

GÖTTINGER STUDIEN ZUR ENTWICKLUNGSÖKONOMIK /  
GÖTTINGEN STUDIES IN DEVELOPMENT ECONOMICS

Boris Branisa Caballero

# **Social Institutions, Gender Inequality, and Regional Convergence in Developing Countries**



PETER LANG

Boris Branisa Caballero

## **Social Institutions, Gender Inequality, and Regional Convergence in Developing Countries**

This book contributes to the understanding of gender and regional inequalities in developing countries. First, it deals with social institutions related to gender inequality and proposes new composite indices to measure them. Using these indices, some interesting empirical connections between social institutions related to gender inequality and several relevant development outcomes are examined at the cross-country level. The second part of the book is concerned with the historical development of another type of inequality which is relevant for developing countries: inequality between regions. The topic of regional convergence in Colombia during the last quarter of the 20th century is analyzed using different approaches and focusing on both income and social indicators.

Boris Branisa Caballero studied Economics and Finance at the University of Geneva and obtained a PhD in Economics from the University of Göttingen. During his PhD studies he completed the PhD Program 'Applied Statistics and Empirical Methods' of the Center of Statistics at the same university, and worked as a consultant for the Organisation for Economic Co-operation and Development (OECD) and the German Agency for International Cooperation (GIZ). He had previously held several positions at the Central Bank of Bolivia and is presently conducting postdoctoral research at the University of Mannheim.

**Social Institutions, Gender Inequality, and Regional Convergence  
in Developing Countries**

**Göttinger Studien zur Entwicklungsökonomik  
Göttingen Studies in Development Economics**

Herausgegeben von/ Edited by Hermann Sautter und/and Stephan Klasen

Bd./Vol. 33



**PETER LANG**

Frankfurt am Main · Berlin · Bern · Bruxelles · New York · Oxford · Warszawa · Wien



Boris Branisa Caballero

**Social Institutions,  
Gender Inequality,  
and Regional Convergence  
in Developing Countries**



**PETER LANG**

Internationaler Verlag der Wissenschaften

**Bibliographic Information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at <http://dnb.d-nb.de>.

Open Access: The online version of this publication is published on [www.peterlang.com](http://www.peterlang.com) and [www.econstor.eu](http://www.econstor.eu) under the international Creative Commons License CC-BY 4.0. Learn more on how you can use and share this work: <http://creativecommons.org/licenses/by/4.0>.



This book is available Open Access thanks to the kind support of ZBW – Leibniz-Informationszentrum Wirtschaft.

Zugl.: Göttingen, Univ., Diss., 2010

Cover design:

© Olaf Glöckler, Atelier Platen, Friedberg

Cover illustration by Rolf Schinke

Gratefully acknowledging financial support  
by Ibero-Amerika-Institut  
für Wirtschaftsforschung, Göttingen.

D 7

ISSN 1439-3395

ISBN 978-3-631-61422-8

ISBN 978-3-631-75354-5 (eBook)

© Peter Lang GmbH

Internationaler Verlag der Wissenschaften

Frankfurt am Main 2012

All rights reserved.

All parts of this publication are protected by copyright. Any utilisation outside the strict limits of the copyright law, without the permission of the publisher, is forbidden and liable to prosecution. This applies in particular to reproductions, translations, microfilming, and storage and processing in electronic retrieval systems.

[www.peterlang.de](http://www.peterlang.de)

**A mi familia**



# Editor's Preface

The international community has agreed that gender equality is one of the main topics in the development agenda. The Millennium Development Goal number 3, for example, is "promote gender equality and empower women". Considerable effort has been devoted recently to quantify and understand gender inequalities at the cross-country level, but most of the existing research concentrates on inequalities based on outcome measures such as education, health or participation, while the institutional basis of these inequalities is often overlooked. For policy action, however, understanding these institutional drivers of gender inequality seems crucial. This book makes a significant contribution in this regard.

In **part 1** of this book, Boris Branisa makes a twofold contribution to the discussion of gender issues and development. The first is related to the measurement and understanding of what Amartya Sen calls substantive freedoms. Under this approach, the success of a society is to be evaluated primarily by the substantive freedoms that people enjoy. In **essay 1** Branisa explores the measurement of social institutions related to gender inequality. These institutions are understood as long-lasting norms, values and codes of conduct that shape everyday life and determine role models that people try to fulfill and satisfy, and as such they are essential to understand gender roles. He uses variables from the OECD Development Centre's Gender, Institutions and Development database and proposes several composite measures of social institutions related to gender inequality. Five subindices combine variables that proxy one dimension of social institutions: Family code, Civil liberties, Physical integrity, Son preference and Ownership rights. The aggregation procedure is based on polychoric principal component analysis. The five one-dimensional measures are then combined to construct the Social Institutions and Gender Index (SIGI) which is a multidimensional measure of social institutions related to gender inequality. The aggregation of the dimensions follows the Foster-Greer-Thorbecke approach to poverty measurement. The SIGI and the five composite measures are helpful to understand the deprivation of women and allow ranking and comparing over 100 developing countries as well as to identify priority areas where action is needed in a given country. The essay also shows that these measures complement existing measures and indicators of gender inequality.

The second main contribution of **part 1** is related to another idea championed by Amartya Sen: Going beyond the intrinsic importance of freedom as the objective of development, one should also consider the instrumental effectiveness of freedom of different kinds to promote human freedom, as greater freedom means that people can exert more influence in their lives and at the societal level. In **essays 2 and 3**, Branisa examines some interesting empirical connections at the cross-country level between social institutions related to gender inequality measured by the composite indices proposed in essay 1, and relevant development outcomes. **Essay 2** reviews some of the existing theoretical literature such as household bargaining models and formulates hypothesis about the potentials effects of social institutions related to gender inequality on female education, child mortality, fertility, and governance measured as rule of law and governance. The empirical results show that among developing countries higher inequality in social institutions is associated with worst development outcomes, even after accounting for differences in religion, geography, political system, and the level of income. The focus of **essay 3** is the link between social institutions related to gender inequality, and corruption. The study contributes to the existing literature on the topic showing that when the opportunities of women to participate in social life are restricted in developing countries, the perceived level of corruption tends to be higher. This empirical result is robust, and holds when one accounts for other possible factors that influence corruption.

**Part 2** of this book is concerned with the evolution over time of another type of inequality which is also pertinent for most developing countries, namely inequality between regions. Branisa specifically deals with the question of regional convergence among departments in Colombia in the last quarter of the 20th century understood as whether departments that were lagging behind the national average have been able to catch up in that period. **Essay 4** presents a sound review of the concepts and of the main econometric approaches to measure convergence empirically, and explores the Colombian case discussing crucial data issues and focusing on the two existing yearly time series of consistent per capita income measures: gross departmental product and gross household disposable income. The results suggest no convergence if one relies on gross departmental product, and that only a very slow convergence took place if one observes gross household disposable income.

**Essay 5** examines convergence among departments in Colombia during a similar period, but concentrating on alternative non-income indicators. As suggested among others by Amartya Sen, it is important to go beyond income measures and focus on social opportunities which contribute to the overall freedom that people have to live as they choose. Branisa discusses relevant public policies and major reforms put in place in Colombia during the period, as well as data and measurement issues concerning social indicators, and empirically examines convergence using variables reflecting outcomes related to education, health and

nourishment. The main results show on one hand that there has been convergence in basic education. On the other hand, no robust evidence of convergence is found using health measures, which seems consistent with the results of **essay 4**.

Taken together, the essays in this book make an important contribution to the understanding of gender and regional inequalities and are of interest to scholars and policy-makers alike.

Prof. Stephan Klasen, Ph.D.

Göttingen, November 2011





# Author's Preface

Completing this research project was only possible through the support of many people, and I would like to thank all of them here.

I express my sincere gratitude to my first supervisor, Professor Stephan Klasen, for accepting me as a doctoral student at the Chair of Development Economics and for his constant support and excellent guidance. I benefited a lot from his expertise as a development economist and as an econometrician, and I really enjoyed working with him.

I wish to thank my second supervisor, Professor Walter Zucchini, for his friendly help, permanent encouragement and advice. He is not only a brilliant man, but also a great teacher and a very supportive person.

Professor Inmaculada Martínez-Zarzoso was the third member of my dissertation committee, and I thank her very much for her time, stimulating questions and suggestions.

Doing research in Göttingen was a valuable experience for me, and allowed me to interact and work with many nice colleagues at the Chair of Development Economics, the Ibero-America Institute for Economic Research, the Courant Research Centre “Poverty, Equity and Growth in Developing Countries”, the Institute for Statistics and Econometrics, and the Centre for Statistics. My sincere thanks go to all of them, especially to my co-authors Maria Ziegler and Adriana Cardozo. Carola Grün, Michael Grimm, Felicitas Nowak-Lehmann, and Axel Dreher, among others, provided useful comments to some of the essays in this book.

In addition, comments from participants of seminars, workshops and conferences in Berlin, Galway, Göttingen, Hannover, La Paz, Mannheim, Montevideo, Paris, Rio de Janeiro, Rome, Sevilla and Tokyo helped improving this book.

Friends and family are very important in my life. I would like to acknowledge their help along the road during these years. I owe my loving thanks to my parents, brothers, and grandparents for always believing in me and continuously encouraging me to reach my goals.

Finally, my deepest and warmest thanks go to my wife Gabriela and my sons Alexander and Santiago for the love they bring into my life, their unconditional support and great patience during this process.

Funding for this research project through a Georg-Christoph Lichtenberg scholarship provided by the Ministry of Science and Culture of Lower-Saxony and granted by the Centre for Statistics at the University of Göttingen is gratefully acknowledged.

**Boris Branisa Caballero**  
Göttingen, November 2011

# Contents

<b>Editor's Preface</b>	<b>vii</b>
<b>Author's Preface</b>	<b>xi</b>
<b>List of Tables</b>	<b>xvii</b>
<b>List of Figures</b>	<b>xxi</b>
<b>List of Abbreviations</b>	<b>xxiii</b>
<b>Introduction and Overview</b>	<b>1</b>
<b>I Social institutions and gender inequality</b>	<b>5</b>
<b>1 The Social Institutions and Gender Index (SIGI)</b>	<b>7</b>
1.1 Introduction . . . . .	8
1.2 The Database . . . . .	10
1.3 Construction of the Subindices . . . . .	12
1.3.1 Measuring the Association between Categorical Variables . . . . .	12
1.3.2 Aggregating Variables to Build a Subindex . . . . .	14
1.4 The Social Institutions and Gender Index (SIGI) . . . . .	15
1.5 Results . . . . .	18
1.5.1 Country Rankings and Regional Patterns . . . . .	18
1.5.2 Simple Correlation with other Gender-related Indices . . . . .	20
1.5.3 Regression Analysis . . . . .	21
1.6 Conclusion . . . . .	22
1.7 Tables . . . . .	24
1.8 Figures . . . . .	43

<b>2</b>	<b>Why care about social inst. related to gender ineq.</b>	<b>49</b>
2.1	Introduction . . . . .	50
2.2	Social Institutions and Household Decisions . . . . .	53
2.2.1	Social Institutions and Female Education . . . . .	54
2.2.2	Social Institutions and Fertility and Child Mortality Rates . . . . .	55
2.3	Social Institutions and the Society: Governance . . . . .	57
2.4	Data . . . . .	57
2.5	Empirical estimation and Results . . . . .	60
2.5.1	Empirical estimation . . . . .	60
2.5.2	Results . . . . .	62
2.6	Conclusion . . . . .	65
2.7	Tables . . . . .	67
<b>3</b>	<b>Reexamining the link between gender and corruption</b>	<b>79</b>
3.1	Introduction . . . . .	80
3.2	Empirical Estimation and Results . . . . .	83
3.2.1	Data . . . . .	83
3.2.2	Empirical Estimation . . . . .	87
3.2.3	Results . . . . .	89
3.3	Conclusion . . . . .	91
3.4	Tables . . . . .	93
3.5	Figures . . . . .	101
<b>II</b>	<b>Regional growth convergence in Colombia</b>	<b>103</b>
<b>4</b>	<b>Regional convergence in Colombia: Income indicators</b>	<b>105</b>
4.1	Introduction . . . . .	106
4.2	Motivation and Background . . . . .	107
4.2.1	Economic Background . . . . .	107
4.2.2	Data Issues Affecting Convergence Results in Colombia . . . . .	111
4.3	The Solow Model and Its Estimation . . . . .	114
4.3.1	The Solow Model . . . . .	114
4.3.2	Absolute Beta-Convergence . . . . .	115
4.3.3	Conditional Convergence . . . . .	116
4.3.4	Parameter Heterogeneity: Are There Different Steady States? . . . . .	117
4.3.5	Sigma-Convergence . . . . .	118
4.4	Distributional Approach: Quah's Critique . . . . .	119

---

4.5	Empirical Estimation and Results . . . . .	121
4.5.1	Sigma-Convergence . . . . .	121
4.5.2	Absolute Beta-Convergence . . . . .	121
4.5.3	Conditional Beta-Convergence Using Control Variables . . . . .	122
4.5.4	Beta-Convergence Using Time-Series Cross-Sectional Data . . . . .	122
4.5.5	Kernel Density Estimators . . . . .	127
4.6	Conclusions . . . . .	128
4.7	Tables . . . . .	132
4.8	Figures . . . . .	146
<b>5</b>	<b>Regional convergence in Colombia: Social indicators</b>	<b>161</b>
5.1	Introduction . . . . .	162
5.2	Motivation . . . . .	163
5.3	Methods for Measuring Convergence . . . . .	166
5.4	Data and Empirical Estimation . . . . .	168
5.4.1	Data . . . . .	168
5.4.2	Empirical estimation . . . . .	169
5.5	Results . . . . .	171
5.5.1	Literacy Rate . . . . .	172
5.5.2	Infant Survival Rate . . . . .	173
5.5.3	Life Expectancy at Birth . . . . .	174
5.5.4	Nourishment . . . . .	175
5.6	Conclusions . . . . .	175
5.7	Tables . . . . .	180
5.8	Figures . . . . .	187
	<b>Appendices</b>	<b>203</b>
	Appendix to Essay 1 . . . . .	205
	<b>Bibliography</b>	<b>211</b>



# List of Tables

1.1	Kendall tau b: Dimension Family Code . . . . .	24
1.2	Kendall tau b: Dimension Civil Liberties . . . . .	24
1.3	Kendall tau b: Dimension Physical Integrity with Missing Women . . . . .	25
1.4	Kendall tau b: Dimension Physical Integrity without Missing Women . . . . .	25
1.5	Kendall tau b: Dimension Ownership Rights . . . . .	26
1.6	Weights from Polychoric PCA . . . . .	26
1.7	Kendall tau b between Subindices . . . . .	27
1.8	Pearson Correlation Coefficient ( $\rho$ ) between the SIGI and the Simple Average of the Five Subindices . . . . .	27
1.9	Comparison of the SIGI and the Simple Average of the Subindices . . . . .	28
1.10	Ranking according to the SIGI and the Five Subindices . . . . .	31
1.11	Regional Pattern of the Composite Index and Subindices . . . . .	37
1.12	Statistical Association between the SIGI and other Gender-related Measures . . . . .	38
1.13	Comparison of Ranks: the SIGI and other Gender-related Indices . . . . .	39
1.14	Linear Regression with Dependent Variables GGG and Ratio GDI to HDI . . . . .	42
2.1	Description and Sources of Variables . . . . .	67
2.2	Descriptive statistics of variables used . . . . .	70
2.3	Pearson Correlation Coefficient between the SIGI and the Subindices . . . . .	71
2.4	Correlation of the SIGI and the Subindices with the Control Variables . . . . .	72
2.5	Linear regressions with dependent variable female secondary school . . . . .	74
2.6	Linear regressions with dependent variable fertility . . . . .	75
2.7	Linear regressions with dependent variable child mortality . . . . .	76
2.8	Linear regressions with dependent variable ‘voice and accountability’ . . . . .	77
2.9	Linear regressions with dependent variable ‘rule of law’ . . . . .	78
3.1	Description and Sources of Variables . . . . .	93
3.2	Descriptive statistics of variables used . . . . .	96

3.3	Pearson Correlation Coefficient between subindex Civil liberties and control variables . . . . .	97
3.4	Pearson Correlation Coefficient ( $\rho$ ) between the Corruption Measures . . .	97
3.5	Variation of the subindex civil liberties over religion . . . . .	98
3.6	Linear regressions with dependent variable CPI . . . . .	99
3.7	Linear regressions with dependent variable ICRG . . . . .	100
4.1	Colombia. Gross Domestic Product (Constant Million Pesos of 1994), Per Capita GDP and Population. 1980-2006. . . . .	132
4.2	Beta Convergence Using Cross-sections and Non Linear Least Squares. Dependent Variable: Average Growth Rate of pc PDB 1975-2000. . . . .	133
4.3	Beta Convergence Using Cross-sections and Non Linear Least Squares. Dependent Variable: Average Growth Rate of pc IDBH 1975-2000. . . . .	133
4.4	NLLS Regressions. Dependent Variable Average Growth Rate of pc PDB. 1975-2000. . . . .	134
4.5	NLLS Regressions. Dependent Variable Average Growth Rate of pc IDBH. 1975-2000. . . . .	135
4.6	OLS Linear Regression. TSCS Data. Dependent Variable $\log(y_{i,t})$ . Relative Per capita PDB. 1975-2000. . . . .	136
4.7	OLS Linear Regression. TSCS Data. Dependent Variable $\log(y_{i,t})$ . Relative Per Capita IDBH. 1975-2000. . . . .	136
4.8	Linear Mixed Model (REML). TSCS Data. Dependent Variable: $\log(y_{i,t})$ . Relative Per Capita PDB. 1975-2000. . . . .	137
4.9	Linear mixed model (REML). TSCS Data. Dependent Variable: $\log(y_{i,t})$ . Relative Per Capita IDBH. 1975-2000. . . . .	137
4.10	Implied Convergence Rates Using TSCS Data and Linear Mixed Models (REML). Per capita PDB. 1975-2000 . . . . .	138
4.11	Implied Convergence Rates Using TSCS Data and Linear Mixed Models (REML). Per capita IDBH. 1975-2000 . . . . .	139
4.12	Autoregressive Processes of Order 1. Dependent Variable: $\log(y_{i,t})$ . Per Capita PDB. 1975-2000. . . . .	140
4.13	Autoregressive Processes of Order 1. Dependent Variable: $\log(y_{i,t})$ . Per capita IDBH. 1975-2000. . . . .	141
4.14	Mixture Model with 3 Components. Fitted with ML. Dependent Variable: $\log(y_{i,t})$ . Relative Per capita PDB. 1975-2000. . . . .	142
4.15	Mixture Model with 3 Components. Fitted with ML. Dependent Variable: $\log(y_{i,t})$ . Relative Per Capita IDBH. 1975-2000. . . . .	143



---

4.16	Bootstrap tests of equality of univ. density estimates, Income indicators . . .	144
4.17	Share of Total PDB and Share of selected PDB Sectors by Department. Years 1975 and 2000. . . . .	145
5.1	Descriptive Statistics of the Variables Used . . . . .	180
5.2	Beta Convergence (OLS) using all available Observations. Dependent Variable: Change in Literacy Rate (LIT) between 1973 and 2005 . . . . .	181
5.3	Beta convergence (OLS) excluding Outliers. Dependent Variable: Change in Literacy Rate (LIT) between 1973 and 2005 . . . . .	181
5.4	Bootstrap tests of equality of univ. density estimates, Social indicators . . .	182
5.5	Beta Convergence (OLS) using all available Observations. Dependent Variable: Change in Infant Survival Rate (ISR) between 1975 and 2000 . . . . .	182
5.6	Beta Convergence (OLS) excluding Outliers. Dependent Variable: Change in Infant Survival Rate (ISR) between 1975 and 2000 . . . . .	183
5.7	Beta Convergence (OLS) using all available Observations. Dependent Variable: Change in Life Expectancy at Birth (LEX) between 1975 and 2000 . . .	183
5.8	Beta Convergence (OLS) excluding Outliers. Dependent Variable: Change in Life Expectancy at Birth (LEX) between 1975 and 2000 . . . . .	184
5.9	Beta Convergence (OLS) using all available Observations. Dependent Variable: Change in Well-nourished Rate (WR) between 1995 and 2005 . . . . .	184
5.10	Beta Convergence (OLS) excluding Outliers. Dependent Variable: Change in Well-nourished Rate (WR) between 1995 and 2005 . . . . .	185
5.11	Poverty Headcount Index (% of Households below the Poverty Line) by Department. 1996 to 2005 . . . . .	186



# List of Figures

1.1	MJCA for the Dimension Family Code . . . . .	43
1.2	MJCA for the Dimension Civil Liberties . . . . .	44
1.3	MJCA for the Dimension Physical Integrity with Missing Women . . . . .	45
1.4	MJCA for the Dimension Physical Integrity without Missing Women . . . . .	46
1.5	MJCA for the Dimension Ownership Rights . . . . .	47
3.1	Scatter plot: Subindex Civil liberties and percentage of Muslim population . . . . .	101
4.1	Map of Colombia. . . . .	146
4.2	Box Plot: Log of Per Capita PDB. 1975-2000. . . . .	147
4.3	Box Plot: Log of Per Capita IDBH. 1975-2000. . . . .	147
4.4	Box Plot: Log of Relative Per Capita PDB. 1975-2000. . . . .	148
4.5	Box Plot: Log of Relative Per Capita IDBH. 1975-2000. . . . .	148
4.6	Sigma Convergence. GDP by Department. . . . .	149
4.7	Sigma Convergence. Per Capita Gross Departmental Product (PDB) and Gross Personal Disposable Income (IDBH). 1975-2000. . . . .	149
4.8	Beta Convergence. Per Capita PDB. 1975-2000. . . . .	150
4.9	Beta Convergence without Nuevos, Chocó and Guajira. Per Capita PDB. 1975-2000. . . . .	150
4.10	Beta Convergence. Per Capita IDBH. 1975-2000. . . . .	151
4.11	Beta Convergence without Guajira. Per Capita IDBH. 1975-2000. . . . .	151
4.12	Log of Relative PDB by Department. 1975-2000. . . . .	152
4.13	Log of relative IDBH by department. 1975-2000. . . . .	153
4.14	Log of Relative PDB. All Departments. 1975-2000. . . . .	154
4.15	Log of Relative IDBH. All Departments. 1975-2000. . . . .	154
4.16	Univariate Kernel Density Estimators of Relative Per Capita PDB. Years 1975 and 2000. Constant Prices of 1994. . . . .	155
4.17	Relative Per Capita PDB Dynamics. Years 1975 and 2000. Constant Prices of 1994. . . . .	156

4.18	Relative per capita PDB Dynamics: Contour Plot. Years 1975 and 2000. Constant Prices of 1994. . . . .	157
4.19	Univariate Kernel Density Estimators of Relative per Capita IDBH. Years 1975 and 2000. Constant Prices of 1994. . . . .	158
4.20	Relative Per Capita IDBH Dynamics. Years 1975 and 2000. Constant Prices of 1994. . . . .	159
4.21	Relative per Capita IDBH Dynamics: Contour Plot. Years 1975 and 2000. Constant Prices of 1994. . . . .	160
5.1	Evolution of Literacy Rate. 1973-2005 . . . . .	187
5.2	Univariate Kernel Density Estimators of Relative Literacy Rate. 1973 and 2005. . . . .	188
5.3	Bivariate Kernel Density Estimators of Relative Literacy Rate. 3D Representation. 1973 and 2005. . . . .	189
5.4	Bivariate Kernel Density Estimators of Relative Literacy Rate. Contour Plot. 1973 and 2005. . . . .	190
5.5	Evolution of Infant Survival Rate. 1975-2000 . . . . .	191
5.6	Univariate Kernel Density Estimators of Relative Infant Survival Rate. 1975 and 2000. . . . .	192
5.7	Bivariate Kernel Density Estimators of Relative Infant Survival Rate. 3D Representation. 1975 and 2000. . . . .	193
5.8	Bivariate Kernel Density Estimators of Relative Infant Survival Rate. Contour Plot. 1975 and 2000. . . . .	194
5.9	Evolution of Life Expectancy at Birth. 1975-2000 . . . . .	195
5.10	Univariate Kernel Density Estimators of Relative Life Expectancy at Birth. 1975 and 2000. . . . .	196
5.11	Bivariate Kernel Density Estimators of Relative Life Expectancy at Birth. 3D Representation. 1975 and 2000. . . . .	197
5.12	Bivariate Kernel Density Estimators of Relative Life Expectancy at Birth. Contour Plot. 1975 and 2000. . . . .	198
5.13	Evolution of Well-Nourished Rate. 1995-2005 . . . . .	199
5.14	Univariate Kernel Density Estimators of Relative Well-nourished Rate. 1995 and 2005. . . . .	200
5.15	Bivariate Kernel Density Estimators of Relative Well-nourished Rate. 3D Representation. 1995 and 2005. . . . .	201
5.16	Bivariate Kernel Density Estimators of Relative Well-nourished Rate. Contour Plot. 1973 and 2005. . . . .	202

# List of Abbreviations

AIC	Akaike Information Criterion
CEDAW	Convention on the Elimination of All Forms of Discrimination against Women
CEGA	Centro de Estudios Ganaderos
CPI	Corruption Perception Index
DANE	Departamento Nacional Administrativo de Estadísticas
DHS	Demographic and Health Survey
DNP	Departamento Nacional de Planeación
EAP	East Asia and Pacific
ECA	Europe and Central Asia
FGT	Foster-Greer-Thorbecke
FPC	First Principal Component
GDI	Gender-Related Development Index
GDP	Gross Domestic Product
GEM	Gender Empowerment Measure
GGG	Global Gender Gap Index
GGI	Gender Gap Index Capped
GID	Gender, Institutions and Development Database
HC	Heteroscedasticity-consistent
HDI	Human Development Index
ICRG	Corruption in Government Index
IDB	Gross Departmental Income
IDBH	Gross Household Disposable Income at the department level
ISR	Infant Survival Rate
LAC	Latin America and the Caribbean
LEX	Life Expectancy at Birth
LIT	Literacy Rate
MENA	Middle East and North Africa
MJCA	Multiple Joint Correspondence Analysis
ML	Maximum Likelihood

NLLS	Non Linear Least Squares
NPV	Net Present Value
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PDB	Gross Departmental Product
REML	Restricted Maximum Likelihood
SA	South Asia
SIGI	Social Institutions and Gender Index
SNA-86	System of National Accounts of 1986
SNA-93	System of National Accounts of 1993
SSA	Sub-Saharan Africa
TSCS	Time-Series–Cross-Section Data
UNDP	United Nations Development Programme
WOSOC	Women’s Social Rights index
WR	Well-nourished rate

# Introduction and Overview

This book is a collection of five empirical essays and is divided into two independent parts. Part one comprises three essays that deal with social institutions related to gender inequality at the cross-country level. Part two investigates whether there was convergence across Colombian departments during the last quarter of the 20th century.

## Part I: Social institutions and gender inequality

The importance of striving for gender equality has been recognized and incorporated in the international development agenda, e.g. in Millennium Development Goal 3 “Promote gender equality and empower women” or in the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW). Nevertheless, when it comes to measurement of gender inequality at the cross-country level, most of the attention centers on measures that proxy gender inequality in well-being or in agency, and which are typically outcome-focused (Klasen, 2006, 2007). Focusing only on outcomes neglects the relevant question of the origins of these inequalities and their great heterogeneity. Gender inequality is the result of human behavior, and how people behave and interact is influenced by institutions. Hence, to understand gender inequality in outcomes, one needs to study the institutional basis of gender inequality.

In **Essay 1** we propose new composite measures that proxy social institutions related to gender inequality in non-OECD countries based on variables of the OECD Gender, Institutions and Development database (Morrison and Jütting, 2005; Jütting et al., 2008). We aggregate the variables into five subindices that each measure one dimension of social institutions related to gender inequality (Family code, Civil liberties, Physical integrity, Son preference and Ownership rights). We combine the subindices into the Social Institutions and Gender Index (SIGI) as a multidimensional measure of the deprivation of women caused by social institutions. Methodologically, the SIGI is inspired by the Foster-Greer-Thorbecke poverty measures. It offers a new way of aggregating gender inequality in several dimensions, penal-

izing high inequality in each dimension and allowing only for partial compensation between dimensions.

The SIGI and the subindices are useful tools to compare the societal situation of women in over 100 non-OECD countries from a new perspective, allowing the identification of problematic countries and dimensions of social institutions that deserve attention by policy makers and need to be scrutinized in detail. Empirical results show that the SIGI provides additional information to that of other well-known gender-related indices. Moreover, regression analysis shows that the SIGI is related to indices that measure outcome gender inequality, even if one takes into account region, religion and level of economic development.

Institutions are a major factor explaining development outcomes in general. **Essay 2** focuses on social institutions related to gender inequality understood as long-lasting norms, values and codes of conduct that shape gender roles, and presents evidence on why they matter for development. We derive hypotheses from existing theories and empirically test them at the cross-country level with linear regressions using the SIGI and its subindices as measures for social institutions. We find that apart from geography, political system, religion, and the level of economic development, one has to consider social institutions related to gender inequality to better account for differences in development. Our results show that social institutions that deprive women of their autonomy and bargaining power in the household, or that increase the private costs and reduce the private returns to investments into girls, are associated with lower female education, higher fertility rates and higher child mortality. Moreover, social institutions related to gender inequality are negatively associated with governance measured as ‘rule of law’ and ‘voice and accountability’.

**Essay 3** reexamines the link between gender inequality and corruption. We review the literature on the relationship between representation of women in economic and political life, democracy and corruption, and bring in a new previously omitted variable that captures the level of discrimination against women in a society: social institutions related to gender inequality. Using a sample of developing countries we regress corruption on the representation of women, democracy and other control variables. Then we add the subindex Civil liberties proposed in **Essay 1**, as it covers social institutions that directly shape the opportunities of women to participate in social life. The results show that corruption is higher in countries where social institutions deprive women of their freedom to participate in social life, even accounting for democracy and representation of women in political and economic life as well as for other variables. Our findings suggest that, in a context where social values disadvantage women, it might not be enough to push democratic reforms and to increase the participation of women to reduce corruption.



## Part II: Regional growth convergence in Colombia

Colombia is the third most populated country in Latin America. According to its macroeconomic performance, it is considered in general a successful story in the region. It is one of the few countries that did not default on its external debt during the ‘lost decade’ of last century and which did not experience hyperinflation. The annual growth rate of per capita GDP between 1975 and 2005 was 1.4 percent, which is twice as much as the Latin American average in the same period. If one takes a broader perspective of development and focuses on other indicators as education or health, Colombia is close to or slightly above the Latin American average.

However, Colombia is also well-known for large regional disparities in income and in social indicators. In **Essays 4 and 5**, we focus on departments, which are important political entities in the country, with elected local governments and separate department assemblies. We investigate whether there was convergence among them, i.e. if departments that were lagging behind have been able to catch up during the last quarter of the 20th century.

**Essay 4** focuses on growth convergence across Colombian departments during the period of 1975 to 2000, following both the regression and the distributional approaches suggested in the literature, and using two income measures computed by Centro de Estudios Ganaderos (CEGA). We also discuss issues related to data provided by Departamento Administrativo Nacional de Estadísticas (DANE) used by previous convergence studies. Our results show no evidence supporting convergence using per capita gross departmental product, but rather persistence in the distribution. Using per capita gross household disposable income, we find some evidence of convergence, but only at a low speed, close to one percent per year. Furthermore, we find no evidence of the existence of different steady states for the two variables considered.

**Essay 5** investigates convergence in social indicators among Colombian departments from 1973 to 2005. We use census data and apply both the regression approach and the distributional approach (univariate and bivariate kernel density estimators). Using literacy rate as a proxy for education, we find convergence between 1973 and 2005, but persistence in the distribution between 1975 and 2000, when we use the infant survival rate and life expectancy at birth as proxies for health. Additionally, using data from Demographic and Health Surveys, we find some evidence of convergence in the rate of children that are well-nourished between 1995 and 2005.



## **Part I**

### **Social institutions and gender inequality**



# Essay 1

## The Institutional Basis of Gender Inequality: The Social Institutions and Gender Index (SIGI) and its Subindices

### Abstract

In this paper we construct the Social Institutions and Gender Index (SIGI) and its five subindices Family code, Civil liberties, Physical integrity, Son Preference and Ownership rights using variables of the OECD Development Centre's Gender, Institutions and Development database. Instead of measuring gender inequality in education, health, economic or political participation, these indices allow a new perspective on gender issues in developing countries. The SIGI and the subindices measure long-lasting social institutions which are mirrored by societal practices and legal norms that frame gender-relevant meanings and form the basis of gender roles. The subindices measure each one dimension of the concept and the SIGI combines the subindices into a multidimensional index of deprivation of women caused by social institutions. Methodologically, the SIGI is inspired by the Foster-Greer-Thorbecke poverty measures. It offers a new way of aggregating gender inequality in several dimensions, penalizing high inequality in each dimension and allowing only for partial compensation between dimensions. The SIGI and the subindices are useful tools to identify countries and dimensions of social institutions that deserve attention. Empirical results confirm that the SIGI provides additional information to that of other well-known gender-related indices.

## 1.1 Introduction

Despite considerable progress in recent decades, gender inequality in the manifold dimensions of well-being remains pervasive in many developing countries. This is an intrinsic issue of equity as the affected women are deprived of their basic freedoms (Sen, 1999). But going beyond this intrinsic feature of gender inequality, there is considerable evidence that it implies high costs for society in the form of lower human capital, worse governance, and lower growth (e.g. World Bank, 2001; Klasen, 2002; Klasen and Lamanna, 2009). The intrinsic and instrumental value of gender equality has been recognized and incorporated in the development agenda, for example in Millennium Development Goal 3 “Promote gender equality and empower women” or in the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW).

To measure the extent of this problem at the cross-country level several gender-related indices have been proposed, e.g. the Gender-Related Development Index (GDI) and the Gender Empowerment Measure (GEM) (United Nations Development Programme, 1995) and more recently the Gender Inequality Index (GII) (United Nations Development Programme, 2010), the Global Gender Gap Index (GGG) from the World Economic Forum (Lopez-Claros and Zahidi, 2005), the Gender Equity Index developed by Social Watch (2005) or African Gender and Development Index (AGDI) proposed by the Economic Commission for Africa (2004). These measures focus on gender inequality in well-being or in agency and they are typically outcome-focused (Klasen, 2006, 2007). Focusing only on outcomes neglects the question of the origins of these inequalities and their great heterogeneity across space and time. Gender inequality is the result of human behavior, and how people behave and interact is influenced by institutions. Thus to understand gender inequality in outcomes, one needs to study the institutional basis of gender inequality.

There are several approaches to institutions. According to North (1990, p. 3 ff.) “institutions are the rules of the game in a society”, they are “humanly devised constraints that shape human interaction”. From an economics perspective, institutions are conceived as the result of collective choices in a society to achieve gains from cooperation by reducing uncertainty, collective action dilemmas and transaction costs. A sociological or cultural perspective, which is complementary to the rational choice one, relates institutions to culture. Institutions in this sense frame meanings and beliefs. People try to satisfy norms rather than to act individually within the rules of the game, i.e. institutions do not canalize preferences of actors, they influence the preferences and shape the role models and identities of the actors themselves. Actors and institutions amalgamate so that actors are often not aware of the guiding principles of their behavior. Legitimacy and appropriateness drive institutional evolution more than efficiency considerations. Cultural authority, power in a society and

community dynamics might be more relevant in shaping such institutions that become taken-for-granted without continuously being evaluated against efficiency considerations (Hall and Taylor, 1996, and references therein).

There is a particular type of institutions that is relevant for gender inequality, *social institutions related to gender inequality*. These institutions are more embedded in the cultural-sociological account although efficiency issues may also matter. We conceive these social institutions as long-lasting norms, values and codes of conduct that find expression in traditions, customs and cultural practices, informal and formal laws. They are at the bottom of gender roles and the distribution of power between men and women in the family, in the market and in social and political life. As social institutions related to gender inequality build an often taken-for-granted basis of people's behavior and interaction in all spheres of life, they shape the social and economic opportunities of men and women, their autonomy in taking decisions (Dyson and Moore, 1983; Abadian, 1996; Hindin, 2000; Bloom et al., 2001) or their capabilities to live the life they value (Sen, 1999). That is why they might affect important development outcomes and contribute to outcome gender inequalities (De Soysa and Jütting, 2007).

There are three measures at the country level that somehow proxy social institutions, which determine how women are treated in society: the Women's Political Rights index (WOPOL), the Women's Economic Rights index (WECON), and the Women's Social Rights index (WOSOC) of the CIRI Human Rights Data Project.<sup>1</sup> These indices take a human rights perspective and measure on a yearly basis whether a number of internationally recognized rights for women are included in law and whether government enforces them. From the three indices, WOSOC is the most encompassing measure covering social relations (Bjornskov et al., 2009). However, it does not allow one to differentiate between different dimensions of social institutions. For example, it is important to distinguish between what happens within the family and what happens in public and social life. Furthermore, other shortcomings of all three indices are that they also cover outcomes of institutions, and they can only take four values from 0 (no rights) to 3 (legally guaranteed and enforced rights) which makes it difficult to compare and rank countries as there are many ties in the data.

In this paper we propose new composite measures that proxy social institutions related to gender inequality in non-OECD countries based on variables of the OECD Development Centre's Gender, Institutions and Development Database (Morrison and Jütting, 2005; Jütting et al., 2008). We aggregate the variables into five subindices that each measure one dimension of social institutions related to gender inequality (Family code, Civil liberties, Physical integrity, Son preference and Ownership rights). We combine the subindices into

---

<sup>1</sup>Information is available on the webpage of the project <http://ciri.binghamton.edu/>.

the Social Institutions and Gender Index (SIGI) as a multidimensional measure of the deprivation of women.

In general, the construction of composite measures requires several decisions, for example about the weighting scheme and the method of aggregation (e.g. Nardo et al., 2005). The subindices as one-dimensional measures are built using the method of polychoric principal component analysis to extract the common information of the variables corresponding to a subindex. When we combine the subindices to construct the SIGI, we use a reasonable methodology to capture the multidimensional deprivation of women caused by social institutions. The formula of the SIGI is inspired by the Foster-Greer-Thorbecke poverty measures (Foster et al., 1984) and offers a new way of aggregating gender inequality in several dimensions measured by the subindices. It is transparent and easy to understand, it penalizes high inequality in each dimension and allows only for partial compensation between dimensions.

The SIGI and the subindices are useful tools to compare the societal situation of women in over 100 non-OECD countries from a new perspective, allowing the identification of problematic countries and dimensions of social institutions that deserve attention by policy makers and need to be scrutinized in detail. Empirical results show that the SIGI provides additional information to that of other well-known gender-related indices. Moreover, regression analysis shows that the SIGI is related to indices that measure outcome gender inequality, even if one takes into account region, religion and level of economic development.

This paper is organized as follows. In section 1.2, we describe the OECD Development Centre's Gender, Institutions and Development Database. Then, in sections 1.3 and 1.4 we focus on the construction of the subindices and of the SIGI. In section 1.5, we present empirical results by country, interesting regional patterns and a comparison between the SIGI and other gender-related measures. Furthermore, using regression analysis we illustrate the relevance of the SIGI for explaining outcome gender inequality. The last section concludes with a discussion of the strengths and weaknesses of the proposed measures.

## 1.2 The OECD GID Database

As input for the composite measures we use variables from the OECD Development Centre's Gender, Institutions and Development (GID) Database (Morrison and Jütting, 2005; Jütting et al., 2008). This is a cross-country database covering about 120 countries with more than 20 variables measuring social institutions related to gender inequality.<sup>2</sup> These variables proxy social institutions through prevalence rates, legal indicators or indicators of social practices.

---

<sup>2</sup>The data are available at the web-pages <http://www.wikigender.org> and <http://www.oecd.org/dev/gender/gid>.



We assume that the concept social institutions related to gender inequality is multidimensional. Following previous work done by the OECD (Jütting et al., 2008) we choose twelve variables that are assumed to measure each one of four dimensions of social institutions.

The *Family code* dimension refers to the private sphere with institutions that influence the decision-making power of women in the household. Family code is measured by the following four variables. *Parental authority* measures whether women have the right to be the legal guardian of a child during marriage, and whether women have custody rights over a child after divorce. *Inheritance* is based on formal inheritance rights of spouses and children. *Early marriage* measures the percentage of girls between 15 and 19 years of age who are/were ever married. *Polygamy* measures the acceptance of polygamy in the population. Countries where this information is not available are assigned scores based on the legality of polygamy.<sup>3</sup>

The public sphere is measured by the *Civil liberties* dimension that captures the freedom of social participation of women and includes the following two variables. *Freedom of movement* measures the level of restrictions women face in moving freely outside their own household. *Freedom of dress* measures the extent to which women are obliged to follow a certain dress code in public, for example being obliged to cover their face or body when leaving the house.

The *Physical integrity* dimension comprises different indicators on violence against women. The variable *violence against women* indicates the existence of laws against domestic violence, sexual assault or rape, and sexual harassment. *Female genital mutilation* is the percentage of women who have undergone female genital mutilation. *Missing women* measures gender bias in mortality. Countries were coded based on estimates of gender bias in mortality for a sample of countries (Klasen and Wink, 2003) and on sex ratios of young people and adults.

The *Ownership rights* dimension covers the economic sphere of social institutions proxied by the access of women to several types of property. *Women's access to land* indicates whether women are allowed to own land. *Women's access to bank loans* measures whether women are allowed to access credits. *Women's access to property other than land* covers mainly access to real property such as houses, but also any other property.

Concerning the *missing women* variable in the *Physical integrity* dimension, it could be argued that it reflects another dimension of gender inequality. Missing women is an extreme manifestation of son preference under scarce resources. 100 million women are not alive

<sup>3</sup>Acceptance of polygamy in the population might proxy actual practices better than the formal indicator legality of polygamy and, moreover, laws might be changed faster than practices. Therefore, the acceptance variable is the first choice for the subindex Family code. The reason for using legality when acceptance is missing is to increase the number of countries.

who should be alive if women were not discriminated against (Sen, 1992; Klasen and Wink, 2003). The other components of *Physical integrity, violence against women and female genital mutilation*, measure particularly the treatment of women which is not only motivated by economic considerations. In the next section, we check with statistical methods if *missing women* measures another dimension as the variables *violence against women* and *female genital mutilation*.

These twelve variables are between 0 and 1. The value 0 means no or very low inequality and the value 1 indicates high inequality. Three of the variables (*early marriage, female genital mutilation and violence against women*) are continuous. The other indicators measure social institutions on an ordinal categorical scale. The chosen variables cover around 120 non-OECD countries from all regions in the world except North America.<sup>4</sup> The choice of the variables is also guided by the availability of information so that as many countries as possible can be ranked by the SIGI. Within our sample 102 countries have information for all twelve variables.

### 1.3 Construction of the Subindices

The objective of the subindices is to provide a summary measure for each dimension of social institutions related to gender inequality. In every subindex we want to combine variables that are assumed to belong to one dimension. The first step is to check the statistical association between the variables. The second step consists in aggregating the variables with a reasonable weighting scheme.

#### 1.3.1 Measuring the Association between Categorical Variables

To check the association between variables, and as most of them are ordinal, we use Kendall Tau b and Multiple Joint Correspondence Analysis (Greenacre, 2007; Nenadić, 2007).

Kendall Tau b is a rank correlation coefficient. These measures are useful when the data are ordinal and thus the conditions for using Pearson's correlation coefficient are not fulfilled. For each variable, the values are ordered and ranked. Then the correspondence between the rankings is measured.<sup>5</sup>

---

<sup>4</sup>The OECD Gender, Institutions and Development Database does not contain variables that capture relevant social institutions related to gender inequality in OECD countries.

<sup>5</sup>For calculating Kendall Tau, one counts the number of concordant and discordant pairs of two rankings, builds the difference and divides this difference by the total number of pairs. A value of 1 means total correspondence of rankings, i.e. the rankings are the same. A value of -1 indicates reverse rankings or a negative association between rankings. A value of 0 means independence of rankings. Kendall Tau b is a variant of Kendall tau

Taking into account tied pairs, the formula for Kendall Tau b is

$$\tau_b = \frac{C - D}{\sqrt{\frac{n(n-1)}{2-T_x} \frac{n(n-1)}{2-T_y}}}, \quad (1.1)$$

where  $C$  is the number of concordant pairs,  $D$  is the number of discordant pairs,  $n$  is the number of observations,  $\frac{n(n-1)}{2}$  is the number of all pairs,  $T_x$  is the number of pairs tied on the variable  $x$  and  $T_y$  is the number of pairs tied on the variable  $y$ . The notation is taken from Agresti (1984). The p-value of tau b under the null hypothesis of no association between the variables is computed with the approximation suggested by Kendall (1976), which is adequate unless ties are very extensive. As in our case many ties are present, we confirm the results with an asymptotically distribution-free confidence interval for Kendall's tau b based on the bootstrap method with 1000 replications (Hollander and Wolfe, 1999).

As a second method to check the association between variables we examine the graphics produced by Multiple Joint Correspondence Analysis (MJCA) (Greenacre, 2007; Nenadić, 2007), after having discretized the three continuous variables. Correspondence Analysis is a method for analyzing and representing the structure of contingency tables graphically. We use MJCA to find out whether variables seem to measure the same.<sup>6</sup>

The results for Kendall tau b are reported in Tables 1.1-1.5. A significant positive value of Kendall tau b is a sign for a positive association between two variables. This is the case for all variables belonging to one dimension, except *missing women* in the subindex *Physical integrity*. The graphs produced with MJCA are shown in Figures 1.1-1.5.<sup>7</sup> The results of MJCA also confirm that within every dimension all the variables seem to measure the same dimension, with the exception of *missing women* in the dimension *Physical integrity*. These results support the argumentation in section 1.2.

---

that corrects for ties, which are frequent in the case of discrete data (Agresti, 1984, chap. 9). We consider Kendall Tau b to be the appropriate measure of rank correlation to find out whether our data are related.

<sup>6</sup>Correspondence Analysis is an exploratory and descriptive method to analyze contingency tables. Instead of calculating a correlation coefficient to capture the association of variables, the correspondence of conditional and marginal distributions of either rows or columns - also called row or column profiles - is measured using a  $\chi^2$ -statistic, that captures the distance between them. These row or column profiles then are plotted in a low-dimensional space, so that the distances between the points reflect the dissimilarities between the profiles. Multiple Joint Correspondence Analysis is an extended procedure for the analysis of more than two variables and considers the cross-tabulations of the variables against each other in a so-called Burt matrix but with modified diagonal sub-tables. This facilitates to figure out whether variables are associated. This is the case when they have similar deviations from homogeneity, and therefore get a similar position in a profile space (Greenacre, 2007; Nenadić, 2007).

<sup>7</sup>The graphs produced with MJCA can be interpreted in the following way. In most cases, one of the axes represents whether there is inequality and the other axis represents the extent of inequality. If one connects the values of a variable one obtains a graphical pattern. If this is similar to the pattern obtained for another variable, then both variables are associated.

We decide to use the variable *missing women* as a fifth subindex called *Son preference*. The artificially higher female mortality is one of the most important and cruel aspects of gender inequality and should not be neglected, as over 100 million women that should be alive are missing (Sen, 1992; Klasen and Wink, 2003). Missing women is the “starkest manifestation of the lack of gender equality” (Duflo, 2005).

### 1.3.2 Aggregating Variables to Build a Subindex

The five subindices *Family code*, *Civil liberties*, *Son preference*, *Physical integrity* and *Ownership rights* use the twelve variables as input that were mentioned in the previous section. Each subindex combines variables that measure one dimension of social institutions related to gender inequality. In the case of *Son preference*, the subindex takes the value of the variable *missing women*. In all other cases, the computation of the subindex values involves two steps.

In the first step, the method of polychoric principal component analysis is used to extract the common information of the variables corresponding to a subindex. Principal component analysis (PCA) is a method of dimensionality reduction that is valid for normally distributed variables (Jolliffe, 1986). This assumption is violated in this case, as the data include variables that are ordinal, and hence the Pearson correlation coefficient is not appropriate. Following Kolenikov and Angeles (2004, 2009) we use polychoric PCA, which relies on polychoric and polyserial correlations. These are estimated with maximum likelihood, assuming that there are latent normally distributed variables that underly the ordinal categorical data. We use the First Principal Component (FPC) as a proxy for the common information contained by the variables corresponding to the subindices, measuring each one of the dimensions of social institutions related to gender inequality. The FPC is the weighted sum of the standardized original variables that captures as much of the variance in the data as possible.<sup>8</sup> The standardization of the original variables is done as follows. In the case of continuous variables, one subtracts the mean and then divides by the standard deviation. In the case of ordinal categorical variables, the standardization uses results of an ordered probit model. The weight that each variable gets in these linear combinations is obtained by analyzing the correlation structure in the data. The weights are shown in Table 1.6.

In the second step, the subindex value is obtained rescaling the FPC so that it ranges from 0 to 1 to ease interpretation. A country with the best possible performance (no inequality) is assigned the value 0 and a country with the worst possible performance (highest inequality) the value 1. Hence, the subindex values of all countries are between 0 and 1. Using the

<sup>8</sup>The proportion of explained variance by the first principal component is 70% for *Family code*, 93% for *Civil liberties*, 60% for *Physical integrity* and 87% for *Ownership rights*.

score of the FPC the subindex is calculated using the following transformation. Country  $X$  corresponds to a country of interest, Country *Worst* corresponds to a country with worst possible performance and Country *Best* is a country with best possible performance.

$$\text{Subindex}(\text{Country } X) = \frac{\text{FPC}(\text{Country } X)}{\text{FPC}(\text{Country } \textit{Worst}) - \text{FPC}(\text{Country } \textit{Best})} - \frac{\text{FPC}(\text{Country } \textit{Best})}{\text{FPC}(\text{Country } \textit{Worst}) - \text{FPC}(\text{Country } \textit{Best})} \quad (1.2)$$

Every subindex is intended to measure a different dimension of social institutions related to gender inequality. To check whether the subindices are empirically non-redundant, so that they provide each additional information, we conduct an empirical analysis of the statistical association between them. In the case of well-being measures, McGillivray and White (1993) suggest using two explicit thresholds to separate redundancy from non-redundancy, that is a correlation coefficient of 0.90 and 0.70. Based on this suggestion we use the threshold 0.80. In Table 1.7 we present Kendall tau b as a measure of the statistical association between the five subindices. In all cases, the subindices are positively correlated, showing that they all measure social institutions related to gender inequality. It must be noted, however, that the correlation is not always statistically significant. Kendall tau b is lower than 0.80 in all cases, which suggests that each subindex measures a distinct aspect of social institutions related to gender inequality.

## 1.4 The Social Institutions and Gender Index (SIGI)

With the subindices described in the last section as input, we build a multidimensional composite index named Social Institutions and Gender Index (SIGI) which reflects the deprivation of women caused by social institutions related to gender inequality. The proposed index is transparent and easy to understand. As in the case of the variables and of the subindices, the index value 0 corresponds to no inequality and the value 1 to complete inequality.

The SIGI is an unweighted average of a non-linear function of the subindices. We use equal weights for the subindices, as we see no reason for valuing one of the dimensions more or less than the others.<sup>9</sup> The non-linear function arises because we assume that inequality in gender-related social institutions leads to deprivation experienced by the affected women,

<sup>9</sup>Empirically, even in the case of equal weights the ranking produced by a composite index is influenced by the different variances of its components. The component that has the highest variance has the largest influence on the composite index. In the case of the SIGI the variances of the five components are reasonably close to each other, *Ownership rights* having the largest and *Physical integrity* having the lowest variance.

and that deprivation increases more than proportionally when inequality increases. Thus, high inequality is penalized in every dimension. The non-linearity also means that the SIGI does not allow for total compensation among subindices, but permits partial compensation. Partial compensation implies that high inequality in one dimension, i.e. subindex, can only be partially compensated with low inequality on another dimension.<sup>10</sup>

For our specific five subindices, the value of the index the SIGI is then calculated as follows.

$$\begin{aligned} \text{SIGI} &= \frac{1}{5} (\text{Subindex Family Code})^2 + \frac{1}{5} (\text{Subindex Civil Liberties})^2 \\ &+ \frac{1}{5} (\text{Subindex Physical Integrity})^2 + \frac{1}{5} (\text{Subindex Son preference})^2 \\ &+ \frac{1}{5} (\text{Subindex Ownership Rights})^2 \end{aligned} \quad (1.3)$$

Using a more general notation, the formula for the SIGI  $I(X)$ , where  $X$  is the vector containing the values of the subindices  $x_i$  with  $i = 1, \dots, n$ , is derived from the following considerations. For any subindex  $x_i$ , we interpret the value 0 as the goal of no inequality to be achieved in every dimension. We define a deprivation function  $\phi(x_i, 0)$ , with  $\phi(x_i, 0) > 0$  if  $x_i > 0$  and  $\phi(x_i, 0) = 0$  if  $x_i = 0$  (e.g. Subramanian, 2007). Higher values of  $x_i$  should lead to a penalization in  $I(X)$  that should increase with the distance  $x_i$  to zero. In our case the deprivation function is the square of the distance to 0 so that deprivation increases more than proportionally as inequality increases.

$$\text{SIGI} = I(X) = \frac{1}{n} \sum_{i=1}^n \phi(x_i, 0) = \frac{1}{n} \sum_{i=1}^n (x_i - 0)^2 = \frac{1}{n} \sum_{i=1}^n (x_i)^2. \quad (1.4)$$

The formula is inspired by the Foster-Greer-Thorbecke (FGT) poverty measures (Foster et al., 1984). The general FGT formula is defined for  $y_i \leq z$  as:

$$\text{FGT}(Y, \alpha, z) = \frac{1}{n} \sum_{i=1}^n \left( \frac{z - y_i}{z} \right)^\alpha, \quad (1.5)$$

where  $Y$  is the vector containing all incomes,  $y_i$  with  $i = 1, \dots, n$  is the income of individual  $i$ ,  $z$  is the poverty line, and  $\alpha > 0$  is a penalization parameter.

<sup>10</sup>Other approaches have been also proposed in the literature, e.g. the non-compensatory approach by Munda and Nardo (2005a,b).

To compute the SIGI, the value 2 is chosen for  $\alpha$  as the square function has the advantage of easy interpretation. With  $\alpha = 2$  the *transfer principle* is satisfied (Foster et al., 1984). In the context of poverty this principle means that a transfer from a person below the poverty line to a person less poor will raise poverty if the set of poor remains unchanged. In the case of the SIGI, the transfer principle means that an increase in inequality in one dimension and a decrease of inequality in another dimension of the same magnitude will raise the SIGI, assuming that both dimensions had equal values at the beginning.

Some differences between the SIGI and the FGT measures must be highlighted. In the case of the SIGI, we are aggregating across dimensions and not over individuals. Moreover, in contrast to the income case, a lower value of  $x_i$  is preferred, and the normalization achieved when dividing by the poverty line  $z$  is not necessary as  $0 \leq x_i \leq 1$ ,  $i = 1, \dots, n$ .

The SIGI fulfills several properties. For a formal presentation of the properties and the proofs, see the Appendix to Essay 1.

- *Support and range*: The value of the index can be computed for any values of the subindices, and it is always between 0 and 1.
- *Anonymity*: Neither the name of the country nor the name of the subindex have an impact on the value of the index.
- *Unanimity or Pareto Optimality*: If a country has values for every subindex that are lower than or equal to those of another country, then the index value for the first country is lower than or equal to the one for the second country.
- *Monotonicity*: If one country has a lower value for the index than a second country, and a third country has the same values for the subindices as the first country, except for one subindex which is lower, then the third country has a lower index value than the second country.
- *Penalization of dispersion*: For two countries with the same average value of the subindices, the country with the lowest dispersion of the subindices gets a lower value for the index.
- *Compensation*: Although the SIGI is not conceived for changes over time this property is more intuitively understood in the following way. Assume there are only two subindices and that a country has the same level of inequality in the two subindices. If the country experiences an increase in inequality by a given amount on one subindex, then it can only have the same value of the index as before, if there is a decrease in inequality on the other subindex that is higher in absolute value than the increase in the first subindex.



To highlight the effects of partial compensation as compared to total compensation we computed the statistical association between the SIGI and a simple arithmetic average of the five subindices that allows for total compensation and compared the country rankings of both measures.<sup>11</sup> The Pearson correlation coefficient between the SIGI and the simple arithmetic average of the five subindices shows a high and statistically significant correlation between both measures (Table 1.8). However, when we compare the ranks of the SIGI with those obtained using a simple arithmetic average of the five subindices in Table 1.9, we observe that there are noticeable differences in the rankings of the 102 included countries. Examples are China and Nepal. China ranks in position 55 using the simple average, but worsens to place 83 in the SIGI ranking. Nepal has place 84 considering the simple average, and improves to rank 65 using the SIGI. For China, this is due to the high value on the subindex *Son preference*, which in the SIGI case cannot be fully compensated with relatively low values for the other subindices. For Nepal we observe the opposite case as all subindices have values reflecting moderate inequality.

## 1.5 Results

### 1.5.1 Country Rankings and Regional Patterns

The subindices are computed for countries that have no missing values on the relevant input variables. In the case of the SIGI only countries that have values for every subindex are considered. The results for the SIGI and its five subindices are presented in Table 1.10. Among the 102 countries considered by the SIGI, Paraguay, Croatia, Kazakhstan, Argentina and Costa Rica have the lowest levels of gender inequality related to social institutions. Sudan is the country that occupies the last position, followed by Afghanistan, Sierra Leone, Mali and Yemen, which means that gender inequality in social institutions is a major problem there.

Rankings according to the subindices are as follows. For *Family code* 112 countries can be ranked. Best performers are China, Jamaica, Croatia, Belarus and Kazakhstan. Worst performers are Mali, Chad, Afghanistan, Mozambique and Zambia. In the dimension *Civil liberties* 123 countries are ranked. Among them 83 share place 1 in the ranking. Sudan, Saudi Arabia, Afghanistan, Yemen and Iran occupy the last five positions of high inequality. 114 countries can be compared with the subindex *Physical Integrity*. Hong Kong, Bangladesh,

---

<sup>11</sup>We cannot compare the SIGI with the results of the non-compensatory index as proposed by Munda and Nardo (2005a,b). The algorithm used for calculating non-compensatory indices compares pairwise each country for each subindex. However, as our dataset includes many countries with equal values on several subindices, the numerical algorithm cannot provide a ranking.



Chinese Taipei, Ecuador, El Salvador, Paraguay and Philippines are at the top of the ranking while Mali, Somalia, Sudan, Egypt and Sierra Leone are at the bottom. In the dimension *Son preference* 88 out of 123 countries rank at the top as they do not have problems with missing women. The countries that rank worst are China, Afghanistan, Papua New Guinea, Pakistan, India and Bhutan. Finally, 122 countries are ranked with the subindex *Ownership rights*. 42 countries share position 1 as they have no inequality in this dimension. On the other hand the four worst performing countries are Sudan, Sierra Leone, Chad and the Democratic Republic of Congo.

To find out whether apparent regional patterns in social institutions related to gender inequality are systematic, we divide the countries in quintiles following the scores of the SIGI and its subindices (Table 1.11). The first quintile includes countries with lowest inequality, and the fifth quintile countries with highest inequality.

For the SIGI, no country of Europe and Central Asia (ECA) or Latin America and the Caribbean (LAC) is found in the two quintiles reflecting social institutions related to high gender inequality. In contrast, most countries in South Asia (SA), Sub-Saharan Africa (SSA), and Middle East and North Africa (MENA) rank in these two quintiles. It is interesting to note that in the most problematic regions two countries rank in the first two quintiles. These are Mauritius (SSA) and Tunisia (MENA). East Asia and Pacific (EAP) has countries in all five quintiles with Philippines, Thailand, Hong Kong and Singapore in the first quintile and China in the fifth quintile.

Going on with the subindices the patterns are similar to the one of the SIGI. As more information is available for the subindices, the number of countries covered by every subindex is different and higher than for the SIGI. In the following some interesting facts are highlighted, especially countries whose scores are different than the average in the region.

- *Family code*: No country in ECA, LAC or EAP shows high inequality. SA, MENA and SSA remain problematic with countries with social institutions related to high gender inequality. Exceptions are Bhutan in SA, Mauritius in SSA, and Tunisia and Israel in MENA.
- *Civil liberties*: Only three groups of countries using the quintile analysis can be generated with the first group including the first three quintiles. In SSA over one-half of the countries are now in the first group. Also in MENA there are some countries with good scores (Israel, Morocco and Tunisia). No country in SA is found in the first three quintiles of low and moderate inequality.

- *Physical integrity*: Most problematic regions are SSA and MENA. Exceptions in these regions are Botswana, Mauritius, South Africa and Tanzania (SSA), and Morocco and Tunisia (MENA).
- *Son preference*: Again only three groups of countries can be built by quintile analysis, with the first group including the first three quintiles. As in the case of Civil liberties most of the countries in SSA do not show problems. Missing women is mainly an issue in SA and MENA. But in both regions there are countries that rank in the first group. These are Sri Lanka in SA, and Israel, Lebanon and Occupied Palestinian Territory in MENA.
- *Ownership rights*: Most problematic regions are SA, SSA and MENA. Nevertheless, there are cases in these regions that rank in the first quintile. These are Egypt, Israel, Kuwait and Tunisia (MENA), Bhutan (SA), and Eritrea and Mauritius (SSA).

### 1.5.2 Simple Correlation with other Gender-related Indices

The SIGI is an important measure to understand gender inequality as it measures institutions that influence the basic functioning of society and explain gender inequality in outcomes. From this perspective, the SIGI has an added value to other gender-related measures irrespective from an empirical redundancy perspective, i.e. whether it provides additional information as compared to other measures.

Nevertheless, one can check whether the index is empirically redundant with an analysis of the statistical association between the SIGI and other well-known gender-related indices. Relying on McGillivray and White (1993) we use a correlation coefficient of 0.80 in absolute value as the threshold to separate redundancy from non-redundancy.

We calculated Pearson correlation coefficient and Kendall tau b as a measure of rank correlation between the SIGI and each of the following indices: the Gender-related Development Index (GDI) and the Gender Empowerment Measure (GEM) from United Nations Development Programme (2006), the Global Gender Gap Index (GGG) from Hausmann et al. (2007) and the Women's Social Rights Index (WOSOC).<sup>12</sup> As the GDI and the GEM have been criticized in the literature (e.g. Klasen, 2006; Schüler, 2006), we also do the analysis for two alternative measures, the Gender Gap Index Capped (GGI) and a revised Gender Empowerment Measure (GEM2) based on income shares proposed by Klasen and Schüler (2009).<sup>13</sup> For all the indices considered both measures of statistical association are lower

<sup>12</sup>Data obtained from <http://ciri.binghamton.edu/>.

<sup>13</sup>The Gender Gap Index Capped (GGI) is a geometric mean of the ratios of female to male achievements in the dimensions health, education and labor force participation. Capped means that every component is capped

than 0.80 in absolute value and statistically significant. We conclude that the SIGI is related to these gender measures but is non-redundant. These results as well as the comparison of the country rankings of the SIGI and these other measures can be found in Tables 1.12 and 1.13.

### 1.5.3 Regression Analysis

The SIGI is aimed to measure the institutional basis of gender inequality. To explore whether the SIGI is associated with gender inequality in outcomes we use linear regressions with two well-known measures as dependent variables and the SIGI as regressor. The first is the Global Gender Gap Index (GGG) that captures gaps in outcome variables related to basic rights such as health, economic participation and political empowerment. The second measure is the ratio of the Gender-Related Development Index (GDI) to the Human Development Index (HDI) as composite measure of gender inequality in the dimensions health, education and income.<sup>14</sup> In both regressions we control for the level of economic development using the log of per capita GDP in constant prices (US\$, PPP, base year: 2005) (World Bank, 2008); for religion using a Muslim majority and a Christian majority dummy, the left-out category being countries that have neither a majority of Muslim nor a majority of Christian population (Central Intelligence Agency, 2009); and for geography and other unexplained heterogeneity that might go together with region using region dummies, the left-out category being Sub-Saharan Africa. As the number of observations is lower than 100, we use HC3 robust standard errors proposed by Davidson and MacKinnon (1993) to account for possible heteroscedasticity in our data.

The regression results are shown in Table 1.14. When GGG is used as the dependent variable, 73 countries are included in the sample and the adjusted coefficient of determination  $R^2$  is 0.62. The SIGI is negatively associated with GGG and significant at the 1% level. In the second regression the ratio of GDI to HDI is the dependent variable. The sample consists of 79 countries and the adjusted  $R^2$  is 0.44. The SIGI is again negatively associated with the

---

at one before calculating the geometric mean. This is necessary as a better relative performance of women, e.g. in the dimension health can be due to a risky behavior of men that should not be rewarded. GGI can be more directly interpreted as a measure of gender inequality while the GDI measures human development penalizing gender inequality. The GEM has three components, political representation, representation in senior positions in the economy, and power over economic resources. The most problematic component is power over economic resources proxied by earned incomes. This component measures female and male earned incomes using income levels adjusted by gender gaps but not the gender gaps themselves. The revised version GEM2 uses income shares of males and females.

<sup>14</sup>As the GDI is not a measure of gender inequality, UNDP recommends using the ratio of GDI to HDI as a proxy of gender inequality ([http://hdr.undp.org/en/statistics/indices/gdi\\_gem/](http://hdr.undp.org/en/statistics/indices/gdi_gem/)).

response variable and this association is statistically significant at the 1% level. The results suggest that gender inequality in well-being and empowerment is strongly associated with social institutions that shape gender roles.

Even if we include control variables in the regressions we cannot rule out omitted variable bias, but as we consider that social institutions related to gender inequality are relatively stable and long-lasting, we consider that endogeneity does not pose a major problem. To check that our findings are not driven by observations that have large residuals and/or high leverage, we also run robust regressions obtaining similar results.<sup>15</sup>

## 1.6 Conclusion

In this paper we present composite indices that offer a new way to approach gender inequality that has been neglected in the literature and by other gender measures that focus mainly on well-being and agency. Instead of measuring gender inequality in education, health, economic or political participation and other dimensions, the proposed measures proxy the underlying social institutions that are mirrored by societal practices and legal norms that might produce inequalities between women and men in developing countries.

Based on 12 variables of the OECD Gender, Institutions and Development (GID) Database (Morrison and Jütting, 2005; Jütting et al., 2008) we construct five subindices each capturing one dimension of social institutions related to gender inequality: *Family code*, *Civil liberties*, *Physical integrity*, *Son preference* and *Ownership rights*. The Social Institutions and Gender Index (SIGI) combines the subindices into a multidimensional index of deprivation of women caused by social institutions related to gender inequality. With these measures over 100 developing countries can be compared and ranked.

When constructing composite indices one is always confronted with decisions and trade-offs concerning for example the choice and treatment of the variables included, the weighting scheme and the aggregation method. We try to be transparent in our choices. As the subindices are intended each to proxy one dimension of social institutions, we use the method of polychoric PCA to extract the common element of the included variables (Kolenikov and Angeles, 2009). The methodology for constructing the multidimensional SIGI is based on the assumption that in each dimension deprivation of women increases more than proportionally when inequality increases, and that each dimension should be weighted equally. The formula of the SIGI is inspired by the FGT poverty measures (Foster et al., 1984) and

---

<sup>15</sup>Results are available upon request. The type of robust regression we perform uses iteratively reweighted least squares and is described in Hamilton (1992). A regression is run with ordinary least squares, then case weights based on absolute residuals are calculated, and a new regression is performed using these weights. The iterations continue as long as the maximum change in weights remains above a specified value.

has the advantage of penalizing high inequality in each dimension and only allowing for partial compensation among the five dimensions. We consider that the formula to compute the SIGI is easy to understand and to communicate.

However, some limitations of the subindices and the SIGI must be noted. First, a composite index depends on the quality of the data used as input. Social institutions related to gender inequality are hard to measure and the work accomplished by the OECD Development Centre in building the GID database is an important step forward. It is worthwhile to continue this endeavor and invest more resources in the measurement of social institutions related to gender inequality. This includes data coverage, coding schemes and the refinement of indicators. It would be useful to exploit data available, for example from Demographic and Health Surveys (DHS)<sup>16</sup>, that specifically address the perception that women have of violence against women, and to finance surveys in countries where data is not available.

Secondly, by aggregating variables and subindices, one inevitably loses some information. Figures and rankings according to the SIGI and the subindices should not substitute a careful investigation of the variables from the database. Furthermore, to understand the situation in a given country additional qualitative information could be valuable.

Thirdly, one should keep in mind that OECD countries are not included in our sample, as social institutions related to gender inequality in these countries are not well captured by the 12 variables used for building the composite measures. This does not mean that this phenomenon is not relevant for OECD countries, but that further research is required to develop appropriate measures.

Nonetheless, the SIGI and its subindices offer a new perspective to understand gender inequality. Empirical results show that the SIGI is statistically non-redundant and adds new information to other well-known gender-related measures. The SIGI and the five subindices can help policy-makers to detect in which developing countries and in which dimensions of social institutions problems need to be addressed. For example, according to the SIGI scores, regions with highest inequality are South Asia, Sub-Saharan Africa, and the Middle East and North Africa. The composite measures can be valuable instruments to generate public discussion. Moreover, the SIGI and its subindices have the potential to influence current development thinking as they highlight social institutions that affect overall development. As is shown in the literature (e.g. Klasen, 2002; Klasen and Lamanna, 2009), gender inequality in education negatively affects overall development. Economic research investigating these outcome inequality should consider social institutions related to gender inequality as possible explanatory factors. Results from regression analysis show that the SIGI is related to gender

---

<sup>16</sup>Information is available on the webpage <http://www.measuredhs.com/>.

inequality in well-being and empowerment, even after controlling for region, religion and the level of economic development.

## 1.7 Tables

Table 1.1: Kendall tau b: Dimension Family Code

		<b>Earmarr</b>	<b>Polyg</b>	<b>Parauth</b>	<b>Inher</b>
<b>Earmarr</b>	Kendall tau b	1			
	Number of obs.	112			
	p-Value				
<b>Polyg</b>	Kendall tau b	0.30	1		
	Number of obs.	112	112		
	p-Value	0.0001			
<b>Parauth</b>	Kendall tau b	0.29	0.48	1	
	Number of obs.	112	112	112	
	p-Value	0.0001	0.0000		
<b>Inher</b>	Kendall tau b	0.23	0.60	0.57	1
	Number of obs.	112	112	112	112
	p-Value	0.0020	0.0000	0.0000	

Earmarr: Early marriage, Polyg: Polygamy, Parauth: Parental authority, Inher: inheritance. See section 1.2 for details. The p-values correspond to the null hypothesis that the two variables are independent.

Table 1.2: Kendall tau b: Dimension Civil Liberties

		<b>Obliveil</b>
<b>Freemov</b>	Kendall tau b	0.61
	Number of obs.	123
	p-Value	0.0000

Freemov: Freedom of movement, Obliveil: Freedom of dress. See section 1.2 for details. The p-value correspond to the null hypothesis that two variables are independent.

Table 1.3: Kendall tau b: Dimension Physical Integrity with Missing Women

		<b>Femmut</b>	<b>Vio</b>	<b>Misswom</b>
<b>Femmut</b>	Kendall tau b	1		
	Number of obs.	114		
	p-Value			
<b>Vio</b>	Kendall tau b	0.16	1	
	Number of obs.	114	114	
	p-Value	0.0382		
<b>Misswom</b>	Kendall tau b	-0.10	0.11	1
	Number of obs.	114	114	114
	p-Value	0.2160	0.1634	

Femmut: Female Genital Mutilation, Vio: Violence against women, Misswom: Missing women. See section 1.2 for details. The p-values correspond to the null hypothesis that the two variables are independent.

Table 1.4: Kendall tau b: Dimension Physical Integrity without Missing Women

		<b>Vio</b>
<b>Femmut</b>	Kendall tau b	0.16
	Number of obs.	114
	p-Value	0.0382

Femmut: Female Genital Mutilation, Vio: Violence against women. See section 1.2 for details. The p-value correspond to the null hypothesis that two variables are independent.

Table 1.5: Kendall tau b: Dimension Ownership Rights

		<b>Womland</b>	<b>Womloans</b>	<b>Womprop</b>
<b>Womland</b>	Kendall tau b	1		
	Number of obs.	122		
	p-Value			
<b>Womloans</b>	Kendall tau b	0.59	1	
	Number of obs.	122	122	
	p-Value	0.0000		
<b>Womprop</b>	Kendall tau b	0.64	0.60	1
	Number of obs.	122	122	122

Womland: Women's access to land, Womloans: Women's access to loans, Womprop: Women's access to property other than land. See section 1.2 for details. The p-values correspond to the null hypothesis that the two variables are independent.

Table 1.6: Weights from Polychoric PCA

	<b>Weights</b>
<b>Family code</b>	
Parental authority	0.5212
Inheritance	0.5404
Early marriage	0.3877
Polygamy	0.5348
<b>Civil liberties</b>	
Freedom of movement	0.7071
Freedom of dress	0.7071
<b>Physical integrity</b>	
Female genital mutilation	0.7071
Violence against women	0.7071
<b>Ownership rights</b>	
Women's access to land	0.5811
Women's access to loans	0.5665
Women's access to other property	0.5843



Table 1.7: Kendall tau b between Subindices

		<b>Family code</b>	<b>Civil liberties</b>	<b>Physical integrity</b>	<b>Son preference</b>	<b>Ownership rights</b>
<b>Family code</b>	Kendall tau b	1				
	Number obs.	112				
<b>Civil liberties</b>	Kendall tau b	0.38	1			
	Number obs.	112	123			
	p-value	0.0000				
<b>Physical integrity</b>	Kendall tau b	0.44	0.26	1		
	Number obs.	103	113	114		
	p-value	0.0000	0.0005			
<b>Son preference</b>	Kendall tau b	0.16	0.43	0.03	1	
	Number obs.	112	122	114	123	
	p-value	0.0317	0.0000	0.7220		
<b>Ownership rights</b>	Kendall tau b	0.55	0.30	0.39	0.10	1
	Number obs.	111	121	112	121	122
	p-value	0.0000	0.0001	0.0000	0.181	

Table 1.8: Pearson Correlation Coefficient ( $\rho$ ) between the SIGI and the Simple Average of the Five Subindices

$\rho$	0.96
Number obs.	102
p-value	0.0000

Table 1.9: Comparison of the SIGI and the Simple Average of the Subindices

Country	SIGI		Simple Aver.		Simple Aver. Rank minus SIGI rank
	Ranking	Value	Ranking	Value	
Paraguay	1	0.00248	2	0.03129	1
Croatia	2	0.00333	1	0.02738	-1
Kazakhstan	3	0.00348	3	0.03143	0
Argentina	4	0.00379	4	0.03548	0
Costa Rica	5	0.00709	5	0.05021	0
Russian Federation	6	0.00725	11	0.05381	5
Philippines	7	0.00788	15	0.06032	8
El Salvador	8	0.00826	16	0.06479	8
Ecuador	9	0.00915	18	0.07005	9
Ukraine	10	0.00969	6	0.05138	-4
Mauritius	11	0.00976	7	0.05219	-4
Moldova	12	0.00980	8	0.05267	-4
Bolivia	13	0.00983	9	0.05300	-4
Uruguay	14	0.00992	10	0.05381	-4
Venezuela, RB	15	0.01043	13	0.05786	-2
Thailand	16	0.01068	17	0.06530	1
Peru	17	0.01213	14	0.05866	-3
Colombia	18	0.01273	24	0.08289	6
Belarus	19	0.01339	12	0.05638	-7
Hong Kong, China	20	0.01465	19	0.07076	-1
Singapore	21	0.01526	20	0.07146	-1
Cuba	22	0.01603	22	0.07502	0
Macedonia, FYR	23	0.01787	23	0.08186	0
Brazil	24	0.01880	21	0.07353	-3
Tunisia	25	0.01906	29	0.10123	4
Chile	26	0.01951	31	0.10653	5
Cambodia	27	0.02202	27	0.08862	0
Nicaragua	28	0.02251	32	0.11175	4
Trinidad and Tobago	29	0.02288	34	0.11434	5
Kyrgyz Republic	30	0.02924	36	0.12716	6
Viet Nam	31	0.03006	25	0.08375	-6
Armenia	32	0.03012	26	0.08456	-6
Georgia	33	0.03069	28	0.09024	-5
Guatemala	34	0.03193	35	0.12440	1
Tajikistan	35	0.03262	37	0.13772	2
Honduras	36	0.03316	33	0.11225	-3
Azerbaijan	37	0.03395	30	0.10590	-7
Lao PDR	38	0.03577	39	0.14164	1

Continued on next page

Table 1.9 – continued from previous page

Country	SIGI		Simple Aver.		Simple Aver. Rank minus SIGI rank
	Ranking	Value	Ranking	Value	
Mongolia	39	0.03912	43	0.16806	4
Dominican Republic	40	0.03984	40	0.14402	0
Myanmar	41	0.04629	42	0.15532	1
Jamaica	42	0.04843	38	0.13998	-4
Morocco	43	0.05344	45	0.19732	2
Fiji	44	0.05450	41	0.15512	-3
Sri Lanka	45	0.05914	47	0.21069	2
Madagascar	46	0.06958	44	0.19385	-2
Namibia	47	0.07502	49	0.24188	2
Botswana	48	0.08102	46	0.20277	-2
South Africa	49	0.08677	53	0.25654	4
Burundi	50	0.10691	52	0.24881	2
Albania	51	0.10720	58	0.27159	7
Senegal	52	0.11041	50	0.24241	-2
Tanzania	53	0.11244	51	0.24452	-2
Ghana	54	0.11269	54	0.25684	0
Indonesia	55	0.12776	57	0.26929	2
Eritrea	56	0.13645	48	0.22890	-8
Kenya	57	0.13704	56	0.26730	-1
Cote d'Ivoire	58	0.13712	59	0.28623	1
Syrian Arab Republic	59	0.13811	74	0.36194	15
Malawi	60	0.14323	65	0.33096	5
Mauritania	61	0.14970	68	0.33362	7
Swaziland	62	0.15655	70	0.34562	8
Burkina Faso	63	0.16161	60	0.30306	-3
Bhutan	64	0.16251	63	0.31967	-1
Nepal	65	0.16723	84	0.39738	19
Rwanda	66	0.16859	61	0.30592	-5
Niger	67	0.17559	72	0.35373	5
Equatorial Guinea	68	0.17597	76	0.36767	8
Gambia, The	69	0.17830	62	0.31775	-7
Central African Republic	70	0.18440	67	0.33231	-3
Kuwait	71	0.18602	79	0.37231	8
Zimbabwe	72	0.18700	78	0.36859	6
Uganda	73	0.18718	80	0.37357	7
Benin	74	0.18899	66	0.33197	-8
Algeria	75	0.19024	87	0.41232	12
Bahrain	76	0.19655	89	0.43106	13
Mozambique	77	0.19954	82	0.38088	5
Togo	78	0.20252	69	0.34352	-9

Continued on next page

Table 1.9 – continued from previous page

Country	SIGI		Simple Aver.		Simple Aver. Rank minus SIGI rank
	Ranking	Value	Ranking	Value	
Congo, Dem. Rep.	79	0.20448	64	0.32770	-15
Papua New Guinea	80	0.20936	83	0.38431	3
Cameroon	81	0.21651	85	0.40132	4
Egypt, Arab Rep.	82	0.21766	81	0.37798	-1
China	83	0.21786	55	0.26056	-28
Gabon	84	0.21892	86	0.40386	2
Zambia	85	0.21939	71	0.35261	-14
Nigeria	86	0.21991	92	0.45401	6
Liberia	87	0.22651	75	0.36290	-12
Guinea	88	0.22803	77	0.36782	-11
Ethiopia	89	0.23325	73	0.35590	-16
Bangladesh	90	0.24465	91	0.44911	1
Libya	91	0.26019	94	0.50580	3
United Arab Emirates	92	0.26575	96	0.50826	4
Iraq	93	0.27524	97	0.52298	4
Pakistan	94	0.28324	95	0.50621	1
Iran, Islamic Rep.	95	0.30436	98	0.52525	3
India	96	0.31811	99	0.52951	3
Chad	97	0.32258	93	0.47332	-4
Yemen	98	0.32705	100	0.55679	2
Mali	99	0.33949	88	0.42266	-11
Sierra Leone	100	0.34245	90	0.44886	-10
Afghanistan	101	0.58230	101	0.74613	0
Sudan	102	0.67781	102	0.80051	0

The data are sorted according to the value of the SIGI.

Table 1.10: Ranking according to the SIGI and the Five Subindices

Country	SIGI		Family code		Civil liberties		Physical integrity		Son preference		Ownership rights	
	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value
Paraguay	1	0.00248	19	0.06890	1	0	3	0.08757	1	0	1	0
Croatia	2	0.00333	3	0.00811	1	0	9	0.12878	1	0	1	0
Kazakhstan	3	0.00348	5	0.02837	1	0	9	0.12878	1	0	1	0
Argentina	4	0.00379	13	0.04864	1	0	9	0.12878	1	0	1	0
Costa Rica	5	0.00709	23	0.08106	1	0	15	0.16999	1	0	1	0
Russian Federation	6	0.00725	35	0.14028	1	0	9	0.12878	1	0	1	0
Philippines	7	0.00788	8	0.04053	1	0	3	0.08757	1	0	53	0.17351
El Salvador	8	0.00826	17	0.06485	1	0	3	0.08757	1	0	43	0.17151
Ecuador	9	0.00914	24	0.08917	1	0	3	0.08757	1	0	53	0.17351
Ukraine	10	0.00969	8	0.04053	1	0	23	0.21635	1	0	1	0
Mauritius	11	0.00976	11	0.04458	1	0	23	0.21635	1	0	1	0
Moldova	12	0.00980	12	0.04701	1	0	23	0.21635	1	0	1	0
Bolivia	13	0.00983	13	0.04864	1	0	23	0.21635	1	0	1	0
Uruguay	14	0.00992	15	0.05269	1	0	23	0.21635	1	0	1	0
Venezuela, RB	15	0.01043	21	0.07295	1	0	23	0.21635	1	0	1	0
Thailand	16	0.01068	41	0.15649	1	0	15	0.16999	1	0	1	0
Peru	17	0.01213	15	0.05269	1	0	33	0.24059	1	0	1	0
Colombia	18	0.01273	21	0.07295	1	0	15	0.16999	1	0	43	0.17151
Belarus	19	0.01339	4	0.02432	1	0	34	0.25756	1	0	1	0
Hong Kong, China	20	0.01465	26	0.10380	1	0	1	0	89	0.25	1	0
Singapore	21	0.01526	25	0.09975	1	0	34	0.25756	1	0	1	0
Cuba	22	0.01603	28	0.11754	1	0	34	0.25756	1	0	1	0
Macedonia, FYR	23	0.01787	39	0.15169	1	0	34	0.25756	1	0	1	0

Continued on next page

Table 1.10 – continued from previous page

Country	SIGI		Family code		Civil liberties		Physical integrity		Son preference		Ownership rights	
	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value
Brazil	24	0.01880	19	0.06890	1	0	48	0.29877	1	0	1	0
Tunisia	25	0.01906	32	0.12738	1	0	9	0.12878	89	0.25	1	0
Chile	26	0.01951	34	0.13909	1	0	23	0.21635	1	0	56	0.17723
Cambodia	27	0.02202	38	0.14433	1	0	48	0.29877	1	0	1	0
Nicaragua	28	0.02251	33	0.12970	1	0	34	0.25756	1	0	43	0.17151
Trinidad & Tobago	29	0.02288	39	0.15169	1	0	15	0.16999	89	0.25	1	0
Kyrgyz Republic	30	0.02924	42	0.15980	1	0	48	0.29877	1	0	56	0.17723
Viet Nam	31	0.03006	6	0.03242	1	0	60	0.38634	1	0	1	0
Armenia	32	0.03012	7	0.03648	1	0	60	0.38634	1	0	1	0
Georgia	33	0.03069	17	0.06485	1	0	60	0.38634	1	0	1	0
Guatemala	34	0.03193	27	0.10538	1	0	54	0.34513	1	0	43	0.17151
Tajikistan	35	0.03262	47	0.25955	1	0	34	0.25756	1	0	43	0.17151
Honduras	36	0.03316	44	0.21610	1	0	54	0.34513	1	0	1	0
Azerbaijan	37	0.03395	37	0.14314	1	0	60	0.38634	1	0	1	0
Lao PDR	38	0.03577	51	0.32034	1	0	23	0.21635	1	0	43	0.17151
Mongolia	39	0.03912	30	0.12001	1	0	48	0.29877	89	0.25	43	0.17151
Dominican Republic	40	0.03984	28	0.11754	1	0	34	0.25756	1	0	58	0.34502
Myanmar	41	0.04629	35	0.14028	1	0	60	0.38634	89	0.25	1	0
Jamaica	42	0.04843	1	0.00405	1	0	54	0.34513	1	0	76	0.35074
Morocco	43	0.05344	48	0.26279	1	0	9	0.12878	89	0.25	58	0.34502
Fiji	44	0.05450	8	0.04053	1	0	60	0.38634	1	0	66	0.34874
Sri Lanka	45	0.05914	46	0.23404	98	0.30069	15	0.16999	1	0	66	0.34874
Madagascar	46	0.06958	70	0.41138	1	0	60	0.38634	1	0	43	0.17151
Namibia	47	0.07502	58	0.35307	1	0	34	0.25756	89	0.25	66	0.34874

Continued on next page

Table 1.10 – continued from previous page

Country	SIGI		Family code		Civil liberties		Physical integrity		Son preference		Ownership rights	
	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value
Botswana	48	0.08102	53	0.32163	1	0	15	0.16999	1	0	79	0.52225
South Africa	49	0.08677	73	0.42326	84	0.29808	23	0.21635	1	0	58	0.34502
Burundi	50	0.10691	57	0.33545	1	0	60	0.38634	1	0	79	0.52225
Albania	51	0.10720	31	0.12288	1	0	60	0.38634	101	0.5	66	0.34874
Senegal	52	0.11041	99	0.60250	1	0	45	0.26455	1	0	58	0.34502
Tanzania	53	0.11244	81	0.49886	1	0	22	0.20151	1	0	79	0.52225
Ghana	54	0.11269	61	0.36621	1	0	80	0.39575	1	0	79	0.52225
Indonesia	55	0.12776	59	0.35405	103	0.59876	79	0.39362	1	0	1	0
Eritrea	56	0.13645	76	0.45538	1	0	106	0.68910	1	0	1	0
Kenya	57	0.13704	63	0.37027	1	0	46	0.28152	1	0	111	0.68473
Cote d'Ivoire	58	0.13712	79	0.49012	1	0	85	0.43455	1	0	77	0.50650
Syrian Arab Republic	59	0.13811	68	0.40269	98	0.30069	34	0.25756	101	0.5	66	0.34874
Malawi	60	0.14323	60	0.36087	84	0.29808	88	0.47362	1	0	79	0.52225
Mauritania	61	0.14970	71	0.42056	98	0.30069	103	0.60183	1	0	58	0.34502
Swaziland	62	0.15655	86	0.52144	84	0.29808	60	0.38634	1	0	79	0.52225
Burkina Faso	63	0.16161	88	0.53939	1	0	104	0.63092	1	0	58	0.34502
Bhutan	64	0.16251	43	0.20513	84	0.29808	54	0.34513	118	0.75	1	0
Nepal	65	0.16723	62	0.36779	84	0.29808	48	0.29877	101	0.5	79	0.52225
Rwanda	66	0.16859	56	0.32974	1	0	91	0.51512	1	0	111	0.68473
Niger	67	0.17559	104	0.64882	1	0	99	0.52482	89	0.25	58	0.34502
Equatorial Guinea	68	0.17597	82	0.50291	84	0.29808	91	0.51512	1	0	79	0.52225
Gambia, The	69	0.17830	103	0.64303	1	0	102	0.59698	1	0	66	0.34874
Central African Rep.	70	0.18440	92	0.55902	1	0	101	0.58029	1	0	79	0.52225
Kuwait	71	0.18602	83	0.50523	103	0.59876	34	0.25756	101	0.5	1	0

Continued on next page

Table 1.10 – continued from previous page

Country	SIGI		Family code		Civil liberties		Physical integrity		Son preference		Ownership rights	
	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value
Zimbabwe	72	0.18700	80	0.49075	84	0.29808	59	0.36937	1	0	111	0.68473
Uganda	73	0.18718	102	0.63697	84	0.29808	81	0.41058	1	0	79	0.52225
Benin	74	0.18899	84	0.50633	1	0	87	0.46877	1	0	111	0.68473
Algeria	75	0.19024	69	0.40501	103	0.59876	60	0.38634	101	0.5	43	0.17151
Bahrain	76	0.19655	52	0.32147	103	0.59876	60	0.38634	101	0.5	66	0.34874
Mozambique	77	0.19954	109	0.69776	84	0.29808	60	0.38634	1	0	79	0.52225
Togo	78	0.20252	96	0.58833	1	0	86	0.44452	1	0	111	0.68473
Congo, Dem. Rep.	79	0.20448	66	0.39038	1	0	81	0.41058	1	0	119	0.83752
Papua New Guinea	80	0.20936	50	0.27697	1	0	60	0.38634	118	0.75	78	0.50825
Cameroon	81	0.21651	89	0.54344	84	0.29808	90	0.48332	1	0	109	0.68175
Egypt, Arab Rep.	82	0.21766	49	0.26647	98	0.30069	111	0.82273	101	0.5	1	0
China	83	0.21786	1	0.00405	1	0	48	0.29877	122	1	1	0
Gabon	84	0.21892	107	0.68387	84	0.29808	91	0.51512	1	0	79	0.52225
Zambia	85	0.21939	108	0.69197	1	0	60	0.38634	1	0	111	0.68473
Nigeria	86	0.21991	71	0.42056	103	0.59876	89	0.47847	89	0.25	79	0.52225
Liberia	87	0.22651	87	0.53470	1	0	107	0.75756	1	0	79	0.52225
Guinea	88	0.22803	105	0.67140	1	0	105	0.64546	1	0	79	0.52225
Ethiopia	89	0.23325	55	0.32726	1	0	109	0.77424	1	0	108	0.67801
Bangladesh	90	0.24465	95	0.58334	103	0.59876	2	0.04121	101	0.5	79	0.52225
Libya	91	0.26019	67	0.39285	103	0.59876	91	0.51512	101	0.5	79	0.52225
Unit. Arab Emirates	92	0.26575	93	0.56197	103	0.59876	100	0.53180	101	0.5	66	0.34874
Iraq	93	0.27524	77	0.47391	103	0.59876	98	0.51997	101	0.5	79	0.52225
Pakistan	94	0.28324	64	0.37821	103	0.59876	47	0.28180	118	0.75	79	0.52225
Iran, Islamic Rep.	95	0.30436	91	0.55792	119	0.78099	91	0.51512	89	0.25	79	0.52225

Continued on next page



Table 1.10 – continued from previous page

Country	SIGI		Family code		Civil liberties		Physical integrity		Son preference		Ownership rights	
	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value
India	96	0.31811	100	0.60655	103	0.59876	15	0.16999	118	0.75	79	0.52225
Chad	97	0.32258	111	0.79330	98	0.30069	84	0.43212	1	0	120	0.84049
Yemen	98	0.32705	97	0.59439	119	0.78099	60	0.38634	101	0.5	79	0.52225
Mali	99	0.33949	112	0.79735	1	0	114	0.97091	1	0	58	0.34502
Sierra Leone	100	0.34245	98	0.60159	1	0	110	0.79849	1	0	121	0.84424
Afghanistan	101	0.58230	110	0.71598	121	0.81777	91	0.51512	122	1	109	0.68175
Sudan	102	0.67781	106	0.67981	122	1	111	0.82273	101	0.5	122	1
Angola		NA	89	0.54344	1	0		NA	89	0.25	79	0.52225
Bosnia & Herzegovina		NA		NA	1	0	34	0.25756	1	0	1	0
Chinese Taipei		NA		NA	1	0	3	0.08757	101	0.5	1	0
Congo, Rep.		NA	101	0.62450	1	0		NA	1	0	79	0.52225
Guinea-Bissau		NA		NA		NA	107	0.75756	1	0	111	0.68473
Haiti		NA	65	0.37837	1	0	54	0.34513	1	0		NA
Israel		NA	45	0.22712	1	0		NA	1	0	1	0
Jordan		NA	85	0.51739	103	0.59876		NA	101	0.5	79	0.52225
Korea, Dem. Rep.		NA		NA	84	0.29808	91	0.51512	1	0	1	0
Lebanon		NA		NA	103	0.59876	60	0.38634	1	0	53	0.17351
Lesotho		NA	94	0.57149	84	0.29808		NA	1	0	79	0.52225
Malaysia		NA	53	0.32163	103	0.59876		NA	1	0	1	0
Occup. Palestinian Terr.		NA	78	0.48607	103	0.59876		NA	1	0	66	0.34874
Oman		NA	74	0.45364	84	0.29808		NA	101	0.5	66	0.34874
Panama		NA		NA	1	0	8	0.11181	1	0	1	0
Puerto Rico		NA		NA	1	0	23	0.21635	1	0		NA
Saudi Arabia		NA	74	0.45364	122	1		NA	101	0.5	79	0.52225

Continued on next page

Table 1.10 – continued from previous page

Country	SIGI		Family code		Civil liberties		Physical integrity		Son preference		Ownership rights	
	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value
Serbia & Montenegro		NA		NA	1	0		NA		NA	43	0.17151
Somalia		NA		NA	103	0.59876	113	0.84213	1	0	111	0.68473
Timor-Leste		NA		NA	1	0	83	0.42755	89	0.25	79	0.52225
Turkmenistan		NA		NA	1	0	60	0.38634	1	0	79	0.52225
Uzbekistan		NA		NA	1	0	60	0.38634	1	0	1	0

Table 1.11: Regional Pattern of the Composite Index and Subindices

	ECA	LAC	EAP	SA	SSA	MENA	Total
<b>SIGI</b>							
Quintile 1	6	10	4	0	1	0	21
Quintile 2	6	8	5	0	0	1	20
Quintile 3	1	1	2	1	14	2	21
Quintile 4	0	0	1	2	13	4	20
Quintile 5	0	0	1	4	10	5	20
<b>Total</b>	<b>13</b>	<b>19</b>	<b>13</b>	<b>7</b>	<b>38</b>	<b>12</b>	<b>102</b>
<b>Family Code</b>							
Quintile 1	7	11	4	0	1	0	23
Quintile 2	5	8	6	1	0	2	22
Quintile 3	1	1	4	3	9	5	23
Quintile 4	0	0	0	0	15	7	22
Quintile 5	0	0	0	3	16	3	22
<b>Total</b>	<b>13</b>	<b>20</b>	<b>14</b>	<b>7</b>	<b>41</b>	<b>17</b>	<b>112</b>
<b>Civil Liberties</b>							
Quintile 1, 2, 3	17	22	14	0	27	3	83
Quintile 4	0	0	1	3	12	3	19
Quintile 5	0	0	2	4	3	12	21
<b>Total</b>	<b>17</b>	<b>22</b>	<b>17</b>	<b>7</b>	<b>42</b>	<b>18</b>	<b>123</b>
<b>Physical Integrity</b>							
Quintile 1	5	13	5	3	4	2	32
Quintile 2	4	4	1	0	3	2	14
Quintile 3	7	5	7	3	6	4	32
Quintile 4	0	0	3	1	13	2	19
Quintile 5	0	0	0	0	14	3	17
<b>Total</b>	<b>16</b>	<b>22</b>	<b>16</b>	<b>7</b>	<b>40</b>	<b>13</b>	<b>114</b>
<b>Missing Women</b>							
Quintile 1, 2, 3	15	21	10	1	38	3	88
Quintile 4	0	1	4	0	4	3	12
Quintile 5	1	0	3	6	1	12	23
<b>Total</b>	<b>16</b>	<b>22</b>	<b>17</b>	<b>7</b>	<b>43</b>	<b>18</b>	<b>123</b>
<b>Ownership Rights</b>							
Quintile 1	12	12	11	1	2	4	42
Quintile 2	2	4	2	0	1	1	10
Quintile 3	2	3	2	1	8	7	23
Quintile 4	1	1	2	4	18	6	32
Quintile 5	0	0	0	1	14	0	15
<b>Total</b>	<b>17</b>	<b>20</b>	<b>17</b>	<b>7</b>	<b>43</b>	<b>18</b>	<b>122</b>

ECA: Europe and Central Asia, LAC: Latin America and the Caribbean, EAP: East Asia and Pacific, SA: South Asia, SSA: Sub-Saharan Africa, MENA: Middle East and North Africa.

Table 1.12: Statistical Association between the SIGI and other Gender-related Measures

<b>GDI</b>	Kendall tau b	-0.50	Pearson Corr. Coeff.	-0.59
Number obs. 79	p-value	0.0000	p-value	0.0000
<b>GGI (capped)</b>	Kendall tau b	-0.51	Pearson Corr. Coeff.	-0.72
Number obs. 85	p-value	0.0000	p-value	0.0000
<b>GEM</b>	Kendall tau b	-0.43	Pearson Corr. Coeff.	-0.70
Number obs. 33	p-value	0.0005	p-value	0.0000
<b>GEM (revised)</b>	Kendall tau b	-0.44	Pearson Corr. Coeff.	-0.75
Number obs. 33	p-value	0.0003	p-value	0.0000
<b>GGG</b>	Kendall tau b	-0.47	Pearson Corr. Coeff.	-0.73
Number obs. 73	p-value	0.0000	p-value	0.0000
<b>WOSOC</b>	Kendall tau b	-0.49	Pearson Corr. Coeff.	-0.53
Number obs. 99	p-value	0.0000	p-value	0.0000

Data for the Gender-related development Index (GDI) and the Gender Empowerment Measure (GEM) are from United Nations Development Programme (2006) and are based on the year 2004. The Gender Gap Index (GGI) capped and the revised Gender Empowerment Measure (GEM revised) are taken from Klasen and Schüler (2009) based on the year 2004. Data for the Global Gender Gap Index (GGG) are from Hausmann et al. (2007). The Women's Social Rights Index (WOSOC) data correspond to the year 2007 and are obtained from <http://ciri.binghamton.edu/>. The p-values correspond to the null hypothesis that the SIGI and the corresponding measure are independent.

Table 1.13: Comparison of Ranks: the SIGI and other Gender-related Indices

Country	SIGI	GDI	GGI (capped)	GEM	GEM (revised)	GGG	WOSOC
Paraguay	1					32	19
Croatia	2	6	16	6	7	3	19
Kazakhstan	3	18	1			10	19
Argentina	4	2	21	2	3	11	3
Costa Rica	5	7	40	3	2	8	3
Russian Federation	6	10	6	22	22	18	19
Philippines	7	22	30	10	8	1	19
El Salvador	8	29	35	13	14	20	19
Ecuador	9			14	11	17	19
Ukraine	10	19	7	23	23	25	19
Mauritius	11	12	46			44	3
Moldova	12						
Bolivia	13	35	24	19	15	41	3
Uruguay	14	5	17	15	17	39	19
Venezuela, RB	15	17	23	11	13	24	
Thailand	16	16	8	20	18	22	19
Peru	17	23	24	8	6	37	3
Colombia	18	15	11	16	16	7	3
Belarus	19	11	3			6	3
Hong Kong, China	20						
Singapore	21			1	11	38	19
Cuba	22		37			5	1
Macedonia, FYR	23	13	32	9	9	13	19
Brazil	24	14	20	20	19	36	3
Tunisia	25	26	72			55	64
Chile	26	3	44	16	20	45	3
Cambodia	27	45	10	28	26	52	3
Nicaragua	28	37	56			49	19
Trinidad and Tobago	29	9	33	4	5	19	1
Kyrgyz Republic	30	34	11			33	19
Viet Nam	31	31	2			15	19
Armenia	32	20	4			34	19
Georgia	33			24	24	30	19
Guatemala	34	39	64			58	19
Tajikistan	35	40	19			40	19
Honduras	36	38	36	12	10	31	19
Azerbaijan	37	28	4			26	19

Continued on next page

Table 1.13 – continued from previous page

Country	SIGI	GDI	GGI		GEM	GGG	WOSOC	
			(capped)	(revised)				
Lao PDR	38	47		45			3	
Mongolia	39	36		27	25	25	27	3
Dominican Republic	40	25		38			29	19
Myanmar	41			14				64
Jamaica	42	30		18			14	3
Morocco	43							19
Fiji	44							3
Sri Lanka	45	24		51	29	28	2	19
Madagascar	46	53		15			48	19
Namibia	47	43		33	5	4	9	19
Botswana	48	46		59	18	21	23	64
South Africa	49	41		42			4	19
Burundi	50	72		24				64
Albania	51							19
Senegal	52							64
Tanzania	53	66		27	7	1	12	19
Ghana	54	48		27			28	19
Indonesia	55	32		39			42	19
Eritrea	56							19
Kenya	57	57		42			43	64
Cote d'Ivoire	58	68		80				64
Syrian Arab Republic	59	33		63			56	64
Malawi	60	70		41			46	19
Mauritania	61	60		48			60	64
Swaziland	62	59		82				64
Burkina Faso	63	76		50			66	64
Bhutan	64							3
Nepal	65	51		61			70	64
Rwanda	66	63		9				3
Niger	67	79		78				19
Equatorial Guinea	68	42		62				19
Gambia, The	69						50	19
Central African Republic	70	75		67				19
Kuwait	71	1		48			51	64
Zimbabwe	72	58		57			47	19
Uganda	73	54		31			21	19
Benin	74	67		73			69	64
Algeria	75							64
Bahrain	76	4		76			64	64
Mozambique	77	71		47			16	64

Continued on next page

Table 1.13 – continued from previous page

Country	SIGI	GDI	GEM		GGG	WOSOC	
			GGI (capped)	GEM (revised)			
Togo	78	61	70			64	
Congo, Dem. Rep.	79	73	60			64	
Papua New Guinea	80	50	22			19	
Cameroon	81	55	54		65	64	
Egypt, Arab Rep.	82			32	31	68	
China	83	20	13		35	64	
Gabon	84					64	
Zambia	85	69	64		54	64	
Nigeria	86	64	66		59	64	
Liberia	87		68			19	
Guinea	88	65	58			19	
Ethiopia	89				62	64	
Bangladesh	90	49	52	27	27	53	
Libya	91		69			64	
United Arab Emirates	92	8	74	30	32	57	
Iraq	93		84			64	
Pakistan	94	51	81	26	28	71	
Iran, Islamic Rep.	95	27	54	31	30	67	
India	96	44	77		63	19	
Chad	97	74	75		72	64	
Yemen	98	62	83	33	33	73	
Mali	99	77	53		61	19	
Sierra Leone	100	78	71			64	
Afghanistan	101		85			19	
Sudan	102	56	79			64	
Number of obs.	102	79	85	33	33	73	99

Data for the Gender-related development Index (GDI) and the Gender Empowerment Measure (GEM) are from United Nations Development Programme (2006) and are based on the year 2004. The Gender Gap Index (GGI) capped and the revised Gender Empowerment Measure (GEM revised) are taken from Klasen and Schüller (2009) based on the year 2004. Data for the Global Gender Gap Index (GGG) are from Hausmann et al. (2007). The Women's Social Rights Index (WOSOC) data correspond to the year 2007 and are obtained from <http://ciri.binghamton.edu/>.

Table 1.14: Linear Regression with Dependent Variables GGG and Ratio GDI to HDI

	<b>GGG</b>	<b>Ratio of GDI to HDI</b>
	coef/se	coef/se
<b>SIGI</b>	-0.284*** (0.089)	-0.054*** (0.017)
log GDP	0.014* (0.008)	0.004 (0.003)
SA	-0.006 (0.032)	-0.001 (0.008)
ECA	-0.012 (0.017)	0.007 (0.005)
LAC	-0.040** (0.017)	-0.000 (0.005)
MENA	-0.043 (0.028)	0.001 (0.011)
EAP	0.005 (0.022)	0.010** (0.005)
Muslim	-0.001 (0.018)	-0.002 (0.006)
Christian	0.026 (0.017)	0.002 (0.004)
constant	0.570*** (0.063)	0.960*** (0.020)
Number of obs.	73	79
Adjusted R2	0.617	0.438
Prob F	0.000	0.000

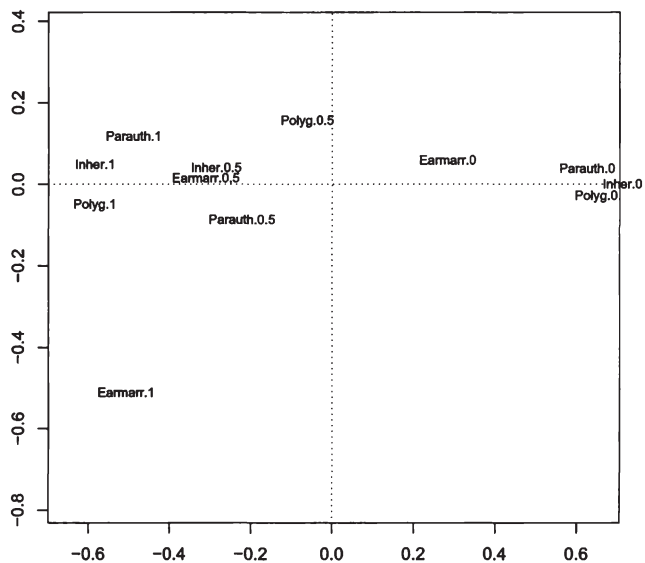
HC3 robust standard errors in brackets.

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1



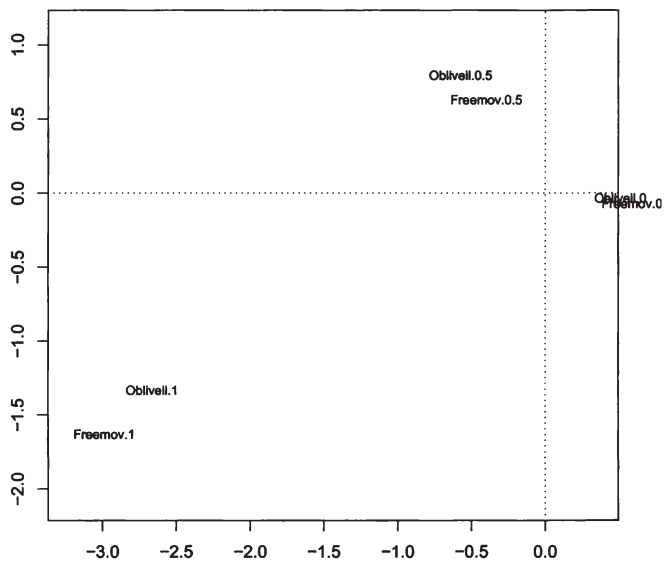
## 1.8 Figures

Figure 1.1: MJCA for the Dimension Family Code



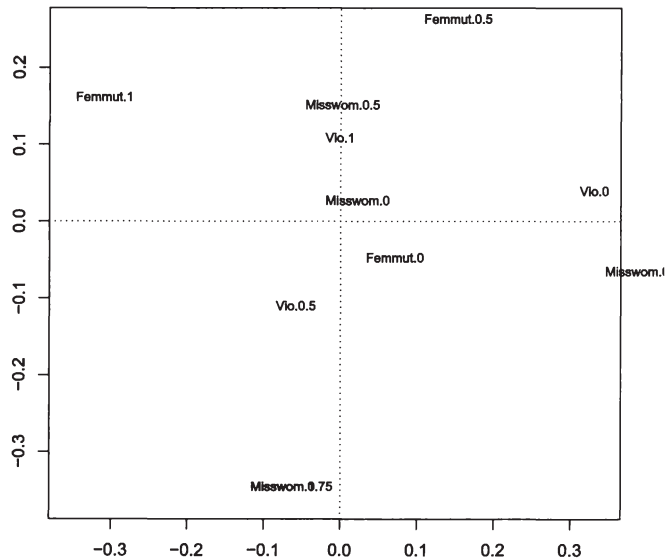
Earmarr stands for the variables Early marriage, Polyg for Polygamy, Parauth is the variable Parental authority and Inher is the variable inheritance. See section 1.2.

Figure 1.2: MJCA for the Dimension Civil Liberties



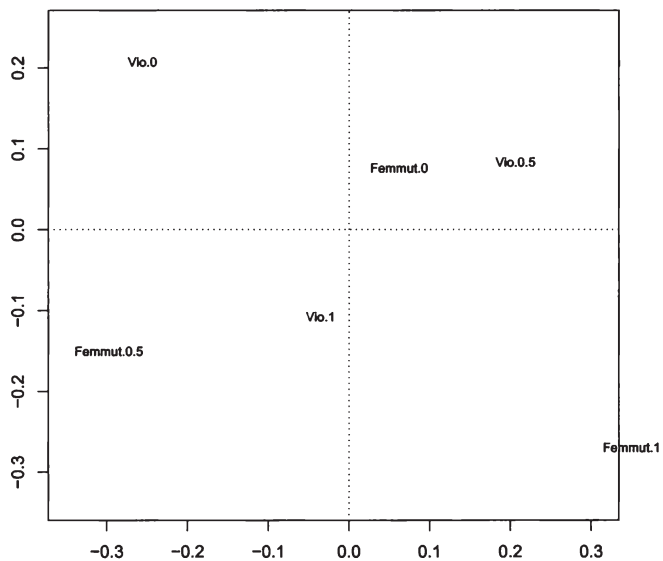
Freemov stands for the variable Freedom of movement. Oblivell is the variable Freedom of dress. See section 1.2.

Figure 1.3: MJCA for the Dimension Physical Integrity with Missing Women



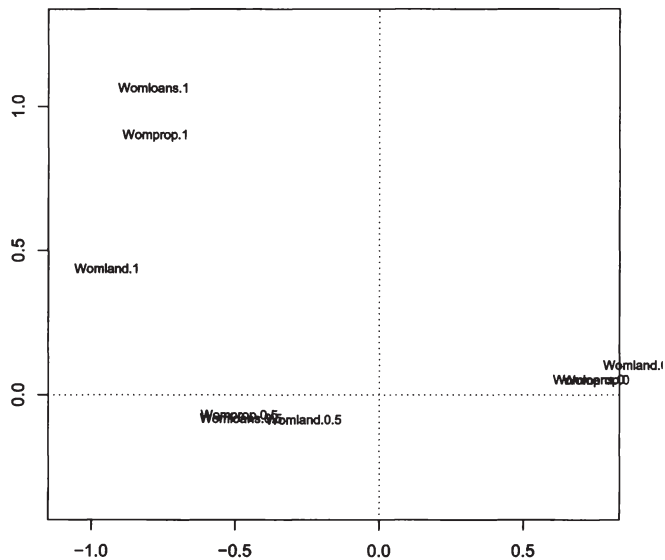
Femmut stands for the variable Female Genital Mutilation, Vio for Violence against women and Misswom is the variable Missing women. See section 1.2.

Figure 1.4: MJCA for the Dimension Physical Integrity without Missing Women



Femmut stands for the variable Female Genital Mutilation and Vio for Violence against women. See section 1.2.

Figure 1.5: MJCA for the Dimension Ownership Rights



Womland stands for the variable Women's access to land. womloan is the variable Women's access to loans and Womprop is the variable Women's access to property other than land. See section 1.2.



## Essay 2

# Why we should all care about social institutions related to gender inequality

### Abstract

Institutions are a major factor explaining development outcomes. This study focuses on social institutions related to gender inequality understood as long-lasting norms, values and codes of conduct that shape gender roles, and presents evidence on why they matter for development. We derive hypotheses from existing theories and empirically test them at the cross-country level with linear regressions using the newly created Social Institutions and Gender Index (SIGI) and its subindices as measures for social institutions. We find that apart from geography, political system, religion, and the level of economic development, one has to consider social institutions related to gender inequality to better account for differences in development. Our results show that social institutions that deprive women of their autonomy and bargaining power in the household, or that increase the private costs and reduce the private returns to investments into girls, are associated with lower female education, higher fertility rates and higher child mortality. Moreover, social institutions related to gender inequality are negatively associated with governance measured as ‘rule of law’ and ‘voice and accountability’.

## 2.1 Introduction

Institutions are a major factor explaining development outcomes. They guide human behavior and shape human interaction (North, 1990). Institutions are humanly devised to reduce uncertainty and transaction cost, they are rooted in culture and history and sometimes they are taken-for-granted and become beliefs (Hall and Taylor, 1996; De Soysa and Jütting, 2007). Our study centers on a special type of institutions and their explanatory value for development outcomes: social institutions related to gender inequality.

It is an established fact that gender inequalities come at a cost. Besides the consequences that the affected women experience because they are deprived of their basic freedoms (Sen, 1999), gender inequalities affect the whole society. They can lead to ill-health, low human capital, bad governance and lower economic growth (e.g. World Bank, 2001; Klasen, 2002). Gender inequalities can be observed in outcomes like education, health and economic and political participation, but they are rooted in gender roles that evolve from institutions that shape everyday life and form role models that people try to fulfill and satisfy. We refer to these long-lasting norms, values and codes of conduct as social institutions related to gender inequality.

We investigate the impact of these social institutions related to gender inequality on development outcomes, controlling for relevant determinants such as religion, political system, geography and the level of economic development. As development outcomes we choose indicators from the fields of education, demographics, health and governance. In particular, we use female secondary schooling, fertility rates, child mortality and governance in the form of ‘rule of law’ and ‘voice and accountability’. We choose these indicators as they are related to economic development and allow us to find out whether social institutions related to gender inequality hinder progress in reaching the Millennium Development Goals.<sup>1</sup>

Most of the studies that have a similar research focus are conducted at the household level and proxy social institutions related to gender with measures of the autonomy or status of women (e.g. Abadian, 1996; Hindin, 2000). At the cross-country level data are scarce and therefore only a few studies are available that center on the development impact of gender-relevant social institutions (e.g. Morrison and Jütting, 2005; Jütting et al., 2008).

Using the *Social Institutions and Gender Index* (SIGI) and its five subindices *Family code*, *Civil liberties*, *Physical integrity*, *Son preference* and *Ownership rights* proposed in Essay 1,

---

<sup>1</sup>In particular, goal 3 “Promote gender equality and empower women”, goal 4 “Reduce child mortality” and goal 5 “Improve maternal health” are relevant here, although the other goals can be at least indirectly linked to our chosen indicators.



we investigate whether social institutions related to gender inequality are associated with the chosen development outcomes at the cross-country level.<sup>2</sup>

These indices cover between 102 and 123 developing countries and are built out of twelve variables of the OECD Gender, Institutions and Development Database that proxy social institutions through prevalence rates, indicators of social practices and legal indicators (Morrison and Jütting, 2005; Jütting et al., 2008).<sup>3</sup> The five subindices of the SIGI each measure one dimension of social institutions related to gender inequality.<sup>4</sup> The *Family code* subindex captures institutions that directly influence the decision-making power of women in the household. It is composed of four variables that measure whether women have the right to be the legal guardian of a child during marriage and whether women have custody rights over a child after divorce, whether there are formal inheritance rights for wives, the percentage of girls between 15 and 19 years of age who are/have been married, and the acceptance of polygamy in the population.<sup>5</sup> The *Civil liberties* subindex covers the freedom of social participation of women and combines two variables, freedom of movement and freedom of dress, which measure the level of restrictions women face in moving freely outside their own household, and the extent to which women are obliged to follow a certain dress code in public. The *Physical integrity* dimension comprises two indicators of violence against women, the existence of laws against domestic and sexual violence and the percentage of women who have undergone female genital mutilation. The subindex *Son preference* measures the economic valuation of women and is based on a ‘missing women’ variable that measures an extreme form of preferring boys over girls based on information about the female population that has died as a result of gender inequality. The last subindex *Ownership rights* covers the access of women to several types of property: land, credits and property other than land. The values of the SIGI and of all the subindices are between 0 and 1. The value 0 means no or very low inequality and the value 1 indicates high inequality.

The SIGI combines the five subindices into a multidimensional measure of deprivation of women in a country. The underlying methodology of construction is inspired by the Foster-

<sup>2</sup>As discussed in Essay 1, an alternative measure of social institutions would be the Women’s Social Rights index (WOSOC) of the CIRI Human Rights Data Project (<http://ciri.binghamton.edu/>), which measures from a human rights perspective the type of institutions we are interested in. We prefer to work with the SIGI and its subindices and not with WOSOC as the latter also covers outcomes of these institutions and does not allow one to differentiate between dimensions of social institutions, e.g. between what happens within the family and what happens in public life. Moreover, WOSOC can only take four values, from 0 to 3, which makes it difficult to compare countries as there are many ties in the data.

<sup>3</sup>The data are available at the web-pages <http://www.wikigender.org> and <http://www.oecd.org/dev/gender/gid>.

<sup>4</sup>To extract the common information of the variables used to construct one subindex the method of polychoric principal component analysis is used (Kolenikov and Angeles, 2009).

<sup>5</sup>Countries where this information is not available are assigned scores based on the legality of polygamy.

Greer-Thorbecke poverty measures (Foster et al., 1984). It leads to penalization of high inequality in each dimension and allows for only partial compensation between dimensions. The value of the SIGI is calculated as follows:

$$\begin{aligned} \text{SIGI} &= \frac{1}{5} (\text{Subindex Family Code})^2 + \frac{1}{5} (\text{Subindex Civil Liberties})^2 \\ &+ \frac{1}{5} (\text{Subindex Physical Integrity})^2 + \frac{1}{5} (\text{Subindex Son preference})^2 \\ &+ \frac{1}{5} (\text{Subindex Ownership Rights})^2 \end{aligned} \quad (2.1)$$

The main shortcoming of these indices is that they cover only developing countries. This is due to the fact that the variables used as input do not measure relevant social institutions related to gender inequalities in OECD countries. Further research is required to develop appropriate measures for developed countries. Nevertheless, these social institutions indicators are innovative measures of the social, economic and political valuation of women and add information to other existing measures of gender inequality in well-being and empowerment.<sup>6</sup> The SIGI and its subindices focus on the roots of gender inequality in a society and not on gender inequality in outcomes. The ranking of countries according to the SIGI and its subindices is presented in Table 1.10.

We proceed as follows. First, we look for relevant theories linking – at least implicitly – social institutions related to gender inequality with development outcomes such as health, demographics, education and the governance of a society. We refer to bargaining household models (e.g. Manser and Brown, 1980; McElroy and Horney, 1981; Lundberg and Pollak, 1993) and models considering the costs and returns of children (e.g. Becker, 1981; King and Hill, 1993; Hill and King, 1995) as well as to contributions from several disciplines on governance and democracy. These contributions focus on differences in behavior between men and women, and on women’s movements countervailing power to personal rule (e.g. Swamy et al., 2001; Tripp, 2001). Secondly, we run several linear regressions with the outcome indicators as dependent variables and the SIGI and its subindices as the main explanatory variables. Our results show that social institutions related to gender inequality matter; higher inequality in social institutions is associated with lower development outcomes. In a related paper, Jütting et al. (2010) follow the same econometric procedure we use here and study the impact of the SIGI and its subindices on gender inequality on labor market outcomes.

The rest of the paper consists of 5 sections. In section 2.2 we review existing theory on household decision-making and incorporate social institutions into the models, deriving

<sup>6</sup>Examples are the Gender-Related Development Index (GDI) and the Gender Empowerment Measure (GEM) from United Nations Development Programme (1995), the Global Gender Gap Index from the World Economic Forum (Lopez-Claros and Zahidi, 2005), the Gender Equity Index developed by Social Watch (Social Watch, 2005), and the African Gender Status Index proposed by the Economic Commission for Africa (Economic Commission for Africa, 2004).

hypotheses on their impact on female education, fertility and child mortality. In section 2.3, we formulate hypotheses on the impact of social institutions on ‘rule of law’, and ‘voice and accountability’ based on the literature on governance, democracy and gender. Data is described in section 2.4. The empirical estimation and the results are presented in section 2.5. Section 2.6 concludes.

## 2.2 Social Institutions and Household Decisions

In this section, we review the existing literature about the potentials effects of social institutions related to gender inequality on development outcomes. It is beyond the scope of this study to develop a formal model that incorporates social institutions and specifies the exact functional relationships. Instead, we use the non-unitary approach to the household and the method of Net Present Value which give hints on how social institutions operate at the household level. These approaches provide the necessary micro-foundation for the empirical analysis which can only be conducted at the macro-level because of the available data.

Non-unitary household models show that household decisions are the result of the distribution of bargaining power in the household. Common to the non-unitary models, initiated by Manser and Brown (1980) and McElroy and Horney (1981), is a game-theoretic approach to the household. Husband and wife have their own utility function,  $U^h(c^h)$  for the husband and  $U^w(c^w)$  for the wife, that depend each on the consumption of private goods  $c$ .<sup>7</sup> They bargain over the allocation of resources to maximize their utility. In the case they do not reach agreement they receive a payoff which corresponds to an individual ‘threat point’,  $P^h(S, Z)$  and  $P^w(S, Z)$  which comprises the utilities associated with non-agreement.<sup>8</sup>  $S$  and  $Z$  are defined below. The implication of non-unitary models is that household members do not simply pool resources and that inequality in power may cause inequality in outcomes (Kanbur, 2003; Pollak, 2003, 2007; Lundberg and Pollak, 2008).<sup>9</sup> Empirical evidence supports this (e.g. Thomas, 1997; Schultz, 1990; Haddad and Hoddinott, 1994; Rasul, 2008).

<sup>7</sup>Certainly, there are public goods in the household that both husband and wife consume within the marriage.

<sup>8</sup>The threat point may be external to the marriage. In this case it corresponds to the individual’s utility outside the family in case of divorce, as it is modeled in the divorce threat models of Manser and Brown (1980) and McElroy and Horney (1981). In the separate spheres bargaining models of Lundberg and Pollak (1993) the threat point is internal to the marriage and is the utility associated with a non-cooperative equilibrium within marriage given by traditional gender roles and social norms, where the spouses receive benefits due to the joint consumption of public goods.

<sup>9</sup>Using Nash-Bargaining a solution to these non-unitary models can be found. Husband and wife maximize the Nash product function  $N = [U^h(c^h - P^h(S, Z))][U^w(c^w - P^w(S, Z))]$ , that is subject to a pooled budget constraint. The result is the demand function  $c^i = f^i(p, y, S, Z)$  with  $p$  for prices,  $y$  for total household income and  $i = w, h$  (Lundberg and Pollak, 2008).

If husband and wife have to take decisions about their sons and daughters which will affect the future then time needs to be considered. The method of the Net Present Value (*NPV*) allows to take into account present and future costs and returns to investments. To simplify the illustration we ignore that bargaining takes place and name the decision-maker 'parents'. The maximization of utility in a multi-period model leads parents to consider the costs and returns of their investment in their children (e.g. King and Hill, 1993). This private calculation of parents at period  $t = 0$  can then be represented with the *NPV* of the investment in a child, with  $NPV = \sum_{t=0}^T \frac{R(S,Z)_t - K(S,Z)_t}{(1+r)^t}$  where  $T$  is the number of time periods considered,  $R$  represents the returns,  $K$  the costs of investments in a child, and  $r$  represents the discount rate. Like the threat point  $P$  in the non-unitary models,  $R$  and  $K$  are functions of  $S$  and  $Z$  that will be explained below. If the *NPV* is positive parents decide to invest in a child. Gender inequality in the investments in boys and girls arises if the *NPV* of boys is larger than the one of girls.<sup>10</sup>

Finally, let us explain  $S$  and  $Z$ .  $S$  can be defined as 'extrahousehold environmental parameters' (McElroy, 1990) or 'gender-specific environmental parameters' (Folbre, 1997) that influence the threat point in the non-unitary household models and the *NPV* of a child. We consider that  $S$  can be best described as *social institutions related to gender inequality*.  $Z$  represents all other influential factors besides  $S$ .

### 2.2.1 Social Institutions and Female Education

The following examples illustrate how social institutions related to gender inequality affect the private costs and returns of educational investments.<sup>11</sup> Social institutions related to gender inequality influence the costs of education as they shape a gendered division of labor and the opportunity costs of educating girls. Opportunity costs include income from child labor and are higher for girls when they are expected to do housework, to care for their younger siblings or to work in agriculture (Hill and King, 1995; Lahiri and Self, 2007). Social institutions related to gender inequality also affect the returns to education. The returns are generally lower for girls than for boys because girls and women are discriminated on the labor market in the form of entry restrictions and wage gaps. Thus, boys are expected to be economically more productive. Furthermore, parents often expect only low returns from

<sup>10</sup>See Pasqua (2005) who considers both perspectives, the non-unitary approach to the household and the cost and returns approach in the case of education of girls.

<sup>11</sup>It must be noted that the private *NPV* of investments in the education of children does not correspond to the social *NPV*. Social returns to education, especially female education, are often higher than the private ones. There is evidence that society benefits from female education as it contributes to overall development and drives economic growth (Hill and King, 1995; Klasen, 2002; Braunstein, 2007; Klasen and Lamanna, 2009). The resulting investment in female education will then often be sub-optimal.

female education because the daughter marries and leaves the house implying that the family loses her labor force. As a consequence sons become the building block of their parents' old-age security (Hill and King, 1995; Pasqua, 2005; Song et al., 2006).<sup>12</sup>

The costs and returns perspective does not rule out that the distribution of decision-making power in the household matters. The non-unitary household approach can be used to explain low female education (Pasqua, 2005). Several empirical studies show that when women dispose of more resources, investments in the education of girls are higher (e.g. Schultz, 2004; Emerson and Souza, 2007).

*Hypothesis 1: Social institutions that deprive women of their autonomy and bargaining power in the household or that increase the private costs and reduce the private returns to investments into female education are associated with lower female education than in a more egalitarian environment.*

## 2.2.2 Social Institutions and Fertility and Child Mortality Rates

Social institutions related to gender inequality that influence female decision-making power in the household and the *NPV* of the investment in girls in comparison to boys are also relevant for fertility levels and child mortality.

Concerning fertility, one can use the non-unitary household approach and argue that the net utility of a woman associated with getting a child might differ from that of a man. If one assumes that man and woman derive the same satisfaction of having a child, the net utility a woman derives is lower than the one of the man as she bears most of the costs of having children. These costs are related to the discomfort and health risks related to pregnancy, and the income losses associated with time spent on child care. This might explain why women want less children than men, but cannot achieve their objectives as social institutions restrict their power in limiting the number of children born. Empirical studies support the hypothesis that reduced female bargaining power leads to shorter time spans between births, a lower use of contraceptives and higher fertility levels (Thomas, 1990; Abadian, 1996; Hindin, 2000; Saleem and Bobak, 2005; Seebens, 2008).

The perspective of the *NPV* provides a second explanation for higher fertility. In the absence of well-functioning insurance markets and pension systems, parents in developing countries may need more children to feel secure. Depending on the costs of a child and

---

<sup>12</sup>In addition to all of these considerations, social institutions related to gender inequality might affect the supply of schooling which might influence the decision to send girls to school if school environments are hostile to the needs of girls (e.g. no female teachers available, long distances to school or prices in favor of boys) (Hill and King, 1995; Alderman et al., 1996; Pasqua, 2005; Lahiri and Self, 2007).

the returns to the investment in a child parents will consider to get more children. As it was explained in the previous subsection on female education, social institutions related to gender inequality affect the *NPV* of investments in children. If these social institutions lower income earning opportunities for girls, the *NPV* of investments in girls will be lower than the *NPV* of investments in boys. Hence, sons yield the promise of more economic security as compared to daughters. As long as parents cannot perfectly control the sex of their offspring, they will bear more children to increase the chance of having more sons (Cain, 1984; Abadian, 1996; Kazianga and Klonner, 2009).

To explain higher child mortality levels with social institutions that disadvantage women one has to bear in mind that mothers are usually the primary caregivers of children. Within the non-unitary framework, if mothers have only limited power in the household, they are constrained in the use of health care or in the access to food and other goods necessary for children. Thus, they cannot take care of their children as they would without those restrictions. This might lead to worse child health and higher child mortality rates (Thomas, 1990, 1997; Bloom et al., 2001; Smith et al., 2002; Maitra, 2004; Shroff et al., 2009).

From the *NPV* perspective it might be rational for parents to invest more in the health and nutrition of boys than in girls who as a consequence could suffer more heavily from health problems and experience higher mortality rates than boys. It is possible that this behavior increases overall child mortality rates. In addition, the limited education that women typically receive in patriarchal societies as a result of past *NPV* calculations of their parents might also lead to worse child health and to higher child mortality figures (Schultz, 2002; Shroff et al., 2009).

*Hypothesis 2: Social institutions that deprive women of their autonomy and bargaining power in the household or that increase the private costs and reduce the private returns of investments into girls are associated with higher fertility levels than in an egalitarian environment.*

*Hypothesis 3: Social institutions that deprive women of their autonomy and bargaining power in the household or that increase the private costs and reduce the private returns of investments into girls are associated with higher child mortality than in an egalitarian environment.*

## 2.3 Social Institutions and the Society: Governance

In societies where social institutions limit the rights of women, and where women's place is restricted to the private sphere, they have no or less say in the public and political domain. What is the impact of social institutions related to gender inequality on governance? We use Kaufmann et al. (2008, p. 7)'s definition of governance "as the traditions and institutions by which authority in a country is exercised. This includes the process by which governments are selected, monitored and replaced; the capacity of the government to effectively formulate and implement sound policies; and the respect of citizens and the state for the institutions that govern economic and social interactions among them."

There are at least two approaches that allow to link social institutions with governance. First, there exist psychological and sociological explanations that state that women are less egoistic than men. Women are more risk-averse, they tend to follow the rules and they are more community-oriented than men (Dollar et al., 2001; Swamy et al., 2001). Countries in which women have more power will have a political system that is more rule-oriented, responsive and accountable. Second, women's movements, being the answer to the exclusion of women from power, play an important role in increasing the quality of political systems by challenging e.g. personal rule (Waylen, 1993; Tripp, 2001). This argumentation suggests that countries with social institutions that hinder women to organize and to express their interests might lack an important oppositional force and therefore have a bad quality of governance.

*Hypothesis 4: Social institutions related to high gender inequality inhibit the building blocks of good governance. In societies with social institutions favoring gender inequality political systems will be less responsive and less open to the citizens, so that 'voice and accountability' will be reduced.*

*Hypothesis 5: Social institutions related to high gender inequality inhibit the building blocks of good governance. In societies with social institutions favoring gender inequality there might be more personal rule in the political system as well as inequality in justice and legal systems, so that the 'rule of law' will be weakened.*

## 2.4 Data

Our investigation uses macro-data at the country level. Table 2.1 gives an overview over the variables used for our estimations, the definitions and the data sources. Descriptive statistics of the variables used are presented in Table 2.2. As main regressors we use the SIGI and its



five subindices *Family code*, *Civil liberties*, *Physical integrity*, *Son Preference* and *Ownership rights* in our estimations to check their explanatory value for the development outcomes female education, fertility, child mortality and governance.

First, we are interested in the impact of social institutions on female education, fertility and child mortality. As dependent variables we use *total fertility rates* from World Bank (2009a) and *child mortality rates* from World Bank (2008). To measure education we choose *female gross secondary school enrollment rates* because this enables important functionings and empowers women. Furthermore we assume that parents take into account that basic education of both boys and girls is necessary for fulfilling tasks related to the household. Data for secondary school enrollment are from World Bank (2009a).

Second, we want to estimate the association between governance and our social institutions measures. We use the Governance Indicators developed by Kaufmann et al. (2008) and choose two of them to capture equality before the law, justice, tolerance and security as well as responsiveness, political openness and accountability in the political system. The *rule of law* index measures the extent to which contracts are enforced and property rights are ensured and the extent to which people trust in the state and respect the rules of the society. The *voice and accountability* index proxies civil and political liberties like freedom of expression, freedom of association, free media and the extent of active and passive political participation of citizens.

In all regressions we control for the level of economic development, religion, region and the political system in a country. The specific variables we use are:

- the log of per capita GDP in constant prices to control for the level of economic development (US\$, PPP, base year: 2005);
- a Muslim majority and a Christian majority dummy to control for the impact of religion, the left-out category being countries that have neither a majority of Muslim nor a majority of Christian population;
- region dummies to capture geography and other unexplained heterogeneity that might go together with region, the left-out category being Sub-Saharan Africa;
- two political institutions variables, the electoral democracy variable and the civil liberties index from Freedom House (2008) that together measure liberal democracy which is assumed to be related to responsiveness to the needs of the public, political openness and tolerance in a country.<sup>13</sup>

We use different additional control variables in each regression following suggestions in the literature. In the fertility and child mortality regressions, we additionally control for

<sup>13</sup>We multiply the civil liberties index by -1 to facilitate interpretation.



- female literacy rates to measure the ability of women to control their reproductive behavior, to care for themselves and their children (e.g. Basu, 2002; Hatt and Waters, 2006; Lay and Robilliard, 2009);
- a dummy proxying for high HIV/AIDS prevalence rates to control for extreme health problems especially in Sub-Saharan Africa due to AIDS (e.g. Foster and Williamson, 2000).

The Governance regressions exclude as control variables the civil liberties index from Freedom House as this index is used to build the voice and accountability index that we choose as dependent variable. We keep the electoral democracy variable because it does not pose a problem. We additionally include as control variables

- the share of literate adult population to control for the population's ability to be informed, to express their needs and to hold politicians' accountable (Keefer and Khemani, 2005);
- ethnic fractionalization as it might disturb governance through identity politics, patronage and distribution conflicts (e.g. Collier, 2001; Tripp, 2001);
- a measure of trade openness as openness increases the incentives to build 'good' institutions to attract trading partners, to join trading agreements etc. (e.g. Al-Marhubi, 2005).

Social institutions, i.e. normative frameworks, change only slowly and incrementally. As the social institutions indicators are not expected to change much over time we have to decide which year or time span should be covered by the other variables. For our response variables we choose to take the average of the existing values over five or six years (2000-2005, 2001-2005). For the control variables we take the averages of the existing values over ten years (1996-2005).<sup>14</sup> The averages provide information that is more stable than using a particular year. Using a longer time span for the control variables than for the response variables allows to capture possible time delays until effects can be observed. Nevertheless, we acknowledge that the choice of the time spans is arbitrary.

---

<sup>14</sup>The ethnic fractionalization variable is constant over time as changes in the ethnic composition of a country at least over 20 and 30 years are rare.

## 2.5 Empirical estimation and Results

### 2.5.1 Empirical estimation

We empirically test with linear regressions whether the composite measures reflecting social institutions related to gender inequality  $s_i$  are associated with each of the response variables  $y_i$ , representing the chosen development outcomes. We estimate regressions in the form

$$y_i = \gamma + \beta s_i + \text{control variables}_i + \varepsilon_i \quad (2.2)$$

using information at the country level. We are mainly interested in testing the null hypothesis that the coefficient  $\beta$  is zero at a statistical significance level of  $\alpha = 10\%$ . If the null hypothesis is rejected, it is reasonable to infer that the measure proxying social institutions related to gender inequality does matter for the given response variable, as predicted in the hypotheses from sections 2.2 and 2.3.

The general procedure used for each of the response variables consists of two steps. First, we start examining the effect of SIGI. We begin our estimation with a simple linear regression with SIGI as the only regressor  $s_i$ . We then run a multiple linear regression adding the main group of control variables that consists of the level of economic development, region dummies, religion dummies and the political system variables. If SIGI is significant in this regression, we continue and, if applicable, estimate the complete model with all control variables to confirm whether SIGI remains significant.

As SIGI is a rather broad measure to rank and compare countries and policy implications are difficult to derive from it, in a second step we focus on the subindices to get a more precise idea about what kind of social institutions might be related to the chosen development outcomes. We estimate the same multiple linear regression(s) described above using the five subindices as  $s_i$  one at a time instead of SIGI to explore which dimension of social institutions related to gender inequality seems to be the most relevant. In the corresponding regression tables we only report the specification with the subindex or subindices that are statistically significant. It must be noted that we keep and show even those control variables that are not statistically significant in the regression, as we want to stress that the social institutions indices are associated with the development outcomes even if we include these control variables.

All regressions are estimated with Ordinary Least Squares (OLS). Regression diagnostics not reported here suggest that heteroscedasticity is a possible issue in our data and that there are influential observations that could drive our results. Concerning the first issue, it is known that if the model is well specified, the OLS estimator of the regression parameters remains unbiased in the presence of heteroscedasticity, but the estimator of the covariance

matrix of the parameter estimates can be biased and inconsistent making inference about the estimated regression parameters problematic. Violations of homoscedasticity can lead to hypothesis tests that are not valid and confidence intervals that are either too narrow or too wide. To deal with heteroscedasticity, we use 'heteroscedasticity-consistent' (HC) standard errors. This means that while the parameters are still estimated with OLS, alternative methods of estimating the standard errors that do not assume homoscedasticity are applied. As the samples we use contain less than 150 observations, we use HC3 robust standard errors proposed by Davidson and MacKinnon (1993), which are better in the case of small samples. These are the standard errors that are presented in the regression Tables 2.5-2.9. Simulation studies by Long and Ervin (2000) have shown that HC standard error estimates tend to maintain test size closer to the nominal alpha level in the presence of heteroscedasticity than OLS standard error estimates that assume homoscedasticity. These authors recommend the use of HC3 robust standard errors, especially for sample sizes less than 250, as they can keep the test size at the nominal level regardless of the presence or absence of heteroscedasticity, with only a minor loss of power associated when the errors are indeed homoscedastic.<sup>15</sup>

In addition to this, we also use bootstrap with 1000 replications to compute a Bias-corrected and accelerated (Bca) 90% confidence interval of the regression coefficients computed with OLS (Efron and Tibshirani, 1993). One of the main advantages of bootstrapping methods is that no assumptions about the sampling distribution or about the statistic are needed. The results are not reported here, but are available upon request, and confirm that all the coefficients that are significant in Tables 2.5-2.9 remain significant when using Bca 90% confidence intervals around them.

To deal with the second issue and check whether influential observations drive the results, we take the estimates of a regression obtained with OLS with standard variance estimator to detect the observations with unusual influence or leverage based on Cook's distance. Cook's distance is a commonly used estimate of the influence of a data point when doing least squares regression. We exclude countries from the sample if the value of Cook's distance is larger than  $4/n$ , with  $n$  being the number of observations, and re-estimate each regression on the restricted sample with HC3 robust standard errors. In all the cases we confirm that even after we exclude influential observations, the results remain basically unchanged.<sup>16</sup> The regressions are not reported here, but are available upon request.

---

<sup>15</sup>Certainly, heteroscedasticity-consistent standard errors are not a panacea for inferential problems under heteroscedasticity. As pointed out by some authors, there are limitations and trade-offs in these estimators (e.g. Kauermann and Carroll, 2001; Wilcox, 2001).

<sup>16</sup>As an alternative procedure we use robust regression with iteratively reweighted least squares as described in Hamilton (1992), and confirm that results are similar.

We consider that the model specification is reasonable. However, possible endogeneity of our main regressors  $s_i$  (the SIGI and its subindices) should be taken into account when interpreting the coefficients of  $s_i$  as they would be biased and inconsistent in this case. Endogeneity is given if  $s_i$  is correlated with the disturbance  $\varepsilon_i$  in equation 2.2. There are three sources of endogeneity: omitted variables, measurement error and simultaneity (Wooldridge, 2002). We have included control variables to minimize omitted variable bias, although it is impossible to completely rule out this problem. Concerning measurement error, we regard the SIGI and the subindices as adequate proxies of social institutions related to gender inequality. It is not very plausible that there are errors in measurement that are related to the unobserved social institutions. The last source, simultaneity, arises when  $s_i$  is determined simultaneously with  $y_i$ . As was discussed in section 2.1, we consider that social institutions related to gender inequality  $s_i$  are relatively stable and long-lasting. Therefore, we think it is unlikely that the response variables  $y_i$  influence  $s_i$ .<sup>17</sup>

## 2.5.2 Results

Before we run the regressions it is necessary to check first the correlation between the subindices to rule out redundancy, and secondly between the subindices and the control variables to check whether the social institutions indices are proxies for these control variables. The Pearson correlation coefficient between the subindices is always positive, but not always significant (Table 2.3). The correlation coefficients are always lower than 0.6, with the exception of the correlation between the subindices *Family Code* and *Ownership rights*, which is equal to 0.75.<sup>18</sup> Table 2.4 shows that the absolute value of the Pearson correlation coefficient between the social institutions indicators and the control variables is always lower than 0.6, except for the SIGI and the subindices *Family code* and *Ownership rights* and the two variables capturing literacy of the whole population and of the female population.

Regression results using *female secondary education* as dependent variable are presented in Table 2.5. Regression (1) with SIGI as the only regressor yields a negative and statistically significant association. Higher levels of inequality are associated with lower levels of female

<sup>17</sup>Social institutions are hard to measure. Therefore, sometimes one has to rely on legal indicators to proxy them, although we acknowledge that this could pose problems as there is for example an international mechanism, the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), that aims at changing social institutions through legal measures. However, the impact of CEDAW on national legislation depends on the willingness of governments to sign and ratify it without reservation and on its willingness and ability to enact the new laws. Given the constituting function of social institutions for a society this could be difficult and depends on many factors.

<sup>18</sup>Table 1.7 shows Kendall tau b between the five subindices and confirms that they are positively correlated, albeit not perfectly.

secondary education. The association vanishes in regression (2) if one includes the level of economic development, religion, region and the political system as control variables. Using the subindex *Family code* instead of SIGI as the main regressor in regression (3) shows a different picture. The subindex is statistically significant even if the control variables are included. The adjusted coefficient of determination  $R^2$  is 0.78. Hence, we find no evidence against Hypothesis 1 that states that social institutions related to high gender inequality are negatively associated with female education.<sup>19</sup>

Results obtained using *total fertility rate* and *child mortality* as response variables are shown in Tables 2.6 and 2.7. In both cases, the simple linear regression (1) using SIGI as the only regressor shows a positive and significant statistical association between SIGI and the response variable. Higher levels of inequality are associated with higher levels of fertility and with higher levels of child mortality. However, once control variables related to the level of economic development, religion, region and the political system in a country are included in regression (2), SIGI is not longer statistically significant. This is not the case when we use the subindex *Family code* as the main regressor, as it is significant in regression (3) which uses the same control variables, and even in regression (4) which adds two additional regressors: the share of literate adult female population and a dummy reflecting high adult HIV/AIDS prevalence. In regression (4) the obtained adjusted  $R^2$  is 0.84 for fertility and 0.82 for child mortality. Hence, we cannot reject Hypotheses 2 and 3, suggesting that social institutions related to high gender inequality are associated with higher fertility levels and higher child mortality.<sup>20</sup> As the subindex *Family code* is the relevant social institutions measure in our empirical estimations it seems that social institutions that deprive women of their autonomy and bargaining power in the family and that might restrict women's possibilities outside the family do matter for female education, fertility and child mortality.

Table 2.8 shows the results obtained for the dependent variable *voice and accountability*. Regression (1) with SIGI as the only regressor shows a negative and statistically significant association: higher levels of gender inequality are associated with lower levels of voice and accountability. This association remains significant in regression (2) where we add the level of economic development, religion, region and the political system<sup>21</sup> as control variables, and in the complete specification shown in regression (3) where we additionally include the

<sup>19</sup>Regressions not reported here, but available upon request, using primary gross completion rates obtained from World Bank (2008) instead of female secondary schooling as the dependent variable yield similar results.

<sup>20</sup>Regressions not shown here, but available upon request, confirm that the results concerning mortality rates hold when using infant mortality rates taken from World Bank (2008) instead of child mortality rates as the dependent variable.

<sup>21</sup>Recall that in the governance regressions we only include the electoral democracy variable of Freedom House (2008) as the civil liberties index is included in the chosen governance indicators which are now the response variables.

proportion of seats held by women in national parliaments, the literacy rate of the population, a measure of openness of the economy, and a measure of ethnic fractionalization. In regression (3), we obtain an adjusted  $R^2$  of 0.69. We explore which dimension of social institutions related to gender inequality is behind this result and find that it is the subindex *Civil liberties*. The specifications with the subindex *Civil liberties* in regressions (4) and (5) show that this subindex is negatively associated with voice and accountability and that this association is statistically significant even with the control variables. In regression (5) the adjusted  $R^2$  is 0.69. Hypothesis 4 cannot be rejected with this evidence suggesting that social institutions related to gender inequality inhibit the building blocks of good governance in the form of voice and accountability. The subindex *Civil liberties* is the relevant social institutions measure in our empirical estimations. The freedom of women to participate in public life seems to increase the quality of governance of a society. Relating back to theory, this could be due to the behavior of women as they tend to be more socially oriented than men and are a group that cross-cuts cleavages in general.

Results for the other component of governance, *rule of law*, are shown in Table 2.9, providing no evidence against Hypothesis 5. Regression (1) shows a negative and statistically significant association between SIGI and rule of law: higher levels of inequality are associated with lower levels of rule of law. This association remains significant in regression (2) where we add the level of economic development, religion, region and the political system as control variables, and in the complete specification in regression (3) where we additionally include the proportion of seats held by women in national parliaments, the literacy rate of the population, a measure of openness of the economy, and a measure of ethnic fractionalization. In this last regression, we obtain an adjusted  $R^2$  of 0.51. Again, we are interested in exploring which dimension of social institutions related to gender inequality is the relevant one for rule of law finding that two subindices matter: *Ownership rights* and *Civil liberties*.<sup>22</sup> The specifications with the subindices yield similar results to those of the SIGI and are presented in regressions (4) and (5) for *Ownership rights* and (6) and (7) for *Civil liberties*. For both subindices the adjusted  $R^2$  obtained for the complete specification is 0.56. As postulated in Hypothesis 5, social institutions related to gender inequality seem to matter for governance inhibiting the rule of law, e.g. through personal rule and inequality in justice. Assuming that women's attitudes are different from those of men and that they countervail injustice, women's power in a society contributes to improve rule of law. The two subindices proxy where this power comes from, with *Ownership rights* measuring economic power through

<sup>22</sup>As shown in Table 2.3 the Pearson Correlation coefficient between the subindices *Ownership rights* and *Civil liberties* is 0.36.

access to property and *Civil liberties* measuring the freedom to participate in and to shape public life.

A reasonable question is whether the social institutions indicators are capturing different religions. In the regressions reported here, we control for religion using a Christian and a Muslim dummy. As the results show, at least one subindex is significant when we control for religion. One could argue that what matters is how religion is practiced in the considered regions, and that the SIGI and the subindices might capture regional practice of religion. Therefore, we re-estimate all regressions including interactions between the religion and region dummies. The results for the SIGI and the subindices remain unchanged suggesting that they capture something different than religion and the regional practice of it.<sup>23</sup>

## 2.6 Conclusion

This study presents several answers to the question why we should care about social institutions related to gender inequality beyond the intrinsic value of gender equality. We derive hypotheses from existing theories and empirically test them with linear regression at the cross-country level using the newly created Social Institutions and Gender Index (SIGI) and its subindices. Our results show that social institutions related to gender inequality are associated with lower female secondary education, higher fertility rates, higher child mortality and lower levels of governance measured as voice and accountability and rule of law. We find that apart from geography, political system, the level of economic development and religion, one has to consider social institutions related to gender inequality to better account for differences in important development outcomes.

The empirical estimation follows a two-step procedure for each outcome measure. First, the focus is to examine the explanatory value of the SIGI. In the specifications including all control variables, the SIGI is significant in the regressions for the governance measures ‘voice and accountability’ and ‘rule of law’. If one interprets the SIGI as a summary measure of lack of power of women in all spheres of society then it seems that when women have more power, governance is better.<sup>24</sup> In the case of female secondary schooling, fertility rate and child mortality the SIGI turns out to be insignificant in the complete specifications.

Secondly, as the SIGI is a broad measure of social institutions related to gender inequality, we investigate which particular dimension of social institutions is significantly related to the chosen development outcomes, using the complete specifications. The subindex *Family code* is negatively associated with female education, fertility and child mortality. These

<sup>23</sup>The results are available upon request.

<sup>24</sup>The association between two composite measures like the SIGI and the governance indicators has to be interpreted carefully.



results suggest that social institutions that deprive women of their autonomy and bargaining power in the family do matter for female education, fertility and child mortality. The subindex *Civil liberties* is the dimension of social institutions that is significantly related to the governance component 'voice and accountability'. The freedom of women to participate in public life seems to increase the quality of governance of a society as women tend to be more socially oriented than men and are a group that cross-cuts cleavages in general. The 'rule of law' component of governance is negatively related to the subindices *Civil liberties* and *Ownership rights*. The two subindices proxy where this power comes from, with *Ownership rights* measuring access to property and *Civil liberties* measuring the freedom to participate in public life. Assuming that women's attitudes are different from those of men and that they countervail personal rule, women's power in a society is a relevant factor to improve 'rule of law'.

Although the subindices *Family code*, *Ownership rights* and *Civil liberties* are the relevant dimensions of social institutions related to gender inequality for the response variables considered in this study, this does not mean that the other two subindices *Son preference* and *Physical integrity* are not important intrinsically or instrumentally for other outcomes. Case studies investigating the mechanisms between social institutions and the outcome variables are necessary. Our study has the limitations of any cross-sectional regression analysis as we cannot rule out omitted variable bias. Causality can never be derived from regression analysis with cross-sectional data unless valid instruments are found. Concerning the results of the subindices, these should be considered exploratory and need to be confirmed with further research, which should also include the elaboration of appropriate theories linking social institutions related to gender inequality with each of the development outcomes used in this study.

Social institutions are long-lasting and deep-seated in people's minds. Changing them is a difficult task and requires approaches tailored to the particular needs and the socio-economic context (Jütting and Morrisson, 2005). The state can certainly help attenuate the effects of social institutions through specific policies. It may set incentives to counteract social institutions, e.g. in the form of laws to fight against discriminatory practices or through the implementation of programs favoring girls and women. Micro-credit programs or subsidies targeted at mothers are good examples here. Nevertheless, changing social institutions needs more than that. It needs a thorough understanding of the power relations in a country and people that are willing to become reform drivers and initiate learning processes that should be complemented by deliberation and public discussion at all levels of society. Be it through internal or external forces, women need help to empower themselves. That is what Sen calls 'agency of women' (Sen, 1999).



## 2.7 Tables

Table 2.1: Description and Sources of Variables

Variables	Definition	Source
<b>Response Variables</b>		
Fertility	Total fertility rate (births per woman) (average of existing values over the last five years)	World Bank (2009a)
Child mortality	Children under five mortality rate per 1,000 live births (year 2005)	World Bank (2008)
Female secondary school	School enrollment, secondary, female (% gross) (average of existing values over the last five years)	World Bank (2009a)
Voice and accountability	Index that combines several data sources based on expert perceptions of "the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media" (Kaufmann et al., 2008); (average of existing values over the last five years)	Kaufmann et al. (2008)
Rule of law	Index that combines several data sources based on expert perceptions of "the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence" (Kaufmann et al., 2008); (average of existing values over the last five years)	Kaufmann et al. (2008)
<b>Regressors</b>		
<b>SIGI</b>	Social Institutions and Gender Index	Branisa et al. (2009a)
<b>Subindex family code</b>	Subindex Family code	Branisa et al. (2009a)

Continued on next page

Table 2.1 – continued from previous page

Variables	Definition	Source
<b>Subindex civil liberties</b>	Subindex Civil liberties	Branisa et al. (2009a)
<b>Subindex physical integrity</b>	Subindex Physical integrity	Branisa et al. (2009a)
<b>Subindex son preference</b>	Subindex Son preference	Branisa et al. (2009a)
<b>Subindex ownership rights</b>	Subindex Ownership rights	Branisa et al. (2009a)
Literacy female	Share of literate adult female population (15+) (%) year 2000 (average of the existing values over the last 10 years)	World Bank (2009a)
Literacy population	Share of literate population (whole) (average of the existing values over the last 10 years)	Human Development Report (HDR) stats office
log GDP	Log of GDP per capita, PPP (constant 2005 international \$) (average over the last 10 years)	World Bank (2008)
FH civil liberties	-1 * Index that measures the extent to which countries ensure civil liberties including freedom of expression, assembly, association, education, and religion as well as personal autonomy. It covers whether there is an established and generally equitable system of rule of law, free economic activity and equality of opportunity. (scale -1 (best) to -7 (worst)) (average of the existing values over the last 10 years)	Freedom House (2008)
Electoral democracy	Index that qualifies countries as electoral democracy when there exist competitive, universal and free and secret elections and a multiparty system that can access the media for political campaigning; (average of the existing values over the last 10 years)	Freedom House (2008)
Parliament	Proportion of seats held by women in national parliaments (%) (average of the existing values over the last 10 years)	World Bank (2009a)
Aids	Adult (15-49) HIV prevalence percent by country, 1990-2007; Countries were coded 1 if Adult (15-49) HIV prevalence rate	UNAIDS/WHO (2008)

Continued on next page

Table 2.1 – continued from previous page

Variables	Definition	Source
Ethnic	<p>exceeds 5 per cent, otherwise 0.</p> <p>The ethnic fractionalization measure gives the probability that two individuals selected at random from a population are members of different groups. It is calculated with data on language and origin using the following formula <math>FRAC_j = 1 - \sum_{i=1}^N s_{ij}^2</math>, where <math>s_{ij}</math> is the proportion of group <math>i = 1, \dots, N</math> in country <math>j</math> going from complete homogeneity (an index of 0) to complete heterogeneity (an index of 1).</p>	Alesina et al. (2003)
Openness	Share of imports of goods and services of total GDP	World Bank (2008)
Muslim	Countries get a 1 if at least 50 % of the population are muslim, 0 otherwise.	Central Intelligence Agency (2009)
Christian	Countries get a 1 if at least 50 % of the population are christian, 0 otherwise.	Central Intelligence Agency (2009)
SA	Countries get a 1 if located in region South Asia, 0 otherwise.	
ECA	Countries get a 1 if located in region Europe and Central Asia, 0 otherwise.	
LAC	Countries get a 1 if located in region Latin America and the Caribbean, 0 otherwise.	
MENA	Countries get a 1 if located in region Middle East and North Africa 0 otherwise.	
EAP	Countries get a 1 if located in region East Asia and Pacific 0 otherwise.	

Table 2.2: Descriptive statistics of variables used

Variable	Observations	Mean	Std. Dev.	Min	Max
SIGI	102	0.126	0.122	0.002	0.678
Subindex Family Code	112	0.326	0.223	0.004	0.797
Subindex Civil Liberties	123	0.160	0.259	0	1
Subindex Physical integrity	114	0.358	0.191	0	0.971
Subindex Son preference	123	0.134	0.240	0	1
Subindex Ownership rights	122	0.298	0.266	0	1
Fertility	121	3.562	1.702	0.933	7.678
Child mortality	119	80.005	67.777	3.758	273.8
Female secondary school	108	59.210	30.484	6.037	113.275
Rule of law	123	-0.563	0.718	-2.142	1.658
Voice and accountability	123	-0.583	0.752	-2.102	1.088
SA	124	0.056	0.232	0	1
ECA	124	0.137	0.345	0	1
LAC	124	0.177	0.384	0	1
MENA	124	0.145	0.354	0	1
EAP	124	0.137	0.345	0	1
Muslim	124	0.331	0.472	0	1
Christian	124	0.435	0.498	0	1
log GDP	115	7.988	1.121	5.609	10.553
Literacy population	121	0.741	0.218	0.173	1
Literacy female	106	0.705	0.251	0.128	0.998
Electoral democracy	120	0.455	0.459	0	1
FH civil liberties	121	-4.366	1.434	-7	-1.4
Parliament	118	10.630	6.925	0	29.556
Aids	116	0.138	0.346	0	1
Openness	119	0.452	0.261	0.013	1.914
Ethnic	120	0.517	0.237	0.039	0.930

Table 2.3: Pearson Correlation Coefficient between the SIGI and the Subindices

		SIGI	Subindex Family	Subindex Civil	Subindex Physical	Subindex Son	Subindex Ownership
SIGI	$\rho$	1					
	Obs.	102					
Subindex Family	$\rho$	0.79	1				
	p-value	0.0000					
	Obs.	102	112				
Subindex Civil	$\rho$	0.71	0.47	1			
	p-value	0.0000	0.0000				
	Obs.	102	112	123			
Subindex Physical	$\rho$	0.66	0.59	0.28	1		
	p-value	0.0000	0.0000	0.0025			
	Obs.	102	103	113	114		
Subindex Son	$\rho$	0.54	0.18	0.53	0.02	1	
	p-value	0.0000	0.0594	0.0000	0.8312		
	Obs.	102	112	122	114	123	
Subindex Ownership	$\rho$	0.74	0.75	0.36	0.51	0.13	1
	p-value	0.0000	0.0000	0.0001	0.0000	0.1504	
	Obs.	102	111	121	112	121	122

Table 2.4: Correlation of the SIGI and the Subindices with the Control Variables

		SIGI	Subindex Family	Subindex Civil	Subindex Physical	Subindex Son	Subindex Ownership
log GDP	$\rho$	-0.34	-0.39	0.20	-0.47	0.16	-0.48
	p-value	0.0005	0.0000	0.0362	0.0000	0.0948	0.0000
	Obs.	98	108	114	105	114	114
Muslim	$\rho$	0.50	0.42	0.57	0.40	0.36	0.23
	p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0122
	Obs.	102	112	123	114	123	122
Christian	$\rho$	-0.39	-0.33	-0.40	-0.27	-0.37	-0.05
	p-value	0.0001	0.0003	0.0000	0.0036	0.0000	0.5662
	Obs.	102	112	123	114	123	122
SA	$\rho$	0.30	0.13	0.33	-0.13	0.49	0.14
	p-value	0.0023	0.1589	0.0002	0.1652	0.0000	0.1319
	Obs.	102	112	123	114	123	122
ECA	$\rho$	-0.32	-0.38	-0.25	-0.17	-0.17	-0.33
	p-value	0.0012	0.0000	0.0057	0.0762	0.0659	0.0002
	Obs.	102	112	123	114	123	122
LAC	$\rho$	-0.42	-0.47	-0.29	-0.36	-0.24	-0.35
	p-value	0.0000	0.0000	0.0012	0.0001	0.0076	0.0001
	Obs.	102	112	123	114	123	122
MENA	$\rho$	0.23	0.16	0.53	0.08	0.42	0.02
	p-value	0.0196	0.0843	0.0000	0.3796	0.0000	0.8501
	Obs.	102	112	123	114	123	122
EAP	$\rho$	-0.19	-0.29	-0.11	-0.15	0.10	-0.28
	p-value	0.0505	0.0017	0.2205	0.1127	0.2934	0.0016
	Obs.	102	112	123	114	123	122

Continued on next page

Table 2.4 – continued from previous page

		SIGI	Subindex Family	Subindex Civil	Subindex Physical	Subindex Son	Subindex Ownership
Electoral democ.	$\rho$	-0.39	-0.38	-0.34	-0.37	-0.22	-0.24
	p-value	0.0001	0.0000	0.0001	0.0001	0.0179	0.0091
	Obs.	101	110	119	111	119	119
FH civil libert.	$\rho$	-0.44	-0.30	-0.42	-0.42	-0.28	-0.25
	p-value	0.0000	0.0016	0.0000	0.0000	0.0021	0.0056
	Obs.	101	110	120	112	120	120
Parliament	$\rho$	-0.15	-0.15	-0.28	-0.18	-0.17	-0.11
	p-value	0.1514	0.1202	0.0023	0.0578	0.0750	0.2611
	Obs.	100	109	117	110	118	117
Literacy pop.	$\rho$	-0.66	-0.70	-0.19	-0.59	-0.25	-0.59
	p-value	0.0000	0.0000	0.0389	0.0000	0.0054	0.0000
	Obs.	102	112	120	112	121	119
Literacy female	$\rho$	-0.64	-0.68	-0.13	-0.58	-0.15	-0.62
	p-value	0.0000	0.0000	0.1891	0.0000	0.1286	0.0000
	Obs.	95	103	106	98	106	105
Openness	$\rho$	-0.20	-0.10	-0.07	-0.13	-0.13	-0.17
	p-value	0.0509	0.2995	0.4465	0.1784	0.1775	0.0605
	Obs.	101	111	118	109	118	117
Ethnic	$\rho$	0.40	0.51	0.08	0.41	-0.11	0.46
	p-value	0.0000	0.0000	0.3918	0.0000	0.2548	0.0000
	Obs.	101	110	119	111	119	119
AIDS	$\rho$	0.12	0.36	0.02	0.02	-0.19	0.36
	p-value	0.2312	0.0002	0.8425	0.8684	0.0381	0.0001
	Obs.	99	108	115	107	115	115

Table 2.5: Linear regressions with dependent variable female secondary school

Specification with SIGI	(1) b/se	(2) b/se	Specification with Subindex	(3) b/se
<b>SIGI</b>	<b>-141.77***</b> (37.31)	<b>-10.91</b> (36.37)	<b>Subindex family code</b>	<b>-39.10**</b> (11.64)
log GDP		12.69*** (3.39)	log GDP	11.46*** (2.61)
Muslim		-2.21 (5.47)	Muslim	3.43 (4.84)
Christian		5.31 (5.48)	Christian	4.18 (4.33)
SA		16.05 (8.75)	SA	12.3 (8.44)
ECA		40.26*** (8.98)	ECA	28.25*** (6.95)
LAC		18.33* (9.07)	LAC	8.64 (7.41)
MENA		33.86** (12.50)	MENA	29.67** (9.69)
EAP		24.73** (8.26)	EAP	14.36* (6.53)
Electoral democracy		8.11 (7.67)	Electoral democracy	6.19 (6.84)
FH civil liberties		1.95 (3.56)	FH civil liberties	2.72 (2.89)
constant	74.75*** (4.12)	-56.71 (37.27)	constant	-27.87 (30.56)
Number of obs.	94	91	Number of obs.	99
Adj. R-Square	0.28	0.75	Adj. R-Square	0.78
Prob>F	0.0003	0.0000	Prob>F	0.0000

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . HC3 robust standard error in brackets.

Regression (2) and (3) with controls for economic development, geography, religion and political system. In this case, this specification corresponds to the complete specification.



Table 2.6: Linear regressions with dependent variable fertility

Specification with SIGI	(1)	(2)	Specification with Subindex	(3)	(4)
	b/se	b/se		b/se	b/se
<b>SIGI</b>	<b>8.25***</b> (2.31)	<b>1.73</b> (2.61)	<b>Subindex family code</b>	<b>1.89**</b> (0.70)	<b>2.03**</b> (0.70)
log GDP		-0.71*** (0.16)	log GDP	-0.60*** (0.12)	-0.43*** (0.12)
Muslim		0.52 (0.27)	Muslim	0.34 (0.27)	0.18 (0.27)
Christian		0.25 (0.26)	Christian	0.24 (0.25)	0.46 (0.26)
SA		-1.89*** (0.37)	SA	-1.73*** (0.41)	-1.88*** (0.38)
ECA		-2.44*** (0.48)	ECA	-2.08*** (0.38)	-1.59*** (0.43)
LAC		-0.96* (0.47)	LAC	-0.68 (0.36)	-0.57 (0.40)
MENA		-1.42* (0.63)	MENA	-1.07* (0.50)	-1.23* (0.48)
EAP		-1.74*** (0.42)	EAP	-1.37*** (0.39)	-1.20** (0.38)
Electoral democracy		-0.2 (0.31)	Electoral democracy	0.02 (0.29)	-0.03 (0.30)
FH civil liberties		-0.02 (0.17)	FH civil liberties	-0.11 (0.13)	-0.14 (0.13)
			Literacy female		-1.62** (0.60)
			Aids		-0.51 (0.30)
constant	2.55*** (0.25)	9.76*** (1.82)	constant	7.89*** (1.30)	7.47*** (1.29)
Number of obs.	100	97	Number of obs.	106	99
Adj. R-Square	0.31	0.82	Adj. R-Square	0.80	0.84
Prob>F	0.0006	0.0000	Prob>F	0.0000	0.0000

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . HC3 robust standard error in brackets.

Regression (2) and (3) with minimum of controls for economic development, geography, religion and political system. Regression (4) with complete specification for fertility.

Table 2.7: Linear regressions with dependent variable child mortality

Specification with SIGI	(1) b/se	(2) b/se	Specification with Subindex	(3) b/se	(4) b/se
<b>SIGI</b>	<b>318.56**</b> (108.81)	<b>50.42</b> (150.58)	<b>Subindex family code</b>	<b>80.14**</b> (25.85)	<b>77.23*</b> (31.50)
log GDP		-22.55** (7.35)	log GDP	-20.24*** (5.34)	-13.82** (5.09)
Muslim		26.61 (14.13)	Muslim	14.23 (13.13)	5.74 (14.50)
Christian		7.49 (11.72)	Christian	9.47 (10.31)	14.27 (10.81)
SA		-68.33*** (18.87)	SA	-61.30*** (17.05)	-71.03*** (16.33)
ECA		-85.65*** (23.82)	ECA	-66.13*** (16.75)	-53.16* (20.65)
LAC		-66.65** (23.84)	LAC	-50.69*** (14.88)	-50.23** (18.89)
MENA		-97.73*** (26.90)	MENA	-86.25*** (21.71)	-93.71*** (23.48)
EAP		-73.44*** (17.23)	EAP	-59.37*** (15.02)	-55.65** (17.85)
Electoral democracy		-0.79 (15.86)	Electoral democracy	7.05 (15.96)	1.75 (14.80)
FH civil liberties		-4.54 (7.86)	FH civil liberties	-8.33 (6.65)	-8.32 (6.44)
			Literacy female		-62.77** (21.39)
			Aids		-19.02 (14.56)
constant	43.38*** (10.80)	272.39** (93.09)	constant	209.47** (66.26)	209.34** (63.27)
Number of obs.	99	97	Number of obs.	106	99
Adj. R-Square	0.28	0.79	Adj. R-Square	0.79	0.82
Prob>F	0.0043	0.0000	Prob>F	0.0000	0.0000

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . HC3 robust standard error in brackets.

Regression (2) and (3) with controls for economic development, geography, religion and political system. Regression (4) with complete specification for child mortality.

Table 2.8: Linear regressions with dependent variable 'voice and accountability'

Specification with SIGI	(1)	(2)	(3)	Specification with Subindex	(4)	(5)
	b/se	b/se	b/se		b/se	b/se
<b>SIGI</b>	<b>-2.60***</b>	<b>-1.42**</b>	<b>-1.59**</b>	<b>Subindex civil liberties</b>	<b>-0.61**</b>	<b>-0.65**</b>
	(0.50)	(0.48)	(0.54)		(0.23)	(0.23)
log GDP		0.27***	0.30***	log GDP	0.31***	0.27***
		(0.06)	(0.06)		(0.05)	(0.06)
Muslim		0.18	0.15	Muslim	0.16	0.21
		(0.13)	(0.14)		(0.13)	(0.14)
Christian		-0.03	-0.04	Christian	-0.05	-0.08
		(0.12)	(0.13)		(0.12)	(0.12)
SA		-0.27	-0.28	SA	-0.12	-0.04
		(0.20)	(0.21)		(0.18)	(0.20)
ECA		-0.64***	-0.56*	ECA	-0.52***	-0.57**
		(0.14)	(0.22)		(0.13)	(0.22)
LAC		-0.40*	-0.41*	LAC	-0.32*	-0.31
		(0.17)	(0.18)		(0.15)	(0.16)
MENA		-0.45	-0.47	MENA	-0.27	-0.23
		(0.23)	(0.25)		(0.19)	(0.24)
EAP		-0.30*	-0.21	EAP	-0.14	-0.21
		(0.14)	(0.21)		(0.13)	(0.18)
Electoral democracy		1.10**	1.07***	Electoral democracy	1.13**	1.14***
		(0.12)	(0.11)		(0.10)	(0.10)
Parliament			0.01	Parliament		0.01
			(0.01)			(0.01)
Literacy population			-0.31	Literacy population		0.24
			(0.42)			(0.37)
Openness			-0.07	Openness		0.23
			(0.36)			(0.22)
Ethnic			-0.07	Ethnic		0.01
			(0.25)			(0.23)
constant	-0.23*	-2.80***	-2.77***	constant	-3.28***	-3.37***
	(0.10)	(0.45)	(0.47)		(0.41)	(0.39)
Number of obs.	102	97	95	Number of obs.	112	108
Adj. R-Square	0.18	0.69	0.69	Adj. R-Square	0.68	0.69
Prob>F	0.0000	0.0000	0.0000	Prob>F	0.0000	0.0000

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . HC3 robust standard error in brackets.

Regression (2) and (4) with controls for economic development, geography, religion and political system.

Regressions (3) and (5) with complete specification for governance/voice and accountability.

Table 2.9: Linear regressions with dependent variable 'rule of law'

	Specification with SIGI			Specification with Subindices					
	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se	(6) b/se	(7) b/se		
SIGI	-1.73*** (0.49)	-1.88*** (0.53)	-1.33* (0.60)	Subindex ownership rights	-0.89*** (0.20)	-0.71** (0.23)	Subindex civil liberties	-0.75** (0.24)	-0.63* (0.25)
log GDP		0.41*** (0.08)	0.36*** (0.07)	log GDP	0.37*** (0.08)	0.30*** (0.07)	log GDP	0.47*** (0.08)	0.36*** (0.07)
Muslim		0 (0.16)	-0.04 (0.16)	Muslim	-0.03 (0.13)	-0.02 (0.14)	Muslim	0.04 (0.14)	0.11 (0.14)
Christian		-0.18 (0.15)	-0.18 (0.14)	Christian	-0.11 (0.14)	-0.14 (0.13)	Christian	-0.22 (0.14)	-0.22 (0.13)
SA		0.18 (0.22)	0.26 (0.24)	SA	0.11 (0.17)	0.21 (0.20)	SA	0.37 (0.22)	0.44 (0.26)
ECA		-0.84*** (0.18)	-0.67* (0.27)	ECA	-0.93*** (0.16)	-0.83*** (0.22)	ECA	-0.71*** (0.15)	-0.74** (0.22)
LAC		-0.74*** (0.19)	-0.54* (0.21)	LAC	-0.78*** (0.19)	-0.61** (0.19)	LAC	-0.58*** (0.17)	-0.51** (0.18)
MENA		-0.14 (0.27)	0.17 (0.32)	MENA	-0.09 (0.25)	0.18 (0.29)	MENA	0.10 (0.24)	0.30 (0.28)
EAP		-0.31 (0.16)	-0.28 (0.23)	EAP	-0.35* (0.15)	-0.36 (0.20)	EAP	-0.12 (0.15)	-0.23 (0.20)
Electoral democracy		0.33* (0.14)	0.40** (0.13)	Electoral democracy	0.38** (0.11)	0.44*** (0.11)	Electoral democracy	0.38** (0.13)	0.46*** (0.12)
Parliament			0.01 (0.01)	Parliament		0.01 (0.01)	Parliament		0.01 (0.01)
Literacy population			-0.29 (0.42)	Literacy population		-0.03 (0.38)	Literacy population		0.20 (0.36)
Openness			0.69* (0.33)	Openness		0.71** (0.27)	Openness		0.73** (0.23)
Ethnic			-0.07 (0.32)	Ethnic		-0.12 (0.28)	Ethnic		-0.13 (0.27)
constant	-0.35*** (0.10)	-3.37*** (0.58)	-3.32*** (0.52)	constant	-3.06*** (0.56)	-2.94*** (0.53)	constant	-4.05*** (0.52)	-3.83*** (0.46)
Number of obs.	102	97	95	Number of obs.	112	108	Number of obs.	112	108
Adj. R-Square	0.09	0.49	0.51	Adj. R-Square	0.53	0.56	Adj. R-Square	0.52	0.56
Prob>F	0.0006	0.0000	0.0000	Prob>F	0.0000	0.0000	Prob>F	0.0000	0.0000

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . HC3 robust standard error in brackets. Regression (2), (4) and (6) with controls for economic development, geography, religion and political system. Regressions (3), (5) and (7) with complete specification for governance/rule of law.

## Essay 3

# Reexamining the link between gender and corruption: The role of social institutions

### Abstract

In this paper we reexamine the link between gender inequality and corruption. We review the literature on the relationship between representation of women in economic and political life, democracy and corruption, and bring in a new previously omitted variable that captures the level of discrimination against women in a society: social institutions related to gender inequality. Using a sample of developing countries we regress corruption on the representation of women, democracy and other control variables. Then we add the subindex civil liberties from the OECD Gender, Institutions and Development Database as the measure of social institutions related to gender inequality. The results show that corruption is higher in countries where social institutions deprive women of their freedom to participate in social life, even accounting for democracy and representation of women in political and economic life as well as for other variables. Our findings suggest that, in a context where social values disadvantage women, it might not be enough to push democratic reforms and to increase the participation of women to reduce corruption.

### 3.1 Introduction

Is there a link between gender inequality and corruption in a society? The studies of Swamy et al. (2001) and Dollar et al. (2001) suggest that countries with greater representation of women in political and economic life tend to have lower levels of corruption. How can this relationship be explained?

This could be attributed to behavioral differences between men and women. As mentioned by Dollar et al. (2001), there are experimental studies and studies using survey data that find that, on average, women are less selfish and might have higher moral and ethical standards than men (e.g. Eagly and Crowley, 1986; Glover et al., 1997; Eckel and Grossman, 1998; Rivas, 2008).<sup>1</sup> If one accepts that women are less selfish and align their actions on higher moral standards than men, having women in important political and economic positions might lead to less corruption in a country.

An alternative explanation is put forward by Swamy et al. (2001), who argue that the negative relationship between women's participation and corruption could be due to self-selection. Only a few women reach powerful positions, and these women possibly gain access to these positions as they are from the "better" part of the women's distribution.

From a historical perspective, Goetz (2007) claims that it is gendered access to political positions that explains why women seem to be less corrupt than men. Excluded from male patronage networks, women are restricted in their opportunities for corrupt behavior. As they are newcomers or only few in the political or business sphere, women lack familiarity with the rules of illicit exchange to their own benefit. They try to assert their position by acting honestly and trustworthily. This all leads to fewer corrupt activities by women, but as time passes and more women get access to power this effect might vanish.

It can also be argued that the observed relationship between women's representation and corruption is spurious. Swamy et al. (2001) and Dollar et al. (2001) warn that even if one controls for other factors in the regression, the observed relationship at the cross-country level could be due to some unobserved variable which influences both female representation and corruption. For example, according to Sung (2003) it might be the political system in the form of liberal democratic institutions that influences both. Sung (2003) argues that institutions of *liberal* democracy increase women's participation in government through values like equality, pluralism, fairness and tolerance. Competitive elections, an independent judiciary and a free press, which are elementary to a liberal democratic system, guarantee transparency

---

<sup>1</sup>There are empirical studies that challenge the finding that women are the "fairer sex" (e.g. Andreoni and Vesterlund, 2001; Alhassan-Alolo, 2007; Alatas et al., 2009). Another investigation highlights that when women are in a powerful position, they take decisions that are closely related to women's needs (Chattopadhyay and Duflo, 2004).

and hold government officials accountable, thereby reducing corruption. Therefore, the negative effect of women's representation in government on corruption is spurious and vanishes when one includes a measure of democracy in the regression, which is empirically confirmed by Sung (2003). Swamy et al. (2001) draw attention to the "level of discrimination against women" as another possible omitted variable that drives both female participation and corruption. They claim that in countries that are more corrupt there is more discrimination against women and argue that in countries where traditions and clientelism prevail, there is a preference for men in power.

In this paper, we focus on the effect of discrimination against women on corruption in a society as we have a new measure of society's attitude towards gender inequality to empirically test this relationship. Swamy et al. (2001) do not explain how this relationship operates, but several papers deal with this issue in a direct or indirect way (Tripp, 2001; Inglehart et al., 2002; Rizzo et al., 2007). These authors claim that society's attitude towards women influences how a political system functions and that it affects the positions women take in this system. Assuming that the level of corruption depends on the functioning of the political system, one could argue that society's attitude towards gender inequality has an impact on corruption.

The study of Tripp (2001) focuses on women's movements as a countervailing force to prevailing practices of corruption in Eastern and South Africa.<sup>2</sup> Political reforms at the beginning of the 1990s, including free and competitive elections, a multi-party system and freedom of expression and association were not enough to give women access to powerful positions and to curtail the practices of patronage, clientelism and personal rule. Women could enter the system, but they were excluded from male-dominated networks and therefore from the benefits of clientelism. However, political reforms allowed the formation of social forces. The disadvantaged women organized in autonomous movements, which were broad-based, multi-ethnic and multi-religious. These movements cross-cut cleavages and started to demand transparency and the removal of clientelistic networks.

A similar perspective is adopted by Inglehart et al. (2002) and Rizzo et al. (2007) who state that when a society favors gender equality, there is more tolerance in general, more personal freedom and individual autonomy. The absence of these values inhibits political reforms towards a democratic system. The study of Inglehart et al. (2002) finds that gender equality is the most important part of "self-expression values" appearing in post-industrialization societies which directly contribute to both democratization and to a greater representation of women in politics. Focusing on Arab and non-Arab Muslim countries, Rizzo et al. (2007) shows that even if democratic political institutions like elections, political parties or checks

---

<sup>2</sup>Waylen (1993) makes a similar point for Latin America.

and balances are put in place, gender inequality can prevent these institutions from functioning well.

We empirically test on a sample of developing countries the relationship between social institutions related to gender inequality and the level of corruption, and contribute to the literature discussed above. We focus on public corruption, which refers to the misuse of public office for private gain. It comprises grand corruption, which refers to activities of top officials and big companies, and petty corruption, which refers to the activities of people at the lower end of hierarchies (Pardo, 2004). To proxy society's attitude towards gender inequality or what Swamy et al. (2001) call "level of discrimination against women" we introduce social institutions related to gender inequality into the analysis. Social institutions are long-lasting norms, traditions and codes of conduct that shape gender roles and influence the opportunities of women and men in a society. As suggested by e.g. De Soysa and Jütting (2007) and in Essay 2, these guiding principles of human behavior affect development outcomes and should not be neglected in the study of a society. We measure social institutions related to gender inequality with the subindex civil liberties proposed by Essay 1, which is based on variables from the OECD Gender, Institutions and Development Database (Jütting et al., 2008). This subindex captures society's attitude with regard to gender roles based on the freedom of women to participate in social life.

Our aim is to investigate whether society's attitude towards gender inequality matters for corruption once one accounts for the representation of women in parliament and business as well as the political system of a country. The hypothesis is that in a society where women's participation in social life is restricted, there is a higher level of corruption.

Even after controlling for democracy and political and economic participation of women, as well as for other factors, we find a robust and significant relationship between the subindex civil liberties and the level of corruption. We show that social institutions related to gender inequality are an important factor for the study of corruption. In societies where women are deprived of their freedoms to participate in social life, corruption is higher. As should be clear from the various existing theories the exact causal mechanism behind this relationship is not obvious and it cannot be established in this study since we conduct a cross-sectional analysis. This implies that one needs to carefully investigate the context, as tackling corruption might require more than pushing democratic reforms and increasing female representation in political and economic positions.

The rest of the paper is organized as follows. Section 3.2 describes the the data used, the empirical estimation and the main results, which are discussed in section 3.3.



## 3.2 Empirical Estimation and Results

### 3.2.1 Data

The definition of all variables and descriptive statistics are presented in Tables 3.1, 3.2 and 3.3. Measuring corruption is a complex task as it has many faces. There is public corruption, which refers to the misuse of public office for private gain, and corruption that comprises the collusion between firms or misuse of corporate assets (Svensson, 2005). Other authors differentiate between grand and petty corruption. Grand corruption refers to activities of top-officials and big companies. Petty corruption refers to the activities of people at the lower end of hierarchies (Pardo, 2004).

We use two different measures of public corruption in our estimations comprising grand and petty corruption. The first measure is the Corruption Perception Index (CPI) of Transparency International.<sup>3</sup> The CPI measures the level of corruption in a country. It is based on various data sources, business surveys and expert panels about perceptions of corruption, and is a comprehensive measure that covers the different forms of grand and petty corruption in business, politics and administration. It is continuous and ranges from 0 meaning high corruption to 10 meaning low corruption (Lambsdorff, 2006).

The second indicator is the Corruption in Government Index from the International Country Risk Guide (ICRG) provided by the Political Risk Services.<sup>4</sup> The ICRG index assesses the political risk associated with corruption and focuses in particular on those types of corruption that lead to instability in the political system as they distort the economic and financial environment, put foreign investments into risk and reduce the efficiency of government and business because people come to power not because of their ability but through patronage and clientelistic practices.<sup>5</sup> Hence, this measure gives the extent of political risk of instability that is assumed to increase with corruption. Therefore, it is only under certain conditions an indicator of the level of corruption. Whether the political risk of instability caused by corruption coincides with the level of corruption depends on the degree of tolerance towards corruption (Lambsdorff, 2006). The ICRG corruption index is a continuous variable that ranges from 0 to 6 with 0 meaning high risk and 6 indicating low risk.

The subindex civil liberties (Subindex Civil lib.) is one of five composite indices (the others being subindex family code, subindex son preference, subindex physical integrity, subindex ownership rights) that measure social institutions related to gender inequality, and were presented in Essay 1. These social institutions are conceived as long-lasting norms,

<sup>3</sup>Data are available at [http://www.transparency.org/policy\\_research/surveys\\_indices/cpi](http://www.transparency.org/policy_research/surveys_indices/cpi).

<sup>4</sup><http://www.prsgroup.com/>.

<sup>5</sup>[http://www.prsgroup.com/ICRG\\_Methodology.aspx#\#PolRiskRating](http://www.prsgroup.com/ICRG_Methodology.aspx#\#PolRiskRating)

traditions and codes-of conduct that find expression in traditions, customs and cultural practices, informal and formal laws and guide people's behavior and interaction. They shape gender roles and therefore the social and economic opportunities of men and women. We use the subindex civil liberties in this study as it covers those social institutions that directly shape the opportunities of women to participate in social life. It therefore reflects better their opportunities to gain power in politics and economics than the other subindices related to gender inequality. Indeed, we find that the subindex civil liberties is the only subindex that is significant in the regression analysis. The subindex civil liberties is built out of two variables of the OECD Gender, Institutions and Development Database (Morrison and Jütting, 2005; Jütting et al., 2008), which are freedom of movement and freedom of dress. Freedom of movement measures the level of restrictions women face in moving freely outside their own household. Freedom of dress measures the extent to which women are obliged to follow a certain dress code in public, for example being obliged to cover their face or body when leaving the house. Both variables are ordinal taking the values 0, 0.5 and 1 with 0 indicating no restrictions and 1 indicating high restrictions on women.<sup>6</sup> They are proxies of civil liberties in a sense that when women are restrained to leave the house it is difficult to imagine that they can actively participate in social, political and economic life. Wearing a veil might be a form of self-determination and expression, and different traditions, styles and customs are connected to it. However, *forced* veiling is incompatible with agency, as it might be a sign of subordination in a society and might hinder interactions with other human beings - either as women cannot interact because they wear a veil or they can only interact if they wear a veil (Macdonald, 2006; Milallos, 2007). The subindex is the rescaled weighted sum of the two variables with the weights obtained from polychoric principal component analysis (Kolenikov and Angeles, 2009). The subindex goes from 0 (no gender inequality) to 1 (high gender inequality). As the subindex civil liberties does not cover developed (OECD) countries, the subsequent empirical analysis focuses on developing countries.

The variables that are contained in the subindex could be considered as proxies for religion and therefore one could think that the subindex civil liberties might be a proxy for religion as well. When investigating the variation of the subindex over religion, one observes that there is more variation within Muslim majority countries than in countries with

---

<sup>6</sup>The variable freedom of dress takes the value 0 if less than 50% of women are obliged to follow a certain dress code, 0.5 if more than 50% of women are forced to follow a certain dress code, and 1 if all women are obliged to follow a certain dress code, or if it is punishable by law not to follow it. The variable freedom of movement is 0 if less than 50% of women face restrictions on their movement outside the home, 0.5 if more than 50% of women face restrictions on their movement outside the home, and 1 if women can never leave home without restrictions (i.e. they need a male companion).

either Christian majority or countries without Christian or Muslim majority (Table 3.5).<sup>7</sup> To further examine whether the subindex measures Muslim religion, we plot the subindex civil liberties against the percentage of Muslim population in a country (Figure 3.1). It is true that countries having less than 50% Muslim population tend to have lower values on the subindex civil liberties with the exception of India which scores 0.6 with about 15% of Muslim population. For countries with more than 50% Muslim population the subindex shows more variation. Noticeably, there are several countries that have more than 70% of Muslim population and the value 0 on the subindex civil liberties.<sup>8</sup> Consequently, there is no perfect correspondence between the subindex and the percentage of Muslim population. Nevertheless, in the regressions we include a Muslim and a Christian dummy (Muslim and Christian) to control for the impact of religion, the left-out category being countries that have neither a majority of Muslim nor a majority of Christian population.<sup>9</sup>

To account for female representation, which is highlighted by e.g. Swamy et al. (2001) and Dollar et al. (2001), we include three measures of female representation. We take data from World Bank (2009a) on the proportion of female legislators (Parliament), the female share in professional, technical, administrative and managerial positions (Managers),<sup>10</sup> and women's share of labor force (Labor force).

To capture democracy we choose the Electoral Democracy index (Electoral democ.) of Freedom House (2008) that takes the value 1 if there are competitive, universal, free and secret elections and a multiparty system. An alternative measure is the Polity2 index of the Polity IV Project that we use to check the robustness of the results as Polity2 measures more closely liberal democracy (Marshall and Jaggers, 2009).<sup>11</sup> Unfortunately, it covers

---

<sup>7</sup>The variable freedom of movement varies over all three religious categories, while the variable freedom of dress has almost no variation in countries having a Christian majority or countries without Christian or Muslim majority, except for India and Sri Lanka.

<sup>8</sup>Albania, Azerbaijan, Gambia, Guinea, Kyrgyz Republic, Mali, Morocco, Niger, Senegal, Sierra Leone, Tajikistan, Tunisia, Turkmenistan, Uzbekistan

<sup>9</sup>As Muslim religion is related to the subindex we also use the percentage of Muslim population instead of the two religion dummies in the regressions. The results are unchanged.

<sup>10</sup>Both indicators have been criticized (Bardhan and Klasen, 1999; Dijkstra, 2002). In some countries, for example communist ones, parliaments lack power and the representation of women in these parliaments does not reflect actual power of women. Moreover, female representation in parliament measures only representation at the national level and ignores women's participation at other levels of the state and in civil society. A similar problem is attached to the representation of women in senior economic positions that only measures formal sectors. In addition, this indicator does not fluctuate much over years. However, given that there is a lack of data available for women's representation at the local and societal level as well as for informal economic participation and to be comparable to other studies, we use both measures.

<sup>11</sup>Current data for the Polity IV Project can be found at

<http://www.systemicpeace.org/polity/polity4.htm>.

fewer countries than the Electoral democracy index.<sup>12</sup> Dollar et al. (2001), Swamy et al. (2001) and Sung (2003) use either the Civil Liberties index<sup>13</sup>, the Political Rights index or the Freedom of the Press index of the Freedom House project as regressors in their empirical analysis to measure or to refine the measurement of democracy. It needs to be stressed that these measures are not without methodological problems as they include questions about bribing and other forms of corrupt behavior and are therefore by construction correlated with corruption. The Civil Liberties index includes questions on corruption that restrains free and independent media. The Political Rights index includes questions related to corruption in government. The Freedom of the Press index includes questions on the impact of corruption and bribery on content of the press. Moreover, Sung (2003) uses a rule of law index that is also problematic as rule of law is closely related to the prevalence of corruption. Therefore, only the Electoral Democracy index of Freedom House (2008) is included in our regressions to account for democracy.

As additional controls we include:

- the log of GDP per capita in constant prices to control for the level of economic development as combatting corruption might be costly, and as poorer people might tend to engage more in corrupt activities (log GDP)<sup>14</sup> (Swamy et al., 2001);
- region dummies to capture geography and other unexplained regional heterogeneity, with Subsaharan Africa as the reference category (SA for South Asia, ECA for Europe and Central Asia, LAC for Latin America and Caribbean, EAP for East Asia and Pacific);
- ethnic fractionalization as it might increase corruption through clientelistic networks, identity politics and patronage along ethnic lines (e.g. Tripp, 2001) (Ethnic frac.);
- literacy rates to control for the knowledge of the population about laws against corruption, and as higher education might come along with less tolerance towards corruption (Swamy et al., 2001) (Literacy pop.);
- a measure of trade openness as trade barriers increase the incentives for corrupt behavior between individuals and customs officials (Ades and Tella, 1997; Gatti, 2004) (Openness);

---

<sup>12</sup>We use averages over ten years to capture stability of democracy. For the 121 countries for which both Electoral democracy and Polity2 are available, the Pearson Correlation Coefficient between them is 0.90 and significant.

<sup>13</sup>The Civil liberties index from Freedom House (2008) measures civil liberties in general and is not to be mixed up with the subindex civil liberties related to gender inequality.

<sup>14</sup>US\$, PPP, base year: 2005.

- a dummy indicating whether a country has never been a colony (Not colony) and a dummy measuring whether a country was a British colony (British colony) based on Correlates of War 2 Project (2003) as corruption might also be linked to the history of colonialism (Swamy et al., 2001).

The subindex civil liberties reflects the information available around the year 2000 and is not expected to change rapidly over time as social institutions are long-lasting and change only slowly and incrementally. For this reason, we use in the case of all other variables averages of the existing values over time to minimize the loss of observations due to missing values and to obtain a more stable value for the indicators used. For the corruption indicators representing our response variables we take averages over the years 2001 to 2005 for the CPI and over the period 2000-2004 in the case of the ICRG. For the other regressors we use averages over ten years (1996-2005), with the exception of ethnic fractionalization as changes in the ethnic composition of a country in less than 20 years are rare (Alesina et al., 2003). Concerning the two democracy variables, choosing averages over ten years has the advantage of capturing the stability of a democratic system, which has been highlighted by Treisman (2007) as important for corruption. In addition, having a difference of five years between response variable and the regressors might help to alleviate endogeneity and capture delays until possible effects can be observed.

### 3.2.2 Empirical Estimation

We empirically test with multiple linear regressions whether the subindex civil liberties  $s_i$ , which measures the freedom of social participation of women, is correlated with a response variable  $y_i$  capturing the level of corruption, after controlling for other factors that have been described in the literature as possible determinants of corruption.<sup>15</sup> As was discussed previously, we consider that social institutions related to gender inequality are relatively stable and long lasting. Therefore, we assume that they do not depend on the response variable for the period considered.<sup>16</sup>

We run regressions as

$$y_i = \alpha + \beta s_i + \text{control variables}_i + \varepsilon_i \quad (3.1)$$

<sup>15</sup>Before conducting the multiple linear regression analysis, we account for the importance of GDP for corruption. We first run a simple linear regression of each corruption measure on log GDP. We then compute the estimated residuals from this regression and use them as the dependent variable in a new simple linear regression where the subindex civil liberties is the only regressor. For both CPI and ICRG we obtain a negative and significant coefficient for the subindex civil liberties which suggests that the subindex is able to account for something that goes beyond GDP when explaining corruption.

<sup>16</sup>In general, social institutions, i.e. normative frameworks, change only slowly and incrementally.

using information at the country level. We are mainly interested in testing the null hypothesis that coefficient  $\beta$  is zero at a statistical significance level of 10%. The control variables included to attenuate omitted variable bias are described in Table 3.1. We acknowledge, however, that it is impossible to entirely rule out this problem.

To reproduce the findings from the literature, we first run a regression without the subindex civil liberties to focus on the effects of democracy and representation of women, which have been largely discussed. In a second step, we add to the regressions the subindex civil liberties as a measure of society's attitude towards gender inequality, as it can be argued that it is a variable that has been omitted in the previous regressions (Swamy et al., 2001). We run each specification for the two measures of corruption and use each time one of the two alternative measures of democracy. At the end, we present four regressions for each corruption indicator.

Preliminary regressions not reported here suggest that heteroscedasticity is a possible issue in our data and that there are influential observations that could drive the results. If our model is well specified, the OLS estimator of the regression parameters remains unbiased in the presence of heteroscedasticity, but the estimator of the covariance matrix of the parameter estimates can be biased and inconsistent, making inference about the estimated regression parameters problematic. Violations of homoscedasticity can lead to hypothesis tests that are not valid and confidence intervals that are either too narrow or too wide. To deal with heteroscedasticity, we run the regressions with OLS and 'heteroscedasticity-consistent' (HC) standard errors. As our sample sizes are less than 150, we use HC3 robust standard errors proposed by Davidson and MacKinnon (1993), which are better with small samples.<sup>17</sup>

For all the regressions, we check whether the results concerning the subindex civil liberties are stable in three ways. First, it is clear that in the multiple regressions, the estimate of the effect of our main variable, the subindex civil liberties, depends on the values of the other explanatory variables included (Mukherjee et al., 1998). We also try a simpler model to confirm that the estimated coefficient of the subindex civil liberties is negative and statistically significant. In this smaller model and based on the arguments presented before, we include as additional regressors the variables capturing the representation of women in so-

---

<sup>17</sup>Simulation studies by Long and Ervin (2000) have shown that HC standard error estimates tend to maintain test size closer to the nominal alpha level in the presence of heteroscedasticity than OLS standard error estimates that assume homoscedasticity. These authors recommend the use of HC3 robust standard errors, especially for sample sizes less than 250, as they can keep the test size at the nominal level regardless of the presence or absence of heteroscedasticity, with only a minor loss of power associated when the errors are indeed homoscedastic. We acknowledge that heteroscedasticity-consistent standard errors are not a panacea for inferential problems under heteroscedasticity. As pointed out by some authors, there are limitations and trade-offs in these estimators (e.g. Kauermann and Carroll, 2001; Wilcox, 2001).



ciety, a measure of democracy, the log GDP, religion dummies and regional dummies. This has the advantage that less parameters have to be estimated with the available observations.

Second, we use bootstrap with 1000 replications to compute a Bias-corrected and accelerated (Bca) 90% confidence interval of the regression coefficients computed with OLS to confirm that the value zero is not contained in the confidence interval around  $\beta$  (Efron and Tibshirani, 1993). One of the main advantages of bootstrapping methods is that one does not make any assumptions about the sampling distribution or about the statistic. Third, we detect observations with high influence or leverage based on the first estimates (OLS with standard variance estimator) using Cook's distance. Cook's distance is a commonly used estimate of the influence of a data point when doing least squares regression, and it measures the effect of deleting a given observation. We exclude the countries identified as outliers from the sample if the value of Cook's distance is larger than  $4/n$ , with  $n$  being the number of observations, and re-estimate regression 3.1 on the restricted sample using HC3 robust standard errors.

One should consider that possible endogeneity of the regressor  $s_i$  (the subindex civil liberties), meaning that  $s_i$  is correlated with the error term  $\varepsilon_i$  in the regression, might lead to an estimated coefficient of  $s_i$  that is biased. Endogeneity might arise due to omitted variables, measurement error and simultaneity (Wooldridge, 2002). The control variables included in the regression aim at minimizing omitted variable bias, albeit one cannot rule out this problem. We do not find it plausible that there are measurement errors in  $s_i$  which are related to the unobserved 'true' social institutions. Simultaneity could arise if  $s_i$  is determined simultaneously with the dependent variable  $y_i$ . As was discussed previously, social institutions related to gender inequality  $s_i$  are relatively stable and long-lasting. Hence, it is unlikely that the response variable  $y_i$  influences  $s_i$ .

### 3.2.3 Results

Results for the CPI as the first measure of corruption are presented in Table 3.6. Specifications (1) and (2) do not include the subindex civil liberties. In both specifications, none of the democracy variables Electoral democracy and Polity2 are significant. From the three measures of representation of women only Parliament is significant and positively related to corruption in specification (1) where Electoral democracy is the measure of democracy. Of the control variables only GDP has a significant and positive coefficient. In specifications (3) and (4) the subindex civil liberties is added as a new regressor to the former specifications. Its coefficient is negative and significant in both. Both democracy variables as well as the measures for participation of women in the economy are not significant. Only Parliament carries a positive and significant coefficient when Electoral democracy is used (specification

(3)). In the same specification (3) two control variables besides log GDP become significant: British colony and the regional dummy for ECA. For all four specifications the adjusted R square is around 0.5.

Table 3.7 shows the results when ICRG is used as the measure of corruption. For all 4 specifications (1-4), none of the variables reflecting representation of women and none of the democracy measures is significant. Interestingly, log GDP is also insignificant in all specifications, whereas it is always significant when the CPI is used as measure of corruption. Openness is the only control variable which is significant in all specifications. Important for the results of this paper, the subindex civil liberties is significant in specifications (3) and (4), and adding it to the corresponding regressions yields values for adjusted R-square that are noticeably larger than without it. It must be noted, however, that the obtained values for adjusted R-square for the regressions with the ICRG are lower than for the CPI (between 0.2 and 0.3 for the ICRG and around 0.5 for the CPI), suggesting that the model is not able to explain much of the variation of the political risk of instability due to corruption.

Using the smaller model yields similar results for the subindex civil liberties and the variables measuring representation of women and democracy.<sup>18</sup> These findings do also withstand the two other robustness checks. First, we confirm with Bias-corrected and accelerated (Bca) confidence intervals that in all cases the value zero is not contained in the 90% confidence interval around the regression coefficient of the subindex civil liberties. Second, excluding outliers (6 to 7 countries) and re-running specifications (3) and (4) for both corruption measures, the subindex civil liberties remains significant in all estimations. It is worth mentioning that for every restricted sample, the adjusted R-square is higher than in the corresponding complete sample.<sup>19</sup>

Summarizing the results, when we do not include the subindex civil liberties we find that from all variables for representation of women only Parliament is significant in the case of the CPI as long as Electoral democracy is used as measure of democracy. If one uses Polity2 instead, Parliament becomes insignificant. None of the democracy measures used turns out to be significant. When we include the subindex civil liberties, the results for representation of women and the democracy variables stay unchanged. Neither representation of women, except Parliament in the case of CPI when Electoral democracy is used, nor the democracy variables are significantly related to corruption. The main result concerning the subindex civil liberties is that even after controlling for democracy and for measures of political and

---

<sup>18</sup>With the smaller model we also run three separate regressions for (i) Muslim majority countries, (ii) Christian majority countries, and (iii) countries without Christian or Muslim majority. As could be expected, the subindex civil liberties is statistically significant only in the regressions with Muslim majority countries, as in other countries there is not enough variation in the subindex.

<sup>19</sup>Results for all the robustness checks are not reported here, but are available upon request.



economic participation of women as well as for other factors, we find a robust and significant relationship between the subindex civil liberties, which reflects society's attitude towards gender inequality, and the level of corruption. Social institutions favoring gender inequality are associated with higher levels of corruption.

### 3.3 Conclusion

The literature investigating the link between gender and corruption finds that there is a relationship between female representation in political and economic life and the level of corruption in a country. However, some studies warn that the observed relationship may be due to omitted variable bias. A possible variable that might influence both participation of women and corruption, is liberal democracy (e.g. Sung, 2003). We introduce a further omitted variable that has either been neglected in the literature or not been adequately dealt with because of insufficient data. Swamy et al. (2001) refer to this as the "level of discrimination against women" and proxy it with the gaps in educational attainment and life expectancy between men and women. We use the subindex civil liberties, which we consider a better proxy of the "level of discrimination against women" as it captures social institutions that restrict women in their freedom to participate in public life and reflect society's attitude towards gender inequality. The subindex measures underlying institutions and not outcomes of these institutions as do the variables used by Swamy et al. (2001).

When we replicate the findings of the literature for our sample of developing countries without the social institutions indicator, the results support the hypothesis of Sung (2003) and others that, when liberal democracy (in our case measured with Polity2) is considered in the regression, the representation of women in political and economic life is insignificant. However, Sung's hypothesis is weakened by the fact that there is no statistically significant association between democracy and corruption. Consequently, our statistical results support neither Sung's arguments nor the arguments put forward by Swamy et al. (2001) and Dollar et al. (2001) that representation of women is negatively related to corruption.<sup>20</sup> These results make it difficult to interpret social institutions related to gender inequality as an omitted variable when one investigates the relationship between representation of women in society, democracy and corruption.<sup>21</sup>

---

<sup>20</sup>Once again, our sample includes only developing countries, while the other studies include developed countries as well.

<sup>21</sup>We have estimated with multivariate regressions, not reported here, whether there is (1) a relationship between democracy and the subindex civil liberties and (2) a relationship between representation of women in society and the subindex civil liberties in our sample of developing countries, but did not find significant results.

Once we include the subindex civil liberties as a regressor, the main finding is that after controlling for representation of women in political and economic life and for democracy, it has a robust negative and significant relationship with corruption. In countries where social institutions inhibit the freedom of women to participate in social life, the level of corruption is higher.

Admittedly, one has to be cautious with these results. Interpretations for these findings in the light of the theories discussed are difficult, and country or regional studies are needed. Measurement is another relevant issue as the concepts of social institutions, democracy, participation of women and corruption are all hard to operationalize. The measures used in this study could be contaminated by measurement error. Finally, it cannot be ruled out that another factor, which has been neglected from the analysis, shapes the results.

Nevertheless, we derive one policy implication from this study, which should be mainly targeted at developing countries. In a context where social institutions deprive women of the freedom to participate in social life, neither political reforms towards democracy nor the representation of women in political and economic positions might be enough to reduce corruption. How women are treated in a society is not only important for them, but has major implications for the functioning of the whole society.

### 3.4 Tables

Table 3.1: Description and Sources of Variables

Variables	Definition	Source
<b>Measures of corruption</b>		
CPI	Corruption Perception Index (CPI); comprehensive measure of the level of corruption in a country that covers the different forms of grand and petty corruption in business, politics and administration. Ranges from 0 (high corruption) to 10 (low corruption). (average of existing values over the last five years)	Transparency International (TI)
ICRG	Corruption in Government Index; assesses corruption within the political system and focuses in particular on those types of corruption that lead to instability in the political system. Ranges from 0 (high risk) to 6 (low risk). (average of existing values over the last five years)	International Country Risk Guide (ICRG)
<b>Representation of women</b>		
Parliament	Proportion of seats held by women in national parliaments (%) (average of the existing values over the last 10 years)	World Bank (2009a)
Managers	Proportion of professional and technical, administrative and managerial positions held by women (%) (average of the existing values over the last 10 years)	World Bank (2009a)
Labor force	Female labor force participation rate (average of the existing values over the last 10 years)	World Bank (2009a)
<b>Democracy</b>		
Electoral democ.	Index that qualifies countries as electoral democracy when there exist competitive, universal and free and secret elections and a	Freedom House (2008)
Continued on next page		

Table 3.1 – continued from previous page

Variables	Definition	Source
Polity2	<p>multiparty system that can access the media for political campaigning, (average of the existing values over the last 10 years) Measure of democracy taking account of competitiveness of participation, institutions and procedures openness and competitiveness of executive recruitment and constraints on the chief executive, ranges from -10 (highly autocratic) to 10 (highly democratic) , score 0 means country is democratic (average of the existing values over the last 10 years)</p>	Marshall and Jaggers (2009)
<b>Social inst. related to gender ineq.</b> Subindex civil lib.	Subindex Civil liberties that captures the freedom of social participation of women	Branisa et al. (2009a)
<b>Control variables</b> log GDP	Log of GDP per capita, PPP (constant 2005 international \$) (average over the last 10 years)	World Bank (2008)
SA	Countries get a 1 if located in region South Asia, 0 otherwise.	
ECA	Countries get a 1 if located in region Europe and Central Asia, 0 otherwise.	
LAC	Countries get a 1 if located in region Latin America and the Caribbean, 0 otherwise.	
MENA	Countries get a 1 if located in region Middle East and North Africa 0 otherwise.	
Continued on next page		

Table 3.1 – continued from previous page

Variables	Definition	Source
EAP	Countries get a 1 if located in region East Asia and Pacific 0 otherwise.	
Muslim	Countries get a 1 if at least 50 % of the population are Muslim, 0 otherwise.	Central Intelligence Agency (2009)
Christian	Countries get a 1 if at least 50 % of the population are Christian, 0 otherwise.	Central Intelligence Agency (2009)
Muslim percentage	percentage of Muslim population	Pew Forum on Religion and Public Life (2009)
Ethnic frac.	The ethnic fractionalization measure gives the probability that two individuals selected at random from a population are members of different groups. It is calculated with data on language and origin. The value 0 means complete homogeneity and 1 complete heterogeneity.	Alesina et al. (2003)
Literacy pop.	Literacy rate for the whole population (average of the existing values over the last 10 years)	Human Development Report (HDR) stats office
Openness	Imports of goods and services (% of GDP)	World Bank (2008)
Not colony	Countries get a 1 if never colonized, 0 otherwise.	Correlates of War 2 Project (2003)
British colony	Countries get a 1 if former British colony, 0 otherwise.	Correlates of War 2 Project (2003)

Table 3.2: Descriptive statistics of variables used

Variable	N	mean	sd	min	max
<b>Measures of corruption</b>					
CPI	115	3.17	1.37	1.22	9.32
ICRG	97	2.17	0.74	0.25	4.32
<b>Representation of women</b>					
Parliament	119	10.76	7.03	0.00	29.56
Managers	120	7.98	5.26	0.00	23.70
Labor force	122	55.10	16.75	10.96	92.96
<b>Democracy</b>					
Electoral democ.	121	0.45	0.46	0.00	1.00
Polity2	98	1.09	6.08	-9.00	10.00
<b>Social inst. related to gender ineq.</b>					
Subindex Civil lib.	124	0.16	0.26	0.00	1.00
<b>Control Variables</b>					
log GDP	116	7.98	1.12	5.61	10.55
SA	125	0.06	0.23	0.00	1.00
ECA	125	0.14	0.34	0.00	1.00
LAC	125	0.18	0.38	0.00	1.00
MENA	125	0.14	0.35	0.00	1.00
EAP	125	0.14	0.35	0.00	1.00
Muslim	125	0.33	0.47	0.00	1.00
Christian	125	0.43	0.50	0.00	1.00
Muslim percentage	121	33.38	39.65	0.00	99.70
Ethnic frac.	121	0.51	0.24	0.04	0.93
Literacy pop.	122	0.74	0.22	0.17	1.00
Openness	120	0.45	0.26	0.01	1.91
Not colony	121	0.21	0.41	0.00	1.00
British colony	121	0.30	0.46	0.00	1.00

Table 3.3: Pearson Correlation Coefficient between subindex Civil liberties and control variables

log GDP	$\rho$	0.20	Muslim	$\rho$	0.57
	p-value	0.036		p-value	0.000
	Number of obs.	114		Number of obs.	123
SA	$\rho$	0.33	Christian	$\rho$	-0.40
	p-value	0.000		p-value	0.000
	Number of obs.	123		Number of obs.	123
ECA	$\rho$	-0.25	Ethnic	$\rho$	0.08
	p-value	0.006		p-value	0.392
	Number of obs.	123		Number of obs.	119
LAC	$\rho$	-0.29	Literacy population	$\rho$	-0.19
	p-value	0.001		p-value	0.039
	Number of obs.	123		Number of obs.	120
MENA	$\rho$	0.53	Openness	$\rho$	-0.07
	p-value	0.000		p-value	0.447
	Number of obs.	123		Number of obs.	118
EAP	$\rho$	-0.11	Not colony	$\rho$	-0.06
	p-value	0.221		p-value	0.549
	Number of obs.	123		Number of obs.	119
Muslim percent.	$\rho$	0.54	British colony	$\rho$	0.36
	p-value	0.000		p-value	0.000
	Number of obs.	120		Number of obs.	119

Table 3.4: Pearson Correlation Coefficient ( $\rho$ ) between the Corruption Measures

		CPI	ICRG
<b>CPI</b>	$\rho$	1	
	obs	115	
<b>ICRG</b>	$\rho$	0.58	1
	p-value	0.0000	
	obs	93	97

Table 3.5: Variation of the subindex civil liberties over religion

Subindex civil liberties	No christian/ Muslim majority	Christian majority	Muslim majority	Total
0	22	46	15	83
0.298	5	8	1	14
0.301	1	0	4	5
0.599	1	0	15	16
0.781	0	0	2	2
0.818	0	0	1	1
1	0	0	2	2
Total	29	54	40	123



Table 3.6: Linear regressions with dependent variable CPI

	Model 1	Model 2	Model 3	Model 4
<b>Representation of women</b>				
Parliament	0.031*	0.033	0.032*	0.037
	(0.018)	(0.023)	(0.018)	(0.023)
Managers	0.025	0.022	0.011	0.006
	(0.029)	(0.032)	(0.031)	(0.034)
Labor force	0.007	0.009	0.001	0.004
	(0.009)	(0.010)	(0.010)	(0.011)
<b>Democracy</b>				
Electoral democ.	0.339		0.263	
	(0.234)		(0.231)	
Polity2		0.039		0.032
		(0.025)		(0.023)
<b>Social inst. related to gender ineq.</b>				
Subindex Civil lib.			-1.730***	-1.624*
			(0.593)	(0.866)
log GDP	0.710***	0.738***	0.766***	0.821***
	(0.197)	(0.212)	(0.193)	(0.209)
Muslim	-0.367	-0.271	0.049	0.107
	(0.319)	(0.394)	(0.305)	(0.363)
Christian	-0.392	-0.240	-0.280	-0.131
	(0.288)	(0.341)	(0.283)	(0.329)
Ethnic frac.	-0.334	-0.364	-0.267	-0.124
	(0.628)	(0.824)	(0.595)	(0.809)
Literacy pop.	-0.928	-1.122	-0.470	-0.831
	(1.070)	(1.193)	(1.009)	(1.091)
Openness	1.457	1.752	1.199	1.455
	(1.106)	(1.435)	(1.063)	(1.378)
Not colony	0.135	0.146	0.331	0.197
	(0.315)	(0.410)	(0.300)	(0.362)
British colony	0.478	0.313	0.611**	0.407
	(0.298)	(0.391)	(0.298)	(0.387)
constant	-3.305**	-3.455*	-3.364**	-3.809*
	(1.634)	(1.964)	(1.687)	(2.108)
Number of obs.	103	86	103	86
R2	0.576	0.580	0.613	0.607
Adjusted R2	0.491	0.474	0.530	0.501
Prob > F	0.000	0.000	0.000	0.000

HC3 robust standard errors in brackets.

Regional dummies included in all estimations.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 3.7: Linear regressions with dependent variable ICRG

	Model 1	Model 2	Model 3	Model 4
<b>Representation of women</b>				
Parliament	0.015 (0.018)	0.012 (0.020)	0.016 (0.014)	0.016 (0.017)
Managers	0.025 (0.021)	0.025 (0.021)	0.010 (0.017)	0.011 (0.019)
Labor force	-0.003 (0.007)	-0.000 (0.008)	-0.009 (0.007)	-0.006 (0.008)
<b>Democracy</b>				
Electoral democ.	0.273 (0.234)		0.221 (0.223)	
Polity2		0.029 (0.025)		0.027 (0.025)
<b>Social inst. related to gender ineq.</b>				
Subindex Civil lib.			<b>-1.488*** (0.425)</b>	<b>-1.260** (0.604)</b>
log GDP	0.122 (0.149)	0.081 (0.182)	0.153 (0.135)	0.123 (0.166)
Muslim	-0.337 (0.293)	-0.229 (0.316)	0.076 (0.261)	0.070 (0.315)
Christian	-0.351 (0.272)	-0.321 (0.338)	-0.300 (0.257)	-0.289 (0.333)
Ethnic frac.	0.507 (0.427)	0.349 (0.465)	0.655 (0.410)	0.652 (0.496)
Literacy pop.	-0.165 (0.930)	0.118 (0.988)	0.404 (0.769)	0.436 (0.873)
Openness	1.277** (0.625)	1.523** (0.650)	0.991* (0.588)	1.274** (0.596)
Not colony	0.033 (0.237)	0.122 (0.304)	0.255 (0.308)	0.177 (0.396)
British colony	-0.022 (0.228)	-0.055 (0.289)	0.131 (0.210)	0.067 (0.293)
constant	0.474 (1.082)	0.529 (1.193)	0.461 (0.924)	0.351 (1.094)
Number of obs.	86	72	86	72
R2	0.361	0.423	0.462	0.482
Adjusted R2	0.201	0.241	0.318	0.306
Prob > F	0.005	0.001	0.000	0.001

HC3 robust standard errors in brackets.

Regional dummies included in all estimations.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$





## **Part II**

# **Regional growth convergence in Colombia**



## **Essay 4**

# **Revisiting the regional growth convergence debate in Colombia using income indicators**

### **Abstract**

This paper investigates growth convergence across Colombian departments during the period of 1975 to 2000, following both the regression and the distributional approaches suggested in the literature, and using two income measures computed by Centro de Estudios Ganaderos (CEGA). We also discuss issues related to data provided by Departamento Administrativo Nacional de Estadísticas (DANE) used by previous convergence studies. Our results show no evidence supporting convergence using per capita gross departmental product, but rather persistence in the distribution. Using per capita gross household disposable income, we find some evidence of convergence, but only at a low speed, close to one percent per year. Furthermore, we find no evidence of the existence of different steady states for the two variables considered.

## 4.1 Introduction

One of the most interesting and disputed questions in the economics discipline during the last half century has been whether or not poor countries tend to catch up with wealthier ones over time or if, on the contrary, the gap between the rich and poor widens. This question also reflects an interest in understanding the distribution of outcomes across countries and, implicitly, the determinants of growth (Durlauf et al., 2005).

Empirical research on this topic is based upon macroeconomic aggregates and has concentrated on testing the neoclassical growth model of Solow (1956) using the estimation method proposed by Barro and Sala-i-Martin (1991) to investigate whether economies with lower capital per person at a certain initial point in time tend to grow more quickly than economies with higher capital per person. If this is the case, there is convergence among economies over the long run.

The convergence question has also been studied within particular countries to analyze how much regional disparities diminish over time. The difference with cross-country convergence analysis is that in such cases it is risky to make assumptions across countries on key model parameters, such as technology, savings, and population growth rates. On the contrary, within a single country, it is plausible to assume that regions exhibit similarities in these and other variables, such as language, institutions, and preferences. This presumed homogeneity has lead researchers to assume that convergence is more likely to hold within, rather than across, countries (Barro and Sala-i-Martin, 2004).

Empirical research supports regional convergence within industrial countries over the long run. Typical examples are given by Barro and Sala-i-Martin (1992b) who find convergence across U.S. states between 1880 and 2000, across Japanese prefectures between 1930 and 1990, and between regions in eight European countries between 1950 and 1990 (see also Barro and Sala-i-Martin (1992a)).

In the case of Colombia, a heterogeneous country at the department level in economic, geographic, and cultural aspects, existing research is contradictory. While some authors argue that Colombia was a successful case of convergence in the second half of the twentieth century, others argue for the persistence of regional disparities.

The objective of this study is to investigate whether or not Colombia was a case of convergence at the department level between 1975 and 2000 using two different income variables: gross departmental product and gross personal disposable income. We consider that the second variable is more appropriate for measuring convergence in well-being.

The study is constructed around three main questions. First, the study evaluates whether departments converged between 1975 and 2000 and if so, if convergence results obtained using the regression approach contradict the results obtained with the distributional approach



suggested by Quah (1997), using bivariate Kernel density estimators. Second, we determine if the assumption of a common steady state for all departments holds or whether there is evidence of heterogeneity in the model parameters. Finally, the study evaluates whether the presence or absence of convergence occurs simultaneously in gross departmental product and in gross personal disposable income.

An important contribution of the study is the first ever test of the convergence hypothesis using time-series cross-sectional Colombian data with different specifications to check the robustness of results. The results are based upon data from Centro de Estudios Ganaderos (CEGA) because those data provide the longest time series (25 years) computed with a consistent methodology.<sup>1</sup>

To summarize our results, we do not find convergence in gross departmental product and find no evidence of different steady states across departments using that variable. When using gross personal disposable income, we find convergence, but a very slow one, and no evidence of different steady states. For both variables, when using the regression approach, we find that the best estimators can be achieved using pooled time-series cross-section data and assuming homogeneity in the parameters. Furthermore, considering both variables, we do not find a contradiction in results obtained using the regression and the distributional approaches. Using bivariate kernel density estimators, we find persistence in the distribution of gross departmental product and slight convergence in gross personal disposable income.

One important policy implication of our results is the need to periodically review whether or not departmental disparities diminish over time based on consistent time series constructed under a single methodology. We explicitly warn that linking different time series computed with different methodologies can lead to incorrect conclusions for interventions, such as poverty-alleviating policies and growth strategies. In keeping with previous studies on this topic (e.g. Bonet and Meisel, 2006a), we consider important the need to have an explicit regional policy in Colombia to foster growth in departments lagging behind national averages, after conducting case studies to assess which policies could be most effective in each case.

## **4.2 Motivation and Background**

### **4.2.1 Economic Background**

One remarkable characteristic of Colombia is the large income inequality which exists at different levels-between individuals, between rural and urban areas, and between departments. The country is currently divided into 32 departments and the capital district of Bogotá. De-

<sup>1</sup>CEGA was a large research center financed by a private financial institution in Colombia.

partments may also be grouped into 5 regions: the Caribbean Region comprising departments with access to the Caribbean Sea; the Pacific Region, with departments in the west coast to the Pacific Ocean; the Central Region, covering the three branches of the Andes mountain chains; Orinoquia, comprising large plains to the south-east of the country; and Amazonia in the south, comprising the Colombian part of the Amazon rainforest (see the map of Colombia in Figure 4.1).

Economic growth over the last 30 years, which was low but stable compared to other countries in the region, comes together with a combination of a high incidence of poverty, inequality, and violence. In 2004, the percentage of people living below the poverty line (headcount index) was 52 percent and the Gini coefficient was 0.58. The homicide rate was 63 per 100,000 people. Evidence shows that growth slowed compared to long-term historical trends after 1970. In fact, after having achieved in 1970 a growth rate of 3.1 percent in per capita gross domestic product, growth between 1980 and 1990 occurred at an average annual rate of only 1.2 percent due primarily by the adverse effects of Latin America's debt crisis. In the 1990s, the average growth rate was similar (1.1 percent), driven by a boom and bust cycle throughout the decade, which concluded in a severe recession in 1999 (per capita GDP contracted by 5.5 percent (Table 4.1)). On the contrary, between 2002 and 2007, favorable external conditions, especially high commodity prices and confidence due to the easing of internal conflict, have contributed to the acceleration of the economy (Tenjo G. and López E., 2003; Cárdenas, 2007).

The heart of economic activity in Colombia lies in the Central or Andean Region which concentrates the largest proportion of population within the major cities. Bogotá and the departments of Cundinamarca and Antioquia account for 42 percent of total GDP with Bogotá having a high level of participation in total production (22 percent). This area concentrates not only manufacturing industry and commerce near the cities, but also coffee plantations and other large-scale agricultural areas.

The GDP of departments in the Caribbean Region is based upon mining, small-scale agriculture, and cattle farming. La Guajira and Cesar are the two largest producers of coal, while Córdoba is the largest nickel producer. Despite having some departments rich in minerals, this region nevertheless has a high incidence of poverty, particularly in Córdoba and Sucre.

The Pacific Region comprises, relative to the Colombian average, three poor departments and one wealthy one (Valle del Cauca). Chocó, which is the poorest department in this region and in the country, is predominantly rural and sparsely populated, with large tropical rain forests and humid areas. It is known as the rainiest area in the country (and even one of the rainiest worldwide) and is geographically isolated from the rest of the country due to a chain of mountains to the east and the ocean to the west. Transport of population living

in the department is largely done by way of its abundant creeks and rivers; road infrastructure is minimal. The scarce literature explaining socio-economic factors in this department argues that the current distribution of population and the quality of institutions may largely be explained by the early settlement of an extractive economy during colonization, at which time colonizers brought slaves to exploit gold mines but did not establish themselves in the department (Bonet, 2007). As opposed to Chocó, Valle del Cauca is the third largest departmental economy in the country after Bogotá and Antioquia and has some of the most productive agricultural areas, as well as a high level of participation in the manufacturing sector.

During the last 30 years, production was driven in some departments by the discovery of important mineral resources, as is the case for the departments of Arauca and Casanare, which have the largest oil fields in the country.<sup>2</sup> The same applies for La Guajira, which has the largest open coal mine in Latin America.

According to Meisel (2007b), the burden of poverty in Colombia is geographically located in the coastal departments and inequality is greater between departments than within them. Meisel argues that the urban-versus-rural divide is not the relevant dimension upon which to design poverty-alleviating programs, but the departmental one. Moreover, Meisel affirms that the already-large disparities have increased over the past 15 years and will not spontaneously disappear merely as a result of market forces.

The level of empirical research addressing regional disparities in Colombia has increased gradually since the early nineties, inspired by the international debate on convergence and the methodology proposed by Barro and Sala-i-Martin (1991). Since then, approximately 20 papers have investigated whether departments, regions, or even major cities have converged over time. Important shortcomings in this field arise due to the absence of consistent time-series data allowing for a long-term perspective. As a consequence, results frequently depend upon how the researcher combined the available time series, as well as on the methodology and control variables used, with no robust and undisputed evidence concerning departmental convergence.

Debate in this field in Colombia revolves around two issues: first, a methodological discussion as to whether or not to rely on the methodology proposed by Barro and Sala-i-Martin (1992a) or on the distributional approach proposed by Quah (1993b), and second, whether

---

<sup>2</sup>These departments are included in our sample as one group named Nuevos Departamentos (Nuevos), meaning new departments. The so-called New Departments are distributed in the south-east lowland plains, the Amazon Region, and the Caribbean islands. Excepting the islands, these departments are large in extension but have low population densities.

one should use information generated by Departamento Nacional de Estadísticas (DANE), rather than by Centro de Estudios Ganaderos (CEGA).<sup>3</sup>

Early studies used Barro-type regressions. The pioneer work of Cárdenas and Pontón (1995), combining early GDP data by department from the National Planning Department with those produced by DANE, concluded that between 1950 and 1990, Colombia was a successful case of convergence with a 4-percent speed of convergence, and that migration played an insignificant role in convergence. Alternative combinations of data from DANE yield different results, despite using the same methodology. For instance, Barón (2003) finds convergence during the eighties but not during the nineties. Research using kernel density estimators concluded that Colombia was a case of polarization with the existence of three groups: a wealthy one that diverges from the average national income, a middle income one that shows convergence inside the group, and a third one that grows more impoverished over time (Birchenall and Murcia, 1997). Using CEGA data, research points to polarization in favor of the capital district of Bogotá, to the detriment of departments located in the peripheries (Bonet and Meisel, 2006b). Almost all studies focus only upon convergence in income, while only three ask for convergence in living standards using social indicators.<sup>4</sup>

The reader is then confronted with the question of whether Colombia is a successful case of convergence or rather an example of hopeless persistence unless strong regional redistributive policies are adopted.<sup>5</sup>

Intuitively, when observing the different geographic conditions of the country and the agglomeration processes around the largest cities, as well as the differences in infrastructure, it is unrealistic to expect that poor departments can catch up with leading departments in terms of per capita product, given that they lack basic infrastructure and have a minor manufacturing and government presence.

However, there are mechanisms that could have promoted convergence among departments in recent years. One of them is fiscal equalization through central government transfers. Starting in the mid eighties, the government implemented a decentralization program to reduce the burden of spending by the central government. This process accelerated with the new constitution of 1991 which established a new system of transfers in order to increase the efficiency of social expenditures, as well as the supply of social services, compensated municipalities with weak financial capacities, and increased political power and the participation of local governments in the implementation of health and educational policies (Departamento

---

<sup>3</sup>DANE is the official statistical agency in Colombia (<http://www.dane.gov.co/>).

<sup>4</sup>A comprehensive list of convergence studies in Colombia can be found in Aguirre (2008). We deal with regional convergence in social indicators in Colombia in Essay 5.

<sup>5</sup>Research using alternative methodologies and looking for linkages among regions found that Colombia has limited spatial interdependency (Haddad et al., 2008).

Nacional de Planeación DNP, 2002; Rojas, 2003; Barrera and Domínguez, 2006). As a result, social spending increased from 7 to 15 percent between 1991 and 2001. Moreover, starting in the eighties, the model of industrialization through import substitution changed to the policy of liberalization of the economy, reduction of tariffs, and integration into the world markets in order to increase competitiveness, productivity, and economic growth. This shift also accelerated after the constitutional reform.

Another possible mechanism for convergence is migration. In general, the country underwent an important urbanization process in recent decades encouraged by industrialization around urban centers and a higher orientation towards export markets. Labor mobility was a combination between voluntary migration for economic gain (which profited from increasing returns to scale in the manufacturing sector) and forced migration due to violence, which migration helped enlarge informal markets. Migration from rural to urban areas accelerated during the twentieth century. The percentage of population which is urban changed from 59 percent in 1973 to 75 percent in 2005 due not only to a transformation from a predominantly agriculture-based economy to a services and industry-based one, but also due to conflict, violence, and a lack of opportunities in rural areas (Murad R., 2003).

In this context, the substantive question we try to empirically answer in this study is whether or not Colombia was a case of convergence at the department level between 1975 and 2000. Thus, if poor departments had greater growth rates than wealthy ones over time and the gap between them decreased. Our interest relies upon closing the debate on the existence of convergence across departments in Colombia by analyzing methodological issues and data sources that may have had affected results up to now. One important motivation of this study is the policy implication that can result as a consequence of wrongly assuming that departments converge automatically over time.

In order to explain the importance of the data used and the possible combinations of time series, we explain in the next subsection the available data sources and the relevance of two variables, gross departmental product and gross personal disposable income, arguing that the second one is more appropriate for measuring convergence for well-being.

#### **4.2.2 Data Issues Affecting Convergence Results in Colombia**

There are two different data sources of departmental accounts in Colombia: Department of Statistics (DANE) and Centro de Estudios Ganaderos (CEGA).

DANE provides per capita GDP by department for three different periods: one for 1980 through 1996 in constant prices as of 1975, one for 1990 through 2005 in constant prices as of 1994, and a final one for 2000 through 2005 in constant prices as of 2000. The first period was calculated applying concepts of the System of National Accounts of 1986 (SNA-86)

and used an indirect method for collecting information. The second period was calculated using the System of National Accounts of 1993 (SNA-93) and combined direct and indirect methods for collecting information.<sup>6</sup> The third period did not include illicit crops in its estimation and is also based upon SNA-93.

The classification of sectors, transactions, concepts, and methodology changed considerably in the SNA-93 and allowed for the inclusion of illegal activities as part of the GDP (DANE, 2008).<sup>7</sup> It must be noted that statistical offices use different techniques to produce consistent time series of national accounts, particularly when international guidelines change.<sup>8</sup> For instance, most (OECD) countries make regular revisions for short time periods (usually of about twenty years) to incorporate new available information and benchmark revisions, in order to provide users with consistent time series. In Latin America, only Chile and Perú offer consistent large time series of regional per capita GDP using statistical or interpolation methods (Serra et al., 2006).

In Colombia, DANE collected information for some overlapping years using both methodologies, but did not construct a consistent time series based only upon one. Although users do not have enough information to consistently recompute long time series, they tend to rebase series and connect them using growth rates, which can be problematic.<sup>9</sup>

Comparison of the series for the overlapping years shows different departmental growth rates and a different evolution of the logarithm of the standard deviation, explaining why convergence results change depending on how and when the researcher linked the different data series. Note in Figure 4.6 that the annual standard deviation of the logarithm of GDP of the three series of DANE yields different patterns. In the series of 1980 to 1996, the standard deviation increases sharply starting in 1990, while in the series from 1990 to 2005, it remains close to 0.36 until 1997 and falls thereafter. Concerning the third time series (2000 to 2005), the trend is similar to the series for 1990 through 2005, but the level of the standard deviation is higher.

---

<sup>6</sup>Direct methods take departmental information by product whenever data sources are available. Indirect ones use national aggregates and assign each department a percentage of those aggregates.

<sup>7</sup>The main changes concern the measurement of value-added taxes, the reclassification of transactions in the government sector, changes to the capital account, and productivity levels for the banking, energy, and insurance sectors.

<sup>8</sup>Techniques can be broadly classified into four groups: detailed reworking, proportion methods, interpolation between benchmarks, and indicator methods.

<sup>9</sup>For instance, the Canadian statistical office explicitly prohibits users from simply rebasing series using growth rates due to the large methodological differences derived from changing to SNA-93, and argues that only the statistics office in charge may compile series using detailed accounting and recomputing information according to the new procedures (Lal, 1999).



The CEGA project compiled information at the departmental level in Colombia from 1975 to 2000 using SNA-93 and presented a simplified system of national accounts. The project used mixed methods for collecting information, classified some particular products differently than DANE, and did not include illicit crops in the agriculture category. Departmental results coincide between CEGA and DANE from 1990 onwards because both use SNA-93 (there are, however, important differences before 1990).

CEGA produced consistent time series of two key variables relevant for convergence analysis, gross departmental product, which we will call henceforth PDB, and gross departmental income, which we will refer to as IDB. The first variable reflects production by residents in each department, while the second reflects the primary income received by those residents. The difference between these variables is the net external income of residents.

CEGA also provided time series of gross household disposable income by department, which we will call IDBH, and which is the result of households' income after subtracting taxes on property and rental income and net payments to the social security, and adding other net current transfers. This variable is a more accurate measure of a population's welfare than per-capita PDB, as it reflects household income after taxes and net public and private transfers.<sup>10</sup>

Due to the advantages of the CEGA database, and as this database provides the only two consistent time series covering a long time span, we present results and discussion on convergence for per capita PDB and IDBH.<sup>11</sup> Our final data set covers the period of 1975 to 2000 for 23 departments, the capital district of Bogotá, and the nine "New Departments" grouped into one observation, for a total of 25 units and 25 years.<sup>12</sup>

To calculate yearly per capita figures of both PDB and IDBH, we use the latest available population data, computed after reconciliation of the census of 2005 with previous censuses. According to the census of 2005, population is less than what had been forecasted using the 1993 census due to a lower birth rate and increased external migration (DANE, 2007). Although in most of the departments population was overestimated, there are some particular cases in which the contrary situation applies.

We use yearly population data at the departmental level from DANE (2007) for the years 1985 to 2000, and for the years 1975 to 1985, we interpolated departmental population using

<sup>10</sup>The abbreviations used refer to the original names in Spanish: Producto Departamental Bruto (PDB), Ingreso Departamental Bruto (IDB) and Ingreso Departamental Bruto (disponible) de los Hogares (IDBH).

<sup>11</sup>As will be explained in the next section, it would be best to work with data expressed as per unit of effective worker, but due to data availability, researchers often use per capita figures.

<sup>12</sup>The New Departments have existed formally since the 1991 constitutional reform when nine former intendancies and commissariats were acknowledged as departments (Amazonas, Arauca, Casanare, Guainía, Guaviare, Putumayo, San Andrés y Providencia, Vaupés, and Vichada).

the annual growth rate from 1973 to 1985 based on the 1973 census. The obtained values show a consistent evolution of population by department once connected to the official estimates from 1985 onwards.

Box plots of per capita PDB and IDBH in logs are shown in Figures 4.2 and 4.3. Box plots of relative PDB and relative IDBH in logs are shown in Figures 4.4 and 4.5. By relative we mean that the variables are expressed as ratios to the national average of the corresponding year. We can see that the ordering of departments is similar in both types of graphs, in levels and relatively, particularly in the upper and lower ends. The five departments with the lowest per capita PDB are Chocó, Sucre, Córdoba, Nariño, and Cauca, four of these being located on the Pacific Coast. Bogotá, Valle, Antioquia, Nuevos Departamentos and Cundinamarca have the five highest PDBs. Concerning per capita IDBH, departments with the lowest values are almost the same, excepting Santander instead of Nuevos.

The box plots show large variability in per capita figures of Guajira in both PDB and IDBH and low variability for Bogotá. This pattern is accentuated in figures relative to the average, as well as for the group of Nuevos and observing PDB. On the contrary, the log of per capita IDBH shows less variation and dispersion of values, but a higher difference between the richest and the poorest departments. Note also that the group of Nuevos Departamentos has large variability in PDB. That variability is not visible in IDBH. In the following two sections, we present two well-known approaches for testing for convergence—the classical approach to convergence analysis and the distributional approach.

## 4.3 The Solow Model and Its Estimation

### 4.3.1 The Solow Model

Empirical testing of convergence across economies is based upon the neoclassical growth model developed by Solow (1956)<sup>13</sup> in which economies have a transition dynamic towards the *steady state*, defined as a situation in which all variables per unit of effective worker remain unchanged over time. In the steady state, the ratio of capital to labor is constant given that the capital stock expands at the same rate as the labor force, and the capital expansion is sufficient to compensate for it.

The neoclassical growth model assumes diminishing returns to factors and constant returns to scale. Due to this assumption, real returns of factors adjust to bring about full employment of labor and capital. Technology is exogenous and is the only force that explains changes in output and capital per worker. Any capital-to-labor ratio different than the one

<sup>13</sup>The neoclassical model was also developed in the original works of Ramsey (1928) and Cass (1965).



needed in the steady state readjusts as time passes so that economies tend towards the steady state. The speed at which this happens is known as the convergence rate and is inversely related to the distance from the steady state (Durlauf, 1996).

Under the framework of the neoclassical growth model, smaller initial values of the capital-to-labor ratio  $k$ , are associated with greater growth rates of the ratio production per worker. Robert Barro and Xavier Sala-i-Martin tested whether economies with lower capital per worker at a certain initial point in time grew more quickly in per-worker terms (Barro and Sala-i-Martin, 1991, 1992a,b, 2004; Sala-i-Martin, 1996), using the following equation:

$$\log[\hat{Y}(t)] = (1 - \exp^{-\beta^*t}) \log(\hat{Y}^*) + \exp^{-\beta^*t} \log[\hat{Y}(0)], \quad (4.1)$$

where  $t$  represents time,  $\beta^*$  indicates how rapidly an economy's output per effective worker  $\hat{Y}$  approaches its steady-state value  $\hat{Y}^*$  in the neighborhood of the steady state. The corresponding definition of  $\beta^*$  with a constant saving rate  $s$  is  $\beta^* = (1 - \alpha)(x + n + \delta)$ , where  $\alpha$  is a constant representing the share of capital in production,  $n$  is the rate of population growth,  $x$  is the rate of exogenous growth, and  $\delta$  is the depreciation rate. The speed of convergence is measured by how much the growth rate decreases as the capital stock increases in a proportional manner.<sup>14</sup> Equation 4.1 implies that the average growth rate of per-capita output  $Y$  over an interval from an initial time 0 to any future time  $T$  (higher than 0) is

$$\frac{\log[Y(T)/Y(0)]}{T} = x + \frac{(1 - \exp^{-\beta T})}{T} \log[\hat{Y}^*]/\hat{Y}(0), \quad (4.2)$$

where  $x$  is the rate of technological progress or the steady-state growth rate.<sup>15</sup> Equation 4.2 also shows that the effect of the initial position  $\hat{Y}(0)$  is conditioned on the steady-state position  $\hat{Y}^*$  (conditional convergence) (Barro and Sala-i-Martin, 2004). The approach suggested by Barro and Sala-i-Martin (2004) is known as the regression approach or as the classical approach to convergence analysis (Sala-i-Martin, 1996; Magrini, 2004). There are two alternatives for applying this concept – testing for absolute convergence or for conditional convergence.

### 4.3.2 Absolute Beta-Convergence

The concept of absolute beta-convergence (also known as unconditional convergence) is relevant for a group of closed economies that are structurally similar; they have the same

<sup>14</sup>Note that  $\beta^*$  is not the same as  $\hat{\beta}$ . It is the convergence rate in the proximity of the steady state and is determined by  $(1 - \alpha)$  for given values of  $x$ ,  $n$ , and  $\delta$ .

<sup>15</sup>Equation 4.2 indicates that the coefficient  $(1 - \exp^{-\beta T})/T$  declines, the higher  $T$  is for a given  $\beta$ , and as long as  $\beta$  is positive. Therefore, the average growth rate of  $Y$  decreases as  $T \rightarrow \infty$  (and thus  $x$ ) dominates the average growth rate. In contrast, for a given  $T$ , a higher  $\beta$  implies a higher coefficient  $(1 - \exp^{-\beta T})/T$ .

values of the parameters  $x$ ,  $s$ ,  $n$ , and  $\delta$ , and thus they have the same production function steady-state values  $k^*$  and  $Y^*$ . The only difference is the initial quantity of capital per person  $k(0)$ , which reflects past disturbances (wars, transitory shocks to production, etc.). Hence, economies with lower values of  $k(0)$  and  $Y(0)$  have unambiguously greater growth rates of  $k$  and  $Y$ . The estimation equation for absolute convergence is equation 4.2, omitting the  $\hat{Y}^*$  term:

$$\frac{\log[Y_{i,t}/Y_{i,t-T}]}{T} = a - \frac{(1 - \exp^{-\beta T})}{T} \log[Y_{i,t-T}] + w_{it,T}, \quad (4.3)$$

where  $w_{it,T}$  represents the effect of the error terms  $w_t$  between dates  $t$  and  $T$ ,  $i$  is the corresponding subscript for each region or country, and  $a = x + (1 - \exp^{-\beta T}) \log(\hat{Y}^*)$ . Absolute convergence arises when the term multiplying the initial income is negative, and implies that poor economies tend to grow more quickly than wealthy ones. One can estimate a regression with non-linear least squares (NLLS) to obtain the speed of convergence  $\beta$  directly.

### 4.3.3 Conditional Convergence

Conditional beta-convergence arises by allowing for heterogeneity across economies, particularly by dropping the assumption that all economies have the same parameters and the same steady state.<sup>16</sup> The main idea is that the further an economy is from its own steady-state value, the more quickly it grows:

$$\frac{\log[Y_{i,t}/Y_{i,t-T}]}{T} = a - \frac{(1 - \exp^{-\beta T})}{T} \log[Y_{i,t-T}] + \gamma X_i + w_{it,T}, \quad (4.4)$$

where  $X_i$  is a set of variables that proxy for the steady-state level of income ( $\hat{Y}_i^*$ ). Empirical studies show little evidence of unconditional convergence for large and heterogeneous samples of countries. Instead, they tend to find conditional convergence in economies with similar structural characteristics (Barro and Sala-i-Martin, 1991) with speeds of convergence usually around 2 percent. However, there is no agreement on which variables to include as proxies for the steady state, and their selection depends mostly upon the researcher interest. An extensive review made by Durlauf et al. (2005) shows a list of about 145 different regressors used in convergence literature and points out that most of them have been found to be statistically significant. These regressors are classified by Durlauf et al. (2005) into 43 distinct growth theories or growth determinants, raising doubts about their usefulness.

<sup>16</sup>Under the assumption of different parameters, Equation 4.3 would provide biased estimates because the steady-state level of income  $\hat{Y}_i^*$  would be correlated with the explanatory variable  $\log[Y_{i,t-T}]$ . To solve this problem, Barro and Sala-i-Martin (1992a) suggest incorporating into the regression a set of variables  $X_i$  as proxies for the steady-state level of income ( $\hat{Y}_i^*$ ) and testing for conditional convergence.

### 4.3.4 Parameter Heterogeneity: Are There Different Steady States?

An alternative way to estimate conditional beta-convergence is to remove the assumption of parameter homogeneity, as suggested by Canova and Marcet (1995) and Maddala and Wu (2000), using time-series cross-sectional (TSCS) data.<sup>17</sup> Advocates of this approach argue that the Barro-type growth regressions create biases in the estimated coefficients by pooling data whenever there is heterogeneity in the parameters. Moreover, cross-sectional regressions lead to a waste of information, since they ignore unit-specific time variations in growth rates and prevent the estimation of a steady state for each region or country separately (e.g. Lee et al., 1997; Temple, 1999; Pritchett, 2000; Durlauf, 2001; Brock and Durlauf, 2001; Masanjala and Papageorgiou, 2004).<sup>18</sup>

Canova and Marcet (1995) propose a way to model heterogeneity and calculate steady states for each unit *without* proxying for the steady state of income with additional variables. The model allows calculation of the speed of adjustment for each unit to its own steady state. A weakness of the approach is the need for the time dimension  $t$  to be large; otherwise, estimates will have large standard errors and their small sample distribution may strongly deviate from the asymptotic one. Using cross-country data, they find an average speed of adjustment to be close to 11 percent, but reject the hypothesis of equal steady states for all cross-sectional units.<sup>19</sup> Using an iterative Bayesian approach with a similar cross-country data set, Maddala and Wu (2000) find average annual convergence rates of around 5 percent and further argue in favor of different steady states for each country.

The estimation relies upon transforming equation 4.2 in discrete time as follows:

$$\log(y_{i,t}) = \alpha + \rho_T \log(y_{i,0}) + \gamma X_i + u_i, \quad (4.5)$$

where  $y_{i,t}$  is *relative* output per worker, which will be defined below,  $\rho_T = \exp^{-\beta T}$ ,  $t = 0, 1, 2, \dots, T$ , and the variables  $X_i$  are introduced to allow for shifts in the limit of the steady state means of  $y_i$ . The key to allow for parameter heterogeneity relies in dropping the assumptions that  $\beta_i = \beta$  and  $\alpha_i = \alpha \forall i$ . The first assumption is expressed by  $\rho_i \neq \rho$ ; that is to

<sup>17</sup>For a description of time-series cross-sectional data, see, for example, Beck (2001) and Beck and Katz (2007).

<sup>18</sup>As indicated by Masanjala and Papageorgiou (2004), parameter heterogeneity in growth regressions has at least three interpretations: there are (i) multiple steady states, i.e., the parameters of a linear growth regression are not constant across countries (e.g. Durlauf, 1996), (ii) omitted growth determinants (e.g. Durlauf and Quah, 1999), and (iii) nonlinearities of the production function, i.e., the identical Cobb-Douglas aggregate production function may be unsuitable. After investigating the third interpretation, Masanjala and Papageorgiou (2004) conclude that using more general constant elasticity of substitution aggregate production functions does not explain away heterogeneity across countries, and they consequently suggest shifting attention to the other two interpretations.

<sup>19</sup>According to Shioji (1997) their convergence rates are high due to the type of Bayesian approach and the short period used (10 years).

say, the convergence rates among all economies are allowed to be different. After grouping  $\alpha_i = \alpha + \gamma X_i$ , the final estimation is

$$\log(y_{i,t}) = \alpha_i + \rho_i \log(y_{i,t-1}) + u_{i,t}. \quad (4.6)$$

Note that both Canova and Marcet (1995) and Maddala and Wu (2000) use relative per worker (capita) output  $y_{i,t}$  for the estimation, defined as  $Y_{i,t}$ , i.e., per capita output of region  $i$  in period  $t$ , divided by the national average of output per capita in year  $t$ . A value higher (or lower) than 1 means that the region has a higher (or lower) per-capita output than the national average. Using  $y_{i,t}$  instead of  $Y_{i,t}$  has the advantage that the linear trend term disappears, as it is assumed that in steady state all  $y_{i,t}$  should grow at the same rate of technological progress, although the levels may vary. It also corrects for problems of serial and residual cross-unit correlation and avoids specifying a process for growth, that is, whether it is trend or unit-root with drift (Maddala and Wu, 2000).

For each region, Equation 4.6 is an AR(1) process of  $\log(y_{i,t})$ . If  $|\rho| < 1$ , the time series is stationary and given that  $E(\log(y_{i,t})) = E(\log(y_{i,t-1}))$ , one could estimate the expected value as

$$\hat{E}(\log(y_{i,t})) = \frac{\hat{\alpha}_i}{1 - \hat{\rho}_i}, \quad (4.7)$$

where  $\hat{\alpha}_i$  and  $\hat{\rho}_i$  are obtained from regressions based on Equation 4.6.

According to Maddala and Wu (2000), the condition  $|\rho| < 1$  ensures that region  $i$  converges towards its own steady state and is equivalent to the definition of beta-convergence in Barro and Sala-i-Martin (1992a). As long as  $|\rho| < 1$ , the speed of adjustment of each unit to its own steady state is given by  $1 - \rho_i$ .

Concerning the empirical estimation, and as discussed by Maddala and Wu (2000), equation 4.6 can be estimated by (i) pooling the data and assuming that  $\forall_i \alpha_i = \alpha$  and  $\rho_i = \rho$ , (ii) running 25 separate regressions, one for each department, allowing for 25  $\alpha_i$  and  $\rho_i$ , or (iii) through shrinkage estimators that assume that  $\alpha_i$  and  $\rho_i$  have two components, one fixed and one random. Additionally, one could estimate Equation 4.7, assuming that there is a fixed number of groups, allowing, for example, for three values of  $\alpha$  and  $\rho$ , in other words,  $\alpha_1, \alpha_2, \alpha_3$  and  $\rho_1, \rho_2, \rho_3$ . The departments that belong to each group should be identified with the appropriate method.

We will estimate equation 4.6 following all the alternatives presented.

### 4.3.5 Sigma-Convergence

An alternative to evaluating beta-convergence is to focus on whether there is a reduction over time in the dispersion of real per-capita income across entities, indicating a more equitable distribution of income. This is called sigma-convergence and arises when for  $T > 0$

$$\sigma_{t+T} < \sigma_t, \quad (4.8)$$

where  $\sigma_t$  is the standard deviation of real per-capita income in period  $t$  (Sala-i-Martin, 1996). The existence of beta-convergence tends to generate sigma-convergence. However, there are cases in which shocks affecting each entity differently lead to the existence of beta-convergence but the lack of sigma-convergence. The example given by Sala-i-Martin (1996) in this regard is clear. Assume two economies, one rich and one poor. The initial poor economy grows so quickly that in the final period its distance from the rich one is the same as before, except that now the poor economy is the wealthier. In such a case, the resulting standard deviation would be the same in the initial and final period. One would observe beta-convergence, given that the poor economy is growing more quickly than the rich one, but no sigma-convergence. Hence, sigma-convergence is an indicator of dispersion of the overall entities, but does not tell much about mobility of each one. Beta-convergence is thus a necessary, but not sufficient, condition for observing sigma-convergence.

## 4.4 Distributional Approach: Quah's Critique

One important critique to the standard regression approach was raised by Danny Quah (Quah, 1993a,b, 1996, 1997), who argues that neither beta nor sigma-convergence can deliver useful answers to the question of whether poor countries or regions are catching up to wealthier ones. Quah argues that the classical approach does not give any information about mobility, stratification, or polarization, and suggests that the typically obtained 2-percent speed of convergence is a statistical artifact that arises in moderate size samples for reasons other than convergence (Durlauf et al., 2005). In his analysis using cross-country data, Quah finds some evidence of convergence clubs, but also evidence of poor countries becoming progressively poorer and wealthy countries, even wealthier.

Quah initially suggested working with a sequence of income distributions and, after discretizing the space of income values, counting the observed transitions into and out of the distinct cell values to construct a transition probability matrix (Quah, 1993a,b). Later, Quah (1997) argued that the discretization could distort dynamics if the underlying observations are indeed continuous variables. He proposed thinking of the distinct cells as tending to-

wards infinity and towards the continuum, with the transition probability matrix tending to a matrix with a continuum of rows and columns, that is, becoming a stochastic kernel.<sup>20</sup>

The methodology is based upon tracking the evolution over time of the entire cross-sectional distributions across regions through the estimation of kernel densities for “relative” variables, which means that the variables of interest are expressed as being relative to the national average, allowing abstraction from changes in the mean when one evaluates how the distribution changes.

Before we define how we proceed to test for convergence using the distributional approach, we briefly present some concepts needed for our estimation.<sup>21</sup>

For the distributional approach, all variables are expressed relative to the Colombian value. Additionally, we take the logarithm of the relative variable, as it facilitates the comparison to the national level. Expressed in logs, a relative value equal to 0 indicates that the department has the same value as the country, while a value that is, for example, equal to -0.05 means that the value of the department is 5 percent lower than the national value.

A univariate kernel density estimate may be regarded as a generalization of a histogram:

$$\hat{f}_h(q) = \frac{1}{mh} \sum_{i=1}^n \kappa \left( \frac{q - Q_i}{h} \right), \quad (4.9)$$

where  $\kappa$  is a kernel,  $m$  is the number of observations, and  $h > 0$  is the bandwidth, also called the smoothing parameter.<sup>22</sup> In the context of growth convergence, we are interested in checking whether we find unimodality or multimodality in the estimated densities of the logarithm of relative income, and in what way the estimated densities change between the starting and the final period.

Bivariate kernel density estimation requires two-dimensional data and a two-dimensional kernel. Here,  $Q = (Q_1, Q_2)^T$  and the kernel  $K$  maps  $\mathbb{R}^2$  into  $\mathbb{R}_+$ . The estimate is

$$\hat{f}_H(q) = \frac{1}{m} \sum_{i=1}^m \frac{1}{\det(H)} K\{H^{-1}(q - Q_i)\}, \quad (4.10)$$

where  $K$  is a bivariate kernel function,  $m$  is the number of observations, and  $H$  is a symmetrical bandwidth matrix.

For the analysis of convergence, we estimate the bivariate kernel density for the relative variable in two periods and check whether or not a large portion of the probability mass remains clustered around the 45-degree diagonal, which would indicate persistence in the

<sup>20</sup>For a technical derivation of a stochastic kernel see Quah (1997, section 4).

<sup>21</sup>A review of the statistical principles of univariate and multivariate kernel density estimations can be found, for example, in Härdle et al. (2004).

<sup>22</sup>Kernel refers to any smooth function satisfying the conditions  $\kappa(q) > 0$ ,  $\int \kappa(q) dq = 1$ ,  $\int q \kappa(q) dq = 0$ , and  $\sigma_\kappa^2 \equiv \int q^2 \kappa(q) dq > 0$  (Wasserman, 2006).

distribution. We present the 3D representation of the estimated bivariate density and a contour plot showing the highest density regions.

## 4.5 Empirical Estimation and Results

We empirically test for convergence in PDB and IDBH, using both the classical and distributional approaches to convergence, as we are interested in checking if, in the Colombian case, there is a contradiction of the results obtained when employing both approaches, as suggested by the existing literature on Colombia. We do not use population weights in our calculations, as we are interested in investigating whether or not departments that were lagging behind have been able to catch up, and consider this to be a pertinent question in the Colombian case where departments are important political entities, with elected local governments and separate department assemblies.

Our empirical analysis begins with the classical approach, testing for sigma and beta-convergence. In the case of beta-convergence, we test absolute and conditional convergence. Conditional convergence is tested with cross-sectional regressions with control variables and also with AR(1) regressions using time-series cross-sectional data for relative income, starting with a pooled model that assumes homogeneity in the parameters and then allows for heterogeneity.

We then follow the distributional approach and compute univariate and bivariate kernel density estimators for relative income in 1975 and 2000.

### 4.5.1 Sigma-Convergence

Results of sigma convergence are presented in Figure 4.7. As may be observed, there exists evidence of sigma-convergence in IDBH but not in PDB. From 1975 to 1984, the standard deviation of the log of both variables remains close to 0.40. From 1985 onwards, IDBH decreases and has a value close to 0.32 in 2000. On the contrary, PDB remains around 0.40. Thus, the distribution of IDBH has become more equitable, while the distribution of PDB has not.

### 4.5.2 Absolute Beta-Convergence

Figure 4.8 shows a weak inverse relationship between the growth rate of per-capita PDB between 1975 and 2000 and its value in 1975. Cross-sectional regression results based upon Equation 4.3 and using NLLS are shown in Table 4.2. We use HC3 robust standard errors as proposed by Davidson and MacKinnon (1993) to account for possible heteroscedasticity,



considering that the number of observations is small (Long and Ervin, 2000). The estimated speed of convergence is 0.7 percent, but it is not significantly different from 0 at the 5 percent level. The adjusted R-squared of the regression is extremely low (0.01) suggesting that this model does not explain departmental PDB growth rates. These results do not change if one excludes Chocó, Nuevos, and Guajira, which have a large influence on results, as suggested by Cook's distance computed after the first regression (Figure 4.9).

In the case of IDBH, Figure 4.10 shows a stronger negative relationship between the growth rate of per-capita IDBH between 1975 and 2000 and its value in 1975. This is confirmed with the regression presented in Table 4.3, where the estimated speed of convergence is 1.2 percent and statistically significant. The adjusted R-squared is 0.35. Excluding Guajira, as suggested by Cook's distance, and then rerunning the regression yields similar results.

Hence, we find evidence of absolute beta-convergence using IDBH, but not using PDB.

### 4.5.3 Conditional Beta-Convergence Using Control Variables

As explained in Subsection 4.3.3, one may drop the assumption that all economies have the same parameters, and hence the same steady state, and try to proxy for the steady-state level of income with a set of variables  $X_i$ , running regressions based upon Equation 4.4.

There is no agreement as to which variables to include as proxies for the steady state with cross-sectional data (Durlauf et al., 2005). We use variables that are based upon theoretical arguments and our choice is limited by data availability at the departmental level. We use the logarithm of population growth and a variable based upon saving rates.<sup>23</sup> Additionally, we use three variables proxying for human capital: log of life expectancy in 1975, log of literacy in 1973, and log of net enrolment rate in 1985. Several specifications for the average growth rate of per-capita PDB are shown in Table 4.4 and for per-capita IDBH in Table 4.5.<sup>24</sup>

Results for PDB show that the speed of convergence remains statistically insignificant in all the specifications, including the variables proxying for the steady state, as was the case with absolute convergence. We find no evidence of conditional convergence using PDB data.

In the case of IDBH, where we find evidence of absolute convergence, once we include variables  $X_i$  proxying for the steady-state level of income, the speed of convergence turns insignificant. We find no evidence of conditional convergence using IDBH data.

<sup>23</sup> As the saving rates that are available from CEGA (2006b,a) include values that are negative, we add a constant to all values, so that the transformed data are all positive and we can compute the logs.

<sup>24</sup> The number of departments included depends upon data availability.



#### 4.5.4 Beta-Convergence Using Time-Series Cross-Sectional Data

Recall that with TSCS data, the regression is based upon Equation 4.6, defined in subsection 4.3.4 as

$$\log(y_{i,t}) = \alpha_i + \rho_i \log(y_{i,t-1}) + u_{i,t},$$

which uses the measure of relative income  $y_{i,t}$ , that is, income of each department expressed as the ratio to the national average. One may estimate the equation in several ways. First, we begin by pooling the data, assuming homogeneity in the parameters. Second, we use linear mixed models where the parameters are assumed to have a fixed component, common to all departments, and a random part. Third, we estimate 25 separate ordinary least squares (OLS) regressions for each entity. Finally, we assume that there are several groups of departments which share the same  $\alpha$  and  $\rho$ , and explore this issue with finite mixture models.

In all cases, the key issue is whether the estimated value for  $\rho$  is lower than 1, which would suggest that there is economic convergence.

##### Pooled Data and OLS

The assumption of  $\alpha_i = \alpha$  and  $\rho_i = \rho \forall i$  in Equation 4.6 is equivalent to assuming that there is a common steady state to all departments. Hence, the results are comparable to those obtained using cross-sectional data when we tested for absolute beta-convergence in subsection 4.5.2.

Tables 4.6 and 4.7 present the results for PDB and IDBH using TSCS pooled data and estimating with OLS. In both cases, the estimated  $\rho$  is less than 1 (0.989 for PDB and 0.986 for IDBH). However, it must be noted that while the value 1 is not included in the 95 percent confidence interval of  $\rho$  for IDBH, it is included for PDB, confirming the evidence of absolute convergence in IDBH, but not in PDB.

For IDBH the implied estimated speed of convergence  $\beta$ , computed with the estimated  $\rho$  value, is 1.4 percent, similar to and only slightly higher than the one observed using cross-sectional data (1.2 percent) in Section 4.5.2.

##### Mixed Models

We follow here a frequentist approach for the estimation of Equation 4.6. Following Maddala et al. (1997) and using matrix notation, we define

$$Z_i = \begin{pmatrix} \log(y_{i,1}) \\ \vdots \\ \log(y_{i,T}) \end{pmatrix}, \quad X_i = \begin{pmatrix} 1 & \log(y_{i,0}) \\ \vdots & \vdots \\ 1 & \log(y_{i,T-1}) \end{pmatrix}, \quad b_i = \begin{pmatrix} \alpha_i \\ \rho_i \end{pmatrix}, \quad \text{and} \quad U_i = \begin{pmatrix} u_{i,1} \\ \vdots \\ u_{i,T} \end{pmatrix},$$

with  $i = 1, \dots, N$ , where  $N$  is the number of regions in the data.

We consider the autoregressive regression model

$$Z_i = X_i b_i + U_i, \quad i = 1, \dots, N, \quad (4.11)$$

with the assumptions  $U_i \sim N(0, \sigma_i^2 I)$ , and  $b_i \sim N(\mu, \Sigma)$ , where  $I$  is the identity matrix and  $\Sigma$  is a nonzero covariance matrix.<sup>25</sup> We further assume that the  $U_i$  are independent across the  $N$  equations, and that  $b_i$  and  $U_i$  are independent for different regions.

We work with a linear mixed model (McCulloch and Searle, 2001). If we write  $b_i$  as  $b_i = \mu + \eta_i$ , with  $\eta_i \sim N(0, \Sigma)$ , we can rewrite  $Z_i$ , ( $i = 1, \dots, N$ ) as

$$\begin{aligned} Z_i &= X_i(\mu + \eta_i) + U_i \\ &= X_i\mu + X_i\eta_i + U_i \end{aligned} \quad (4.12)$$

$$= X_i\mu + w_i, \quad (4.13)$$

with  $w_i \sim N(0, \Omega_i)$ ,  $\Omega_i$  being the variance covariance matrix defined as

$$\Omega_i = X_i \Sigma X_i' + \sigma_i^2 I. \quad (4.14)$$

In Equation 4.12, the vector  $\mu$  represents the fixed effects and  $\eta_i$  represent the random effects. In linear mixed models, fixed effects are used for modeling the mean of the response variable and the random effects are used to model the variance-covariance structure of it (McCulloch and Searle, 2001). The parameters in our linear mixed model are then  $\mu$ ,  $\Sigma$ , and  $\sigma_i^2$ . The last two parameters are in fact variance components, as presented in Equation 4.14.

One can obtain an estimator for  $\mu$  and best-linear unbiased predictors for the random effects  $\eta_i$  with maximum likelihood or restricted maximum likelihood (REML).<sup>26</sup> Here, we prefer REML for three reasons: (i) the estimators are based upon taking into account the degrees of freedom for the fixed effects in the model, (ii) because of its unbiasedness in the case of balanced panels, and (iii) as REML estimators seem to be less sensitive to outliers in the data.<sup>27</sup> With the obtained values for  $\mu$  and  $\eta_i$ , one could compute predicted values for the  $N$  different  $\alpha_i$  and  $\rho_i$ .

We are interested in the estimation of the fixed effects. As was mentioned before, the literature suggests that in some cases, the estimated speed of convergence can be substantially higher than the one obtained by assuming there are no random effects. We also compare the results with those assuming homogeneity in the parameters using likelihood ratio tests

<sup>25</sup>The results of the estimation assume no special structure of the matrix  $\Sigma$ .

<sup>26</sup>For the algorithms used for obtaining maximum likelihood and restricted maximum likelihood estimates in the case of a linear mixed model, see Pinheiro and Bates (2000).

<sup>27</sup>For a review of linear mixed models and a discussion of the estimation with maximum likelihood and REML, see McCulloch and Searle (2001).

and the Akaike information criterion (AIC) in order to investigate if a more flexible model allowing for heterogeneity in the parameters should be preferred.

Results for PDB are presented in Table 4.8. The estimated coefficients for the fixed effects are similar to the coefficients estimated when assuming homogeneity in the parameters (Table 4.6). In the case of  $\rho$ , the estimated value for the linear mixed model is 0.984, close to the value 0.989 obtained with OLS and assuming no random effects. It must be noted that the standard error of the fixed effect of  $\rho$  is higher than for the coefficient estimated in the model assuming homogeneity in the parameters. The estimated standard deviations of both random effects are quite low, especially the one for  $\alpha$ , with a value close to 0, suggesting there is no evidence of different steady states. The value for the Akaike information criterion for the linear mixed model is larger than for the simpler model, assuming parameter homogeneity, and hence the simpler model is preferred. This is also corroborated by a likelihood ratio test.

Table 4.9 shows the results for IDBH. Once again, the coefficients for the fixed effects are close to the ones obtained with the model in the previous section, in which we assumed parameter homogeneity (Table 4.7), with  $\rho$  equal to 0.986 in both cases. The estimated standard deviations of both random effects are low, in particular the one for  $\alpha$ , which is close to 0, giving no support for the existence of different steady states. The AIC suggest that the simpler model is better, which is confirmed with a likelihood ratio test.<sup>28</sup>

### Separate Regressions for Each Department

We also treat all departments as separate entities and run an AR(1) regression for each one. These separate regressions shed light upon the effect of past values on current values, but due to the low amount of observations for each department (25 years), estimations are not reliable. In Table 4.12, we present results for PDB. The slope coefficient  $\rho$  is lower than 1 for all departments but has large standard errors and is not significant at the 5-percent level for Cauca and Boyacá.<sup>29</sup> The resulting speeds of convergence are implausibly high with values ranging from 10 to 60 percent in the case of PDB, a result influenced by the fact that the period only covers 25 years. Results for IDBH are similar (Table 4.13).

The graphical analysis of each time series is more informative. In Figure 4.12, we plot the individual time series for relative PDB in logs for all departments. We observe that in almost all departments, the values change little over time and the series seem stationary. They remain either above or below the national average with the exception of Guajira and

<sup>28</sup>Although it is possible to calculate the implied speed of convergence for each department, the interpretation is difficult. For illustrative purposes, we present them in Tables 4.10 for PDB and 4.11 for IDBH. The associated speeds of convergence have a larger variability for PDB than for IDBH. The average speed of convergence is 1.6 percent for PDB and 1.4 percent for IDBH).

<sup>29</sup>The expected value can be calculated when  $|\rho| < 1$ , i.e. when the time series is stationary.

Nuevos. The time series do not become closer to the national value over time, except for Guajira, indicating a lack of economic convergence among departments.

Results for IDBH (Figure 4.13) show that most of the time series seem stationary. Interestingly, the wealthiest department, Bogotá, moves slightly closer to the national average, as does as the poorest department, Chocó. Guajira, although becoming closer to the national average, still remains below it.

### Mixture Models

In the previous sections, we estimated a model assuming that  $\alpha$  and  $\rho$  are the same for all departments. We then allowed these parameters to be different for each department, in the context of a linear mixed model, where the parameters are assumed to have a fixed component, common to all departments, and a random part. Then, we estimated 25 separate AR(1) regressions, one for each department.

Another possibility is that there are several groups of departments which share the same  $\alpha$  and  $\rho$ . We explore this possibility with a finite mixture model, as described in Leisch (2004) and Grün and Leisch (2008). These types of models can be applied, assuming that observations originate from various groups, where the group affiliations are unknown. Finite mixture models with a fixed number of components are estimated with the expectation-maximization (EM) algorithm within a maximum likelihood framework.

We assume three groups and fit the model with the statistical software R (R Development Core Team, 2008) and the package *flexmix* (Leisch and Grün, 2008). Results for PDB and IDBH are presented in Tables 4.14 and 4.15. We show estimated  $\alpha$  and  $\rho$  for each group of departments, as well as the departments composing each group.

Results for PDB (Table 4.14) show that Group 1 includes many of the poorest departments (e.g., Chocó, Sucre, Nariño, and Córdoba), Group 2 is composed of Nuevos Departamentos and La Guajira, and Group 3 includes the richest departments (e.g., Bogotá, Valle, and Antioquia).<sup>30</sup> Estimated values for  $\alpha$  and  $\rho$  are similar for Groups 1 and 3, with  $\alpha$  being negative and close to 0 and  $\rho$  being close to 0.99, a result that is similar to the estimated value obtained in subsection 4.5.4, assuming homogeneity in the parameters. The implied speed of convergence for these two groups is close to 1 percent. If one believes in the validity of the estimated expected value of the time series, one would expect that departments belonging to Group 1 would remain well below the national average over time, while those from Group 3 would remain below, as well, but would be closer to it. As was discussed before, Nuevos Departamentos and La Guajira experienced high growth rates between 1975 and 2000, as-

<sup>30</sup>Mixture models are only identifiable up to a permutation of the component labels (Leisch, 2004). The names, Group 1, Group 2, etc., have no special meaning here, and the order of the groups is irrelevant.

sociated with the production of oil and coal. The model captures this, suggesting that both departments are far from their steady states, showing a large implied speed of convergence (10 percent), and predicting that both would remain above the national average.

Concerning IDBH (Table 4.15), the grouping of departments is similar as above, with Group 1 including many of the poorest departments and Group 3 including the richest ones. Group 2 now includes Nuevos Departamentos, La Guajira, and Sucre. Groups 1 and 3 have values for the estimated  $\alpha$  that are quite similar to one another, and close to 0. Values for the estimated  $\rho$  are also similar with 0.98 for Group 1 and 0.99 for Group 3, both being close to the estimated value obtained, assuming homogeneity in the parameters (Subsection 4.5.4). Nuevos Departamentos, La Guajira, and Sucre have values for  $\alpha$  and  $\rho$  that are different than those from the other two groups (-0.01 for  $\alpha$  and 0.96 for  $\rho$ ). Once again, the model suggests that these departments are far from the steady state, with an implied speed of convergence of 4 percent, which speed is greater than that for Groups 1 (2 percent) and 3 (1 percent). Once again, with an analyzed time period of only 25 years, it is questionable whether one should rely upon the estimated expected values.

#### 4.5.5 Kernel Density Estimators

All the results for kernel density estimations were computed with the statistical software R (R Development Core Team, 2008) and the package *ks*.<sup>31</sup> For both univariate and bivariate kernel density estimations, we use Gaussian kernels and smoothed cross validation bandwidth selectors<sup>32</sup> (Jones et al., 1991; Duong and Hazelton, 2005). In the univariate case, and as suggested by Bowman and Azzalini (1997), we use for both years the same bandwidth, which is computed as the mean of the two selected bandwidths obtained for each year separately. In the bivariate case, the smoothed cross validation is unconstrained, meaning that we do not impose that the (nonsingular) bandwidth matrix  $H$  has to be diagonal in Equation 4.10. Hence, we are able to handle correlation between components, as we allow kernels to have an arbitrary orientation (Wand and Jones, 1995). As we are especially interested in checking whether a large portion of the probability mass remains clustered around the 45-degree diagonal, this flexibility is relevant for us. If we were to impose a diagonal matrix  $H$ , only kernels which are oriented to the coordinate axes would be allowed.

Univariate kernel density estimations of the logarithm of relative departmental PDB for the years 1975 and 2000 are shown in Figure 4.16. Both densities seem unimodal and are

<sup>31</sup>*ks* is currently the most comprehensive kernel density estimation package in R (Duong, 2008). All the estimations were done with the function *kde*.

<sup>32</sup>We have also tried direct plug-in methods as suggested by Sheather and Jones (1991) and obtained results that are not very dissimilar.

very similar. Thus, according to this figure, there were almost no changes in the distribution. Table 4.16 shows the results of a formal bootstrap test of equality of the estimated densities for both years. We cannot reject the hypothesis that the two densities are identical. Bivariate kernel-density estimators are presented in Figures 4.17 and 4.18. Both figures make clear that most of the mass is concentrated along the 45-degree diagonal and hence support persistence in the distribution. Departments with a relative PDB that was above (or below) average in the year 1975 tend to remain above (or below) average in 2000. Two interesting cases are La Guajira and Nuevos Departamentos, as they show some mobility. Nuevos Departamentos was close to the national average in 1975 and is clearly above the average in 2000, while La Guajira was clearly below the national average in 1975 and is quite close to it in 2000.

Turning to results using the logarithm of relative departmental IDBH, Figure 4.19 presents the univariate kernel estimators for the years 1975 and 2000, showing a slight shift of the distribution to the right in 2000. The distribution narrowed between 1975 and 2000 and the two modes observed in 1975 in the left and right tails of the distribution almost disappeared in 2000. However, according to the bootstrap test of equality of the estimated densities for both years, we cannot reject the hypothesis that the two densities are identical (Table 4.16). Bivariate kernel-density estimators are presented in Figures 4.17 and 4.18. Bivariate kernel density estimators in Figures 4.20 and 4.21 show some mobility, as well. In the contour plot (Figure 4.21), the mass of the distribution rotates slightly clockwise, suggesting mild convergence in the distribution.

## 4.6 Conclusions

Returning to the questions raised at the beginning of the study, we do not find absolute or conditional convergence in PDB using the regression approach. The distributional approach shows persistence in the distribution, i.e., relative to the average, each department remains in the position where it was located in 1975. Results of both methods point in the same direction – there is no convergence but persistence in PDB does exist.

Analysis of IDBH shows absolute convergence using the regression approach. After testing different models allowing for parameter heterogeneity, we found that there is no evidence of the existence of different steady states. The pooled model using TSCS provides our preferred estimators. Bivariate kernel density estimators show some improvements in the distribution. However, the changes are small and consistent with the low speed of convergence of around 1 percent.

Different factors explain our results. Differences in geography, infrastructure, and population density among departments are relevant factors to explain lack of convergence in

PDB, as are differences in production structures and value added by department. Excepting for the mining departments, the different production structures remained almost unchanged between 1975 and 2000 (Table 4.17).<sup>33</sup> However, mineral exploitation in Colombia is relatively recent, going back only to the mid eighties, and this fact explains why the group of Nuevos and the department of La Guajira are the only initial poor departments that grew more quickly than the wealthier departments, according to PDB data. Previous literature had already pointed to the fact that once the mining departments are excluded, any hint of convergence disappears (Birchenall and Murcia, 1997) and that departments with a high share of agricultural production had the lowest growth rates (Bonet, 1999). Three departments concentrated at least 50 percent of PDB in both evaluated years: Antioquia, Bogotá, and Valle del Cauca. These three departments combined produced 65 percent of the manufacturing output in 1975 and 60 percent in 2000. The stability of the shares in other sectors is also remarkable, indicating departmental concentration and low mobility of production factors across the country.

At least two of the assumptions of the Solow model, which is the usual theoretical framework for studying convergence, seem problematic for application to the Colombian case. First, the neoclassical model assumes mobility of factors, which is in this case constrained by geographic, climatic, and infrastructural issues, as well as by the internal conflict issue. For instance, several productive sectors periodically suffer from attacks by violent groups, not only on physical capital, but also human capital through kidnapping and extortion. Second, the assumption of constant returns to scale is an oversimplification that does not hold for all sectors in the economy. As has been argued by World Bank (2009b), while returns to scale in agriculture tend to be constant, those in manufacturing and services are increasing.

The slow convergence observed in IDBH could be explained by recent redistributive policies. Recall that per capita IDBH is a measure of households' income after subtracting taxes on property and rental income and net payments to the social security, and adding other net current transfers. We must also note that Colombia has experienced higher public spending in social sectors and infrastructure. Literature dealing with the direct link between convergence and public spending is scarce, but suggests that it affected the relative position of some departments, although not the distribution as a whole (Ardila Rueda, 2004), and that efficiency of public spending has been decreasing over time, mainly due to political interests and corruption.

---

<sup>33</sup>Nuevos Departamentos increased its participation in the Colombian mining sector from 11 percent in 1975 to 55 percent in 2000 (Table 4.17).



---

**Summary of Results**


---

	Per capita income measure used	
	PDB	IDBH
<b>Classical Approach: Convergence?</b>		
Sigma	No	Yes
Absolute Beta	No	Yes
Conditional Beta Cross Sections	No	No
Conditional Beta Pooled TSCS assuming homogeneity of parameters	No	Yes
<b>Distributional Approach</b>		
Univariate Kernel Estimators	Distribution	Distribution
	Unchanged	Unchanged
Bivariate Kernel Estimators	Persistence in the Distribution	Suggests slow Convergence

*Note:* Results for conditional beta convergence with TSCS data and for the distributional approach based on relative values, i.e., ratios to the national level.

Increased social spending has also benefited from mining sector revenues which are distributed across all departments through the fiscal system.<sup>34</sup> IDBH of mining departments is still very low and did not exhibit the high growth rates observed in PDB.<sup>35</sup> One reason for this is that fiscal decentralization began in the late eighties and the reforms are thus still too recent to be fully evaluated. A second reason is that financial resources from mining sectors are not efficiently spent because of corruption and are not sufficient to compensate for the low starting point in income of these departments. Recall that in 1975, La Guajira was the second poorest department in Colombia and that a large part of its population is indigenous and poorly linked to the departmental economy. Previous research suggests that even if rev-

<sup>34</sup>Oil revenues are divided between direct and indirect revenues and correspond to about eight to 25 percent of total extracted crude oil income. Direct revenues are those given to producing departments, municipalities, and ports of exports basically to finance investment in social sectors, and account for about 76 percent of oil revenues. Indirect revenues are those distributed among non-producing departments (Hernández, 2004).

<sup>35</sup>Producing departments are obliged to spend at least 50 percent of the received mining revenues on social investment until having achieved certain minimum thresholds for infant mortality, health care, education, water, and sanitation. Indirect revenues are distributed according to projects presented through territorial entities (Law 141 of 1994).



venues of coal exports in La Guajira were distributed efficiently and without any corruption-related loss (corruption levels seem to be particularly high in mining departments), IDB of that department would still be about 60 percent of national IDB in 2000 (Meisel, 2007a).

Although overall public spending has increased, the transfer system bears some disadvantages for poor departments. Evidence shows that after totaling all public revenue (not only that directed to social sectors), there is no fiscal equalization in Colombia and the system is regressive; wealthy municipalities have the highest shares of public funds.

Two other issues have to be taken into consideration for interpreting the results of both PDB and IDBH. One is that in 2000, our last year of analysis, the country was experiencing a large economic crisis which affected public and private finances. Transfers from the central government were thus also affected by the crisis. A second issue is related to the domestic conflict. Between 1998 and 2002, violence escalated dramatically when the groups involved in the war were fighting one other for control of strategic areas. Sánchez and Palau (2006), who deal directly with this last issue, argue that decentralization policies, political and fiscal, affected the interests of armed groups and even strengthened them through the sharp increase in local resources. The higher political autonomy at the local level increased the ability of armed groups to intimidate politicians and to extract rents from public funds. Guerrillas relocated in strategic zones with greater levels of prosperity, the facility for processing illicit drugs, and an intimidated local population (Sánchez and Palau, 2006).

One of the policy implications of this study is the necessity of monitoring the efficiency of social spending and enforcing decentralization policies so that a faster convergence in IDBH can be achieved. Concerning convergence in PDB, reallocation of productive sector resources is not easy to achieve and could yield to efficiency losses, but the state can, for example, encourage the accumulation of human capital and improve infrastructure in lagging departments, which would help attract investments in the long run. Additionally, it is crucial to find a way out of the internal conflict to foster factor mobility in Colombia, particularly in those areas without significant state presence. We consider it vital to have an explicit regional policy in Colombia to foster growth in departments lagging behind national averages, after conducting case studies to assess which policies could be most effective in each case.

Finally, for monitoring convergence across departments in the future, it is essential to have consistent time series constructed under a single methodology. Unfortunately, the work done by CEGA for the period 1975 to 2000 was not continued for the years after 2000. Such a project is of high policy relevance for the country.

## 4.7 Tables

Table 4.1: Colombia. Gross Domestic Product (Constant Million Pesos of 1994), Per Capita GDP and Population. 1980-2006.

Year	GDP (million)	Per capita GDP	Growth rate	Population	Growth rate
1980	40822304	1503335		27154504	
1981	41846404	1503069	-0.02	27840636	2.53
1982	42160220	1476873	-1.74	28546950	2.54
1983	42820420	1462737	-0.96	29274176	2.55
1984	44217404	1472781	0.69	30023068	2.56
1985	45475604	1476748	0.27	30794424	2.57
1986	48189708	1533078	3.81	31433316	2.07
1987	50775504	1582200	3.20	32091720	2.09
1988	52808848	1611804	1.87	32763808	2.09
1989	54544940	1630958	1.19	33443488	2.07
1990	56873928	1666658	2.19	34124536	2.04
1991	58222936	1671462	0.29	34833548	2.08
1992	60757528	1710026	2.31	35530176	2.00
1993	64226880	1773819	3.73	36208244	1.91
1994	67532864	1832015	3.28	36862624	1.81
1995	71046216	1895088	3.44	37489664	1.70
1996	72506824	1904234	0.48	38076640	1.57
1997	74994024	1940536	1.91	38646044	1.50
1998	75421328	1923949	-0.85	39201320	1.44
1999	72250600	1817821	-5.52	39745712	1.39
2000	74363832	1846071	1.55	40282216	1.35
2001	75458112	1849177	0.17	40806312	1.30
2002	76917224	1861165	0.65	41327460	1.28
2003	79884488	1908947	2.57	41847420	1.26
2004	83772432	1977279	3.58	42367528	1.24
2005	87727928	2045484	3.45	42888592	1.23
2006	93881688	2162904	5.74	43405388	1.20

Source: Own calculations based on National Accounts and Census 2005, DANE

Table 4.2: Beta Convergence Using Cross-sections and Non Linear Least Squares. Dependent Variable: Average Growth Rate of pc PDB 1975-2000.

Variable	Robust HC3			
	Coefficient	Std. Err.	95% conf. interval	
Intercept	0.1055	0.1259	-0.1548	0.3659
$\beta$	0.0067	0.0108	-0.0155	0.0290
$\beta$ (%)	0.67			
Number of observations	25			
Adj.R-squared	0.0112			

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Source: HC3 standard errors calculated according to Davidson and MacKinnon (1993).

Table 4.3: Beta Convergence Using Cross-sections and Non Linear Least Squares. Dependent Variable: Average Growth Rate of pc IDBH 1975-2000.

Variable	Robust HC3			
	Coefficient	Std. Err.	95% conf. interval	
Intercept	0.1533	0.0392	0.0721	0.2345
$\beta$	0.0119	0.0039	0.0038	0.0200
$\beta$ (%)	1.19			
Number of observations	25			
Adj.R-squared	0.3514			

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Source: HC3 standard errors calculated according to Davidson and MacKinnon (1993).

Table 4.4: NLLS Regressions. Dependent Variable Average Growth Rate of pc PDB. 1975-2000.

Regressors	(1)	(2)	(3)	(4)	(5)
	b/se	b/se	b/se	b/se	b/se
constant	0.1157 (0.1306)	0.1534 (0.4804)	0.3533 (0.3071)	0.2753 (0.2533)	0.2271 (0.3062)
speed of convergence $\beta$	0.0076 (0.0114)	0.0401 (0.0809)	0.0186 (0.0250)	0.0226 (0.0278)	0.0375 (0.0622)
log (life expectancy 1975)		0.0402 (0.1344)	-0.0092 (0.0785)		
log (literacy 1973)		0.0355 (0.0645)		0.015 (0.0367)	
log (transformed saving rate)		0.0017 (0.0059)	0.0047 (0.0064)	0.0036 (0.0086)	0.0043 (0.0128)
log (population growth + 0.05)		0.0378 (0.0443)	0.032 (0.0373)	0.0293 (0.0371)	-0.0009 (0.0313)
log (net enrollment rate 1985)					0.0342 (0.0414)
Number of observations	24	24	24	25	24
R-square	0.06	0.31	0.28	0.24	0.31
Adjusted R-square	0.02	0.12	0.13	0.09	0.16

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Note: HC3 robust standard errors (Davidson and MacKinnon, 1993) in brackets.

Table 4.5: NLLS Regressions. Dependent Variable Average Growth Rate of pc IDBH. 1975-2000.

Regressors	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se
constant	0.1574*** (0.0402)	0.1272 (0.2917)	0.335 (0.1817)	0.1400 (0.0962)	0.1233 (0.1038)
speed of convergence $\beta$	0.0123** (0.0040)	0.0282 (0.0280)	0.0135 (0.0110)	0.0254 (0.0126)	0.0262 (0.0155)
log (life expectancy 1975)		0.0063 (0.0777)	-0.0443 (0.0485)		
log (literacy 1973)		0.031 (0.0309)		0.0279 (0.0161)	
log (transformed saving rate)		0.0007 (0.0033)	0.0036 (0.0046)	0.0009 (0.0021)	0.0019 (0.0055)
log (population growth + 0.05)		-0.001 (0.0224)	-0.0079 (0.0181)	-0.0037 (0.0183)	-0.0227 (0.0127)
log (net enrollment rate 1985)					0.0241 (0.0171)
Number of observations	24	24	24	25	24
R-square	0.40	0.58	0.53	0.56	0.64
Adjusted R-square	0.37	0.47	0.43	0.48	0.56

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Note: HC3 robust standard errors (Davidson and MacKinnon, 1993) in brackets.

Table 4.6: OLS Linear Regression. TSCS Data. Dependent Variable  $\log(y_{i,t})$ . Relative Per capita PDB. 1975-2000.

Variable	Coefficient	Std. Err.	95% conf. interval	
Intercept	-0.0023	0.0032	-0.0086	0.0040
$\log(y_{i,t-1})$	0.9891	0.0069	0.9755	1.0027
Implied $\beta$	1.09%			
Number of observations	625			
R-squared	0.9730			
AIC	-1632			

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Table 4.7: OLS Linear Regression. TSCS Data. Dependent Variable  $\log(y_{i,t})$ . Relative Per Capita IDBH. 1975-2000.

Variable	Coefficient	Std. Err.	95% conf. interval	
Intercept	-0.0014	0.0017	-0.0048	.0020
$\log(y_{i,t-1})$	0.9862	0.0046	0.9771	0.9953
Implied $\beta$	1.38%			
Number of observations	625			
R-squared	0.9867			
AIC	-2183			

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Table 4.8: Linear Mixed Model (REML). TSCS Data. Dependent Variable:  $\log(y_{i,t})$ . Relative Per Capita PDB. 1975-2000.

<b>Fixed effects</b>		
<b>Variable</b>	<b>Coefficient</b>	<b>Std. Err.</b>
Intercept	-0.0027	0.0034
$\log(y_{i,t-1})$	0.9836	0.0083
<b>Random effects</b>		
<b>Standard deviation</b>	<b>Estimate</b>	
Intercept	6.1e-09	
$\log(y_{i,t-1})$	0.0163	
Number of observations	625	
Number of groups	25	
AIC	-1606	

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Table 4.9: Linear mixed model (REML). TSCS Data. Dependent Variable:  $\log(y_{i,t})$ . Relative Per Capita IDBH. 1975-2000.

<b>Fixed effects</b>		
<b>Variable</b>	<b>Coefficient</b>	<b>Std. Err.</b>
Intercept	-0.0014	0.0022
$\log(y_{i,t-1})$	0.9859	0.0047
<b>Random effects</b>		
<b>Standard deviation</b>	<b>Estimate</b>	
sd(Intercept)	0.0000	
$\log(y_{i,t-1})$	0.0026	
Number of observations	625	
Number of groups	25	
AIC	-2155	

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Table 4.10: Implied Convergence Rates Using TSCS Data and Linear Mixed Models (REML). Per capita PDB, 1975-2000

Department	Intercept $\alpha$	Slope $\rho$	Implied $\beta$ (%)	Expected Value
Nuevos Departamentos	-0.003	0.976	2.4	-0.111
Antioquia	-0.003	0.984	1.6	-0.168
Atlántico	-0.003	0.983	1.7	-0.156
Bogotá D. C.	-0.003	0.986	1.4	-0.197
Bolívar	-0.003	0.982	1.8	-0.149
Boyacá	-0.003	0.981	1.9	-0.140
Caldas	-0.003	0.981	1.9	-0.137
Caquetá	-0.003	0.990	1.0	-0.265
Cauca	-0.003	0.986	1.4	-0.193
Cesar	-0.003	0.986	1.4	-0.187
Córdoba	-0.003	0.989	1.1	-0.242
Cundinamarca	-0.003	0.983	1.7	-0.161
Chocó	-0.003	0.994	0.6	-0.483
Huila	-0.003	0.983	1.7	-0.162
La Guajira	-0.003	0.957	4.3	-0.063
Magdalena	-0.003	0.986	1.4	-0.194
Meta	-0.003	0.981	1.9	-0.143
Nariño	-0.003	0.988	1.2	-0.220
Norte Santander	-0.003	0.986	1.4	-0.192
Quindío	-0.003	0.977	2.3	-0.119
Risaralda	-0.003	0.982	1.8	-0.148
Santander	-0.003	0.982	1.8	-0.152
Sucre	-0.003	0.998	0.2	-1.095
Tolima	-0.003	0.983	1.7	-0.158
Valle	-0.003	0.985	1.5	-0.174
Mean			1.6	
Median			1.7	

Source: Own calculations based on data from CEGA. Constant prices of 1994.



Table 4.11: Implied Convergence Rates Using TSCS Data and Linear Mixed Models (REML). Per capita IDBH. 1975-2000

Department	Intercept $\alpha$	Slope $\rho$	Implied $\beta$ (%)	Expected Value
Nuevos Departamentos	-0.001	0.986	0.0	-0.10
Antioquia	-0.001	0.986	1.4	-0.10
Atlántico	-0.001	0.986	1.4	-0.10
Bogotá D. C.	-0.001	0.986	1.4	-0.10
Bolívar	-0.001	0.986	1.4	-0.10
Boyacá	-0.001	0.986	1.4	-0.10
Caldas	-0.001	0.986	1.4	-0.10
Caquetá	-0.001	0.986	1.4	-0.10
Cauca	-0.001	0.986	1.4	-0.10
Cesar	-0.001	0.986	1.4	-0.10
Córdoba	-0.001	0.986	1.4	-0.10
Cundinamarca	-0.001	0.986	1.4	-0.10
Chocó	-0.001	0.987	1.3	-0.11
Huila	-0.001	0.986	1.4	-0.10
La Guajira	-0.001	0.985	1.5	-0.09
Magdalena	-0.001	0.986	1.4	-0.10
Meta	-0.001	0.986	1.4	-0.10
Nariño	-0.001	0.986	1.4	-0.10
Norte Santander	-0.001	0.986	1.4	-0.10
Quindío	-0.001	0.986	1.4	-0.10
Risaralda	-0.001	0.986	1.4	-0.10
Santander	-0.001	0.986	1.4	-0.10
Sucre	-0.001	0.987	1.3	-0.11
Tolima	-0.001	0.986	1.4	-0.10
Valle	-0.001	0.986	1.4	-0.10
Mean			1.4	
Median			1.4	

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Table 4.12: Autoregressive Processes of Order 1. Dependent Variable:  $\log(y_{i,t})$ . Per Capita PDB. 1975-2000.

Department	Intercept ( $\alpha$ )		Slope ( $\rho$ )		Regression Adjusted $R^2$	Expected Value	Implied $\beta$ (%)
	Coefficient	Std. Error	Coefficient	Std. Error			
Nuevos Departamentos	0.03	0.03	0.85	0.14	0.62	0.21	14.75
Antioquia	0.06	0.03	0.65	0.16	0.41	0.16	35.48
Atlántico	-0.02	0.01	0.85	0.06	0.88	-0.15	14.99
Bogotá D. C.	0.08	0.09	0.84	0.17	0.52	0.48	16.10
Bolívar	-0.05	0.02	0.69	0.15	0.48	-0.14	31.16
Boyacá	-0.09	0.03	0.29	0.21	0.08		
Caldas	-0.11	0.03	0.39	0.18	0.17	-0.17	61.42
Caquetá	-0.16	0.08	0.76	0.13	0.58	-0.64	24.10
Cauca	-0.26	0.13	0.62	0.18	0.33	-0.68	38.31
Cesar	-0.07	0.04	0.85	0.10	0.77	-0.45	15.34
Córdoba	-0.18	0.09	0.77	0.12	0.66	-0.78	22.95
Cundinamarca	0.01	0.00	0.87	0.13	0.68	0.06	12.51
Chocó	-0.37	0.17	0.67	0.16	0.45	-1.11	33.14
Huila	-0.11	0.05	0.66	0.16	0.43	-0.31	34.25
La Guajira	0.01	0.03	0.90	0.06	0.92	0.06	10.36
Magdalena	-0.29	0.11	0.52	0.18	0.27	-0.59	48.31
Meta	-0.04	0.02	0.67	0.15	0.48	-0.11	32.80
Nariño	-0.30	0.14	0.63	0.16	0.40	-0.83	36.65
Norte Santander	-0.47	0.11	0.09	0.21	0.01		
Quindío	-0.04	0.03	0.64	0.16	0.41	-0.12	36.26
Risaralda	-0.02	0.01	0.74	0.15	0.52	-0.07	25.89
Santander	0.00	0.01	0.86	0.11	0.71	0.01	13.89
Sucre	-0.24	0.09	0.75	0.10	0.72	-0.96	25.32
Tolima	-0.09	0.05	0.71	0.15	0.51	-0.30	29.06
Valle	0.05	0.05	0.80	0.20	0.41	0.23	19.61
Mean	-0.11		0.68				
Median	-0.07		0.71				

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Note: Expected values ( $\alpha/(1-\rho)$ ) calculated only for  $\rho$  significant at 5% level and lower than 1.

Table 4.13: Autoregressive Processes of Order 1. Dependent Variable:  $\log(y_{i,t})$ . Per capita IDBH. 1975-2000.

Department	Intercept ( $\alpha$ )		Slope ( $\rho$ )		Regression Adjusted $R^2$	Expected Value	Implied $\beta$ (%)
	Coefficient	Sdt.Error	Coefficient	Std.Error			
Nuevos Departamentos	-0.02	0.02	0.71	0.17	0.44	-0.08	29.39
Antioquia	0.02	0.01	0.86	0.11	0.74	0.11	14.22
Atlántico	-0.01	0.01	0.84	0.09	0.78	-0.07	15.57
Bogotá D. C.	-0.04	0.03	1.04	0.04	0.96		
Bolívar	-0.06	0.03	0.83	0.10	0.76	-0.35	16.92
Boyacá	-0.04	0.03	0.60	0.23	0.23	-0.11	39.76
Caldas	0.00	0.02	0.93	0.06	0.91	-0.01	6.81
Caquetá	-0.14	0.08	0.76	0.13	0.59	-0.58	23.54
Cauca	-0.02	0.06	0.95	0.12	0.75	-0.32	5.03
Cesar	-0.01	0.03	0.95	0.10	0.79	-0.19	5.30
Córdoba	-0.13	0.07	0.79	0.11	0.68	-0.60	20.87
Cundinamarca	-0.01	0.01	0.90	0.07	0.88	-0.07	9.80
Chocó	-0.08	0.12	0.92	0.10	0.78	-1.07	7.87
Huila	-0.09	0.04	0.68	0.13	0.52	-0.27	32.45
La Guajira	-0.07	0.05	0.86	0.07	0.86	-0.50	13.79
Magdalena	-0.03	0.07	0.95	0.12	0.73	-0.49	5.46
Meta	-0.05	0.02	0.71	0.10	0.67	-0.17	28.72
Nariño	-0.23	0.13	0.71	0.16	0.46	-0.82	28.68
Norte Santander	-0.05	0.05	0.87	0.10	0.75	-0.40	12.91
Quindío	-0.04	0.02	0.74	0.12	0.61	-0.15	25.57
Risaralda	-0.04	0.02	0.63	0.14	0.47	-0.11	36.51
Santander	-0.02	0.01	0.65	0.18	0.35	-0.05	35.48
Sucre	-0.19	0.09	0.74	0.13	0.60	-0.71	26.37
Tolima	-0.01	0.03	0.96	0.09	0.82	-0.27	3.53
Valle	0.03	0.02	0.79	0.16	0.51	0.14	20.62
Mean	-0.05		0.82				
Median	-0.04		0.83				

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Note: Expected values ( $\alpha/(1-\rho)$ ) calculated only for  $\rho$  significant at 5% level and lower than 1.

Table 4.14: Mixture Model with 3 Components. Fitted with ML. Dependent Variable:  $\log(y_{i,t})$ . Relative Per capita PDB. 1975-2000.

Department	Group	Intercept $\alpha$	Slope $\rho$	Implied $\beta$ (%)	Expected value
Bolívar	1	-0.007	0.988	1.168	-0.618
Boyacá	1				
Caldas	1				
Caquetá	1				
Cauca	1				
Cesar	1				
Córdoba	1				
Chocó	1				
Magdalena	1				
Meta	1				
Nariño	1				
Quindío	1				
Risaralda	1				
Sucre	1				
Tolima	1				
Nuevos Departamentos	2	0.015	0.900	9.986	0.153
La Guajira	2				
Antioquia	3	-0.001	0.990	1.045	-0.139
Atlántico	3				
Bogotá D.C	3				
Cundinamarca	3				
Huila	3				
Norte Santander	3				
Santander	3				
Valle	3				

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Table 4.15: Mixture Model with 3 Components. Fitted with ML. Dependent Variable:  $\log(y_{i,t})$ . Relative Per Capita IDBH. 1975-2000.

Department	Group	Intercept $\alpha$	Slope $\rho$	Implied $\beta$ (%)	Expected value
Bolívar	1	0.000	0.982	1.757	-0.004
Caquetá	1				
Cauca	1				
Cesar	1				
Córdoba	1				
Chocó	1				
Magdalena	1				
Nariño	1				
Quindío	1				
Nuevos Departamentos	2	-0.013	0.961	3.893	-0.325
La Guajira	2				
Sucre	2				
Antioquia	3	-0.003	0.988	1.174	-0.239
Atlántico	3				
Bogotá D.C	3				
Boyacá	3				
Caldas	3				
Cundinamarca	3				
Huila	3				
Meta	3				
Norte Santander	3				
Risaralda	3				
Santander	3				
Tolima	3				
Valle	3				

Source: Own calculations based on data from CEGA. Constant prices of 1994.

Table 4.16: Bootstrap tests of equality of univariate density estimates, Income indicators

Variable	Years compared	p-value
Per capita PDB	1975, 2000	0.998
Per capita IDBH	1975, 2000	0.844

Test proposed by Bowman and Azzalini (1997). 1000 replications.

*Source:* Own calculations based on data from CEGA.

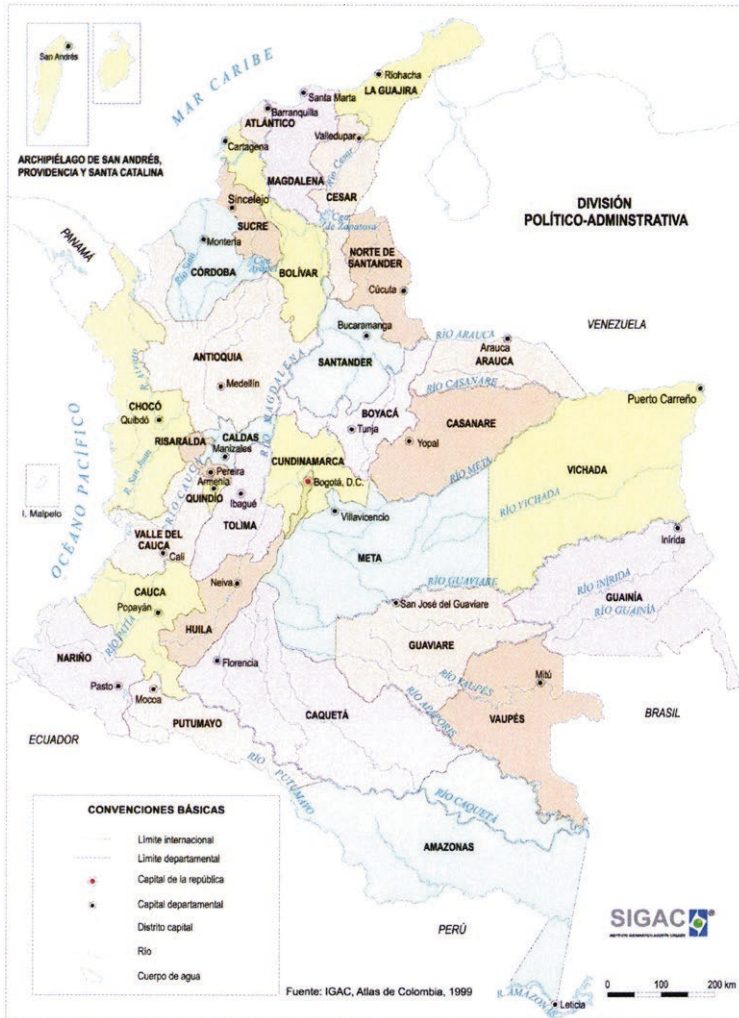
Table 4.17: Share of Total PDB and Share of selected PDB Sectors by Department. Years 1975 and 2000.

Department	Percentage of total PDB		Agriculture		Mining		Manufacturing		Financing		Government		Commerce	
	1975	2000	1975	2000	1975	2000	1975	2000	1975	2000	1975	2000	1975	2000
	Antioquia	15.3	14.4	12.2	13.9	21.3	4.0	21.0	18.0	14.1	15.7	15.5	10.9	18.4
Atlántico	5.3	4.3	0.8	0.7	0.0	0.0	7.5	6.1	4.8	4.0	2.7	3.4	6.4	5.1
Bogotá D.C	22.4	23.6	0.7	0.0	0.5	0.0	25.4	22.6	40.1	45.6	27.8	24.8	21.2	18.3
Bolívar	4.0	4.1	3.3	2.9	2.3	0.9	4.8	5.6	2.0	1.8	3.4	2.9	4.1	6.3
Boyacá	3.8	2.9	7.2	7.1	13.1	2.8	2.5	1.8	1.9	1.2	3.7	4.6	3.2	2.1
Caldas	2.5	2.0	3.7	2.6	0.7	0.1	2.2	2.0	2.3	1.8	2.4	2.9	2.0	1.9
Caquetá	0.5	0.5	1.2	1.6	0.0	0.0	0.1	0.0	0.1	0.2	0.7	0.9	0.2	0.2
Cauca	1.8	1.9	4.0	3.3	0.5	0.2	1.0	2.2	0.9	0.7	2.3	3.0	1.5	2.3
Cesar	1.6	1.5	3.7	3.0	0.0	5.5	0.6	0.4	1.3	0.5	1.2	1.8	1.5	0.9
Chocó	0.4	0.3	0.5	1.0	6.0	0.3	0.0	0.0	0.2	0.1	1.2	0.7	0.2	0.2
Córdoba	2.0	1.9	6.2	3.7	1.5	2.8	0.1	1.9	1.0	0.7	2.1	2.2	1.2	1.6
Cundinamarca	4.8	5.7	8.4	12.7	7.5	1.2	4.2	8.2	1.6	1.4	4.7	7.1	5.2	8.6
Guajira	0.4	1.3	0.6	0.6	3.6	12.3	0.1	0.0	0.4	0.3	0.6	0.8	0.3	0.1
Huila	1.6	1.7	3.1	2.5	1.5	5.4	0.4	0.6	1.2	1.0	2.0	2.3	1.1	1.1
Magdalena	1.8	1.5	3.6	3.4	1.9	0.0	0.4	0.3	1.1	0.7	1.8	1.9	1.0	0.8
Meta	1.2	1.5	2.4	2.8	0.3	5.1	0.4	0.7	1.2	0.9	1.0	1.3	0.9	1.4
Nariño	1.6	1.6	2.9	3.8	1.0	0.1	0.4	0.4	1.3	1.0	1.7	2.5	1.2	1.1
Norte de Santander	2.0	1.8	3.1	3.3	5.3	0.6	0.9	0.7	1.7	1.4	2.9	2.7	1.3	1.0
Nuevos	1.3	4.7	2.4	3.9	10.7	55.0	0.1	0.1	1.2	1.0	2.0	2.9	0.5	0.8
Quindío	1.2	1.1	2.2	2.3	0.0	0.0	0.6	0.3	1.1	1.1	1.4	1.2	1.2	0.6
Risaralda	2.0	1.9	2.1	1.5	0.1	0.0	2.2	2.3	2.0	1.6	1.6	1.7	2.0	2.0
Santander	5.2	5.0	6.2	6.9	18.9	2.1	5.2	6.5	3.8	3.5	4.0	5.3	5.0	7.7
Sucre	1.0	0.7	3.0	1.6	0.3	0.1	0.2	0.2	0.6	0.3	1.0	0.9	0.6	0.4
Tolima	3.2	2.6	7.0	6.0	0.9	1.3	1.1	1.9	2.3	1.6	4.1	3.1	2.7	2.7
Valle	13.2	11.4	9.3	8.8	2.0	0.3	18.5	17.2	12.0	11.8	7.8	8.3	17.1	16.1

Source: Own calculations based on data from CEGA. Constant prices of 1994.

### 4.8 Figures

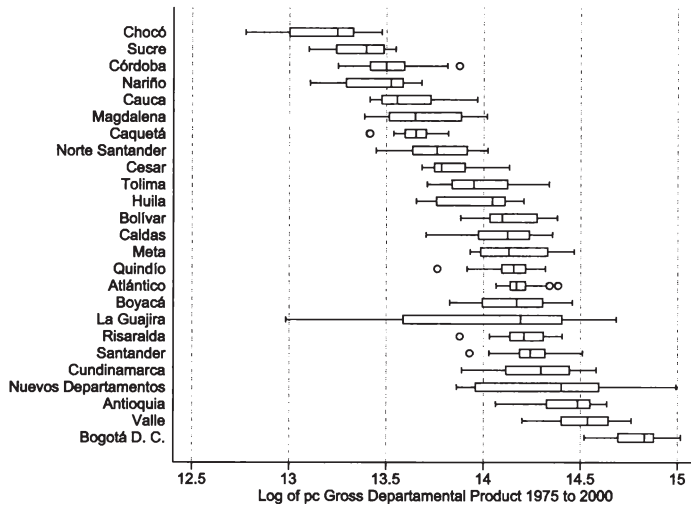
Figure 4.1: Map of Colombia.



Source: Instituto Geográfico Agustín Codazzi.  
 Labels in its order of appearance: International Limit, Departmental Limit, Country Capital, Capital District, River, Water.

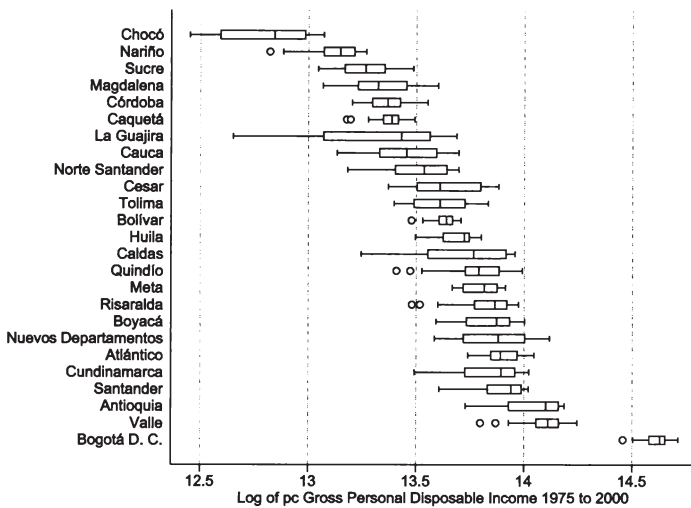


Figure 4.2: Box Plot: Log of Per Capita PDB. 1975-2000.



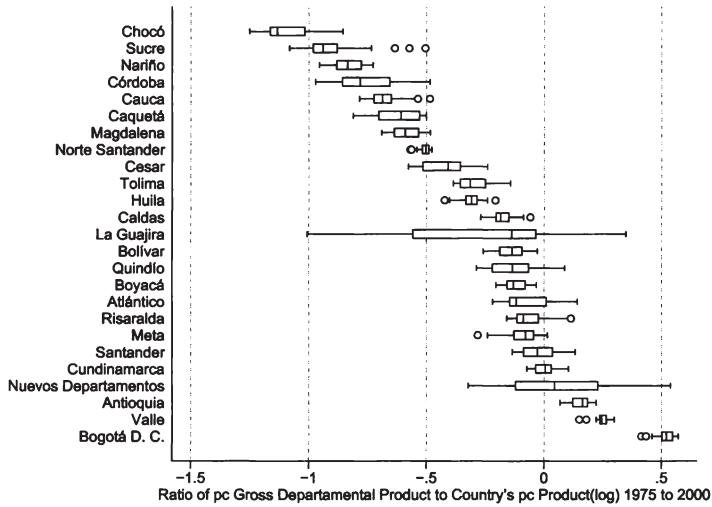
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.3: Box Plot: Log of Per Capita IDBH. 1975-2000.



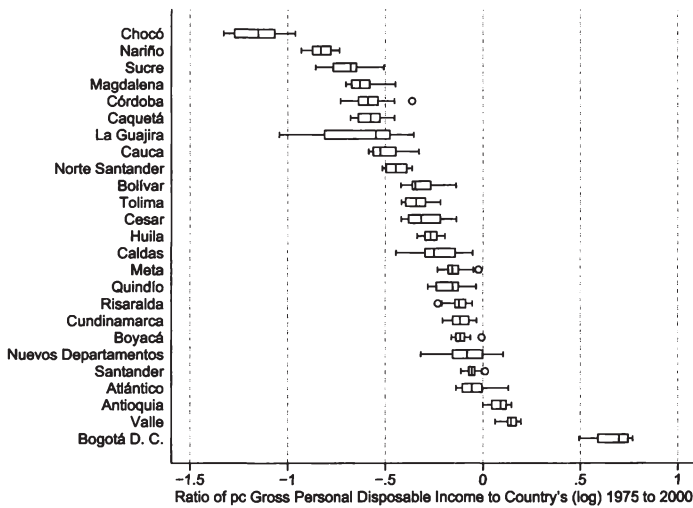
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.4: Box Plot: Log of Relative Per Capita PDB. 1975-2000.



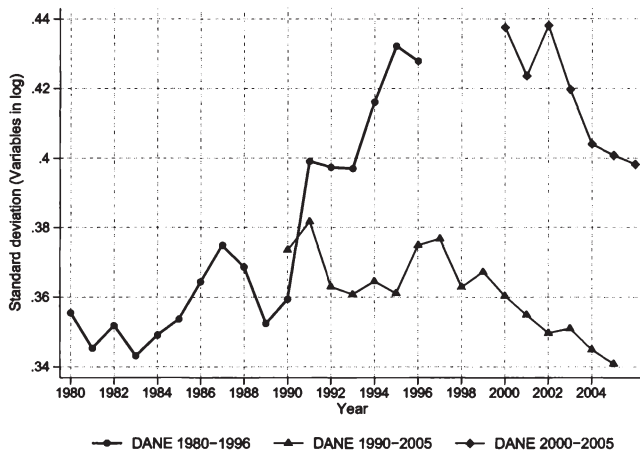
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.5: Box Plot: Log of Relative Per Capita IDBH. 1975-2000.



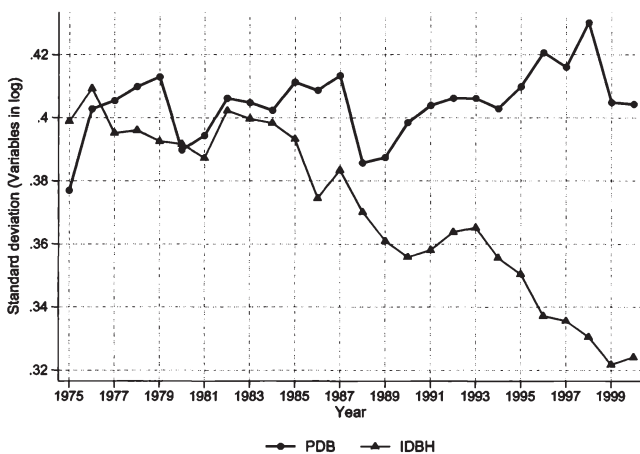
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.6: Sigma Convergence. GDP by Department.



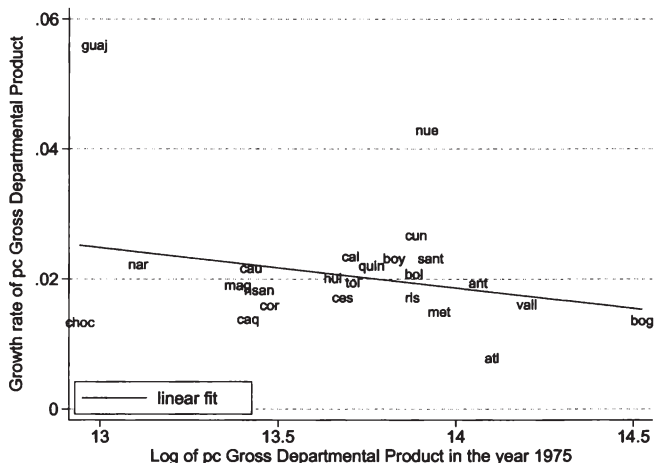
Source: Own calculations based on data from DANE.

Figure 4.7: Sigma Convergence. Per Capita Gross Departmental Product (PDB) and Gross Personal Disposable Income (IDBH). 1975-2000.



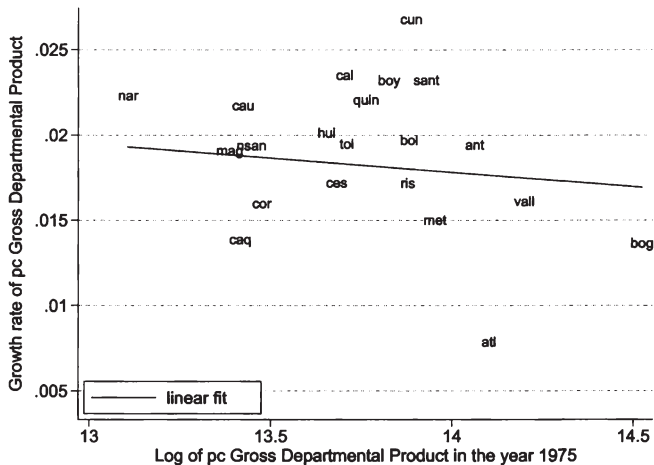
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.8: Beta Convergence. Per Capita PDB. 1975-2000.



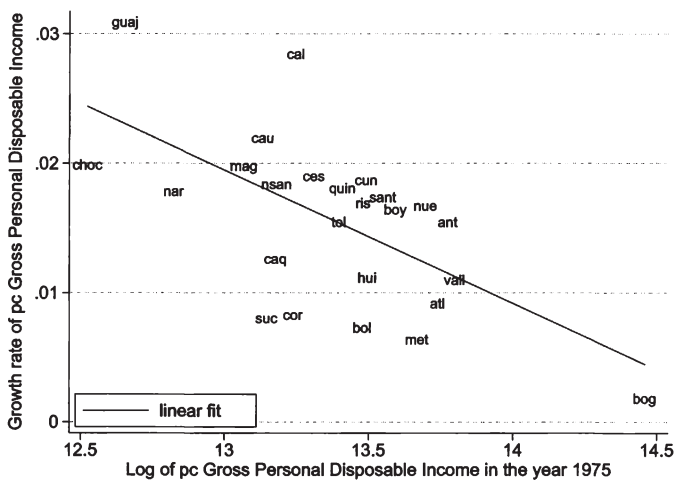
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.9: Beta Convergence without Nuevos, Chocó and Guajira. Per Capita PDB. 1975-2000.



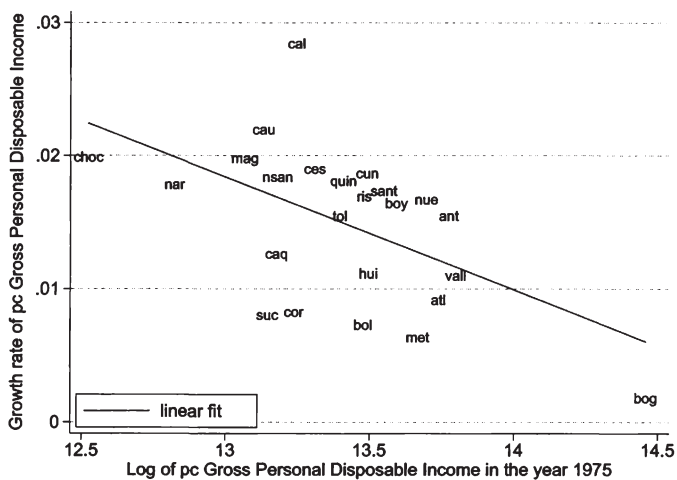
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.10: Beta Convergence. Per Capita IDBH. 1975-2000.



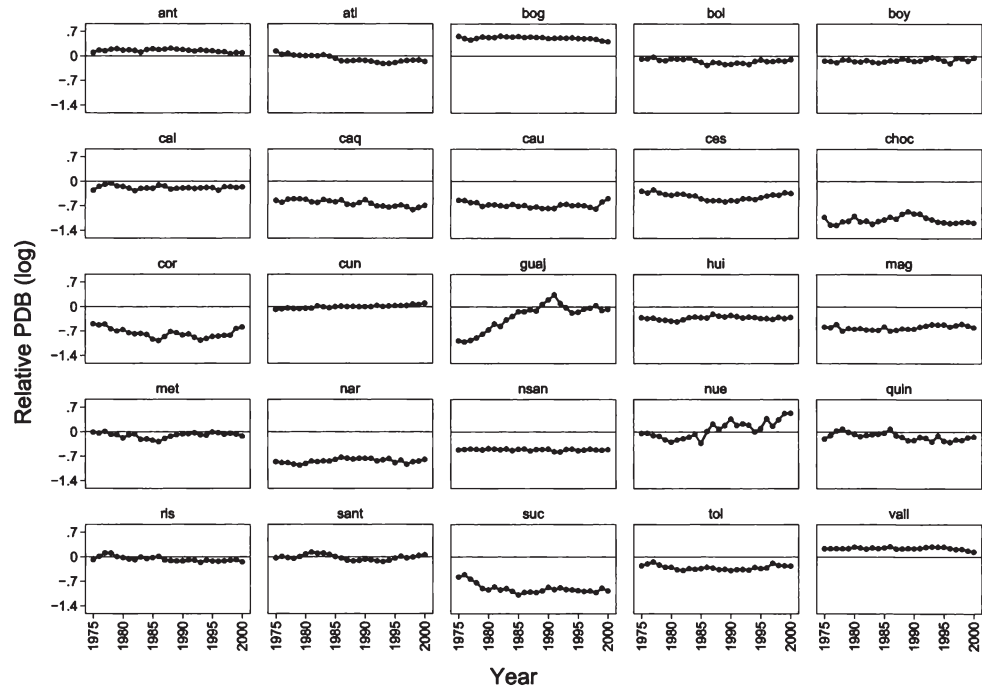
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.11: Beta Convergence without Guajira. Per Capita IDBH. 1975-2000.



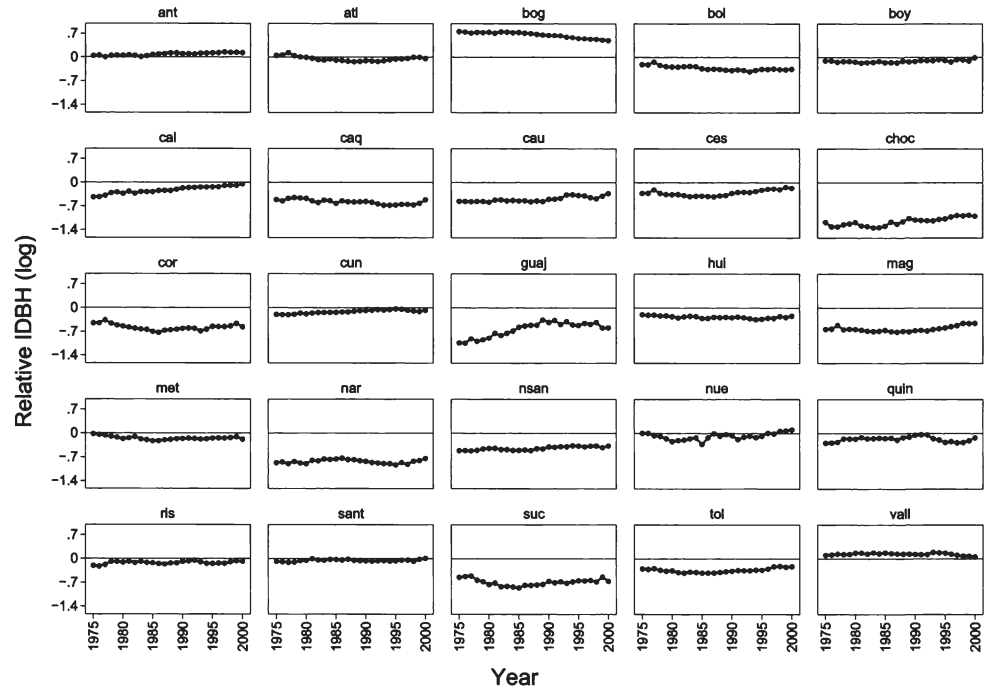
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.12: Log of Relative PDB by Department. 1975-2000.



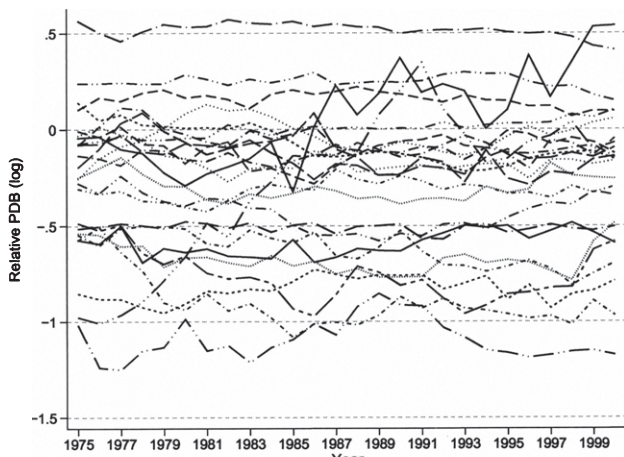
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.13: Log of relative IDBH by department. 1975-2000.



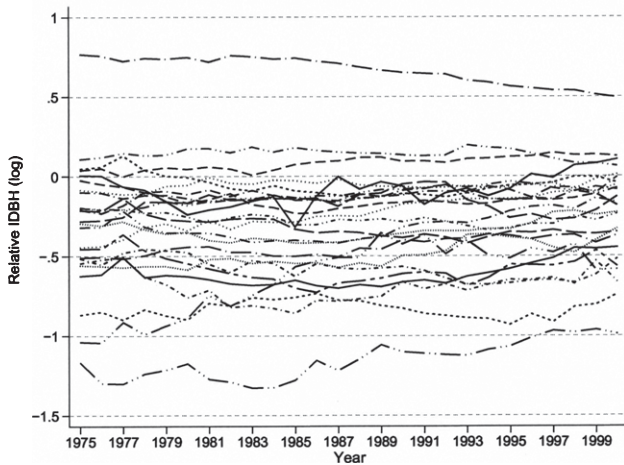
Source: Own calculations based on data from CEGA. Constant prices of 1994.

Figure 4.14: Log of Relative PDB. All Departments. 1975-2000.



Source: Own calculations based on data from CEGA. Constant prices of 1994.

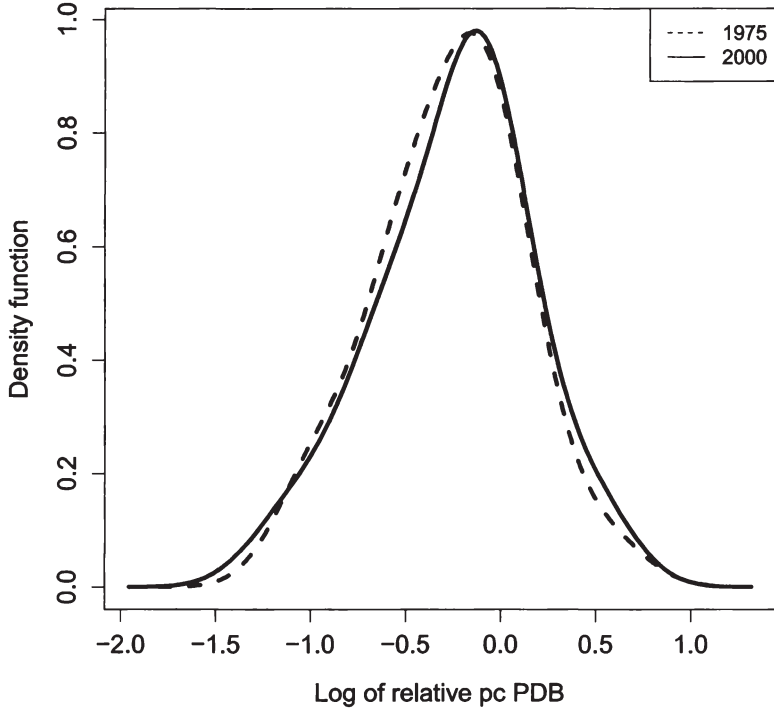
Figure 4.15: Log of Relative IDBH. All Departments. 1975-2000.



Source: Own calculations based on data from CEGA. Constant prices of 1994.

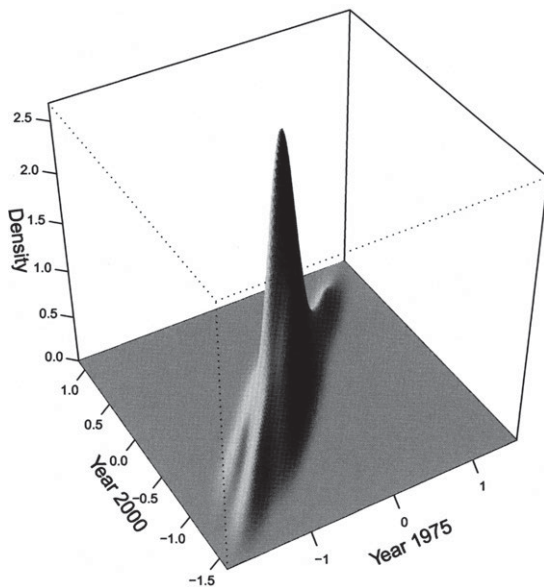


Figure 4.16: Univariate Kernel Density Estimators of Relative Per Capita PDB. Years 1975 and 2000. Constant Prices of 1994.



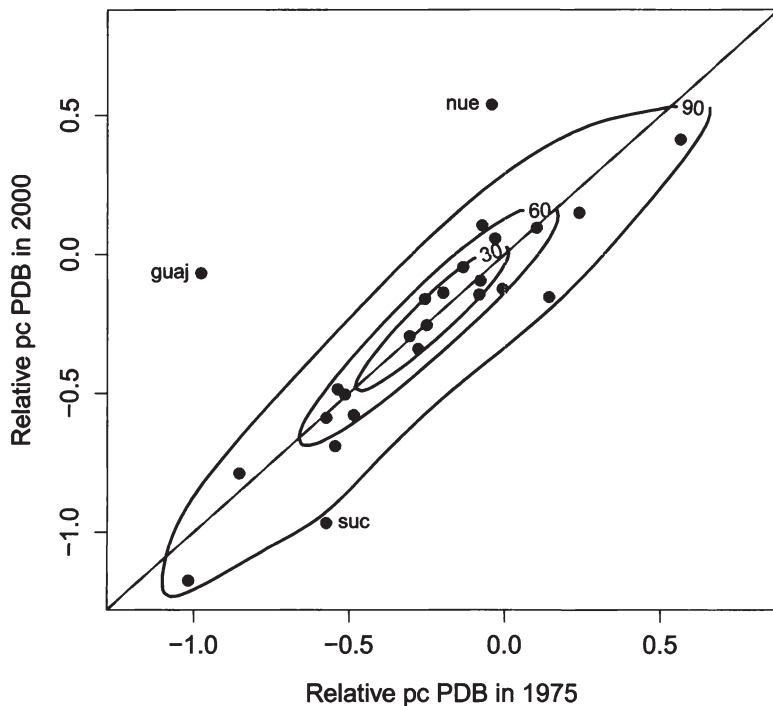
Source: Own calculations based on data from CEGA. Variables in logs

Figure 4.17: Relative Per Capita PDB Dynamics. Years 1975 and 2000. Constant Prices of 1994.



Source: Own calculations based on data from CEGA. Variables in logs

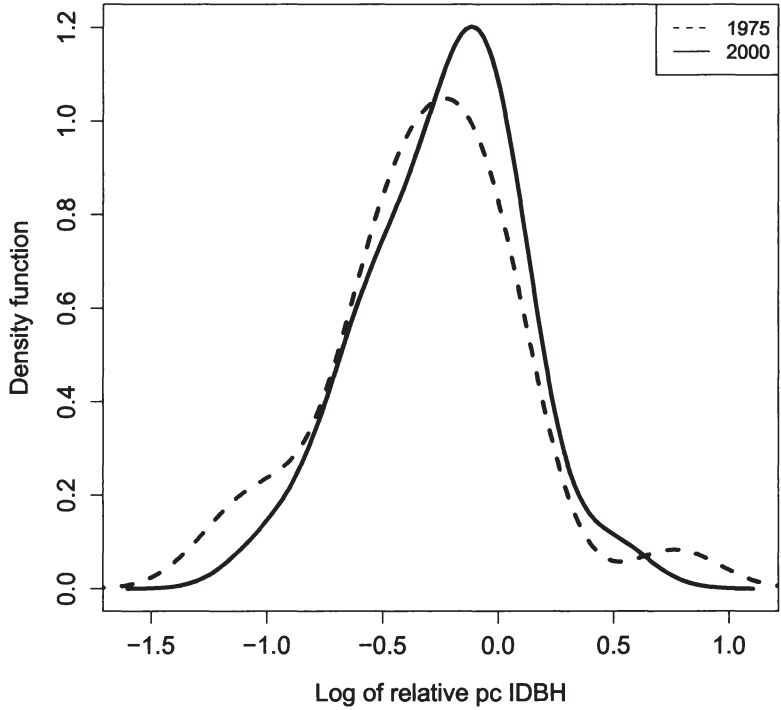
Figure 4.18: Relative per capita PDB Dynamics: Contour Plot. Years 1975 and 2000. Constant Prices of 1994.



Source: Own calculations based on data from CEGA. Variables in logs

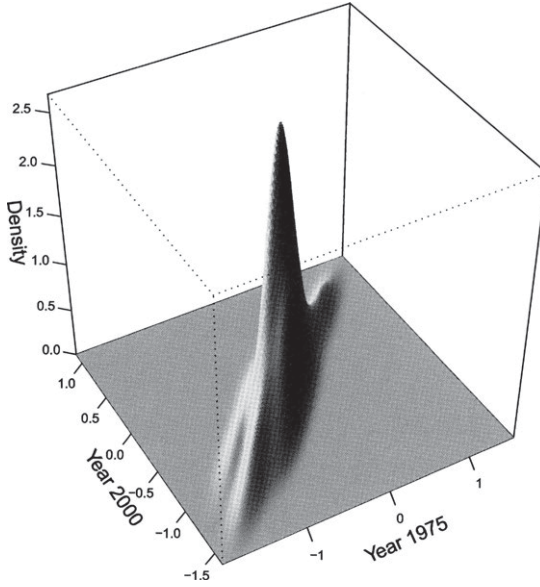
Note: Contours are drawn at 30%, 60% and 90%. The points represent the 25 observations. Points outside the contour level curves are identified. A 45 degree line is added to the plot.

Figure 4.19: Univariate Kernel Density Estimators of Relative per Capita IDBH. Years 1975 and 2000. Constant Prices of 1994.



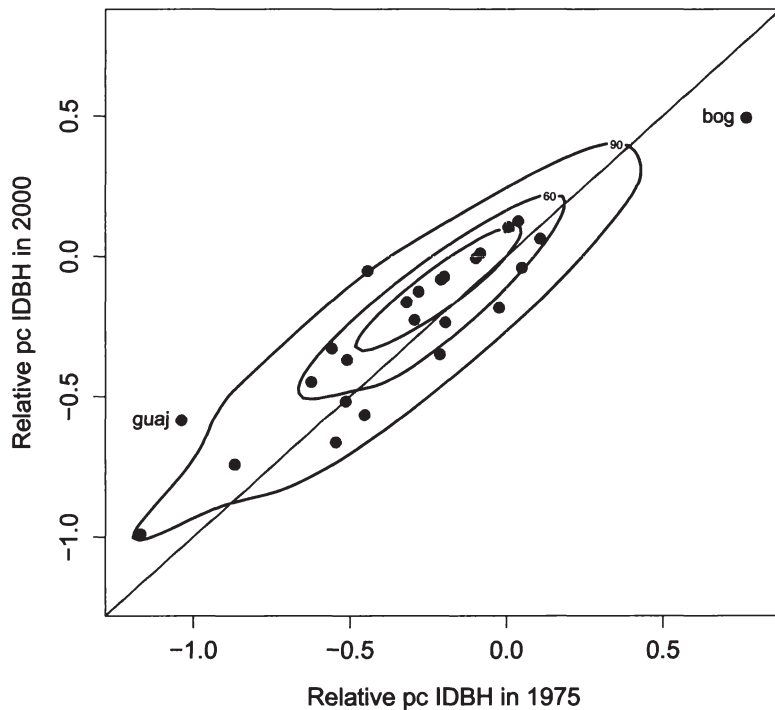
Source: Own calculations based on data from CEGA. Variables in logs

Figure 4.20: Relative Per Capita IDBH Dynamics. Years 1975 and 2000. Constant Prices of 1994.



Source: Own calculations based on data from CEGA. Variables in logs

Figure 4.21: Relative per Capita IDBH Dynamics: Contour Plot. Years 1975 and 2000. Constant Prices of 1994.



Source: Own calculations based on data from CEGA. Variables in logs

Note: Contours are drawn at 30%, 60% and 90%. The points represent the 25 observations. Points outside the contour level curves are identified. A 45 degree line is added to the plot.

## **Essay 5**

# **Regional growth convergence in Colombia using social indicators**

### **Abstract**

This paper investigates convergence in social indicators among Colombian departments from 1973 to 2005. We use census data and apply both the regression approach and the distributional approach (univariate and bivariate kernel density estimators). Using literacy rate as a proxy for education, we find convergence between 1973 and 2005, but persistence in the distribution between 1975 and 2000, when we use the infant survival rate and life expectancy at birth as proxies for health. Additionally, using data from Demographic and Health Surveys, we find some evidence of convergence in the rate of children that are well-nourished between 1995 and 2005.

## 5.1 Introduction

The majority of studies on convergence use macroeconomic aggregates to study whether poor countries are catching up with wealthier countries, and are particularly interested in the speed at which this process occurs, when it does occur. A large part of the empirical analysis of convergence is based on the neoclassical model developed by Solow (1956) and the estimation procedure suggested by Barro and Sala-i-Martin (1992a). Most of these studies use cross-sectional regressions based on per capita gross domestic product (GDP) to investigate if poor regions have higher rates of growth as they develop than wealthier areas.

In recent years, some authors have also tested if there is convergence in living standards, given that it is well-being that really matters, and arguing that per capita GDP is not the appropriate indicator for it.<sup>1</sup> For example, Neumayer (2003) and Kenny (2004) conclude that even in the absence of convergence in per capita GDP, there is convergence in living standards among poor and rich countries, a phenomenon praised by Neumayer as one of the greatest success stories of development in the last century.

This argument seems to be valid for cross-country analysis, but few studies exist investigating regional convergence in living standards in developing countries. Within a particular country, convergence analysis is important, not only to focus development assistance towards regions lagging behind in economic growth, but also to evaluate the efficiency and scope of public policies. Lack of convergence and the persistence of inequalities inside a country can lead to political instability, social unrest, and violence, given that people are concerned not only with their own improvements, but also with meeting the living standards of the wealthiest regions (Kenny, 2004). In Colombia, high regional inequality in living standards is one of the possible underlying causes of the ongoing domestic conflict, which is fueled by drug trafficking and corruption, particularly in isolated regions with a low level of governmental presence and a high incidence of poverty.

Over the past 30 years, Colombia has witnessed important reforms in social areas, particularly in health care and education. These reforms were marked by increased governmental intervention during the mid seventies, such as the creation of a national health care system and vaccination campaigns to eradicate tropical diseases like malaria, and, in the early eighties, policy formulations to reduce mortality and increase primary education. At the beginning of the nineties, decentralization accelerated in order to reduce the fiscal burden on the central government, making municipalities more responsible for generating and administrating their own resources, and partially moving the provision of health services to the private sector (Hernández and Obregón, 2002).

---

<sup>1</sup>In this paper, we will refer indistinctly to non-income indicators, social indicators, quality-of-life variables, and living standards.



The link between decentralization policies and the improvement in living standards in Colombia is a topic of extensive debate. While some researchers argue that the health policies have categorically failed, others conclude that they have been successful, although recognizing that there is still a large margin for improvement, particularly given that the system in place is still too recent to make a general judgement (Homedes and Ugalde, 2005; Barrera and Domínguez, 2006). Evidence shows that even though social spending increased considerably and reforms were designed to compensate municipalities with weak fiscal capacities, fiscal equalization has not been achieved.

Although Colombia has experienced improvements in living standard indicators at the national level, particularly those related to health, education, and access to public services, there is evidence of a heterogeneous distribution of improvements. Thus, investigating regional convergence in social indicators is relevant, as it is a relatively under-researched topic. Out of 20 research documents about regional convergence in Colombia produced in the last 15 years (Aguirre, 2008), only two published articles that deal with convergence in non-income indicators (Aguirre, 2005; Meisel and Vega, 2007).

The objective of this study is to analyze whether departments that were lagging behind in social indicators in the 1970s were able to catch up by the year 2000. We use both the regression approach suggested by Barro and Sala-i-Martin (1991), among others, and the distributional approach pioneered by Quah (1997) to test for convergence in three variables, based on census data. Using literacy rate as a proxy for education, we find regional convergence between 1973 and 2005, but persistence in the distribution between 1975 and 2000 when we use the infant survival rate and life expectancy at birth as proxies for health. Additionally, using data from Demographic and Health Surveys (DHS), we find convergence between 1995 and 2005 in the percentage of children that are well-nourished.

This paper is divided into six sections. After this introduction, we argue in section 5.2 that it is fundamental to focus upon indicators other than income and that the analysis of convergence at the department level is relevant for Colombia. We present two methods to empirically test for convergence in section 5.3. Section 5.4 describes the data used and the details of our empirical estimation. Results are presented in section 5.5 and discussed in section 5.6.

## 5.2 Motivation

The importance of analyzing social indicators, instead of focusing only upon income, has been extensively discussed in the work of Amartya Sen, who argues that *social opportunities* are one of the five types of instrumental freedoms that contribute to the overall freedom

people have to live as they choose (Sen, 1999). Social opportunities are understood to be the arrangements that a society makes which influence the individual's substantive freedom to live better, such as providing education and health care, which are important, not only for the citizens in the conduct of their private lives, but also for more effective participation in societal economic and political activities.

Public policies have a crucial influence on social opportunities (e.g., they can influence longevity through epidemiological policies and education through the provision of the corresponding facilities). Thus, it is important to shift attention to elements that affect individual well-being and freedom but which are not captured by income statistics (Sen, 1997).

Particular attention shall be given herein to health status when analyzing social inequalities. Sen (1998) argues that mortality is a key economic indicator, given that it mirrors the success or failure of a society. Indicators like the infant survival rate, which responds very rapidly to public health policies, are central to that kind of analysis (Mazumdar, 2003). Along the same lines, indicators referring to the level of nutrition reflect one of the most basic needs for survival, namely access to adequate food (Sen, 2002).

In Colombia, awareness of the influence and scope of the government on health and education gradually increased in the second half of the 20<sup>th</sup> century, and has translated into programs, as well as legislation, aimed at achieving universal access to health care and primary education. Policies for both sectors experienced important changes in the last quarter of the century, reflecting the transition from a centralist system to a locally managed one.

In the seventies, policies on education focused upon reducing illiteracy and increasing the coverage of primary education, particularly in rural areas, where public schools and teachers were almost nonexistent.<sup>2</sup> However, rural areas continued to lag behind. Adult education and literacy campaigns continued through the eighties, together with the expansion of secondary and pre scholar education.

Due to macroeconomic imbalances at the end of the eighties, the government initiated a decentralization program to reduce the financial burden on the central government, and transferred substantial revenues and responsibilities to local administrations. The country has been praised as a leading example of fiscal decentralization, which policy had the objective of increasing social expenditures and efficiency in social sectors, as well as financially compensating territories with weak fiscal capacities. This process accelerated with the new constitution in 1991 (Rojas, 2003; Barrera and Domínguez, 2006). As a result, social spending increased from seven to 15 percent of GDP between 1991 and 2001. Concerning education, evidence shows that reforms were beneficial for the urban sector, but less so for

---

<sup>2</sup>With society facing limited resources, secondary and tertiary education was left largely to the private sector under the argument that spending in public universities would favor middle and upper income groups.

the rural sector, which lags behind the former on the issues of quality of education and net enrolment rates, particularly for secondary education (Velez et al., 2003).

Concerning health policies, by the mid seventies, the government had increased its role as a provider of health care and had implemented programs to improve neonatal care and nutrition, eliminate tropical diseases through mass vaccinations, and promote reproductive health. At the same time, coverage of public services in the country expanded. An important result of these campaigns was the reduction in infant mortality, due in part to better access to drinking water. This reduction led to higher life expectancy at birth (Profamilia, 2005).

In the nineties, the policy orientation shifted; the government increased the role of the private sector as health provider and tried to strengthen the national health care system through the decentralization of services at the local level. Implementation of a dual system, combining the contribution of formal sector employees with subsidies for the population outside the system, yielded an increase in health coverage, which attained 58 percentage in 2000 (Hernández and Obregón, 2002).

Statistics on health and education at the national level show significant improvements in the last quarter of the 20<sup>th</sup> century, and indicate that the country is a successful case within Latin America. Colombia ranks well within the Latin American average in many social indicators, or even slightly above. As an example, in 1970, the average adult literacy rate was 78 percentage and increased to 93 percentage in 2005, which is higher than the Latin American average of 90 percentage (World Bank, 2008).<sup>3</sup> However, these gains do not seem to have been homogeneously distributed across society. Inequality is anchored within different levels - between departments (the main administrative units), between urban and rural areas, and also inside urban areas.<sup>4</sup> Given that Colombia is a land of contrasts, where one can find well-developed modern metropolitan areas with reasonable infrastructure and large underdeveloped low density areas, lacking even basic infrastructure, looking at variables at the national level clearly masks differences.

Colombia is divided into 32 departments and the capital district of Bogotá, a division that has existed formally since the 1991 constitutional reform, when nine former intendancies and commissariats, sparsely populated areas, were acknowledged as departments (Amazonas, Arauca, Casanare, Guainía, Guaviare, Putumayo, San Andrés y Providencia, Vaupés,

---

<sup>3</sup>Similarly, in 1980, live expectancy at birth was 66 years, and in 2005, it reached 73 years, which is the average for the region. Infant mortality for each 1,000 live births fell from 68 deaths in 1970 to 17 in 2005, compared to a regional average of 23 (World Bank, 2008).

<sup>4</sup>As an example of inequality, the income Gini coefficient, which had modest decreases in the eighties, reached 0.56 in 2004, almost as high as the value for Brazil for the same year.

and Vichada).<sup>5</sup> Departments are important political entities in Colombia, with elected local governments and separate department assemblies. Hence, it is relevant to analyze the performance in social indicators by department, investigating whether those that were lagging behind in the early seventies have been able to catch up in the year 2000.

In this paper, we test the hypothesis that among Colombian departments, there was convergence in social indicators in the period from 1973 through 2005. This hypothesis is based on the following facts. First, starting in the mid seventies, national policies aimed at reducing illiteracy and improving neonatal care were put into place. Second, decentralization reforms were conducted, starting in the late eighties, to increase the efficiency of spending on education and health care. Third, after decentralization, social spending, as a share of total gross domestic product, doubled. Fourth, as will be discussed later, social indicators are naturally bounded, which fact facilitates convergence if better-off departments are close to the upper bound in the initial year. Finally, empirical results presented in Essay 4, using per capita gross disposable income<sup>6</sup> as a proxy for well-being, suggest slow convergence among Colombian departments between 1975 and 2000.

We explain in the next two sections the methods used to test for convergence and the variables included in the analysis.

### 5.3 Methods for Measuring Convergence

We will consider two alternatives to empirically test for convergence. The first one is the *regression approach* (Magrini, 2004), also called the *classical approach to convergence analysis* (Sala-i-Martin, 1996), which is the most frequently used analysis in the literature. Robert Barro and Xavier Sala-i-Martin are among the best known authors associated with it (Barro and Sala-i-Martin, 1991, 1992a,b, 2004; Sala-i-Martin, 1996). The second alternative is the *distributional approach to convergence*, pioneered by, among others, Danny Quah (Quah, 1993a,b, 1996, 1997). Both approaches are presented briefly in this section.

Within the classical approach to convergence analysis, the concepts of beta-convergence and sigma-convergence are relevant. Beta-convergence is related to the mean-reversion of the variable of interest. This is typically done by regressing the average growth rate of the variable of interest at the initial level.<sup>7</sup> If the regression coefficient is negative and statis-

<sup>5</sup>As it is often done in studies with Colombian data at the department level due to lack of data, we group them as one unit under the name Nuevos departamentos.

<sup>6</sup>IDBH is the abbreviation, based on the Spanish denomination Ingreso Departamental Bruto disponible de los Hogares (CEGA, 2006b).

<sup>7</sup>To be precise, we are discussing here *absolute* or unconditional beta-convergence, which assumes that regions are structurally similar.

tically significant, this means that the variable tends to grow more quickly in regions that lagged behind at the beginning of the period considered.

A reduction over time in the dispersion of the variable of interest across entities (in our case, departments) indicates a more equitable distribution and is known as sigma-convergence. Testing for sigma-convergence is performed by checking the evolution of the standard deviation over time, or the coefficient of variation if the mean of the variable changes. The existence of beta-convergence tends to generate sigma-convergence. Beta-convergence is a necessary, but not a sufficient condition for observing sigma-convergence. Sigma-convergence is an indicator of dispersion of departments, but does not tell much about the mobility of each one.

Quah (1997) criticizes the classical approach to convergence analysis, arguing that neither beta-convergence nor sigma-convergence can deliver useful answers to the question of whether poor countries or regions are catching up to wealthy ones. Quah argues that the classical approach does not provide any information about mobility, stratification, or polarization, and suggests analyzing the distribution dynamics directly. One alternative proposed by him is to work with a sequence of distributions and, after discretizing the space of values, to count the observed transitions into and out of the distinct cell values and construct a transition probability matrix (Quah, 1993a,b).

Quah (1997) warns, however, that a discretization could distort dynamics if the underlying observations are indeed continuous variables. He therefore suggests not to discretize at all, and rather to think of the distinct cells as tending to infinity and to the continuum, with the transition probability matrix tending to a matrix with a continuum of rows and columns, i.e., becoming a stochastic kernel.<sup>8</sup> In particular, the proposed methodology is based upon tracking the evolution of the entire cross-section distributions across regions over time through the estimation of kernel densities for 'relative' variables. By relative variables, we mean that the variables of interest are expressed as relative to the national average, which allows abstraction from changes in the mean when we look at how the distribution changes.

Empirically, in a graph showing how the cross-sectional distribution of the relative variable of interest changes between two periods, if most of the mass of the estimated bivariate kernel density is concentrated along the 45-degree diagonal, then regions basically remain where they started. We will refer to this situation as *persistence* in the distribution of the relative variable of interest.

---

<sup>8</sup>For a technical derivation of a stochastic kernel, see Quah (1997, section 4).

## 5.4 Data and Empirical Estimation

The convergence analysis in our paper is done at the department level. Using two cross-sections per variable, we treat each department as an observation and do not use population weights. We are interested in investigating whether departments that were lagging behind were able to catch up, and consider this to be a pertinent question in the Colombian case where, as mentioned in section 5.2, departments are important political entities.

In this section, we deal briefly with the selection of variables, the transformation of the variables needed in some cases, and the particular choices we apply for the empirical estimation.

### 5.4.1 Data

As discussed by Micklewright and Stewart (1999), many quality-of-life variables have a complement (e.g., the infant survival rate is 1,000 minus the infant mortality rate). They warn that sigma-convergence results may depend upon whether one uses a variable or its complement. Kenny (2004) argues for measuring convergence towards a maximum and not towards zero, claiming that the latter approach favors small absolute changes, close to zero, above large absolute changes, further from zero. Additionally, he claims that convergence towards the maximum (i.e., a positive value) is what the majority of the literature on global trends does. We follow these arguments and use in this study '*positive*' variables. By this we mean that we transform the variables so that they are, in theory, positively correlated with living standards.

We have tried to obtain data at the department level for years close to 1975 and 2000, as we additionally want to compare our regional convergence results with those found for two income measures in Essay 4 for the period from 1975 to 2000. Evaluating a period of 25 years to investigate convergence seems reasonable as the time span roughly represents a generation. Our main source of data at the department level is DANE.<sup>9</sup> It kindly provided illiteracy rate data that were computed from information obtained in censuses for the years 1973 and 2005. For health data, we obtained infant mortality rates and life expectancy at birth. Data for the year 1975 are from DANE (1990) and for 2000, from DANE (2007). As explained before, we transform two of the variables so that they are 'positive' - we work with literacy rates instead of illiteracy rates, and with infant survival rates instead of infant mortality rates.

---

<sup>9</sup>Departamento Administrativo Nacional de Estadística. DANE is the official statistical agency in Colombia (<http://www.dane.gov.co/>).

The literacy rate (LIT) is the percentage of literate population above age 5. Being illiterate can be considered a deprivation of a very basic capability. As argued by Sen (1999), basic education can additionally be considered a semi-public good, as it is not only the literate person who benefits from it, but society in general, for example through the reduction in fertility and mortality.

The infant survival rate (ISR) is the number of babies that survive until their first birthday out of every 1,000 live-born babies during a particular year. It is a measure of nutrition and hygiene in the first months of life, and also reflects the degree of the existence of contagious diseases (Mazumdar, 2003). There are many empirical studies that show that women's education and literacy tend to increase the survival rates of children (Sen, 1999).

Life expectancy at birth (LEX) is the average number of years that a newborn is expected to live if current mortality rates continue to apply. This variable reflects the level of health care, nutrition, and income. However, at least at the cross-country level, some studies show that the connection between income and life expectancy works mainly through two channels, public expenditure on health care and the success of poverty eradication efforts (Sen, 1999).

In addition to the data provided by DANE, we use data from Demographic and Health Surveys for Colombia for the years 1995 and 2005 containing information about child nutrition at the department level. As this variable covers a shorter time period than the other three variables and is based upon data that are not always representative at the department level, we will treat the results based on the 'positive' variable, well-nourished rate, carefully.

Our well-nourished rate (WR) is defined as 100 minus the percentage of children which are underweight. Underweight means insufficient weight for age and is commonly used as a summary indicator of undernutrition (UNICEF, 1998). Undernutrition depends upon both food intake and the ability to make nutritive use of it, which ability is influenced by general health conditions that depend on health care and public health provisions (Drèze and Sen, 1989; Sen, 1999).

### 5.4.2 Empirical estimation

As the mean of each variable of interest has increased in the period considered, to test for sigma-convergence we look at the standard deviation in the initial period and in the final period.

For testing for beta-convergence, we follow one of the estimations used by Bloom and Canning (2007). We run regressions as

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad (5.1)$$



where  $x_i$  is the initial value of the variable of interest for department  $i$ , and  $y_i$  is the change in the variable of interest in the period considered.<sup>10</sup> We assume that  $\varepsilon_i \sim N(0, \sigma^2)$  and estimate the regressions with ordinary least squares (OLS). We use HC3 robust standard errors, as proposed by Davidson and MacKinnon (1993), to account for possible heteroscedasticity, considering that the number of observations is relatively small (Long and Ervin, 2000). We are interested in checking whether the estimated coefficient  $\beta$  is negative and statistically different from 0 at the 5 percentage level, meaning that lower initial levels of the variable of interest are associated with larger improvements in the periods considered. To check whether the results are robust, after the regression, we compute Cook's distance to detect observations that have an unusual influence or leverage, and re-run the regressions on the restricted sample, excluding those observations.

For the distributional approach, all variables are expressed relative to the Colombian value, as was explained in section 5.3. We additionally take the logarithm of the relative variable, as it facilitates the comparison to the national level. Expressed in logs, a relative value that is equal to 0 means that the department has the same value as the country, while a value that is, for example, equal to -0.05 means that the value of that department is 5 percentage lower than the national value.

Before we define how we proceed to test for convergence using the distributional approach, we briefly present some concepts needed for our estimation.<sup>11</sup>

A univariate kernel density estimate can be regarded as a generalization of a histogram. It has the form

$$\hat{f}_h(q) = \frac{1}{nh} \sum_{i=1}^n \kappa\left(\frac{q - Q_i}{h}\right), \quad (5.2)$$

where  $\kappa$  is a kernel<sup>12</sup>,  $h > 0$  is the bandwidth, also called the smoothing parameter, and  $n$  is the number of observations.

In the context of convergence, we are interested in checking whether we find unimodality or multimodality in the estimated univariate densities of the relative variable of interest in both periods, and in determining how the estimated densities changed.

<sup>10</sup>We have also tried a specification proposed by Barro and Sala-i-Martin (1992a) that uses the average growth rate as the dependent variable and a function of the logarithm of the initial value as a regressor, obtaining similar results.

<sup>11</sup>A review of the statistical principles of univariate and multivariate kernel density estimation can be found in Härdle et al. (2004), for example.

<sup>12</sup>Kernel refers to any smooth function satisfying the conditions  $\kappa(q) > 0$ ,  $\int \kappa(q) dq = 1$ ,  $\int q \kappa(q) dq = 0$ , and  $\sigma_\kappa^2 \equiv \int q^2 \kappa(q) dq > 0$  (Wasserman, 2006). In kernel density estimation, the choice of the kernel does not have a large impact on the estimation, but the choice of the bandwidth does.



Bivariate kernel density estimation requires two-dimensional data and a two-dimensional kernel. Here,  $Q = (Q_1, Q_2)^T$ , and the kernel  $K$  maps  $\mathbb{R}^2$  into  $\mathbb{R}_+$ . The estimate has the form

$$\hat{f}_H(q) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\det(H)} K\{H^{-1}(q - Q_i)\}, \quad (5.3)$$

where  $K$  is a bivariate kernel function,  $H$  is a symmetric bandwidth matrix, and  $n$  is the number of observations.

For the analysis of convergence, we estimate the bivariate kernel density for the relative variable in two periods and check whether a large portion of the probability mass remains clustered around the 45-degree diagonal, which would indicate persistence in the distribution. We present the 3D representation of the estimated bivariate density and a contour plot showing the highest density regions.

All the results for kernel density estimation presented in section 5.5 were computed with the statistical software R (R Development Core Team, 2008) and the package *ks*.<sup>13</sup> For both univariate and bivariate kernel density estimation, we use gaussian kernels and smoothed cross validation (SCV) bandwidth selectors<sup>14</sup> (Jones et al., 1991; Duong and Hazelton, 2005). As recommended by Bowman and Azzalini (1997), in the univariate case we use for both years the same bandwidth, which is computed as the mean of the two selected bandwidths obtained for each year separately. In the bivariate case, the smoothed cross validation in unconstrained, i.e., we do not impose the requirement that the (nonsingular) bandwidth matrix  $H$  has to be diagonal. Hence, we are able to handle correlation between components, as we allow kernels to have an arbitrary orientation (Wand and Jones, 1995). As we are especially interested in checking whether a large portion of the probability mass remains clustered around the 45-degree diagonal, this flexibility is relevant for us. If we were to impose a diagonal matrix  $H$ , only kernels which are oriented to the coordinate axes would be allowed.

## 5.5 Results

We address in this paper the question of regional “positive convergence” in Colombia (Micklewright and Stewart, 1999); that is, we investigate whether departments which were lagging behind at the beginning of the period in certain variables of interest proxying for health care and education have been able to catch up in a period that has been one of improvement in

<sup>13</sup>*ks* is currently the most comprehensive kernel density estimation package in R (Duong, 2008). All the estimations were done with the function *kde*.

<sup>14</sup>We also tried direct plug-in methods for bandwidth selection suggested by Sheather and Jones (1991) and obtained results that are not very different.

average.<sup>15</sup> As can be observed in the descriptive statistics of the variables (Table 5.1), there has been a general improvement in all the variables chosen as proxies for living standards.

### 5.5.1 Literacy Rate

Table 5.1 shows that between 1973 and 2005, the average departmental literacy rate (LIT) increased from 70 percent to 86 percent. We find strong evidence of convergence in literacy rates. Both sigma-convergence (Table 5.1) and beta-convergence are observed (Tables 5.2 and 5.3). In the OLS regression with all available observations, Bogotá and La Guajira are identified as having an unusual influence on regression results; the coefficient of the initial level remains negative and statistically significant, however, if one excludes these two departments. In Bogotá literacy rates were already high in 1973 (90 percent), making it more difficult to achieve further improvements. La Guajira is in the opposite position, a department with very low literacy rates in 1973 and with a minor improvement in the 32 years of analysis. As suggested by Meisel (2007a) in a study of this department, illiteracy is widespread in the indigenous population, consisting of the Wayúu group, which predominantly lives in rural areas. Estimations of this author show that around 80 percent of the Wayúus had not even finished primary school in 2005.

The univariate density estimates in Figure 5.2 show that the distribution has narrowed between 1973 and 2005. The distribution in 2005 shows three modes, suggesting that some departments lag behind even if the dispersion decreased. According to the results of a bootstrap test of equality of the estimated densities for both years (Bowman and Azzalini, 1997) in Table 5.4, we can soundly reject the hypothesis that the two densities are identical. In Figure 5.4, one can observe a clear pattern of convergence, as most of the mass of the estimated bivariate density is concentrated in an axis that is flatter than the 45-degree line. Nevertheless, the case of La Guajira raises attention; it was among the worst relative performers in 1973, and in 2005, it was the worst relative performer. La Guajira had literacy rates that were 28 percent lower than the national average in 1973 and 33 percent in 2005.

Although one can praise improvements for the other departments that were lagging behind in 1973, it is important to note that the literacy rate only indicates the existence of a basic education level, which is definitely important, but probably not adequate.<sup>16</sup> Even considering this very basic indicator, it is worrisome that in many departments (La Guajira, Chocó,

---

<sup>15</sup>“Negative convergence,” (European Commission, 1996) a situation of general deterioration towards the standard of the worst, is not relevant for Colombia in the period considered.

<sup>16</sup>Additionally, as discussed in Velez et al. (2003), there is some evidence that despite large and increasing public expenditures on public education, particularly after decentralization, the quality of education is decreasing.

Sucre, Córdoba, Magdalena, Caquetá, Cesar, Nariño, and Bolívar), more than 15 percent of the population is still illiterate. Unfortunately, we have not been able to access data at the department level that proxies higher levels of education.

### 5.5.2 Infant Survival Rate

Between 1975 and 2000, the average departmental infant survival rate per thousand live births improved, increasing from 936 to 962 (Table 5.1). Results show that the standard deviation of this indicator decreased slightly, suggesting mild sigma-convergence. However, this result tells us little about how exactly the distribution of departments changed.

When looking at beta-convergence, we find a negative coefficient for infant survival rate in 1975, but it is not statistically significant unless we exclude the department of Chocó from the regression (Tables 5.5 and 5.6). This exclusion is based on unusual influence or leverage and indicates that this department has an important influence on the lack of convergence. Chocó had the lowest performance in 1975 and even though infant survival rate improved there, by 2000, it was the furthest from the departmental average. Notice that the infant survival rate in Chocó in 2000 was lower than the departmental average in 1975 (Table 5.1).

Figure 5.5 shows that after Chocó, departments with the lowest starting infant survival rates were Nariño, Caldas, Risaralda, and Quindío. Of these, Caldas, Risaralda, and Quindío achieved the largest improvements during the period of analysis. Changes in departments that were close to the average in 1975 vary considerably. Antioquia, Norte de Santander, and Valle experienced important improvements, while Cauca, Córdoba, Guajira, and Bolívar stagnated.

Kernel density estimates allow a closer look at changes in the distribution in relative terms. In Figure 5.6 we observe that the density was slightly narrower in 2000 than in 1975. Both years had a bimodal distribution. The bootstrap test of equality of the estimated densities for both years shows that we cannot reject the hypothesis that the two densities are identical (Table 5.4). The bivariate kernel in Figure 5.7 and the corresponding contour plot in Figure 5.8 suggest persistence in the infant survival rate, as most of the estimated density is concentrated along the 45-degree diagonal. Thus, in relative terms, departments basically remained where they were in 1975. The departments of Chocó, Nariño, and Cauca are outside of the 90 percent contour.

To summarize, coastal regions have the lowest infant survival rates, although the rates have improved over time. The infant survival rate is particularly low in 2000 in the Pacific region (e.g., 912 children per 1,000 births in Chocó) where population density is low and a large share of the population lives in rural areas with precarious sanitation infrastructure. This department is also prone to a higher prevalence of tropical diseases, given that it is one

of the rainiest and humid regions in the world. Transport between small rural villages along the river shores, which are prone to floods, is possible only within the extensive network of rivers. This also explains why the the scope and reach of health programs are limited.<sup>17</sup> In contrast, an important increase in infant survival rates has been observed in central regions located along the main transport corridors of the country. Increased vaccination was achieved in large cities and in departments where agglomeration around urban centers allows easier access to the population.

### 5.5.3 Life Expectancy at Birth

In Colombia, the average departmental life expectancy at birth was 62 years in 1975 and increased to 70 years in 2000 (Table 5.1). The standard deviation decreased, suggesting sigma-convergence. Beta-convergence analysis shows a negative relationship between the starting value in 1975 and its change up to 2000 (Figure 5.9, Tables 5.7 and 5.8). The regression coefficient is negative and statistically significant when including all observations, but it is insignificant once we exclude influential observations, in this case Chocó and Nariño, which experienced large improvements. In Nariño, people born in 2000 were expected to live 14 years longer than those born in 1975, and in Chocó the gain was 13 years. Once again the coffee-growing region (Caldas, Quindío, and Risaralda) had outstanding improvements. Sucre, located in the Atlantic coast, had the longest life expectancy at birth by 2000, even exceeding that of the capital district of Bogotá, which ranks frequently in first place for economic and well-being indicators. On the contrary, the set of departments that we group as Nuevos Departamentos in the eastern and southern parts of the country experienced modest improvements, taking into account their low starting position.

The univariate kernel density estimates of life expectancy at birth for 1975 and 2000 in Figure 5.10 show that the distribution has become narrower, as was expected from the sigma-convergence result. Even if both distributions seem bimodal, the mode on the left of the distributions is much closer to the main mode in 2000 than in 1975. Nevertheless, according to the bootstrap test of equality of the estimated densities for both years in Table 5.4, we cannot reject the hypothesis that the two densities are identical.

In Figures 5.11 and 5.12, we observe the bivariate kernel density estimator, which is computed using life expectancy at birth relative to the national average for both years. Once

---

<sup>17</sup>According to the Demographic and Health Survey from 2005, 20 percent of women in La Guajira did not have any kind of prenatal care before delivery. These rates are also very high for Caquetá (20 percent), Cauca (15 percent), Chocó (15 percent), and Córdoba(14 percent). In Chocó, 40 percent of births were attended at home (usually by a midwife). The corresponding figures for Caquetá and Cauca are 32 percent and 31 percent, respectively (Profamilia, 2005).

again the results suggest persistence, rather than convergence, as most of mass of the estimated bivariate density is concentrated along the 45-degree diagonal. In relative terms, departments basically remained where they were in 1975. Chocó, Nariño, and Nuevos are the three departments outside of the 90 percent contour. As mentioned before, Chocó and Nariño improved dramatically in their relative positions, while Chocó remained the worst performer both years.

#### 5.5.4 Nourishment

As was mentioned before, we must treat the results concerning nourishment with care, as we only have data for 1995 and 2005, a period which is shorter than that used for other variables, and as we are using data for 23 departments<sup>18</sup> from Demographic and Health Surveys which are not always representative at the department level. Nevertheless, we consider that some insight can be gained investigating convergence in the rate of well-nourished children.

Table 5.1 shows that between 1995 and 2005, the departmental average of the well-nourished rate (WR) improved from 93 percent to 95 percent. Both sigma-convergence and beta-convergence are observed. Figure 5.13 plots the value of WR in 1995 against the change in WR between 1995 and 2005. There is a clear negative relationship between them which is confirmed with the regressions presented in Tables 5.9 and 5.10.

Figure 5.14 shows the univariate kernel density estimates of the log of relative WR and confirms that the distribution is less skewed in 2005, but the bimodality observed in 1995 remains in 2005. The bootstrap test of equality of the estimated densities for both years suggests that there are no systematic differences between them (Table 5.4). Bivariate kernel density estimators shown in Figures 5.15 and 5.16 suggest mild convergence in this indicator. Most of the mass of the estimated bivariate density is concentrated in an axis that is flatter than the 45-degree line.

## 5.6 Conclusions

Several points are important for the discussion. First, unlike in the case when one is dealing with income indicators, social indicators have natural upper bounds (Neumayer, 2003; Kenny, 2004). In the case of the three rates we use (infant survival rate, literacy rate, and well-nourished rate), the upper bound is evident and constant (e.g., no department can have a more than 100 percent literacy rate). In the case of life expectancy at birth, the upper bound

<sup>18</sup>No information is available for Caquetá and Nuevos Departamentos.

can be thought of as variable, but a slow moving one.<sup>19</sup> As discussed by Neumayer (2003) and Kenny (2004), among others, convergence is more likely to be observed in variables with an upper bound when some departments are close to that bound in the initial year. Our results show that this is the case for the classical approach to convergence, where sigma-convergence is found in all cases, and beta-convergence also in all, with the exception of the infant survival rate. Working with relative values and using the distributional approach, it is possible to make a more precise evaluation if one uses bivariate kernel estimators. For example, the distributional approach yields persistence in the distribution of life expectancy at birth, while the beta-convergence analysis shows convergence. Obviously, regression results can be driven by outliers, as is the case for this variable. If one excludes outliers, no evidence of beta-convergence is observed.

Second, it has been argued that at least at the cross-country level, one can observe convergence in social indicators even in the absence of convergence in income. Two reasons advanced for this are high returns to small marginal increases in income at low income levels and causal relationships between the social indicators, themselves (Kenny, 2004), for example between literacy and mortality. Additionally, the dispersion of best practices in health care can lead to improvement in health outcomes, even without income convergence. This is relevant for us, as results in Essay 4 suggest that there has been persistence in the distribution of departmental per capita GDP between 1975 and 2000.<sup>20</sup>

Third, the role of urbanization should also be highlighted, as it is relevant in our case. The percentage of total population in Colombia that lived in urban areas increased from 59 percent in 1973 to 75 percent in 2005. Urbanization facilitates convergence in social indicators, as it is easier to provide social services to urban residents than to rural (Kenny, 2004). As shown by Bettencourt et al. (2007), cities make economies of scale in infrastructure possible. Nevertheless, we stress that living in an urban area does not guarantee access to services.

Fourth, departments in Colombia differ according to climatic and geographic conditions, which conditions have been historically determinant for agglomeration and the availability of infrastructure, particularly roads. Two important consequences are that some diseases only affect departments that are located in the tropics and that access to sanitation services is more limited in isolated areas.

---

<sup>19</sup>There is no consensus among scientists concerning the upper bound of life expectancy at birth and values between 85 years (which is the value currently used in the Human Development Reports) and 100 years have been mentioned. Olshansky et al. (1990), for example, claim that it seems unlikely that life expectancy at birth will exceed the age of 85, while other studies, based on extrapolations from historical trends, predict that it could attain 100 years in developed countries by 2060 (Oeppen and Vaupel, 2002) or by 2300 (United Nations, 2004).

<sup>20</sup>PDB is the abbreviation used in Essay 4 for departmental GDP, based on the Spanish denomination "Producto Departamental Bruto" (CEGA, 2006b).

Fifth, and related to the last point, the internal conflict in Colombia is not of the same intensity across all departments.<sup>21</sup> Although it is a widespread problem affecting the whole country, sparsely populated departments, with predominantly rural areas having limited access due to geographic conditions and in which state presence has been historically low or weak, are more prone to the presence of illegal groups.

With these issues in mind, we begin the discussion of the results concerning convergence in education. In the final quarter of the last century, we find clear evidence of regional convergence in literacy rates in Colombia. As discussed by Barrera and Domínguez (2006), since the seventies, there have been policies in place having the objective of reducing illiteracy and increasing the coverage of primary education, which policies are partially responsible for our result. Urbanization probably also played a role here. While some of the departments with the best indicators in 1973 had a population that was primarily urban, many of the worst performers in that year had a population that was mainly rural and urbanization increased relatively more in the worst performers.<sup>22</sup> Another possible reason for the convergence result is that our indicator considers the population above five years of age in both periods. The elder population in 1973, which was no longer alive in 2005, likely had relatively high illiteracy rates, particularly in the departments which were lagging behind in 1973. Thus, there is a generational effect; the older cohorts included in 1973 are no longer visible in the statistics in 2005, while younger cohorts, who benefited from improved educational resources and literacy campaigns, are included.

Even if regional convergence in literacy rates seems to be a robust result, one should keep in mind that it only reflects very basic education and we do not know whether convergence was observed in education at higher levels. The department of La Guajira deserves special attention as it still lags far behind in literacy rates.<sup>23</sup>

---

<sup>21</sup>The current internal conflict began almost 50 years ago with the emergence of leftist guerrilla groups, the root motivations of which were mainly ideological. Up until 1980, the military capacity of these groups was limited and was concentrated in marginal areas of the country. Parallel to those, paramilitary groups developed slowly in the eighties to defend isolated areas from guerrilla attacks. During the coca bonanza ("bonanza coquera") and the consolidation of drug trafficking in the eighties, illegal armed groups found new ways of financing operations and expanding through the control of areas where illegal crops were grown, as well as in territories that are rich in natural resources, particularly oil (Díaz and Sánchez, 2004).

<sup>22</sup>As examples for well performing departments in terms of literacy rates in 1973, consider Antioquia, which had a share of urban population of 62 percent in 1973 and 78 percent in 2005, and Valle, which had 76 percent in 1973 and 87 percent in 2005. Examples for poorly performing departments in 1973 are Chocó, which had a share of urban population of 26 percent in 1973 and 54 percent in 2005, and Sucre, which had 47 percent in 1973 and 64 percent in 2005.

<sup>23</sup>As mentioned before, this department has a large indigenous population (44 percent) of which population 80 percent has not attained any educational degree (Meisel, 2007a).



Considering indicators of the health status of the population and using infant survival rates (ISR) as a proxy, we find no evidence of beta-convergence. Following the distributional approach and values expressed relative to the national levels, we find evidence of persistence in the distribution. In 2005 departments are basically where they were in 1973, in relative terms. These results can also be explained by a low prevalence of prenatal care in departments of the Pacific and Atlantic coasts (Profamilia, 2005).

Using life expectancy at birth as a proxy for health yields similar results (persistence) according to the distributional approach. In relative terms, in the year 2005, departments are basically where they started in 1973. However, the two departments that had the lowest values at the beginning of the period improved substantially in relative terms. Beta-convergence is driven by these two departments.

Results regarding nourishment show regional convergence in the 10 years studied (1995 through 2005). (As previously mentioned, estimates have to be considered cautiously because the sample of 1995 is not representative for all departments.) Two interesting issues emerge. First, the well-nourished rate deteriorated slightly for some departments that were close to the upper bound in the initial year. Second, the three poorest departments in the Pacific coast improved considerably in relative terms, as pointed by Profamilia (2005). Interestingly, in the Atlantic coast, nourishment relative to the national average stagnated.

Our results differ from those in the scarce literature on convergence among departments in Colombia which used health and education indicators. Some of our results contradict those obtained by Aguirre (2005), who finds convergence in life expectancy at birth, but not in education. The difference concerning life expectancy can be explained as follows. We find that beta-convergence is driven by two influential observations, and once we use bivariate kernel estimators with the variable expressed as a ratio to the the national value, we find persistence in the distribution of life expectancy rather than convergence. The difference concerning education could be due to the fact that Aguirre uses *illiteracy* rate for the analysis, while we use literacy rates. Note that while Aguirre also computes univariate kernel density estimators for the variables in both periods, they are based on absolute values (i.e., not relative to the national average), making judgements as to distributional changes more difficult given that the means vary.

Finally, the study of Meisel and Vega (2007) considers a much longer period (1870 through 2003) for investigating regional convergence in Colombia, using the evolution of adult height over time as an alternative perspective on the standard of living. They find both sigma and beta-convergence among departments and highlight that nutritional improvements are among the main explanations of this result. Even if our results concerning nourishment cover a much shorter period (1995 through 2005), they point to the same conclusion. It could



be interesting to test for convergence among departments with the distributional approach using the sample of Meisel and Vega (2007).

Lack of convergence among Colombian departments in the two variables proxying for health raises some doubts as to the effectiveness of current policies, as convergence is what one would expect, for the reasons explained above. Understanding why some departments still lag behind is relevant, as it is not clear whether the reasons are due mainly to the differences in per capita income, climatic and geographic conditions, infrastructure, or behavior, or to still other factors. It is crucial to understand the main causes of infant and adult morbidity and mortality in the departments lagging behind to assess which specific policies could improve living conditions in each case.

### Summary of Results

	Social indicator used			
	Literacy rate	Infant survival rate	Life expectancy at birth	Well-nourished rate
<b>Classical Approach:</b>				
<b>Convergence?</b>				
Sigma	Yes	Yes	Yes	Yes
Absolute Beta (all obs.)	Yes	No	Yes	Yes
Absolute Beta (excl. outliers)	Yes	Yes	No	Yes
<b>Distributional Approach</b>				
Univariate Kernel Estimators	Dispersion decreases	Distribution unchanged	Distribution unchanged	Distribution unchanged
Bivariate Kernel Estimators	Convergence	Persistence in the distribution	Persistence in the distribution	Suggests slow convergence

*Note:* Results for the distributional approach based on relative values, i.e. ratios to the national level.

## 5.7 Tables

Table 5.1: Descriptive Statistics of the Variables Used

Variable	Literacy rate		Infant survival rate		Life expectancy at birth		Well-nourished rate	
	1973	2005	1975	2000	1975	2000	1995	2005
mean	70.3	85.8	935.9	962.3	61.7	69.9	93.0	95.3
median	69.0	87.4	940.1	967.8	62.5	69.9	93.8	95.7
stand. deviation	9.38	6.8	16.6	14.4	3.6	2.6	4.2	2.3
skewness	-0.0	-1.8	-2.0	-2.0	-1.6	-1.0	-1.4	-0.9
kurtosis	2.5	6.5	7.8	7.6	6.6	4.4	4.8	3.9
range	37.3	30.2	81.3	64.8	17.6	11.5	17.8	9.8
minimum	52.2	63.1	876.7	911.5	49.7	62.4	80.9	88.9
maximum	89.6	93.4	958.1	976.3	67.3	73.8	98.7	98.7
Number of obs.	25	25	24	24	24	24	23	23

Source: Own calculations based on data at the department level from DANE and DHS.

Literacy rate measured as the percentage of literate population above age 5.

Infant survival rate measured as per thousand live births.

Life expectancy at birth measured in years.

Well-nourished rate measured as the percentage of population under age 5 not underweight.

Table 5.2: Beta Convergence (OLS) using all available Observations. Dependent Variable: Change in Literacy Rate (LIT) between 1973 and 2005

Variable	Robust		p-value
	Coefficient	Std. error	
Intercept	43.42	9.56	0.000
LIT in 1973	-0.40	0.13	0.005
Number of observations	25		
R-squared	0.50		

Source: Own calculations based on data from DANE.

Note: HC3 robust standard errors calculated following Davidson and MacKinnon (1993).

Table 5.3: Beta convergence (OLS) excluding Outliers. Dependent Variable: Change in Literacy Rate (LIT) between 1973 and 2005

Variable	Robust		p-value
	Coefficient	Std. error	
Intercept	49.36	5.93	0.000
LIT in 1973	-0.47	0.08	0.000
Number of observations	23		
R-squared	0.81		

Source: Own calculations based on data from DANE.

Notes: Two departments (Bogotá and La Guajira) were excluded using Cook's distance to detect for unusual influence or leverage after the regression with all observations.

HC3 robust standard errors calculated following Davidson and MacKinnon (1993).

Table 5.4: Bootstrap tests of equality of univariate density estimates, Social indicators

Variable	Years compared	p-value
LIT	1973, 2005	0.004
ISR	1975, 2000	0.968
LEX	1975, 2000	0.733
WR	1995, 2005	0.118

Test proposed by Bowman and Azzalini (1997). 1000 replications.

Source: Own calculations based on data from DANE and DHS.

Table 5.5: Beta Convergence (OLS) using all available Observations. Dependent Variable: Change in Infant Survival Rate (ISR) between 1975 and 2000

Variable	Coefficient	Robust	
		Std. error	p-value
Intercept	370.75	297.25	0.257
ISR in 1975	-0.37	0.32	0.225
Number of observations	24		
R-squared	0.28		

Source: Own calculations based on data from DANE.

Note: HC3 robust standard errors calculated following Davidson and MacKinnon (1993).

Table 5.6: Beta Convergence (OLS) excluding Outliers. Dependent Variable: Change in Infant Survival Rate (ISR) between 1975 and 2000

Variable	Coefficient	Robust	
		Std. error	p-value
Intercept	660.00	195.01	0.003
ISR in 1975	-0.68	0.21	0.004
Number of observations	23		
R-squared	0.41		

*Source:* Own calculations based on data from DANE.

*Notes:* One department (Chocó) was excluded using Cook's distance to detect for unusual influence or leverage after the regression with all observations.

HC3 robust standard errors calculated following Davidson and MacKinnon (1993).

Table 5.7: Beta Convergence (OLS) using all available Observations. Dependent Variable: Change in Life Expectancy at Birth (LEX) between 1975 and 2000

Variable	Coefficient	Robust	
		Std. error	p-value
Intercept	34.69	6.67	0.000
LEX in 1975	-0.43	0.11	0.001
Number of observations	24		
R-squared	0.48		

*Source:* Own calculations based on data from DANE.

*Note:* HC3 robust standard errors calculated following Davidson and MacKinnon (1993).

Table 5.8: Beta Convergence (OLS) excluding Outliers. Dependent Variable: Change in Life Expectancy at Birth (LEX) between 1975 and 2000

Variable	Coefficient	Robust	
		Std. error	p-value
Intercept	27.48	10.46	0.016
LEX in 1975	-0.32	0.17	0.070
Number of observations	22		
R-squared	0.19		

*Source:* Own calculations based on data from DANE.

*Notes:* Two departments (Chocó and Nariño) were excluded using Cook's distance after the regression with all observations to detect for unusual influence or leverage.

HC3 robust standard errors calculated following Davidson and MacKinnon (1993).

Table 5.9: Beta Convergence (OLS) using all available Observations. Dependent Variable: Change in Well-nourished Rate (WR) between 1995 and 2005

Variable	Coefficient	Robust	
		Std. error	p-value
Intercept	0.86	0.23	0.001
WR in 1995	-0.90	0.24	0.001
Number of observations	23		
R-squared	0.74		

*Source:* Own calculations based on data from DHS.

*Note:* HC3 robust standard errors calculated following Davidson and MacKinnon (1993).

Table 5.10: Beta Convergence (OLS) excluding Outliers. Dependent Variable: Change in Well-nourished Rate (WR) between 1995 and 2005

Variable	Coefficient	Robust	
		Std. error	p-value
Intercept	0.99	0.16	0.000
WR in 1975	-1.04	0.17	0.000
Number of observations	21		
R-squared	0.68		

*Source:* Own calculations based on data from DHS.

*Notes:* Two departments (Chocó and La Guajira) were excluded using Cook's distance after the regression with all observations to detect for unusual influence or leverage.

HC3 robust standard errors calculated following Davidson and MacKinnon (1993).

Table 5.11: Poverty Headcount Index (% of Households below the Poverty Line) by Department, 1996 to 2005

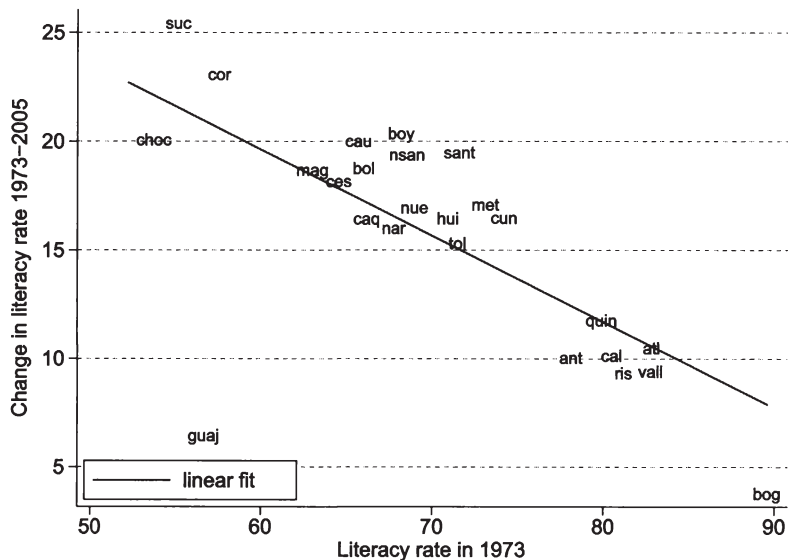
<b>Department</b>	<b>1996</b>	<b>1999</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>
Antioquia	54.1	57.8	58.9	55.6	54.1
Atlántico	49.5	57.9	53.2	52.1	48.2
Bogotá	37.8	46.3	36.1	34.2	29.5
Bolívar	60.3	59.9	67.8	51.5	54.6
Boyacá	65.0	62.6	72.3	70.3	71.5
Caldas	54.3	54.0	59.6	58.8	57.7
Caquetá	58.0	59.9	53.5	54.5	56.8
Cauca	61.4	73.3	64.5	69.0	63.0
Cesar	49.7	53.7	67.2	61.6	59.3
Chocó	70.9	78.0	62.6	70.3	71.6
Córdoba	74.3	72.6	68.5	66.5	70.8
Cundinamarca	44.5	50.9	58.4	51.9	53.6
Huila	62.9	62.7	74.4	69.7	66.3
Guajira	49.3	50.2	68.4	54.6	52.8
Magdalena	62.6	62.7	66.4	55.4	55.0
Meta	50.9	52.4	47.9	44.3	42.5
Nariño	68.1	71.7	70.7	71.2	67.3
N.Santander	61.3	58.2	57.3	57.3	57.9
Quindío	45.4	51.4	49.3	41.3	47.3
Risaralda	52.1	51.5	47.9	45.3	44.7
Santander	48.9	55.4	50.2	48.6	48.6
Sucre	47.9	64.0	69.4	56.5	65.7
Tolima	59.3	58.4	60.6	58.8	60.1
Valle	47.4	47.6	44.1	37.4	38.9

Source: Own calculations based on Household Surveys from DANE.



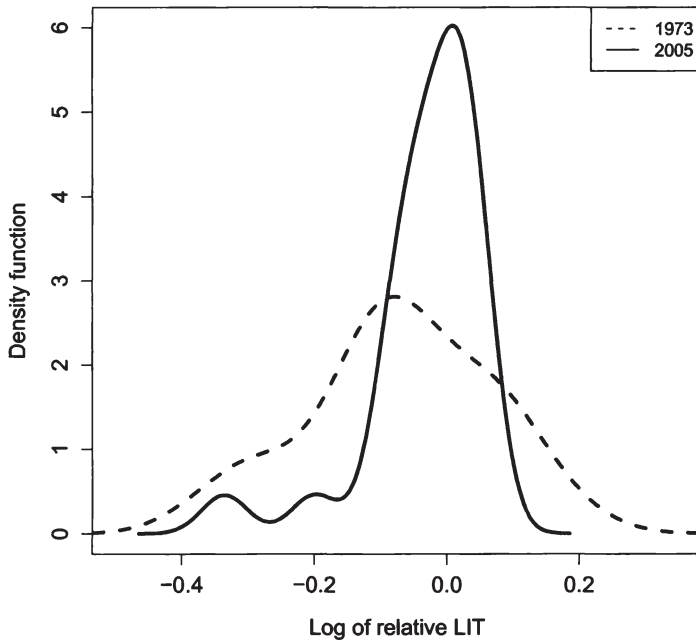
## 5.8 Figures

Figure 5.1: Evolution of Literacy Rate. 1973-2005



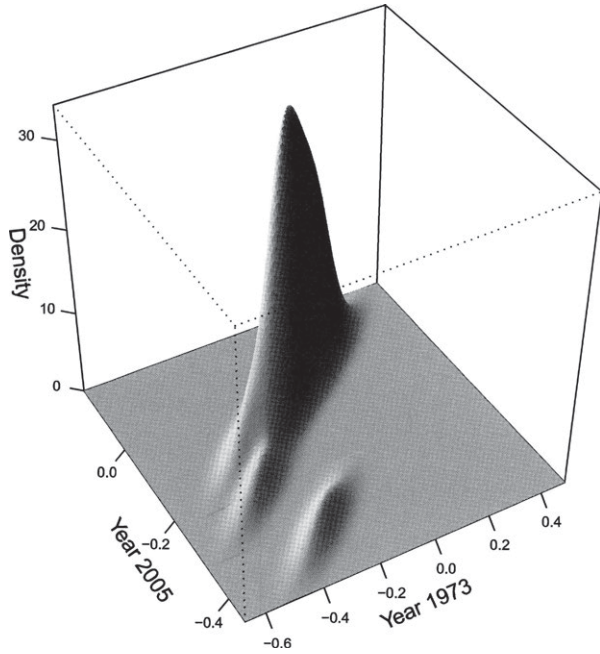
Source: Own calculations based on data from DANE.

Figure 5.2: Univariate Kernel Density Estimators of Relative Literacy Rate. 1973 and 2005.



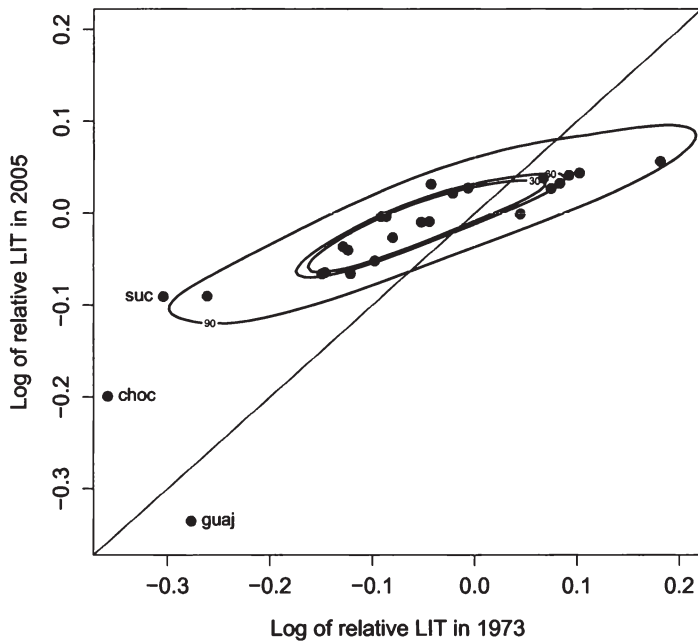
Source: Own calculations based on data from DANE. Variables relative to the national average and in logs.

Figure 5.3: Bivariate Kernel Density Estimators of Relative Literacy Rate. 3D Representation. 1973 and 2005.



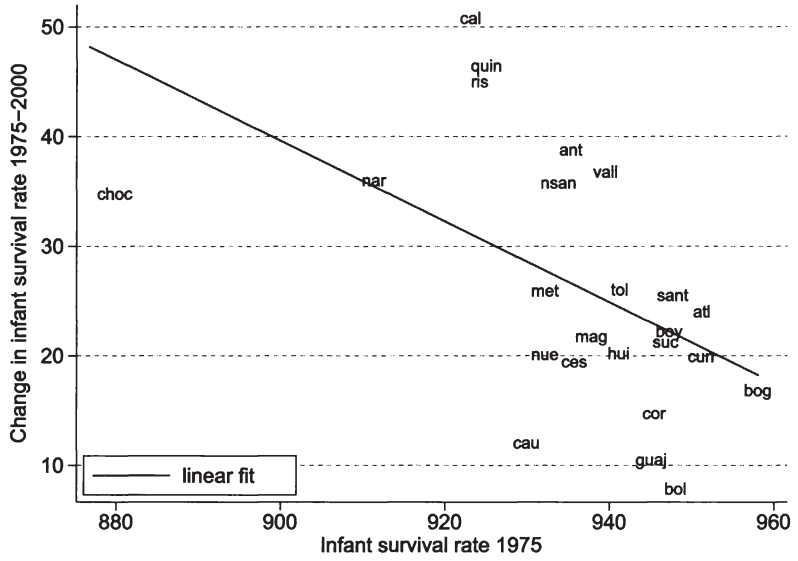
Source: Own calculations based on data from DANE. Variables relative to the national average and in logs.

Figure 5.4: Bivariate Kernel Density Estimators of Relative Literacy Rate. Contour Plot. 1973 and 2005.



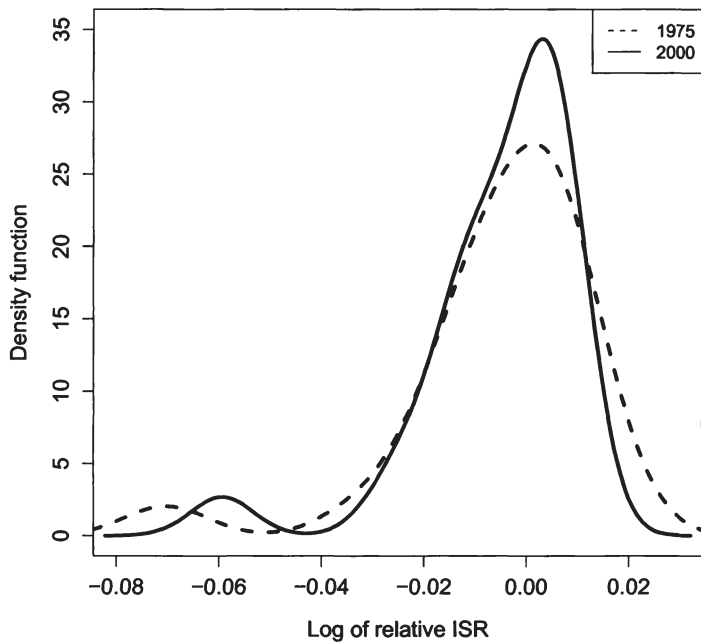
Source: Own calculations based on data from DANE. Variables in logs. Contours are drawn at 30%, 60% and 90%. The points represent the 25 observations. Points outside the contour level curves are identified. A 45 degree line is added to the plot.

Figure 5.5: Evolution of Infant Survival Rate. 1975-2000



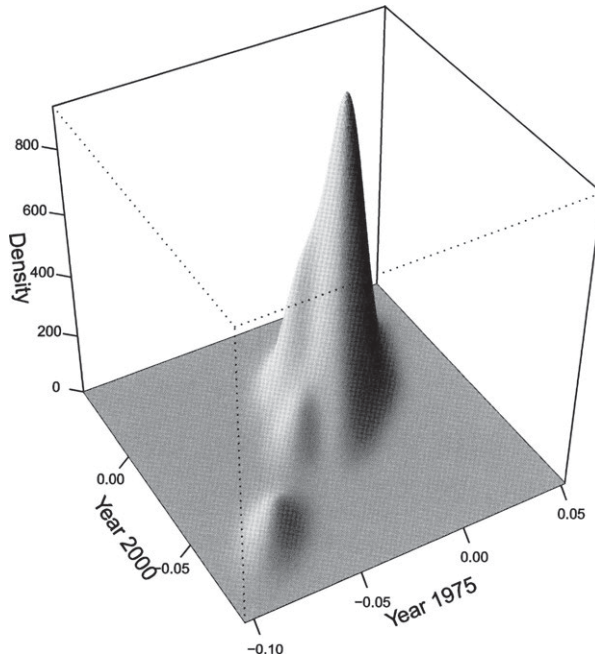
Source: Own calculations based on data from DANE.

Figure 5.6: Univariate Kernel Density Estimators of Relative Infant Survival Rate. 1975 and 2000.



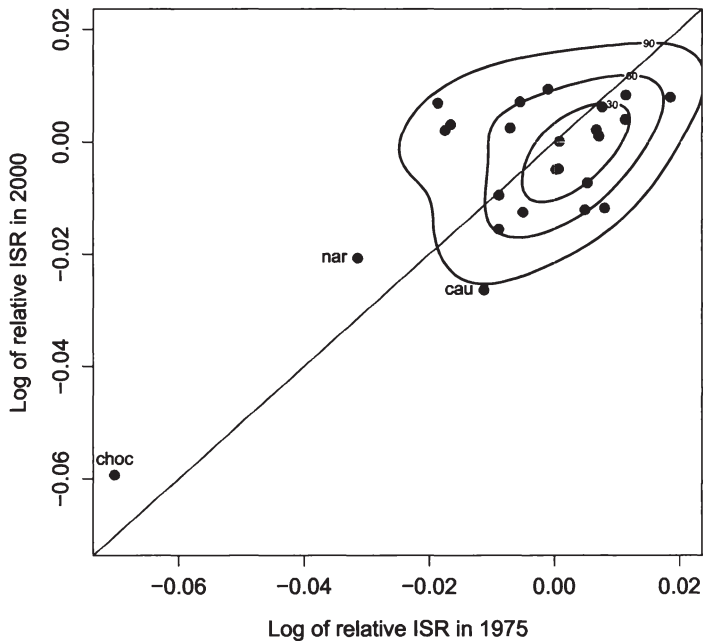
Source: Own calculations based on data from DANE. Variables relative to the national average and in logs.

Figure 5.7: Bivariate Kernel Density Estimators of Relative Infant Survival Rate. 3D Representation. 1975 and 2000.



*Source:* Own calculations based on data from DANE. Variables relative to the national average and in logs.

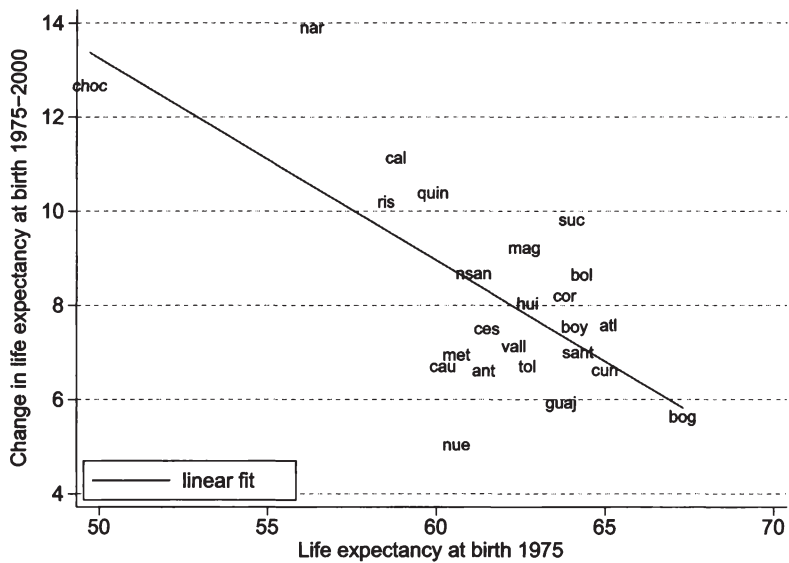
Figure 5.8: Bivariate Kernel Density Estimators of Relative Infant Survival Rate. Contour Plot. 1975 and 2000.



Source: Own calculations based on data from DANE. Variables in logs. Contours are drawn at 30%, 60% and 90%. The points represent the 25 observations. Points outside the contour level curves are identified. A 45 degree line is added to the plot.

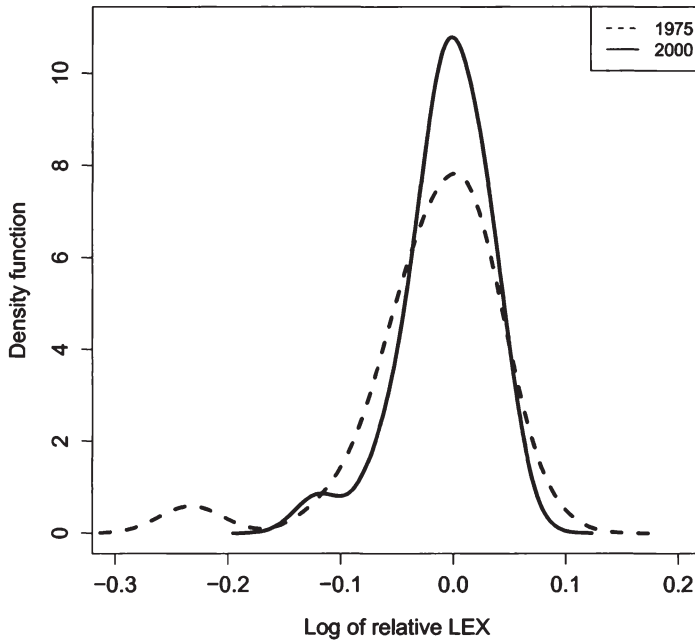


Figure 5.9: Evolution of Life Expectancy at Birth. 1975-2000



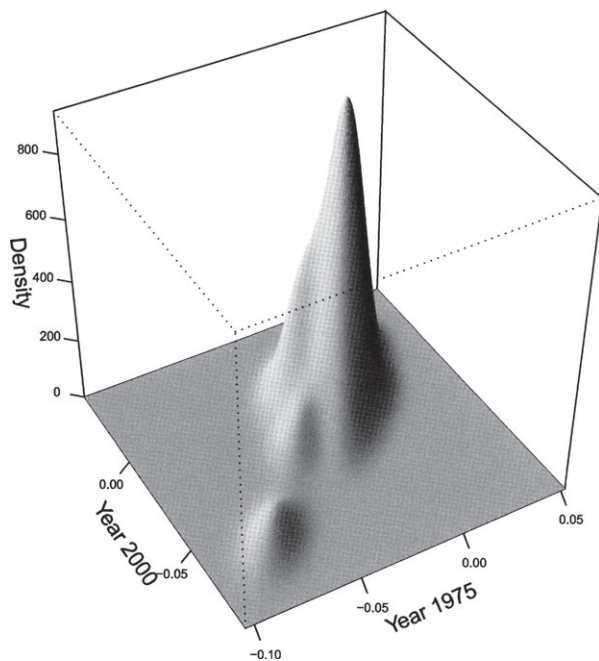
Source: Own calculations based on data from DANE.

Figure 5.10: Univariate Kernel Density Estimators of Relative Life Expectancy at Birth, 1975 and 2000.



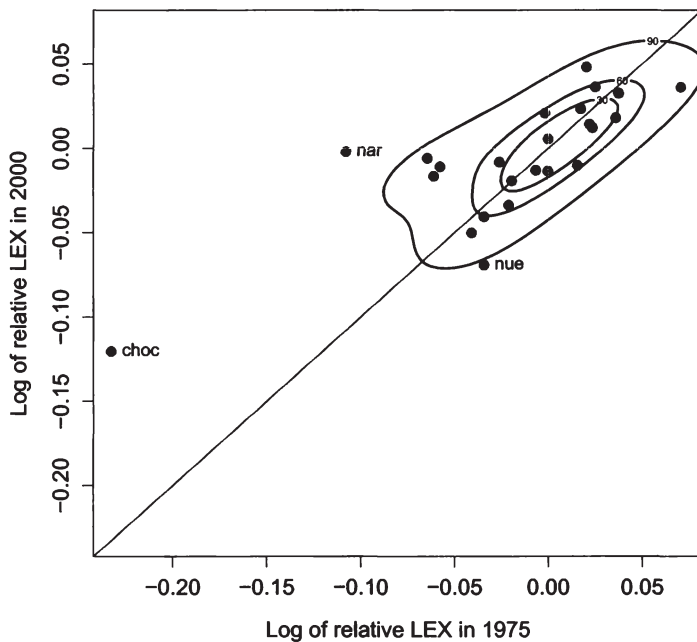
Source: Own calculations based on data from DANE. Variables relative to the national average and in logs.

Figure 5.11: Bivariate Kernel Density Estimators of Relative Life Expectancy at Birth. 3D Representation. 1975 and 2000.



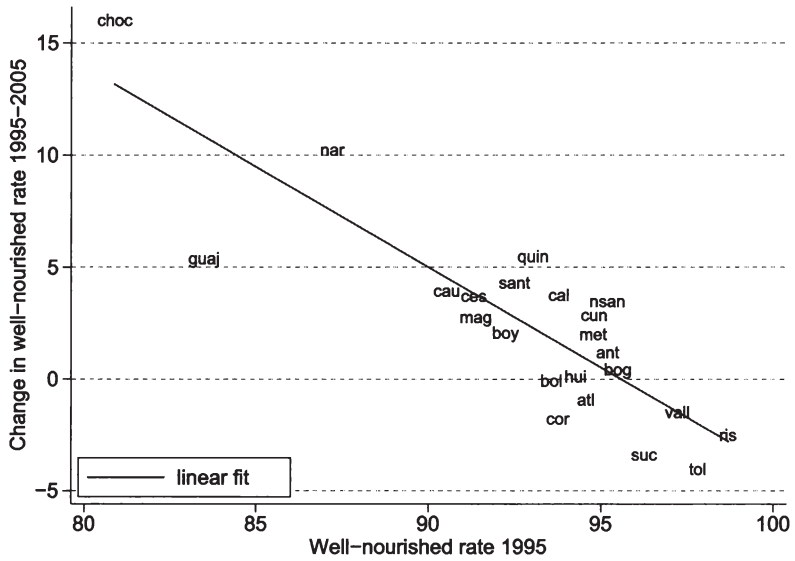
*Source:* Own calculations based on data from DANE. Variables relative to the national average and in logs.

Figure 5.12: Bivariate Kernel Density Estimators of Relative Life Expectancy at Birth. Contour Plot, 1975 and 2000.



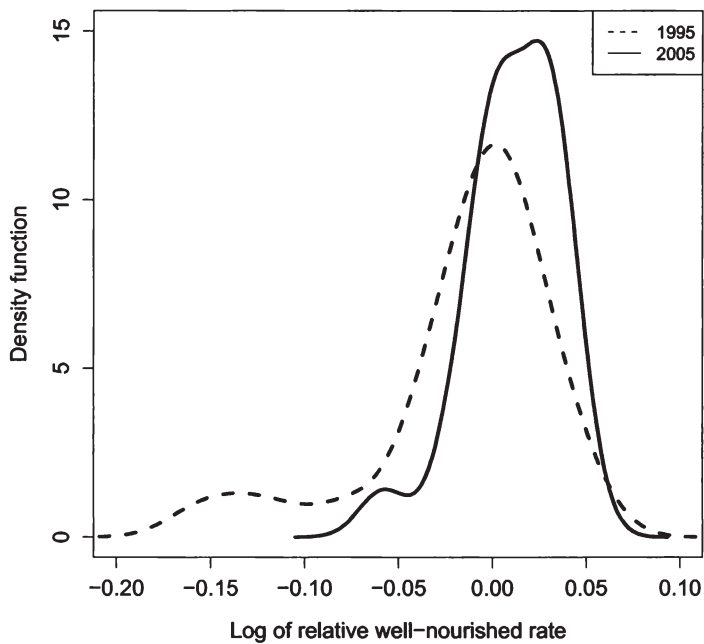
*Source:* Own calculations based on data from DANE. Variables in logs. Contours are drawn at 30%, 60% and 90%. The points represent the 25 observations. Points outside the contour level curves are identified. A 45 degree line is added to the plot.

Figure 5.13: Evolution of Well-Nourished Rate. 1995-2005



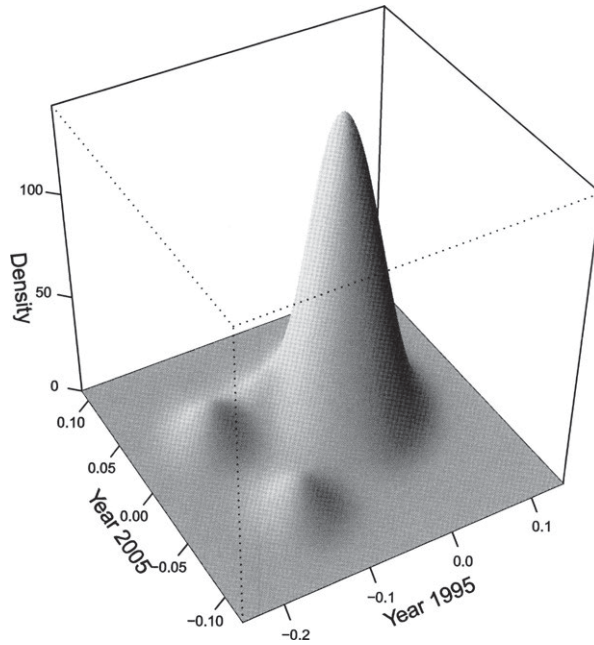
Source: Own calculations based on data from DHS

Figure 5.14: Univariate Kernel Density Estimators of Relative Well-nourished Rate. 1995 and 2005.



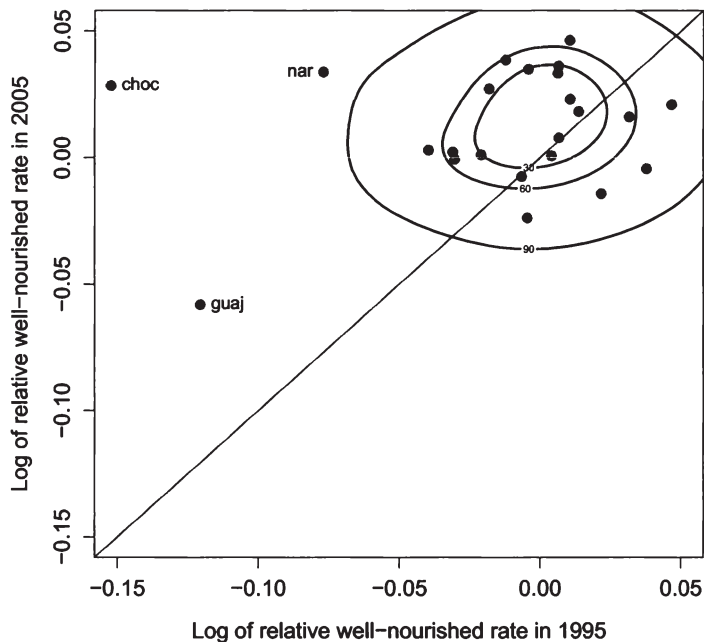
Source: Own calculations based on data from DHS. Variables relative to the national average and in logs.

Figure 5.15: Bivariate Kernel Density Estimators of Relative Well-nourished Rate. 3D Representation. 1995 and 2005.



*Source:* Own calculations based on data from DHS. Variables relative to the national average and in logs.

Figure 5.16: Bivariate Kernel Density Estimators of Relative Well-nourished Rate. Contour Plot. 1973 and 2005.



*Source:* Own calculations based on data from DHS. Contours are drawn at 30%, 60% and 90%. The points represent the 25 observations. Points outside the contour level curves are identified. A 45 degree line is added to the plot.



# **Appendices**



## Appendix to Essay 1

### Objectives, Properties and Proofs

In this appendix, we present the objectives and properties that we consider relevant for any composite index related to social institutions related to gender inequality. Moreover, we show that the proposed index fulfills all of them.

We use the following notation. Let  $X^j$ , with  $j = A, B$ , be the vector containing the values of the subindices  $x_i^j$ , with  $i = 1, \dots, n$ , for the country  $j$ <sup>24</sup>.  $I(X)$  represents the composite index.

### Objectives of the Index

The objectives of the index are the following:

1. The index  $I(X)$  should represent the level of gender inequality, so that countries can be ranked.
2. The interpretation of  $I(X)$  should be straightforward. As in the case of the subindices  $x_i$ , the value 0 should correspond to no inequality and the value 1 to complete inequality.
3. For any subindex  $x_i$ , we interpret the value 0, i.e. no inequality, as the goal to be achieved. The value zero can be thought of as a poverty line (see Ravallion, 1994; Deaton, 1997; Subramanian, 2007, and references therein). We define a deprivation function  $\phi(x_i, 0)$ , with  $\phi(x_i, 0) > 0$  if  $x_i > 0$ , and  $\phi(x_i, 0) = 0$  if  $x_i = 0$ . Higher values of  $x_i$  should lead to a penalization in  $I(X)$  that should increase with the distance  $x_i$  to zero, i.e.  $\frac{\partial I(X)}{\partial x_i} > 0$ , and  $\frac{\partial^2 I(X)}{\partial x_i^2} > 0$ .
4.  $I(X)$  should not allow for total compensation among subindices, but permit partial compensation. This somehow relates to the transfer axioms that should be fulfilled by inequality as well as poverty measures (see Atkinson, 1970; Kakwani, 1984; Shorrocks and Foster, 1987; Subramanian, 2007; Alkire and Foster, 2008, and references therein). Assuming that two subindices have the same value, a decrease in  $x_i$ , i.e. less inequality, is rewarded more in  $I(X)$  than an equivalent increase in the other subindex  $x_k$ .
5.  $I(X)$  should be easy to compute and transparent.

<sup>24</sup>In what follows, the superscript  $j$  will only be used if it is necessary to distinguish countries.

## Properties of the Index

Some of the properties that any index should fulfill are:

1. **Support and range of  $I(X)$ :**
  - $I(X)$  must be defined for  $0 \leq x_i \leq 1$ ,  $i = 1, \dots, n$ .
  - $0 \leq I(X) \leq 1$  must hold for any  $X$ .
  - If  $x_i = 0 \forall i$ , then  $I(X) = 0$ . If  $x_i = 1 \forall i$ , then  $I(X) = 1$ .
2. **Anonymity (symmetry):** The value of  $I(X^j)$  does not depend either on the names of the subindices nor on the name of the country ( $j$ ).
3. **Unanimity (Pareto Optimality):** If  $x_i^A \leq x_i^B \forall i$ , then  $I(X^A) \leq I(X^B)$ .
4. **Monotonicity:** If considering  $X^A$  and  $X^B$  country  $A$  is preferred to country  $B$ , and only  $x_i^A$  improves (i.e. decreases) for a given  $i$ , while  $x_i^B \forall i$  remains unchanged, then country  $A$  should still be preferred over country  $B$ .
5. **Penalization of inequality in the case of equal means:** Let the mean of  $X^A$  be equal to the mean of  $X^B$ . If the dispersion of  $X^A$  is smaller than the dispersion of  $X^B$ , then  $I(X^A) < I(X^B)$ .
6. **Compensation property:** In an example with two subindices, let  $x_1 = x_2 > 0$ ,  $\Delta x_1 \leq 1 - x_1$ , and  $\Delta x_2 \leq 1 - x_2$ . Assume that  $x_1$  increases by  $|\Delta x_1|$  and  $x_2$  decreases by  $|\Delta x_2|$ 
  - (a) If  $|\Delta x_1| = |\Delta x_2|$ , then  $I(X)$  must increase.
  - (b) For  $I(X)$  to remain unchanged, we must have  $|\Delta x_2| > |\Delta x_1|$ .

## Proofs

The composite index  $I(X)$  is defined as

$$I(X) = \frac{1}{n} \sum_{i=1}^n (x_i - 0)^2.$$

The index proposed fulfills all the stated properties.

1. **Support and range of  $I(X)$** 
  - $I(X)$  is defined for  $0 \leq x_i \leq 1$ ,  $i = 1, \dots, n$ .
  - For any  $X$ , we have that  $0 \leq I(X) \leq 1$ .

- If  $x_i = 0 \forall i$ , then  $I(X) = 0$ . If  $x_i = 1 \forall i$ , then  $I(X) = 1$ .

## 2. Anonymity (symmetry)

The value of  $I(X^j)$  does not depend either on the names of the subindices nor on the name of the country ( $j$ ).

## 3. Unanimity (Pareto Optimality)

If we assume that  $\forall i$

$$x_i^A \leq x_i^B,$$

then we can show that

$$\begin{aligned} (x_i^A)^2 &\leq (x_i^B)^2 \\ \frac{1}{n} \sum_{i=1}^n (x_i^A - 0)^2 &\leq \frac{1}{n} \sum_{i=1}^n (x_i^B - 0)^2 \\ I(X^A) &\leq I(X^B). \end{aligned}$$

## 4. Monotonicity

We assume that

$$\begin{aligned} I(X^A) &\leq I(X^B) \\ \frac{1}{n} \sum_{i=1}^n (x_i^A - 0)^2 &\leq \frac{1}{n} \sum_{i=1}^n (x_i^B - 0)^2. \end{aligned}$$

Let us suppose, without loss of generality, that subindex  $x_1$  improves (decreases) by  $\delta > 0$  for country  $A$ . Then we have that

$$\frac{1}{n} (x_1^A - \delta - 0)^2 + \frac{1}{n} \sum_{i=2}^n (x_i^A - 0)^2 \leq \frac{1}{n} \sum_{i=1}^n (x_i^A - 0)^2,$$

and hence

$$\frac{1}{n} (x_1^A - \delta - 0)^2 + \frac{1}{n} \sum_{i=2}^n (x_i^A - 0)^2 \leq \frac{1}{n} \sum_{i=1}^n (x_i^B - 0)^2.$$

This means that

$$I(X^{A^*}) \leq I(X^B)$$

with  $X^{A^*}$  defined as the vector corresponding to country  $A$  with only one subindex having improved (decreased) by  $\delta$ .

### 5. Penalization of inequality in the case of equal means

If we assume equal means, so that

$$\mu = \frac{1}{n} \sum_{i=1}^n (x_i^A) = \frac{1}{n} \sum_{i=1}^n (x_i^B),$$

then we also have

$$\sum_{i=1}^n (x_i^A) = \sum_{i=1}^n (x_i^B).$$

If we assume that the variance of  $X^A$  is smaller than the variance of  $X^B$  so that

$$\frac{1}{n} \sum_{i=1}^n (x_i^A - \mu)^2 < \frac{1}{n} \sum_{i=1}^n (x_i^B - \mu)^2,$$

we can show that

$$\begin{aligned} \sum_{i=1}^n [(x_i^A)^2 - 2\mu x_i^A + \mu^2] &< \sum_{i=1}^n [(x_i^B)^2 - 2\mu x_i^B + \mu^2], \\ \sum_{i=1}^n (x_i^A)^2 - 2\mu \sum_{i=1}^n x_i^A + n\mu^2 &< \sum_{i=1}^n (x_i^B)^2 - 2\mu \sum_{i=1}^n x_i^B + n\mu^2. \end{aligned}$$

As  $\sum_{i=1}^n (x_i^A) = \sum_{i=1}^n (x_i^B)$ , we have that

$$\begin{aligned} \sum_{i=1}^n (x_i^A)^2 &< \sum_{i=1}^n (x_i^B)^2 \\ \frac{1}{n} \sum_{i=1}^n (x_i^A - 0)^2 &< \frac{1}{n} \sum_{i=1}^n (x_i^B - 0)^2 \\ I(X^A) &< I(X^B). \end{aligned}$$

### 6. Compensation property

In an example with two subindices, let  $x_1 = x_2 = x > 0$ ,  $\Delta x_1 \leq 1 - x_1$ , and  $\Delta x_2 \leq 1 - x_2$ .

(a) We can show that if  $\Delta x_1 = \Delta x_2 = \delta > 0$ , then

$$\begin{aligned} x_2 &< x_1 + \delta \\ 0 &< x_1 - x_2 + \delta \\ 0 &< 2\delta(x_1 - x_2 + \delta) \\ x_1^2 + x_2^2 &< x_1^2 + x_2^2 + 2\delta(x_1 - x_2 + \delta) \\ \frac{1}{2}(x_1^2 + x_2^2) &< \frac{1}{2}(x_1^2 + 2\delta x_1 + \delta^2 + x_2^2 - 2\delta x_2 + \delta^2) \\ \frac{1}{2}(x_1^2 + x_2^2) &< \frac{1}{2}[(x_1^2 + \delta)^2 + (x_2^2 - \delta)^2] \\ I(x_1, x_2) &< I(x_1 + \delta, x_2 - \delta), \end{aligned}$$

and hence we have shown that if  $x_1$  increases by  $\delta$  and  $x_2$  decreases by  $\delta$ , then  $I(X)$  must increase.

- (b) We will show that if  $x_1$  increases by  $\Delta x_1$  and  $x_2$  decreases by  $\Delta x_1$  and the value of the index remains unchanged, the increase of  $x_1$  must be smaller than the absolute value of the decrease in  $x_2$ .

$$\begin{aligned} I(x_1, x_2) &= I(x_1 + \Delta x_1, x_2 - \Delta x_2) \\ \frac{1}{2}(x_1^2 + x_2^2) &= \frac{1}{2}[(x_1 + \Delta x_1)^2 + (x_2 - \Delta x_2)^2] \\ x_1^2 + x_2^2 &= x_1^2 + 2x_1\Delta x_1 + (\Delta x_1)^2 + x_2^2 - 2x_2\Delta x_2 + (\Delta x_2)^2 \\ 0 &= 2x_1\Delta x_1 + (\Delta x_1)^2 - 2x_2\Delta x_2 + (\Delta x_2)^2 \end{aligned}$$

Using the fact that  $x_1 = x_2 = x$ , we can rewrite this as

$$\begin{aligned} 0 &= 2x\Delta x_1 + (\Delta x_1)^2 - 2x\Delta x_2 + (\Delta x_2)^2 \\ 0 &= 2x(\Delta x_1 - \Delta x_2) + (\Delta x_1)^2 + (\Delta x_2)^2. \end{aligned}$$

As  $2x > 0$ ,  $(\Delta x_1)^2 > 0$ , and  $(\Delta x_2)^2 > 0$ , we must have that

$$\begin{aligned} \Delta x_1 - \Delta x_2 &< 0 \\ \Delta x_1 &< \Delta x_2. \end{aligned}$$





# Bibliography

- Abadian, S. (1996). Women's autonomy and its impact on fertility. *World Development* 24(12), 1793–1809.
- Ades, A. and R. D. Tella (1997). The new economics of corruption: A survey and some new results. *Political Studies* 45(3), 496–515.
- Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley and Sons.
- Aguirre, K. (2005). Convergencia en indicadores sociales en Colombia: Una aproximación desde los enfoques tradicional y no paramétrico. *Desarrollo y Sociedad* (56), 147–176.
- Aguirre, K. (2008). Convergencia: Del análisis del nivel de actividad económica a las variables sociales. Una revisión de la literatura del caso colombiano. Documentos de CERAC 8, Centro de Recursos para el Análisis de Conflictos.
- Al-Marhubi, F. A. (2005). Openness and governance: Evidence across countries. *Oxford Development Studies* 33(3), 453–471.
- Alatas, V., L. Cameron, A. Chaudhuri, N. Erkal, and L. Gangadharan (2009). Gender, culture, and corruption: Insights from an experimental analysis. *Southern Economic Journal* 75(3), 663–680.
- Alderman, H., J. R. Behrman, D. R. Ross, and R. Sabot (1996). Decomposing the gender gap in cognitive skills in a poor rural economy. *Journal of Human Resources* 31(1), 229–254.
- Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth* 8(2), 155–194.
- Alhassan-Alolo, N. (2007). Gender and corruption: Testing the new consensus. *Public Administration and Development* 27, 227–237.

- Alkire, S. and J. E. Foster (2008). Counting and multidimensional poverty measurement. OPHI Working Paper 7, Oxford Poverty and Human Development Initiative, Queen Elizabeth House, University of Oxford.
- Andreoni, J. and L. Vesterlund (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics* 116(1), 293–312.
- Ardila Rueda, L. (2004). Gasto público y convergencia regional en Colombia. *Ensayos sobre Política Económica* (45).
- Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory* 2(3), 244–263.
- Bardhan, K. and S. Klasen (1999). UNDP's Gender-related indices: A critical review. *World Development* 27, 985–1010.
- Barón, J. (2003). ¿Qué sucedió con las disparidades económicas regionales en Colombia entre 1980 y el 2000? Documentos de Trabajo sobre Economía Regional 38, Banco de la República.
- Barrera, F. and C. Domínguez (2006). Educación básica en Colombia: Opciones futuras de política. Technical report, Departamento Nacional de Planeación, DNP-MERDP.
- Barro, R. J. and X. Sala-i-Martin (1991). Convergence across states and regions. *Brookings Papers on Economic Activity* 1, 107–182.
- Barro, R. J. and X. Sala-i-Martin (1992a). Convergence. *Journal of Political Economy* 100, 223–251.
- Barro, R. J. and X. Sala-i-Martin (1992b). Regional growth and migration: A Japan-United States comparison. *Journal of Japanese and International Economies* 6, 312–346.
- Barro, R. J. and X. Sala-i-Martin (2004). *Economic Growth* (Second ed.). Cambridge, MA: MIT Press.
- Basu, A. M. (2002). Why does education lead to lower fertility? A critical review of some of the possibilities. *World Development* 30(10), 1779–1790.
- Beck, N. (2001). Time-series-cross-section data: What have we learned in the past few years? *Annual Review of Political Science* 4, 271–293.
- Beck, N. and J. N. Katz (2007). Random coefficient models for time-series-cross-section data: Monte Carlo experiments. *Political Analysis* 15, 185–195.

- Becker, G. S. (1981). *A Treatise on the Family* (Enlarged 1991 ed.). Cambridge: Harvard University Press.
- Bettencourt, L. M. A. M., J. Lobo, D. Helbing, C. Kühnert, and G. B. B. West (2007). Growth, innovation, scaling, and the pace of life in cities. *PNAS* 104(17), 7301–7306.
- Birchenall, J. and G. Murcia (1997). Convergencia regional: una revisión del caso colombiano. *Desarrollo y Sociedad* (40), 273–308.
- Bjornskov, C., A. Dreher, and J. A. Fischer (2009). On gender inequality and life satisfaction: Does discrimination matter? Mimeo.
- Bloom, D. E. and D. Canning (2007). Mortality traps and the dynamics of health transitions. *PNAS* 104(41), 16044–16049.
- Bloom, S. S., D. Wypij, and M. Das Gupta (2001). Dimensions of women's autonomy and the influence on maternal health care utilization in a north Indian city. *Demography* 38(1), 67–78.
- Bonet, J. (1999). El crecimiento regional en Colombia, 1980-1996: Una aproximación con el método shift-share. Documentos de Trabajo sobre Economía Regional 10, Banco de la República.
- Bonet, J. (2007). ¿Porqué es pobre el Chocó? Documentos de Trabajo sobre Economía Regional 90, Banco de la República.
- Bonet, J. and A. Meisel (2006a). El legado colonial como determinante del ingreso per cápita departamental en Colombia. Documentos de Trabajo sobre Economía Regional 75, Banco de la República.
- Bonet, J. and A. Meisel (2006b). Polarización del ingreso per cápita departamental en Colombia, 1975 - 2000. Documentos de Trabajo sobre Economía Regional 76, Banco de la República.
- Bowman, A. W. and A. Azzalini (1997). *Applied Smoothing Techniques for Data Analysis*. Number 18 in Oxford Statistical Science Series. Oxford: Clarendon Press.
- Branisa, B. and A. Cardozo (2009a). Regional growth convergence in Colombia using social indicators. Ibero America Institute for Econ. Research (IAI) Discussion Papers 195, University of Goettingen.

- Branisa, B. and A. Cardozo (2009b). Revisiting the regional growth convergence debate in Colombia using income indicators. Ibero America Institute for Econ. Research (IAI) Discussion Papers 194, University of Goettingen.
- Branisa, B., S. Klasen, and M. Ziegler (2009a). New measures of gender inequality: The social institutions and gender index (SIGI) and its subindices. Courant Research Centre: Poverty, Equity and Growth - Discussion Papers 10, University of Goettingen.
- Branisa, B., S. Klasen, and M. Ziegler (2009b). Why we should all care about social institutions related to gender inequality. Courant Research Centre: Poverty, Equity and Growth - Discussion Papers 15, University of Goettingen.
- Braunstein, E. (2007). The efficiency of gender equity in economic growth: Neoclassical and feminist approaches. GEM-IWG Working Paper 07-4, The Gender and Macro International Working Group.
- Brock, W. and S. N. Durlauf (2001). Growth economics and reality. Mimeo, Department of Economics, University of Wisconsin.
- Cain, M. (1984). *On Women's Status, Family Structure, and Fertility in Developing Countries*. New York: New York Center for Policy Studies, The Population Council.
- Canova, F. and A. Marcet (1995). The poor stay poor: Non-convergence across countries and regions. CEPR Working Paper.
- Cárdenas, M. (2007). Economic growth in Colombia: A reversal of fortune? Departamento Nacional de Planeación - Archivos de Economía No. 2402.
- Cárdenas, M. and A. Pontón (1995). Growth and convergence in Colombia: 1950-1990. *Journal of Development Economics* 47, 5-37.
- Cass, D. (1965). Optimum growth in an aggregation model of capital accumulation. *Review of Economic Studies* 32, 233-240.
- CEGA (2006a). *Ingreso, Consumo y Ahorro en los Departamentos de Colombia: 1975-2000*, Volume 2. Bogotá: Universidad de Los Andes/Centro de Estudios Ganaderos y Agrícolas.
- CEGA (2006b). *Sistema Simplificado de Cuentas Departamentales de Colombia: 1975-2000*, Volume 1. Bogotá: Universidad de Los Andes/Centro de Estudios Ganaderos y Agrícolas.

- Central Intelligence Agency (2009). The world factbook. Electronic publication. <https://www.cia.gov/library/publications/the-world-factbook/>.
- Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica* 72(5), 1409–1443.
- Collier, P. (2001). Implications of ethnic diversity. *Economic Policy Analysis* 16(32), 129–166.
- Correlates of War 2 Project (2003). Colonial / dependency contiguity data v3.0. Electronic publication. <http://correlatesofwar.org/>.
- DANE (1990). *La Población de Colombia en 1985: Estudios de Evaluación de la Calidad y Cobertura del XV Censo Nacional de Población y IV de Vivienda*. Bogotá.
- DANE (2007). *Colombia. Proceso de Conciliación Censal 1985-2005*. Bogotá.
- DANE (2008). Metodología de cuentas departamentales. Technical report, Departamento Administrativo Nacional de Estadística.
- Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford Univ. Press.
- De Soysa, I. and J. Jütting (2007). Informal institutions and development: How they matter and what makes them change. In J. Jütting, D. Drechsler, S. Bartsch, and I. de Soysa (Eds.), *Informal Institutions. How Social Norms Help or Hinder Development*, pp. 29–43. Paris: OECD Development Centre.
- Deaton, A. (1997). *The analysis of household surveys: A microeconomic approach to development policy*. Baltimore: John Hopkins University Press.
- Departamento Nacional de Planeación DNP (2002). Evaluación de la descentralización municipal en Colombia: Balance de una década. Technical report, Departamento Nacional de Planeación DNP, Dirección de Planeación Nacional, Bogotá.
- Díaz, A. M. and F. Sánchez (2004). A geography of illicit crops (coca leaf) and armed conflict in Colombia. Documentos CEDE 2004-19, Universidad de los Andes-CEDE.
- Dijkstra, A. G. (2002). Revisiting UNDP's GDI and GEM: Towards an alternative. *Social Indicators Research* 57, 301–338.
- Dollar, D., R. Fisman, and R. Gatti (2001). Are women really the “fairer” sex? Corruption and women in government. *Journal of Economic Behavior & Organization* 46(4), 423–429.

- Drèze, J. and A. Sen (1989). *Hunger and Public Action*. WIDER Studies in Development Economics. Oxford: Oxford University Press.
- Duflo, E. (2005). Gender equality in development. BREAD Policy Paper 011, Bureau for Research in Economic Analysis of Development.
- Duong, T. (2008). *ks: Kernel Smoothing*. R package version 1.5.6.
- Duong, T. and M. Hazelton (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics* 32, 485–506.
- Durlauf, S. and D. Quah (1999). *The New Empirics of Economic Growth*, Volume 1, Chapter 4, pp. 235–308. North Holland Press.
- Durlauf, S. N. (1996). Controversy on the convergence and divergence of growth rates. *The Economic Journal* 106, 1016–1018.
- Durlauf, S. N. (2001). Manifesto for a growth econometrics. *Journal of Econometrics* 100, 65–69.
- Durlauf, S. N., A. Johnson, Paul, and R. W. Temple, Jonathan (2005). Growth econometrics. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1A, pp. 555–677. Elsevier.
- Dyson, T. and M. Moore (1983). On kinship structure, female autonomy and demographic behavior in India. *Population and Economic Review* 9(1), 35–60.
- Eagly, A. H. and M. Crowley (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin* 100(3), 283 – 308.
- Eckel, C. C. and P. J. Grossman (1998). Are women less selfish than men? Evidence from dictator experiments. *The Economic Journal* 108(448), 726–735.
- Economic Commission for Africa (2004). *The African Gender and Development Index*. Addis Ababa: ECA.
- Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Emerson, P. M. and A. P. Souza (2007). Child labor, school attendance, and intrahousehold gender bias in Brazil. *The World Bank Economic Review* 21(2), 301–316.
- European Commission (1996). *First Report on Economic and Social Cohesion*. Luxembourg: Office for Official Publications of the European Communities.

- Folbre, N. (1997). Gender coalitions: Extrafamily influences on intrafamily inequality. In L. Haddad, J. Hoddinott, and H. Alderman (Eds.), *Intrahousehold Resource Allocation*. Baltimore, MD: The Johns Hopkins Press.
- Foster, G. and J. Williamson (2000). A review of current literature of the impact of HIV/AIDS on children in Sub-Saharan Africa. *AIDS* 14(3), 275–284.
- Foster, J. E., J. Greer, and E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica* 52, 761–766.
- Freedom House (2008). Freedom in the world 2008. Technical report, Freedom House. <http://www.freedomhouse.org>.
- Gatti, R. (2004). Explaining corruption: Are open countries less corrupt? *Journal of International Development* 16(6), 851–861.
- Glover, S. H., M. A. Bumpus, J. E. Logan, and J. R. Ciesla (1997). Re-examining the influence of individual values on ethical decision making. *Journal of Business Ethics* 16, 1319–1329.
- Goetz, A. M. (2007). Political cleaners: Women as the new anti-corruption force? *Development and Change* 38(1), 87–105.
- Greenacre, M. (2007). *Correspondence Analysis in Practice* (second ed.). Interdisciplinary Statistics. Boca Raton: Chapman and Hall.
- Grün, B. and F. Leisch (2008). FlexMix Version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software* 28(4), 1–35.
- Haddad, E. A., F. S. Perobelli, J. Bonet, and G. J. D. Hewings (2008). Structural interdependence among Colombian departments. Borradores de Economía 517, Banco de la República.
- Haddad, L. and J. Hoddinott (1994). Women's income and boy-girl anthropometric status in Cote d'Ivoire. *World Development* 22(4), 543–553.
- Hall, P. A. and R. C. R. Taylor (1996). Political science and the three New Institutionalisms. *MPIFG Discussion Paper* 96/6.
- Hamilton, L. C. (1992). How robust is robust regression? *Stata Technical Bulletin* 1(2).
- Härdle, W., M. Müller, S. Sperlich, and A. Werwartz (2004). *Nonparametric and Semiparametric Methods*. Springer Series in Statistics. Berlin: Springer.

- Hatt, L. E. and H. R. Waters (2006). Determinants of child morbidity in Latin America: A pooled analysis of interactions between parental education and economic status. *Social Science and Medicine* 62, 375–386.
- Hausmann, R., L. D. Tyson, and S. Zahidi (2007). *The Global Gender Gap Report 2007*. Geneva: World Economic Forum.
- Hernández, G. H. (2004). Impacto de las regalías petroleras en el departamento del Meta. Ensayos sobre economía regional, Banco de la República, Centro Regional de Estudios Económicos Villavicencio.
- Hernández, M. and D. Obregón (2002). *La Organización Panamericana de la Salud y el Estado Colombiano: Cien Años de Historia, 1902-2002*. Bogotá: Organización Panamericana de la Salud.
- Hill, M. A. and E. M. King (1995). Women's education and economic well-being. *Feminist Economics* 1(2), 21–46.
- Hindin, M. J. (2000). Women's autonomy, women's status and fertility-related behavior in Zimbabwe. *Population Research and Policy Review* 19, 255–282.
- Hollander, M. and D. A. Wolfe (1999). *Nonparametric Statistical Methods* (Second ed.). Wiley Series in Probability and Statistics. New York: John Wiley and Sons.
- Homedes, N. and A. Ugalde (2005). Las reformas de salud neoliberales en América Latina: Una revisión crítica a través de dos estudios de caso. *Pan American Journal of Public Health* 17(3), 210–220.
- Inglehart, R., P. Norris, and C. Welzel (2002). Gender equality and democracy. *Comparative Sociology* 1(3–4), 321–345.
- Jolliffe, I. T. (1986). *Principal component analysis*. New York: Springer.
- Jones, M., J. Marron, and B. Park (1991). A simple root n bandwidth selector. *Annals of Statistics* 19, 1919–1932.
- Jütting, J., A. Luci, and C. Morrisson (2010). Why do so many women end up in bad jobs? A cross-country assessment. Working Paper 287, OECD Development Centre.
- Jütting, J., C. Morrison, J. Dayton-Johnson, and D. Drechsler (2008). Measuring gender (In)Equality: The OECD gender, institutions and development data base. *Journal of Human Development* 9(1), 65–86.



- Jütting, J. and C. Morrisson (2005). Changing social institutions to improve the status of women in developing countries. Technical Report Policy Brief 27, OECD Development Centre.
- Kakwani, N. C. (1984). Issues in measuring poverty. In R. L. Basman and J. G. F. Rhodes (Eds.), *Advances in Econometrics*, Volume 3. Greenwich, CT and London: JAI Press.
- Kanbur, R. (2003). Education, empowerment and gender inequalities. In N. Stern and B. Pleskovic (Eds.), *The New Reform Agenda / Annual World Bank Conference on Development Economics*. Oxford: Oxford University Press for The World Bank.
- Kauermann, G. and R. J. Carroll (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 96(456), 1387–1396.
- Kaufmann, D., A. Kraay, and M. Mastruzzi (2008). Governance matters VII: Aggregate and individual governance indicators, 1996–2007. World Bank Policy Research Working Paper 4654, World Bank.
- Kazianga, H. and S. Klöner (2009). The intra-household economics of polygyny: Fertility and child mortality in rural Mali. MPRA Paper 12859, University Library of Munich, Germany.
- Keefer, P. and S. Khemani (2005). Democracy, public expenditure, and the poor: Understanding political incentives for providing public services. *The World Bank Research Observer* 20(1), 1–27.
- Kendall, M. (1976). *Rank Correlation Methods* (Fourth ed.). London: Griffin.
- Kenny, C. (2004). Why are we worried about income? Nearly everything that matters is converging. *World Development* 33(1), 1–19.
- King, E. M. and M. A. Hill (1993). *Women's Education in Developing Countries*. Baltimore: John Hopkins University Press.
- Klasen, S. (2002). Low schooling for girls, slower growth for all? *World Bank Economic Review* 16(3), 345–373.
- Klasen, S. (2006). UNDP's gender-related measures: Some conceptual problems and possible solutions. *Journal of Human Development* 7(2), 243–274.
- Klasen, S. (2007). Gender-related indicators of well-being. In M. McGillivray (Ed.), *Human Well-Being: Concept and Measurement*, Studies in Development Economics and Policy, Chapter 7, pp. 167–192. New York, NY: Palgrave Macmillan.

- Klasen, S. and F. Lamanna (2009). The impact of gender inequality in education and employment on economic growth in developing countries: Updates and extensions. *Feminist Economics* 15(3), 91–132.
- Klasen, S. and D. Schüler (2009). Reforming the gender-related development index (GDI) and the gender empowerment measure (GEM): Some specific proposals. Ibero-America Institute for Economic Research (IAI) Discussion Papers 186, University of Goettingen.
- Klasen, S. and C. Wink (2003). Missing women: Revisiting the debate. *Feminist Economics* 9, 263–300.
- Kolenikov, S. and G. Angeles (2004). The use of discrete data in PCA: Theory, simulations, and applications to socioeconomics indices. CPC/MEASURE Working paper WP-04-85, Carolina Population Center.
- Kolenikov, S. and G. Angeles (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth* 55(1), 128–165.
- Lahiri, S. and S. Self (2007). Gender bias in education: The role of inter-household externality, dowry and other social institutions. *Review of Development Economics* 11(4), 591–606.
- Lal, K. (1999). Implementation of the 1993 SNA in Canada - backcasting issues. Technical report, System of National Accounts Branch. Statistics Canada.
- Lambsdorff, J. G. (2006). Measuring corruption - The validity and precision of subjective indicators (CPI). In C. Sampford, A. Shacklock, C. Connors, and F. Galtung (Eds.), *Measuring Corruption*, Law, Ethics and Governance Series, Chapter 5, pp. 81–99. Aldershot: Ashgate.
- Lay, J. and A.-S. Robilliard (2009). The complementarity of MDG achievements: The case of child mortality in Sub-Saharan Africa. Policy Research Working Paper 5062, World Bank.
- Lee, K., M. H. Pesaran, and R. Smith (1997). Growth and convergence in a multi-country empirical stochastic Solow model. *Journal of Applied Econometrics* 12, 357–392.
- Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software* 11(8).
- Leisch, F. and B. Grün (2008). *flexmix: Flexible Mixture Modeling*. R package version 2.2-3.

- Long, J. S. and L. H. Ervin (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician* 54(3), 217–224.
- Lopez-Claros, A. and S. Zahidi (2005). *Women's Empowerment: Measuring the Global Gender Gap*. Davos: World Economic Forum.
- Lundberg, S. and R. A. Pollak (1993). Separate spheres bargaining and the marriage market. *Journal of Political Economy* 101(6), 988–1010.
- Lundberg, S. and R. A. Pollak (2008). Family decision making. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (Second ed.). Basingstoke: Palgrave Macmillan.
- Macdonald, M. (2006). Muslim women and the veil. *Feminist Media Studies* 6(1), 7–23.
- Maddala, G. S., R. P. Trost, H. Li, and F. Joutz (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics* 15, 90–100.
- Maddala, G. S. and S. Wu (2000). Cross-country growth regressions: Problems of heterogeneity, stability and interpretation. *Applied Economics* 32, 635–642.
- Magrini, S. (2004). Regional (di)convergence. In J. V. Henderson and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4, Chapter 62, pp. 2741–2796. Elsevier.
- Maitra, P. (2004). Parental bargaining, health inputs and child mortality in India. *Journal of Health Economics* 23(2), 259–291.
- Manser, M. and M. Brown (1980). Marriage and household decision making: A bargaining analysis. *International Economic Review* 21, 31–44.
- Marshall, M. G. and K. Jagers (2009). Polity IV project: Political regime characteristics and transitions, 1800–2008. Electronic publication. <http://www.systemicpeace.org/polity/polity4.htm>.
- Masanjala, W. H. and C. Papageorgiou (2004). The Solow model with CES technology: Nonlinearities and parameter heterogeneity. *Journal of Applied Econometrics* 19, 171–200.
- Mazumdar, K. (2003). Do standards of living converge? A cross-country study. *Social Indicators Research* 64, 29–50.

- McCulloch, C. E. and S. R. Searle (2001). *Generalized, Linear, and Mixed Models*. Wiley Series in Probability and Statistics. New York, NY: John Wiley and Sons.
- McElroy, M. B. (1990). The empirical content of Nash-bargained household behavior. *Journal of Human Resources* 25, 559–583.
- McElroy, M. B. and M. J. Horney (1981). Nash bargained household decisions. *International Economic Review* 22, 333–349.
- McGillivray, M. and H. White (1993). Measuring development? The UNDP's Human Development Index. *Journal of International Development* 5(2), 183–192.
- Meisel, A. (2007a). La Guajira y el mito de las regalías redentoras. Documentos de Trabajo sobre Economía Regional 86, Banco de la República.
- Meisel, A. (2007b). ¿Por qué se necesita una política económica regional en Colombia? Documentos de Trabajo sobre Economía Regional 100, Banco de la República.
- Meisel, A. and M. Vega (2007). The biological standard of living (and its convergence) in Colombia, 1870–2003 - A tropical success story. *Economics and Human Biology* 5(1), 100–122.
- Micklewright, J. and K. Stewart (1999). Is the well-being of children converging in the European Union? *The Economic Journal* 109(459), 692–714.
- Milallos, M. T. R. (2007). Muslim veil as politics: Political autonomy, women and Syariah Islam in Aceh. *Contemporary Islam* 1(3), 289–301.
- Morrison, C. and J. P. Jütting (2005). Women's discrimination in developing countries: A new data set for better policies. *World Development* 33(7), 1065–1081.
- Mukherjee, C., H. White, and M. Wuyts (1998). *Econometrics and Data Analysis for Developing Countries*. London: Routledge.
- Munda, G. and M. Nardo (2005a). Constructing consistent composite indicators: The issue of weights. Technical Report EUR 21834 EN, European Commission.
- Munda, G. and M. Nardo (2005b). Non-compensatory composite indicators for ranking countries: A defensible setting. Technical Report EUR 21833 EN, European Commission.
- Murad R., R. (2003). Estudio sobre la distribución espacial de la población en Colombia. Technical Report 48, CELADE/CEPAL.

- Nardo, M., M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, and E. Giovannini (2005). Handbook on constructing composite indicators: Methodology and user guide. Technical Report 2005/3, OECD.
- Nenadić, O. (2007). *An Implementation of Correspondence Analysis in R and its Application in the Analysis of Web Usage*. Ph. D. thesis, University of Goettingen, Goettingen.
- Neumayer, E. (2003). Beyond income: Convergence in living standards, big time. *Structural Change and Economic Dynamics* 14, 275–296.
- North, D. C. (1990). *Institutions, Institutional Change, and Economic Performance*. New York: Cambridge University Press.
- Oeppen, J. and J. W. Vaupel (2002). Broken limits to life expectancy. *Science* 296(5570), 1029–1031.
- Olshansky, S. J., B. A. Carnes, and C. Cassel (1990). In search of Methuselah: Estimating the upper limits to human longevity. *Science* 250(4981), 634–640.
- Pardo, I. (2004). *Between Morality And The Law: Corruption, Anthropology And Comparative Society*. Aldershot: Ashgate.
- Pasqua, S. (2005). Gender bias in parental investments in children's education: A theoretical analysis. *Review of Economics of the Household* 3(3), 291–314.
- Pew Forum on Religion and Public Life (2009). Mapping the global Muslim population: A report on the size and the distribution of the world's Muslim population.
- Pinheiro, J. C. and D. M. Bates (2000). *Mixed-Effects Models in S and S-Plus*. Statistics and Computing. New York, NY: Springer.
- Pollak, R. A. (2003). Gary Becker's contributions to family and household economics. *Review of Economics of the Household* 1, 111–141.
- Pollak, R. A. (2007). Bargaining around the hearth. NBER Working Papers 13142, National Bureau of Economic Research.
- Pritchett, L. (2000). Patterns of economic growth: Hills, plateaus, mountains, and plains. *World Bank Economic Review* 14, 221–250.
- Profamilia (2005). Salud social y reproductiva en Colombia, encuesta nacional de demografía y salud. Technical report, Profamilia.

- Quah, D. (1993a). Empirical cross-section dynamics in economic growth. *European Economic Review* 37(2-3), 426–434.
- Quah, D. (1993b). Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics* 95(4), 427–443.
- Quah, D. (1996). Empirics for economic growth and convergence. *European Economic Review* 40(6), 1353–1375.
- Quah, D. (1997). Empirics for growth and distribution: Stratification, polarization and convergence clubs. *Journal of Economic Growth* 2(1), 27–59.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ramsey, F. P. (1928). A mathematical theory of saving. *Economic Journal* 38, 543–559.
- Rasul, I. (2008). Household bargaining over fertility: Theory and evidence from Malaysia. *Journal of Development Economics* 86, 215–241.
- Ravallion, M. (1994). *Poverty Comparisons: A Guide to Concepts and Methods*. Chur: Harwood Academic.
- Rivas, M. F. (2008). An experiment on corruption and gender. ThE Papers 08/10, Department of Economic Theory and Economic History of the University of Granada.
- Rizzo, H., A.-H. Abdel-Latif, and K. Meyer (2007). The relationship between gender equality and democracy: A comparison of Arab versus non-Arab muslim societies. *Sociology* 4(6), 1151–1170.
- Rojas, F. (2003). At the crossroads of decentralization: Recentralization, federalization. In M. M. Giugale, O. Lafourcade, and C. Luff (Eds.), *Colombia: The Economic Foundation of Peace*, Chapter 31, pp. 871–896. Washington D.C.: World Bank.
- Sala-i-Martin, X. (1996). The classical approach to convergence analysis. *The Economic Journal* 106, 1019–1036.
- Saleem, S. and M. Bobak (2005). Women's autonomy, education and contraception use in Pakistan: A national study. *Reproductive Health* 2(8).
- Sánchez, F. and M. D. M. Palau (2006). Conflict, decentralisation and local governance in Colombia, 1974–2004. Documentos CEDE 2006-20, Universidad de los Andes-CEDE.

- Schüler, D. (2006). The uses and misuses of the Gender-related Development Index and the Gender Empowerment Measure: A review of the literature. *Journal of Human Development* 7(2), 161–182.
- Schultz, T. P. (1990). Testing the neoclassical model of family labor supply and fertility. *Journal of Human Resources* 25(4), 599–634.
- Schultz, T. P. (2002). Why governments should invest more to educate girls. *World Development* 30(2), 207–225.
- Schultz, T. P. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of Development Economics* 74, 199–250.
- Seebens, H. (2008). *The Economics of Gender and the Household in Developing Countries*. Frankfurt: Peter Lang.
- Sen, A. (1992). Missing women. *British Medical Journal* 304, 586–7.
- Sen, A. (1997). From income inequality to economic inequality. *Southern Economic Journal* 64(2), 383–401.
- Sen, A. (1998). Mortality as an indicator of economic success and failure. *Economic Journal* 108(446), 1–25.
- Sen, A. (1999). *Development as Freedom*. New York, NY: Alfred A. Knopf.
- Sen, A. (2002). *El Derecho a No Tener Hambre*. Bogotá: Universidad Externado de Colombia.
- Serra, M. I., M. F. Pazmino, G. Lindow, B. Sutton, and G. Ramirez (2006). Regional convergence in Latin America. IMF Working Paper 125, International Monetary Fund.
- Sheather, S. J. and M. C. Jones (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society* 53, 683–690.
- Shioji, E. (1997). It's still 2 percent: Evidence on convergence from 116 years of the US states panel data. Economics Working Papers 236, Department of Economics and Business, Universitat Pompeu Fabra.
- Shorrocks, A. F. and J. E. Foster (1987). Transfer sensitive inequality measures. *Review of Economic Studies* 54(1), 485–497.

- Shroff, M., P. Griffiths, L. Adair, C. Suchindran, and M. Bentley (2009). Maternal autonomy is inversely related to child stunting in Andhra Pradesh, India. *Maternal and Child Nutrition* 5, 64–74.
- Smith, L. C., U. Ramakrishnan, A. Ndiaye, L. Haddad, and R. Martorell (2002). *The Importance of Women's Status for Child Nutrition in Developing Countries*. Research Report 131. Washington, DC: International Food Policy Research Institute.
- Social Watch (2005). *Roars and Whispers Gender and Poverty: Promises versus Action*. Montevideo: Social Watch.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70, 65–94.
- Song, L., S. Appleton, and J. Knight (2006). Why do girls in rural China have lower school enrollment? *World Development* 34(9), 1639–1653.
- Subramanian, S. (2007). Indicators of inequality and poverty. In M. McGillivray (Ed.), *Human Well-Being: Concept and Measurement*, Studies in Development Economics and Policy, Chapter 6, pp. 135–166. New York, NY: Palgrave Macmillan.
- Sung, H.-E. (2003). Fairer sex or fairer system? Gender and corruption revisited. *Social Forces* 82(2), 703–723.
- Svensson, J. (2005). Eight questions about corruption. *Journal of Economic Perspectives* 19(3), 19–42.
- Swamy, A., S. Knack, Y. Lee, and O. Azfar (2001). Gender and corruption. *Journal of Development Economics* 64(1), 25–55.
- Temple, J. (1999). The new growth evidence. *Journal of Economic Literature* 37, 112–156.
- Tenjo G., F. and E. López E. (2003). Credit bubble and stagnation in Colombia, 1990–2001. *Colombian Economic Journal* 1(1), 151–191.
- Thomas, D. (1990). Intra-household resource allocation: An inferential approach. *Journal of Human Resources* 25(4), 635–64.
- Thomas, D. (1997). Incomes, expenditures and health outcomes: Evidence on intrahousehold resource allocation. In L. Haddad, J. Hoddinott, and H. Alderman (Eds.), *Intra-household Resource Allocation in Developing Countries: Models, Methods and Policy*. Baltimore, MD: The Johns Hopkins Press.



- Treisman, D. (2007). What have we learned about the causes of corruption from ten years of cross-national empirical research? *Annual Review of Political Science* 10, 211–244.
- Tripp, A. M. (2001). Women's movements and challenges to neopatrimonial rule: Preliminary observations from Africa. *Development and Change* 32, 33–54.
- UNAIDS/WHO (2008). Report on the global AIDS epidemic. Technical report, UNAIDS/WHO, Geneva.
- UNICEF (1998). *The State of World's Children: Focus on Nutrition*. New York: UNICEF.
- United Nations (2004). *World Population to 2300*. United Nations Population Studies No. 236. New York: United Nations.
- United Nations Development Programme (1995). *Human Development Report 1995. Gender and Human Development*. New York: Oxford University Press.
- United Nations Development Programme (2006). *Human Development Report 2006. Beyond scarcity: Power, poverty and the global water crisis*. New York: Oxford University Press.
- United Nations Development Programme (2010). *Human Development Report 2010. The Real Wealth of Nations: Pathways to Human Development*. New York: Palgrave Macmillan.
- Velez, E., D. Harding, and A. Sarmiento (2003). Education. In M. M. Giugale, O. Lafourcade, and C. Luff (Eds.), *Colombia: The Economic Foundation of Peace*, Chapter 21, pp. 611–651. Washington D.C.: World Bank.
- Wand, M. P. and M. C. Jones (1995). *Kernel Smoothing*. Monographs on Statistics and Applied Probability 60. London: Chapman and Hall.
- Wasserman, L. (2006). *All of nonparametric statistics*. Springer Texts in Statistics. New York: Springer.
- Waylen, G. (1993). Women's movements and democratization in Latin America. *Third World Quarterly* 14(3), 573 – 587.
- Wilcox, R. (2001). Comment. *The American Statistician* 55(4), 374–375.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.
- World Bank (2001). *Engendering Development: Through Gender Equality in Rights, Resources and Voices*. New York, NY: World Bank/Oxford University Press.

World Bank (2008). *World Development Indicators 2008*. Technical report, World Bank, Washington, DC.

World Bank (2009a). *GenderStats*. Electronic publication. <http://genderstats.worldbank.org>.

World Bank (2009b). *World Development Report 2009: Reshaping Economic Geography*. Washington, DC: The World Bank.

**Göttinger Studien zur Entwicklungsökonomik  
Göttingen Studies in Development Economics**

Herausgegeben von / Edited by Hermann Sautter  
und / and Stephan Klasen

Die Bände 1-8 sind über die Vervuert Verlagsgesellschaft (Frankfurt/M.) zu beziehen.

- Bd./Vol. 9 Hermann Sautter / Rolf Schinke (eds.): Social Justice in a Market Economy. 2001.
- Bd./Vol. 10 Philipp Albert Theodor Kircher: Poverty Reduction Strategies. A comparative study applied to empirical research. 2002.
- Bd./Vol. 11 Matthias Blum: Weltmarktintegration, Wachstum und Innovationsverhalten in Schwellenländern. Eine theoretische Diskussion mit einer Fallstudie über „Argentinien 1990-1999“. 2003.
- Bd./Vol. 12 Jan Müller-Scheeßel: Die Privatisierung und Regulierung des Wassersektors. Das Beispiel Buenos Aires / Argentinien. 2003.
- Bd./Vol. 13 Ludger J. Löning: Economic Growth, Biodiversity Conservation, and the Formation of Human Capital in a Developing Country. 2004.
- Bd./Vol. 14 Silke Woltermann: Transitions in Segmented Labor Markets. The Case of Brazil. 2004.
- Bd./Vol. 15 Jörg Stosberg: Political Risk and the Institutional Environment for Foreign Direct Investment in Latin America. An Empirical Analysis with a Case Study on Mexico. 2005.
- Bd./Vol. 16 Derk Bienen: Die politische Ökonomie von Arbeitsmarktreformen in Argentinien. 2005.
- Bd./Vol. 17 Dierk Herzer: Exportdiversifizierung und Wirtschaftswachstum. Das Fallbeispiel Chile. 2006.
- Bd./Vol. 18 Jann Lay: Poverty and Distributional Impact of Economic Policies and External Shocks. Three Case Studies from Latin America Combining Macro and Micro Approaches. 2007.
- Bd./Vol. 19 Kenneth Hartgen: Empirical Analysis of Determinants, Distribution and Dynamics of Poverty. 2007.
- Bd./Vol. 20 Stephan Klasen / Felicitas Nowak-Lehmann: Poverty, Inequality and Migration in Latin America. 2008.
- Bd./Vol. 21 Isabel Günther: Empirical Analysis of Poverty Dynamics. With Case Studies from Sub-Saharan Africa. 2007.
- Bd./Vol. 22 Peter Dung: Malaysia und Indonesien: Wirtschaftliche Entwicklungsstrategien in zwei Vielvölkerstaaten. 2008.
- Bd./Vol. 23 Thomas Otter: Poverty, Income Growth and Inequality in Paraguay During the 1990s. Spatial Aspects, Growth Determinants and Inequality Decomposition. 2008.
- Bd./Vol. 24 Mark Misselhorn: Measurement of Poverty, Undernutrition and Child Mortality. 2008.
- Bd./Vol. 25 Julian Weisbrod: Growth, Poverty and Inequality Dynamics. Four Empirical Essays at the Macro and Micro Level. 2008.
- Bd./Vol. 26 Johannes Gräß: Econometric Analysis in Poverty Research. With Case Studies from Developing Countries. 2009.
- Bd./Vol. 27 Sebastian Vollmer: A Contribution to the Empirics of Economic and Human Development. 2009.
- Bd./Vol. 28 Wokia-azi Ndangle Kumase: Aspects of Poverty and Inequality in Cameroon. 2010.

- Bd./Vol. 29 Adriana Rocío Cardozo Silva: Economic Growth and Poverty Reduction in Colombia. 2010.
- Bd./Vol. 30 Ronald Kröker: Ansätze zur Implementierung von RSE (CSR) in einem lateinamerikanischen Entwicklungsland. Das Beispiel Paraguay – Eine wirtschafts- und unternehmensethische Untersuchung. 2010.
- Bd./Vol. 31 Maria Ziegler: Institutions, Inequality and Development. 2011.
- Bd./Vol. 32 Melanie Grosse: Measurement of Trends in Wellbeing, Poverty, and Inequality with Case Studies from Bolivia and Colombia. 2011.
- Bd./Vol. 33 Boris Branisa Caballero: Social Institutions, Gender Inequality, and Regional Convergence in Developing Countries. 2012.

[www.peterlang.de](http://www.peterlang.de)



