on investment, but also there are public benefits of an educated population represented by great tax revenue and social contributions that reduce the amount of public transfers to social targets. Further investing in education can foster economic growth, enhance productivity, contributing to personal and social development, beside reducing social inequality through public subsidies to household. Other benefits include health improvement, individuals well-being, facilitating social cohesion, and increasing the social consciousness of constituents within community (Beach, 2009).

Public funds devoted to education can come from various levels of government. In some cases the central government assumes the main responsibility for education related costs. For example in 2008 OECD countries had spent (6.1%) of their GDP on educational institutions. More than (3/4) of these amounts come from the public sector

(OECD, 2011b). Education expenditure as a percentage of GDP in advanced countries is more important than those of emerging and developing countries which accounts for (4.5%) of GDP, table (D.2) in appendix (D) provides the average of both education and health expenditure as a % of GDP between 1990 – 2011. Since the governments' size in emerging economies are lower compared to those in more advanced, the fraction of education expenditure as a percentage of the whole public expenditure is about (16.5%), larger by (3.5%) of those in advanced countries (Grigoli, 2014). Thus, public sector is the main creditor of education system. This sector's expenditures may not be limited to public schools since some are forwarded to the private sector either under direct support to private education institutions, or through support for educational institutions that are channeled through individuals, households and other private entities, or finally as grants and loans targeted by merit or equipment.

Expenditure per student in all education system levels has risen, at a foster rate rather than GDP growth, by (17) points across OECD members between 2005 – 2011, (OECD, 2011b). The huge amounts of public and private funds allocated to education sector give rise to the important question of the efficient use of these resources in light of the financial crisis that many countries are going through. This crucial question gains more attention when considering countries circumstances and constraints, like for example in the case of European Union where countries are bound to fiscal discipline through the *Stability and Growth Pact (SGP)*. This also is the case in countries where public funds are not amply available like in the case of developing economies. Moreover, nowadays capital and taxpayers mobility that has been fostered due to globalization phenomena, exert more pressure on governments to manage and use funds efficiently (Afonso et al., 2010). Thereby, education expenditure is a financial engagement that the whole society is responsible for, especially after the last worldwide financial crisis and the need to consolidate public budget across countries.

Measuring efficiency in health and education economics has gained much interest in recent years. The core reason of this raising interest comes not only from the fact that they represent a large percentage of national income, but also from the fact that the local government is the dominant provider of financial resources (Afonso and St Aubyn, 2004). In this area, authors in Gupta and Verhoeven (2001) have assessed the efficiency of government spending on education and health in (37) countries in Africa from 1984 to 1995 using *Free Disposal Hull (FDH)*, and *Data Envelopment Analysis (DEA)* techniques. Authors had considered expenditure in *purchasing power parity (PPP)* as an input and gross enrollment rates in primary and secondary schools as outputs. They have found that rich countries tend to be in the less efficient group. In this study, different countries know different efficiency levels depending on regional factors. In

another study (Gupta et al., 2002), Gupta uses cross-sectional data for (50) countries to reassess whether increased public spending on education and health care system matters. Gupta found that the overall level of public spending and intra-sectoral allocation matters, especially shifted spending towards primary and secondary education. Authors in Afonso and St Aubyn (2004) measured the efficiency in education and health sectors in OECD countries using the non-parametric methods of DEA and FDH. Authors main findings concerning education sector was that, the average input efficiency varies between (0.859) and (0.866) depending on the used method. In a sample of (140) developing economies over 1996 – 2002, authors in Pang and Herrera (2005) have examined the efficiency of government spending in health and education using FDH and DEA, they consider education expenditure as the input and gross primary enrollment rate and completion as outputs. Their main outcome was that countries with higher expenditure levels register lower efficiency score. Since measuring education expenditure efficiency using the parametric and non-parametric methods is a challenging subject, Grigoli (2014) has offered a hybrid approach that accounts for the impact of the level of development on education outcomes, to estimate secondary education expenditure efficiency in emerging and developing countries. Grigoli hybrid approach is composed of three stages where in the first he uses the DEA, then in second a *locally weighted scatter plot smoothing (LOWESS)* through the efficient points, in the last one in order to match the point off the frontier to the closet data point he uses the Mahalanobis¹ match. His main finding was that secondary education is inefficient in many developing countries, especially in *Africa*.

Including education and health expenditure the topic of estimating public spending efficiency has enjoyed a remarkable attention of economists over a long time, where estimation approaches is generally made through comparing allocated spent amounts with the education system outputs such as the number of student completing a grade or school enrollment rate. Contrary to the last short background on education expenditure efficiency studies, here *our goal is to assess the contribution of education expenditure as one of human capital proxies in the whole economy efficiency² where estimation model accounts for units fiscal stance, by inducing debt and fiscal deficit to model*. This study is different

1. Mahalanobis distance is a measure of the distance between a point P and a distribution D . It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D . This distance is zero if P is at the mean of D , and grows as P moves away from the mean: along each principal component axis, it measures the number of standard deviations from P to the mean of D .

2. Here by computing economy efficiency, we mean technical efficiency: which represents the pure link between inputs, outputs and the outcomes, in considering at same time the production possibility frontier, thereby the gain in technical efficiency can be showed as a movement toward production frontier. Thereby, allocative efficiency represents the link between the optimal of inputs after taking costs and benefits in consideration (Mandl et al., 2008).

of existing literature in some points: in technical side, inspired from Behr (2010), we use a semi-parametric efficiency estimation method, the *Quantile Regression approach (QR)*, that avoids the aforementioned methods drawbacks, that we will discuss later in this chapter. Secondly, the impact of education expenditure is not limited to one or two outputs otherwise it has a primacy impact in many areas such in income generating, fertility, health status and further social indicators where typically, parametric and non-parametric, estimation methods are not capable to take them all into consideration. Beside the fact that the final objective of public education expenditure is to improve growth rates and development level, in a two stage efficiency estimation procedure we measure the impact of education expenditure on economic growth among other regressors. This proceeding allows us to cover as wide as possible the direct and indirect impact of education expenditure on economy process and functioning . Thirdly, estimation procedure is to be carried out in two steps for policy implications, where in the second step we regress efficiency score on a new set of regressors using the *Bayesian Model Averaging (BMA)*.

The importance of this study comes firstly from the increasing attention accorded to the investment in human capital, as this last promotes economic prosperity, fuller employment and social cohesion. Also because education represents a future collective investment for nations and societies, where the volume of education expenditure in public budget continues its climb compared to other public spending, see table (D.2) in the Appendix (D). Another important motivation for this study is that governments, principally those of developing economies, have limited public sources and an increasing number of public targets to be satisfied. All these reasons make this kind of studies a real need through which policy makers may conclude results and modify policies.

The structure of this chapter is as follows: Section (4.2) presents the limits of different methods used in such studies measuring the economic performance. In Section (4.3) we present our sample of data where discussed estimation methods will be applied, results are discussed in Section (4.4), final conclusion in Section (4.5).

4.2 Methodological strategy

For a long time the issue of assessing government expenditure efficiency has occupied and is still of interests for researchers and policy makers aiming to allocate efficiently the decreasing amounts of public funds. To this respect, there are many parametric and non-parametric methods that have been used by economists. We first present an intuitive overview of the alternative efficiency estimation approaches and

their limits, then we pass to describe the estimation method we use in this chapter.

4.2.1 Limits of traditional efficiency evaluation methods

It is worth remembering that a combination of input-output (x, y) is said to be efficient in technology T if and only if this combination cannot be replaced by another one $(\hat{x}, \hat{y}) \in T$, thereafter this efficient mix TE cannot be improved, (Bogetoft and Otto, 2010).

A variety of parametric and non-parametric efficiency estimation methods are available to estimate input-output efficiency, each one of these methods has some limits and features, then researcher have to choose the appropriate method according their study circumstances.

In non-parametric methods, DEA is a widely used method to estimate efficiency like in Ventelou and Bry (2006); Hsu (2013). The first appearance of this method has been made by the seminal work of Farrell (1957), then DEA gained popularity thanks to Charnes et al. (1978) who had published an article in European Journal of Operational Research, with the purpose of evaluating the relative efficiency of several decision making units. DEA provides a mathematical programming platform to estimate optimal production frontier and evaluate the relative efficiency of different entities (Bogetoft and Otto, 2010). It is weakly restricted, except from certain technological assumptions specific to DEA, those are free disposability, and some form of convexity, where always there is some convexity assumption in economic models.

On the other hand, DEA has some drawbacks making it less attractive. Firstly, being sensitive to outliers, assuming data to be free of measurement error which is far from reality. Using DEA efficiency estimates are usually false when data set is not sufficiently big and there are many inputs or outputs³. Moreover, this method does not give any estimates of the marginal productivity of the inputs (Liu et al., 2008). Finally, DEA does not consider a general relationship but it only relies on the input-output ratio optimization, that makes the implementation of DEA not appropriate for all kind of efficiency estimations studies. For example, that is the case in estimating education expenditure efficiency, where not all outputs are quantitatively measured. Consequently, inefficiency is not associated with wasted resources (Thanassoulis, 2001).

Another non-parametric method is the Free Disposal Hull, this method has been developed to empirically estimate the relative efficiency of production units in a mar-

3. This case will result many individuals situated at the frontier which is far from being true.

ket environment. FDH constitutes an intuitive tool that can be used to identify the best practice in government spending and to assess how governments are faring in comparison with these best practices. One producer is efficient when there is no other producers that could produce the same output quantity with an equal amount of input. It is worth noting that an efficient producer in FDH can be highly inefficient in DEA, implying that DEA is stricter than FDH. Another common drawback for non-parametric methods is that omitted and irrelevant inputs or outputs will affect estimates of inefficiency, as well as does a small size sample on efficiency estimation (Sutherland et al., 2007).

Similarly to DEA, the Stochastic Frontier Analysis is the most used parametric estimation method with more economics and econometrics oriented background. Contrary to DEA, SFA is based on classical statistical principles, most notably the maximum likelihood principle. This method proposes a new formulation with a composed error term⁴.

In SFA with the inefficiency term, all estimated production levels are always less than the optimal because of inefficiency term. Whereas each data above the line, production frontier, is considered as being by pure chance and has no influence on what the technology set looks like Bogetoft and Otto (2010). The composed error term has a non-zero expectation, which means the estimated frontier turns out as a shifted conditional mean regression, therefore the average production relation is maintained even for the most efficient individuals. This contradicts the fact that production/or cost function might vary at different efficiency levels making the using of quantile regression method more interesting (Behr, 2010).

When the main purpose of a study is to evaluate system wide shift like the impact of some policy change, it will be better to apply the stochastic parametric approach rather than the non-parametric one DEA which can give results insensitive to random variation in data.

A common point between DEA and SFA is that both methods assume that production function is common to all units, while FDH does not provide any functional form. Thereby estimated efficiency level is identical all over the dependent variable distribution, moreover it cannot be negative. SFA estimates are based on the maximum likelihood estimation, which in its turn based on the conditional mean. That means

4. SFA mathematical formulation is:

$$Y_{it} = f(X_{it}; \beta) + v - u, \quad (4.1)$$

where u is the inefficiency term following half normal distribution law $u \sim N_+(0, \sigma_u^2)$ with solely positive side, and v is the term which takes care of the stochastic nature of the production process and possible measurement error in the inputs and output, follows a normal distribution law. u and v are independents.

SFA does not take into consideration the possible differences in production technologies between units at the upper tail of output conditional distribution and the units of the bottom representing the less efficient units, implying thereafter overestimation of efficiency. Moreover upper tail observations are rather considered as outliers (Kaditi and Nitsi, 2010), where both of DEA and SFA are sensitive to outliers.

After this brief comparison of some efficiency estimation methods, it is clear that more flexibility to parametric method and going deep in non-parametric methods would be desirable, that is why presenting the following method of Quantile Regression approach QR to estimate efficiency motivated by the work of Behr (2010) would be interesting. QR is an ad hoc method that provides a robust efficiency score. In the next subsection we detail this method and present its features.

4.2.2 Quantile regression approach to estimate economy efficiency

Quantile Regression is a semi-parametric method of estimation that had been used to model the link between economic growth and its factors firstly by Koenker and Bassett Jr (1978). It describes the response variable as a conditional function of a set of covariates (Bernini et al., 2004). Demonstrating how explanatory variables will change location, form and scale of the response variable (Alexander and Lamarche, 2011). Thereby, it extends the analysis from mean or median to the full range of conditional quantile functions, providing an analytical description of an ordered set of technological relationship corresponding to different levels of efficiency (Bernini et al., 2004).

Using QR approach to estimate production function efficiency is recommended particularly when the conditional distribution does not have a standard shape, like being an asymmetric or truncated one (Behr, 2010). In this case the use of QR allows to obtain a reliable inferences on efficiency that does not rely on particular distributional assumption (Kumbhakar et al., 2015). Further stress on this advantage has been made while estimating the effect of government subsidies on financial institutions' performance by Wang et al. (2008), who have proposed the approach of quantile regression as a new *distribution-free* estimation method. Moreover, QR is well suited for efficiency estimation when there is considerable heterogeneity in data, which is the case in our data base⁵. Implementing QR in efficiency studies allows the interpretation of different levels of efficiency over the spectrum of the production function (Bernini et al., 2004), which as we have seen is not available using any of the parametric and non-parametric

5. It is worth noting that when data is homogenous, estimating efficiency by QR will not give different outputs from those obtained using SFA method.

methods. Beside avoiding the criticism addressed to DEA, [Liu et al. \(2008\)](#) states that similarly to non-parametric method, and it is more robust than SFA and DEA against outliers .

In a similar way to parametric methods, QR estimation permits the inclusion of two types of variables, those related to efficiency estimation, and those necessary to explain dependent variable. Moreover, QR uses heterogeneity information stocked in the error term that often present in empirical studies, ([Behr, 2010](#)). The use of this semi-parametric approach has an important and rarely advantages compared to other approaches, that QR allows the estimation of negative efficiency score which is completely consistent with our case, and in estimating profit efficiency score, that because profit efficiency score can be negative⁶. The only drawback that QR has is that, against SFA method, QR does not allow for random error to explain efficiency differences among individuals, ([Kumbhakar et al., 2015](#)). The choice of τ , the conditional quantile of Y , depends on the size of the sample, the bigger the sample is the higher τ can be set. Estimating individual efficiency is simply made by comparing each quantile outputs with those of individual located upper tail, e.g., $\tau = 0.95$, of response variable conditional distribution.

QR have been used in many efficiency studies such as [Chidmi et al. \(2011\)](#) to estimate the sources of technical efficiency in dairy farms in Wisconsin by applying a process of two steps. In a first step, using Cobb-Douglas production function, they estimate the stochastic production frontier following the frame work proposed by [Aigner et al. \(1977\)](#). In the second step, they regress the inefficiency term $TI = 1 - TE$ on a set of variables Z using QR. Equally used by [Kaditi and Nitsi \(2010\)](#) in estimating farm efficiency, the author argues that QR yields a robust efficiency score estimation in the agricultural sector.

Along the same line with [Behr \(2010\)](#), we use the same QR procedure to estimate frontier production function and compute efficiency score, then in next stage, through a Bayesian averaging model, efficiency scores are regressed on a set of covariates in order to identify efficiency determinants.

6. Since firms can throw away more than 100% of their potential profits, then profit efficiency score can be negative, [Berger and Mester \(1997\)](#).

4.3 Data

Intuitively, since governments are the principal responsible of offering education system inputs including education expenditure, that is engaged to form human capital, then there is a potential link between public education expenditure and economic growth. Until now, there is no definitive empirical validation to this relation. In [Blankenau and Simpson \(2004\)](#), authors found that the direct effect of increasing the fraction of outputs devoted to education public expenditure is an increment in the steady state growth rate, whereas general equilibrium adjustment could have contrariwise effects i.e., when labor is taxed at a higher rate than capital then any public education expenditure lowers growth. Thereby, this relationship is non monotonic.

In the present chapter we will carry an empirical application on a large international longitudinal data set of (76) countries during the period from 1990 to 2011, the same data base in Chapter (3), to estimate public education expenditure participation in economy efficiency.

The large number of countries reflects heterogeneity aspects especially when one knows that the development levels of these countries are divergent, where the mixture of developed and developing economies increases the power of statistical estimates because of the high variation they induce to estimation model. That is why the use of quantile regression approach with such data sounds interesting since this method accounts for data heterogeneity.

Data set has already been imputed (100) and experimented in Chapter (3), in Chapter (1) more details on imputation concept and procedure, data before and after imputation and imputation diagnostics tests are available. Model component selection is based on the wide range of empirical papers on economic growth determinants, where the result set is a battery of (24) explanatory variables reflecting countries economic environment such as: public education expenditure, public health expenditure, labor force, gross fixed capital formation, to name a few. The collecting of available data on model predictors has been made mainly on the site of the World Bank Data.

Only the first (5) after-imputation complete data sets have been used to estimate model and then calculate efficiency score. In Section (4.4) results and discussion are shown.

4.4 Efficiency estimation

A first estimation of economy efficiency had been made using the parametric method SFA, where mean efficiency for the whole is (0.164). This score does not reflect individual heterogeneity, that is why we are interested in applying the quantile regression estimation as it gives variant coefficients of all regressors according to the conditional quantile of dependent variable. We estimate the following model:

$$Y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + \lambda S_{it} + \epsilon_{it}, \quad (4.2)$$

where X is a vector of interest variables, that includes: public education expenditure as a % of GDP, and average years of schooling for person above (25) year. Vector of control variables Z contains: public debt as a % of GDP, gross fixed capital formation as a % of GDP, current account, unemployment rate, primary sold, military expenditure as a % of GDP, health expenditure as a % of GDP, saving rate, life expectancy, exportation as a % of GDP, importation as a % of GDP, foreign direct investment at current dollar, interest rate, inflation rate, labor as a % of total population aged more than (15) year, price investment, population growth rate, purchasing power parity, general government revenue as a % of GDP and agricultural land. S vector of time invariant variables contains population and GDP at 1990 year, finally ϵ representing the error term.

4.4.1 First stage: QR estimated efficiency score

In quantile regression analysis of GDP annual growth rates, estimates are made on quantiles members separately, table (D.1) in the Appendix D shows countries classification in quantiles according to their GDP growth rates. Here, we select four quantiles of dependent variable $\tau = 0.25, 0.5, 0.75, 0.95$, each of them has (19) country. Since the data sample is big enough, we set $\tau = 0.95$ as the *benchmark level*. First quantile members have in majority advanced economies and issued from the OECD like: Germany, France, Belgium, at the same time there are other members with less development levels such as: Ivory Coast and Latvia.

Estimating the variant regressors parameters has been done thanks to *rqpd-R* package (Koenker, 2004), results are presented in table (D.3) in the Appendix D. Using these estimates we will calculate the estimated dependent variable of each quantile. Then, estimating efficiency score can simply be achieved by comparing the estimated output of each quantile with that of the benchmark quantile $\tau = 0.95$ as being the most efficient, where countries are located at the highest level of the conditional distribution of

growth rates. As we have mentioned earlier that only (5) complete data sets will be estimated, efficiency scores in the (5) data sets are equals and no significant differences are present as confirms Anova-one factor test ⁷.

Table (4.1) shows efficiency score of quantiles' members:

$\tau = 0,25$		0,5		0,75	
Country	Mean Eff	Country	Mean Eff	Country	Mean Eff
Moldova	-0.735	Sweden	0.21	Zambia	0.812
Ukraine	-0.405	UK	0.262	Australia	0.5
Russia	-0.199	Austria	0.244	El Salvador	0.473
Latvia	-0.319	Canada	0.286	Uruguay	0.355
Kyrgyzstan	-0.41	Netherlands	0.234	Ecuador	0.644
Hungary	-0.669	Iceland	0.11	Malta	0.312
Italy	-0.748	Slovak Rep	0.419	Honduras	0.622
Romania	-0.267	Spain	0.225	Cyprus	0.379
Bulgaria	-0.361	USA	0.139	Colombia	0.562
Japan	-0.895	Norway	0.401	Guatemala	0.54
Ivory Coast	-0.473	South Africa	0.281	Philippines	0.664
Jamaica	-0.369	Algeria	0.586	Morocco	0.588
Denmark	-0.982	Brazil	0.163	Luxembourg	0.558
France	-0.903	Mexico	0.485	Bolivia	0.638
Germany	-0.891	Paraguay	0.424	Trinidad and Tobago	0.410
Greece	-1.208	Venezuela	0.336	Lesotho	0.642
Portugal	-0.995	Nicaragua	0.362	Pakistan	0.527
Belgium	-0.719	Kenya	0.483	Benin	0.571
Finland	-0.677	Albania	0.367	Nepal	0.772
Mean Efficiency	-0.643		0.317		0.556

Table 4.1: Efficiency Estimation

Average efficiency score for the whole sample ranges from (-0.643) for ($Q = 0.25$), to (0.556) for ($Q = 0.75$), while members of the upper quantile are supposed to have a unit efficiency term. First quantile average efficiency is negative, this is completely natural as we estimate efficiency in term of annual economy growth rate. Note that Such negative efficiency score is not possible using SFA.

The immediate and direct contribution of education expenditure on growth rates and economy performance is negative in all quantiles, it goes between (-0.259) and (-0.383) implying that any supplement increase of education expenditure tends to decrease the annual level of economic growth effective. Table (4.2) shows, through average schooling years, the immediate and lagged impact of education expenditure in each quantile. Since standardized coefficients ⁸ allow the comparison of the relative effects

7. Anova One Factor Analysis,calculated Fisher test is (0.009) where critical value is (2.404) with probability of (0.999). Alternative hypothesis of this test supposes that the different estimated values are not equals, therefore we have to analysis separately the 5 sets of data. According to Fisher test results we retain null hypothesis where all obtained results are equals.

8. X Standardized coefficient= $\beta_X * (SD_X / SD_Y)$.

4.4. Efficiency estimation

of predictors measured in different scales also to determine precisely the standardized effect size statistics, we also calculate those of education expenditure and schooling stock.

Quantiles	Education Expenditure % of GDP	Education Coef	Education Stand Coef	Schooling Average Years	Schooling Coef	Schooling Stand Coef
$\tau = 0.25$	5.04	-0.377	-0.099	8.89	0.139	0.047
$\tau = 0.50$	5.2	-0.247	-0.108	8.4	0.165	0.116
$\tau = 0.75$	4.24	-0.215	-0.223	6.4	0.184	0.174
$\tau = 0.95$	4.34	-0.271	-0.127	6.68	0.518	0.251

Table 4.2: Public Education Expenditure fraction and coefficients

Table (4.2) notes that quantiles at the upper tail of growth rates conditional distribution equally the most efficient have the lowest share of education expenditure in GDP and the lowest stock of schooling years per human capital. But, any supplement year of schooling in the tail conditional growth rate members will increase growth rate by (0.518) percent, thus education expenditure can act positively if any spending rising will lead to increase schooling years.

Elsewhere, education expenditure and schooling years are not the main source behind the high growth rates in quantiles. This constatation has also been confirmed by the negative education expenditure coefficient, the adverse impact of education expenditure on the whole economy gets bigger in the highest quantiles $\tau = 0.75, 0.95$. Moreover, the negative impact of education expenditure is not associated with the amount of these expenditure, countries having the bigger negative impact of education expenditure own lower level of education expenditure in GDP compared to the rest.

At this stage, holding constant the impact of any growth factor, results suggest a negative direct contribution of education expenditure in efficiency estimation process, against a positive role played by the stock of human capital. Unsurprisingly, higher level of education expenditure cannot insure a better performance by education system, because countries spending an equate expenditure amounts do not necessarily have the same economic development level, nor the same education policies. That can result by different levels of outputs implying different effects on the whole economy thereby on these expenditure efficiency level.

To further policy implication, in the next subsection, we pursuit the process of efficiency estimation in a second stage where further variables have been added to the model.

4.4.2 Second stage analysis: Bayesian model averaging

Education expenditure represents a policy tool through which government intervenes in economic environment to achieve some economic and social objectives. Hence, it would be logical and fruitful to identify the variables which contributes either positively or negatively to the performance of these expenditure, then governments can devote more attention to guarantee the best results. In the first precedent stage we were cautious in interpreting outputs since results are influenced by factors other than education spending and human capital stock. That is why in this second stage we aim to analyze how education system factors contribute in improving efficiency score that is why we run a BMA regression, to identify in light of results the policies to be established by authorities.

In this stage, economy efficiency scores derived at different quantiles, $\tau = 0.25, 0.50, 0.75$, will be regressed on a new set of covariates, beside the first stage variables, that are supposed also to have an impact on efficiency score. Those factors have been already used in second stage analysis in [Grigoli \(2014\)](#) study while estimating the efficiency of public spending on education.

Then, the variables we test here are the following:

1. **Student-to-teacher ratio** at primary level, this indicator as recognized by the OECD is an important indicator of the level of resources devoted to education. It expresses the relationship between the number of full-time equivalent students enrolled at a given education level and the full-time equivalent teachers at that level and in the similar institutions. The ratio of student per teachers is importantly related to class size, that for long time it has been considered as its proxy. Class size and student per teacher ratio are the most discussed aspects of education system ([OECD, 2011a](#)). When the number of students per teacher is high, the government is considered to have a more efficient education expenditure in advanced countries, while it is the inverse in less developed and emerging economies. The World Bank 2005 indicates that a number of empirical studies identify an optimal threshold of 25 students per teacher, ([Grigoli, 2014](#)). Moreover to capture any possible nonlinear relationship between this ratio and education expenditure efficiency, we include its quadratic term. The term of student per teacher as education expenditure efficiency factor has received long analysis. Early research found no association between class size and student achievement as a measure of efficiency. Rather, [Gilbert \(1995\)](#) state that class size is neither necessary nor sufficient to ensure students learning quality, growth and development, instead they found that student, instructor, course organization, and man-

agement are more important than class size.

2. **Level of adults literacy**, the more parents are educated, the more they intend to educate their children, resulting a higher enrollment rate. Many studies have controlled for this measure like in [Pang and Herrera \(2005\)](#). The two previous factors are directly related to the performance of education expenditure.
3. **Rural population** as a share of total population. We include this variable for two reasons. Firstly, more this fraction increases, less efficient education expenditure becomes and their returns lower. That is because in rural regions transportation system is not sufficiently developed as in urban ones. Thus, as a consequence of difficulty attaining school and a decreasing enrollment rate ([Grigoli, 2014](#)). [Jayasuriya and Wodon \(2003\)](#) had argued that urbanization and bureaucracy are strongly and significantly correlated to education and health efficiency. Secondly, rural population indicator reflects economy dispersion that may influence any sort of economic activities. Rural zones, contrary to urban areas, do not highly attract population as no wider variety of cultural institutions are disposable. Moreover, the improved mobility in urban areas attract human capital and stimulate research and development.
4. **Government effectiveness**, this measure as defined by the World Bank, captures the perception of the quality of public service, the quality of civil service, and the degree of its independence from political pressure. Therefore, it implicitly measures the perception of public education services. At the same time, improving the quality of government institutions will imply an improving of its services efficiency, e.g., education expenditure. Moreover empirical results in [Mandl et al. \(2008\)](#) state that modern and efficient public administration have a positive impact on productivity and growth, that is why EU members have already reformed public administration to achieve efficiency gains. The inclusion of such factor has been motivated by the work of [Grigoli \(2014\)](#), since he argues that better institutions are associated with higher efficiency of spending.
5. **Political stability**, the uncertainty associated with unstable political environment may reduce economic development and investment. The author in [Barro \(1991\)](#), has already proven that policy instability reduces per capita growth rates. Moreover, it reduces productivity growth, and the accumulation of human and physical capital as suggested by [Aisen and Veiga \(2013\)](#)⁹.
6. **Income inequality indicator, GINI**. Equally distributed incomes imply that a large share of the population has a sufficient income to attend school rising by that the enrollment rate, and later the participation in market labor ([Pang and](#)

9. The source of government effectiveness and political stability is: www.govindicators.org

Herrera, 2005).

7. Being an OECD member implies specific characteristics that affect economic performance, to capture any potential effect we include an OECD dummy variable.
8. A time dummy variable to control shocks.
9. We also control for a non-linear effect by adding a quadratic term, and interaction term with the OECD dummy.

One would say that regarding the subject of education expenditure contribution in efficiency score, there are many more indicators than those cited in the above list like: enrollment rate and education expenditure by level. Unfortunately, we have to ignore them as they have an important missing data fraction. To the same missingness issue, and due to the important fraction of missing values in second stage variables, we impute them using the procedure of *multiple imputation* by setting $m = 5$. Those variables statistics, including the amount of missing data, before and after imputation procedure are available in table (2.3) in Chapter 2.

In this stage to estimate variables linkage to efficiency score we use the *Bayesian Model Averaging* approach. BMA has been deployed in various branches of applied sciences that suffer from model uncertainty, where there exist a variability of models which may all be statistically reasonable but most likely result in different conclusions about the question of interest to the researcher. To this aim BMA calculates variables posterior probability to be in the final model, (Amini and Parmeter, 2011). This technique provides a posterior, a likelihood, and a prior distribution for each model. Using bayes' rule, BMA enables researchers to derive a probability statement about what we do not know depending on what we already know: which is the collected data. In what concerns the prior, many studies use the diffuse prior setting which gives the same probability to all possible models thereby imposing a mean prior model size of $K/2$, (Sala-I-Martin et al., 2004).

R program offers the possibility to implement the Bayesian model averaging through the BMS¹⁰ package developed by Zeugner (2011), that we use here.

The implementation of Bayesian approach of averaging across models, so called after Sala-I-Martin et al. (2004) Bayesian Averaging of Classical Estimate (BACE)¹¹ to select growth determinants has been made in Raftery (1995). Then in a pure Bayesian spirit, Fernandez et al. (2001) have applied the BMA approach for the same objective.

10. BMS package offers the Bayesian model averaging for linear models with wide choice of customizable priors.

11. Sala and others in Sala-I-Martin et al. (2004) have proposed a model averaging technique to determine the importance of variables in cross country growth regression. They call it BACE, which combines the averaging of estimates across models which is a Bayesian concept, with classical OLS estimation which comes from the assumption of diffuse prior.

In the same line with previous studies, we implement BMA on a panel data with countries specific fixed effects in order to simultaneously address model uncertainty and endogeneity issues.

In table (4.3), we present BMA with individual fixed effects estimates:

	PIP	Post Mean	Post SD	Cond.Pos. Sign
RPOP	0.941029	$-9.833798e - 02$	0.0623411653	0.00000000
GOVEFF ²	0.818542	$-8.690845e - 02$	0.0620203807	0.00000000
RPOP.OECD	0.713486	$6.159060e - 02$	0.0604026485	0.99923054
RPOP ²	0.673559	$5.070110e - 04$	0.0006001466	0.93045895
STU-TEACH	0.605039	$-2.168370e - 02$	0.0325784604	0.00000000
OECD	0.559829	$-2.454113e + 00$	3.5044974238	0.03216875
STU – TEACH ²	0.485167	$1.745207e - 04$	0.0004426250	0.64043309
STU-TEACH.OECD	0.420686	$8.208835e - 03$	0.0208129453	0.87612376
GOVEFF	0.417065	$4.571938e - 02$	0.1202729608	0.98771654
GOVEFF.OECD	0.410460	$8.827519e - 02$	0.2693907244	0.86302685
LITER	0.372293	$-4.142606e - 03$	0.0167931773	0.00749410
POLIST ²	0.367737	$1.250741e - 02$	0.0429785710	0.93395008
LITER ²	0.361124	$9.224299e - 06$	0.0001062513	0.38019905
LITER.OECD	0.360845	$-1.481593e - 03$	0.0123201019	0.33829761
GINI ²	0.352391	$2.591195e - 05$	0.0001370821	0.92997267
POLIST	0.340199	$8.006668e - 03$	0.0743588168	0.76534910
GINI	0.340091	$-9.455207e - 04$	0.0128470620	0.58070046
GINI.OECD	0.335373	$-1.512141e - 04$	0.0099255713	0.49397238
POLIST.OECD	0.335335	$3.975820e - 03$	0.1543737792	0.53951720

Table 4.3: BMA estimates

Results of first column in table (4.3)¹², represent the *posterior inclusion probability* (*PIP*) of each variable in the candidate model reflecting its importance in explaining dependent variable. The value of (*PIP*) goes between (0) and (1), estimates show that final model contains all individual fixed effects whose (*PIP*) is completely (100%) , see in the Appendix D table (D.5), plus variables whose (*PIP*) is higher or equals 0.50%, those variables are: *RPOP*, *GOVEFF²*, *RPOP.OECD*, *RPOP²*, *STU – TEACH*, *OECD*, *STU – TEACH²*¹³.

The second column labeled (*PostMean*) displays the coefficients averaged over all models including models wherein these variable were not included. According to this column's values (*RPOP*, *RPOP.OECD*, *STU – TEACH²*, and *POLIST*) are supposed

12. Results in table (4.3), are carried out the first imputed data set. BMA estimates on the rest (4) imputed data sets are exposed in the appendix. Observing BMA results on the five complete data bases, there is no major differences between results. Consequently, we retain the estimates of the first after-imputation complete data base.

13. BMA estimates on the all (5) after-imputation data sets are reported in the appendix D, see tables (D.6, D.7, D.8, and D.9). Where is these tables, we expose only variables with *PIP* equals or close to 50%.

to have a positive contribution in the candidate model. This has been emphasized by last column values (*Cond.Pos.Sign*) that represent the expected sign of each variable. When (*Cond.Pos.Sign*) is close to the unity variable's sign, it is positive otherwise it will be negative. The correlation between iteration counts and analytical posterior model probability (*PMPs*) for the best 1000 model tested is far from perfect but in the same time reflects a good degree of convergence ($CorPMP = 0.9019$).

With (94.1%,0.61%) respectively, virtually all of the posterior models mass rests on model that includes (*RPOP*) and (*STU – TEACH*). Consistent to [Grigoli \(2014\)](#) results each of rural population part in total population and student per teacher ratio has a negative association with efficiency score of (-0.098) and (-0.022), thereby on the contribution of education expenditure in this efficiency. [Gupta and Verhoeven \(2001\)](#) also has identified a positive association between urbanization and higher education expenditure efficiency level.

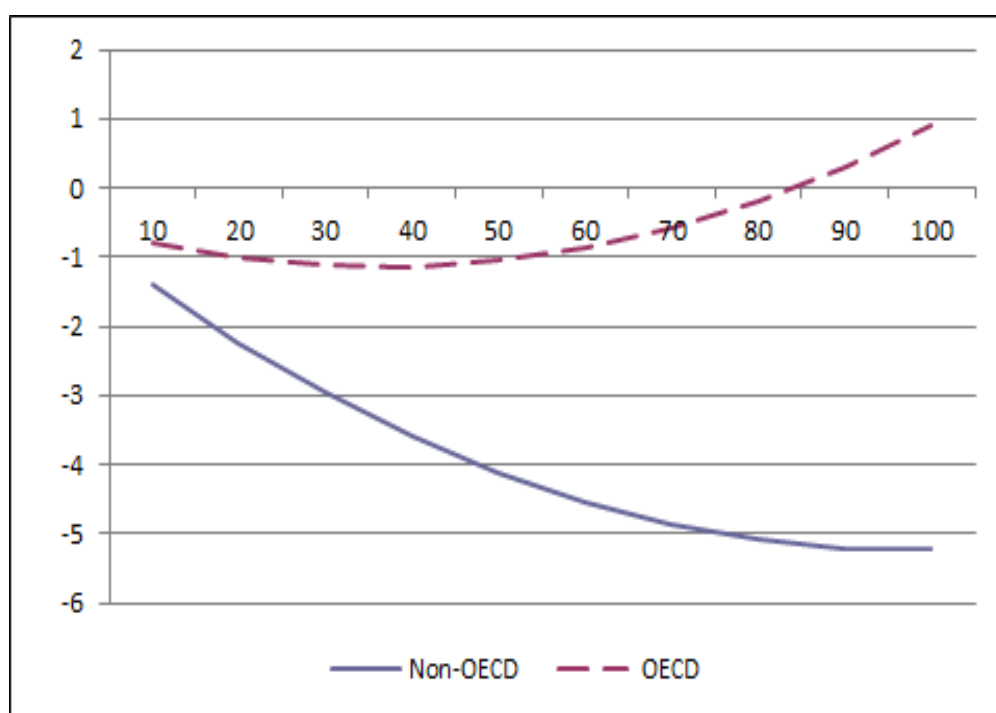


Figure 4.1: Rural population in % of total population effect on efficiency score, in OECD and Non-OECD countries

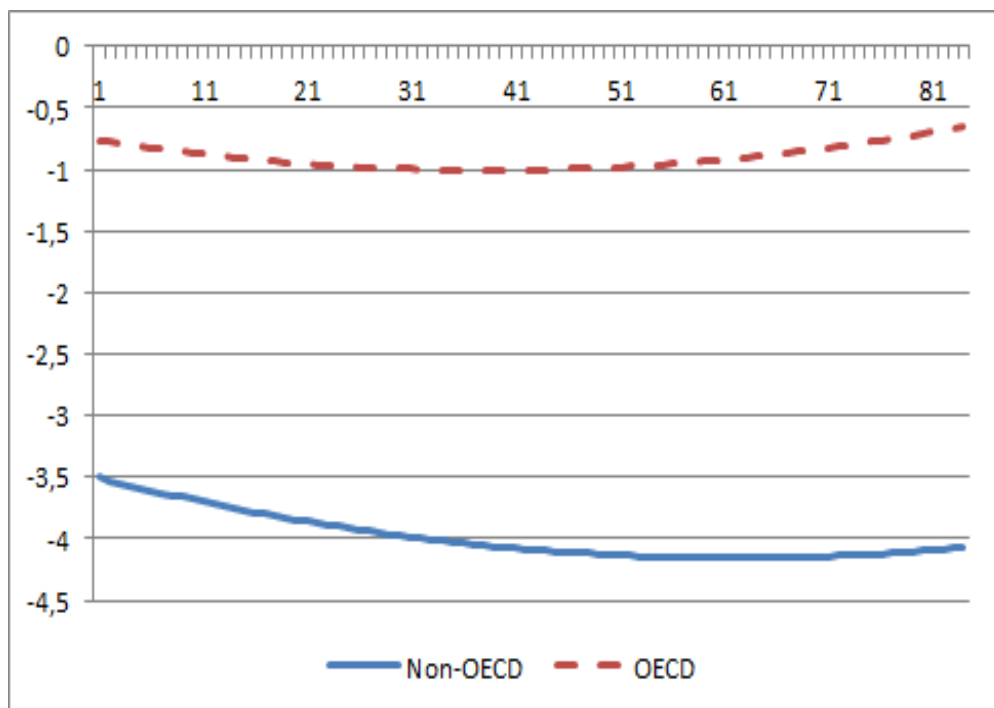


Figure 4.2: Student per teacher effect on efficiency score, in OECD and Non-OECD countries

Figure (4.2) shows a U-inverted relationship, indicating that the raise of students number per teacher will lead to a reduction in economy efficiency. Thereby education expenditure are less efficient. More students per teacher has a higher negative impact in non-OECD countries where the mean of student per teacher is of (27.839), against (16.426) in OECD members. Once students number attains the level of (39) in OECD members and (69) in non-OECD countries, any further expansion in class size will contribute positively in improving economy efficiency.

The same constatation in figure (4.1), that rural population impact shows a negative association between efficiency scores and the fraction of non urbanized population, notably for non-OECD countries. Figure (4.1) shows that, in OECD members, after the level of (50%), the fraction of rural population has a positive contribution in economy efficiency. With a probability of (71.4%) the interaction term rural population and OECD dummy variable is positively associated with efficiency score, where the rural population net impact becomes $(-0.098 + 0.062 = -0.036)$ in OECD members. This attenuation of the negative linkage between rural population and efficiency score in OECD countries, could come from the fact that the quality and quantity of public services they provide to rural population, implying transportation system and telecommunication, are higher than in other countries. OECD dummy variable is supposed to be

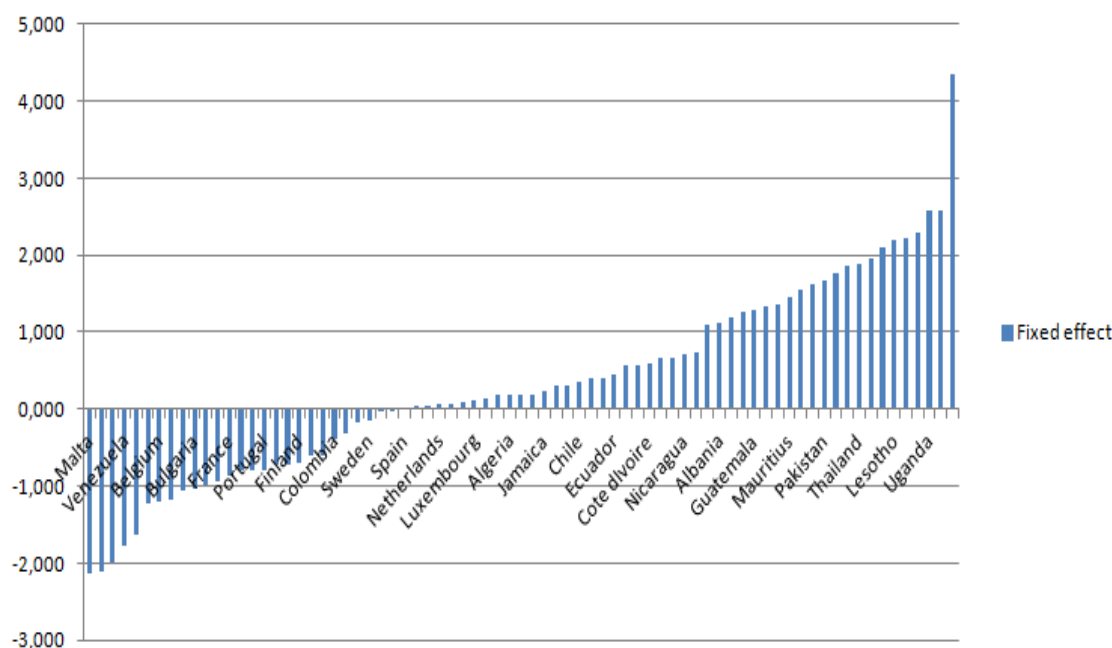


Figure 4.3: Countries fixed effects

a potential regressor in candidate model that has an adverse link to economy efficiency score with a probability of (55.98%). Non-linear impact of public services quality perception is negative with inclusion posterior probability of (81.85%). Surprisingly, each of policy stability and literacy rate does not seem to be potential determinants to economy efficiency. Policy stability and government effectiveness are positively correlated to efficiency score, while income inequality is negatively associated with efficiency as it is supposed to be, but all those factors with (*PIP*) lower than(50%) do not belong to final model equally for the rest of regressors and their interactions. BMA estimates of individual fixed effects are shown in figure (4.3):

Countries under the horizontal line in figure (4.3), that count for (27) countries¹⁴, are those with negative fixed effects regarding efficiency score, whereas (48) residual countries have a positive individual fixed effects. Countries with negative fixed effects include 14 OECD members, against (10) in the group of positive individual fixed effects, see table (D.5) in the Appendix D.

BMA estimated fixed effects in general account for adverse relationship between efficiency score and countries economic development level which should not be the case, as shown in the histogram (4.3). As we tried to explain this inconsistency in the

14. Reference country in the sample is the United Kingdom, the choice of this country comes from its nature of advanced country and as a member of the European Union and OECD organization.

two precedent steps, we see in the convergence hypothesis the most plausible answer.

Convergence hypothesis

Regarding countries fixed effects figure (4.3), development level seems to have an impact on economy efficiency thereby on the function of education expenditure. The authors in [Pang and Herrera \(2005\)](#) note that there is a positive relationship between expenditure and country development level, that implies price levels in wealthier countries tend to be higher than those in poorer ones¹⁵. In another study that addresses the impact of development level on education expenditure efficiency, ([Gupta and Verhoeven, 2001](#)), results suggest that rich countries tend to be among the least efficient ones.

In the attempt to understand the negative link between development level and efficiency score, we find in the convergence theorem a possible interpretation. This theory was the fruit of two principal works in 1952 and 1956, the first appearance was in the essay of Alexander Gerschenkron titled *Economic Backwardness In Historical Perspective*, ([Gerschenkron et al., 1962](#)). This essay showed that the relative backwardness of a country and the absence of great blocks to industrialization can contribute to its future development if supported by adequate endowments of usable resources. The second was produced in the work of ([Solow, 1956](#)) *A Contribution To The Theory Of Economic Growth*. His work pioneering the idea of the generality of unconditional convergence worldwide, ([Korotayev and Zinkina, 2014](#)). Several paths may help the convergence phenomena to take place such as: research and development activities, international trade and human capital, that we have included their proxies in the estimation model. These variables constitute a channel that promotes technological transfer through which convergence happens.

In line with the convergence theory, the major number of countries with negative fixed effects are OECD countries, see table (D.5). As shown in the histogram (4.3), countries with negative fixed effects regarding efficiency score are in majority advanced countries with the (14) OECD members. While the rest of the listed countries are developing and emerging economies. Developing countries are more capable of creating growth than advanced, which explains their positive impact on the efficiency score.

Developing and emerging countries can draw upon the more productive technologies of the leaders implying the convergence of these countries toward the technological frontier. While the diminishing returns per worker to capital of those low and middle

15. [Pang and Herrera \(2005\)](#) explains this situation by the Balassa-Samuelson effect.

income countries has an important implication, that it is easier for a country to grow fast if it starts out relatively poor, this effect is called the catch-up effect, (Korotayev and Zinkina, 2014). In this area, education efforts can play a crucial role in fostering the catch-up effect by offering the basic formation necessary to the workforce for the adoption of new imported technologies from the west, which leads to facilitate the technological diffusion from the more advanced countries to the developing ones. Thereby, the hypothesis of convergence is the most likely one.

4.5 Conclusion and policy implications

This chapter provides a two-stage efficiency estimation approach that we implement it on after imputation complete data set. Our main goal was to estimate the contribution of public education expenditure in economy efficiency of (76) countries around the world. In first stage, inspired from the ease and robust use of QR in efficiency studies proposed by Behr (2010), we used a longitudinal quantile regression approach to calculate the frontier production function, then we compute efficiency score for countries located in four quantiles $\tau = 0.25, 0.50, 0.75, 0.95$.

Estimation function is composed of GDP growth rate as the output with a battery of inputs under indebtedness circumstances, including mainly: education expenditure, and average years of schooling, together with further growth determinants. Economy efficiency scores suggest an interesting variation across quantiles, with negative scores for countries at the lower tail of growth conditional distribution. Public education spending has an immediate negative influence on economy performance, against positive lagged influence measured by schooling years for person aged above (25) years. The size of education expenditure impact does not widely differ between countries at the upper or lower tail of dependent variable conditional distribution, whereas the marginal impact of schooling years is higher in most efficient members.

The second stage is accomplished using BMA approach with individual fixed effects, with which we check the association between educational system related variables and economy efficiency. Results have shown a negative non-linear U-inverted association between each of: rural population, student per teacher ratio with efficiency score, while OECD dummy variable confirms a negative link between its members and efficiency realization that also has been confirmed through the estimated individual fixed effects. The negative association of rural population is attenuated for OECD countries.

From second stage estimation results emerge many policy implications in order to improve the function of education expenditure then the economy efficiency. First off, it is true that rural areas are disadvantaged regarding public services provision. That because providing distant and isolated areas with services, including education, is extremely difficult and expensive. At the same time, government needs to reduce public spending add further challenges to decision makers. This problem, of rural areas, extends to advanced countries as well as developing countries, where the fraction of rural population is higher. The rural-urban distortion is responsible for generating much inequality within countries, which is represented by divergent living standards¹⁶. While [Young \(2013\)](#) did not find a major role of education in explaining this gap, in Vietnam, [Fesselmeyer and Le \(2010\)](#) state that education is one of the primary explanations for the widening gap. Government policies are the primary responsible of this distortion, and of creating more benefits for urban areas at the expense of rural dwellers. That is why, public policies should be adjusted in favor of rural areas, through establishing new policies that take the needs of rural dwellers into account, and assign expenditures in a manner to offer the appropriate environment to enhance education system outputs and reduce rural-urban inequality. Such as investing in the transportation sector in order to alleviate attending school costs, rising schools numbers, or guarantee an easier access to public services.

Second, since a rising number of student per teacher has a negative influence on educational system outputs it would be advisable to reduce the number of students per teacher by devoting more resources to hiring. Equally proved in [Angrist and Lavy \(1999\)](#) who state that a reduction in predicted class size of ten students is associated with a 0.25 standard deviation increase in fifth graders' scores. While [Rivkin et al. \(2005\)](#), consider achievement gains are related to teaches and schools characteristics rather than numbers.

Third, even though policy stability does not belong to the model, its stills an important condition for the good function of all economy expenditures, mainly education ones, that decision makers have to assure. Additionally, returns to investment in education are more pronounced after economic reforms, supported in [Dreze and Sen \(2002\)](#) study. They note that even though China has invested much more heavily in education than India, they both have the same growth rates, before China has started market reforms, which includes the expansion of rural infrastructure. European Union (EU) members had also applied some policies reforms to improve public expenditure

16. Living standards reflect a selection based upon unobserved skills and human capital, where one out of five or four better educated rural workers migrate to urban areas to enjoy higher wages compared to urban residents, ([Young, 2013](#))

efficiency, here some of EU reforms as shown in [Mäkeläinen \(2010\)](#):

1. First of all, EU members have started by reforming public administration in order to achieve efficiency gain.
2. Governors should consider all budgetary resources, then focus on medium-term budgetary planning to improve public sector efficiency.
3. Reforming human resources management, like adopting flexible working opportunities, and performance pay.
4. Reducing administrative costs and enhancing the quality of delivered services by creating the possibility of the interaction with the public via internet.

In addition to the institutional reform, which find its support in [Ramani \(2014\)](#), who states that one of the real drivers of innovation for economic growth and inclusive development is the *institutional reform*¹⁷.

Moreover, reducing the portion of illiterate people in the population would have an incentive role in rising education expenditure efficiency, since educated families will afford private resources to offer their children more education. Finally, all previous implications should be considered with respect to countries specific characteristics and needs particularly, economic development level. Without ignoring the crucial condition of transparency that government should assume while practicing public spending in general in order to avoid any possible deviation of public targets. Expenditure cut off policy is an ultimate need to face the last financial crisis over the world, but decision makers should account for the evidence that reducing education expenditure like reducing salaries or cutting the number of teachers and other staff as a way to assume efficiency, can have a counterproductive effect because it discourages teachers from wanting to enter or remain in the profession, ([OECD, 2011a](#)).

Finally, such a cross country efficiency analysis, allows the comparison of countries' performance. An individual analysis could be more interesting from policy makers viewpoint.

17. To which, [Ramani \(2014\)](#) adds also: focussed targets, *collectively shared vision* and the presence of techno-entrepreneur leaders.

Chapter 5

Summary and General Discussion

Since four years, we have started our research work in this thesis, motivated by the importance of two key elements: *public debt and human capital vis-à-vis* of economic growth. These two factors are decisive for countries' economic growth sustainability. Human capital stock and accumulation play a crucial growth-enhancing role that has been recognized since the birth of endogenous growth theory, (Romer, 1986; Change, 1990; Lucas Jr, 1988). Henceforth, thanks to human capital, no more diminishing returns to capital¹, moreover human capital can lead to long-run per capital growth in the absence of exogenous technological progress, (Barro and Sala-i Martin, 1995). It has been usual to identify the relationship relating human capital to economic growth, in both theoretical and empirical studies, through the formal education.

Increasing public debt levels caused by last financial crisis have nourished the historical debate about debt influence on subsequent growth rates, as well as on human capital accumulation. That high public debt level threatens economic stability and functioning through different channels, such as: rising the risk of fiscal crisis, imposing costs on the economy represented by high cost of borrowing, discouraging private and foreign investment as a result of lack of confidence in economy credibility, and constraining the flexibility of fiscal policy. As a consequent, growth levels decrease. Moreover, after-debt adjusting policies could hurt the process of human capital production and accumulation, also proved through the theoretical model developed by Greiner (2008b), which has alluded to the adverse impact of rising debt level followed by a strict adjusting policies on education expenditure then on human capital formation.

Nonetheless, both of economic growth and human capital are necessary elements

1. Capital is composed of both physical and human capital

to reduce debt levels. Economic growth represents a soft exit for countries, that reduces or even eliminates the need of a painful restructuring, repression or inflation, (Reinhart and Rogoff, 2013). While, assuming high levels of human capital accumulation contributes, from one hand, in improving economic growth. From another hand, this accumulation reduces debt levels through guaranteeing a catch up effect for poor countries and sustainable growth rates for rich countries.

Thus, and due to the importance of the aforementioned elements and the return of fiscal crisis accompanied with high debt levels, our main goal in this thesis was to explore economic growth, debt and human capital relationships. Then, in second place, we aimed to evaluate the contribution of education expenditure in economy performance in order to derive some policy implications that assure an efficient use of public education expenditures. In our pursue to achieve thesis goals, we started the long research path by firstly, adopting a growth baseline model augmented with human capital and public debt, that constitutes a part of the literature on economic growth and in particular the endogenous growth models with human capital accumulation. Then we selected growth determinants to estimate aimed relationships. Doing so, we were, like many other researchers, interested in economic growth analysis, in face of the issue of guidance absence to select growth model and its determinants. Looking through growth literature, we have proceeded to construct our model's elementary inputs. But since there is a huge register of growth published papers, we have been inspired to implement a content analyzing using a new kind of analysis tools, namely *Textual Analysis*, on economic growth studies, focusing on human capital as growth deriving force. Thus, using statistical technics on textual data we could explore the content of (65) published papers' abstracts, in order to detect the hidden research streams.

Afterwards, we proceeded to the collection of data on the chosen macroeconomic indicators, that will be later submitted to different estimation approaches. At this step again, we had a well known concern: *Missing Values*, that is recognized to have serious consequences and influences on results validity. Even though Missing Values is a very common issue in macroeconomic data bases, through our work of this thesis, we did not mention any reference to it in the different empirical papers we referenced in this thesis. That is why we think that handling this issue in our data set using the most appropriate method, *Multiple Imputation*, developed by the seminal work of Rubin (1976) , gives our inferences further validity and features compared to others. Coming to estimation methods, we mainly employ the semi-parametric approach of *Quantile Regression*. In a linear mode, QR regresses model explanatory variables on the whole conditional distribution of the dependent variable, allowing thereby, to account for non-linearity. It also proposes a specific estimator that fits panel data structure,

that is *Quantile Regression with Penalized fixed effects*. This QR extension is well suited when there is an important risk of correlation between model covariates. While, *Quantile Regression with Correlated Random Effects* is more appropriate to address *endogeneity* issue. Moreover, this semi-parametric method is the most appropriate to estimate frontier production function and efficiency score. Further QR advantages evoked earlier in this thesis, underlie also behind the choose of this method.

Consequently, four years' efforts have resulted in three empirical chapters, where two databases have been constructed and exploited using different and appropriate estimation methods. The first database is textual composed of (65) articles' abstracts published in (15) influent journals between 2011 to 2014, while the second database contains data on (38)² macroeconomic indicator across (76) countries from 1990 to 2011. List (5.1) presents briefly the main contribution of this thesis which are of empirical nature:

Chapter	Contribution
Global contributions The empirical contributions of	Methodologically, this thesis is built on a vast economic literature concerning each of growth, debt and human capital. this thesis represent the input and the motivation of a next theoretical work that is more able to simulate reality thanks to improved quality data and results.
Chapter 2	In this chapter, the main empirical contribution consists of providing an empirical application of using textual analysis in economic growth literature. Which to our knowledge, represents one of the first qualitative analysis of economic growth papers.
Chapter 3	In consistence with each of Ferreira (2009) and Woo and Kumar (2010) findings, the principal contribution of this chapter consists of adding new empirical evidence, to the existing ones, of the negative and heterogeneous debt impact on economic growth under different indebtedness conditions and levels.
Chapter 4	Thanks to the superior use of quantile regression approach in estimating the frontier production function. Through the whole growth rate conditional distribution, we could identify a general negative contribution of education expenditure in economy efficiency. Also, student per teacher ratio and the rural population fraction have a U-inverted impact on efficiency level.

Table 5.1: Thesis contributions

2. (28) variables used in the second chapter, (5) in the third and (5) are used only in imputation model.

In Chapter 2, we remind the issue of growth empirics uncertainty where no common consensus supports the presence of a universal growth model and determinants set, to which the abundance of growth literature adds further difficulty. Following all these facts, we were motivated to explore the content of a number of growth published papers using *textual analysis*. Textual analysis is an interesting analysis technic implemented recently in economic literature. Doing so we do not claim resolving uncertainty problem, but at least we provide an analytical and explorative application of a huge number of growth papers simultaneously which appears long and difficult to do at once. Thereby, thanks to textual analysis visual outputs and contingency tables, we could identify the presence of (6) growth determinants research streams, where *human capital* as a growth engine occupies two of them reflecting the ascending weight associated to human capital since the seminal works of [Change \(1990\)](#) and [Lucas Jr \(1988\)](#). The presence of human capital as growth factors has two contexts: in the first one it is consistent with the classical neoclassical vision of human capital as labor force. Whereas the second context focuses on human capital production process represented by the educational system and the formation activities. Through the (6) streams, we had a stunning and paradoxical remark concerns the absence of human capital from the durable and environmental economic growth research stream, where it is supposed to be a crucial factor to sustain economic growth with respect to environment.

In Chapter 3, motivated by the last financial crisis all over the world and the increasing importance devoted to human capital since the last revelation of the endogenous growth theory, we analyze the interaction between economic growth, human capital and public debt. Our interest was to evaluate how could debt impact economic growth and the relationship between growth and human capital, implementing an appropriate estimation approach that accounts for the multiplicity of growth trajectories and the likelihood of non-linear debt-growth relationship through a simple linear framework. Thus, contrarily to the latent class stochastic frontier model employed by [Bos et al. \(2010\)](#) to identify growth clusters, we account for both growth heterogeneity paths using the semi-parametric approach of quantile regression where sample members are gathered in homogenous groups depending on GDP growth rates, and growth uncertainty here for each quantile we will be able to identify the most important growth determinants. Far from studying debt-growth causality which is confirmed by [Ferreira \(2009\)](#) to be bi-directional, our final results show heterogeneous and significant correlation between debt and growth in all quantiles, negative correlation in general, starting from debt average of (50.87%)³ to GDP, in consistence with the constatation of [Eberhardt and Presbitero \(2013\)](#) who claims that the common threshold of (90%) is likely

3. The minimal debt average over the four quantiles $\tau = 0.25, 0.50, 0.75, 0.95$ equals (50.87%)

to be the outcome of empirical misspecification. Moreover, ignoring public debt from growth determinants has led to a negative contribution of schooling years in growth production, which becomes positive once we account for debt again. Debt heterogeneous slopes between quantiles imply the rejection of linearity hypothesis, neglecting this aspect can cause a quite serious problem as argued by [Ul Haque et al. \(1999\)](#). Because henceforth, adjustment policies are not identical for all analyzed units, rather this means policies should be customized in light of each country needs. Human capital measured by education expenditure and schooling years, shows a negative, variant, and statistically different from zero effect on the different levels of growth rates, the larger effect was that of countries with lower growth rates and higher debt levels. While surprisingly, schooling average years has a positive and significant impact. The marginal effect of a supplement schooling year is the largest for countries on the top of growth distribution where countries have high growth rates and low schooling years levels. Robustness check has been made through the application of different estimation methods such as *panel data model*, *stochastic frontier analysis*, *quantile regression with correlated random effects*, and *penalized fixed effects*. Also by augmenting estimation model with different sets of growth factors. PFE-QR results are robust to the different robustness check. Especially that PFE-QR approach is well suited with the high number of predictors, which could be correlated either together or with the dependent variable, which is a famous issue in growth empirics, including our case. Controlling for unobserved heterogeneity, CRE-QR estimates have revealed the absence of endogeneity caused by the correlation between model regressors and unobserved time invariant specific effects.

To sum up, these results imply different policy implications. First of all, countries with low growth rates and high average debt should establish a public debt consolidation policies depending on further investing human capital, since every additional schooling year increases growth rates whatever growth rates are, in order to sustain growth and attenuate debt effects and levels, all in respecting the recommendation of [Greiner \(2008b\)](#) theoretical model results which suggest that government should pay an equilibrium attention to stabilize debt, without being very restrictive in order to avoid the loss of human capital formation positive effects, or very lose to sustain debt levels. To this objective, education expenditure should be devoted in a way to increase the stock of schooling years, beside offering the appropriate environment that eases agents to attain objective.

Given the importance of investing in human capital to improve growth rates thereafter reduces debt levels, together from the fact of the importance education expenditure size in GDP, in Chapter 4 we have proceeded to estimate the contribution of education expenditure in economy efficiency. Equally to [Behr \(2010\)](#), we use the *quantile*

regression method to evaluate the frontier production function, that QR has many technical advantages regarding the other estimation methods such the parametric approach so called SFA and the non-parametric methods such DEA and FDH. Our evaluation procedure has been made in two steps, where the aim of second step is to elaborate some policies to ameliorate economy efficiency and improve the contribution of education expenditure in.

In counterpart of the mass of efficiency estimation methods, implementing QR approach allows the investigation of a battery of the response variables determinants which results in a variant efficiency scores at a scale of countries such the case, and accounts for direct and indirect effects of education expenditure on economy efficiency. Estimates suggest a negative direct impact of education expenditure on economic performance, mainly in the most efficient top quantiles $\tau = 0.75, 0.95$. *Bayesian Model Averaging with fixed effects* second stage analysis suggests significant countries fixed effects, where a high number of advanced countries have a negative effect on efficiency score while less advanced economies have a positive impact confirmed through the convergence hypothesis. Rural population and student per teacher ratio are two potential determinants that are negatively associated with economy efficiency. These results are more interesting for developing countries which represent (55.2%) of our sample where the fraction of rural population and the number of students per teacher (46.62) and (29.624) respectively are higher compared to those in developed economies, consequently developing countries need to put more effort in managing education expenditure in order to rise their efficiency. Through allocating more expenditure to rise teachers' number, and facilitating the access to schools in rural areas, moreover public policies should adjust rural-urban gap through establishing new policies that take the needs of rural dwellers in account, and assign expenditures in a manner to offer the appropriate environment to enhance education system outputs and reduce rural-urban inequality. Such investing in transportation sector to alleviate attending school costs, rising schools numbers, or guarantee an easier access to public services. Together with the application of the World Bank suggestions on macro level as denoted by [Altinok \(2005\)](#), in order to rise educational system efficiency. The World Bank proposes to governments to guarantee concurrence through the contribution of private sector and civil society in educational process, liberty of educational choices with providing qualified services, orienting educational policies toward poor population, encouraging a decentralized educational system which enables poor of affecting policies to meet their needs.

5.1 Limitations and future perspectives

From one side, while preparing this thesis we have faced some limitations that mainly concern data collecting. For instance, data set in Chapter 2 is restrained to include the abstract of only (65) published papers, concerned with economic growth subject and its determinants, due to the specific corpus construction procedure that we have already detailed in Chapter 2. Consequently, concluding results could be biased given the specificity of data collection. Whereas, the main limitation concerning Chapters 3 and 4' data set was tied to research topic, where the access to data on *public debt and education expenditure* is not allowed or easily obtained. This difficulty may reflect the rigidity of some authorities about publishing public data and having a good level of transparency, especially in what concerns public debt records.

As a result, our sample was built to contain only (76) countries, where further countries could be added if and only if we were willing to have shorter than 1990 – 2011 as time series, and (28) variables where adding further variables such as: *expenditure per educational level, and enrolment rate per level*, in Chapter 4, it could be useful to establish more precise policy implication. Finally, it is true that we find in multiple imputation procedure a practical solution to overcome the handicap of data missingness and have more credible inferences, even though we caution readers that more deepening verification on final results would be welcomed.

From another side, at the end of this thesis where many technical tools and concepts have been explored, we see through it the start of further future works. **First proposition** could be made in line with Chapter 2 methodological path, where the application of textual analysis technic could be generalized in growth literature to explore for instance the presence of co-citation network between growth empirical published papers. That because a network design a mapping of the most cited authors in a specific research field. Therefor, the co-citation concept refers to a statistical approach that allows to locate beside the most prominent works, authors of a specific research field, the intensity of association and the ramification entre researchers, (Boissin et al., 2000).

The second one is tied firstly to the importance of debt return, and secondly to the proved performance and superiority of quantile regression approach. Thus, it would be interesting for policy implication to explore any heterogeneity of debt determinants within countries and through time. **Finally**, from theoretical perspective, using this thesis empirical results in a theoretical framework could provide an interesting, illustrative, theoretical model. Especially, it might be the first time that theory comes after the application. That is why, in this thesis our main focus was to obtain reliable results by

carrying out the most appropriate econometric technics, to our eyes at least, on after-imputation, improved quality data.

Bibliography

- Abayomi, K., Gelman, A., and Levy, M. (2008). Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(3):273–291.
- Abrevaya, J. and Dahl, C. M. (2008). The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics*, 26(4):379–397.
- Afonso, A., Schuknecht, L., and Tanzi, V. (2010). Public sector efficiency: evidence for new eu member states and emerging markets. *Applied Economics*, 42(17):2147–2164.
- Afonso, A. and St Aubyn, M. (2004). Non-parametric approaches to education and health expenditure efficiency in oecd countries. *ISEG-UTL Economics Working Paper*, (1).
- Agarwala, R. (1983). *Price distortions and growth in developing countries*, volume 1. World Bank-Washington, DC.
- Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, 6(1):21–37.
- Aisen, A. and Veiga, F. J. (2013). How does political instability affect economic growth? *European Journal of Political Economy*, 29:151–167.
- Aizenman, J. and Glick, R. (2006). Military expenditure, threats, and growth. *Journal of International Trade & Economic Development*, 15(2):129–155.
- Alexander, H. and Lamarche (2011). Quantile regression for time-series-cross-section-data. *International Journal of Statistics and Management System*, 6(1-2):47–72.
- Allison, P. D. (2001). *Missing data*, volume 136. SAGE publications.
- Altinok, N. (2005). La banque mondiale et l'éducation en afrique subsaharienne. analyse normative du meilleur système éducatif. *Les cahiers de l'IREDU-CNRS*, 64.

Bibliography

- Amini, S. M. and Parmeter, C. F. (2011). Bayesian model averaging in r. *Journal of Economic and Social Measurement*, 36(4):253–287.
- Andini, M. and Andini, C. (2014). Finance, growth and quantile parameter heterogeneity. *Journal of Macroeconomics*, 40:208–322.
- Angrist, J. D. and Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575.
- Armstrong, H. and Read, R. (2004). The economic performance of small states and islands: the importance of geography. In *islands of the world VIII international conference, Taiwan*.
- Bache, S. H. M., Dahl, C. M., and Kristensen, J. T. (2013). Headlights on tobacco road to low birthweight outcomes. evidence from a battery of quantile regression estimators and a heterogeneous panel. *Empirical Economics*, 44:1593–1633.
- Bachellerie, A. and Couillault, B. (2005). Soutenabilité de la dette publique et crises des pays émergents: présentation des concepts et des instruments de diagnostic. *Revue de la stabilité financière*, (6):69–86.
- Balassone, F., Francese, M., and Pace, A. (2011). Public debt and economic growth in italy. *Bank of Italy Economic History Working Paper*, (11).
- Baldacci, E., Clements, B., Gupta, S., and Cui, Q. (2008). Social spending, human capital, and growth in developing countries. *World development*, 36(8):1317–1341.
- Baril, E. and Garnier, B. (2013). Iramuteq 0.6, interface de r pour les analyses multidimensionnelles de textes et de questionnaires. <http://www.iramuteq.org/>.
- Barnard, J. and Meng, X.-L. (1999). Applications of multiple imputation in medical studies: from aids to nhanes. *Statistical Methods in Medical Research*, 8(1):17–36.
- Barro, R. J. (1979). On the determination of the public debt. *Journal of Political Economy*, 87(5):940–971.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *The quarterly journal of economics*, 106(2):407–443.
- Barro, R. J. and Lee, J.-W. (1993). International comparisons of educational attainment. *Journal of monetary economics*, 32(3):363–394.
- Barro, R. J. and Lee, J.-w. (1994a). *Data Set for a Panel of 138 Countries*. Harvard Univ.

- Barro, R. J. and Lee, J.-W. (1994b). Sources of economic growth. In *Carnegie-Rochester conference series on public policy*, volume 40, 1–46. Elsevier.
- Barro, R. J. and McCleary, R. (2003). Religion and economic growth. Technical report, National Bureau of Economic Research.
- Barro, R. J. and Sala-i Martin, X. (1990). Economic growth and convergence across the united states. Technical report, National Bureau of Economic Research.
- Barro, R. J. and Sala-i Martin, X. (1995). Economic growth. *Boston, MA*.
- Barro, R. J., Sala-i Martin, X., Blanchard, O. J., and Hall, R. E. (1991). Convergence across states and regions. *Brookings papers on economic activity*, 107–182.
- Bauer, M. (2003). Análisis de textos asistidos con programas computacionales. 101–112.
- Baum, A., Checherita-W, C., and Rother, P. (2013). Debt and growth: new evidence for the euro area. *Journal of International Money and Finance*, 32:809–821.
- Beach, J. M. (2009). A critique of human capital formation in the us and the economic returns to sub-baccalaureate credentials. *Educational Studies*, 45(1):24–38.
- Beauchemin, K. R. (2001). Growth or stagnation? the role of public education. *Journal of Development Economics*, 64(2):389–416.
- Behr, A. (2010). Quantile regression for robust bank efficiency score estimation. *European Journal of Operational Research*, 200(2):568–581.
- Bennett, D. A. (2001). How can i deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25(5):464–469.
- Berger, A. N. and Mester, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking & Finance*, 21(7):895–947.
- Bernini, C., Freo, M., and Gardini, A. (2004). Quantile estimation of frontier production function. *Empirical Economics*, 29(2):373–381.
- Blankenau, W. F. and Simpson, N. B. (2004). Public education expenditures and growth. *Journal of Development Economics*, 73(2):583–605.
- Blomstrom, M., Lipsey, R. E., and Zejan, M. (1993). Is fixed investment the key to economic growth? Technical report, National Bureau of Economic Research.
- Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143.

Bibliography

- Bogetoft, P. and Otto, L. (2010). *Benchmarking with DEA, SFA, and R*, volume 157. Springer.
- Boissin, J.-P., Castagnos, J.-C., and Guieu, G. (2000). Stratégie d'entreprise et organisation industrielle: pratiques de recherche. *Revue d'économie industrielle*, 91(1):25–54.
- Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese economic journal*, 1(2):141–162.
- Bos, J. W. B., Economidou, C., Koetter, M., and Kolari, J. W. (2010). Do all countries grow alike? *Journal of Development Economics*, 91(1):113–127.
- Budina, N. and Fiess, N. (2005). Public debt and its determinants in market access countries. *The World Bank, Washington, DC, United States*.
- Buhai, S. (2004). Quantile regression: overview and selected applications. *Ad-Astra-The Young Romanian Scientists' Journal*, 4.
- Bun, M. J. and Windmeijer, F. (2010). The weak instrument problem of the system gmm estimator in dynamic panel data models. *The Econometrics Journal*, 13(1):95–126.
- Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386.
- Caselli, F., Esquivel, G., and Lefort, F. (1996). Reopening the convergence debate: a new look at cross-country growth empirics. *Journal of economic growth*, 1(3):363–389.
- Castellacci, F. and Natera, J. M. (2011). A new panel dataset for cross-country analyses of national systems, growth and development (cana). *innovation and Development*, 1(2):205–226.
- Cecchetti, S. G., Mohanty, M. S., and Zampolli, F. (2011). The real effects of debt.
- Cerra, V. and Saxena, S. C. (2002). What caused the 1991 currency crisis in india? *IMF Staff Papers*, 395–425.
- Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5–46.
- Chandra, A. (2010). Does government expenditure on education promote economic growth? an econometric analysis. Technical report, University Library of Munich, Germany.
- Chang, T. and Chiang, G. (2009). The behavior of oecd public debt: A panel smooth transition regression approach. *Empirical Economics Letters*, 8.

- Change, E. T. (1990). Endogenous technological change. *journal of Political Economy*, 98(5):2.
- Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6):429–444.
- Checherita-W, C. and Rother, P. (2012). The impact of high government debt on economic growth and its channels: An empirical investigation for the euro area. *European Economic Review*, 56:1392–1405.
- Checherita-W, C. D., Hughes Hallett, A. J., and Rother, P. (2012). Fiscal sustainability using growth-maximising debt targets.
- Chernozhukov, V. and Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142(1):379–398.
- Chidmi, B., Solís, D., and Cabrera, V. E. (2011). Analyzing the sources of technical efficiency among heterogeneous dairy farms: A quantile regression approach. *Journal of Development and Agricultural Economics*, 3(7):318–324.
- Ciccone, A. and Jarociński, M. (2010). Determinants of economic growth: Will data tell? *American Economic Journal: Macroeconomics*, 2(4):222–246.
- Cobb, C. W. and Douglas, P. H. (1928). A theory of production. *The American Economic Review*, 139–165.
- Cohen, D. (1993). Low investment and large ldc debt in the 1980's. *The American Economic Review*, 437–449.
- Cohen, D. (1997). *Growth and external debt: A new perspective on the African and Latin American tragedies*. Number 1753. Citeseer.
- Constâncio, V. (2012). Contagion et crise de la dette européenne. *RSF Revue De La Stabilité Financière*, 121.
- Corman, S. R., Kuhn, T., McPhee, R. D., and Dooley, K. J. (2002). Studying complex discursive systems. *Human communication research*, 28(2):157–206.
- Crawford, R. D. (1991). *In the Era of Human Capital: The Emergence of Talent, Intelligence and Knowledge as the Worldwide Economic Force and What it Means to Managers and Investors*. Harper Business.
- Daniel, J., Callen, T., Terrones, M. E., Debrun, X., and Allard, C. (2003). Public debt in emerging markets: Is it too high? *World Economic Outlook*, 113.

Bibliography

- De la Escosura, L. P. and Rosés, J. R. (2010). Human capital and economic growth in Spain, 1850–2000. *Explorations in Economic History*, 47(4):520–532.
- De Looze, M.-A., Roy, A., Coronini, R., Reinert, M., and Jouve, O. (1999). Two measures for identifying the perception of risk associated with the introduction of transgenic plants. *Scientometrics*, 44(3):401–426.
- Dension, E. (1962). Sources of growth in the United States and the alternatives before us (supplement paper 13). *New York: Committee for Economic Development*.
- Dollar, D. (1992). Outward-oriented developing economies really do grow more rapidly: evidence from 95 LDCs, 1976–1985. *Economic development and cultural change*, 523–544.
- Dong, Y. and Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2(1):1–17.
- Doppelhofer, G., Miller, R. I., and Sala-i Martin, X. (2000). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. Technical report, National Bureau of Economic Research.
- Dreze, J. and Sen, A. K. (2002). *India: Development and participation*. Oxford University Press.
- Durlauf, S. N. (2001). Manifesto for a growth econometrics. *Journal of Econometrics*, 100(1):65–69.
- Durlauf, S. N., Johnson, P. A., and Temple, J. R. (2005). Growth econometrics. *Handbook of economic growth*, 1:555–677.
- Dwyer, J. (2011). *Keynes's economics and the question of public debt*. PhD thesis, The University of Sydney.
- Easterby, W. (2005). National policies and economic growth. *Handbook of Economic Growth*, Elsevier, North-Holland.
- Easterly, W. and Easterly, W. R. (2001). *The elusive quest for growth: economists' adventures and misadventures in the tropics*. MIT press.
- Easterly, W. and Levine, R. (2003). Tropics, germs, and crops: how endowments influence economic development. *Journal of monetary economics*, 50(1):3–39.
- Easterly, W. and Rebelo, S. (1993). Fiscal policy and economic growth. *Journal of monetary economics*, 32(3):417–458.

- Eberhardt, M. and Presbitero, A. (2013). This time they're different: Heterogeneity and nonlinearity in the relationship between debt and growth macroeconomics, working paper 13/10).
- Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological methods*, 8(3):322.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 253–290.
- Fernandez, C., Ley, E., and Steel, M. F. (2001). Model uncertainty in cross-country growth regressions. *Journal of applied Econometrics*, 16(5):563–576.
- Ferreira, M. C. (2009). Public debt and economic growth: a granger causality panel data approach. (24).
- Fesselmeyer, E. and Le, K. T. (2010). Urban-biased policies and the increasing rural–urban expenditure gap in vietnam in the 1990s. *Asian Economic Journal*, 24(2):161–178.
- Frey, L., Botan, C. H., and Kreps, G. (2000). Investigating communication. NY: *Allyn & Bacon*.
- Futagami, K., Iwaisako, T., and Ohdoi, R. (2008). Debt policy rule, productive government spending, and multiple growth paths. *Macroeconomic Dynamics*, 12(04):445–462.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(4060):471–479.
- Gerschenkron, A. et al. (1962). Economic backwardness in historical perspective. *Economic backwardness in historical perspective*.
- Gilbert, S. (1995). Quality education: Does class size matter? *Canadian Society for the Study of Higher Education. Professional File*.
- Gourinchas, P.-O. and Jeanne, O. (2013). Capital flows to developing countries: The allocation puzzle. *The Review of Economic Studies*, rdt004.
- Granato, J., Inglehart, R., and Leblang, D. (1996). The effect of cultural values on economic development: theory, hypotheses, and some empirical tests. *American journal of political science*, 607–631.
- Greiner, A. (2008a). Does it pay to have a balanced government budget? *Journal of Institutional and Theoretical Economics (JITE)*, 164:460–476.

Bibliography

- Greiner, A. (2008b). Human capital formation, public debt and economic growth. *Journal of Macroeconomics*, 30(1):415–427.
- Greiner, A. (2011). Economic growth, public debt and welfare: Comparing three budgetary rules. *German Economic Review*, 12(2):205–222.
- Greiner, A. (2013). Debt and growth: Is there a non-monotonic relation? *Economics Bulletin*, 33(1):340–347.
- Greiner, A. and Fincke, B. (2009). *Public debt and economic growth*, volume 11. Springer Science & Business Media.
- Greiner, A. and Flaschel, P. (2010). Public debt and public investment in an endogenous growth model with real wage rigidities. *Scottish Journal of Political Economy*, 57(1):68–84.
- Greiner, A. and Semmler, W. (2000). Endogenous growth, government debt and budgetary regimes. *Journal of Macroeconomics*, 22(3):363–384.
- Grigoli, F. (2014). *A Hybrid Approach to Estimating the Efficiency of Public Spending on Education in Emerging and Developing Economies*. Number 14-19. International Monetary Fund.
- Grossman, G. M. and Helpman, E. (1991). Innovation and growth in the global economy. *Innovation and Growth in the Global Economy*.
- Guérin-Pace, F. (1997). La statistique textuelle. un outil exploratoire en sciences sociales. *Population (french edition)*, 865–887.
- Gujarati, D. N. (2003). *Basic econometrics*. 4th.
- Gupta, S. and Verhoeven, M. (2001). The efficiency of government expenditure: experiences from africa. *Journal of policy modeling*, 23(4):433–467.
- Gupta, S., Verhoeven, M., and Tiongson, E. R. (2002). The effectiveness of government spending on education and health care in developing and transition economies. *European Journal of Political Economy*, 18(4):717–737.
- Han, H. and Miao, Y. (2010). Calculation of local health expenditure efficiencies and empirical study on influencing factors: Dea-tobit analysis based on panel data of 31 provinces in china. *J. Fin. Econ*, 36(5):4–16.
- Harding, M. and Lamarche, C. (2009). A quantile regression approach for estimating panel data models using instrumental variables. *Economics Letters*, 104(3):133–135.

- Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 1251–1271.
- Helliwell, J. F. (1996). Economic growth and social capital in asia. Technical report, National Bureau of Economic Research.
- Helliwell, J. F. and Putnam, R. D. (1995). Economic growth and social capital in italy. *Eastern Economic Journal*, 295–307.
- Hendry, D. F. (1995). *Dynamic econometrics*. Oxford University Press.
- Hendry, D. F. and Krolzig, H.-M. (2004). We ran one regression*. *Oxford bulletin of Economics and Statistics*, 66(5):799–810.
- Hénin, P.-Y. and Ralle, P. (1993). Les nouvelles théories de la croissance: quelques apports pour la politique économique. *Revue économique*, 75–100.
- Herndon, T., Ash, M., and Pollin, R. (2014). Does high public debt consistently stifle economic growth? a critique of reinhart and rogoﬀ. *Cambridge Journal of Economics*, 28:257–279.
- Herrera, S. and Pang, G. (2005). *Efficiency of public spending in developing countries: an efficiency frontier approach*, volume 3645. World Bank Publications.
- Hineline, D. R. (2008). Parameter heterogeneity in growth regressions. *Economics Letters*, 101:126–129.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382–401.
- Honaker, J. and King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581.
- Honaker, J., King, G., and Blackwell, M. (2011). Amelia ii: A program for missing data. *Journal of Statistical Software*, 45(7):1–47.
- Horton, N. J. and Lipsitz, S. R. (2001). Multiple imputation in practice: comparison of software packages for regression models with missing variables. *The American Statistician*, 55(3):244–254.

Bibliography

- Howitt, P., Blanchard, O., Bolton, P., Dewatripont, M., Griliches, Z., and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*.
- Hsu, Y.-C. (2013). The efficiency of government spending on health: Evidence from Europe and Central Asia. *The Social Science Journal*, 50(4):665–673.
- Jawadi, F. and Sousa, R. M. (2014). The relationship between consumption and wealth: a quantile regression approach. *Revue d'économie politique*, 124(4):639–652.
- Jayasuriya, R. and Wodon, Q. (2003). *Efficiency in reaching the millennium development goals*. World Bank Publications.
- Jones, C. (1998). *Introduction to Economic Growth* 2nd edition.
- Kaditi, E. A. and Nitsi, E. I. (2010). Applying regression quantiles to farm efficiency estimation. Technical Report 61081, Agricultural and Applied Economics Association.
- Kaldor, N. (1970). The case for regional policies. *Scottish Journal of Political Economy*, 17(3):337–348.
- Kelley, A. C. and Schmidt, R. M. (1995). Aggregate population and economic growth correlations: the role of the components of demographic change. *Demography*, 32(4):543–555.
- King, G., Honaker, J., Joseph, A., and Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95(1):49–70.
- Knack, S. and Keefer, P. (1995). Institutions and economic performance: Cross-country tests using alternative institutional measures. *Economics & Politics*, 7(3):207–227.
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91:74–89.
- Koenker, R. (2005). *Quantile Regression*. Number 38. Cambridge University Press.
- Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica*, 33–50.
- Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, 1583–1612.
- Kormendi, R. C. and Meguire, P. G. (1985). Macroeconomic determinants of growth: cross-country evidence. *Journal of Monetary Economics*, 16(2):141–163.

- Korotayev, A. and Zinkina, J. (2014). On the structure of the present-day convergence. *Campus-Wide Information Systems*, 31(2/3):139–152.
- Krueger, A. (1998). Reassessing the view that american schools are broken. *Economic Policy Review*, 4(1).
- Krugman, P. (1988). Financing vs. forgiving a debt overhang. *Journal of development Economics*, 29(3):253–268.
- Kumbhakar, S. C., Wang, H.-J., and Horncastle, A. (2015). *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*. Cambridge University Press.
- Kwon, D.-B. (2009). Human capital and its measurement. In *The 3rd OECD World Forum on Statistics, Knowledge and Policy Charting Progress, Building Visions, Improving Life Busan, OECD World Forum, Korea, October, 27–30*.
- Lazaridis, D. (2014). Mixed-effects models with penalized fixed effects.
- Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review*, 31–43.
- Lebart, L. and Salem, A. (1994). *Statistique textuelle*, volume 1.
- Lee, J.-W. and Francisco, R. (2012). Human capital accumulation in emerging asia, 1970–2030. *Japan and the World Economy*, 24(2):76–86.
- Lensink, R. (2001). Financial development, uncertainty and economic growth. *De Economist*, 149(3):299–312.
- Lensink, R., Bo, H., and Sterken, E. (1999). Does uncertainty affect economic growth? an empirical analysis. *Weltwirtschaftliches Archiv*, 135(3):379–396.
- Lensink, R. and Morrissey, O. (2006). Foreign direct investment: Flows, volatility, and the impact on growth*. *Review of International Economics*, 14(3):478–493.
- Levine, R. and Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *The American economic review*, 942–963.
- Ley, E. and Steel, M. F. (2007). Jointness in bayesian variable selection with applications to growth regression. *Journal of Macroeconomics*, 29(3):476–493.
- Ley, E. and Steel, M. F. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, 24(4):651–674.

Bibliography

- Liu, C., Laporte, A., and Ferguson, B. S. (2008). The quantile regression approach to efficiency measurement: insights from monte carlo simulations. *Health economics*, 17(9):1073–1087.
- Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, 22(1):3–42.
- MacMurray, E. and Shen, L. (2010). Textual statistics and information discovery: Using co-occurrences to detect events. In *VSST Conference, Toulouse, France*.
- Mäkeläinen, U. (2010). Efficiency and effectiveness of public expenditure on tertiary education in the eu, country fiche finland, joint report by the economic policy committee (quality of public finances) and the directorate-general for economic and financial affairs'. *European Economy Occasional Papers*, (70).
- Mandl, U., Dierx, A., and Ilzkovitz, F. (2008). The effectiveness and efficiency of public spending. Technical report, Directorate General Economic and Monetary Affairs (DG ECFIN), European Commission.
- Mankiw, N. G., Romer, D., and Weil, D. N. (1992a). A contribution to the empirics of economic growth. *The quarterly journal of economics*, 107(2):407–437.
- Mankiw, N. G., Romer, D., and Weil, D. N. (1992b). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2):407–437.
- Masters, W. A. and McMillan, M. S. (2001). Climate and scale in economic growth. *Journal of Economic Growth*, 6(3):167–186.
- Mello, M. and Perrelli, R. (2003). Growth equations: a quantile regression exploration. *Quarterly Review of Economics and Finance*, 43:643–667.
- Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, 538–558.
- Montgomery, J. M. and Nyhan, B. (2010). Bayesian model averaging: Theoretical developments and practical applications. *Political Analysis*, 18(2):245–270.
- Moral-Benito, E. (2012). Determinants of economic growth: a bayesian panel data approach. *Review of Economics and Statistics*, 94(2):566–579.
- Moral-Benito, E. (2014). Growth empirics in panel data under model uncertainty and weak exogeneity. *Journal of Applied Econometrics*.

- Moudatsou, A. (2003). Foreign direct investment and economic growth in the european union. *Journal of Economic Integration*, 18(4):689–707.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 69–85.
- Murtagh, F. (2005). *Correspondence analysis and data coding with Java and R*. CRC Press.
- Myrdal, G. (1957). *Economic Theory and Under-Development Regions*. Gerarld Duckworth.
- OECD (2011a). Education at a glance 2011: Oecd indicators.
- OECD (2011b). Oecd factbook 2011-2012.
- Orchard, T., Woodbury, M. A., et al. (1972). A missing information principle: theory and applications. In *Proceedings of the 6th Berkeley Symposium on mathematical statistics and probability*, volume 1, 697–715. University of California Press Berkeley, CA.
- Pang, G. and Herrera, S. (2005). Efficiency of public spending in developing countries: an efficiency frontier approach. *World Bank Policy Research Working Paper*, (3645).
- Panizza, U. and Presbitero, A. F. (2014). Public debt and economic growth: Is there a causal effect? *Journal of Macroeconomics*, 41:21–41.
- Panizza, U., Presbitero, A. F., et al. (2013). Public debt and economic growth in advanced economies: A survey. *Swiss Journal of Economics and Statistics*, 149(2):175–204.
- Peng, C.-Y. J., Harwell, M., Liou, S.-M., Ehman, L. H., et al. (2006). Advances in missing data methods and implications for educational research. *Real data analysis*, 31–78.
- Perotti, R. (2012). The "austerity myth": gain without pain? In *Fiscal policy after the financial crisis*, 307–354. University of Chicago Press.
- Pescatori, A., Sandri, D., and Simon, J. (2014). *Debt and growth: is there a magic threshold?* Number 14-34. International Monetary Fund.
- Petrakos, G., Arvanitidis, P., and Pavleas, S. (2007). Determinants of economic growth: the experts' view. In *2nd Workshop of DYNREG in Athens*, 9–10.
- Pinto, B. and Ulatov, S. (2010). Russia 1998 revisited: Lessons for financial globalization.
- Pirotte, A. (2011). *Econométrie des données de panel: théorie et applications*. Economica.
- Plosser, C. (1992). The search for growth. *Proceedings - Economic Policy Symposium - Jackson Hole*, 57–86.

- Poirson, M. H., Ricci, M. L. A., and Pattillo, M. C. A. (2004). *What are the channels through which external debt affects growth?* Number 4-15. International Monetary Fund.
- Popescu, I.-I., Mačutek, J., and Altmann, G. (2009). *Aspects of word frequencies*. RAM-Verlag Lüdenscheid.
- Prasetyo, A. D. and Zuhdi, U. (2013). The government expenditure efficiency towards the human development. *Procedia Economics and Finance*, 5(0):615–622.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, 25:111–164.
- Ram, R. (2008). Parametric variability in cross-country growth regressions: An application of quantile-regression methodology. *Economics Letters*, 99:387–389.
- Ramani, S. V. (2014). *Nanotechnology and Development: What's in it for Emerging Countries?* Cambridge University Press.
- Rebelo, S. (1991). Long-Run Policy Analysis and Long-Run Growth. *Journal of Political Economy*, 99(3):500–521.
- Reinert, M. (1987). Classification descendante hierarchique et analyse lexicale par contexte-application au corpus des poesies d'arthur rimbaud. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 13(1):53–90.
- Reinert, M. (1993). Les mondes lexicaux et leur logique à travers l'analyse statistique d'un corpus de récits de cauchemars. *Langage et société*, 66(1):5–39.
- Reinert, M. (1995). Quelques aspects de choix des unités d'analyse et de leur contrôle dans la méthode alceste. *JADT 1995*, 1:27–34.
- Reinhart, C. and Rogoff, K. (2010a). Debt and growth revisited.
- Reinhart, C. and Rogoff, K. (2013). Financial and sovereign debt crises: some lessons learned and those forgotten.
- Reinhart, C. M., Reinhart, V. R., and Rogoff, K. S. (2012). Public debt overhangs: advanced-economy episodes since 1800. *The Journal of Economic Perspectives*, 26(3):69–86.
- Reinhart, C. M. and Rogoff, K. S. (2010b). Growth in a time of debt. Technical report, National Bureau of Economic Research.
- Reinhart, C. M., Rogoff, K. S., and Savastano, M. A. (2003). Debt intolerance. Technical report, National Bureau of Economic Research.

- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 417–458.
- Romer, P. M. (1986). Increasing returns and long-run growth. *The journal of political economy*, 1002–1037.
- Rubin, D. B. (1976). Noniterative least squares estimates, standard errors and f-tests for analyses of variance with missing data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 270–274.
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.
- Sahin, I. E., Mucuk, M., et al. (2014). The effect of current account deficit on economic growth: The case of turkey. In *Proceedings of International Academic Conferences*, number 0301828. International Institute of Social and Economic Sciences.
- Saint-Paul, G. (1992). Fiscal policy in an endogenous growth model. *The Quarterly Journal of Economics*, 1243–1259.
- Sala-i Martin, X. (1997). I just ran two million regressions. *The American Economic Review*, 178–183.
- Sala-I-Martin, X., Doppelhofer, G., and Miller, R. I. (2004). Determinants of long-term growth: A bayesian averaging of classical estimates bace approach. *American Economic Review*, 94(4):813–835.
- Saviotti, P., De Looze, M.-A., and Maupertuis, M. (2005). Knowledge dynamics, firm strategy, mergers and acquisitions in the biotechnology based sectors. *Economics of Innovation and New Technology*, 14(1-2):103–124.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.
- Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, 8(1):3–15.
- Schafer, J. L. and Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2):147.
- Schclarek, A. (2004). Debt and economic growth in developing and industrial countries. *Lund University Department of Economics Working Paper*, 2005:34.
- Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, 51(1):1–17.

Bibliography

- Schultz, T. W. (1988). On investing in specialized human capital to attain increasing returns. *The state of development economics*, 339–52.
- Sinha, P., Arora, V., and Bansal, V. (2011). Determinants of public debt for middle income and high income group countries using panel data regression.
- Sleezer, C. M., Conti, G. J., and Nolan, R. E. (2004). Comparing cpe and hrd programs: Definitions, theoretical foundations, outcomes, and measures of quality. *Advances in developing human resources*, 6(1):20–34.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics*, 70(1):65–94.
- Sutherland, D., Price, R. W., Joumard, I., and Nicq, C. (2007). Performance indicators for public spending efficiency in primary and secondary education. Technical report, OECD Publishing.
- Tabachnick, B. G., Fidell, L. S., et al. (2001). Using multivariate statistics.
- Temple, J. (2002). Growth effects of education and social capital in the oecd countries. *Historical Social Research/Historische Sozialforschung*, 5–46.
- Thanassoulis, E. (2001). *Introduction to the theory and application of data envelopment analysis*. Springer.
- Trevisani, M. and Tuzzi, A. (2014). A portrait of jasa: the history of statistics through analysis of keyword counts in an early scientific journal. *Quality & Quantity*, 1–18.
- Turnovsky, S. J. (2000). *Methods of macroeconomic dynamics*. Mit Press.
- Turnovsky, S. J. and Fisher, W. H. (1995). The composition of government expenditure and its consequences for macroeconomic performance. *Journal of Economic Dynamics and Control*, 19(4):747–786.
- Tyler, W. G. (1981). Growth and export expansion in developing countries: some empirical evidence. *Journal of Development Economics*, 9(1):121–130.
- Ul Haque, N., Pesaran, M., and Sharma, S. (1999). Neglected heterogeneity and dynamics in cross-country savings regressions.
- Ulasan, B. (2011). Cross-country growth empirics and model uncertainty: An overview. *Economics Discussion Paper*, (2011-37).
- Ulku, H. (2004). R and d innovation and economic growth: An empirical analysis. *IMF Working Paper 185*.

- Uzawa, H. (1965). Optimum technical change in an aggregative model of economic growth. *International economic review*, 6(1):18–31.
- Van Buuren, S. (2012). *Flexible imputation of missing data*. CRC press.
- Van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C., and Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of statistical computation and simulation*, 76(12):1049–1064.
- Ventelou, B. and Bry, X. (2006). The role of public spending in economic growth: Envelopment methods. *Journal of Policy Modeling*, 28(4):403–413.
- Wang, H.-J., Chang, C.-C., and Chen, P.-C. (2008). The cost effects of government-subsidised credit: Evidence from farmers credit unions in taiwan. *Journal of Agricultural Economics*, 59(1):132–149.
- Wayman, J. C. (2003). Multiple imputation for missing data: What is it and how can i use it. In *Annual Meeting of the American Educational Research Association, Chicago, IL*, 2–16.
- Whitt Jr, J. A. (1996). The mexican peso crisis. *Economic Review*, (Jan):1–20.
- Woo, J. and Kumar, M. S. (2010). *Public debt and growth*. International Monetary Fund.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
- Young, A. (2013). Inequality, the urban-rural gap and migration*. *The Quarterly Journal of Economics*, qjt025.
- Zak, P. J. and Knack, S. (2001). Trust and growth. *The economic journal*, 111(470):295–321.
- Zeugner, S. (2011). Bayesian model averaging with bms. Technical report, mimeo, Available at <http://cran.rproject.org/web/packages/BMS/vignettes/bms.pdf>.
- Zotteri, S. (2002). Heterogeneity in human capital and economic growth. Technical report, Bank of Italy, Economic Research and International Relations Area.

Bibliography

Appendices

Appendix A

Introduction Chapter Appendix

A.1 IRaMuteQ Description

IRaMuteQ handles corpus in text format, each text in the corpus should be started by four stars, that could be followed by some starry variables. These illustrative variables can make sign to some important information that researcher wants to highlight, thus enabling him to do supplement researches.

Corpus treatment in *IRaMuteQ* is achieved by text segments: the whole text is divided into segments. The size of these segments depends of the number of units {either characteristics, or occurrences} by text. The default number of characteristics by text is 40. The first manipulation of segments is the cleaning, then *IRaMuteQ* realizes text lemmatization using specific dictionary for each language where all verbs are brought to the infinitive, nouns to the singular.

Using *IRaMuteQ* functionality, one could obtain textual statistics for all forms contained in the corpus: {Active forme, supplement forme, hapax, and all form types}. Another functionality provided by this program is the *Factorial Correspondence Analysis* by forms, it is also possible to do this analysis by variable. Further data classification is possible by *Reinert Method*, where sorted data is classified in classes. A detailed manual about *IRaMuteQ* description realized by Lucie Loubère and Pierre Ratinaud¹ is available on program's web site². In this thesis we use *IRaMuteQ version 0.7 alpha 2*.

1. <http://www.iramuteq.org/Members/pierre.ratinaud>

2. <http://www.iramuteq.org/>

A.2 R-environment

As described on R-project web site {www.r-project.org}, R is a language and environment for statistical, computing and graphics within which many modern and classical statistical techniques have been executed, many of these techniques are supplied as *packages*. R is a free software similar to S, that is designed around true computer language. R was initially written by Ross Ihaka and Robert Gentleman at the Department of Statistics of the University of Auckland in Auckland, New Zealand. In addition, a large group of individuals has contributed to R by sending code and bug reports. R is highly extensible, It allows users to add further functionalities by defining new functions. R is an integrated suite of softwares facilities for data manipulation, calculation and graphical display, it provides a wide range of statistical and graphical techniques.

An important difference between S (and hence R) and other statistical tools, while SAS and SPSS give out a copious outputs, R give only a minimal outputs and stores the rest in objects for subsequent interrogations. A number of books about R by R core team members are available now to describe and ease the use of R, like: Software for Data Analysis: Programming with R by John M. Chambers (2008), or Introductory Statistics with R by Peter Dalgaard. (2008). R packages are freely available from the Comprehensive R Archive Network (CRAN) at: {<http://www.r-project.org>}.

Used packages:

1. Amelia package: Amelia II "multiply imputes" missing data in a single cross-section from a time series, or from a time-series-cross-sectional data set. Amelia II is usually considerably faster than existing approaches and can handle many more variables. The program also generalizes existing approaches by allowing for trends in time series across observations within a cross-sectional unit, as well as priors that allow experts to incorporate beliefs they have about the values of missing cells in their data. Amelia II also includes useful diagnostics of the fit of multiple imputation models. The program works from the R command line or via a graphical user interface that does not require users to know R.
2. Package *rqpd* regression quantile for panel data, this package has been used to fit fixed-effects and correlated-random-effects quantile regression models, it is also can do (bootstrap) inference. This package authors are: Roger Koenker Stefan Holst Bache.
3. Developed by Yves Croissant and Giovanni Millo, the package *plm-panel linear models* contains a set of tests and estimators for panel data. have developed *plm*
4. Package *SFA*, it fits stochastic frontier analysis models. SFA has been developed

thanks to: Ariane Straub, under the supervision of Torsten Hothorn.

5. Package *BMS*, the authors of this package are: Martin Feldkircher and Stefan Zeugner. *BMS* fits is designed to estimate linear models with a wide choice of (customizable) priors.

A.3 R packages for multiple imputation

Many softwares offer packages that implement MI procedure one of them is R³ software which provides some packages that implement MI, like: BaBooN; cat; Hmisc; MImix, etc. The most known imputation packages in R are *Multiple Imputation by Chained Equation (MICE)*⁴ and *Amelia*. MICE package executes a multiple imputation by chained equation, it requires the identification of imputation model that corresponds to the nature of each variable in the model. Hence, the imputation model of each variable is unique. The predictors used to impute missed values for one variable are not the same to for another variable. MICE represents *Fully Conditional Specification (FCS)* method of multiple imputation, for more information about FCS, ([Van Buuren et al., 2006](#)). The main drawbacks of this method are summarized by three points:

- (i) The first one is that, each conditional density has to be specified separately, as a consequent very substantial modeling effort can be needed for data sets with many variables as in our case.
- (ii) The second one, is that FCS⁵ is often computationally more intensive than joint modeling, typical computational shortcuts may not be applied.
- (iii) The last and the most important point is that less information are known about the quality of the resulted imputations because the implied joint distribution may not exist theoretically by that convergence criteria are ambiguous. Convergence can be guaranteed under compatibility of a condition that is often difficult to verify in practice, ([Van Buuren et al., 2006](#)).

Different from MICE, *Amelia* offer many advantages to improve imputes plausibility and facility especially the possibility to have a unique imputation model for all variables, keeping at the same time the ability to have some specification for each variable such as; the use of future and past values of each variable (*Lags, Leads*). Moreover, one can impose logic constraints on variables, where certain variables might be bounded to $[0, 1]$ thus restricting imputes to this bound is required. Adding prior information

3. R- environment description is available in the appendix [A](#).

4. MICE is a library distributed for S-Plus and R.

5. Fully Conditional Specification specifies a separate conditional densities for each variable. These densities are used to impute missed values using the observed ones.

is also possible via *Ridge Prior* option, especially when cross section dimension is small or slightly bigger than parameter's number. After imputation step, the new selected values are drawn randomly after the bootstrapping step. Also, different from Amelia first version, *Amelia II* provides a support for analysis imputed data sets and combining results.

A.4 Imputation Codes

A.4.1 Imputation code, m=100

```
library(Amelia)
base<-read.table(file="base.txt",header=TRUE, sep="\t ")
summary(base)

# lags
lags2<-c("GDP", "DEBT", "EDUC", "HEXP", "REVENU", "MEXP", "EXPO",
"IMPO", "SCH", "INTEREST", "LABOR", "GFCF", "INF", "LIFE", "PINV",
"SOLD", "SAVE", "UNEMP", "POP", "CACNT", "FDI", "PPP", "AGR", "CONS",
"XRAT", "OPEN", "EXPENS", "INV")
# leads
leads2<-c("GDP", "DEBT", "EDUC", "HEXP", "REVENU", "MEXP", "exp",
"IMPO", "SCH", "INTEREST", "LABOR", "GFCF", "INF", "LIFE", "PINV",
"SOLD", "SAVE", "UNEMP", "POP", "CACNT", "FDI", "PPP", "AGR", "CONS",
"XRAT", "OPEN", "EXPENS", "INV")

# bounds
# bordures logiques (0/100, 0/1)
# if there is no maximal logic bordre then : max=variable observed
# maximum variable number min max
# DEBT 5 0 235.6
# EDUC 6 0 16.06
# HEXP 8 0 17.673
# REVENU 11 0 67.05
# MEXP 12 0 19.09
# EXPO 13 0 241.40
# IMPO 14 0 219.07
# GFCF 18 0 76.69
```

A.4. Imputation Codes

```
# PINV 21 0 374.82
# UNEMP 24 0 39.3
# PPP 28 0 3863.15
#LAND 30 0 16389950
#AGR 31 0 85.46
#XRAT 37 0 10389.9
#OPEN 38 0 433.05
#EXPENS 39 0 78.09
#INV 40 0 64.10

bds <- matrix(c(5, 6, 8, 11, 12, 13, 14, 18, 21, 24, 28, 30, 31,37,
38,39, 40, 0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 235.6,
16.06, 17.673, 67.05, 19.09, 241.4, 219.07, 76.69, 374.82, 39.3,
3863.15, 16389950, 85.46, 10389.9, 433.05,78.09, 64.1),
nrow = 17, ncol = 3)

# Prior for missing data in HEXP from (1990 to 1994)
# new value equals 1995 +/- 2 standard deviation

pr<-matrix(c(1,2,3,4,5, 23,24,25,26,27, 45,46,47,48, 49, 67,68,69,70,
71,89, 90,91,92,93,111,112,113,114,115,133,134,135,136,137,155,156,
157,158,159,177, 178,179,180,181, 199,200,201,202,203,221,222,223,
224,225,243,244,245,246,247, 265,266,267,268,269,287,288,289,290,
291,309,310,311,312,313,331,332,333,334,335, 353,354,355,356,357,
375,376,377,378,379,397,398,399,400,401,419,420,421,422,423, 441,
442,443,444,445,463,464,465,466,467,485,486,487,488,489,507,508,
509,510,511, 529,530,531,532,533,551,552,553,554,555,573,574,575,
576,577,595,596,597,598,599, 617,618,619,620,621,639,640,641,642,
643,661,662,663,664,665,683,684,685,686,687,705,706,707,708,709,
727,728,729,730,731,749,750,751,752,753,771,772,773,774,775,793,
794,795,796,797,815,816,817,818,819,837,838,839,840,841,859,860,
861,862,863,881,882,883,884,885,903,904,905,906,907,925,926,927,
928,929,947,948,949,950,951, 969,970,971,972,973,991,992,993,994,
995,1013,1014,1015,1016,1017,1035,1036,1037, 1038,1039,1057,1058,
1059,1060,1061,1079,1080,1081,1082,1083,1101,1102,1103,1104,1105,
1123,1124,1125,1126,1127,1145,1146,1147,1148,1149,1167,1168,1169,
1170,1171, 1189,1190,1191,1192,1193,1211,1212,1213,1214,1215,1233,
```



```

library(Amelia)
base<-read.table(file="F:/Articles-code/code+tables+article2/
imputation-base1.txt",header=TRUE, sep="\t" )
summary(base)
attach(base)
edit(base)
# lags
lags2<-c("GDP", "DEBT", "EDUC", "HEXP", "REVENU", "MEXP", "EXPO", "IMPO",
"SCH", "INTEREST", "LABOR", "GFCF", "INF", "LIFE", "PINV", "SOLD",
"SAVE", "UNEMP", "POP", "CACNT", "FDI", "PPP", "AGR", "CONS", "XRAT",
"OPEN", "EXPENS", "INV", "GOVEFF", "LITER", "POLIST", "STU-TEACH", "GINI")
# leads
leads2<-c("GDP", "DEBT", "EDUC", "HEXP", "REVENU", "MEXP", "EXPO",
"IMPO", "SCH", "INTEREST", "LABOR", "GFCF", "INF", "LIFE", "PINV",
"SOLD", "SAVE", "UNEMP", "POP", "CACNT", "FDI", "PPP", "AGR", "CONS",
"XRAT", "OPEN", "EXPENS", "INV", "GOVEFF", "LITER", "POLIST", "STU-TEACH",
"GINI")

# bounds
# bordures logiques (0/100, 0/1)
# if there is no maximal logic bordre then : max=variable observed
# maximum variable number min max
# DEBT 5 0 235.6
# EDUC 6 0 16.06
# HEXP 8 0 17.673
# REVENU 11 0 67.05
# MEXP 12 0 19.09
# EXPO 13 0 241.40
# IMPO 14 0 219.07
# GFCF 18 0 76.69
# PINV 21 0 374.82
# UNEMP 24 0 39.3
# PPP 28 0 3863.15
#LAND 30 0 16389950
#AGR 31 0 85.46
#XRAT 37 0 10389.9
#OPEN 38 0 433.05

```

A.4. Imputation Codes

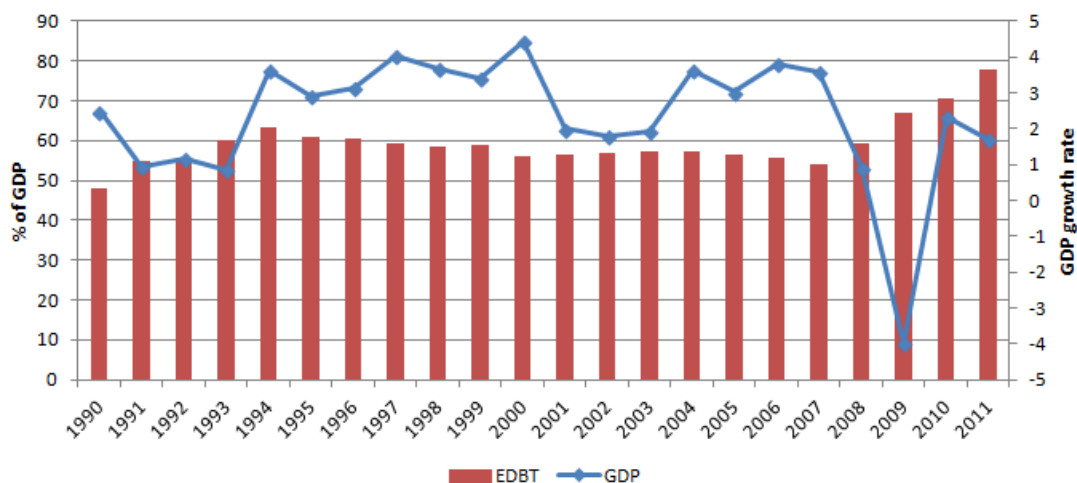
```
#EXPENS 39 0 78.09
#INV 40 0 64.10
#GOVEFF 41 -2.5 2.5
#LITER 42 0 100
#POLIST43 -2.812 2.5
#STU-TEACH 44 0 69,28601
#GINI 46 0 100

bds <- matrix(c(5, 6, 8, 11, 12, 13, 14, 18, 21, 24, 28, 30, 31,37,38,
39, 40, 41,42,43,44,46, 0,0,0,0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, -2.5,0, -2.812,0,0, 235.6, 16.06, 17.673, 67.05, 19.09, 241.4, 219.07,
76.69,374.82, 39.3, 3863.15, 16389950,85.46, 10389.9, 433.05,78.09,
64.1,2.5,100,2.5,69.28601,100), nrow = 22, ncol = 3)

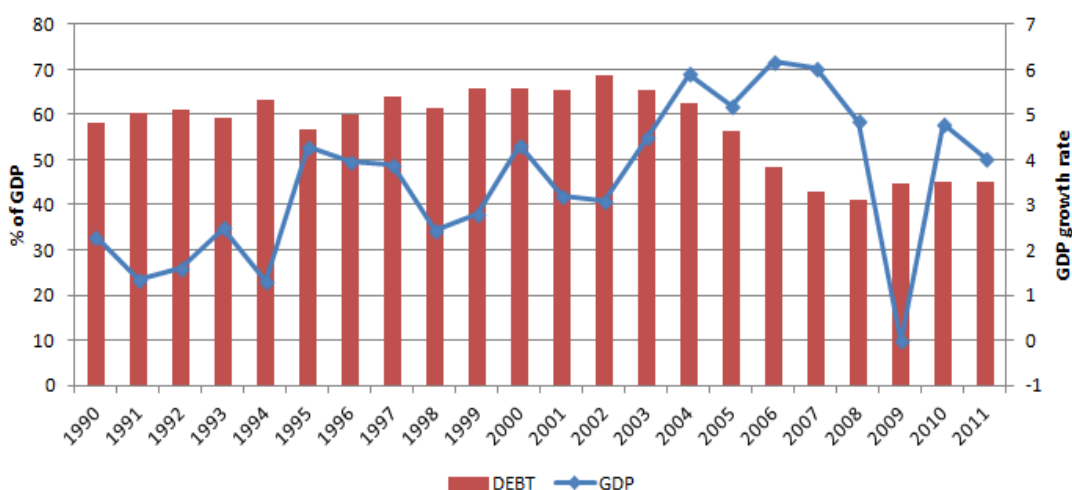
# Prior for missing data in HEXP from (1990 to 1994)
# new value equals 1995 +/- 2 standard deviation

pr<-matrix(c(1,2,3,4,5, 23,24,25,26,27, 45,46,47,48, 49, 67,68,69,
70,71, 89,90,91,92,93,111,112,113,114,115,133,134,135,136,137,155,
156,157,158,159,177,178,179,180,181,199,200,201,202,203,221,222,223,
224,225,243,244,245,246,247,265,266,267,268,269,287,288,289,290,291,
309,310,311,312,313,331,332,333,334,335,353,354,355,356,357,375,376,
377,378,379,397,398,399,400,401,419,420,421,422,423,441,442,443,444,
445,463,464,465,466,467,485,486,487,488,489,507,508,509,510,511,529,
530,531,532,533,551,552,553,554,555,573,574,575,576,577,595,596,597,
598,599,617,618,619,620,621,639,640,641,642,643,661,662,663,664,665,
683,684,685,686,687,705,706,707,708,709,727,728,729,730,731,749,750,
751,752,753,771,772,773,774,775,793,794,795,796,797,815,816,817,818,
819,837,838,839,840,841,859,860,861,862,863,881,882,883,884,885,903,
904,905,906,907,925,926,927,928,929,947,948,949,950,951,969,970,971,
972,973,991,992,993,994,995,1013,1014,1015,1016,1017,1035,1036,1037,
1038,1039,1057,1058,1059,1060,1061,1079,1080,1081,1082,1083,1101,1102,
1103,1104,1105,1123,1124,1125,1126,1127,1145,1146,1147,1148,1149,1167,
1168,1169,1170,1171,1189,1190,1191,1192,1193,1211,1212,1213,1214,1215,
1233,1234,1235,1236,1237,1255,1256,1257,1258,1259,1277,1278,1279,1280,
1281,1299,1300,1301,1302,1303,1321,1322,1323,1324,1325,1343,1344,1345,
```


A.5 Introduction Chapter Figures



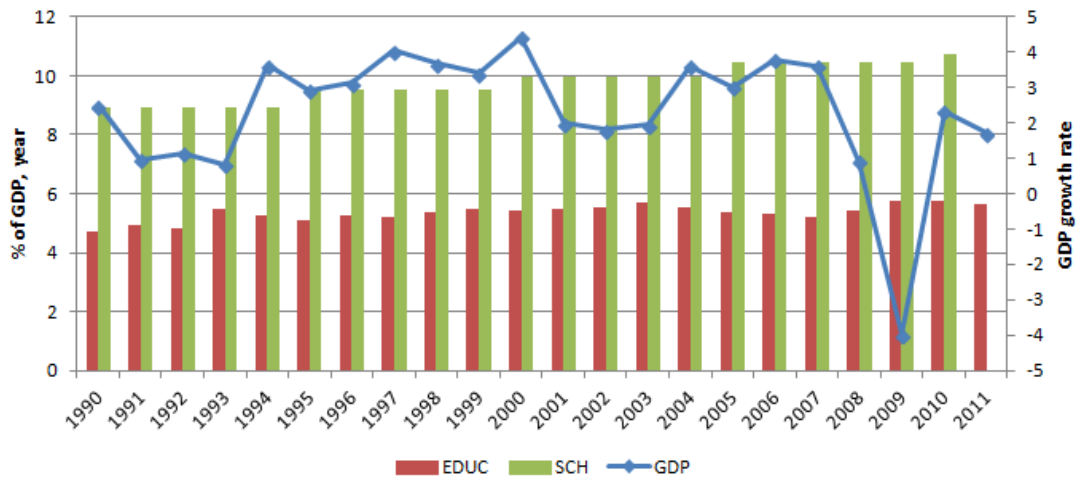
(a) OECD GDP growth rate, and public debt as a % of GDP (1990 – 2011)



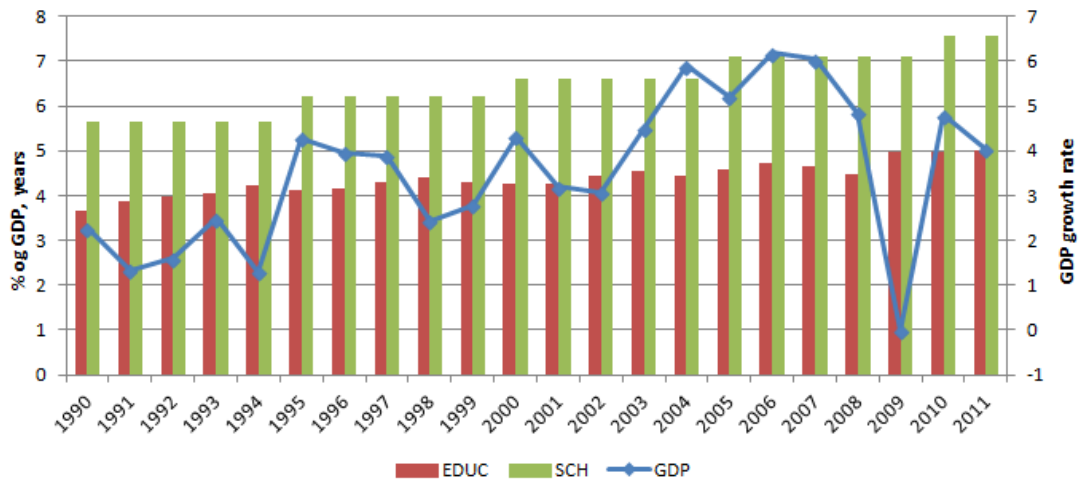
(b) Non-OECD GDP growth rate, and public debt as a % of GDP (1990 – 2011)

Figure A.1: OECD and Non-OECD growth rate, and public debt as a % of GDP (1990 – 2011)

A.5. Introduction Chapter Figures

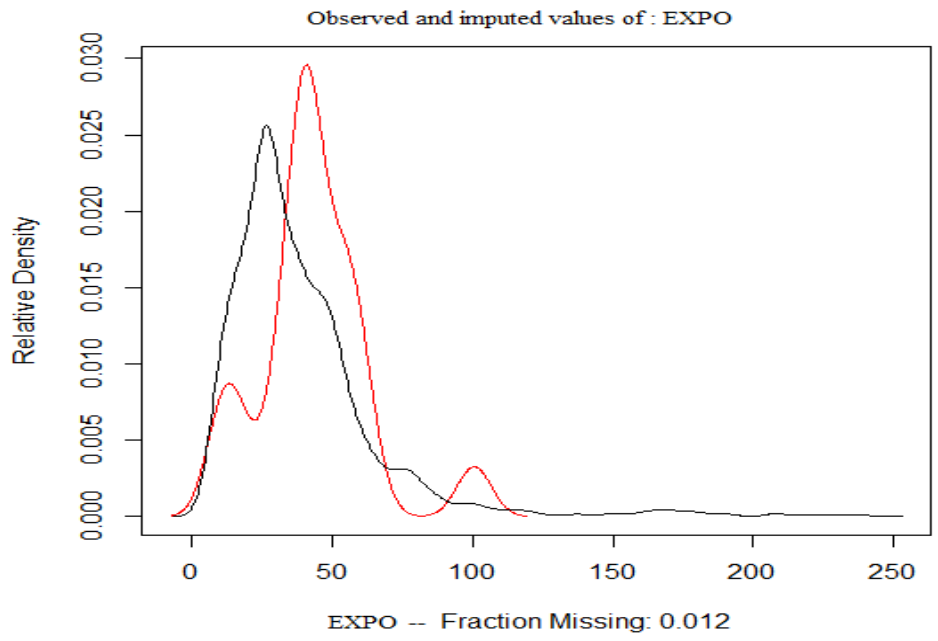


(a) OECD GDP growth rate, and public education expenditure as a % of GDP (1990 – 2011)

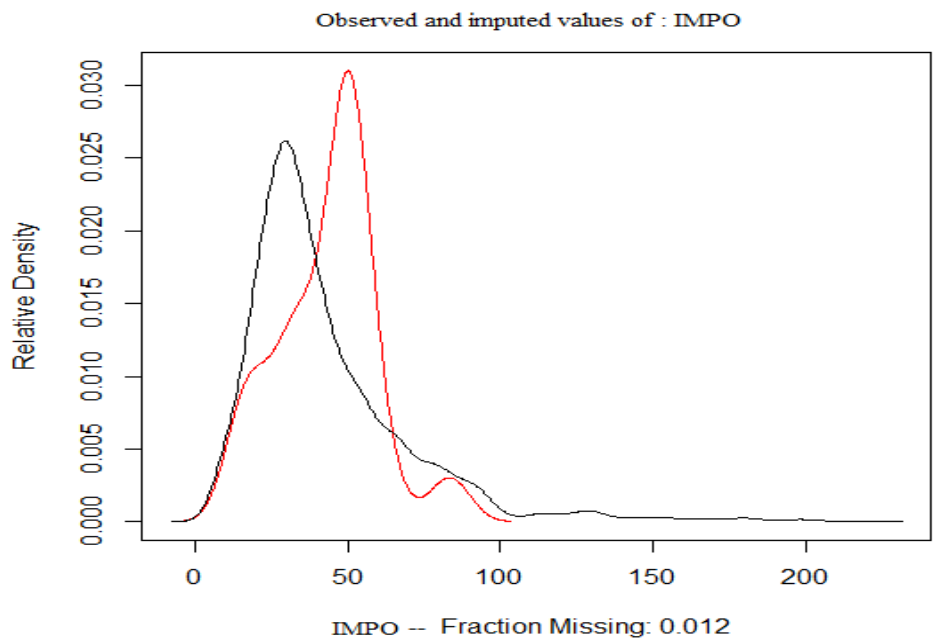


(b) Non-OECD GDP growth rate, and public education expenditure as a % of GDP (1990 – 2011)

Figure A.2: OECD and Non-OECD growth rate, and public education expenditure as a % of GDP (1990 – 2011)

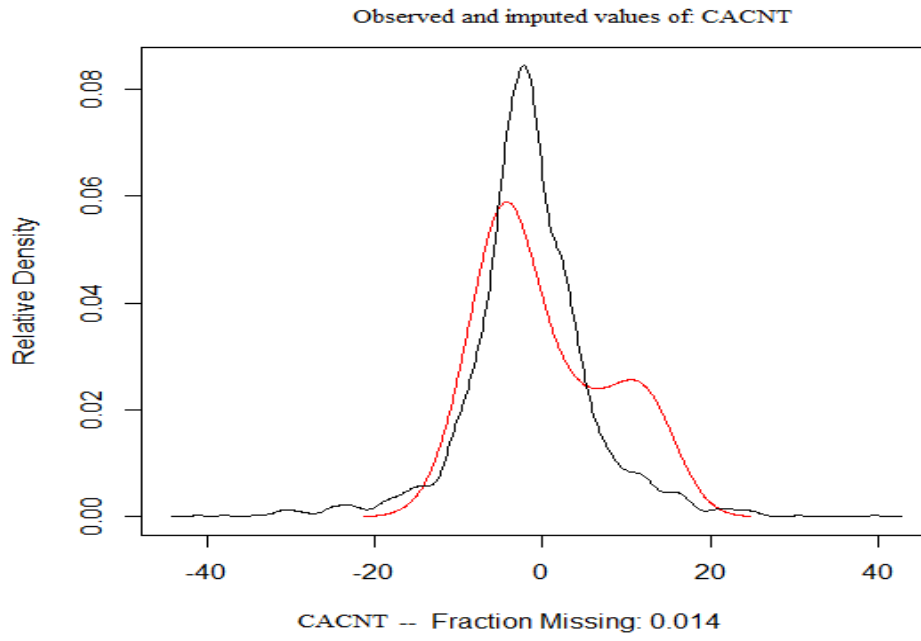


(a)

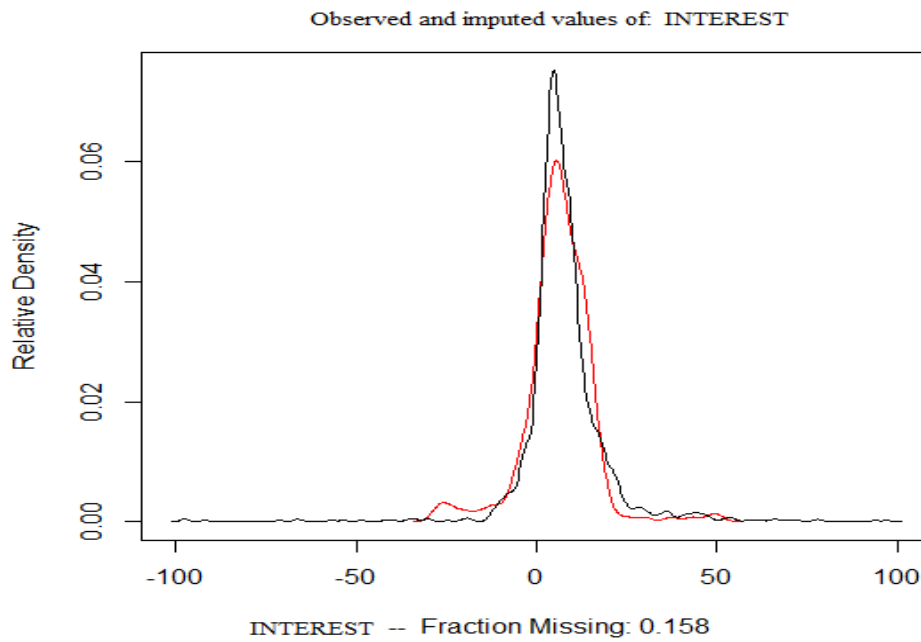


(b)

Figure A.3: Densities Diagnostic IV, $m=100$

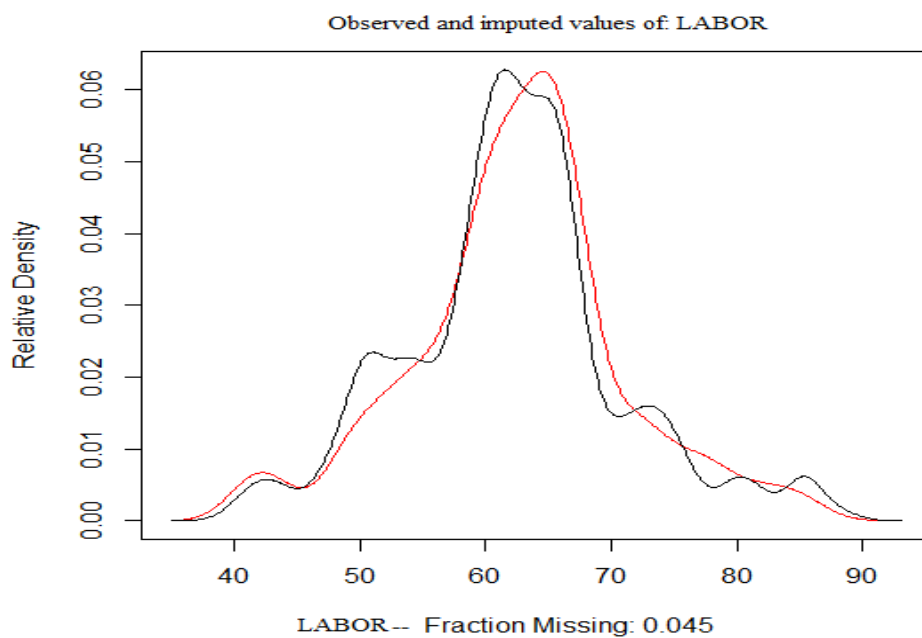


(a)

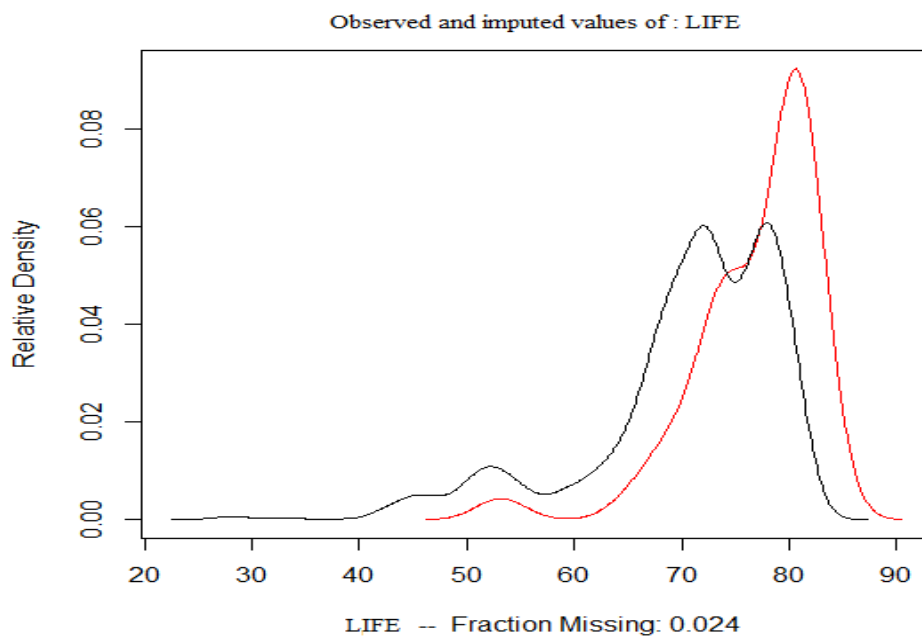


(b)

Figure A.4: Densities Diagnostic V , $m=100$

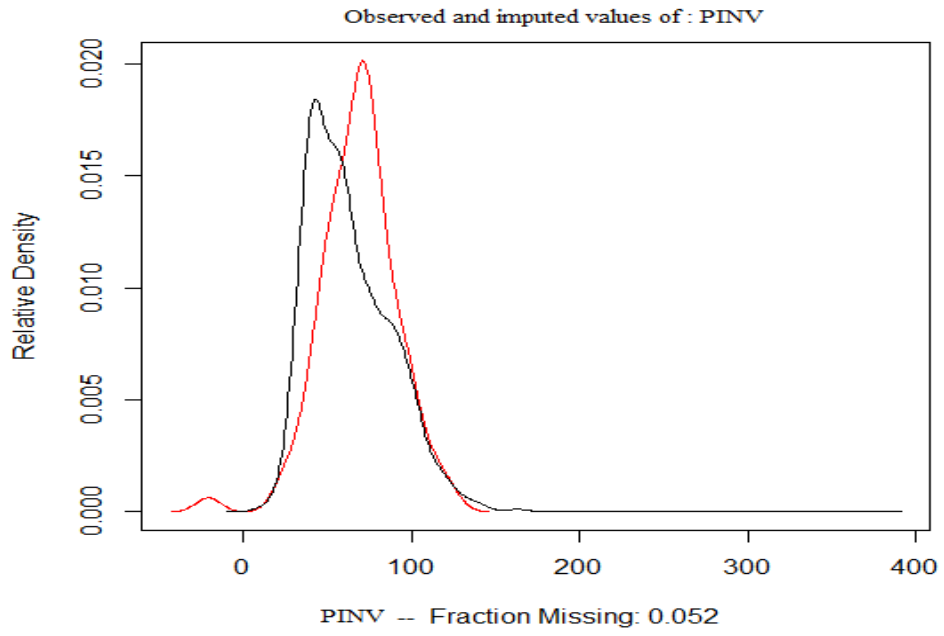


(a)

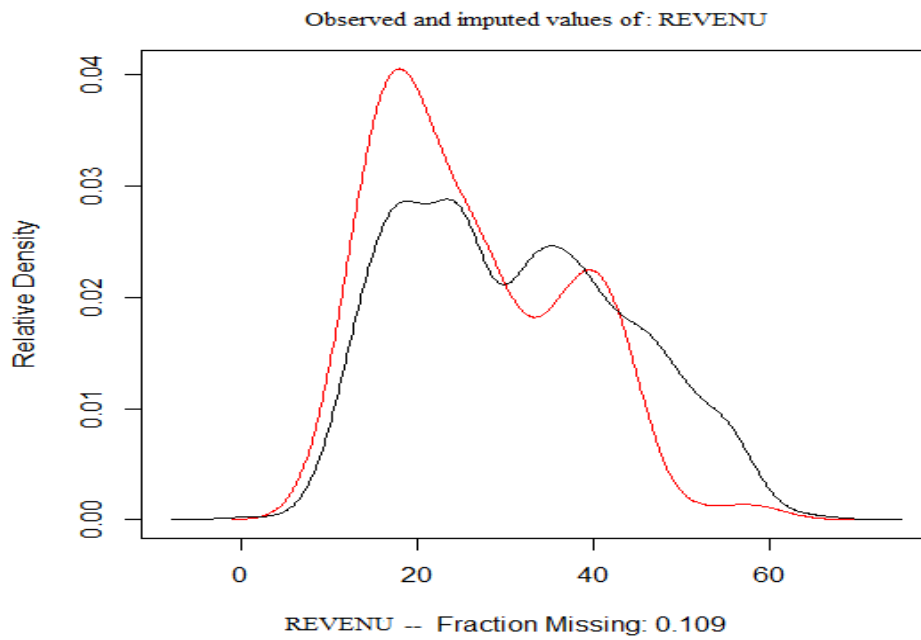


(b)

Figure A.5: Densities Diagnostic VI, $m=100$

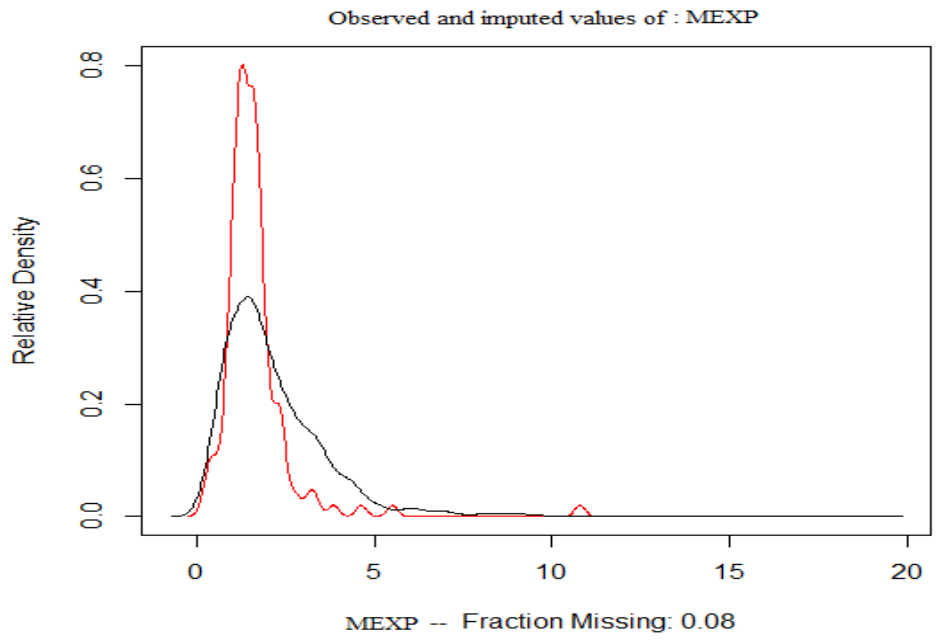


(a)

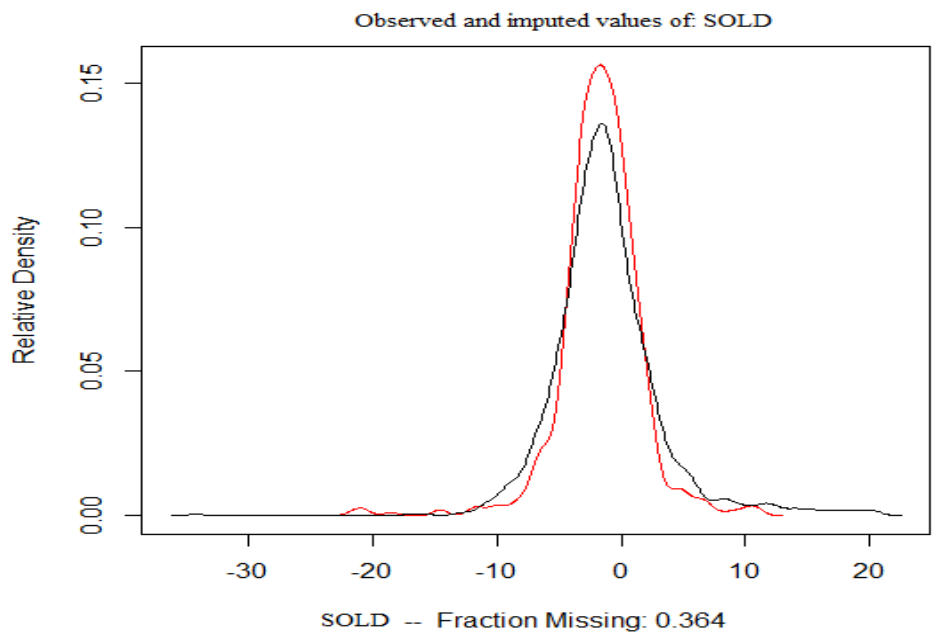


(b)

Figure A.6: Densities Diagnostic VII, $m=100$

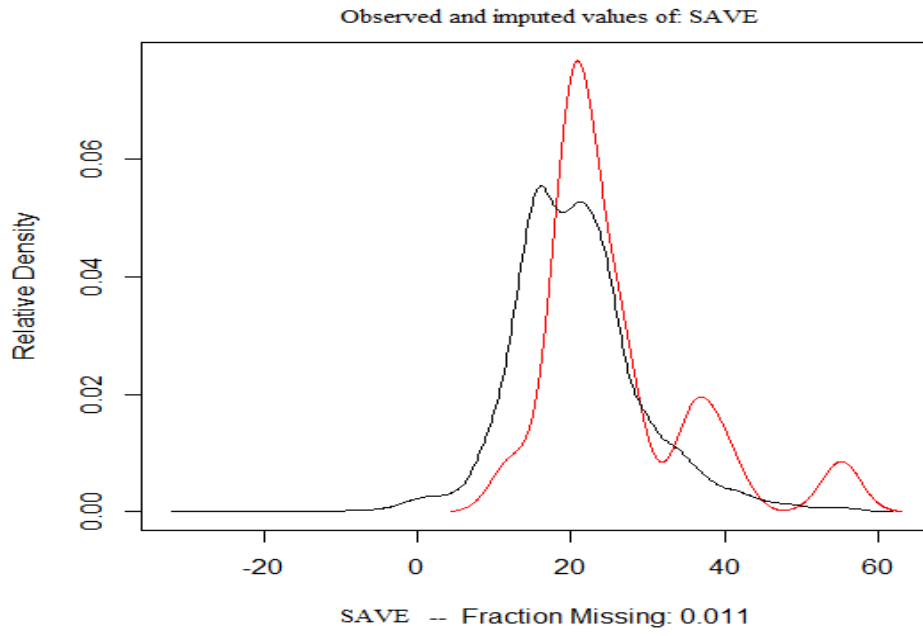


(a)

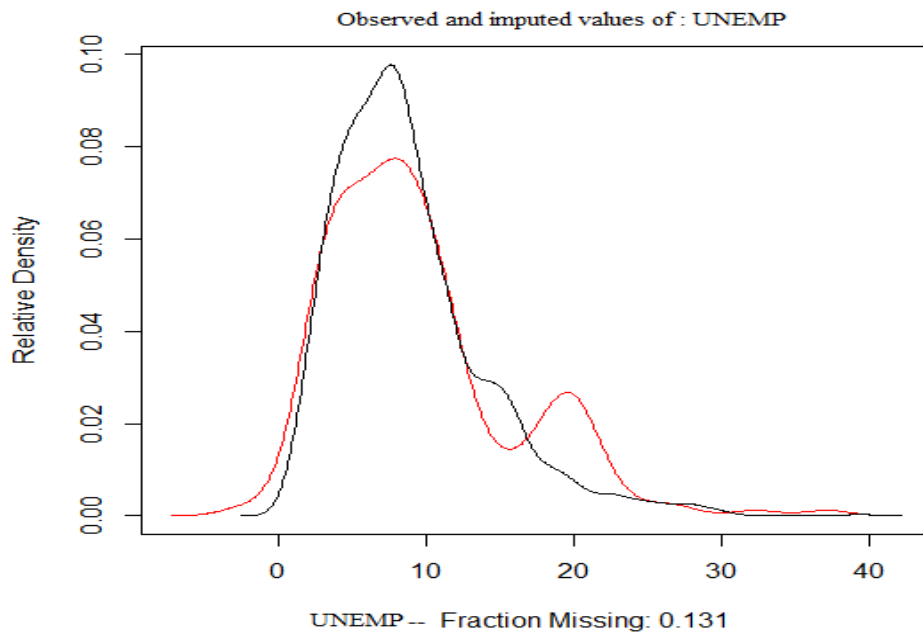


(b)

Figure A.7: Densities Diagnostic VIII, $m=100$

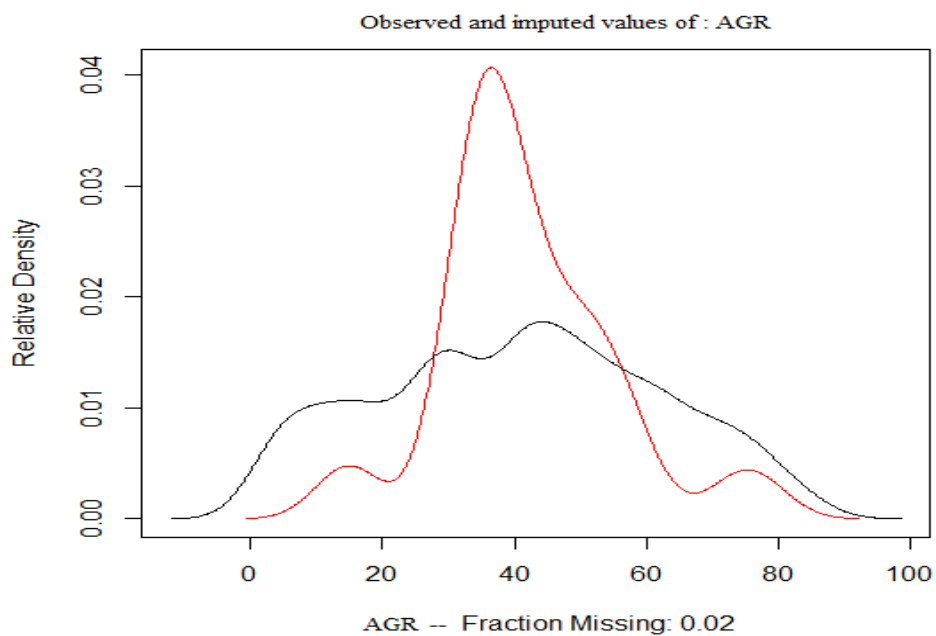


(a)

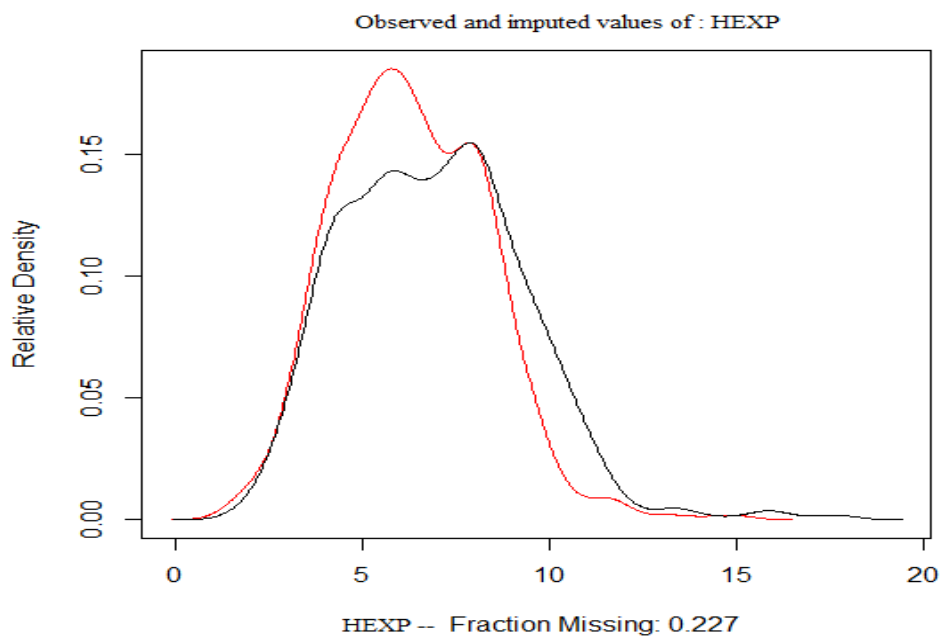


(b)

Figure A.8: Densities Diagnostic IX, $m=100$

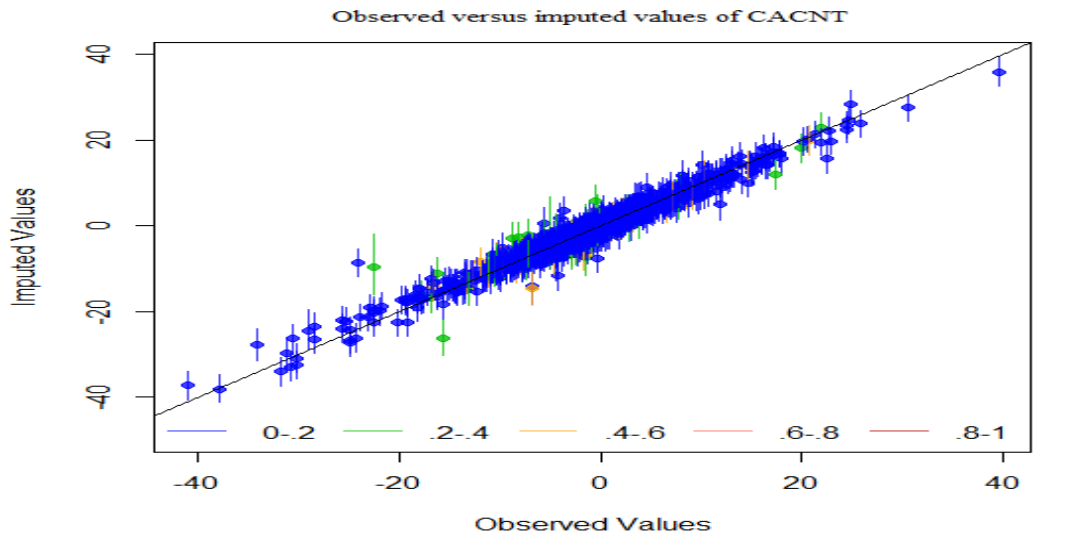


(a)

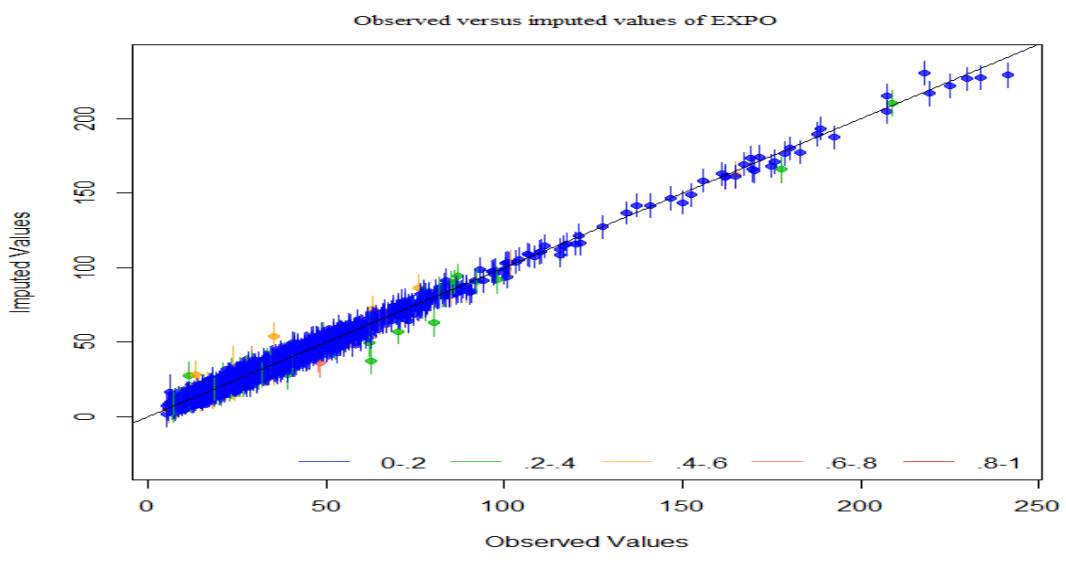


(b)

Figure A.9: Densities Diagnostic X, $m=100$

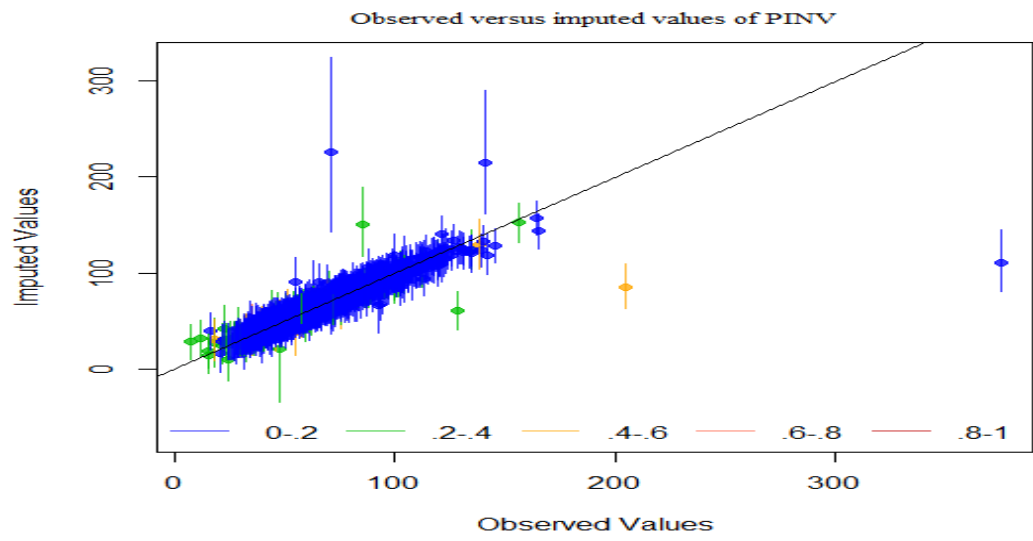


(a)

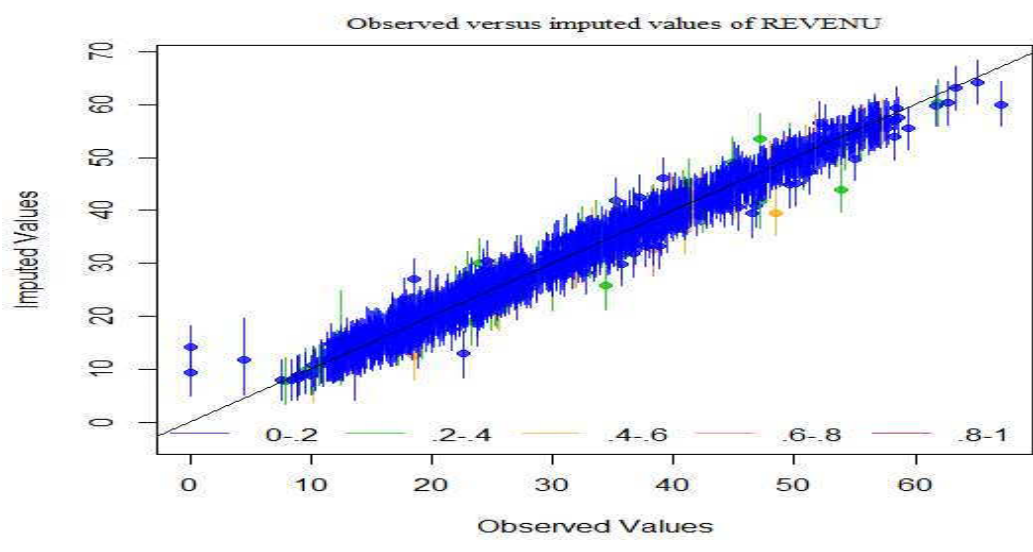


(b)

Figure A.10: Overimputation Diagnostic II, $m=100$

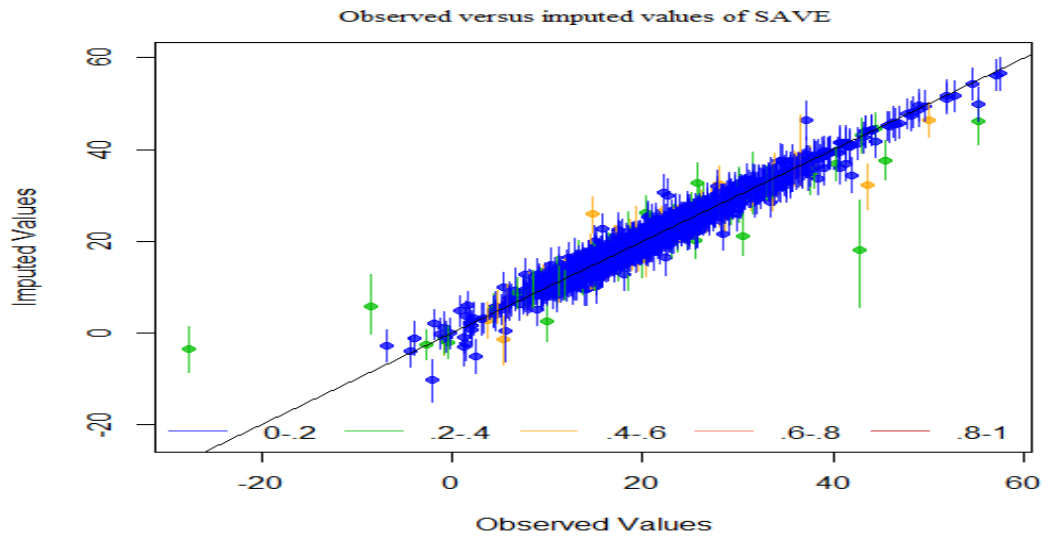


(a)

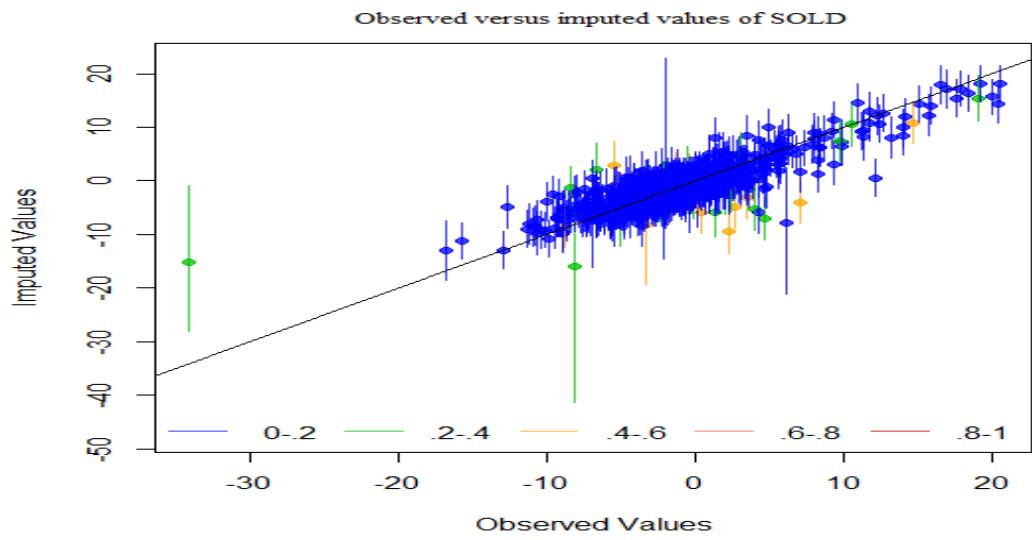


(b)

Figure A.11: Overimputation Diagnostic III, $m=100$



(a)



(b)

Figure A.12: Overimputation Diagnostic IV, $m=100$

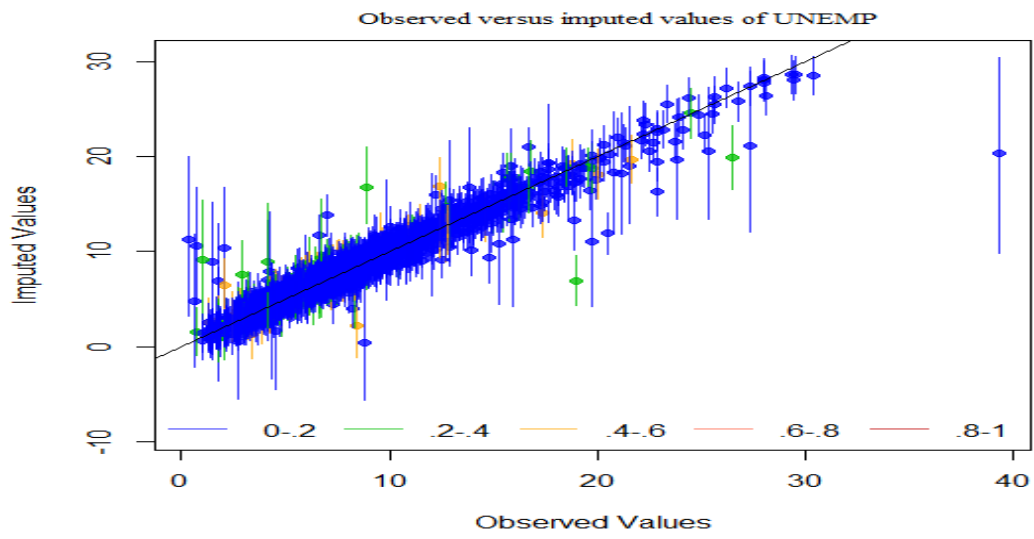
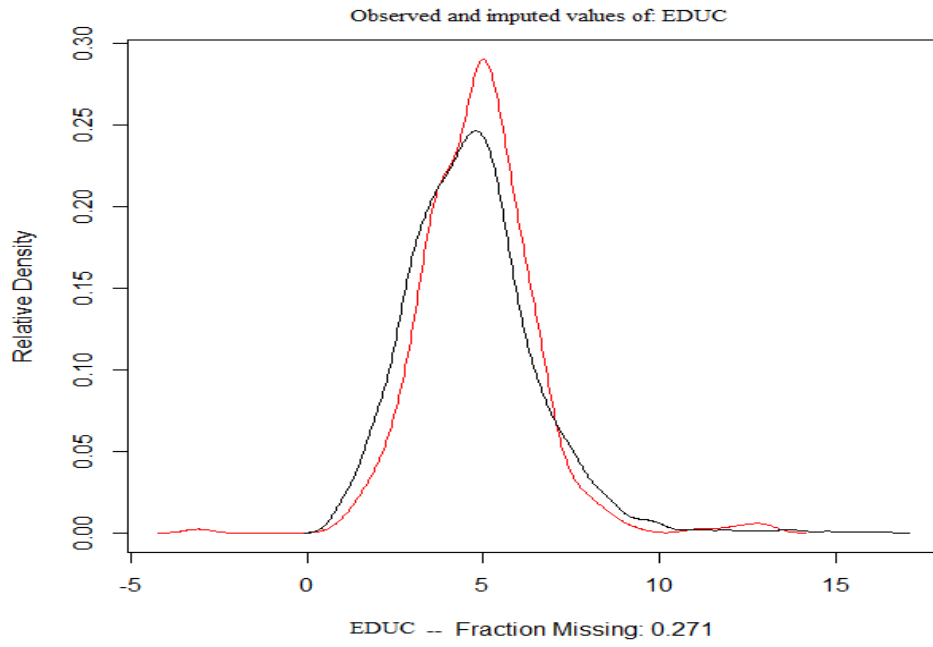
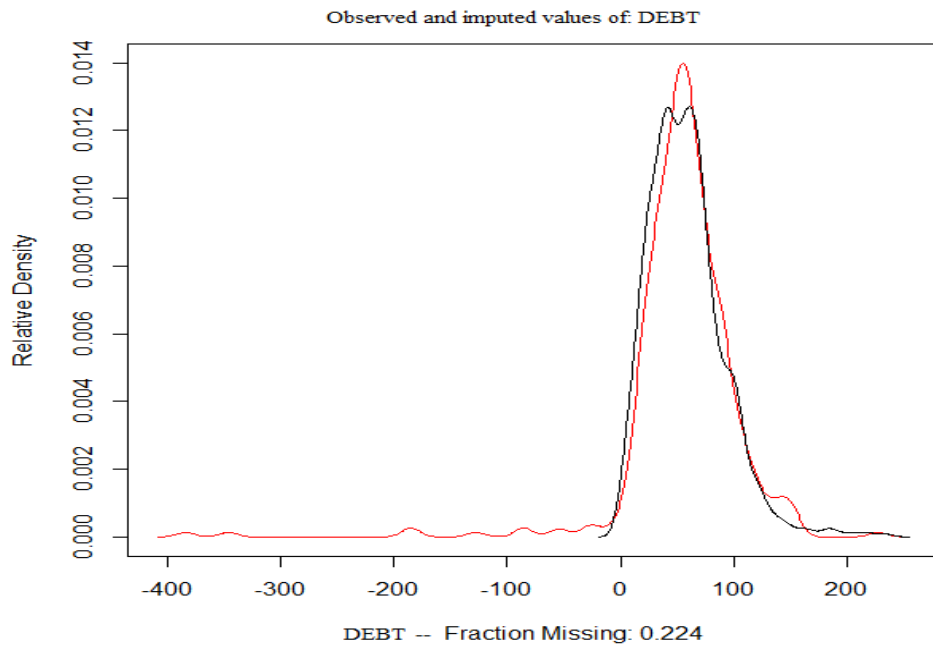


Figure A.13: Overimputation Diagnostic V, $m=100$

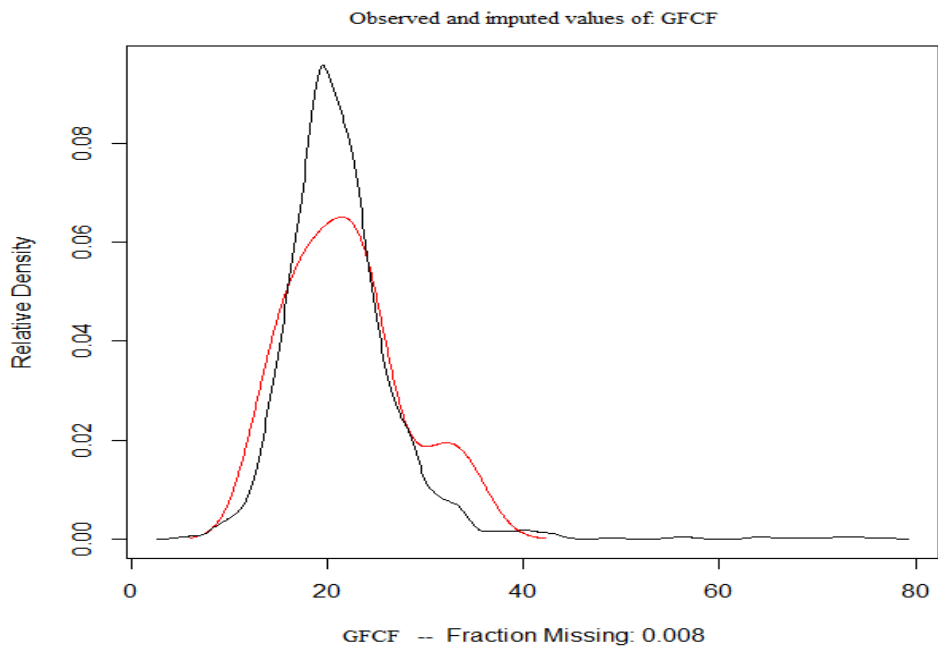


(a)

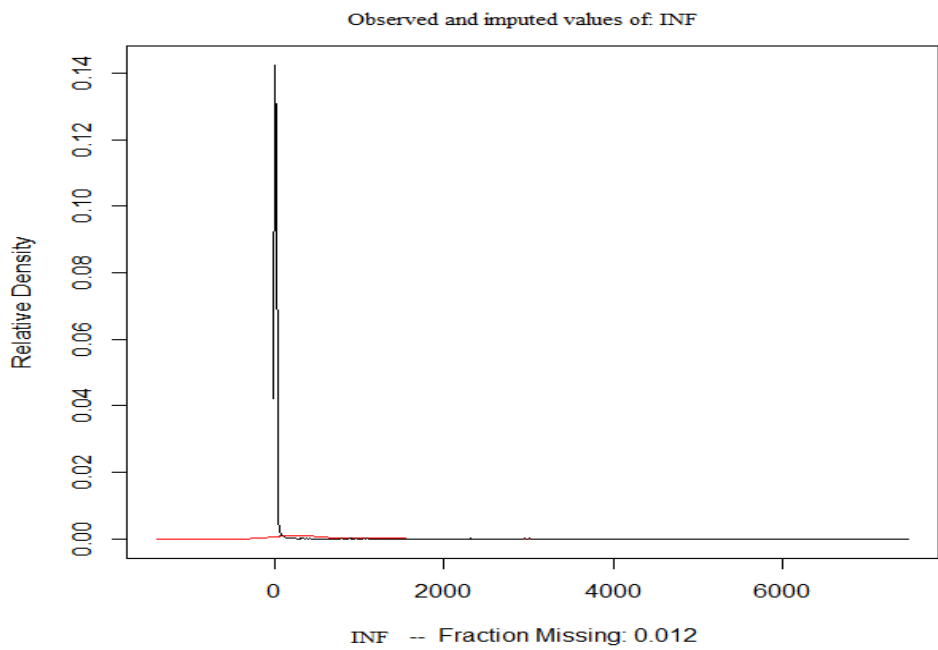


(b)

Figure A.14: Densities Diagnostic I, $m=5$

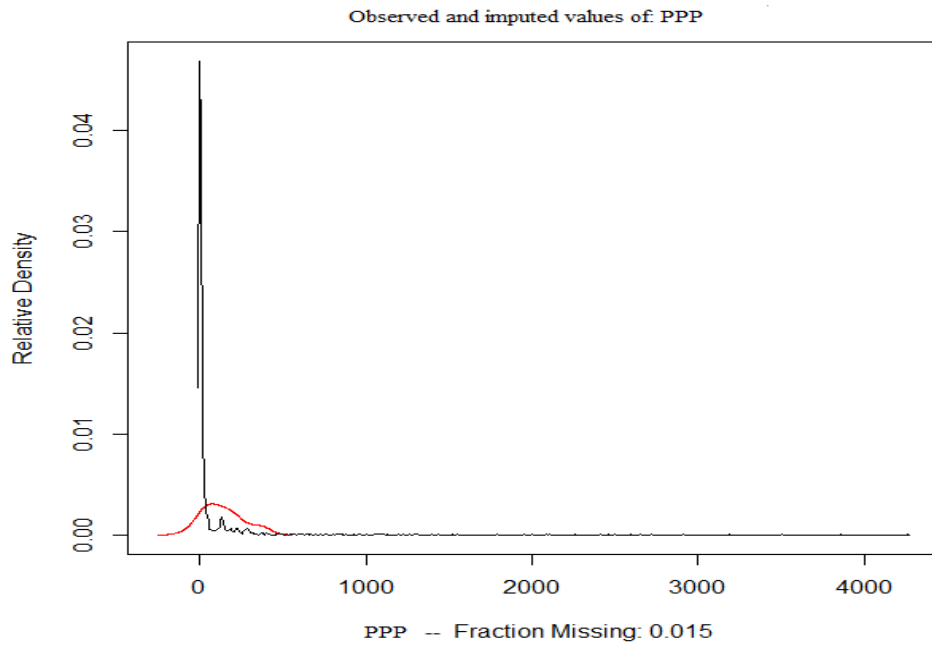


(a)

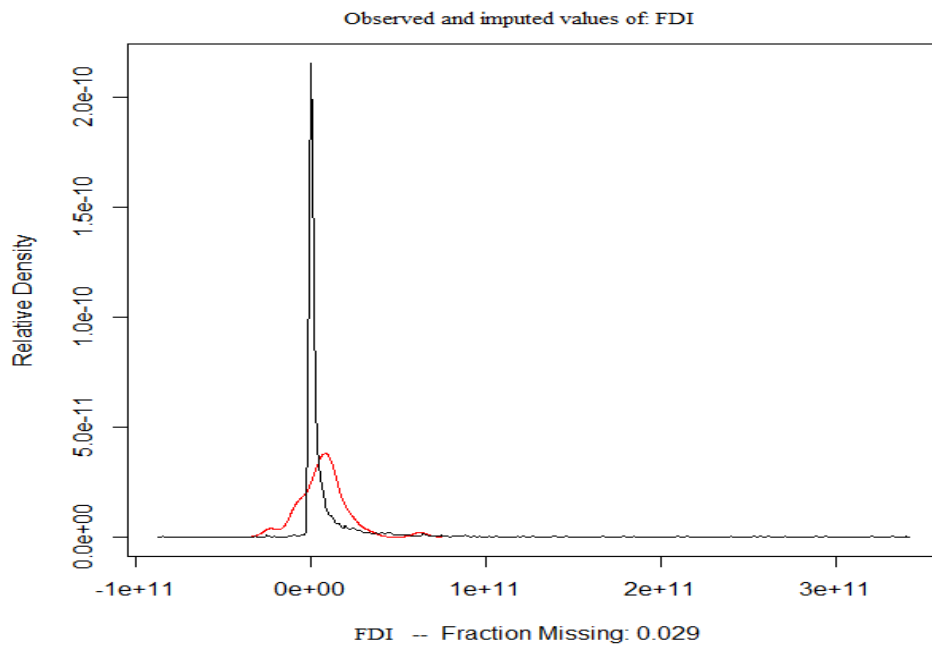


(b)

Figure A.15: Densities Diagnostic II, $m=5$

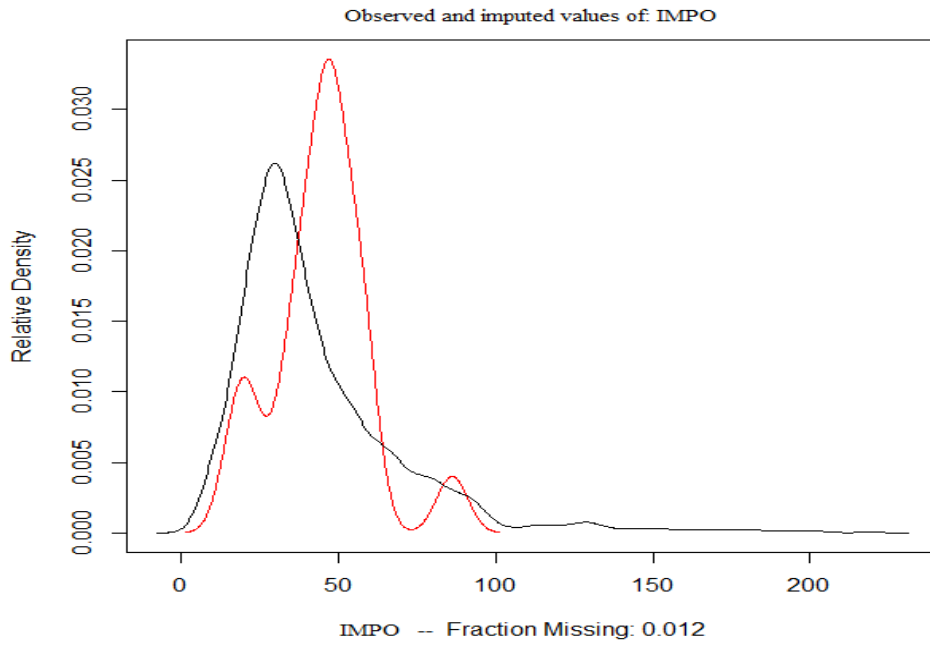


(a)

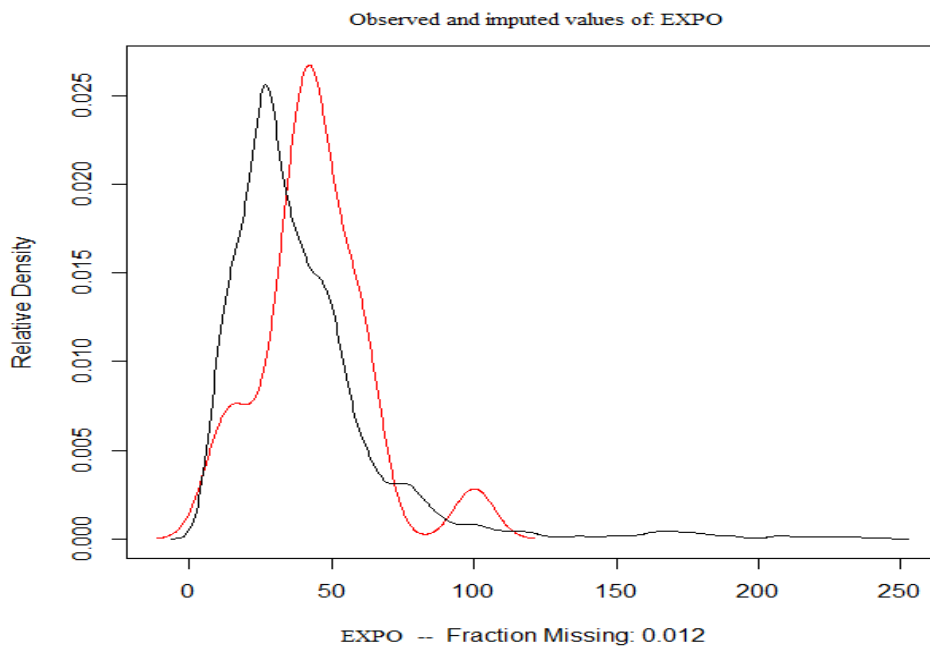


(b)

Figure A.16: Densities Diagnostic III, $m=5$

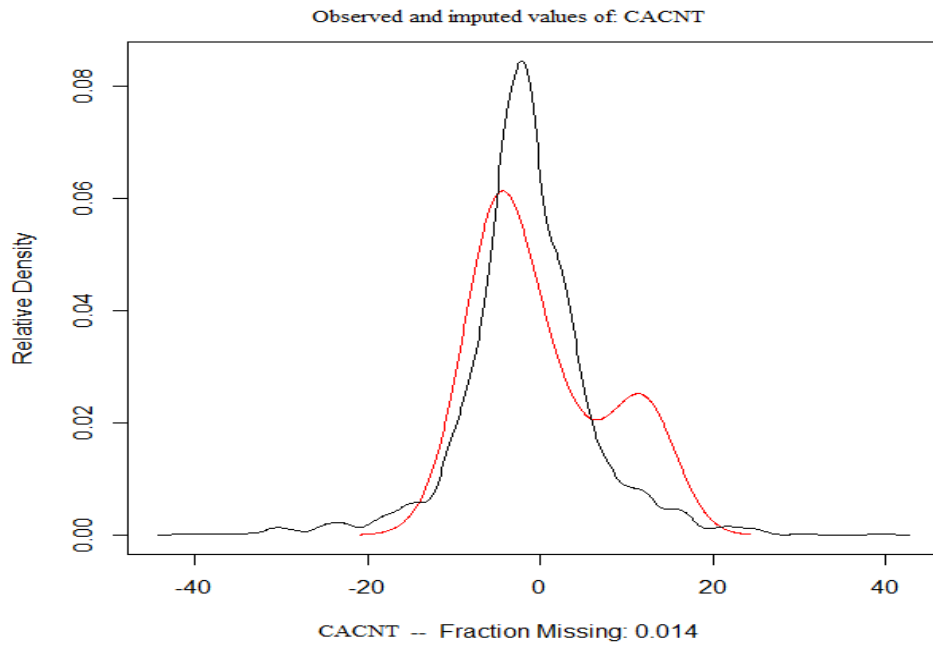


(a)

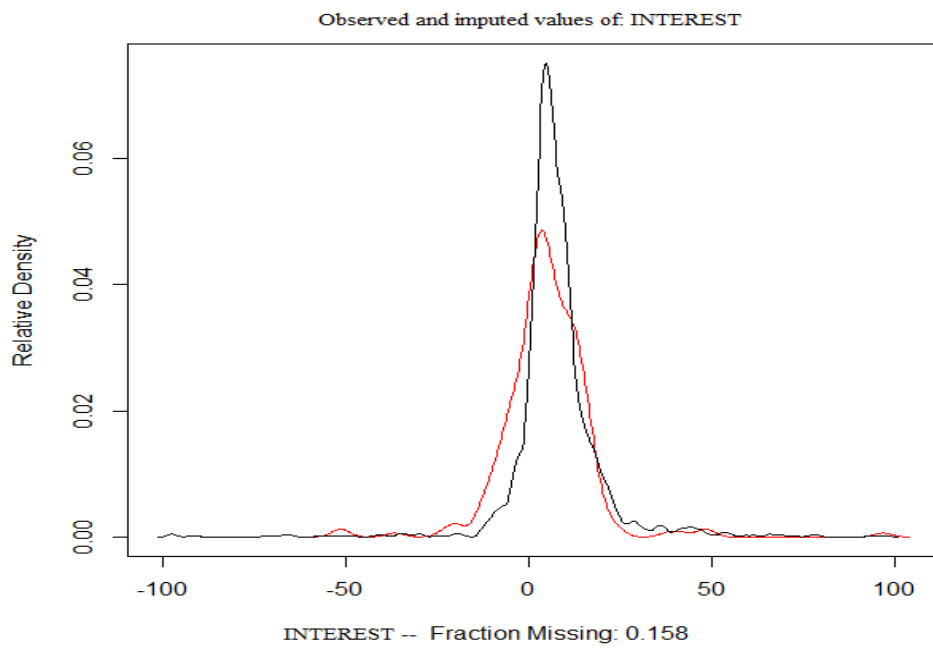


(b)

Figure A.17: Densities Diagnostic IV, $m=5$

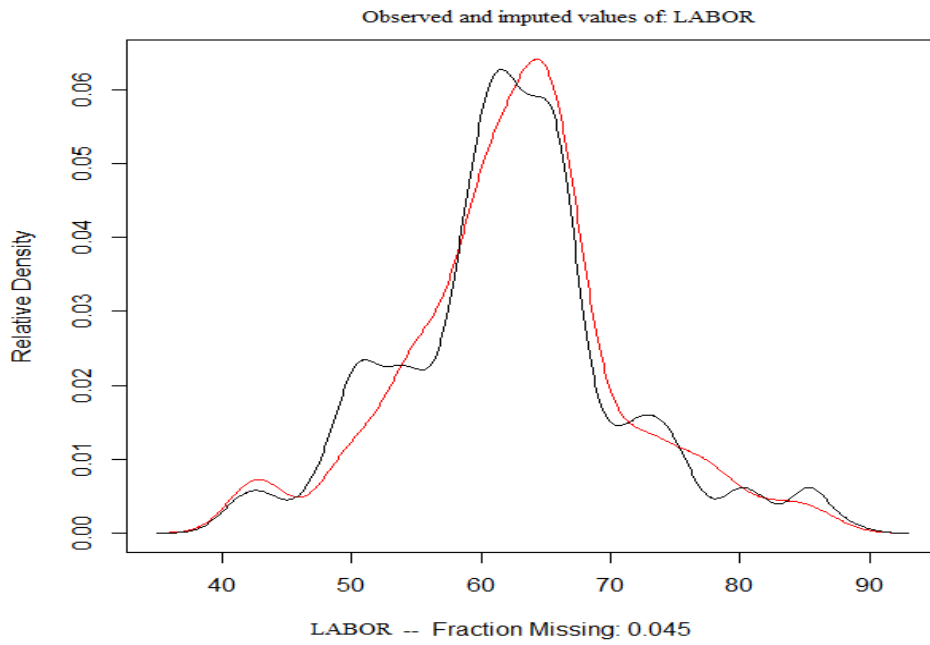


(a)

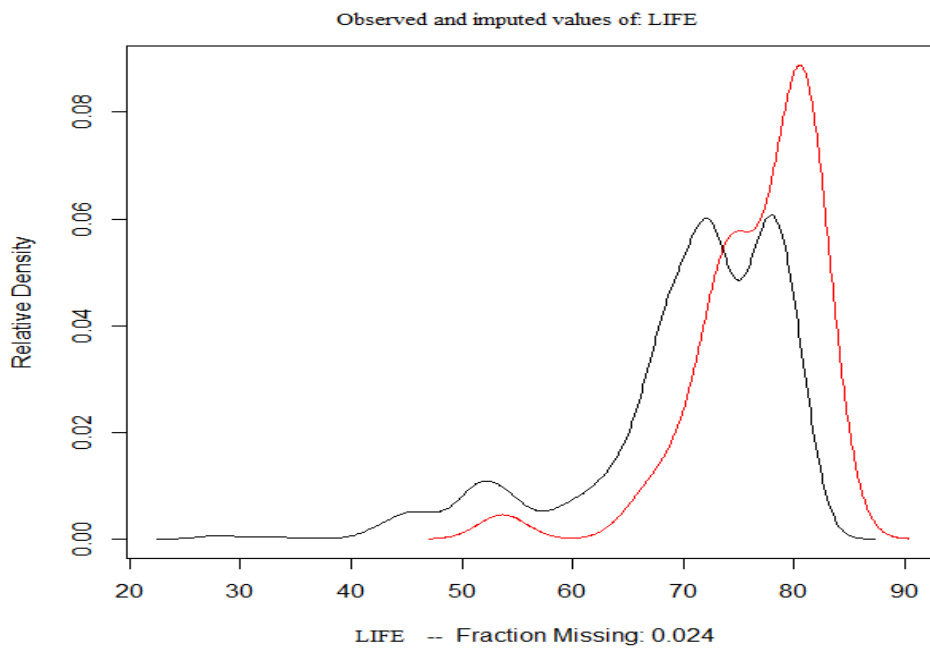


(b)

Figure A.18: Densities Diagnostic $V, m=5$

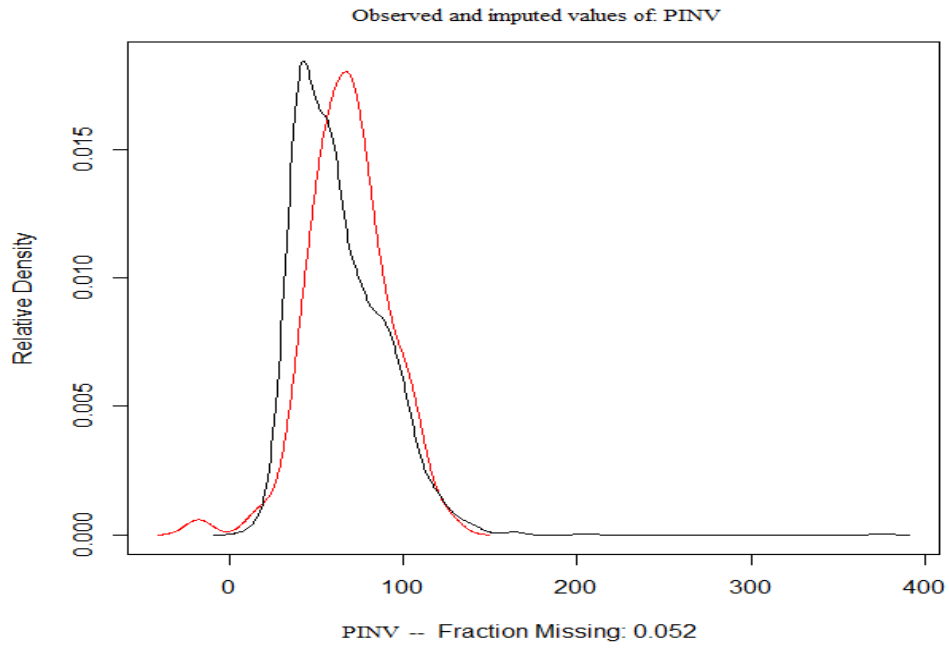


(a)

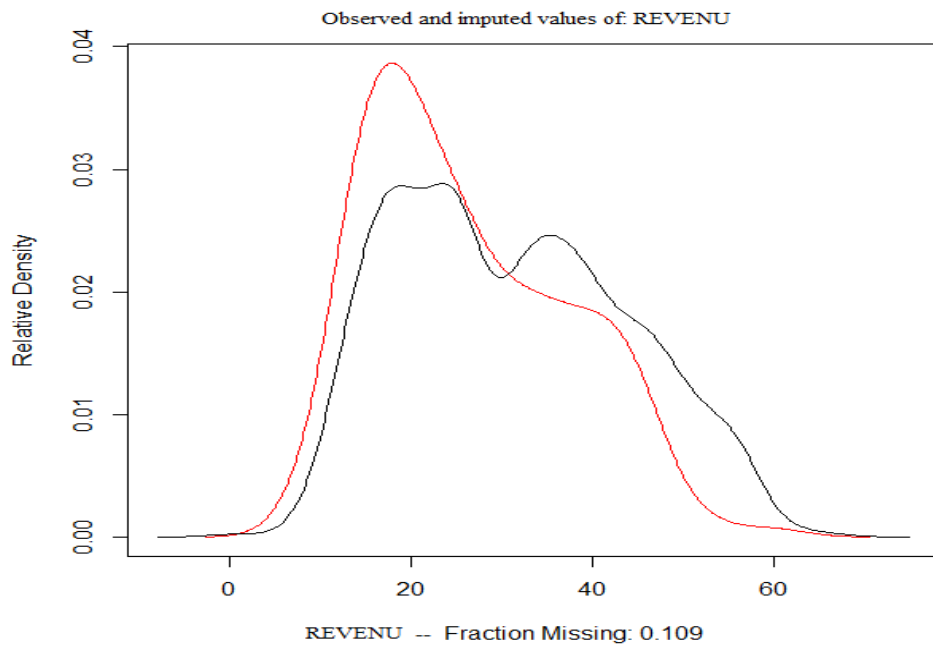


(b)

Figure A.19: Densities Diagnostic VI, $m=5$

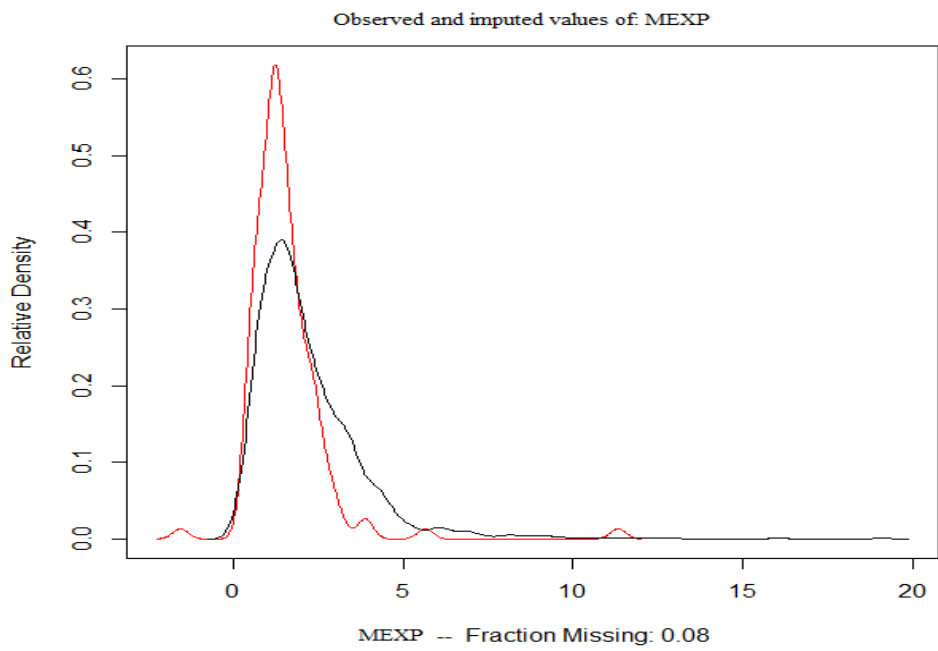


(a)

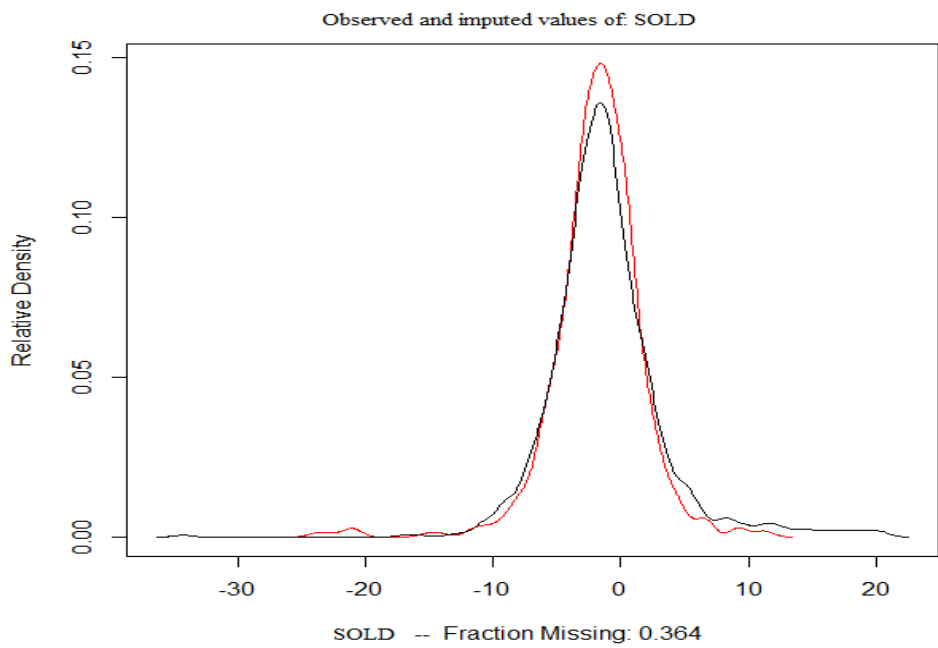


(b)

Figure A.20: Densities Diagnostic VII, $m=5$

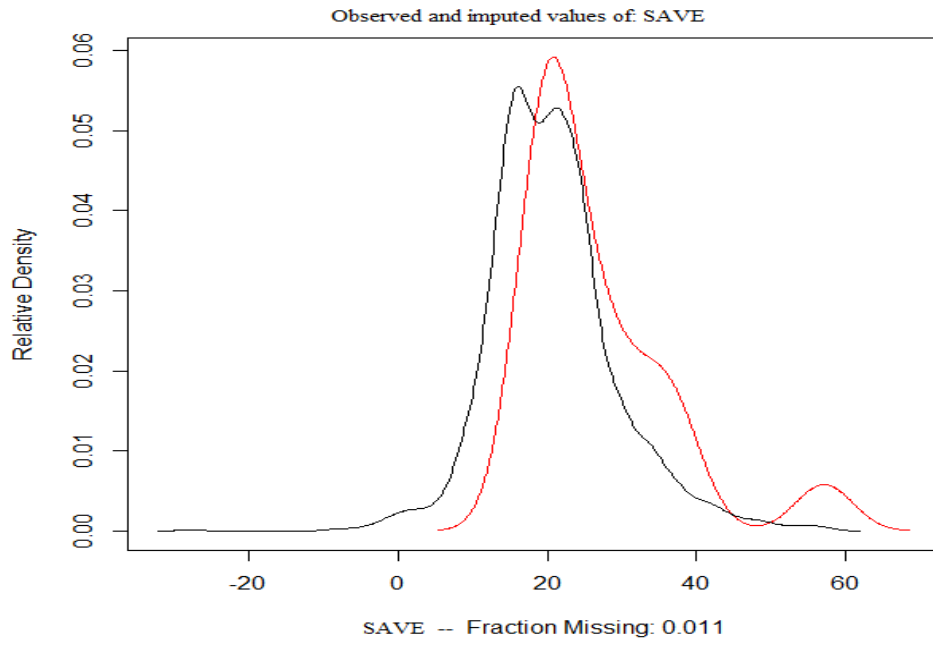


(a)

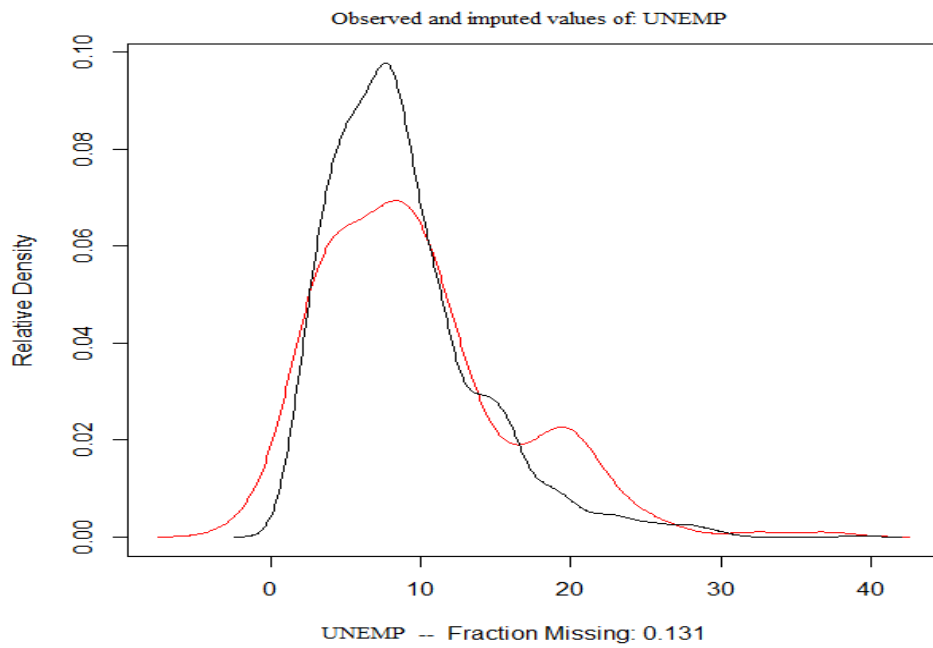


(b)

Figure A.21: Densities Diagnostic VIII, $m=5$

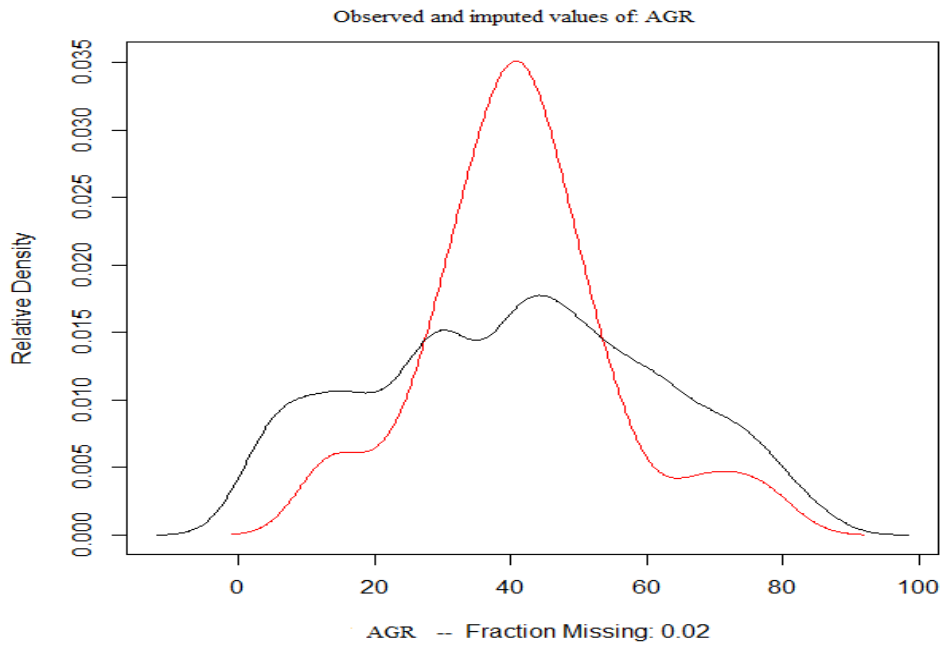


(a)

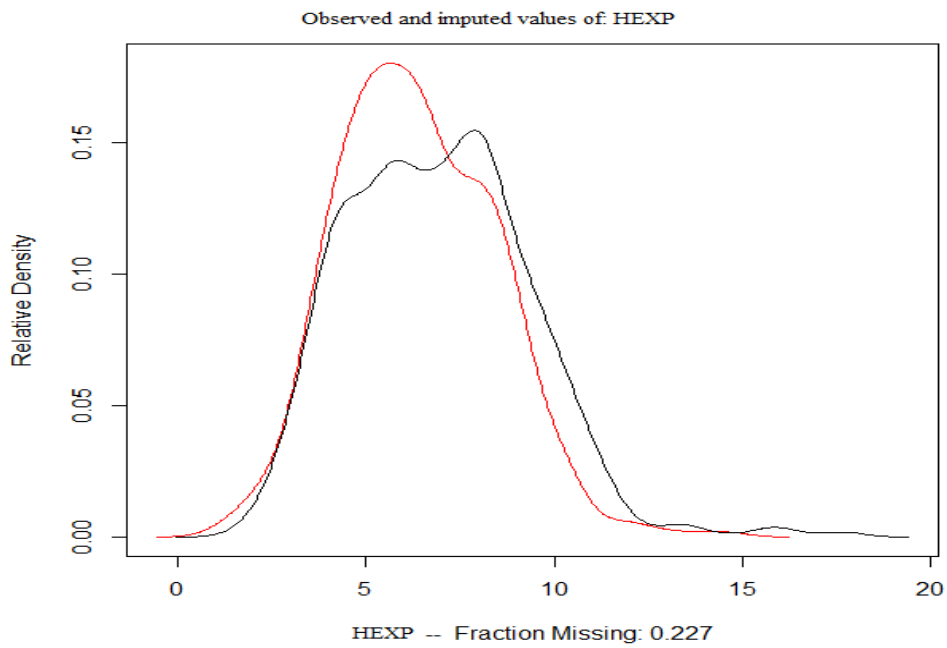


(b)

Figure A.22: Densities Diagnostic IX, $m=5$

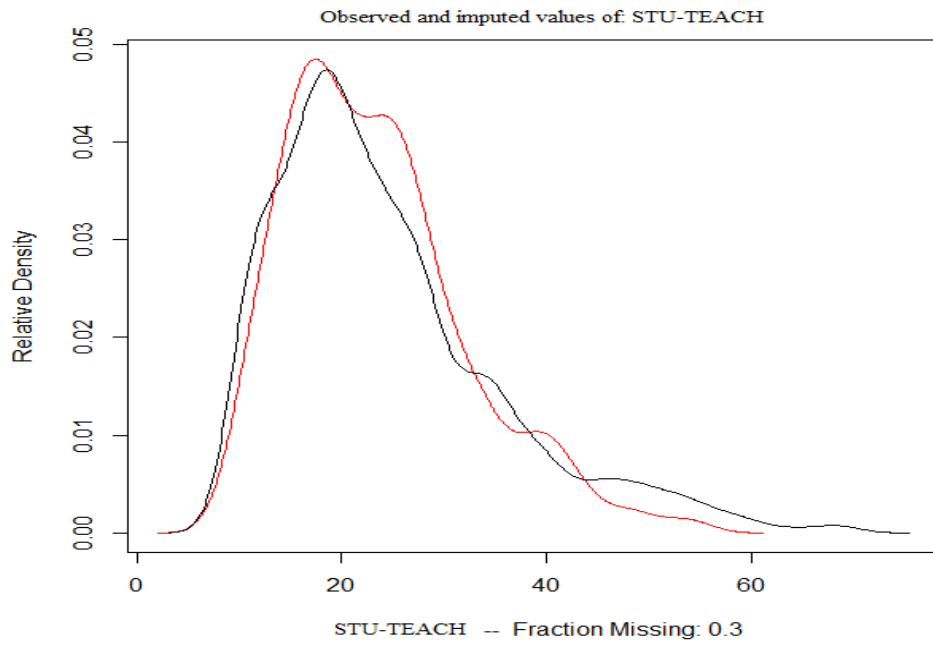


(a)

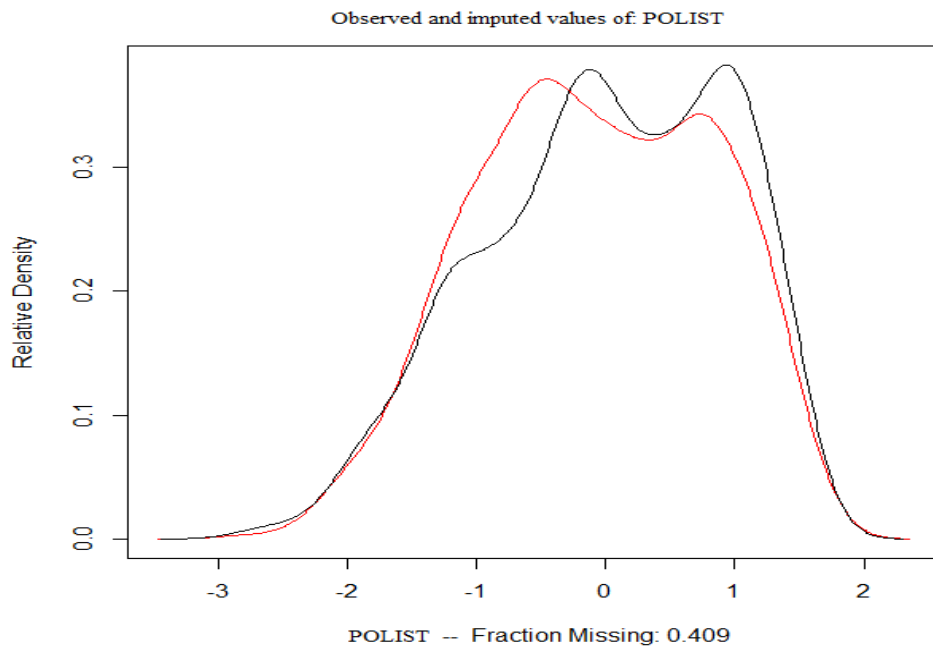


(b)

Figure A.23: Densities Diagnostic X, $m=5$

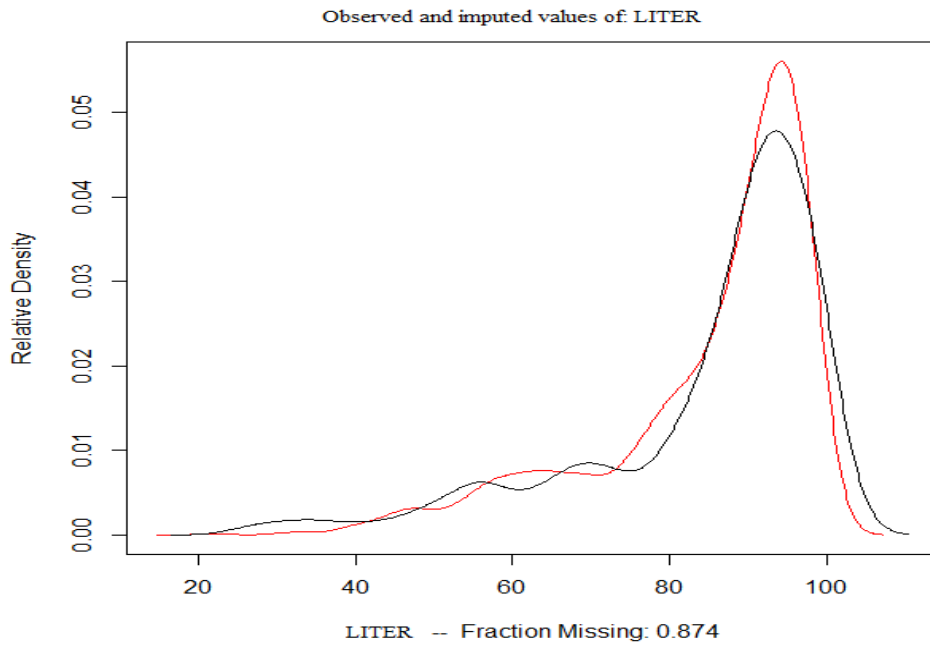


(a)

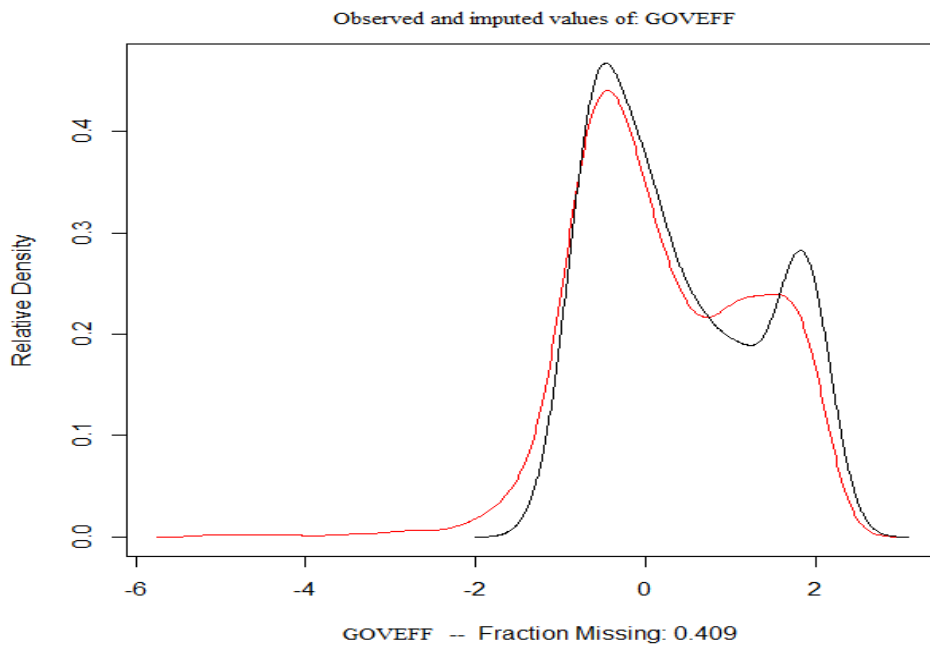


(b)

Figure A.24: Densities Diagnostic XI, $m=5$



(a)



(b)

Figure A.25: Densities Diagnostic XII, $m=5$

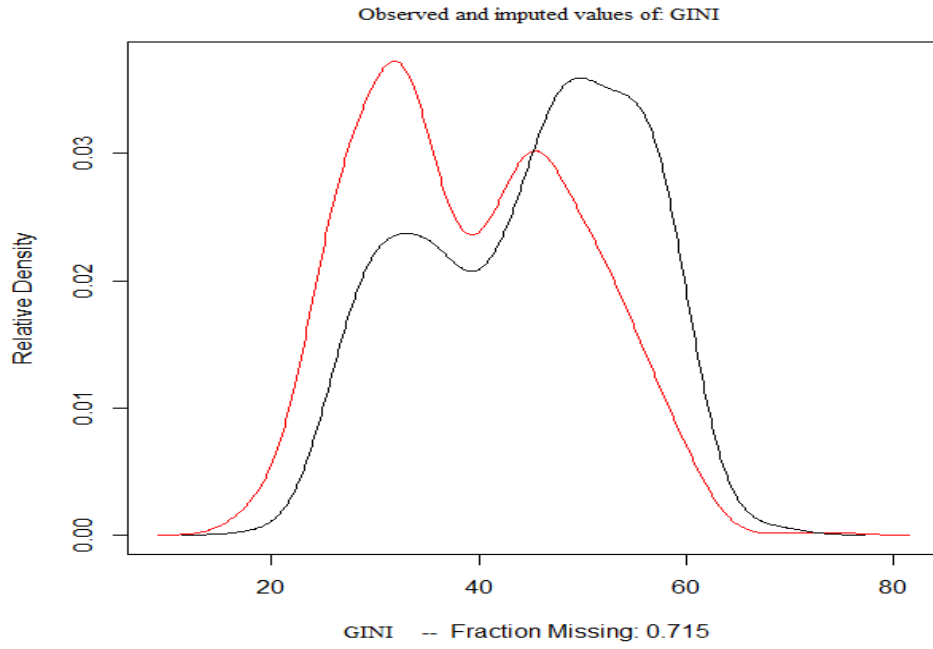


Figure A.26: Densities Diagnostic XIII, $m=5$

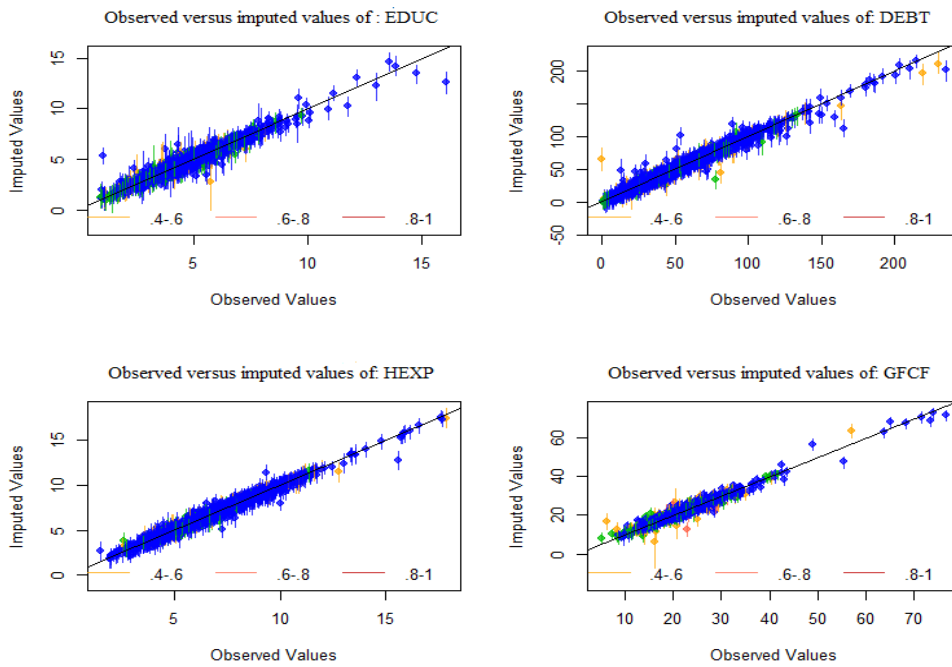


Figure A.27: Overimputation Diagnostic I, $m=5$

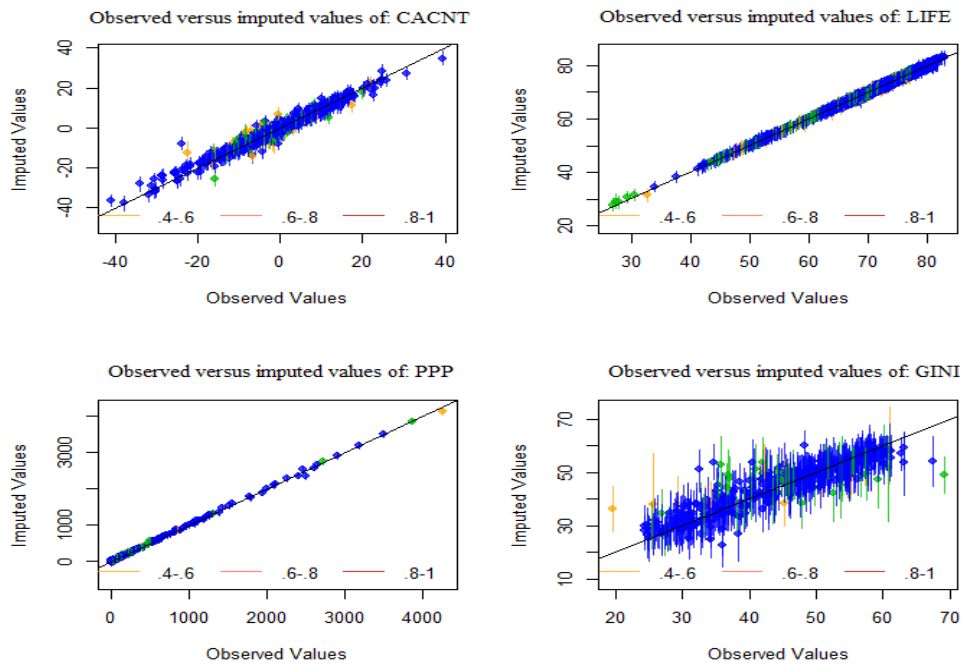


Figure A.28: Overimputation Diagnostic II, $m=5$

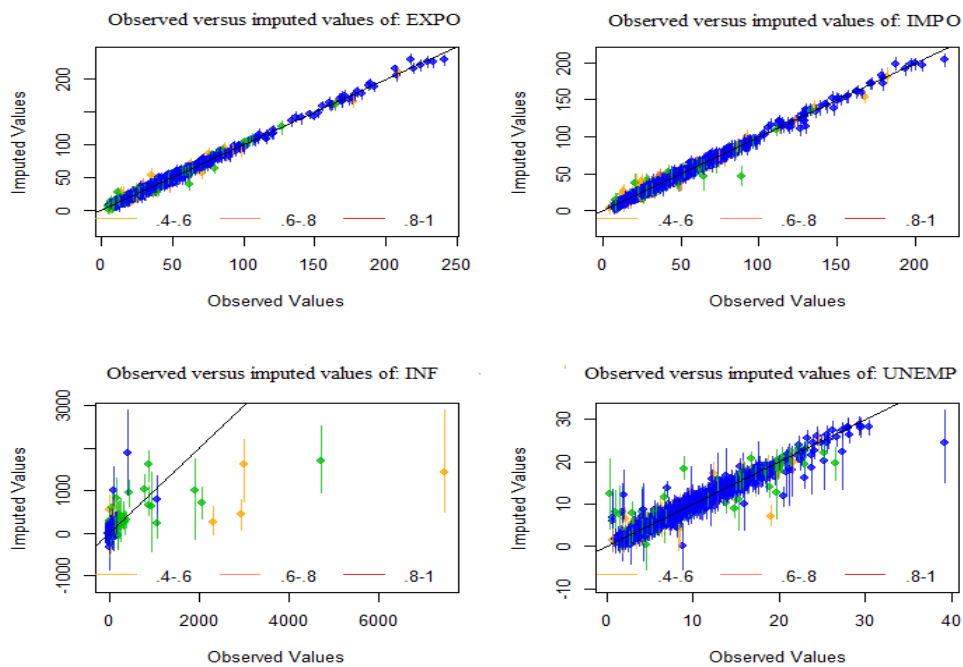


Figure A.29: Overimputation Diagnostic III, $m=5$

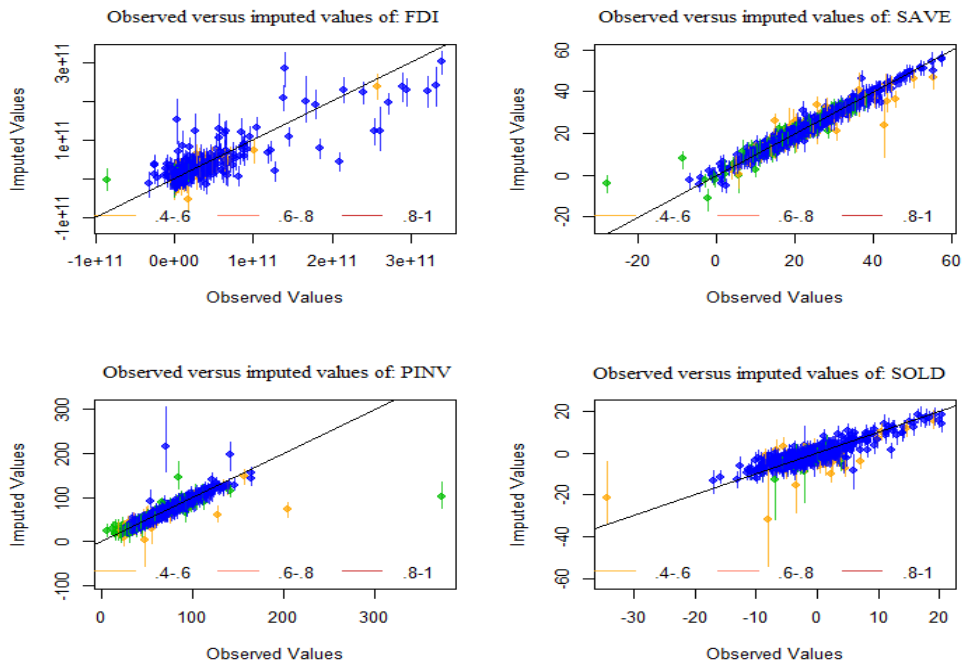


Figure A.30: Overimputation Diagnostic IV, $m=5$

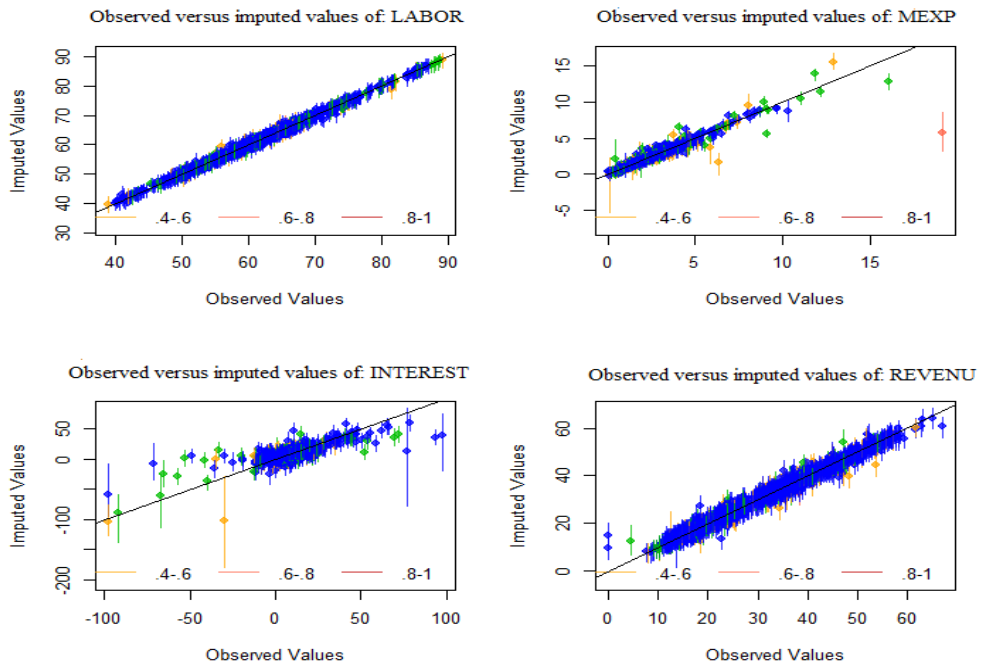


Figure A.31: Overimputation Diagnostic V, $m=5$

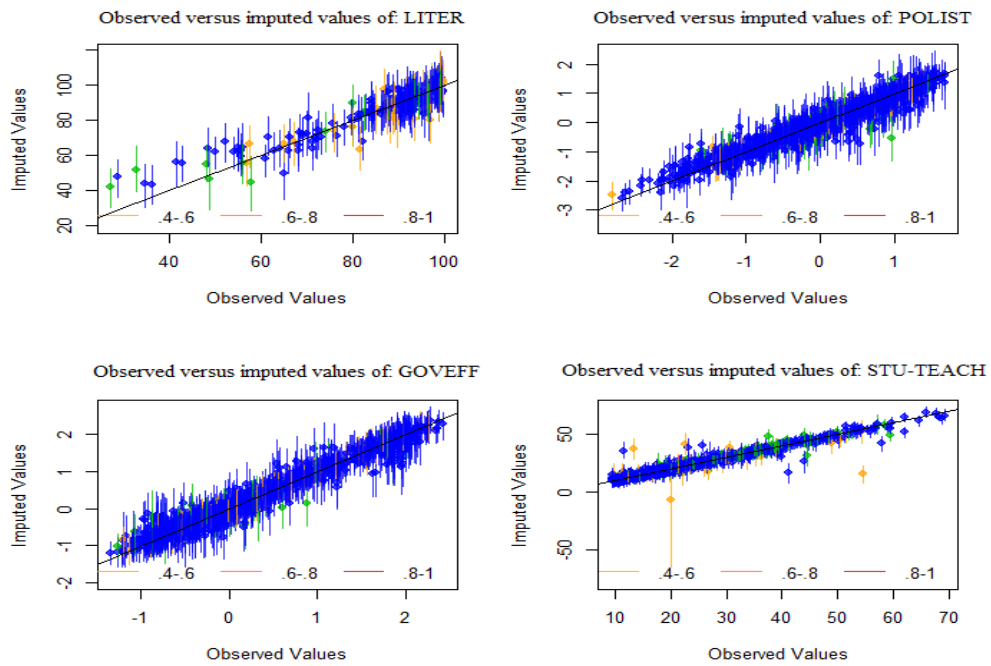


Figure A.32: Overimputation Diagnostic VI, $m=5$

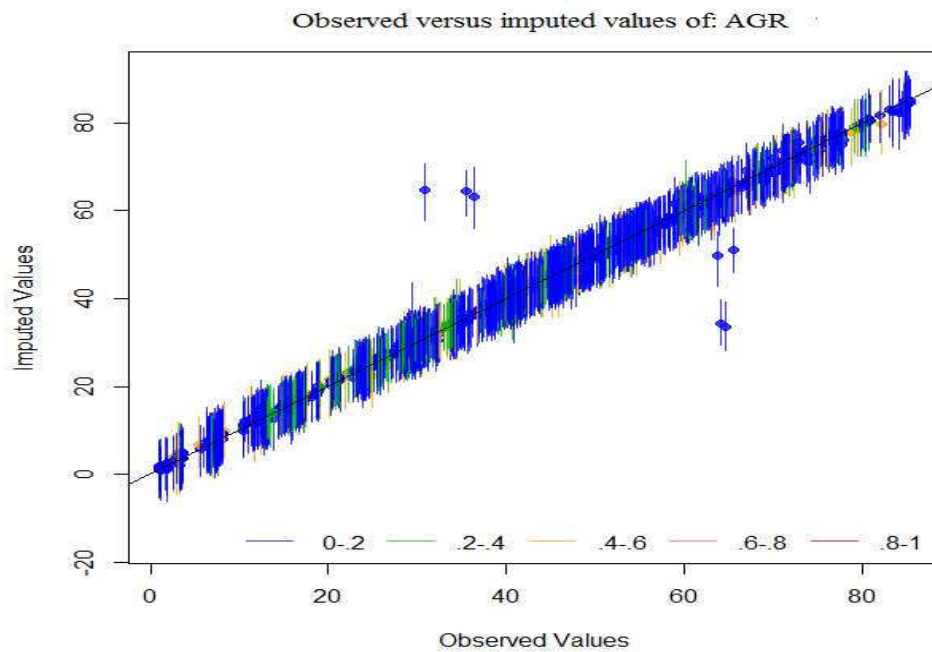


Figure A.33: Overimputation Diagnostic VII, $m=5$

Appendix B

Chapter 2 Appendix

B.1 Chapter 2 Tables

Article title	Publishing journal	Citation number	Authors
Globalization, Brain Drain, and Development	Journal of Economic Literature	199	Frédéric Docquier, Hillel Rapoport
Geography, institutions, and the making of comparative development	Journal of Economic Growth	5	Raphael Auer
Agricultural productivity, structural change, and economic growth in post-reform China	Journal of Development Economics	6	Kang Hua Cao a, Javier A. Birchenall
Services Trade and Policy	Journal of Economic Literature	272	Francois Joseph, Bernard Hoekman
How Deep Are the Roots of Economic Development?	Journal of Economic Literature	85	Enrico Spolaore, Romain Wacziarg
Law and Finance "at the Origin"	Journal of Economic Literature	63	Malmendier Ulrike
Schooling, educational achievement, and the Latin American growth puzzle	Journal of Development Economics	71	Hanushek Eric A., Ludger Woessmann
Trade, innovation and growth	American Economic Review	495	Grossman Gene M., Elhanan Helpman
The Determinants Of Growth	Economic Journal	279	Stern, Nicholas.
Where does microfinance flourish? Microfinance institution performance in macroeconomic context	Journal of Development Economics	141	Ahlin Christian, Jocelyn Lin,
Economic structure, development policy and environmental quality	Ecological Economics	17	Michael Maio He Jie, Hua Wang
An empirical analysis of environmental Kuznets curves with Chinese municipal data			
Energy, property, and the industrial revolution narrative	Ecological Economics	13	Barca Stefania
Growth Accelerations	Journal of Economic Growth	952	Hausmann Ricardo, Lant Pritchett, Dani Rodrik
Borders and Growth	Journal of Economic Growth	99	Spolaore Enrico, Romain Wacziarg
Foreign Aid, the Real Exchange Rate, and Economic Growth in the Aftermath of Civil Wars	World Bank Economic Review	45	Elbadawi Ibrahim A., Linda Kaltani, Klaus Schmidt-Hebbel
Enterprise Restructuring in Transition: A Quantitative Survey	Journal of Economic Literature	1118	Djankov Simeon, Peter Murrell
Inter-provincial migration and inequality during Vietnam's transition	Journal of Development Economics	52	Phan Diep, Ian Coxhead
Does the Profile of Income Inequality Matter for Economic Growth ?	Journal of Economic Growth	171	Voitchovsky Sarah
Institution building and growth in transition economies	Journal of Economic Growth	121	Beck Thorsten, Luc Laeven
A century of economic growth in Latin America	Journal of Development Economics	23	Astorga Pablo
Corporate Governance, Economic Entrenchment, and Growth	Journal of Economic Literature	1002	Morck Randall, Daniel Wolfenzon, Bernard Yeung
Inequality and Economic Growth:	Journal of Economic Literature	1854	Aghion Philippe, Eve Caroli,

Table B.1: Articles' authors & titles I

Article title	Publishing journal	Citation number	Authors
The Perspective of the New Growth Theories			Cecilia Garcia-Penalosa
Environment, human development and economic growth	Ecological Economics	124	Costantini Valeria, Salvatore Monni
Aid effectiveness: the role of the local elite	Journal of Development Economics	60	Angeles Luis, Kyriakos C. Neanidis
Institutions and parameter heterogeneity	Journal of Macroeconomics	20	Minier Jenny
Scale Economies and Industry Agglomeration Externalities	American Economic Review	172	Paul Catherine J. Morrison,
A Dynamic Cost Function Approach			Donald S. Siegel
Measuring and Explaining the Impact of Productive Efficiency on Economic Development	The World Bank Economic Review,	13	Jayasuriya Ruwan, Quentin Wodon
Nonlinearities in cross-country growth regressions: A Bayesian Averaging of Thresholds (BAT) approach	Journal of Macroeconomics	61	Cuaresma Jesus Crespo, Gernot Doppelhofer
Some empirics on economic growth under heterogeneous technology	Journal of Macroeconomics	35	Phillips Peter CB, Donggyu Sul.
The Varieties of Resource Experience: Natural Resource Export Structures and the Political Economy of Economic Growth	The World Bank Economic Review,	548	Isham Jonathan, Michael Woolcock, Lant Pritchett, Gwen Busby
Economic growth and greenhouse gas emissions	Ecological Economics	85	Ansuategi Alberto, Marta Escapa
Oligarchy, democracy, inequality and growth	Journal of development Economics	538	Bourguignon François, Thierry Verdier
Public education expenditures and growth	Journal of development Economics	216	Blankenau William F., Nicole B. Simpson
Growth Empirics and Reality	The World Bank Economic Review,	490	Brock, William A., Steven N Durlauf
Economic growth in Latin America	Journal of development Economics	606	De Gregorio José
Inherited Trust and Growth	The American Economic Review	321	Algan Yann, Pierre Cahuc
The 2008 LAWRENCE R. KLEIN Lecture-Comparative Economic Development: Insights From Unified Growth Theory	International Economic Review	50	Galor Oded
Growth and Agglomeration	International Economic Review	460	Martin Philippe, Gianmarco IP Ottaviano
Credit constraints, equity market liberalization, and growth rate asymmetry	Journal of Development Economics	2	Popov Alexander
Crisis and recovery:	Journal of Macroeconomics	21	Tsangarides Charalambos G
Role of the exchange rate regime in emerging market economies			

Table B.2: Articles' authors & titles II

Article title	Publishing journal	Citation number	Authors
Do all countries grow alike?	Journal of development Economics	44	Bos Jaap WB, Claire Economidou, Michael Koetter, James W. Kolari
Unraveling the fortunes of the fortunate: An Iterative Bayesian Model Averaging (IBMA) approach : A rise by any other name?	Journal of Macroeconomics	39	Eicher Theo S., Chris Papageorgiou, Oliver Roehn
Sensitivity of growth regressions to data source	Journal of Macroeconomics	22	Hanousek Jan, Dana Hajkova, Randall K. Filer
The Lost Decades: Developing Countries' Stagnation in Spite of Policy Reform 1980-1998	Journal of Economic Growth	689	Easterly William
How Robust is the Growth-Openness Connection? Historical Evidence	Journal of Economic Growth	256	Vamvakidis Athanasios
Economic growth in a cross section of countries	The Quarterly Journal of Economics,	12309	Barro Robert J
Debt and growth: New evidence for the euro area	Journal of International Money Finance	86	Baum Anja Cristina Checherita-Westphal, Philipp Rother
Trade, Growth, and Poverty	The Economic Journal	928	Dollar David, Aart Kraay
Human capital formation, public debt and economic growth	Journal of Macroeconomics	28	Greiner Alfred
A Sensitivity Analysis of Cross-Country Growth Regressions	The American economic review	6201	Levine Ross, David Renelt
Manifesto for a growth econometrics	Journal of Econometrics	169	Steven N. Durlauf
Determinants of economic growth will data tell?	American Economic Journal: Macroeconomics	83	Ciccone Antonio, Marek Jarocinski
Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach	American Economic Review	1332	Doppelhofer Gernot, Ronald I. Miller, Xavier Sala-i-Martin
Long-run policy analysis and long-run growth	Journal of Political Economy	3693	Rebelo Sergio T
Education for Growth: Why and For Whom?	Journal of Economic Literature	1981	Krueger Alan B., Mikael Lindahl
Endogenous Growth, Government Debt and Budgetary Regimes	Journal of Macroeconomics	86	Greiner Alfred, Willi Semmler
Endogenous growth, welfare and budgetary regimes	Journal of Macroeconomics	57	Ghosh Sugata, Iannis A. Mourmouras
Economic integration and endogenous growth	The Quarterly Journal of Economics	1876	Rivera-Batiz Luis A., Paul M. Romer
Human Capital Prices, Productivity, and Growth	The American Economic Review,	20	Bowlus Audra J., Chris Robinson
Economic growth and the relative price of capital	Journal of Monetary Economics	252	Jones Charles I
Finance, entrepreneurship, and growth: Theory and evidence	Journal of Monetary economics,	2934	King Robert G., Ross Levine
Growth empirics: A panel data approach	The Quarterly Journal of Economics	3111	Islam Nazrul
Empirics for Economic Growth and Convergence in Asian Economies: A Panel Data Approach	Journal of Economic Development	9	Kim Ji Uk
Africa's Growth Tragedy: Policies And Ethnic Divisions	The Quarterly Journal of Economics	4303	Easterly William, Ross Levine
I just ran two million regressions	The American Economic Review	2000	Xavier Sala-i-Martin

Table B.3: Articles' List

Journal Name	Abreviation
Journal-of-Development-Economics	J-Dev-Eco
Journal-of-Macroeconomics	J-Macr
Journal-of-Economic-Literature	J-Eco-Lit
Journal-of-Economic-Growth	J-Eco-Gr
American-Economic-Review	A-Eco-Rev
World-Bank-Economic-Review	W-B-Eco-Rev
Ecological-Economics	Ecol-Eco
Quarterly-Journal-of-Economics	Q-J-Eco
The Economic-Journal	Eco-J
International-Economic-Review	Int-Eco-Rev
Journal-of-Monetary-Economics	J-Mon-Eco
Journal-of-Econometrics	J-Ect
Journal-of-International-Money-Finance	J-Int-M-F
Journal-of-Political-Economy	J-Pol-Eco
Macroeconomics	Macr

Table B.4: Journals abbreviation

B.2 Chapter 2 Figures

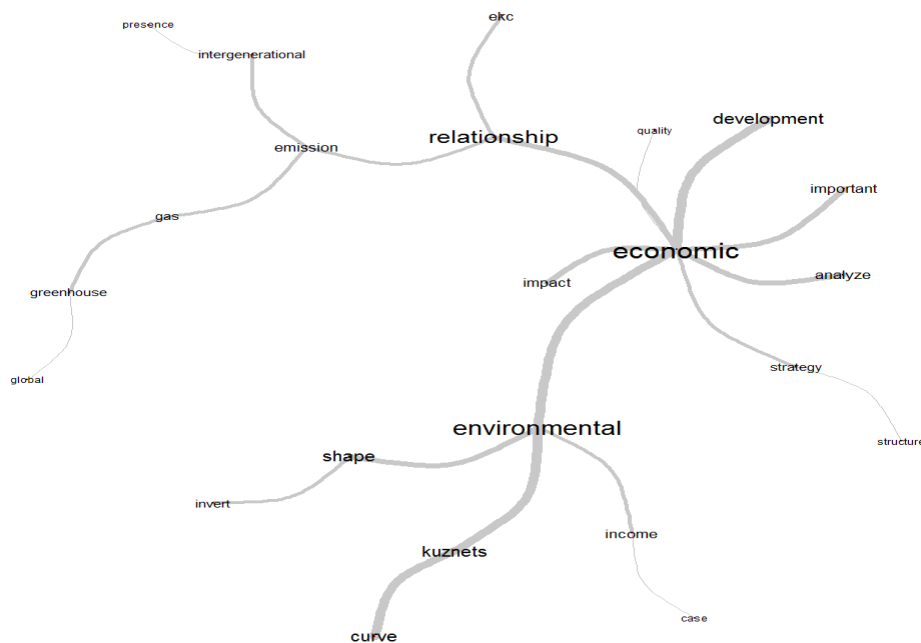


Figure B.1: Class1: Graph of words: class 1

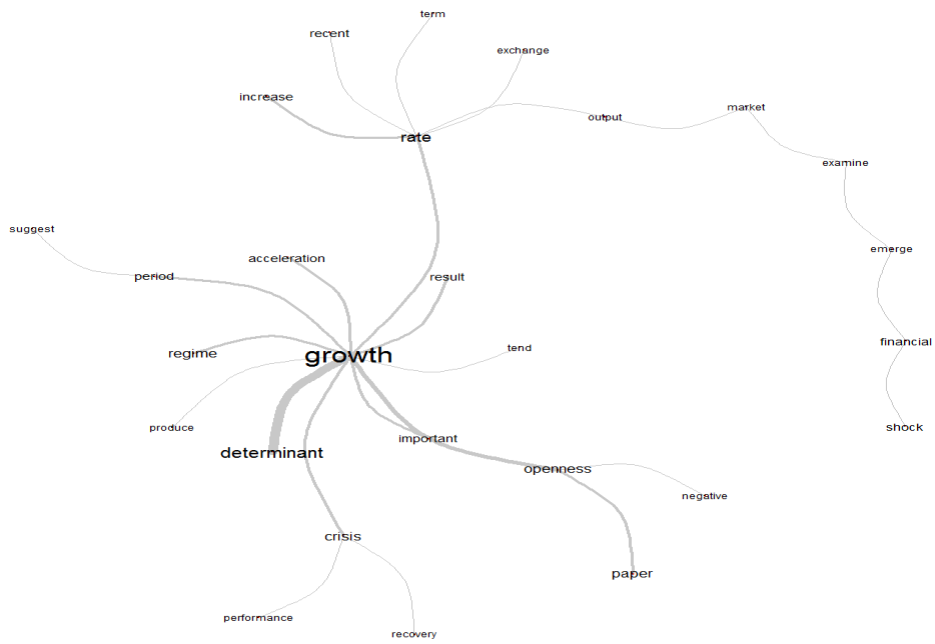


Figure B.2: Class2: Graph of words: class 2

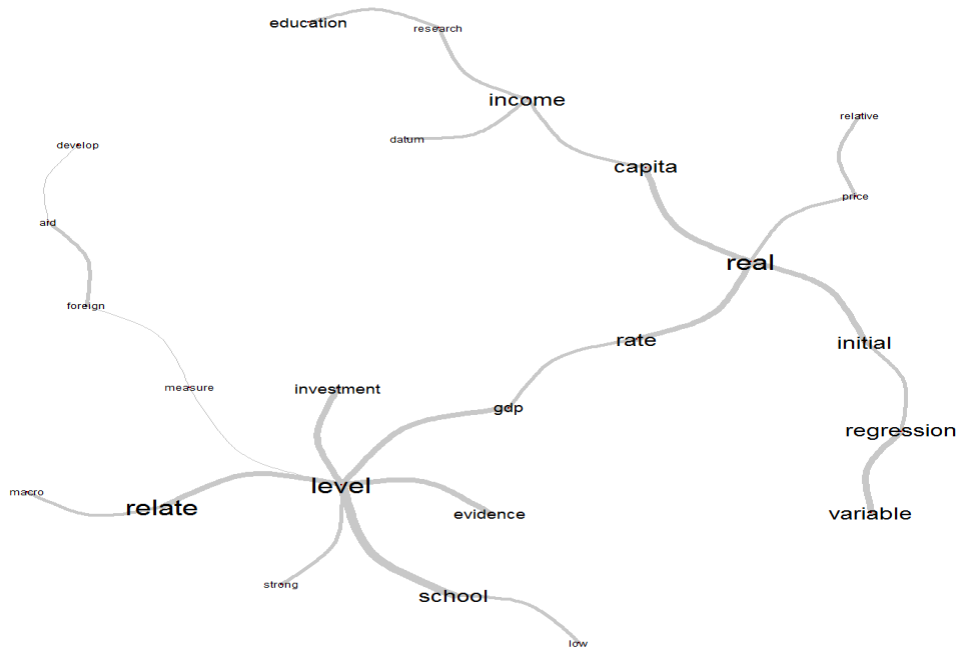


Figure B.3: Class3: Graph of words: class 3

Appendix C

Chapter 3 Appendix

C.1 Chapter 3 Tables

Developing	Developed	In Transition
Algeria, Pakistan	Australia	Albania
Argentina, Paraguay	Austria	Kyrgyzstan
Belize, Paraguay	Belgium	Lesotho
Benin, Peru	27 Bulgaria	Malaysia
Bolivia, Philippines	Canada	Moldova
Botswana, Romania	Cyprus	Russia
Brazil, Rwanda	Denmark	Ukraine
Chile , Singapore	Finland	
Colombia, South Africa	France	
Costa Rica, Thailand	Germany	
Ivory Coast, Trinidad and Tobago	Greece	
Dominican Rep, Uganda	Hungary	
Ecuador, Uruguay	Iceland	
Egypt, Venezuela	Israel	
El Salvador, Zambia	Italy	
Guatemala	Japan	
Honduras	Latvia	
India	Luxembourg	
Indonesia	Malta	
Jamaica	Netherlands	
Jordan	Norway	
Kenya	Portugal	
Mauritius	Slovak Rep	
Mexico	Spain	
Morocco	Sweden	
Nepal	UK	
Nicaragua	US	
42	27	7

Countries in bold are members of OECD.

Data source: World Economic Situation and Prospects WESP, 2012.

Table C.1: Country classification

C.1. Chapter 3 Tables

Indicator code	Unity	Source	Source note
GDP	annual %	The World bank/WDI	Annual percentage growth rate of GDP at market prices based on constant local currency. Aggregates are based on constant 2005 U.S. dollars. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products.
EDUC SCH	% of GDP Year	The World bank/WDI Barro et Lee (2000)	Public spending on education, total Average years of schooling for persons over 25 years
DEBT	% of GDP	WEO database 2012/IMF/	General government gross debt
GFCF	% of GDP	The World bank/WDI	Gross Fixed Capital Formation % of GDP, includes land improvements (fences, ditches, drains , and so on); plant, machinery, and equipment purchases; and the construction of roads, railways, and the like, including schools, offices, hospitals, private residential dwellings, and commercial and industrial buildings.
HEXP	% of GDP	The World bank/WDI	Total health expenditure % of GDP
MEXP	% of GDP	The World bank/WDI	Military expenditure
POP	annual %	The World bank/WDI	Population growth (annual %) is the exponential rate of growth of midyear population from year $t - 1$ to t , expressed as a percentage
RPOP	% of total population	World bank open data	market prices based on constant local currency The part of rural population in percentage of all population
REVENU	% of GDP	WEO database 2012/IMF/	General government revenue
EXPO	% of GDP	The World bank/WDI	Exports of goods and services
IMPO	% of GDP	The World bank/WDI	Imports of goods and services
INTEREST	%	WEO database 2012/IMF/	Real interest rate
LABOR	% of total population ages 15+	World bank open data	Labor force participation rate is the proportion of the population ages 15 older that is economically active: all people who supply labor for the production of goods and services during a specified period
INF	Consumer prices, annual %	The World bank/WDI	Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. The Laspeyres formula is generally used.
LIFE	Years	The World bank/WDI	Life expectancy at birth, total
PINV	PPP over investment / exchange rate	Penn world table7.1	Price Level of Investment
SOLD	% of GDP	Total expenditure-Total income	Primary deficit
SAVE	% of GDP	WEO database 2012/IMF/	Gross national savings

Table C.2: Variables Description I

Appendix C. Chapter 3 Appendix

Indicator code	Unity	Source	Source note
UNEMP	% of total labor force	WEO database 2012/IMF/	Unemployment rate, As defined by the International Labour Organization, unemployed workers are those who are currently not working but are willing and able to work for pay, currently available to work, and have actively searched for work.
CACNT	% of GDP	WEO database 2012/IMF/	Current account balance, Current account is all transactions other than those in financial and capital items. The major classifications are goods and services, income and current transfers.
FDI	BoP, current US\$	The World bank/WDI	Foreign direct investment, net inflows
PPP	National currency per current international \$	WEO database 2012/IMF/	Implied PPP conversion rate
LAND	sq.km	The World bank/WDI	Land area is a country's total area
AGR	% of total land area	The World bank/WDI	Agricultural Land
POP90	population in 1990	The World bank/WDI	the population census in 1990
DEBT90	Debt in 1990 as%GDP	The World bank/WDI	Initial level of Debt
GDP90	current US \$	The World bank/WDI	The initial level of GDP in 1990 excluding area under inland water bodies, national claims to continental shelf and exclusive economic zones
STU.TEACH	Pupil-teacher ratio, primary	The World bank/WDI	Pupil-teacher ratio in primary, is the number of pupils enrolled in primary school divided by the number of primary school teachers.
GOVEFF	Ranges from -2.5to2.5	WGI	Government effectiveness captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.
GINI	From 0 – to – 100	The World bank/WDI	GINI index measures the extent to which the distribution of income or consumption expenditure among individuals or households within an economy deviates from a perfectly equal distribution.
LITER	% of people ages 15+	The World bank/WDI	Adult (15+) literacy rate (%). Total is the percentage of the population age 15 and above who can, with understanding, read and write a short, simple statement on their everyday life. Generally, literacy also encompasses numeracy, the ability to make simple arithmetic calculations.
POLIST	Ranges from -2.5to2.5	WGI	Policy Stability and absence of violence, reflects perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including politically-motivated violence and terrorism.
XRAT	National currency per dollar	Penn world table 7.1	Exchange Rate to US\$
INV	% of GDP	Penn world table 7.1	Investment Share of PPP Converted GDP Per Capita at 2005 constant prices
CONS	% of GDP	Penn world table 7.1	Consumption Share of PPP Converted GDP Per Capita at 2005 constant prices
OPEN	% of GDP	Penn world table 7.1	Openness at 2005 constant prices

Table C.3: Variables Description II

Quantiles	τ	0.25	0.5	0.75	0.95
Model A	INTERCEPT	2.892 ^(*)	3.542 ^(***)	4.870 ^(***)	10.052 ^(***)
	EDUC	-0.315 ^(**)	-0.239 ^(**)	-0.197 ^(**)	-0.402 ^(***)
	SCH	-0.143 ^(*)	-0.141 ^(**)	-0.157 ^(.)	-0.148
	DEBT	-0.012	-0.016 ^(*)	-0.014 ^(.)	-0.005
	GFCF	0.118 ^(*)	0.144 ^(**)	0.141 ^(**)	0.081 ^(.)

Significance code: ./*/**/***, at 1%/5%/10%/0.1%, respectively.

Table C.4: Penalized quantile regression with fixed effect estimates, Model A

Quantiles	τ	0.25	0.5	0.75	0.95
Interest Variables	INTERCEPT	9.987 ^(**)	9.362 ^(**)	10.06 ^(**)	19.333 ^(*)
	EDUC	-0.413 ^(**)	-0.247 ^(*)	-0.24 ^(**)	-0.315 ^(.)
	SCH	0.148 ^(.)	0.184 ^(*)	0.177 ^(*)	0.47 ^(*)
	DEBT	-0.014 ^(***)	-0.015 ^(***)	-0.014 ^(**)	-0.007
	GFCF	-0.007	-0.031	-0.045	-0.076
Control Variables	CACNT	-0.078 ^(*)	-0.112 ^(.)	-0.134 ^(.)	-0.099
	UNEMP	-0.079 ^(*)	-0.048 ^(.)	-0.028	0.027
	SOLD	0.119 ^(*)	0.061 ^(*)	0.081 ^(***)	0.076 ^(.)
	MEXP	-0.125	0.145	0.169 ^(*)	-0.038
	HEXP	-0.069	-0.121	-0.093	-0.165
	SAVE	0.104 ^(*)	0.124 ^(*)	0.155 ^(*)	0.171 ^(*)
	LIFE	-0.061 ^(*)	-0.064 ^(*)	-0.067 ^(*)	-0.159
	EXPO	-0.02	-0.012	-0.015	-0.041
	FDI	$7.64e - 12(.)$	$2.22e - 12$	$-2.56e - 12$	$-9.33e - 12(.)$
	INTEREST	-0.02	-0.01	-0.001	-0.002
	IMPO	0.025	0.022	0.024	0.044
	INF	-0.01 ^(***)	-0.004 ^(*)	-0.003 ^(***)	-0.002 ^(*)
	LABOR	-0.051 ^(.)	-0.016	-0.0004	0.034
	PINV	-0.014 ^(**)	-0.011	-0.011	-0.008
	POP	0.752 ^(**)	0.317	0.201	0.525 ^(.)
	RPOP	0.004	-0.004	-0.005	-0.040
	PPP	0.001	$6.2e - 5$	$-3e - 4$	-0.002
	REVENU	0.021	-0.008	-0.028	-0.096 ^(**)
	LAND	$-2.72e - 8$	$-2.23e - 8$	$-1.99e - 8$	$-8.14e - 8$
	AGR	-0.003	-0.009	-0.012	-0.009

Significance code: ./*/**/***, at 1%/5%/10%/0.1%, respectively.

Table C.5: Penalized quantile regression with fixed effect estimates, Model B

Model	Benchmark	A	B	C	
Interest Variables	EDUC	-0.314 ^(***)	-0.315 ^(***)	-0.394 ^(***)	-0.46 ^(***)
	SCH	-0.243 ^(***)	-0.243 ^(***)	0.135 ^(*)	0.099 ^(*)
	DEBT		0.001	-0.016 ^(**)	-0.025 ^(***)
	GFCF	0.156 ^(***)	0.157 ^(***)	0.098 ^(**)	0.108 ^(**)
Control Variables	CACNT			-0.0197	0.021
	UNEMP			-0.059 ^(**)	-0.028 ^(*)
	SOLD			0.200 ^(***)	0.152 ^(***)
	MEXP			0.035	-0.037
	HEXP			-0.050	-0.077
	SAVE			0.013	-0.019
	LIFE			-0.044 ^(*)	-0.03
	EXPO			-0.012	-0.022
	FDI			-3.6e-13	-1.86e-13
	INTEREST			0.007	0.005
	IMPO			0.006	0.02
	INF			-0.003 ^(***)	-0.033 ^(***)
	LABOR			-0.054 ^(***)	-0.043 ^(**)
	PINV			-0.007	-0.005
	POP			0.976 ^(***)	1.047 ^(***)
	RPOP			0.006	0.002
	PPP			-0.000	-0.000
REVENU			0.006	0.026	
LAND			-5.5e-08	-9.32e08 ^(.)	
AGR			-0.005	-0.009	
Time Invariant Variables	AFR				0.0757
	AMR				0.482 ^(.)
	ASI				-0.0673
	DEBT90				0.007 ^(.)
	GDP90				2.38e-15
	POP90				2.93e-09 ^(*)
R	0.083	0.083	0.269	0.274	
Adj-R	0.082	0.082	0.262	0.266	

Significance code: ./ ** / ***, at 1%/5%/10%/0.1%, respectively.

Table C.6: Panel data model estimation: TFE

Quantiles	τ			
	0.25	0.5	0.75	0.95
INTERCEPT	15.165 ^(*)	15.097 ^(**)	14.146 ^(**)	16.976
EDUC	-0.284 ^(.)	-0.244 ^(**)	-0.215 ^(.)	-0.194
SCH	0.167	0.128	0.153	0.155
DEBT	-0.028 ^(**)	-0.014 ^(**)	-0.01 ^(*)	-0.009
GFCF	-0.044	-0.038	0.023	0.031
CACNT	-0.121	-0.123 ^(*)	-0.093	-0.101 ^(.)
UNEMP	-0.013	0.005	0.011	0.032
SOLD	0.154 ^(**)	0.102 ^(***)	0.085 ^(**)	0.076 ^(*)
MEXP	-0.322	-0.058	0.103	0.183

HEXP	-0.135	-0.152	-0.155 ^(*)	-0.135
SAVE	0.148 ^(*)	0.147 ^(**)	0.13 ^(*)	0.131 ^(*)
LIFE	-0.081	-0.102 ^(*)	-0.108 ^(.)	-0.134
EXPO	-0.034	-0.026	-0.022	0.010
FDI	4.81e - 12	3.71e - 12	3.16e - 12 ^(*)	1.1e - 12
INTEREST	-0.018	-0.02	-0.015	-0.005
IMPO	0.046	0.052 ^(*)	0.046	0.014
INF	-0.008 ^(**)	-0.004 ^(**)	-0.003 ^(*)	-0.002 ^(**)
LABOR	-0.138 ^(.)	-0.089 ^(.)	-0.042	-0.030
PINV	-0.015 ^(**)	-0.01	-0.010	-0.001
POP	0.677 ^(.)	0.248	0.072	0.056
RPOP	-0.001	-0.003	-0.009	-8.53e - 05
PPP	0.001	0.001	0.000	0.000
REVENU	0.027 ^(.)	0.022 ^(.)	-0.001	-0.03
LAND	1.48e - 07 ^(.)	1.2e - 07	3.47e - 08	-2.28e - 08
AGR	-0.008	-0.016	-0.019	-0.032
ASI	0.353	0.272	-0.311	-0.46
DEBT90	-0.002	-0.006	0.002	0.011
GDP90	-1.94e - 16	-1.93e - 15	-5.10e - 15	-3.6e - 15
POP90	1.19e - 09	2.25e - 09	2.83e - 09	1.64e - 09
AD1.DEBT	0.003	0.002	0.003 ^(*)	0.007
AD1.EDUC	-0.25	-0.276	-0.261	-0.497
AD2.DEBT	-0.001	-0.005	-0.008	-0.011
AD2.EDUC	0.236	0.268	0.267	0.541
AD3.DEBT	0.006	0.002	-0.001	-0.006
AD3.EDUC	-0.149	-0.105	-0.027	0.077
AD4.DEBT	0.004	0.006	0.003	0.004
AD4.EDUC	0.285	0.196	0.206	0.173
AD5.DEBT	0.006 ^(.)	0.009	0.009	0.009
AD5.EDUC	-0.97 ^(**)	-1.001 ^(**)	-1.000 ^(**)	-1.52
AD6.DEBT	-0.006	-0.000	0.004	0.007
AD6.EDUC	0.236	0.713 ^(*)	0.839 ^(*)	1.215
AD7.DEBT	-0.002	-0.001	-0.000	0.007
AD7.EDUC	-0.038	-0.302	-0.375	-0.575
AD8.DEBT	0.008	0.008	0.007	0.012
AD8.EDUC	0.17	0.095	0.194	0.026
AD9.DEBT	-0.006	-0.005	-0.002	-0.002

AD9.EDUC	-0.304	-0.325	-0.203	0.373
AD10.DEBT	-0.021	-0.01	-0.015	-0.011
AD10.EDUC	0.098	0.206	0.006	-0.298
AD11.DEBT	0.048	0.035	0.026	-0.033
AD11.EDUC	-0.326	-0.364	-0.256	-0.082
AD12.DEBT	-0.054	-0.081	-0.076	-0.052
AD12.EDUC	0.541	0.304	0.146	-0.036
AD13.DEBT	-0.029	-0.004	0.023	0.069
AD13.EDUC	0.231	0.300	0.419	0.669
AD14.DEBT	0.069	0.063	0.058	0.057
AD14.EDUC	-0.344	-0.381	-0.565	-0.81
AD15.DEBT	-0.063	-0.032	-0.064	-0.113
AD15.EDUC	0.543	0.466	0.466	0.502
AD16.DEBT	0.106	0.082	0.068	0.069
AD16.EDUC	0.114	0.116	-0.039	-0.225
AD17.DEBT	-0.006	-0.012	0.0138	0.047
AD17.EDUC	0.182	0.080	0.02	0.028
AD18.DEBT	-0.089	-0.078	-0.053	-0.039
AD18.EDUC	-0.154	-0.109	0.058	0.322
AD19.DEBT	0.024	-0.041	-0.065	-0.11
AD19.EDUC	-0.2254	-0.06	-0.058	-0.023
AD20.DEBT	0.075	0.108	0.114	0.090
AD20.EDUC	0.116	0.256	0.346	0.318
AD21.DEBT	-0.028	-0.033	-0.03	0.007
AD21.EDUC	0.02	-0.033	-0.065	-0.045
AD22.DEBT	-0.022	-0.010	-0.020	-0.024
AD22.EDUC	-0.102	-0.186	-0.222	-0.285

Significance code: ./*/**/***, at 1%/5%/10%/0.1%, respectively.

Table C.7: Quantile regression with correlated random effects, Model C

Appendix D

Chapter 4 Appendix

D.1 Chapter 4 Tables

$\tau = 0.25$		0.5		0.75		0.95	
Country	Mean GDP Growth Rate	Country	Mean GDP Growth	Country	Mean GDP Growth	Country	Mean GDP Rate
Moldova	-1.54	Sweden	2.169	Zambia	3.151	Argentina	4.497
Ukraine	-1.474	United Kingdom	2.176	Australia	3.183	Egypt	4.505
Russia	0.652	Austria	2.276	El Salvador	3.253	Peru	4.521
Latvia	0.825	Canada	2.317	Uruguay	3.253	Israel	4.581
Kyrgyzstan	0.861	Netherlands	2.317	Ecuador	3.284	Thailand	4.601
Hungary	1.004	Iceland	2.4	Malta	3.498	Mauritius	4.663
Italy	1.021	Slovak Rep	2.422	Honduras	3.529	Costa Rica	4.744
Romania	1.059	Spain	2.436	Cyprus	3.563	Botswana	5.042
Bulgaria	1.064	United States	2.441	Colombia	3.632	Indonesia	5.093
Japan	1.091	Norway	2.53	Guatemala	3.713	Belize	5.1
Ivory Coast	1.309	South Africa	2.57	Philippines	3.793	Rwanda	5.144
Jamaica	1.476	Algeria	2.636	Morocco	3.794	Chile	5.147
Denmark	1.575	Brazil	2.746	Luxembourg	3.879	Dominican Rep	5.161
France	1.605	Mexico	2.841	Bolivia	3.91	Jordan	5.415
Germany	1.697	Paraguay	3.011	Trinidad and Tobago	4.078	Malaysia	6.045
Greece	1.75	Venezuela	3.05	Lesotho	4.082	Panama	6.073
Portugal	1.759	Nicaragua	3.098	Pakistan	4.212	India	6.491
Belgium	1.883	Kenya	3.13	Benin	4.285	Singapore	6.561
Finland	1.968	Albania	3.147	Nepal	4.446	Uganda	6.965

Table D.1: Countries by quantiles

Country	GDP rate	EDUC	HEXP	SCH	Country	GDP rate	EDUC	HEXP	SCH
Albania	3.147	3.352	5.7	9.932	Australia	3.183	4.836	8.268	11.845
Algeria	2.636	8.42	3.772	5.124	Austria	2.276	5.524	10.245	8.963
Argentina	4.497	4.36	8.517	8.503	Belgium	1.883	5.537	9.132	9.986
Belize	5.1	5.543	4.567	8.606	Canada	2.317	5.927	9.762	10.931
Benin	4.285	3.393	4.503	2.627	Chile	5.147	3.473	7.096	8.756
Bolivia	3.91	5.537	5.339	7.486	Denmark	1.575	7.977	9.5	9.954
Botswana	5.042	8.368	5.289	7.202	Finland	1.968	6.428	8.073	9.082
Brazil	2.746	2.731	7.646	5.173	France	1.605	5.582	10.792	8.787
Bulgaria	1.064	4.044	6.616	9.415	Germany	1.697	4.588	10.68	10.262
Colombia	3.632	3.736	6.818	6.34	Greece	1.75	3.031	9.369	8.791
Costa Rica	4.744	4.335	8.019	7.65	Hungary	1.004	5.304	7.626	10.611
Ivory Coast	1.309	4.512	5.239	2.64	Iceland	2.4	6.603	9.455	9.225
Cyprus	3.563	5.443	6.187	9.026	Israel	4.581	6.475	7.737	11.424
Dominican Rep	5.161	1.87	5.631	5.887	Italy	1.021	4.568	8.413	8.625
Ecuador	3.284	2.755	6.106	6.972	Japan	1.091	3.565	7.980	10.612
Egypt	4.505	4.364	5.089	4.636	Luxembourg	3.879	3.653	7.123	9.489
El Salvador	3.253	2.559	7.235	5.59	Mexico	2.841	4.346	5.565	6.916
Guatemala	3.713	2.067	5.868	3.495	Portugal	1.759	4.948	9.345	6.99
Honduras	3.529	3.649	7.179	5.193	Netherlands	2.317	5.039	9.584	10.674
India	6.491	3.174	4.124	3.534	Norway	2.53	7.133	9.084	11.626
Indonesia	5.093	2.441	2.456	4.087	Slovak Rep	2.422	4.28	6.846	11.188
Jamaica	1.476	4.913	4.849	7.994	Spain	2.436	4.328	8.107	8.261
Jordan	5.415	4.905	8.985	6.9	Sweden	2.169	6.744	8.834	10.875
Kenya	3.13	6.538	4.398	5.568	UK	2.176	5.04	7.876	8.72
Kyrgyzstan	0.861	5.065	6.054	8.793	US	2.441	5.35	15.218	12.564
Latvia	0.825	5.399	6.381	9.117					
Lesotho	4.082	10.281	8.325	4.897					
Malaysia	6.045	5.491	3.561	7.957					
Malta	3.498	5.069	7.607	8.919					
Mauritius	4.663	3.649	4.414	6.316					
Moldova	-1.54	7.079	9.330	8.812					
Morocco	3.794	5.557	4.868	3.174					
Nepal	4.446	2.97	5.594	2.411					
Nicaragua	3.098	4.515	8.207	4.577					
Pakistan	4.212	2.521	3.059	3.359					
Panama	6.073	4.465	7.570	8.361					
Paraguay	3.011	3.742	7.637	6.295					
Peru	4.521	2.951	4.741	8.578					
Philippines	3.793	2.97	3.504	7.808					
Romania	1.059	3.426	4.790	9.711					
Russia	0.652	3.608	5.777	8.368					
Rwanda	5.144	4.577	6.784	2.295					
Singapore	6.561	3.198	3.456	7.232					
South Africa	2.569	5.882	8.417	7.485					

Appendix D. Chapter 4 Appendix

Thailand	4.601	3.889	3.694	5.335				
Trinidad and Tobago	4.078	3.754	4.824	8.222				
Uganda	6.965	4.103	7.678	3.681				
Ukraine	-1.474	5.069	6.536	10.419				
Uruguay	3.253	2.212	9.868	7.717				
Venezuela	3.05	3.9	5.264	5.47				
Zambia	3.151	2.123	6.152	6.458				
Non-OECD Mean	3.563	4.323	5.965	6.497	OECD Mean	2.439	5.532	8.913 9.806

Table D.2: Education and health expenditure in % of GDP, GDP growth rates over 1990 – 2011

Quantile	0.25	0.5	0.75	0.95
INTERCEPT	9.581	7.516	8.006	10.781
EDUC	-0.383	-0.259	-0.274	-0.38
SCH	0.139	0.165	0.184	0.518
DEBT	-0.016	-0.0157	-0.016	-0.007
GFCF	-0.019	-0.054	-0.030	-0.047
CACNT	-0.069	-0.12	-0.118	-0.112
UNEMP	-0.067	-0.031	-0.003	0.091
SOLD	0.127	0.063	0.079	0.105
MEXP	-0.175	0.123	0.184	0.027
HEXP	-0.091	-0.133	-0.106	-0.09
SAVE	0.084	0.125	0.133	0.173
LIFE	-0.051	-0.047	-0.055	-0.112
EXPO	-0.028	-0.015	-0.015	-0.012
FDI	5.76e-12	6.02e-13	-3.490e-12	-1.26e-11
INTEREST	-0.024	-0.012	-0.003	0.001
IMPO	0.037	0.03	0.027	0.014
INF	-0.009	-0.004	-0.003	-0.002
LABOR	-0.05	-0.009	0.010	0.038
PINV	-0.014	-0.009	-0.011	-0.001
POP	0.771	0.366	0.241	0.632
PPP	0.001	0.000	-0.000	-0.001
REVENU	0.027	0.003	-0.017	-0.072
AGR	-0.005	-0.009	-0.011	0.005
POP90	3.24e-09	2.45e-09	1.59e-09	-7.23e-10
GDP90	2.58e-15	7.48e-18	-3.36e-15	-6.47e-15

Table D.3: Quantile regression estimation

Quantile	EDUC	SCH	DEBT	HEXP	LIFE
0.25	5.03	8.89	71.94	7.57	72.78
0.50	5.2	8.4	54.24	7.82	73.37
0.75	4.24	6.37	50.87	6.14	67.32
0.95	4.34	6.68	58.53	5.65	68.75

Table D.4: Some variables averages by quantiles

Country	PIP	Post Mean	Post SD	Cond.Pos.Sign	OECD
Albania	1.000	1.118	2.166	0.708	0
Algeria	1.000	0.191	2.188	0.535	0
Argentina	1.000	-1.633	2.586	0.220	0
Australia	1.000	0.045	0.678	0.638	1
Austria	1.000	0.041	0.783	0.506	1
Belgium	1.000	-1.202	0.979	0.000	1
Belize	1.000	1.359	2.161	0.959	0
Benin	1.000	1.556	2.213	0.976	0
Bolivia	1.000	0.295	2.183	0.535	0
Botswana	1.000	1.100	2.181	0.694	0
Brazil	1.000	-1.225	2.372	0.241	0
Bulgaria	1.000	-1.040	2.209	0.238	0
Canada	1.000	-0.021	0.587	0.328	1
Chile	1.000	0.365	0.715	0.920	1
Colombia	1.000	-0.388	2.264	0.530	0
Costa Rica	1.000	4.360	2.167	1.000	0
Ivory Coast	1.000	0.590	2.185	0.545	0
Cyprus	1.000	-0.184	2.213	0.532	0
Denmark	1.000	-0.944	0.632	0.000	1
Dominican Rep	1.000	2.297	2.184	1.000	0
Ecuador	1.000	0.441	2.174	0.536	0
Egypt	1.000	1.332	2.166	0.938	0
El Salvador	1.000	0.579	2.175	0.542	0
Finland	1.000	-0.691	0.583	0.000	1
France	1.000	-0.877	0.555	0.000	1
Germany	1.000	-0.719	0.605	0.000	1
Greece	1.000	-0.798	1.005	0.123	1
Guatemala	1.000	1.286	2.171	0.877	0
Honduras	1.000	1.270	2.172	0.866	0

Hungary	1.000	-0.597	0.799	0.060	1
Iceland	1.000	-0.325	0.813	0.147	1
India	1.000	2.098	2.283	1.000	0
Indonesia	1.000	1.630	2.177	1.000	0
Israel	1.000	0.075	0.848	0.611	1
Italy	1.000	-0.728	0.763	0.007	1
Jamaica	1.000	0.228	2.156	0.525	0
Japan	1.000	-0.789	0.604	0.000	1
Jordan	1.000	-0.594	2.345	0.532	0
Kenya	1.000	1.872	2.416	0.992	0
Kyrgyzstan	1.000	0.667	2.205	0.605	0
Latvia	1.000	-0.993	2.203	0.261	0
Lesotho	1.000	2.188	2.430	1.000	0
Luxembourg	1.000	0.117	0.619	0.773	1
Malaysia	1.000	0.569	2.183	0.541	0
Malta	1.000	-2.143	2.665	0.096	0
Mauritius	1.000	1.458	2.177	0.979	0
Mexico	1.000	0.175	0.711	0.801	1
Moldova	1.000	0.301	2.155	0.525	0
Morocco	1.000	0.727	2.161	0.555	0
Nepal	1.000	2.214	2.567	0.999	0
Netherlands	1.000	0.070	0.569	0.776	1
Nicaragua	1.000	0.716	2.162	0.554	0
Norway	1.000	0.147	0.582	0.952	1
Pakistan	1.000	1.664	2.232	0.998	0
Panama	1.000	0.399	2.206	0.541	0
Paraguay	1.000	0.677	2.162	0.546	0
Peru	1.000	0.088	2.267	0.541	0
Philippines	1.000	1.183	2.162	0.789	0
Portugal	1.000	-0.783	1.231	0.245	1
Romania	1.000	0.193	2.146	0.524	0
Russia	1.000	-1.188	2.256	0.197	0
Rwanda	1.000	2.584	2.611	1.000	0
Singapore	1.000	-1.988	2.983	0.212	0
Slovak Rep	1.000	0.194	1.186	0.535	1
South Africa	1.000	0.412	2.177	0.534	0
Spain	1.000	0.002	0.572	0.482	1

Sweden	1.000	-0.138	0.620	0.251	1
Thailand	1.000	1.889	2.251	0.999	0
Trinidad and Tobago	1.000	1.968	2.652	0.904	0
Uganda	1.000	2.574	2.596	1.000	0
Ukraine	1.000	-1.067	2.194	0.223	0
US	1.000	-0.013	0.600	0.411	1
Uruguay	1.000	-2.121	2.619	0.090	0
Venezuela	1.000	-1.785	2.591	0.174	0
Zambia	1.000	1.755	2.219	1.000	0
RPOP	0.941	-0.098	0.062	0.000	
GOVEFF⁽²⁾	0.819	-0.087	0.062	0.000	
RPOP * OECD	0.713	0.062	0.060	0.999	
RPOP⁽²⁾	0.674	0.001	0.001	0.930	
STU – TEACH	0.605	-0.022	0.033	0.000	
OECD	0.560	-2.454	3.504	0.032	
STU – TEACH²	0.485	0.000	0.000	0.640	

Table D.5: BMA estimates (Imputation 1)

Country	PIP	Post Mean	Post SD	Cond.Pos.Sign
Albania	1.000	1.084	2.126	0.800
Algeria	1.000	0.324	2.121	0.551
Argentina	1.000	-1.551	2.504	0.209
Australia	1.000	0.092	0.671	0.676
Austria	1.000	0.019	0.776	0.473
Belgium	1.000	-1.102	0.979	0.000
Belize	1.000	1.423	2.114	0.998
Benin	1.000	1.671	2.160	0.999
Bolivia	1.000	0.367	2.124	0.550
Botswana	1.000	1.144	2.118	0.803
Brazil	1.000	-1.140	2.303	0.234
Bulgaria	1.000	-0.963	2.151	0.215
Canada	1.000	-0.010	0.582	0.397
Chile	1.000	0.464	0.706	0.977
Colombia	1.000	-0.326	2.191	0.502
Costa Rica	1.000	4.382	2.114	1.000
Ivory Coast	1.000	0.710	2.127	0.600
Cyprus	1.000	-0.245	2.156	0.506

Denmark	1.000	-0.952	0.636	0.000
Dominican Rep	1.000	2.356	2.123	1.000
Ecuador	1.000	0.480	2.117	0.554
Egypt	1.000	1.441	2.117	0.998
El Salvador	1.000	0.638	2.116	0.566
Finland	1.000	-0.679	0.581	0.000
France	1.000	-0.836	0.555	0.000
Germany	1.000	-0.715	0.605	0.000
Greece	1.000	-0.762	0.989	0.143
Guatemala	1.000	1.335	2.117	0.994
Honduras	1.000	1.301	2.120	0.982
Hungary	1.000	-0.542	0.785	0.087
Iceland	1.000	-0.232	0.806	0.263
India	1.000	2.176	2.232	1.000
Indonesia	1.000	1.644	2.133	0.998
Israel	1.000	0.187	0.849	0.667
Italy	1.000	-0.664	0.757	0.014
Jamaica	1.000	0.211	2.107	0.526
Japan	1.000	-0.704	0.595	0.000
Jordan	1.000	-0.494	2.270	0.488
Kenya	1.000	1.875	2.359	0.997
Kyrgyzstan	1.000	0.636	2.165	0.620
Latvia	1.000	-0.969	2.147	0.215
Lesotho	1.000	2.155	2.372	0.999
Luxembourg	1.000	0.150	0.608	0.830
Malaysia	1.000	0.508	2.127	0.556
Malta	1.000	-2.105	2.580	0.097
Mauritius	1.000	1.478	2.133	0.997
Mexico	1.000	0.273	0.684	0.962
Moldova	1.000	0.205	2.120	0.516
Morocco	1.000	0.870	2.103	0.631
Nepal	1.000	2.219	2.508	0.999
Netherlands	1.000	0.053	0.568	0.702
Nicaragua	1.000	0.785	2.109	0.596
Norway	1.000	0.118	0.579	0.909
Pakistan	1.000	1.799	2.176	0.999
Panama	1.000	0.389	2.145	0.558

Paraguay	1.000	0.661	2.113	0.564
Peru	1.000	0.110	2.196	0.558
Philippines	1.000	1.175	2.110	0.893
Portugal	1.000	-0.769	1.216	0.230
Romania	1.000	0.197	2.106	0.520
Russia	1.000	-1.128	2.188	0.179
Rwanda	1.000	2.588	2.542	1.000
Singapore	1.000	-2.213	2.891	0.127
Slovak Rep	1.000	0.198	1.163	0.548
South Africa	1.000	0.420	2.116	0.547
Spain	1.000	0.032	0.572	0.688
Sweden	1.000	-0.135	0.619	0.251
Thailand	1.000	1.862	2.208	0.997
Trinidad and Tobago	1.000	1.837	2.609	0.901
Uganda	1.000	2.563	2.536	1.000
Ukraine	1.000	-1.157	2.141	0.161
US	1.000	-0.011	0.597	0.442
Uruguay	1.000	-2.056	2.536	0.098
Venezuela	1.000	-1.708	2.504	0.176
Zambia	1.000	1.774	2.169	0.999
RPOP	0.929	-0.096	0.063	0.000
RPOP * OECD	0.728	0.063	0.060	1.000
RPOP⁽²⁾	0.676	0.001	0.001	0.924
STU – TEACH	0.577	-0.019	0.031	0.000
OECD	0.540	-2.123	3.384	0.079
GOVEFF	0.512	0.092	0.142	1.000
STU – TEACH⁽²⁾	0.474	0.000	0.000	0.623

Table D.6: BMA estimates (Imputation 2)

Country	PIP	Post Mean	Post SD	Cond.Pos.Sign
Albania	1.000	1.110	2.138	0.742
Algeria	1.000	0.201	2.143	0.551
Argentina	1.000	-1.754	2.531	0.163
Australia	1.000	-0.030	0.705	0.548
Austria	1.000	0.056	0.781	0.525
Belgium	1.000	-1.252	0.987	0.000
Belize	1.000	1.275	2.130	0.899

Benin	1.000	1.554	2.171	0.996
Bolivia	1.000	0.221	2.139	0.549
Botswana	1.000	0.951	2.139	0.662
Brazil	1.000	-1.449	2.321	0.164
Bulgaria	1.000	-1.014	2.169	0.216
Canada	1.000	-0.109	0.634	0.196
Chile	1.000	0.286	0.751	0.843
Colombia	1.000	-0.477	2.213	0.491
Costa Rica	1.000	4.296	2.125	1.000
Ivory Coast	1.000	0.617	2.146	0.570
Cyprus	1.000	-0.168	2.172	0.537
Denmark	1.000	-0.957	0.645	0.000
Dominican Rep	1.000	2.228	2.139	1.000
Ecuador	1.000	0.358	2.135	0.552
Egypt	1.000	1.365	2.132	0.974
El Salvador	1.000	0.510	2.130	0.555
Finland	1.000	-0.680	0.592	0.000
France	1.000	-0.869	0.554	0.000
Germany	1.000	-0.712	0.603	0.000
Greece	1.000	-0.779	1.009	0.140
Guatemala	1.000	1.195	2.135	0.816
Honduras	1.000	1.158	2.137	0.783
Hungary	1.000	-0.582	0.799	0.073
Iceland	1.000	-0.368	0.828	0.102
India	1.000	2.075	2.249	1.000
Indonesia	1.000	1.637	2.150	1.000
Israel	1.000	0.119	0.828	0.640
Italy	1.000	-0.698	0.765	0.011
Jamaica	1.000	0.113	2.119	0.531
Japan	1.000	-0.799	0.600	0.000
Jordan	1.000	-0.591	2.288	0.499
Kenya	1.000	1.716	2.388	0.951
Kyrgyzstan	1.000	0.615	2.179	0.614
Latvia	1.000	-0.963	2.163	0.234
Lesotho	1.000	1.954	2.400	0.993
Luxembourg	1.000	0.115	0.620	0.769
Malaysia	1.000	0.538	2.143	0.561

Malta	1.000	-2.196	2.605	0.073
Mauritius	1.000	1.403	2.140	0.978
Mexico	1.000	0.156	0.714	0.778
Moldova	1.000	0.285	2.128	0.538
Morocco	1.000	0.764	2.119	0.584
Nepal	1.000	2.111	2.546	0.990
Netherlands	1.000	0.070	0.573	0.761
Nicaragua	1.000	0.711	2.121	0.574
Norway	1.000	0.156	0.588	0.949
Pakistan	1.000	1.739	2.196	0.999
Panama	1.000	0.263	2.160	0.555
Paraguay	1.000	0.575	2.129	0.558
Peru	1.000	0.038	2.212	0.557
Philippines	1.000	1.138	2.125	0.761
Portugal	1.000	-0.786	1.225	0.243
Romania	1.000	0.215	2.116	0.537
Russia	1.000	-1.212	2.210	0.170
Rwanda	1.000	2.470	2.577	1.000
Singapore	1.000	-2.091	2.920	0.170
Slovak Rep	1.000	0.177	1.180	0.551
South Africa	1.000	0.271	2.136	0.546
Spain	1.000	0.021	0.574	0.590
Sweden	1.000	-0.128	0.621	0.234
Thailand	1.000	1.789	2.224	0.997
Trinidad and Tobago	1.000	1.725	2.634	0.791
Uganda	1.000	2.392	2.575	1.000
Ukraine	1.000	-1.039	2.157	0.201
US	1.000	-0.070	0.611	0.259
Uruguay	1.000	-2.220	2.560	0.058
Venezuela	1.000	-1.875	2.527	0.132
Zambia	1.000	1.677	2.184	0.999
RPOP	0.948	-0.103	0.063	0.000
GOVEFF⁽²⁾	0.847	-0.089	0.058	0.000
RPOP * OECD	0.722	0.062	0.060	1.000
RPOP⁽²⁾	0.703	0.001	0.001	0.944
STU – TEACH	0.560	-0.015	0.024	0.000
OECD	0.537	-2.196	3.398	0.077

STU – TEACH ⁽²⁾	0.447	0.000	0.000	0.563
----------------------------	-------	-------	-------	-------

Table D.7: BMA estimates (Imputation 3)

Country	PIP	Post Mean	Post SD	Cond.Pos.Sign
Albania	1.000	0.927	2.116	0.655
Algeria	1.000	0.147	2.113	0.560
Argentina	1.000	-1.792	2.490	0.153
Australia	1.000	0.026	0.677	0.649
Austria	1.000	0.055	0.780	0.500
Belgium	1.000	-1.177	0.966	0.000
Belize	1.000	1.186	2.106	0.829
Benin	1.000	1.457	2.152	0.966
Bolivia	1.000	0.117	2.113	0.552
Botswana	1.000	0.870	2.124	0.654
Brazil	1.000	-1.432	2.290	0.160
Bulgaria	1.000	-1.114	2.137	0.163
Canada	1.000	-0.064	0.590	0.185
Chile	1.000	0.301	0.721	0.887
Colombia	1.000	-0.608	2.185	0.421
Costa Rica	1.000	4.219	2.104	1.000
Ivory Coast	1.000	0.535	2.118	0.571
Cyprus	1.000	-0.241	2.141	0.512
Denmark	1.000	-0.940	0.638	0.000
Dominican Rep	1.000	2.138	2.114	1.000
Ecuador	1.000	0.237	2.108	0.559
Egypt	1.000	1.330	2.107	0.973
El Salvador	1.000	0.359	2.104	0.566
Finland	1.000	-0.673	0.581	0.000
France	1.000	-0.860	0.554	0.000
Germany	1.000	-0.724	0.607	0.000
Greece	1.000	-0.772	1.000	0.136
Guatemala	1.000	1.104	2.114	0.759
Honduras	1.000	1.023	2.115	0.717
Hungary	1.000	-0.587	0.797	0.066
Iceland	1.000	-0.306	0.805	0.143
India	1.000	2.019	2.223	0.999
Indonesia	1.000	1.532	2.123	0.996

Israel	1.000	0.128	0.841	0.665
Italy	1.000	-0.687	0.759	0.007
Jamaica	1.000	0.017	2.098	0.513
Japan	1.000	-0.747	0.598	0.000
Jordan	1.000	-0.669	2.254	0.425
Kenya	1.000	1.613	2.358	0.887
Kyrgyzstan	1.000	0.446	2.159	0.578
Latvia	1.000	-1.084	2.132	0.172
Lesotho	1.000	1.784	2.381	0.960
Luxembourg	1.000	0.125	0.614	0.796
Malaysia	1.000	0.420	2.117	0.572
Malta	1.000	-2.177	2.564	0.073
Mauritius	1.000	1.316	2.122	0.938
Mexico	1.000	0.134	0.700	0.770
Moldova	1.000	0.090	2.108	0.515
Morocco	1.000	0.720	2.092	0.593
Nepal	1.000	2.032	2.503	0.995
Netherlands	1.000	0.088	0.576	0.826
Nicaragua	1.000	0.598	2.095	0.579
Norway	1.000	0.150	0.582	0.964
Pakistan	1.000	1.652	2.169	0.998
Panama	1.000	0.158	2.138	0.562
Paraguay	1.000	0.410	2.104	0.562
Peru	1.000	-0.118	2.183	0.548
Philippines	1.000	0.987	2.103	0.683
Portugal	1.000	-0.776	1.218	0.242
Romania	1.000	0.053	2.092	0.516
Russia	1.000	-1.308	2.174	0.138
Rwanda	1.000	2.359	2.539	0.999
Singapore	1.000	-2.142	2.872	0.141
Slovak Rep	1.000	0.181	1.176	0.564
South Africa	1.000	0.166	2.118	0.548
Spain	1.000	0.012	0.570	0.534
Sweden	1.000	-0.096	0.619	0.332
Thailand	1.000	1.701	2.198	0.994
Trinidad and Tobago	1.000	1.625	2.595	0.742
Uganda	1.000	2.328	2.533	0.999

Ukraine	1.000	-1.033	2.126	0.188
USt	1.000	-0.103	0.616	0.159
Uruguay	1.000	-2.255	2.524	0.046
Venezuela	1.000	-1.929	2.493	0.116
Zambia	1.000	1.515	2.161	0.979
GOVEFF⁽²⁾	0.966	-0.070	0.027	0.000
RPOP	0.935	-0.098	0.063	0.000
RPOP * OECD	0.716	0.061	0.059	1.000
RPOP⁽²⁾	0.684	0.001	0.001	0.930
OECD	0.523	-1.983	3.321	0.081
STU – TEACH	0.509	-0.013	0.024	0.000
GINI⁽²⁾	0.437	0.000	0.000	1.000

Table D.8: BMA estimates (Imputation 4)

Country	PIP	Post Mean	Post SD	Cond.Pos.Sign
Albania	1.000	1.117	2.129	0.773
Algeria	1.000	0.349	2.121	0.567
Argentina	1.000	-1.461	2.501	0.233
Australia	1.000	0.035	0.696	0.635
Austria	1.000	0.054	0.794	0.531
Belgium	1.000	-1.170	0.971	0.000
Belize	1.000	1.407	2.110	0.984
Benin	1.000	1.653	2.152	0.999
Bolivia	1.000	0.402	2.115	0.566
Botswana	1.000	1.160	2.110	0.739
Brazil	1.000	-1.104	2.294	0.237
Bulgaria	1.000	-0.938	2.150	0.245
Canada	1.000	-0.052	0.611	0.242
Chile	1.000	0.389	0.740	0.919
Colombia	1.000	-0.241	2.189	0.555
Costa Rica	1.000	4.402	2.107	1.000
Ivory Coast	1.000	0.699	2.128	0.587
Cyprus	1.000	-0.109	2.150	0.552
Denmark	1.000	-0.969	0.634	0.000
Dominican Rep	1.000	2.402	2.117	1.000
Ecuador	1.000	0.535	2.113	0.568
Egypt	1.000	1.435	2.117	0.988

El Salvador	1.000	0.665	2.108	0.572
Finland	1.000	-0.690	0.586	0.000
France	1.000	-0.851	0.556	0.000
Germany	1.000	-0.720	0.608	0.000
Greece	1.000	-0.769	0.999	0.122
Guatemala	1.000	1.376	2.111	0.990
Honduras	1.000	1.366	2.112	0.987
Hungary	1.000	-0.586	0.800	0.062
Iceland	1.000	-0.311	0.809	0.150
India	1.000	2.167	2.229	1.000
Indonesia	1.000	1.662	2.138	0.998
Israel	1.000	0.107	0.850	0.640
Italy	1.000	-0.708	0.768	0.008
Jamaica	1.000	0.284	2.101	0.549
Japan	1.000	-0.763	0.601	0.000
Jordan	1.000	-0.450	2.268	0.548
Kenya	1.000	1.918	2.354	0.995
Kyrgyzstan	1.000	0.627	2.168	0.614
Latvia	1.000	-0.925	2.145	0.252
Lesotho	1.000	2.197	2.363	1.000
Luxembourg	1.000	0.102	0.620	0.750
Malaysia	1.000	0.626	2.120	0.575
Malta	1.000	-2.003	2.577	0.105
Mauritius	1.000	1.475	2.127	0.987
Mexico	1.000	0.205	0.702	0.875
Moldova	1.000	0.252	2.119	0.541
Morocco	1.000	0.873	2.100	0.617
Nepal	1.000	2.218	2.508	0.998
Netherlands	1.000	0.047	0.574	0.648
Nicaragua	1.000	0.812	2.101	0.592
Norway	1.000	0.122	0.592	0.855
Pakistan	1.000	1.759	2.179	0.999
Panama	1.000	0.467	2.136	0.573
Paraguay	1.000	0.735	2.104	0.575
Peru	1.000	0.196	2.191	0.575
Philippines	1.000	1.224	2.108	0.874
Portugal	1.000	-0.754	1.219	0.245

Romania	1.000	0.208	2.106	0.542
Russia	1.000	-1.068	2.189	0.203
Rwanda	1.000	2.644	2.535	1.000
Singapore	1.000	-1.872	2.888	0.207
Slovak Rep	1.000	0.204	1.171	0.562
South Africa	1.000	0.501	2.107	0.564
Spain	1.000	0.017	0.574	0.545
Sweden	1.000	-0.139	0.618	0.186
Thailand	1.000	1.871	2.207	0.997
Trinidad and Tobago	1.000	1.873	2.599	0.883
Uganda	1.000	2.595	2.531	1.000
Ukraine	1.000	-1.063	2.142	0.194
US	1.000	-0.021	0.617	0.381
Uruguay	1.000	-1.981	2.532	0.096
Venezuela	1.000	-1.608	2.507	0.194
Zambia	1.000	1.808	2.160	1.000
RPOP	0.925	-0.094	0.063	0.000
RPOP * OECD	0.698	0.058	0.059	0.999
RPOP⁽²⁾	0.669	0.000	0.001	0.919
GOVEFF⁽²⁾	0.661	-0.069	0.073	0.000
STU – TEACH	0.577	-0.018	0.029	0.000
OECD	0.536	-2.127	3.348	0.062
STU – TEACH⁽²⁾	0.463	0.000	0.000	0.591

Table D.9: BMA estimates (Imputation 5)

Country group	STU-TEACH	RPOP
Developed	15.38	23.264
In Transition	23.429	50.27
Developing	29.624	46.62

Table D.10: Student per teacher ratio & Rural population levels by development level

Thèse de Doctorat

Maya Murched

Capital Humain, Dette Publique et Croissance Economique à Long Terme

Résumé

La croissance économique et ses moteurs représentent le principal sujet préoccupant les chercheurs en macroéconomie depuis longtemps. Investir en capital humain à travers le système éducatif joue un rôle important pour stimuler la croissance et le développement économique, cette accentuation a pris place depuis la naissance innovante de la théorie de la croissance endogène.

L'attention et les efforts dévoués à l'investissement dans le capital humain peuvent être déstabilisés par le retour global et récent de la crise de la dette souveraine dans plusieurs pays, dette qui poursuit son ascension depuis 2007, et les politiques d'ajustement nécessaires d'après-crise. Des judicieuses politiques de redressement devraient être composées d'un mélange des activités encourageant la croissance économique, y compris l'investissement dans le capital humain, l'austérité et le long terme.

L'objectif principal de cette thèse est de fournir des nouvelles évidences empiriques sur la relation dette-croissance économique et leurs externalités sur la formation de capital humain, les estimations sont réalisées sur un jeu de données récent et complet couvrant 22 années et 76 countries dans le monde.

L'ensemble des variables utilisées englobe de nombreux agrégats macroéconomiques tel que: taux de croissance annuel du PIB, la dette publique en % de PIB, les dépenses publiques d'éducation en % de PIB, le moyen d'année de scolarité, le taux d'inflation, et d'autres.

En utilisant une technique d'estimation semi-paramétrique appropriée qui offre des solutions pour de nombreux problèmes concernant les données, les résultats empiriques suggèrent un impact négatif et hétérogène de dette et des dépenses d'éducation publiques sur la croissance du PIB. Là où, l'utilisation des dépenses d'éducation dans l'ensemble de l'échantillon est inefficace, les décideurs politiques devraient ajuster et bien gérer la fonction de ces dépenses en même temps de viser des efforts publics pour réduire les niveaux élevés d'endettement et d'augmenter la croissance économique.

Nous montrons également que l'utilisation des outils d'analyse textuelle en économie, offre une lecture rapide et globale des courants de recherche contenus dans la littérature empirique et théorique de la croissance économique.

Mots clés

Croissance économique, capital humain, dette publique, dépenses d'éducation, efficacité, imputation multiple, régression par quantile.

Abstract

Economic growth and its driving forces have been the main topic preoccupying economic researchers since long time in macroeconomic branch. Public investment in human capital through educational system plays an ultimate role in boosting economic growth and development, this role has taken a place since the innovative dawn of endogenous growth theory.

The focus and efforts of investing in human capital could be destabilized by the global and recent return of sovereign debt crisis in several countries, which continues its rise since the early 2007, and the after-crisis necessary adjusting policies. Getting back wise policies should be composed of mixture of growth fostering activities, including the investment in human capital, austerity and forbearance.

The main purpose of this thesis is to provide new empirical inferences on debt-growth relationship and its interaction with human capital formation. Estimates are carried on a recent and complete data set that spans over 22 years and involves 76 countries worldwide.

The range of invested variables encompasses many macroeconomic aggregates such as: GDP annual growth rates, public debt to GDP ratio, and public education expenditure to GDP ratio, average schooling years, inflation rate, and others.

Using a superior estimation semi-parametric technic which accounts for some data issues, the empirical results suggest a heterogeneous impact of public debt and education expenditures levels on GDP growth rates. Henceforth, the use of education expenditure in the whole sample is inefficient, where policy makers should adjust and well manage the function of these expenditure in line with the public efforts to reduce debt high levels and rise economic growth.

We also show that the use of textual analysis tools in economic studies, such in growth literature, offers a rapid and total lecture of the hidden research trends embodied in the huge empirical and theoretical literature of economic growth.

Key Words

Economic growth, human capital, public debt, education expenditure, efficiency, multiple imputation, quantile regression, panel data.