2005

Health and Growth

Petia Stoianova Stoytcheva
Louisiana State University and Agricultural and Mechanical College, pstoyt1@lsu.edu

HEALTH AND GROWTH

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Economics

by
Petia Stoianova Stoytcheva
B.S., University of National and World Economy, Sofia, Bulgaria, 2000
M.S., Louisiana State University, 2002
December 2005

To my family

Acknowlegements

At the close of this project, it is my pleasure and pride to express my debt of gratitude to my advisor, Professor Chris Papageorgiou who put me on this path three years ago and provided me with constant guidance and assistance. I would like to thank him for his vision, encouragement, and unerring advice. I also owe an enormous debt of gratitude to Professor Marios Zachariadis without whose direction and encouragement this work would not have come to fruition.

It is a pleasure to acknowledge my research experience and fruitful discussions with Professor Carter Hill. His professionalism and discipline guided me for the last three years for which I am very grateful. I am equally indebted to Professor Douglas McMillin for his comments and invaluable insights which helped to make this dissertation more coherent and presentable.

I also would like to thank Professor Areendam Chanda,and Professor Christopher Kenny from the Department of Political Science for serving on my dissertation committee.

I would like to thank Professor Mario Stoitsov for his help on the final preparation of this manuscript and encouragement the past few years. I would like to express my personal thanks to Mohd Al-Azzam for his great friendship and encouragement for the past four years. I am thankful to Kalin Drumev and Viera Chemelarova who made my time here more enjoyable. Thanks also go to the recent and former graduate students in the Economics Department at Louisiana State University.

I wish to acknowledge financial support from the Louisiana State University Graduate School.

I dedicate this thesis to my parents and my sister, for their love, unwavering dedication, and support.

Table of Contents

List of Tables

List of Figures

Abstract

Inarguably the most important question is about the unequal distribution of income among countries. Development economists have recently turned to health for an answer. This dissertation investigates the effect of health on cross-country income.

The first essay sheds new light on the impact of AIDS on cross-country income levels. We control for a variety of factors that are potentially related to income as suggested by our empirical model and existing related literature. Using the extended (for human capital) Solow model as our baseline empirical specification, we consider cross-sectional and panel estimation. For the full sample it is shown that AIDS has a negative and significant effect on the level of income in both the cross-sectional, and panel estimations. When we arbitrarily split our full sample into OECD and non-OECD countries, we find that the AIDS coefficient continues to be negative and significant for the non-OECD subsample.

The second essay constructs gender-specific human capital inequality measures using the Gini coefficient. It also considers a new channel through which infant mortality affects economic growth-female human capital inequality. It is inequality in education among women that affects infant mortality and the latter affects economic growth and development. We consider cross-sectional and panel data analysis and use common instruments to correct for endogeneity of infant mortality. Our analysis suggests diverting general education subsidy money directly into the education of the least educated women, especially in less-developed countries.

Chapter 1

Introduction

In the last 40 years economists have been trying to explain a set of puzzles: Why are there gaps in income between rich and poor countries? Why have some countries experienced rapid economic growth, while others stagnated in poverty traps? Are there any specific country-related characteristics that explain this? What are the different economic policies implemented in each country? Development economists have recently turned to health for an answer. Understanding the effect of health on income is important for two reasons (Shastry and Weil (2002)). First, if health does have a large effect on income per capita, then this would be an important additional benefit of health improvement, and, second, accounting for health differences will reduce the size of the unexplained residual variance in income among countries that is currently attributed to productivity.

This Ph.D. thesis is a combination of two essays which empirically test the effects of health on economic growth. It attempts to investigate the following: (i) the effects of infectious diseases like AIDS on cross-country income and development and (ii) the relationship between female human capital inequality and infant mortality, and the effect of the latter on economic growth.

The first essay, *What Do We Know About the Impact of AIDS on Cross-Country Income So Far?,* examines empirically the potential effect of AIDS on cross-country income. The scope of the world-wide AIDS epidemic is staggering (Kalemli-Ozcan, 2005). The World Heath Organization (WHO) estimated that in December 2002, 42 million people were living with the human immunodeficiency virus (HIV) or the acquired immune deficiency syndrome (AIDS). The newly infected with HIV in 2002 totaled 5 million and AIDS related deaths in 2002 were 3.1 million. HIV/AIDS now ranks as the world's fourth largest cause of death, after heart disease, strokes and acute lower respiratory infections (Dixon, McDonald, Roberts (2002)). It is feared that AIDS will soon surpass malaria,

which has been around for at least a millennium and is considered as the most deadly infectious disease.

Four characteristics distinguish AIDS and make the economic impact far greater than all other diseases (Kalemli-Ozcan, 2005). First, it is always fatal. Second, AIDS in Africa is affecting prime-aged adults in their most productive years. Third, it is the leading cause of death in Africa. Fourth, AIDS affects the educated and the upper class individuals. A central point of analysis for economists is to evaluate the impact of AIDS on economic welfare and in particular on per capita income. There is a small but rapidly expanding literature related to the economic effects of AIDS. Several theoretical papers suggest large negative economic consequences of the pandemic. For example, Cuddington (1993), based on a simulation of a modified Solow model concluded that AIDS, via its impact on morbidity and mortality rates, would likely reduce GDP in Tanzania in 2010 by 15 to 20 percent relative to a counterfactual no-AIDS scenario. Similarly, Cuddington and Hancock (1994) using a similar methodology simulated the impact of AIDS on the Malawian economy and found that the average annual real per capita GDP growth over the 1985-2010 period is projected to be 0.2-0.3 percentage points lower compared to the alternative no-AIDS scenario.

More recently, Ferreira and Pessoa (2003) have proposed a model in which AIDS reduces income by affecting the incentives for schooling attainment due to shorter expected longevity. Based on their model, the most affected countries in sub-Saharan Africa are predicted to become about 25 percent poorer than they would have been without AIDS, with schooling declining by about 50 percent. Finally, Corrigan, Glomm, and Mendez (2003) constructed and fully studied a model that exhibited substantial negative growth effects of the AIDS epidemic, mainly through the detrimental impact of lower life expectancy on investment combined with a sizable number of orphans created by AIDS. Even though the above papers have contributed to our understanding of the problem, they are based on theoretical models that are taken to the data by means of numerical simulation exercises and do not utilize the full information that potentially exists in existing AIDS data. Other recent notable theoretical papers include Levy (2002), Auld (2003), Clark and Vencatachellum, and Oster (2004).

On the empirical side, the little work that exists has focused on the use of micro data – at the village or country level; see e.g. Wachter, Knodel and VanLandingham (2003), de Walque (2004), and Young (2004). In his interesting and highly controversial paper, Young (2004) attempts to

calculate the impact of the AIDS epidemic on future living standards in South Africa. He concludes that from the perspective of per capita living standards, the AIDS epidemic endows society with additional resources which in turn could be used to care for the afflicted and provide higher living standards to future generations. An exception is an important contribution by Bloom and Mahal (1997). These authors use standard epidemiological models to estimate the number of AIDS incidents from information on HIV prevalence at a point in time. Utilizing their rather scarce cross-country estimates of AIDS incidents and using novel econometric techniques, these authors arrive at the conclusion that the AIDS epidemic has had an insignificant effect on the growth rate of per capita income.

This essay sheds new light on the potential effect of AIDS on cross-country income. In principle, it follows the lead of Bloom and Mahal (1997) and contributes to the embryonic literature that studies empirically the potential impact of AIDS on economic aggregates. In particular, our empirical analysis is based on the extended (for human capital) Solow specification. Making use of Penn World Table version 6.1, we extend the Mankiw, Romer and Weil (1992) (MRW hereafter) dataset until the year 2000 and consequently merge this dataset with our AIDS dataset. We obtain results using cross-sectional and panel techniques based on the extended Solow model with AIDS as an additional explanatory variable. In addition, we employ the data splitting methodology proposed by Hansen (2000) to examine whether AIDS is a valid threshold variable that can cluster countries into groups obeying different statistical models.

Our main findings show that AIDS incidents have a negative and significant effect on the level of income for the full sample in both the cross-sectional and panel estimations. When we arbitrary split our full sample into OECD and non-OECD countries, we find that the AIDS coefficient continues to be negative and significant for the non-OECD subsample, but not for the OECD subsample. Second, when we use Hansen's (2000) endogenous splitting methodology, we find that AIDS is a threshold variable that can split countries into four different regimes. Third, exploiting a nice feature of our dataset that allows us to disaggregate the data in four different age groups, we find that only the AIDS coefficient corresponding to the age group 16-34 is negative and significant. Finally, a thorough robustness analysis shows that our results are quite robust to different subsamples and regression specifications.

The second essay, *Female Human Capital Inequality, Infant Mortality and Growth*, investigates the relationship between female human capital inequality, infant mortality and economic growth. More specifically, we empirically test the following two-step relationship. In the first step, inequality in education among women leads to higher infant mortality. We conjecture that higher inequality in education among women is responsible for higher infant mortality because mothers fail to provide adequate care of their infants. In the second step, higher infant mortality is partly responsible for low growth and development experienced by many developing countries.

Due to the lack of available data on human capital inequality, little attention has been devoted to the influence of human capital distribution on economic growth in empirical studies. Following Castelló and Domenech (2002), we construct gender-specific Gini coefficients as a measure of human capital inequality. In our analysis, instead of using the level of education, we use the Gini coefficient which is a better measure of the distribution of education. This approach allows us to include the least educated women and draw conclusions about the potential impact of female human capital inequality on infant mortality. We provide evidence on the effect of female human capital inequality on infant mortality and the negative effect of the latter on economic growth. We are the first to construct female and male human capital inequality measures using the Barro and Lee dataset. The novelty of our data is an important part of our contribution to the literature. We are also the first to propose inequality in education among women as an explanation of higher infant mortality across countries. We provide evidence that inequality in education among women, men and total inequality have been decreasing over time.

There is a small, but rapidly growing macro literature explaining the decline in infant and child mortality. Schultz (1993) argues that women's education is the most significant determinant of child mortality. As another explanation for the mortality decline Jamison, Sandbu and Wang (2001) propose technological progress. Countries may differ in how close their health systems come to utilizing the technology available at any given time. In a recent paper Lorentzen, McMillan and Wacziarg (2004) focus on adult mortality as an explanation of low growth. They argue that poverty leads to high mortality, which leads to low growth.

This essay contributes to the literature in two ways. First, we construct gender-specific measures of human capital inequality and show that female human capital inequality has a positive effect on

infant mortality. Second, we show that in cross-sectional and panel regressions infant mortality has a negative and significant effect on economic growth.

Chapter 2

What Do We Know About the Impact of AIDS on Cross-Country Income So Far?

The World Heath Organization (WHO) estimated that in December 2002, 42 million people were living with the human immunodeficiency virus (HIV) or the acquired immune deficiency syndrome (AIDS). The newly infected with HIV in 2002 totaled 5 million and AIDS related deaths in 2002 were 3.1 million. HIV/AIDS now ranks as the world's fourth largest cause of death, after heart disease, strokes and acute lower respiratory infections (Dixon, McDonald, Roberts (2002)).[1] It is feared that AIDS will soon surpass malaria, which has been around for at least a millennium, and considered as the most deadly infectious disease. AIDS may be a relatively new infectious disease, only quarter of a century old, but its negative impact is felt most profoundly in sub-Saharan Africa in which it is erasing decades of progress made in extending quantity and improving quality of life.[2]

AIDS' alarming infection rate coupled with no known cure has very important social, political, demographic and certainly economic implications. A central point of analysis for economists is to evaluate the impact of AIDS on economic welfare and in particular on per capita income. There is a small but rapidly expanding literature related to the economic effects of AIDS. Several theoretical papers suggest large negative economic consequences of the pandemic. For example, Cuddington (1993), simulating a modified Solow model, concluded that AIDS, via its impact on morbidity and mortality rates, would likely reduce GDP in Tanzania in 2010 by 15 to 20 percent relative to a

[1] For a very insightful introduction to AIDS and the various ways that is embedded within social, cultural, political, ideological and economic contexts see the book by Kalipeni et al. (2004). Extensive information on the AIDS epidemic and its economic consequences is available online at: *http://www.worldbank.org/aids-econ/*.

[2] Average life expectancy at birth in sub-Saharan countries is now 47 years, when according to experts it could have been as high as 62 without AIDS.

counterfactual of no-AIDS scenario. Similarly, Cuddington and Hancock (1994) using a similar methodology simulated the impact of AIDS on the Malawian economy and found that the average annual real per capita GDP growth over the 1985-2010 period is projected to be 0.2-0.3 percentage points lower compared to the alternative no-AIDS scenario.

More recently, Ferreira and Pessoa (2003) have proposed a model in which AIDS impacts negatively on income by affecting the incentives for schooling attainment due to shorter expected longevity. Based on their model, the most affected countries in sub-Saharan Africa are predicted to become about 25 percent poorer than they would have been without AIDS, with schooling declining by about 50 percent. Finally, Corrigan, Glomm, and Mendez (2003) constructed and fully studied a model that exhibited substantial negative growth effects of the AIDS epidemic, mainly through the detrimental impact of lower life expectancy on investment combined with a sizable number of orphans created by AIDS. Even though the above papers have contributed to our understanding of the problem, they are based on theoretical models that are taken to the data by means of numerical simulation exercises and do not utilize the full information that potentially exists in existing AIDS data.[3]

At the empirical side, the little work that exists has focused on the use of mirco data – at the village or country level; see e.g. Wachter, Knodel and VanLandingham (2003), de Walque (2004), and Young (2004).[4,5] An exception is an important contribution by Bloom and Mahal (1997). These authors use standard epidemiological models to estimate the number of AIDS incidents from information on HIV prevalence at a point in time. Utilizing their rather scarce cross-country estimates of AIDS incidents and using novel econometric techniques these authors arrive to the conclusion that the AIDS epidemic has had an insignificant effect on the growth rate of per capita income.

The main goal of this essay is to provide new evidence on the potential effect of AIDS on cross-country income. In principle this work follows the lead of Bloom and Mahal (1997) and makes a

[3]Other recent notable theoretical papers include Levy (2002), Auld (2003), Clark and Vencatachellum, and Oster (2004).

[4]In his interesting and highly controversial paper, Young (2004) attempts to calculate the impact of the AIDS epidemic on future living standards in South Africa. He concludes that from the perspective of per capita living standards, the AIDS epidemic endows society with additional resources which in turn could be used to care for the afflicted and provide higher living standards to future generations.

[5]For updates on recent academic and nonacademic papers, surveys, and field studies on HIV/AIDS in developing countries visit the website of the International AIDS Economics Network at: *http://www.iaen.org/papers/*.

contribution to the embryonic literature that studies empirically the potential impact of AIDS on economic aggregates. There are two main differences between our work and that of Bloom and Mahal relating to the focus of the analysis and the data used in estimation. First, in order to address the economic implications of the disease on welfare, our framework focuses on levels rather than growth of per capita income.[6] Second, we use an alternative more comprehensive dataset on officially reported AIDS cases compiled by WHO and UNAIDS for the period 1979-2000 across 116 countries. This enables us to consider both cross-sectional regression and panel techniques to study the impact of the disease on the level of income.

In particular, our empirical analysis is based on the extended (for human capital) Solow specification. Making use of Penn World Table version 6.1 we extend the Mankiw, Romer and Weil (1992) (MRW hereafter) dataset until the year 2000 and consequently merge this dataset with our AIDS dataset. We obtain results using cross-sectional and panel techniques based on the extended Solow model with AIDS as an additional explanatory variable. In addition, we employ the data splitting methodology proposed by Hansen (2000) to examine whether AIDS is a valid threshold variable that can cluster countries into groups obeying different statistical models.

Our main findings are as follows: First, we show that AIDS incidents has a negative and significant effect on the level of income for the full sample in both the cross-sectional and panel estimations. When we arbitrary split our full sample into OECD and non-OECD countries, we find that the AIDS coefficient continues to be negative and significant for the non-OECD subsample, but not for the OECD subsample. Second, exploiting a nice feature of our dataset that allows us to disaggregate the data in four different age groups, we find that only the AIDS coefficient corresponding to the age group 16-34 is negative and significant. Third, when we use Hansen's (2000) endogenous splitting methodology, we find that AIDS is a threshold variable that can split countries into regimes that obey different statistical models. Finally, robustness analysis shows that our results are quite robust to different subsamples and regression specifications.

The remainder of the essay is organized as follows. Section 2.2 takes a first look at the AIDS data used in our empirical analysis. Section 2.3 presents our baseline cross-sectional and panel estimation

[6]For this and other arguments in favor of using levels rather than growth regressions, see Hall and Jones (1999, pp. 85-86). Others papers that use level regressions include Frankel and Romer (1999), Acemoglu, Johnson and Robinson (2001), and Caselli and Wilson (2004), just to name a few.

results for the full sample and various exogenously and endogenously determined subsamples of countries. Section 2.4 examines the robustness of our baseline results by considering alternative subsamples and regression specifications. Section 2.5 discusses of our main results with particular emphasis in interpretation, and potential caveats of our analysis. Section 2.6 concludes.

2.1 A Look at the Data

We begin by describing the AIDS data used in our estimation. Later on, we explain how we update the MRW original dataset to obtain the rest of the data needed for our analysis.

2.1.1 The AIDS Dataset

We constructed the AIDS dataset which includes 116 countries over 1979-2000 using the officially reported cases from the UNAIDS/WHO Global Surveillance fact sheets.[7] The WHO "case" definition for AIDS surveillance is as specified in "Weekly Epidemiological Record," WHO, Geneva (1994).[8] For each country in the sample we start from the year during which a case was reported. We multiply the number of reported incidents by 100,000 and divide by total population in each year (data on population is from the *World Development Indicators* (2002)) to obtain incidence per 100,000 per country per year. The officially reported AIDS cases represent the number of new AIDS infections, occurring each year. Thus, we obtain AIDS incidence, which is a flow measure. Due to data constraints associated with explanatory variables necessary for our empirical analysis other than AIDS, our sample is reduced from 116 countries to 89.[9] Regarding the cross-sectional estimation, for each country in the sample we average AIDS incidents, starting from the year in which a case was reported (usually 1979) up to the year 2000. For the panel estimation, we average the data into 5 year periods for which the disturbance terms are less likely to be influenced by business cycle fluctuations. Thus, we construct three non-overlapping five-year time intervals 1985-1990, 1990-1995 and 1995-2000.

Next, we take a first look at the AIDS dataset by presenting correlations and descriptive statis-

[7] Of note is the exclusion of South Africa from our dataset due to the gross under-reporting observed and documented by many field researchers. We thank participants at the North East Universities Development Consortium (NEUDC) 2004 conference and in particular Mark Gersovitz, Damien de Walque, Désiré Vencatachellum, for their insights on the substantial measurement errors present in the South African AIDS dataset.

[8] For a detailed description of the definition, see Appendix B.

[9] More on the sample used in our empirical estimation later on. For more information about the sample of countries and relevant variables used in the estimation, see Appendix A, Table A1.

Table 2.1: Regional descriptive statistic

Regions	Variable	Mean	Stand. Dev.	Min.	Max.
Africa	GDP per worker ($)	2195	2395	461	10294
	AIDS cases per 100,000	22.317	37.632	0.021	173.043
Americas	GDP per worker ($)	6192	5234	1075	22934
	AIDS cases per 100,000	6.326	6.734	0.217	26.818
Asia	GDP per worker ($)	7951	6799	1004	21205
	AIDS cases per 100,000	1.129	3.596	0.001	17.047
Europe	GDP per worker ($)	15322	5595	4424	29274
	AIDS cases per 100,000	2.046	2.127	0.022	8.412
Oceania	GDP per worker ($)	10566	7855	3152	19424
	AIDS cases per 100,000	1.433	1.120	0.162	2.872
World	**GDP per worker ($)**	**7153**	**6888**	**461**	**29274**
	AIDS cases per 100,000	**9.938**	**24.355**	**0.001**	**173.043**

Notes: The mean, standard deviation, minimum and maximum values presented above are computed for 41 countries in Africa, 25 countries in the Americas, 22 countries in Asia, 24 countries in Europe, 4 countries in Oceania. GDP per worker and AIDS incidents represent averages since an AIDS case was reported annually from 1979 until 2000.

tics at the regional and country levels. In addition, we exploit a nice feature of our dataset and disaggregate our data into AIDS incidents by four age groups (0-4, 5-15, 16-34, 35-60+). We present examples from this disaggregated dataset for selected countries.

Table 2.1 presents the mean, standard deviation, minimum and maximum of AIDS and mean GDP per worker for five regions and the world.[10] The main reason for grouping countries into regions is to examine whether geographical location matters. We note that the mean for AIDS in Africa (22.317) is much higher than in all other regions/continents. Another interesting observation is the quite high incidence of AIDS in the Americas (with mean 6.326). It is much higher than in Europe, where the mean incidence of AIDS is 2.016. Finally, it is readily seen that Asia and Oceania are experiencing considerably lower AIDS incidents than Africa, the Americas and Europe even though, as the standard deviation reveals, there also exists substantial variation between countries in these

[10] **Africa:** Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, C. African Rep., Chad, Comoros, Congo, Egypt, Gabon, Gambia, Ghana, Guinea, Guinea-Bis., Kenya, Lesotho, Madagascar, Malawi, Mali, Mauritania, Mauritius, Morocco, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Seychelles, Sierra Leone, Sudan, Swaziland, Tanzania, Togo, Tunisia, Uganda, Zambia, Zimbabwe. **Americas:** Argentina, Barbados, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Dominican Rep., Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Tri.&Tobago, USA, Uruguay, Venezuela. **Asia:** Bangladesh, China, Cyprus, Hong Kong, India, Indonesia, Iran, Israel, Japan, Jordan, Korea, Malaysia, Oman, Pakistan, Philippines, Saudi Arabia, Singapore, Sri Lanka, Syria, Thailand, Turkey, Yemen. **Europe:** Austria, Belgium, Czech Republic, Denmark, Finland, France, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, Netherlands, Malta, Norway, Poland, Portugal, Romania, Russian Fed., Slovakia, Spain, Sweden, Switzerland, UK. **Oceania:** Australia, Fiji, New Zealand, Papua N.G..

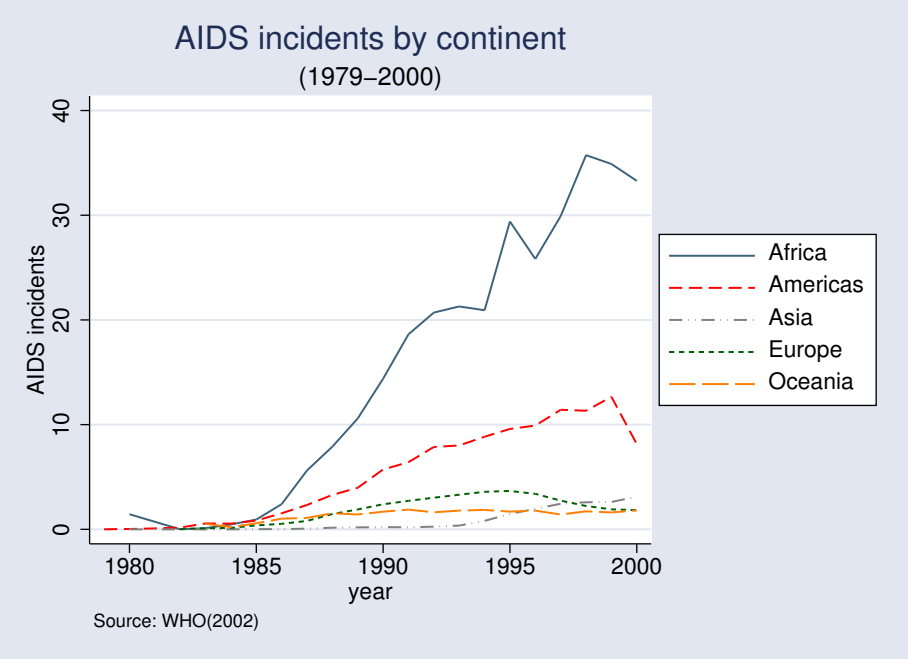

Figure 2.1: Mean AIDS incidents in 5 regions across time

regions. The world mean AIDS incidents is quite large at 9.938 but obviously upward biased by the African subsample.

Figure 2.1 adds a dynamic element to the descriptive statistics of Table 2.1 by illustrating the rate by which the infectious disease spread in each region. Three features stand out in Figure 2.1. First, is the rapid spread of the disease in Africa. This is a concern that is well-documented in the literature and echoed loudly in the public media. Second, is the observed reversal of AIDS spread rate in Africa and Latin America after 1997, and in Europe after 1995. A plausible explanation for this slowdown is that policies and educational programs for promoting AIDS awareness initiated by many local, national and international agencies may have started to pay off. Third, is the recent increase in AIDS incidence in Asia. This is a major concern because AIDS in particular South Asian countries (i.e. Thailand and China) have increased at an alarming rate over the last few years.

Next, we present AIDS incidents for individual countries to highlight the great variation that exists among them. Table 2.2 presents the top and bottom 25 countries in our sample of 116 countries. Among the countries with highest AIDS incidents 20 are located in sub-Saharan Africa. This speaks directly to the major concerns raised by international organizations, such as the World

Table 2.2: Countries with highest and lowest AIDS incidents

Top 25			Bottom 25		
Country	Rank	AIDS incidents Mean (1979-2000)	Country	Rank	AIDS incidents $Mean(1979-2000)$
Namibia*	1	173.043	Bolivia	92	0.217
Congo	2	168.600	Morocco	93	0.207
Botswana	3	57.084	Poland*	94	0.164
Zimbabwe	4	55.472	Fiji*	95	0.162
Lesotho*	5	49.333	S. Arabia*	96	0.158
Malawi	6	40.971	Jordan	97	0.147
Zambia	7	39.767	Algeria	98	0.116
Swaziland*	8	38.525	Yemen*	99	0.109
Burundi	9	27.484	Czech Rep.*	100	0.096
Barbados*	10	26.818	Japan	101	0.095
Tanzania	11	26.060	India	102	0.073
Kenya	12	24.953	Sri Lanka	103	0.047
Gabon*	13	22.013	Philippines	104	0.042
Togo	14	21.910	Turkey	105	0.038
C.African Rep.	15	20.396	Iran*	106	0.037
Uganda	16	19.119	Syria	107	0.036
Rwanda	17	18.540	Korea	108	0.031
Guyana*	18	17.806	Egypt	109	0.029
Thailand	19	17.047	Slovakia	110	0.028
Ghana	20	16.679	Russian Fed.*	111	0.022
Tri.&Tobago	21	15.906	Madagascar	112	0.021
USA	22	14.809	Indonesia	113	0.016
Honduras	23	13.256	Pakistan	114	0.012
Chad	24	12.769	China*	115	0.004
Burkina Faco	25	13.589	Bangladesh	116	0.001

Notes: (*) denotes countries not in our estimation sample.

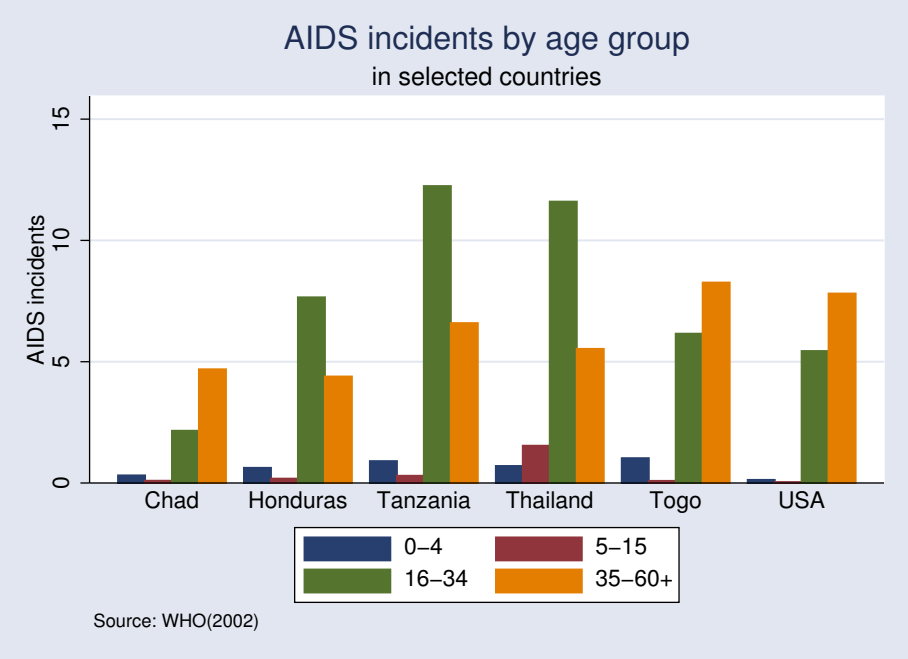

Figure 2.2: AIDS incidents by four age groups in selected countries

Bank, WHO and UN, and governments of advanced nations like the U.K., Germany and the U.S.[11] It is interesting to notice however that the U.S. and Thailand are also part of the top 25 list. This suggests that AIDS may be different from other determinants of economic development that typically are inherently dependent on per worker income. This argument is reinforced by looking at the list with the bottom 25 countries as many developing and less developed countries experience very low AIDS incidents. A notable feature of the low-AIDS-incidence list is that the primary religion in 12 out of the 25 countries is Islam. This is consistent with the hypothesis that religion may be influential to the culture of these countries keeping AIDS incidents very low.

Finally, we take advantage of a nice feature of our dataset and present AIDS incidents by four different age groups for selected countries. This disaggregation reveals that there is significant variability in the way AIDS affects different age groups across countries. For example, Figure 2.2 illustrates that for countries like the U.S., Togo and Chad the most affected age group is 35-60+ whereas for Tanzania, Thailand and Honduras the most affected age group is 16-34. This variability is explored further in our empirical analysis.

[11] For example, during their campaign for the November 2004 U.S. presidential election both president Bush and senator Kerry highlighted AIDS in sub-Saharan Africa as one of the most stressing socioeconomic and humanitarian problems of modern times.

2.1.2 Extending the MRW Dataset

Since our empirical analysis is based on the Solow specification, we have extended the MRW original dataset (PWT version 4.0) until the year 2000 for their non-oil sample. Our data sources are the World Development Indicators (WDI-2002) for working age population growth, Barro and Lee (2001) and Bernanke and Gürkaynak (2001) for human capital, and PWT version 6.1 for the remaining variables. Due to data constraints with variables necessary for our estimation other than AIDS, our sample was reduced from 116 to 89 countries (our sample is reduced further to 81 countries in the panel estimation).

It is important to clarify that for human capital we use the Bernanke and Gürkaynak datatset[12] for our cross-sectional estimation and the Barro and Lee dataset for our panel estimation. We do that because the former dataset offers more observations for our cross-sectional estimation, whereas the latter dataset offers more entries for our panel estimation.

2.2 Estimation and Results

In this section we present our baseline results. First, we present the cross-sectional results for the full sample and arbitrarily chosen subsamples as well as endogenously chosen subsamples.

2.2.1 Cross-Sectional Estimation

Our empirical analysis is based on the extended unrestricted Solow specification in which we consider AIDS as a productivity shock. Specifically, we consider the following regression equation:

$$\ln y_{i,2000} = a_0 + a_1 \ln s_{ik} + a_2 \ln(n_i + g + \delta) + a_3 \ln s_{ih} + a_4 AIDS_i + \varepsilon_i, \tag{2.1}$$

where $y_{i,2000}$ is output per working age person in country i in 2000,[13] s_{ik} is the ratio of average investment to GDP over 1979-2000, s_{ih} is secondary school enrollment of working-age population,

[12] Bernanke and Gürkaynak (2001) follow MRW and obtain their human capital measure by multiplying the fraction of population in the ages of 12-17 that is enrolled in secondary school by the fraction of the working-age population that is of school age (15-19). We average human capital for the period 1970-1995.

[13] Results are insensitive to using output per capita.

Table 2.3: Cross-country regressions for the full sample and OECD and non-OECD subsamples

Dependent variable: ln(GDP per worker in 2000)						
Specification	Extended Solow Model (PWT 6.1)			Extended Solow Model with AIDS ($PWT\ 6.1 - WHO\ 2002$)		
	Non-oil	OECD	Non-OECD	Non-oil	OECD	Non-OECD
Constant	4.7111***	10.2405***	5.9796***	4.8387***	10.0434***	6.1577***
	(0.9751)	(2.0989)	(1.6396)	(0.9673)	(2.0069)	(1.6224)
$\ln s_{ik}$	0.6190***	0.4973	0.5893***	0.6040***	0.5142	0.5732***
	(0.1276)	(0.3342)	(0.1396)	(0.1281)	(0.3173)	(0.1401)
$\ln(n_i + g + \delta)$	−2.7775***	−1.3014*	−2.2274***	−2.7292***	−1.3294**	−2.1595***
	(0.3094)	(0.6683)	(0.6366)	(0.3062)	(0.5799)	(0.6290)
$\ln s_{ih}$	0.6283***	1.2455***	0.6060***	0.6289***	1.2162***	0.6067***
	(0.0789)	(0.3071)	(0.0832)	(0.0755)	(0.2401)	(0.0801)
AIDS				−0.0031*	0.0247	−0.0032*
				(0.0019)	(0.0174)	(0.0020)
Adj. R^2	0.849	0.584	0.724	0.852	0.653	0.731
Obs.	89	21	68	89	21	68

Notes: Standard errors are in parentheses. It is assumed that $g + \delta = 0.05$ as in MRW. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level. Investment and population growth rates are averages for the period 1979-2000. s_h is the average percentage of the working-age population in secondary school for the period 1970-1995.

n_i is average population growth, $g + \delta = 0.05$ as in MRW, AIDS_i is the AIDS incidence per 100,000 people averaged for the period 1979-2000, and ε is an error term.[14]

Table 2.3 presents estimates for the extended Solow model for the period 1979-2000 for the full sample and arbitrarily chosen OECD and non-OECD subsamples using ordinary least squares (OLS). First, we estimate the MRW specification with our extended data. These results are consistent with MRW using data from PWT 4.0 for the period 1960-1985. They are also qualitatively similar to Bernanke and Gürkaynak (2001) who extend the data until 1995, using PWT 6.0. Next we add AIDS as a regressor, therefore treating it as a productivity parameter.

When we reestimate the MRW specification using PWT 6.1 for the full sample of 89 countries, we find that the model explains 84.9% of the overall variation in per worker income (column 2). Adding AIDS into the regression improves Adj. R^2 slightly to 85.2% (column 5). The estimates from the two models have the expected signs, but differ a bit in magnitude. The estimated coefficient for physical capital decreases from 0.6190 in the model without AIDS to 0.6040 in the model with AIDS, keeping the same significance level at 1%. The coefficient for human capital remains almost identical in

[14]Following Gallup and Sachs (2000) and McCarthy, Wolf and Wu (2002), AIDS_i enters the regressions in levels.

magnitude at 0.63 in both models and significant at 1%. The estimated coefficient for $\ln(n_i + g + \delta)$ is -2.7775 in the model without AIDS and increases to -2.7292 in the model with AIDS, remaining highly significant at the 1% level. Most importantly, for our full sample the estimated coefficient on AIDS is negative (-0.0031) and significantly different from zero at the 10% level. This result suggests that each additional AIDS incident per 100,000 people per year is associated with a 0.0031 percentage point reduction in per worker income. This is first evidence that AIDS has a negative impact on cross-country income.

Next, we examine our results by arbitrarily splitting the full sample into OECD and non-OECD countries. In the model without AIDS, for the non-OECD countries, we obtain a positive and highly significant coefficient for $\ln(s_{ik})$, 0.5893, a positive and highly significant coefficient for $\ln(s_{ih})$, 0.6060, and a negative and significant coefficient for $\ln(n_i + g + \delta)$, -2.2274 (column 4). There is little change in the coefficient estimates between the specification with and without AIDS (column 7). What is important to notice is that the coefficient estimate for AIDS remains negative (-0.0032) and significant at the 10% level.

When we compare the coefficient estimates from the models without and with AIDS for the OECD countries (columns 3 and 6, respectively) we find that the coefficient on s_{ik} increases from 0.4973 to 0.5142, but remain insignificant. The coefficient on s_{ih} remains almost identical in terms of magnitude (1.2) and highly significant. The estimated coefficient for $\ln(n_i + g + \delta)$ is -1.3014 and significant at the 10% level in the model without AIDS, and decreases to -1.3294 and significant at the 5% level when we include AIDS. The estimated coefficient for AIDS (quite surprisingly) changes sign but is insignificant, suggesting that the epidemic has no significant impact on the level of income for developed countries.[15]

A possible explanation for this result may be that AIDS in non-OECD countries affects those in their most productive ages who can not afford treatment. More precisely, since people in advanced countries can afford treatment using antiretroviral drugs, this can increase productivity, delay the transmission of the disease, and potentially cause positive externalities by protecting other people.[16]

[15] We have also reestimated all of the specifications in Table 2.3 excluding Botswana, Congo, Malawi, Zimbabwe and Zambia (the countries in our sample with the highest concentration of the epidemic). Results from this exercise appear in Figure D1 and Table D1 in Appendix D. The main result is that when we exclude these countries with highest AIDS incidence, the coefficient estimate for AIDS remains negative and increases in magnitude and significance for the non-OECD subsample.

[16] However, the impact of antiretrovirals on the spread of the epidemic is yet unclear (Kremer (2002)). Advocates

In developing countries, the effect of the pandemic may be different. People cannot afford the expensive drugs and because of the very low level of education, they are not even familiar with the basic protection measure – the use of a condom. Kalemli-Ozcan (2004) provides new evidence on the empirical relationship between the mortality rate changes and the quality-quantity trade-off for a panel of African countries, where parents choose to have more children and provide them with less education facing a high probability of getting infected with AIDS.

2.2.2 Panel Estimation

This section extends our baseline cross-sectional results to consider estimation of the extended Solow equation using panel data techniques. Even thought AIDS data since 1979 exists for some countries in our sample, we consider the period 1985-2000 because for most countries 1985 was the starting year for reporting AIDS incidents. This enables us to evaluate the impact of the epidemic across different countries and over time. Following much of the literature on cross-country panel estimation, we average the data in five-year time intervals; 1985-1990, 1990-1995 and 1995-2000. Due to data constraints our full sample is now reduced to 81 countries with a maximum of three and a minimum of one time observations for each country. Our panel dataset is therefore unbalanced with a total of 238 observations.

Our regression equation is:

$$\ln y_{it} = a_0 + a_1 \ln s_{itk} + a_2 \ln(n_{it} + g + \delta) + a_3 \ln s_{ith} + a_4 AIDS_{it} + \varepsilon_{it}, \tag{2.2}$$

where $\ln y_{it}$ is income per worker and $i = 1, 2, ..., 81$ indexes each country and $t = 0, 1, 2$ indexes time-year periods, s_{itk} is the ratio of average investment to GDP, s_{ith} is investment in human,[17] n_{it} is the average population growth of the working age population, and $g + \delta$ is assumed as previously to be 0.05. As in the cross-country regressions, we add AIDS in the panel regressions.

Table 2.4 presents results from the panel data analysis for the full sample under different spec-

of antiretroviral drugs for HIV/AIDS support the view that the effect of these drugs is expected to lead to prevention and slowdown of transmission. Alternatively, there exists the possibility that due to the availability of such drugs people choose to have more and riskier sexual contacts.

[17] Our measure of human capital is taken from Barro and Lee (2001) and is the percentage of secondary school attained in the total population. We use the Barro and Lee (2001) human capital dataset (instead of the Bernanke-Gürkaynak (2001) dataset) which provides data for five-year periods from 1960-2000 for most (81) of the countries in our sample.

ifications. First we consider the Between Estimator (BE).[18] In a recent paper Hauk and Wacziarg (2004) argue that using an OLS estimator applied to a single cross-section of variables averaged over time (BE) performs best in terms of the extent of bias on each of the estimated coefficients. Consistent with the cross-sectional analysis, the coefficient on AIDS is -0.0050 and significant at the 5% level (column 2). The remaining estimated coefficients for $\ln(s_{itk})$, $\ln(n_{it}+g+\delta)$ and $\ln(s_{ith})$ have the expected signs and are significant at the 1% level.

Table 2.4: Panel regressions

Dependent variable: ln(GDP per worker for 1985-1990, 1990-1995 and 1995-2000)					
Specification	Extended Solow Model with AIDS (PWT 6.1 – WHO 2002)				
	Non-oil with Between Estimator	Non-oil with time effects	Non-oil with time effects & country effects	Non-oil with dOECD & interaction term	Non-oil with time effects, dOECD & interaction term
Constant	6.0352***	6.6639***	7.9758***	8.8015***	8.7002***
	(1.0192)	(0.5649)	(0.3246)	(0.6604)	(0.6330)
$\ln s_{itk}$	0.6714***	0.6524***	−0.0710	0.5462***	0.5592***
	(0.1113)	(0.0664)	(0.0505)	(0.0666)	(0.0638)
$\ln(n_i+g+\delta)$	−1.9045***	−1.5976***	−0.0375	−0.7212***	−0.6996***
	(0.3462)	(0.1920)	(0.1051)	(0.2480)	(0.2374)
$\ln s_{ith}$	0.5218***	0.5318***	−0.3727***	0.5350***	0.5161***
	(0.0795)	(0.0514)	(0.0943)	(0.0498)	(0.0479)
AIDS_{it}	−0.0050**	−0.0045***	−0.0008	−0.0040***	−0.0046***
	(0.0022)	(0.0013)	(0.0006)	(0.0013)	(0.0012)
d91		0.1808**			0.1653**
		(0.0723)			(0.0675)
d96		0.3243***			0.3229***
		(0.0723)			(0.0672)
IT				0.0273**	0.0253**
				(0.0114)	(0.0110)
dOECD				0.4566***	0.4694***
				(0.1120)	(0.1073)
Adj. R^2	0.84	0.81	0.45	0.82	0.84
Obs.	81	238	238	238	238

Notes: d91 and d96 denote time dummies for 1991 and 1996 respectively, IT denotes an interaction term between AIDS and an OECD dummy variable, and dOECD denotes an OECD dummy variable. Standard errors are in parentheses. It is assumed that $g+\delta = 0.05$ as in MRW. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

To allow for the possibility of time effects, we have also estimated the model by adding $(T-1)$

[18] We refer the interested reader to Green (2000, Ch.14, pp. 562-565) for further information on the Between Estimator.

time dummies, where d91 and d96 are dummy indicators for the years 1991 and 1996, respectively.[19] These dummies are meant to captures exogenous shocks specific to each five-year period. The results (column 3) are similar in terms of the magnitude and significance level to those obtained from estimating the model with BE. There is a slight decrease in the magnitude of the AIDS coefficient (-0.0047) but significance increase to the 1% level.

To account for the possibility of country-specific effects as well as time effects, we estimate a two-way fixed-effect specification that involves the addition of 80 country-specific dummy variables and 2 time dummy variables. However, as there are more coefficients to estimate, we lose a large number of degrees of freedom which clearly biases our estimates. This is obvious from the results presented in column 4 as there is a stark change in terms of the magnitude and significance of the coefficient estimates. In particular, the estimate on $\ln(s_{itk})$ becomes insignificant, and the estimate on $\ln(s_{ith})$ changes from positive and significant into negative and significant. The coefficient on AIDS is still negative (-0.0008) but not significantly different from zero. We believe that these radical changes in the estimates is due to the substantial loss of degrees of freedom. In addition, as Griliches and Hausman (1986) note, in regressions using panel data with fixed effects specifications, measurement error in the explanatory variables can lead to coefficient estimates that are "too low" and therefore insignificant; in controlling for the various fixed effects, the relative importance of measurement errors in the explanatory variables becomes greatly exacerbated, biasing coefficient estimates.

In order to allow for the effect of AIDS to differ among OECD and non-OECD countries, we add an interaction term (IT) between AIDS and an OECD dummy variable (column 5). All of the estimates are significant and have the expected signs. In particular the key coefficient estimate for AIDS is -0.0040 and is significant at the 5% level which corresponds with our cross-sectional results. Finally, in addition to the interaction term, we include time specific dummies (d91 and d96) to allow for the effect of AIDS to differ across time (column 6). The coefficient estimate for AIDS continues to be negative (-0.0046) but is now significant only at the 10% level, whereas the IT coefficient estimate is positive and significant at the 5% level and the dummy for OECD is positive and significant at the 1% level.

[19] In order to avoid perfect collinearity we drop the dummy variable on the first five-year period.

In summary, our panel estimation is generally supportive of our cross-sectional results. In particular, with the exception of the model with fixed and time effects the impact of AIDS on income obtained from the panel estimation is shown to be negative and similar in magnitude to that obtained from our cross-sectional estimation.

2.3 Robustness

This section examines the robustness of our baseline results to alternative subsamples of AIDS incidents by age group, and panel estimation that consider the problem of endogeneity.

2.3.1 AIDS by Age Groups

In addition to obtaining data on annual AIDS incidents, we were also able to assemble data on the officially reported AIDS incidents for the period of study on different age groups. In particular we were able to disaggregate our original AIDS dataset into four age-group samples as follows: AIDS[0-4] (*infancy period*), AIDS[5-15] (*schooling period*), AIDS[16-34] (*productive period*) and AIDS[35-60+] (*less productive period*). Due to data constraints our original sample was reduced from 89 to 63 countries.[20]

Some interesting observations become apparent from exploiting this dimension of our data. Two of the four groups, AIDS[16-34] and AIDS[35-60+], are affected most by the disease. More precisely, the most affected group in Africa is AIDS[16-34] which can have disastrous economic consequences since it affects people in their most productive stage of their lives. The same occurs in Europe and Latin American countries like Argentina, Brazil and Mexico. Interestingly, and in contrast to most countries, in the US the most affected group is AIDS[35-60+].

Due to the high correlation between AIDS[0-4] and AIDS[16-34], 0.825, and AIDS[0-4] and AIDS[16-34], 0.812, we decided to exclude AIDS[0-4] from our regression to reduce the possibility of multicolinearity.[21] Table 2.5 presents regression results using AIDS incidents by the three age groups. The estimates on $\ln(s_{ik})$, $\ln(s_{ih})$ and $\ln(n_i + g + \delta)$, are all significant at the 1% level of significance with the expected sign. The main result from this exercise is that only the coefficient

[20]These countries are marked with an asterisk in Table A1 in Appendix A. A detailed explanation of how we construct AIDS incidence by age group appears in Appendix C.

[21]This high correlation is present because infants till the age of 4 are infected almost exclusively by their parents who are HIV positive or they are already infected by AIDS.

Table 2.5: Cross-country regression using AIDS by age group

Dep. var.: ln(GDP per worker in 2000)	
Specification	AIDS by age group
Constant	5.2621^{***}
	(1.0457)
$\ln s_{ik}$	0.7231^{***}
	(0.1461)
$\ln(n_i + g + \delta)$	-2.5612^{***}
	(0.3184)
$\ln s_{ih}$	0.4986^{***}
	(0.0880)
AIDS[5-15]	-0.0230
	(0.2010)
AIDS[16-34]	-0.0961^{*}
	(0.0030)
AIDS[35-60+]	0.0584
	(0.0782)
Adj. R^2	0.85
Obs.	63

Notes: Standard errors are in parentheses. It is assumed that $g + \delta = 0.05$ as in MRW. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

on AIDS[16-34] is significant (albeit marginally at the 7% level) with a negative sign. It is also important to notice that the magnitude of the AIDS[16-34] coefficient estimate (-0.0961) has more than doubled compared to respective cross-sectional estimate. This finding is quite intriguing as it promotes the idea that the negative impact of AIDS on income is primarily due to arguably the most productive age group, AIDS[16-34], being infected by AIDS.

2.3.2 Panel-IV Estimation

Our regression model is potentially subject to the well-known endogeneity problem. A common way to correct the endogeneity problem in much of the existing literature is to use instrumental variables. However, as Islam (1995) and many subsequent papers have pointed out, it is difficult to come up with a set of "good" instruments that will be correlated with the potentially endogenous variable (in our case AIDS) but not correlated with other regressors.[22] An alternative solution to

[22]Nevertheless, we have considered instrumenting AIDS with initial AIDS in our cross-sectional analysis. However, since initial AIDS is very likely measured with very large errors (especially due to under-reporting), this can substantially bias our estimates toward zero.

the endogeneity problem is the use of panel data and in particular the use of lags of the right-hand side variables as predetermined or weakly exogenous instruments in panel-data regressions.[23,24]

In this section we extend our panel data results presented in the previous section by using instrumental variables (a panel-IV approach) to correct for the potential endogeneity of AIDS. In particular, we use the first lag of AIDS and schooling (s_{ih}) as instrumental variables for AIDS. We use schooling because there are empirical and theoretical grounds to expect that past values of human capital play an important role in explaining the effect of AIDS on economic performance.[25] The downside of this analysis is that our sample is reduced from 238 to 157 observations.

To examine the validity of our instruments we test the overidentifying restrictions for every regression specification considered in our panel-IV estimation. Results are presented in Panels A and B in Table 2.6. For the specifications in column 2 and 3, the endogenous variable, AIDS, is explained with two instruments; the first lag of AIDS and the first lag of schooling. This results in one over-identifying restriction. For the next two specifications, presented in columns 4 and 5, in addition to AIDS we allow for another potentially endogenous variable; the interaction term between AIDS and a dummy variable for OECD (IT). Therefore, as suggested by Woolridge (2002), we include in our set of instruments an interaction term between a dummy variable for OECD and the first lag of AIDS.[26] This, once again, results in one over-identifying restriction.

The first row of Panel B in Table 2.6 reports the p-values from χ^2 Sargan's (1958) test. This is a test of the joint null hypothesis that the excluded instruments are valid instruments. A rejection casts doubt on the validity of the instruments. In all the specifications considered we fail to reject the null of no correlation between the instruments and the error term, indicating that our over-identifying instruments are satisfactory. In the bottom row of Panel B in Table 2.6 we use the Hausman test to determine whether AIDS should be treated as exogenous or endogenous. In two of the specifications, with dOECD and interaction term (column 4), and with dOECD, interaction

[23]The first paper that examined cross-country regressions adjusting for both the fixed-effects problem as well as for the endogeneity problem is Caselli et al. (1996).

[24]Despite these advantages, panel data with instrumental variable techniques have also been criticized for obtaining estimates that are quite biased. For further discussion on these issues see Durlauf and Quah (1999), and Hauk and Wacziarg (2004).

[25]See e.g., Corrigan, Glomm and Mendez (forthcoming), and Kalemli-Ozcan (2004).

[26]We thank Carter Hill who suggested to us this instrument. Ressler et al. (2002) use a similar instrument in an attempt to test their hypothesis of a positive relationship between the size of welfare payments per recipient and the heterosexual HIV infection rate in the United States.

Table 2.6: Panel with Instrumental-Variable regressions

IV Regressions of ln(GDP per worker for 1985-1990, 1990-1995 and 1995-2000)				
Specification	Panel A: Two Stage Least Squares			
	Non-oil with time effects	Non-oil with time effects & country effects	Non-oil with dOECD & interaction term	Non-oil with time effects, dOECD & interaction term
Constant	7.1240***	10.4529***	9.3990***	9.3595***
	(0.6595)	(3.4310)	(0.8114)	(0.8043)
$\ln s_{itk}$	0.5965***	−0.2999	0.5206***	0.5202**
	(0.0853)	(0.2956)	(0.0843)	(0.0836)
$\ln(n_{it} + g + \delta)$	−1.4968 ***	−0.2396	−0.5341*	−0.5212*
	(0.2174)	(0.5187)	(0.3016)	(0.2989)
$\ln s_{ith}$	0.5862*	1.6255	0.5575***	0.5556***
	(0.0696)	(2.8414)	(0.0676)	(0.0670)
AIDS_{it}	−0.0081***	−0.0333	−0.0083***	−0.0088***
	(0.0035)	(0.0482)	(0.0019)	(0.0019)
d96	0.1359*	0.1580***		0.1543***
	(0.0771)	(0.0537)		(0.0715)
IT			0.0404**	0.0363**
			(0.0173)	(0.0168)
dOECD			0.4423***	0.4599***
			(0.1495)	(0.1476)
Adj. R^2	0.79	0.51	0.82	0.82
Obs.	157	157	157	157
	Panel B: Specification Tests (p-values)			
Overidentifying Restrictions	0.304	0.875	0.462	0.788
Hausman Test	0.177	0.993	0.064	0.052

Notes: d96 denotes a time dummy for 1996, IT denotes an interaction term between AIDS and an OECD dummy variable, and dOECD denotes an OECD dummy variable. Standard errors are in parentheses. It is assumed that $g + \delta = 0.05$ as in MRW. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

term and time effects, we are able to reject the null at the 10% level of significance that AIDS and the potentially endogenous interaction term are correlated with the error term. This implies that we can apply Two Stage Least Squares (2SLS) and correct for endogeneity. For the specifications in columns 2 and 3 we are not able to reject the null.

To evaluate the quality of our instruments, we further test their validity by estimating reduced form regressions of AIDS on the explanatory instrumental variables. Subsequently we test the joint significance of the coefficients on the instruments in each of our specifications. In all the regressions, we reject the null hypothesis of zero coefficients at the 1% level of significance. This shows that our instruments provide useful information in addition to that provided by the explanatory variables.

The panel-IV results are presented in Panel A of Table 2.6. In all specifications, the coefficients on $\ln(s_{itk}), \ln(s_{ith})$ and $\ln(n_{it} + g + \delta)$, as well as the 1996 dummy variable (d96), the interaction term (IT), and the OECD dummy variable (dOECD) are qualitatively similar to those obtained in the panel estimation without instrumental variables. With the exception of the model with country and time specific effects (column 3), the coefficient estimates for AIDS are negative and highly significant, and in fact larger in magnitude than previous results. Therefore these results provide evidence suggesting that our baseline results are robust to correcting for potential endogeneity.

2.3.3 Endogenous Sample Splitting

Following the emerging literature on *parameter heterogeneity* in cross-country regressions we are able to examine whether AIDS is a threshold variable.[27] In particular, we employ Hansen's (2000) splitting methodology and allow the data to endogenously select regimes using AIDS as a potential threshold variable.[28] The advantage of Hansen's methodology over the regression-tree methodology used in Durlauf and Johnson (1995) is that it is based on an asymptotic distribution theory. Our threshold estimation uses the Solow level regression equation (2.1).[29]

In the first round of splitting the bootstrap p-value was 0.008, implying that there may be a

[27]Papers in this literature include, Durlauf and Johnson (1995), Liu and Stengos (1999), Durlauf, Kourtellos and Minkin (2001), Kalaitzidakis et al. (2001), and Masanjala and Papageorgiou (2004), just to name a few. For a more comprehensive discussion on parameter heterogeneity see Durlauf and Quah (1999, Vol. 1, Ch. 4), and Durlauf, Johnson and Temple (forthcoming, Part II, Ch. 7), and references therein.

[28]We use average AIDS (1979-2000) rather than initial AIDS because we expect initial AIDS data to be much more prone to measurement error than subsequent periods.

[29]The GAUSS programs used for threshold estimation are available from the authors upon request.

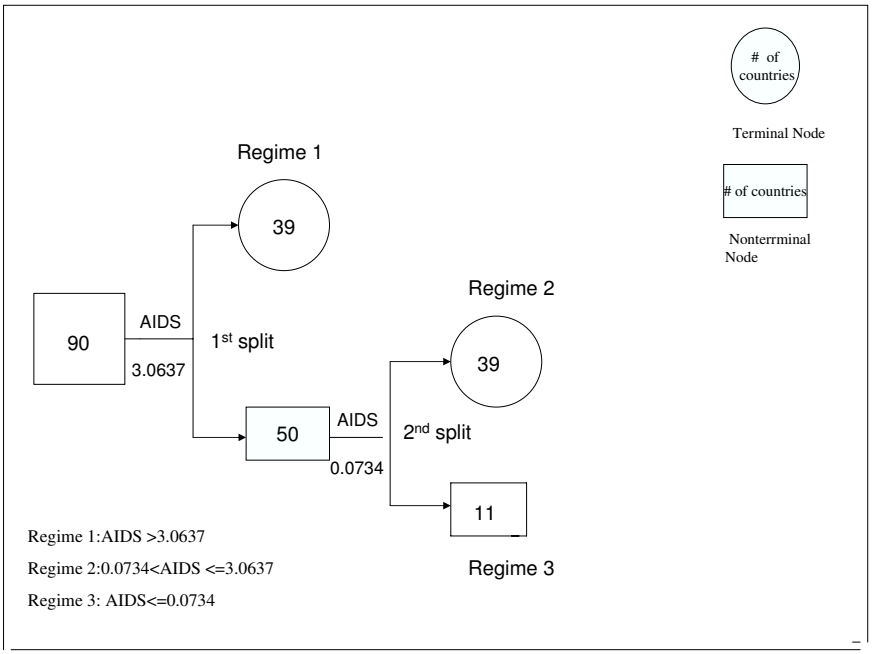

Figure 2.3: Regression tree diagram

sample split based on AIDS. The threshold estimate was $\gamma = 3.0637$ with asymptotic 95% confidence set $[0.0734, 7.4395]$. AIDS as a threshold divided the full sample (89 countries) into two subsamples: one, containing 50 countries (AIDS ≤ 3.0637) and another, with 39 countries (AIDS > 3.0637).

We tried to further split the group with the higher AIDS incidence (AIDS > 3.0637), but the bootstrap test statistic was insignificant. However, the bootstrap test statistic for the sample with 50 countries with AIDS ≤ 3.0637 was significant (0.035), showing a possible sample split. More precisely, $\gamma = 0.0734$ and the confidence set is $[0.0360, 0.4024]$. This implies that AIDS further splits our subsample into two additional regimes: one, with 11 countries (AIDS ≤ 0.0734) and another, with 39 countries (AIDS > 0.0734). No more splits were possible using the new regimes as we obtained bootstrap test statistics that were insignificant.

Figure 2.3 presents a regression tree diagram that illustrates these results. Non-terminal nodes are illustrated by squares whereas terminal nodes are illustrated by circles. The numbers inside the squares and circles show the number of countries in each node. The point estimates for the threshold variable are presented on the rays connecting the nodes. Table E1 in Appendix E presents the countries included in each of the three regimes.

In general we interpret our threshold estimation results as further evidence of parameter heterogeneity; countries can be grouped according to different statistical models. More importantly, we have shown evidence supporting the idea that AIDS is a threshold variable.

2.4 Discussion

Summary of results: Our results can be summarized as follows: a) When using the full sample of 89 countries we find a negative and statistically significant effect of AIDS on cross-country per worker income. b) When we arbitrarily split our entire sample into OECD and non-OECD subsamples the negative relationship continues to exist when using the non-OECD subsample but vanishes in the OECD subsample. c) When using AIDS incidents by age group we find that there exists quantifiable negative impact of AIDS on income only for people in the ages 16-34. d) Panel estimation results (without or with instrumental variables) are consistent with those obtained in the cross-sectional analysis. e) Using Hansen (2000) we also find that AIDS is a threshold variable that can split our full sample into four regimes obeying different statistical models.

Interpretation of results: Beyond the negative impact of AIDS on income that emerges from our estimation results it is important to examine the magnitude of this impact. It works out that the coefficient estimates for AIDS from various alternative estimation specifications (cross-sectional, panel) and samples (full, non-OECD) are surprisingly quite stable at around -0.003 to -0.004. This implies that for the period 1979-2000 each additional AIDS incident per 100,000 people per year was associated with a 0.003 to 0.004 percentage point reduction income per worker income. Using the most conservative AIDS estimate of -0.003 we are able to back out "lower bound" cost estimates for the epidemic. Table 2.7 reports total cost to GDP ratio, cost per worker, cost per capita, and cost per new case in year 2000 for nine non-OECD countries grouped in three categories by AIDS severity. As expected the total cost to GDP ratio varies with the epidemic's severity across countries (column 4). In particular, total cost to GDP ratio was 0.23% for Botswana with the second highest incidence rate in our sample, whereas the same ratio was only 0.0001% for South Korea. Cost per worker and cost per capita (columns 5-6) indicate the difference in individual welfare loss in countries with a range of AIDS incidence. Finally, the last column reports estimates of the cost per case in selected countries. Cost per case calculated using our estimates increases with AIDS incidence but

also with per capita income. For example even though AIDS incidence is much lower in Hong Kong ($0.0494/100,000$) than in Botswana ($57.084/100,000$), the cost per case is more than three times higher in the former than the later country. Of course, thinking about these estimates in relation to individual welfare would be the appropriate metric for this exercise. Overall, these calculations show that the impact of AIDS vary dramatically across countries in our sample and can have devastating effects especially in those countries with high incidence but low per capita income.

Reconciling our results with those of Bloom and Mahal (1997): In an influential paper Bloom and Mahal (1997) reach the conclusion that "... there is more flash than substance to the claim that AIDS impedes national economic growth." A criticism of this paper is that given the scarcity of the data used (authors use *estimated* AIDS cases for 51 countries for the period 1980-1992) it is too early to tell what the impact of AIDS on growth may be. In addition to the problem of data scarcity, it is the problem of quality of early data on HIV/AIDS which forced the authors to resort to estimates of AIDS cases using epidemiological models. Even though measurement errors associated with HIV/AIDS data are likely to be large primarily due to lack of adequate reporting, early on these errors are very likely to be significantly larger. Given the severe criticism of this paper in the literature and public media we decided to reexamine Bloom and Mahal's result using our data and model specification. More precisely, in addition to the level regressions, we examine the effect of AIDS on growth of GDP per worker for the period 1979-2000. We present the results of this exercise in Table F1 in Appendix F. It is shown that standard growth regressors ($\ln y_{i0}$, $\ln(s_{ik})$, $\ln(s_{ih})$ and $\ln(n_i + g + \delta)$) in the alternative samples and specifications considered are consistent with those obtained in other growth regressions commonly found in the literature. When we include AIDS in the regressions, the AIDS coefficients are found *not* to be significantly different from zero for the full and non-OECD samples. For the OECD sample the coefficient is positive and significant which may indicates an endogeneity problem being present. In general, these results suggest that AIDS has an insignificant impact on cross-country growth and therefore are supportive of the evidence and main conclusion in Bloom and Mahal (1997). This then leads us to the key question: How can it be that the Bloom and Mahal results hold, indicating an insignificant impact of AIDS on growth, yet in our host of level regressions AIDS is robustly negatively related with income? The difference in the two results comes down to the central question asked; on the one hand, we are interested in the effect

of AIDS on income, thinking that income is a good proxy for welfare. On the other hand Bloom and Mahal were interested in the effect of AIDS on growth, thinking that growth is a good proxy for the development process. Our analysis suggests that the only criticism that Bloom and Mahal (1997) may be subject to is that by using per worker income growth as the dependent variable the potential effect of AIDS on aggregate output may be masked (see Hall and Jones (1999, p.85)).

Limitations: Our work is certainly not without limitations. Even though one can point to other caveats we want to focus on limitations due to quality and quantity of our AIDS dataset. We recognize that the quality of the UNAIDS/WHO data is questionable on the grounds of cross-country comparability, variable under-reporting and other methodological issues relating to data collection and the definition of AIDS. In addition, we admit that AIDS epidemic is still a transitory phenomenon and therefore as more data become available we will be in a better position to reach more definite conclusions about its effect on income.

Table 2.7: Cost of AIDS in selected countries

Country	AIDS	GDP/worker	Cost/GDP	Cost/worker	Cost/capita	Cost/case
	Severe					
Botswana	57.084	$14,769.7	0.22834%	$33.73	$18.57	32,530.70
Thailand	17.0473	$9,858.	0.06819%	$6.72	$4.58	26,861.10
Honduras	13.256	$3,947.2	0.05302%	$2.09	$1.15	8,657.80
	Medium					
Nigeria	3.148	$1,592.5	0.01259%	$0.20	$0.10	3,305.80
Venezuela	2.647	$11,757.8	0.01059%	$1.25	$0.77	28,930.10
Hong Kong	0.49	$38,179.1	0.00198%	$0.75	$0.55	111,570.50
	Low					
Bolivia	0.217	$5,205.1	0.00087%	$0.05	$0.03	11,735.50
India	0.073	$4,360.6	0.00029%	$0.01	$0.01	10,734.40
Korea	0.031	$20,719.5	0.00012%	$0.03	$0.02	59,747.10

2.5 Conclusion

In this essay, we investigate the impact of AIDS on cross-country income levels. Contrary to previous work on AIDS, we make use of the officially reported AIDS incidents from UNAIDS/WHO on 89 countries for the period 1979-2000, during which the AIDS epidemic has spread across the world.

Using the extended Solow model as the basis of our empirical analysis we first show that in the full sample and non-OECD subsample, the coefficient estimate for AIDS is negative and marginally significant. For the OECD countries, we obtain an insignificant coefficient estimate, which implies that AIDS has no quantifiable effect on the income level for these countries. We also utilize the time dimension of our data and employ panel-data techniques on the extended Solow model with AIDS as a regressor. AIDS enters negative and highly significant in all of the specifications considered except from the specification with country and time effects, where the estimate is insignificant.

Regression analysis using AIDS by age group reveals that only the coefficient on AIDS between the ages 16-34 is significant with a negative sign. In addition, the magnitude of the AIDS[16-34] coefficient estimate has more than doubled compared to that obtained when using the aggregated AIDS data. Finally, we employ Hansen's (2000) threshold methodology that attempts to endogenously split countries in different regimes. This methodology successfully identifies AIDS as a threshold variable. An extensive robustness analysis establishes robustness of our baseline results to various alternative specifications and subsamples.

Obviously, we do not claim to have the last word on the effect of the AIDS epidemic on income but merely to have shed new light on the effects of an unraveling epidemic.

Chapter 3

Female Human Capital Inequality, Infant Mortality and Growth

In recent years we have witnessed the emergence of a vast literature trying to explain the unequal distribution of income among countries.[1] Due to the growing consensus that improving health can have direct and indirect payoffs in terms of better and longer lives leading to higher economic growth, economists have recently turned to health for an answer.[2] This paper aims to contribute to this effort by examining the relationship between female human capital inequality, infant mortality and economic growth. More specifically, we empirically test the following two-step relationship: In the first step, inequality in education among women leads to higher infant mortality. We conjecture that higher inequality in education among women is responsible for higher infant mortality because mothers fail to provide adequate care of their infants. In the second step, higher infant mortality is partly responsible for low growth and development experienced by many developing countries.

To test these hypotheses we use a novel dataset. Following Castelló and Domenech (2002) we develop gender-specific Gini coefficients as a measure of human capital inequality. Due to the lack of available data on human capital inequality, little attention has been devoted to the influence of human capital distribution on economic growth in empirical studies.[3] In our analysis, instead of using the level of education, we use the Gini coefficient which is a better measure of the distribution

[1]The literature starts with theories of neoclassical growth (Solow, 1956; Swan, 1956), reemerges with theories of endogenous growth (i.e. Lucas, 1988, and Romer, 1991) and continues with theories of convergence (i.e. Mankiw, Romer and Weil, 1992; Barro and Sala-i-Martin, 1992; Caselli, Esquivel and Lefort, 1996).

[2]For example, Bloom and Sachs (1998) argue that the growth rate in sub-Saharan Africa could have been higher by 2% in the absence of malaria incidence in the continent. Although malaria infects more than 300 million people annually and is used as an explanation about the continent's poverty, it is expected that AIDS will soon surpass malaria as the most deadly infectious disease (see, Stoytcheva, Chapter 2).

[3]Some exceptions are Birdsall and Londoño (1997) and López, Thomas and Wang (1998). The first study analyzes a sample of 43 countries and uses the standard deviation of years of education as the measure of human capital inequality. The second study uses a wider range of human capital inequality indicators but focuses on a reduced number of 12 Asian and Latin American countries.

of education and allows us to capture not just the mean, but also the different quintiles of the distribution. This approach allows us to include the least educated women and draw conclusions about the potential impact of female human capital inequality on infant mortality. To the best of our knowledge, we are the first to construct female and male human capital inequality measures at the aggregate level using the Barro and Lee (2001) dataset.

In 1991, less than 40% of the 330 million women in India aged 7 and over were literate, which means today there is an estimate of over 200 million illiterate women in India.[4] This low level of literacy not only has a negative impact on women's welfare but also on their families, their cities and beyond. In 1971, only 22% of women and 46% of men were literate worldwide. By 1991, 39% of women and 64% of men were literate. Thus, there has been a large increase in the proportion of women who are literate in just 20 years. Despite these improvements, the large gap between the literacy levels of men and women continues to be significant. A very small proportion of both men and women have a college education, just over 3% of men and 1% of women. Finally, there are dramatic differences in literacy rates by place of residence, with rates in rural areas lagging behind rates in urban areas. In 1991, the urban literacy rate was more than twice that of the rural rate, 64% and 31%, respectively.

There is a small, but rapidly growing macro literature explaining the decline in infant and child mortality over the last few decades.[5] This literature suggests that women's education is the most significant determinant of infant and child mortality (see, i.e. Schultz, 1993). At the sample mean, a one-year increase in women's education is associated with a 5-percent decline in child mortality. Mother's schooling is considered to be the most important determinant, presumably because she manages child-care and administers the child's food and medical care.

Technological progress also is considered to be one of the key reasons that explains the mortality decline as argued by Jamison, Sandbu and Wang (2001). Countries may differ in how close their health systems come to utilizing the technology available at any given time. Their analysis relaxes the assumption that technological progress is constant across countries. Mortality affects investment

[4]These data are taken from the U.S. Department of Commerce, Economics and Statistics Administration, 1998.

[5]There is also substantial evidence from micro-development studies that supports female education as a positive tool in lowering infant mortality. For example, Breierova and Duflo (2002), taking advantage of a massive school construction program that took place in Indonesia between 1973 and 1978, show that female education is a stronger determinant of age at marriage and early fertility than male education.

through rates of return because mortality considerations affect schooling decisions and those countries that differ only in health capital do not converge to similar living standards−threshold effects may occur (see, i.e. Chakraborty, 2003).

In a recent paper Lorentzen, McMillan and Wacziarg (2004) argue that poverty leads to high mortality, which leads to low growth. The paper's main argument is that people who expect to die young will fail to take actions−saving and getting educated−that generate long-term benefits for short term costs. The main focus is on adult mortality, rather than the commonly used variables of infant mortality or life expectancy at birth. It is argued that adult mortality explains almost all of the Africa's growth tragedy over the past forty years.

Finally, in an important contribution, Waldmann (1992) reached the conclusion that comparing two countries in which the poor have equal real incomes, the one in which the rich are wealthier is likely to have a higher infant mortality rate. What is more surprising is that infant mortality appears to be positively related to the income share of the rich, defined as the upper 5 percent of the income distribution. This result suggests that the measured real incomes may be a poor measure of social welfare.

The rest of the essay is organized as follows: Section 3.2 takes a closer look at the data and emphasizes the novelty of female and male human capital inequality measures. Section 3.3 presents cross-sectional estimation of our two main equations and addresses the endogeneity issue. Section 3.4 presents robustness analysis of the cross-sectional results and considers panel-data estimation for the full sample. Section 3.5 concludes.

3.1 A First Look at the Data

We begin by describing the data used in our estimation. First, we present the variables, used in our cross-sectional and panel estimation. Next, we explain how we have constructed the measures of female and male human capital inequality using the Barro and Lee (2001) dataset. Third, we present descriptive statistics of the Gini coefficient of male and female human capital for different geographic regions.

3.1.1 Data

In this essay we estimate two equations: one, where the dependent variable is infant mortality rate,[6] averaged for the period 1960-2000 (data is taken from the World Development Indicators, 2002), and a second growth equation that is derived from augmenting the Solow growth model. For the baseline estimation of our first equation, we include in addition to the Female Gini coefficient, the number of physicians per $1,000$ people (WDI) and malaria in 1966, which is the percentage of country area with malaria (John L. Gallup, Andrew D. Mellinger, and Jeffrey D. Sachs' geography dataset) as regressors. We also include a measure of the level of human capital in addition to the human capital inequality measure; this is the average schooling years in the female population (Barro and Lee, 2001), fifteen years and above, averaged over 1960-2000.[7]

For the baseline estimation of our second equation, we extend the Mankiw, Romer and Weil (1992) (MRW hereafter) original dataset (Penn World Table version 4.0) until the year 2000, using PWT version 6.1 for the non-oil sample consisting of 98 countries. Due to constraints with the human capital data, our sample size was reduced to 73 countries. Data on real gross domestic product (RGDP) per capita are from the PWT (6.1). We average the population growth of the working-age population n for the period 1960-2000 and add $g + \delta$, which is assumed to be 0.05. Following MRW, the saving rate s_k is the ratio of average investment to GDP over the 1960-2000 period (PWT 6.1). We add a variable called *Human* to proxy for s_h that measures the percentage of the working-age population that is in secondary school and is taken from Barro and Lee (2001). For our panel regressions, we average the data into five-year time intervals. For the growth regressions we include initial GDP; this is GDP per worker in 1960 in the cross-sectional analysis and at GDP per worker at the beginning of each five-years period in the panel estimation.

In examining the robustness of our baseline results we have considered a set of carefully selected explanatory variables. More specifically, in the robustness analysis of our first (infant mortality-female human capital inequality) equation we considered Gini Male, Tropics, Latitude, Gini Income and Public Health Expenditure. In the robustness of our second (growth-infant mortality) equation

[6] Infant mortality rate is the number of infants dying before reaching one year of age per 1,000 live births in a given year.

[7] We present values of the relevant variables in Table A1, Appendix A.

we considered Government Expenditure, and three regional dummies (Latin America, Asia, Africa). We briefly discuss these variables in the robustness section.

3.1.2 Measuring Female and Male Human Capital Inequality

In our estimation we include a new variable−female human capital inequality. We follow Castelló and Domenech (2002) and construct the Gini coefficient of male and female human capital inequality for 108 countries, using the Barro and Lee (2001) dataset. There are different ways of computing the Gini coefficient. The Gini coefficient is a summary statistic of the Lorenz curve and a measure of inequality in a population, calculated as the mean of the difference between every possible pair of individuals, divided by the mean size μ,

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \mu}.$$

Since the Barro and Lee dataset provides information of the average years and attainment levels, the human capital coefficient (G^h) can be computed as follows:

$$G^h = \frac{1}{2\overline{H}} \sum_{i=0}^{3} \sum_{j=0}^{3} |\widehat{x_i} - \widehat{x_j}| \, n_i n_j, \tag{3.1}$$

where $\overline{H}$ are the average schooling years of the population aged 15 years and over, i and j are the different levels of education, n_i and n_j are the shares of population with a given level of education, and $\widehat{x_i}$ and $\widehat{x_j}$ are the cumulative average schooling years of each educational level. Castelló and Domenech consider the four levels of education used in Barro and Lee (2001): no schooling (0), primary (1), secondary (2) and higher education (3). Defining x_i as the average schooling years of each educational level i, the cumulative average schooling years of each level can be written as

$$\widehat{x_0} \equiv x_0 = 0, \quad \widehat{x_1} \equiv x_1, \quad \widehat{x_2} \equiv x_1 + x_2, \quad \widehat{x_3} \equiv x_1 + x_2 + x_3 \tag{3.2}$$

Substituting equation (3.1) in equation (3.2), we obtain for the Gini coefficient the following:[8]

[8] For more details, refer to Castello and Domenech (2001, pp. C189-C190).

$$G^h = n_0 + \frac{n_1 x_2 (n_2 + n_3) + n_3 x_3 (n_1 + n_2)}{n_1 x_1 + n_2 (x_1 + x_2) + n_3 (x_1 + x_2 + x_3)}. \tag{3.3}$$

The Barro and Lee dataset provides the estimates for two different age groups - over age 15 and over age 25 - and a breakdown by sex at five-year intervals for the years 1960-2000. This allows us to compute the Gini coefficient for the two sexes over age 15.[9] Using equations (3.2) and (3.3), the Gini Female can be computed in the following way:

$$G^h = n_0^f + \frac{n_1^f x_2^f (n_2^f + n_3^f) + n_3^f x_3^f (n_1^f + n_2^f)}{n_1^f x_1^f + n_2^f (x_1^f + x_2^f) + n_3^f (x_1^f + x_2^f + x_3^f)}, \tag{3.4}$$

where $n_0 = luf15$, $n_1 = lpf15, n_2 = lsf15$, $n_3 = lhf15$, $\overline{H} = tyrf15$, $x_0 = 0$, $x_1 = pyrf15/(lpf15 + lsf15 + lhf15)$, $x_2 = syrf15/(lsf15 + lhf15)$ and $x_3 = hyrf15/lh15$.[10]

Similarly, the Gini Male can be computed in the following way:

$$G^h = n_0^m + \frac{n_1^m x_2^m (n_2^m + n_3^m) + n_3^m x_3^m (n_1^m + n_2^m)}{n_1^m x_1^m + n_2^m (x_1^m + x_2^m) + n_3^m (x_1^m + x_2^m + x_3^m)}, \tag{3.5}$$

where $n_0 = lum15, n_1 = lpm15, n_2 = lsm15, n_3 = lhm15$, $\overline{H} = tyrm15$, $x_0 = 0, x_1 = pyrm15/(lpm15 + lsm15 + lhm15)$, $x_2 = syrm15/(lsm15 + lhm15)$ and $x_3 = hyrm15/lhm15$.[11]

Next, we present descriptive statistics of the Gini Female and Male. Table 3.1 presents the mean, the standard deviation and the minimum and maximum of the two measures on human capital inequality for six geographic regions and the world.[12] We average the Gini Female and Male for the period 1960-2000.

We notice that the two regions with the highest female and male human capital inequality are Middle East & North Africa and sub-Saharan Africa. Factors like religion, culture, institutions have

[9] Since our sample is composed mainly from developing countries, we consider and construct Gini coefficient for individuals over age 15.

[10] We follow the Barro-Lee dataset: LUF is the percentage of "no schooling" in the female population; LPF is the percentage of "primary school attained" in the female population; LSF is the percentage of "secondary school attained" in female population; LHF is the percentage of "higher school attained" in female population; TYRF is the average schooling years in the female population; PYRF is the average years of primary schooling in the female population; SYRF is the average years of secondary schooling in the female population.

[11] We follow the notation from the Barro and Lee dataset: LUM is the percentage of "no schooling" in the male population; LPM is the percentage of "primary school attained" in the male pop.; LSM is the percentage of "secondary school attained" in male pop.; LHM is the percentage of "higher school attained" in male population; TYRM is the average schooling years in the male population; PYRM is the average years of; primary schooling in the male population; SYRM is the average years of secondary schooling in the male population.

[12] The classification is taken from the WDI (2002).

Table 3.1: Descriptive statistics

Geographic Regions		Mean	Stand. Dev.	Min.	Max.
East, South Asia & Pacific	Gini Female	0.533	0.252	0.161	0.961
	Gini Male	0.432	0.206	0.179	0.808
Europe	Gini Female	0.268	0.113	0.146	0.620
	Gini Male	0.251	0.077	0.136	0.441
Latin America & Caribbean	Gini Female	0.410	0.151	0.219	0.818
	Gini Male	0.380	0.123	0.198	0.651
Middle East & North Africa	Gini Female	0.661	0.148	0.295	0.798
	Gini Male	0.540	0.128	0.257	0.690
North America	Gini Female	0.256	0.042	0.226	0.286
	Gini Male	0.286	0.026	0.268	0.305
Sub-Saharan Africa	Gini Female	0.683	0.182	0.278	0.941
	Gini Male	0.576	0.156	0.290	0.877
World	**Gini Female**	**0.507**	**0.238**	**0.146**	**0.961**
	Gini Male	**0.431**	**0.187**	**0.136**	**0.877**

Notes: The mean, the standard deviation, the minimum and the maximum values presented above are computed for 19 countries in East, South Asia&Pacific, 22 countries in Europe, 23 countries in Latin America&Caribbean, 9 countries in Middle East&North Africa, 2 countries in North America, 29 countries in sub-Saharan Africa-see Appendix B for countries in geographic regions.

definitely influenced female education over the years. The mean for Gini Female is the highest in sub-Saharan Africa, 0.683, while the Gini Male is 0.576. Sub-Saharan Africa has very high Gini Female and Gini Male coefficients. This is the continent with the highest rates of infant mortality and mortality under five years of age.[13] Sub-Saharan Africa is the worst affected region from the AIDS epidemic and tropical diseases like malaria. Another interesting thing to notice is that Europe and North America are the two regions with the lowest human capital inequality. This shows again that human capital is one of the crucial determinants of economic growth and development. Finally, it is readily seen that the Latin America & Caribbean region are experiencing considerably high female and male human capital inequality than Europe and North America, as the standard deviation reveals. There also exists substantial variation among countries in these regions.

Figure 3.1 presents the distribution of female (top panel), male (middle panel) and total human capital inequality (bottom panel) for the period 1960-2000. These distributions are constructed by non-parametric estimation of the density functions of Gini using a truncated gaussian kernel for a distribution in the interval $[0, 1]$.

As we can see, the density concentrates around a GiniF coefficient of 0.3, whereas the density

[13]We use infant mortality and mortality under five years of age. The two measures are per 1,000 live births.

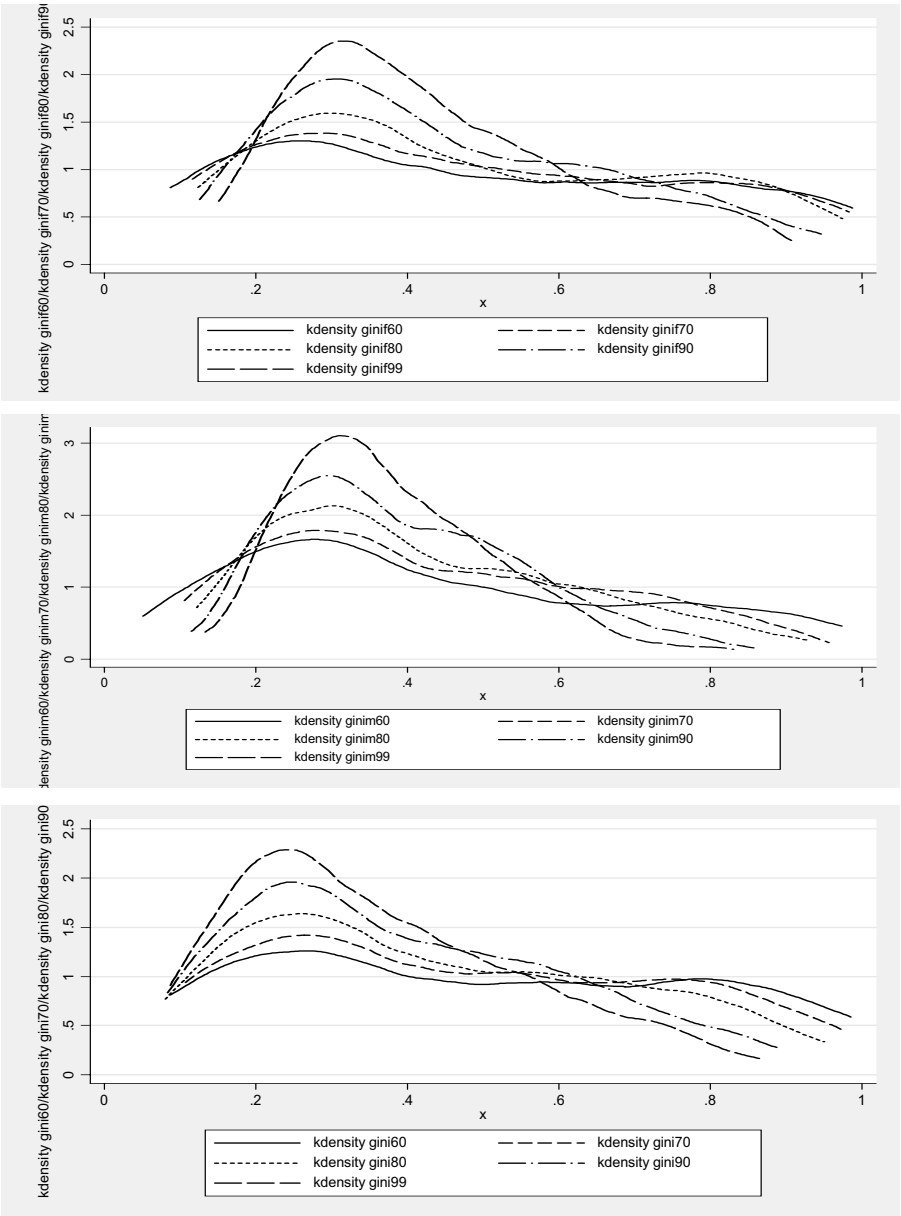

Figure 3.1: Density functions for Gini Female, Gini Male and Gini for the period 1960-2000

functions of the GiniM.[14] Using data on total human capital inequality from Castelló and Domenech (2002) we present the density functions of the GiniH. We can see that the male human capital inequality starts with lower inequality in 1960, but picks up very fast and exhibits much higher inequality by 2000 than female human capital inequality and total human capital inequality.

In summary, these graphs provide support of the hypothesis about convergence in human capital inequality across countries. More importantly, the gender-specific human capital inequality measures illustrated in Figure 3.1 suggest that inequality in education among women (and men) has been decreasing over time.

3.1.3 Cross-Sectional Correlations of All Variables

Next, we present the cross-sectional correlations between the variables used in our estimation for 73 countries (Table 3.2). The choice of these variables is clearly critical. By many accounts, these are the most frequently used variables in cross-country growth regression exercises as they have been found (in various degrees) to matter for growth and health. Infant mortality is determined by income, geography and health, which is proxied by malaria or tropics. In our analysis we use malaria in 1966 because we think that it has undoubtedly significant effect on infant mortality in subsequent years.

We note that the correlation between $GiniF_i$ and $Infant_i$ is 0.85. This shows a very strong positive effect of female human capital inequality on infant mortality. Another point is worth noticing-the correlation between GiniM and Infant is 0.81. Whether male human capital inequality matters for infant mortality is an empirical question and we will address it in the subsequent sections. The correlation coefficient between infant mortality (Infant) and Growth is -0.46. Because of reverse causality−low growth causes high infant mortality or high infant mortality causes low growth, we will correct for endogeneity of infant mortality.

A strong negative effect has Phys (the number of physicians per 1,000 people) on infant mortality-the correlation coefficient is -0.81. Countries that are located near to the tropics tend to have higher infant mortality. We use a dummy variable for tropics taken from Sachs and Weiner dataset. The

[14] GiniF denotes female human capital inequality, GiniM denotes male human capital inequality and GiniH denotes total human capital inequality.

Table 3.2: Cross-sectional correlations

	Inf.	Gr.	GF	GM	Schf.	Phys.	Tr.	Mal.	$(Y/L)_{60}$	Inv.	Pop.	H.
Infant	1											
Growth	-0.46	1										
GF	0.85	-0.39	1									
GM	0.81	-0.42	0.94	1								
Schf.	-0.84	0.30	-0.76	-0.77	0.97							
Phys.	-0.81	0.35	-0.68	-0.62	0.28	1						
Tr.	0.62	-0.42	0.40	0.37	0.03	-0.74	1					
Mal.	0.79	-0.25	0.66	0.62	-0.24	-0.75	0.69	1				
$(Y/L)_{60}$	-0.84	0.16	-0.75	-0.64	0.26	0.81	-0.56	-0.77	1			
Inv.	-0.74	0.48	-0.68	-0.68	0.38	0.62	-0.49	-0.60	0.56	1		
Pop.	0.72	-0.41	0.63	0.59	-0.23	-0.76	0.62	0.65	-0.67	-0.53	1	
Human	-0.84	0.47	-0.87	-0.90	0.55	0.69	-0.48	-0.69	0.73	0.72	-0.51	1

Notes:All variables are in natural algorithms. The sample size used here is 73 countries. GF is average Gini Female , GM is average Gini Male, Tr. is Tropics, Mal. is Malaria, Schf. is Schoolf.

correlation is 0.62, which shows a very high positive relationship between the two. The correlation coefficient between malaria and Infant is high and positive (0.79).[15]

We also examine the correlation between growth rate of GDP per worker and initial income ($(Y/L)_{60}$, population growth (Pop), schooling (Human) and infant mortality. We obtain correlations of $0.16, -0.41, 0.47$, and -0.46 correspondingly. Another interesting thing to notice is the correlation between female human capital inequality (GiniF) and the number of physicians (Phys). The correlation is -0.68, while the correlation between Tropics and GiniF is 0.40. Female human capital inequality is positively correlated with Malaria (0.66), positively correlated with population growth (Pop) -0.63 and negatively correlated with investment (Inv.) and schooling (Human): the correlations are -0.68 and -0.87, correspondingly.

To summarize the most important trends, the correlation between Infant and GiniF is very high and positive (0.85). This shows that higher female human capital inequality is positively correlated with high infant mortality and this is a preliminary evidence of this relationship. We will investigate this result further when we consider both cross-sectional and panel-data analysis and address endogeneity in the relationship between infant mortality and development.

[15]Recently it is feared that AIDS will soon surpass malaria, which has been around for at least a millennium as the most deadly infectious disease. AIDS may be a relatively new infectious disease, only quarter of a century old, but its negative impact is felt most profoundly in sub-Saharan Africa in which it is erasing decades of progress made in extending quantity and improving the quality of life.

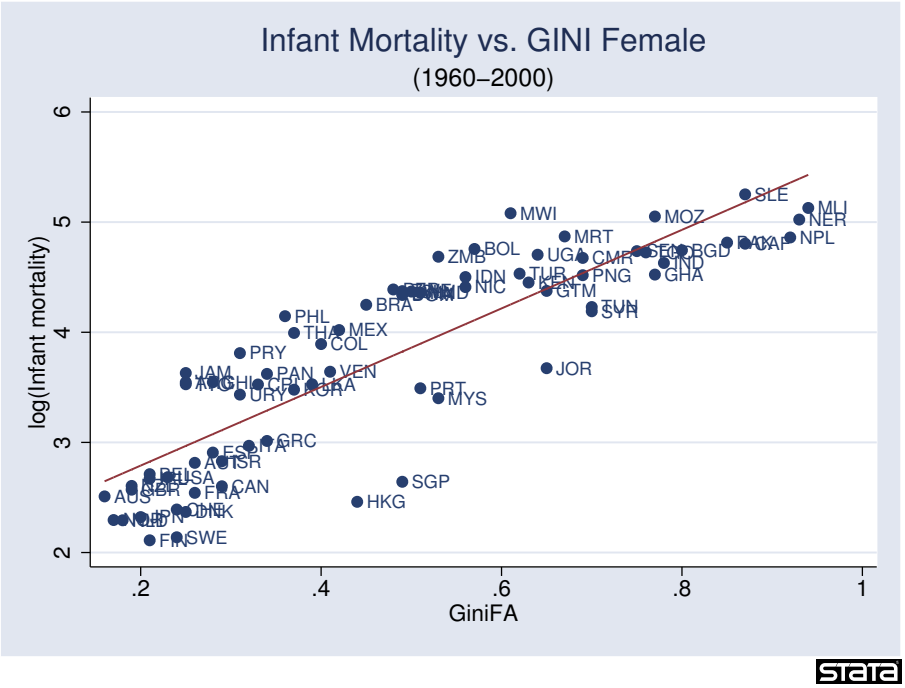

Figure 3.2: Scatter plot of Infant Mortality vs. Gini Female

3.2 Cross-Sectional Estimation and Baseline Results

In this section we present cross-sectional results for the two equations (infant mortality-female human capital inequality, and growth-infant mortality) for the full sample, which consists of 73 countries.[16]

We start our investigation of the first relationship by presenting a scatter plot (in Figure 3.2) of Gini Female vs. Infant mortality as a preliminary evidence of the positive relationship between female human capital inequality and infant mortality. Indeed Figure 3.2 provides evidence of a strong positive relationship (correlation) between these two variables.

Our main estimable equation is:

$$\ln(\inf ant)_i = \alpha_0 + \alpha_1 GiniF_i + \alpha_2 Growth_i + \alpha_3 \ln(phys)_i + \alpha_4 \ln(schoolf)_i + \alpha_5 malaria_i + \varepsilon_i, \quad (3.6)$$

Table 3.3 presents our basic results. Since our variable of interest is $GiniF_i$, in column 1 we esti-

[16] We average the right-hand side variables since the classical (white noise) measurement error gets averaged away at least partially. Hauk and Wacziarg (2004), using Monte Carlo simulations, show that averaging the right-hand side variables is very effective in reducing biases attributable to measurement error. Similarly, Lorentzen, McMillan and Wacziarg (2004) argue that averaging variables over time drastically reduces the incidence of measurement error compared to the case where they are entered at their values for any given year.

Table 3.3: Cross-country regressions

Dependent variable: $\ln(\text{Infant})_i$					
	(1)	(2)	(3)	(4)	(5)
Constant	2.3524^{***}	3.2187^{***}	3.7124^{***}	4.9329^{***}	3.3298^{***}
	(0.1726)	(0.1777)	(0.3007)	(0.0935)	(0.3039)
GiniF_i	3.2876^{***}	2.1699^{***}	1.5755^{***}	-	1.5140^{***}
	(0.2387)	(0.2333)	(0.3746)	-	(0.3696)
Growth_i	-0.2271^{**}	-0.1781^{*}	-0.1447	-0.1264	-0.2000^{*}
	(0.1083)	(0.1002)	(0.1001)	(0.1051)	(0.1072)
$\ln(\text{phys}_i)$		-0.3602^{***}	-0.3412^{***}	-0.3763^{***}	-0.2218^{***}
		(0.0672)	(0.0664)	(0.0721)	(0.0730)
$\ln(\text{schoolf}_i)$			-0.2000	-0.5508^{***}	-0.1252
			(0.1304)	(0.0915)	(0.1127)
malaria_i					0.5176^{***}
					(0.1502)
Adj. R^2	0.73	0.83	0.83	0.80	0.85
Obs.	73	73	73	73	72

Notes: Standard errors are in parentheses. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

mate the equation with only a constant, $GiniF_i$ and $Growth_i$.[17] We obtain a positive and significant estimate at the 1% level showing that higher female human capital inequality leads to higher infant mortality. We obtain a negative and highly significant coefficient estimate at the 5% estimate for *Growth*, showing that higher growth of GDP leads to lower infant mortality.

In column (2) we add $\ln(phys_i)$. Our results show that an increase of 1 physician per $1,000$ individuals leads to a reduction in infant mortality of 0.360 percentage points.[18] The estimate on $GiniF_i$ is positive and significant but falls in terms of its magnitude, and the coefficient on $Growth_i$ remains negative and significant.

In column (3) we add $\ln(schoolf_i)$,which is the average schooling years in the female population (Barro and Lee, 2001). The estimate on $\ln(schoolf_i)$ is negative, but not significant. The estimate on $GiniF_i$ continues to be positive and highly significant, but decreases its magnitude to 1.5755. This implies that a 0.1 unit increase in $GiniF_i$ leads to 1.5755 percentage change in infant mortality. The estimate on $Growth_i$ is negative, but not significant, and the estimate on $\ln(phys_i)$ is negative

[17] Growth is $\ln(\text{Y/L}_{2000})$-$\ln(\text{Y/L}_{1960})$.

[18] Since our dependent variable is in logs and the independent variables are in logs, the estimates can be interpreted in elasticity terms.

and stays almost the same in terms of the magnitude. In column (4) we exclude $GiniF_i$ but retain $ln(schoolf_i)$. We notice that the estimate on $ln(schoolf_i)$ is negative and significant at the 1% level, confirming the negative relationship between infant mortality and education. In column (5)−representing our benchmark estimable equation (3.6), we add $malaria_i$. We obtain a positive and significant estimate on $malaria_i$, suggesting that an increase in a country's area with malaria by 1% leads to an increase in $Infant_i$ by 0.5176 percentage points. We notice that the estimate on $GiniF_i$ continues to be positive and highly significant, while the estimate on $\ln(schoolf_i)$ is insignificant. This shows that our female human capital inequality measure successfully captures the (lower end of) distribution of education among women because when it is included with the $\ln(schoolf_i)$, it overcomes the effect of the mean level of female education.[19]

To summarize, our key estimate on female human capital inequality is found to be significant in the different specification. Even when it is included along with $\ln(schoolf_i)$, it continues to be positive and highly significant. This confirms our main hypothesis that higher inequality in education among women is a key determinant for higher infant mortality. Also, our results show that growth has a negative and statistically significant effect on infant mortality when included in the model with other regressors.

Next we turn to our second (growth) equation that is derived from using the augmented Solow model (with human capital). Specifically, we consider the following regression equation:[20]

$$\begin{aligned} \ln(Y/L)_{i,2000} - \ln(Y/L)_{i,1960} &= \beta_0 + \beta_1 \ln(Y/L)_{i,1960} + \beta_2 \ln(s_{ik}) + \beta_3 \ln(n_i + g + \delta)_i \\ &\quad + \beta_4 \ln(s_{ih}) + \beta_5 \ln(Infant)_i + \beta_6 \ln(Gov)_i \\ &\quad + \beta_7 GiniF_i + \beta_8 Inter_i + \varepsilon_i, \end{aligned} \tag{3.7}$$

where our dependent variable is growth of GDP per working age person, averaged over 1960-2000, s_{ik} is the ratio of average investment to GDP, s_{ih} is secondary school enrollment of working-age population, n_i is average population growth, $g + \delta = 0.05$ as in MRW, ε is an error term.

Table 3.4 presents the results from our estimation. In column (1) we estimate the standard Solow

[19] We estimated the different specifications with ln(GiniF). Our results are robust to the inclusion of Gini in logs. Results are available upon request.

[20] This equation is also consistent with the estimation equation in Domenech and Castelló (2002).

Table 3.4: Cross-country growth regressions

Dependent variable: $ln(Y/L)_{i,2000} - \ln(Y/L)_{i,1960}$						
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	3.6035^{***}	6.1072^{***}	6.4300^{***}	6.0424^{***}	6.0815^{***}	6.6044^{***}
	(0.6785)	(1.5477)	(1.4916)	(1.4847)	(1.447)	(1.5438)
$\ln(Y/L)_{i,1960}$	-0.4837^{***}	-0.5976^{***}	-0.5968^{***}	-0.5913^{***}	-0.5926^{***}	-0.5817^{***}
	(0.0963)	(0.1267)	(0.1233)	(0.1225)	(0.1205)	(0.1241)
$\ln(s_{ik})$	0.1869	0.0979	0.0474	0.0391	0.0464	-0.0629
	(0.1737)	(0.1769)	(0.1957)	(0.1928)	(0.1715)	(0.1815)
$\ln(n_i+g+\delta)$	-0.3211^{***}	-0.2404^{***}	-0.1914^{**}	-0.2174^{**}	-0.2148^{**}	-0.1015
	(0.0661)	(0.0883)	(0.0892)	(0.0891)	(0.0849)	(0.0951)
$\ln(s_{ih})$	0.5203^{***}	0.3708^{***}	0.3768^{***}	0.5197^{***}	0.5324^{**}	0.4291^{**}
	(0.0835)	(0.1173)	(0.1206)	(0.1650)	(0.2113)	(0.2112)
$\ln(infant)_i$		-0.3109^{*}	-0.3257^{*}	-0.3692^{**}	-0.3885^{**}	-0.3260^{*}
		(0.1831)	(0.1722)	(0.1719)	(0.1619)	(0.1895)
$\ln(gov)_i$			-0.0099	-0.0105	-0.0104	-0.0134^{*}
			(0.0077)	(0.0077)	(0.0075)	(0.0075)
$GiniF_i$				0.7279	0.4546	-0.1621
				(0.5556)	(1.6972)	(1.6305)
Inter.					0.0676	0.1013
					(0.4145)	(0.4363)
latin						-0.2672^{*}
						(0.1464)
subsaharan						-0.5047^{***}
						(0.1895)
asia						-0.0721
						(0.1522)
Adj. R^2	0.45	0.47	0.49	0.49	0.49	0.52
Obs.	73	73	73	73	73	73

Notes: Standard errors are in parentheses. It is assumed that $g + \delta = 0.05$ as in MRW. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

growth model with human capital, investment, population growth and initial income. The estimate on $\ln(Y/L)_{i,1960}$ is negative and significant at the 1% level, the estimate on $\ln(s_{ik})$ is positive and insignificant, the estimate on $\ln(n_i + g + \delta)$ is negative and significant at the 1%, and the coefficient on $\ln(s_{ih})$ is positive and significant at the 1% level. Our results are in accordance with previous studies and support the hypothesis of conditional convergence.

Next, we add our key variable, infant mortality. The estimate on $ln(infant)_i$ is negative and significant at the 10% level, the estimates on $\ln(s_{ih})$ and $\ln(n_i+g+\delta)$ have the expected signs and are significant, and the coefficient on $\ln(s_{ik})$ is insignificant. When we add government consumption, we notice that the estimate $\ln(infant)_i$ is again significant and negative,

the estimate on government consumption is negative and insignificant, and the other estimates have the expected signs and are significant. More interesting, in column (4) we add $GiniF$ and obtain an insignificant estimated coefficient. We see that the estimate on $\ln(infant)_i$ stays negative and significant. Our results show that an increase in infant mortality by 1% leads to a reduction in growth by 0.369 percentage points. We conjecture that female human capital inequality is one of the determinants of infant mortality and although the estimate on $GiniF_i$ is insignificant to growth due possibly to endogeneity problems, the estimate on $\ln(infant)$ remains negative and significant. Later on we will try to test for the indirect link between $GiniF_i$ and economic growth via infant mortality.

In column (5) we interact $GiniF$ and $\ln(infant)_i$ to allow infant mortality to depend on the degree of female human capital inequality. The coefficient is insignificant. The estimate on infant mortality stays negative and significant. All other estimates have the expected signs and there is not a big change in terms of the significance level. Finally, in column (6) we add three dummy variables to represent Latin America, sub-Saharan Africa and Asia. We notice that the dummy variables for Latin America and sub-Saharan Africa have negative and significant estimated coefficients consistent with previous literature. Even in the presence of the dummy variables, the estimate on $\ln(infant)_i$ remains negative and significant.

In summary, our empirical results obtained by estimating two separate equations show that female human capital inequality measured by the Gini coefficient leads to higher infant mortality and the later has a negative and significant effect on economic growth. We notice that the estimate on $\ln(infant)_i$ is very stable in the different specifications in Table 3.4. The estimate on $GiniF_i$ changes its magnitude from column 1 to column 2 and stays roughly the same in columns 3 and 4 in Table 3.3.

3.2.1 Addressing the Endogeneity Issue

In this section we examine the indirect effect of female human capital inequality on economic growth. In particular, we estimate our two key equations as a system in order to address the problem of reverse causality between infant mortality and growth. We argue that female human capital inequality affects economic growth through its impact on infant mortality as follows:

Gini Female$\Longrightarrow$Infant Mortality$\Longrightarrow$Growth

A common way to correct the endogeneity problem in much of the existing literature is to use instrumental variables. However, as it is well-known in the growth literature, it is very difficult to come up with a set of "good" instruments that are correlated with the potentially endogenous variable (in our case $Infant$), but not correlated with the error term.

We formulate the following structural model:

$$Growth = \alpha + \beta I + Z\eta + \varepsilon, \tag{3.8}$$

and

$$I = \gamma + \delta Growth + X\phi + \upsilon, \tag{3.9}$$

where I denotes $Infant$ and X and Z denote other explanatory variables. The equation of interest to us is equation (3.8). Specifically, we are interested in whether $Infant$ has a direct effect on Growth. To estimate equation (3.8) it is important that the order and rank conditions for identification are met. We further argue that female human capital inequality affects economic growth only through its effect on infant mortality.

The recent literature on income levels has proposed several historical or geographic instruments. Hall and Jones (1999) argued that European influence affects income only through its effect on "social infrastructure" and can be used as an instrument of social infrastructure on growth. Following this literature, we consider three instrumental variables for $Infant$: $ENGLISH$ (the share of the population speaking English), $EUROPE$ (the share of population speaking one of the major languages of Western Europe: English, French, German, Portuguese, or Spanish), and $LATITUDE$(the absolute value of latitude in degrees divided by 90 and is taken from Frankel and Romer (1999)).

To examine the validity of our instruments we test the over-identifying restrictions where the endogenous variable, $Infant$, is explained by the three instruments, $ENGLISH, EUROPE$ and $LATITUDE$. This implies that we have two over-identifying restrictions. Panel B in Table 3.5 reports the p-value from χ^2 Sargan's (1958) test. This is a test of the joint hypothesis that the

Table 3.5: Instrumental Variable regressions

IV Regressions of $ln(Y/L)_{i,2000} - \ln(Y/L)_{i,1960}$	
Panel A: Two Stage Least Squares	
Specification	Full Sample
Constant	7.2313^{***}
	(2.4482)
$\ln(Y/L)_{i,1960}$	-0.6466^{***}
	(0.1382)
$\ln(s_{ik})$	-0.0080
	(0.2538)
$\ln(n_i + g + \delta)_i$	-0.1796^{*}
	(0.1055)
$\ln(s_{ih})_i$	0.4674^{***}
	(0.1561)
$\ln(Infant)_i$	-0.5293^{*}
	(0.3031)
GiniF_i	0.8351
	(0.6283)
$\ln(\text{gov})_i$	-0.0110
	(0.0083)
Adj.R^2	0.48
Obs.	73
	Panel B: Specification Tests (p value)
Overidentifying Restrictions	0.3203

Notes: Standard errors are in parentheses. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

included instruments are valid instruments. We fail to reject the null of no correlation between the instruments and the error term, indicating that our over-identifying instruments are satisfactory.

To evaluate the quality of our instruments, we test their validity by estimating reduced form regressions of $Infant$ on the instrumental variables and the exogenous variables. We test the joint significance of the coefficients on the instruments and we are able to reject the null of zero coefficients at the 1% level of significance. This suggest that our instruments provide useful information in addition to that provided by the explanatory variables.

We present the results from this exercise in Panel A of Table 3.5. Our results show that $Infant_i$ has a negative and statistically significant effect on economic growth. If we compare our relevant estimate with the those obtained in the cross-sectional estimation (Table 3.4), it is readily seen that it is larger than OLS estimates. This suggests that measurement error seems to be as or more important than reverse causality and omitted variables biases. As expected, the estimate on $\ln(Y/L)_{i,1960}$ is negative and significant at the 1% level, and the estimates on the other regressors are not significant. Our estimate of interest GiniF_i has no direct effect on growth. The coefficient is insignificant.

Our finding that infant mortality has a negative effect on growth after correcting for endogeneity is quite reassuring for two reasons. On one hand, it provides evidence that infant mortality matters for growth and development. On the other hand, it supports our main hypothesis that female human capital inequality indirectly affects growth through its effect on infant mortality.

3.3 Robustness Analysis

In this section we examine the robustness of our cross-sectional results to the inclusion of income inequality, male human capital inequality, tropics, latitude and government public expenditures. In addition, we examine the robustness of our baseline results when we replace our dependent variable with infant mortality in 2000. We present these results in Tables 3.6, 3.7 and 3.8. Finally, we attempt to extend the cross-sectional analysis to panel estimation using the full sample.

3.3.1 Cross-Sectional Robustness Analysis

We begin the robustness analysis by including $GiniF_i$ *and* $GiniM_i$ in the regressions. In the basic regression equation of Table 3.6 column (1) the estimate on $GiniF_i$ is significant at the 1% level, while the estimate on $GiniM_i$ is not significant. This result confirms our main hypothesis that it is inequality among women that leads to higher infant mortality. When we include $Growth_i$ in column (2) in addition to our inequality measures, it remains negative and highly significant. The estimate on $GiniF_i$ continues to be positive and significant while the coefficient on $GiniM_i$ is insignificant, providing evidence on the direct relationship between female human capital inequality and infant mortality. In column (3) we incorporate $GiniM_i$ in our benchmark regression equation that includes also $\ln(\text{phys}_i)$, $\ln(\text{schoolf}_i)$ and malaria_i. Results are similar to those in Table 3.3 column (5) in our benchmark analysis.

Table 3.6: Robustness analysis with Gini-Male

Dependent variable:$\ln(\text{Infant})_i$			
	(1)	(2)	(3)
Constant	2.0377^{***}	2.3299^{***}	3.2523^{***}
	(0.1469)	(0.2073)	(0.2969)
GiniF_i	3.0855^{***}	3.1211^{***}	1.3415^{**}
	(0.7545)	(0.7295)	(0.6560)
GiniM_i	0.5964	0.2389	0.3250
	(0.9910)	(0.9994)	(0.7606)
Growth_i		-0.2221^{**}	-0.1965^{*}
		(0.1120)	(0.1053)
$\ln(\text{phys}_i)$			-0.2252^{***}
			(0.0729)
$\ln(\text{schoolf}_i)$			-0.1046
			(0.1046)
malaria_i			0.5197^{***}
			(0.1512)
Adj. R^2	0.72	0.73	0.85
Obs.	73	73	72

Notes: Standard errors are in parentheses. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

In Table 3.7 we explore the robustness of our results to the inclusion of more relevant variables motivated by theory and found in the literature. In column (1) we use Deininger and Squire (1996) measure of income inequality. We average the data since using the initial income inequality data

Table 3.7: Robustness analysis

Dependent variable:ln(Infant)$_i$				
	(1)	(2)	(3)	(4)
Constant	2.5269***	3.2360***	3.4976***	3.6959***
	(0.5841)	(0.4233)	(0.3483)	(0.3543)
GiniF$_i$	1.6497**	1.5694***	1.4179***	1.3111***
	(0.7272)	(0.4418)	(0.4262)	(0.4187)
GiniY$_i$	0.0188***			
	(0.0050)			
Growth$_i$	−0.2130*	−0.1743	−0.1237	−0.1423
	(0.1139)	(0.1364)	(0.1093)	(0.1083)
ln(phys$_i$)	−0.1592*	−0.1945**	−0.1835**	−0.1911***
	(0.0821)	(0.0781)	(0.0715)	(0.0699)
ln(schoolf$_i$)	−0.1645	−0.1358	−0.2066*	−0.2175*
	(0.2029)	(0.1059)	(0.1239)	(0.1250)
malaria$_i$	0.3935**	0.4640**	0.4222***	0.3988***
	(0.1778)	(0.2034)	(0.1594)	(0.1513)
tropics$_i$		0.1062		
		(0.2181)		
latitude$_i$			−0.0045**	−0.0044***
			(0.0018)	(0.0017)
public$_i$				−3.1379*
				(1.8750)
Adj. R^2	0.86	0.85	0.86	0.86
Obs.	59	72	72	72

Notes: Standard errors are in parentheses. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

could reduce our sample size drastically. Now our sample size is reduced to 59 countries. We notice that even though the $GiniY_i$ coefficient is positive and highly significant, the estimate on $GiniF_i$ continues to be positive and highly significant.

In addition to our regressors from the benchmark model, we add $tropics_i$. The dummy variable for tropics is from the Sachs and Weiner dataset. Tropical climate is measured by a variable that takes the value 1 for a country in which the entire land area is subject to a tropical climate, and 0 for a country with no land area subject to a tropical climate. Countries between the two extremes are assigned a fraction representing the approximate proportion of land area subject to a tropical climate. Although the coefficient is not significant, the estimate on $malaria_i$ is still positive and significant. This result can be interpreted as evidence of the sizable effect of infectious diseases, like malaria, on economic development.

We also test the robustness of our finding to the inclusion of two more variables: latitude and public health expenditures as a percentage of GDP. Our measure of latitude is taken from Hall and Jones (1999) and measures distance from the equator as the absolute value of latitude in degrees divided by 90 to place it on a 0 to 1 scale. It is widely known that economies further from the equator are more successful in terms of per capita income.[21]

The results reported in Column (3) suggest a negative effect of $latitude_i$ on $infant_i$. This finding is supportive of the idea that those countries that are further from the equator are more developed than countries like Gabon, Congo, Somalia, Kenya, Uganda, just to name a few. We also obtain a negative and significant estimate on $\ln(schoolf_i)$ and positive and significant estimate on $GiniF_i$. Column (4) reports results when we add public health expenditure. Our measure of public health expenditures is taken from WDI (2000) and is averaged over 1960-2000. Our finding is that this variable is found to be significant and negatively related to infant mortality and that $GiniFi$ remains very significant and stable in magnitude at around 1.4.

Next, we examine the robustness of our results to replacing average infant mortality with infant mortality in 2000 as our dependent variable. Results reported in Table 3.8 are based on our main estimable equation (3.6). In column (1) we obtain results similar to the results from our benchmark model. We notice that when we use $GiniF_i$ and $GiniM_i$ in the same regression (column 2), they

[21] For a more detailed discussion, please refer to Hall and Jones (1999).

Table 3.8: Robustness analysis

Dependent variable:ln(Infant in 2000)$_i$				
	(1)	(2)	(3)	(4)
Constant	2.7598***	2.5237***	2.9193***	3.0906***
	(0.3803)	(0.3892)	(0.3804)	(0.3893)
GiniF$_i$	1.4135***	0.8878	1.3221***	1.2298**
	(0.5024)	(0.9037)	(0.4907)	(0.5029)
GiniM$_i$		0.9901		
		(1.1406)		
Growth$_i$	−0.3869***	−0.3763**	−0.3145**	−0.3305**
	(0.1533)	(0.1498)	(0.1572)	(0.1545)
ln(phys$_i$)	−0.3071***	−0.3176***	−0.2707***	−0.2772***
	(0.0820)	(0.0792)	(0.0779)	(0.0780)
ln(schoolf$_i$)	−0.1703	−0.1076	−0.2478*	−0.2571*
	(0.1352)	(0.1173)	(0.1336)	(0.1361)
malaria$_i$	0.8027***	0.8090***	0.7121***	0.6918***
	(0.2064)	(0.2107)	(0.2260)	(0.2146)
latitude$_i$			−0.0043**	−0.0042**
			(0.0019)	(0.0018)
public$_i$				−2.7115
				(2.2454)
Adj. R^2	0.85	0.85	0.85	0.85
Obs.	72	72	72	72

Notes: Standard errors are in parentheses. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

are both insignificant. In addition to that, our estimate on $\ln(phys_i)$ stays negative and significant. The signs and significance of the other coefficients remain unchanged.

In column (3) we add $latitude_i$. The estimate is negative and significant. We notice as well that the estimate $\ln(schoolf_i)$ is negative and significant, even though the coefficient on $GiniF_i$ is positive and highly significant. Similar results are obtained in column (4), where we include $public_i$. The estimate on $public_i$ is insignificant, but still negative.

In summary, we find that our cross-sectional results are in general quite robust to the inclusion of different variables. Our estimate on $GiniF_i$ (with only one exception; Table 3.8, column 2)) is positive and significant in the different specifications, confirming the positive relationship between female human capital inequality and infant mortality.

3.3.2 Panel Robustness Analysis

This section extends our baseline cross-sectional results to consider panel data techniques. The main advantage of panel data technique is that it allows one to control for unobserved heterogeneity across countries.[22] Following much of the literature on cross-country panel estimation, we average the data in five-year time intervals. Our panel dataset is therefore unbalanced with a total of 396 observations with a maximum of 8 and minimum of 1 observation.

Our benchmark infant regression equation takes the form:

$$\ln(infant)_{it} = \alpha_0 + \alpha_1 GiniF_{it} + \alpha_2 Growth_{it} + \alpha_3 \ln(phys)_{it} + \alpha_4 \ln(schoolf)_{it} + \alpha_5 malaria_i + \varepsilon_{it}, \tag{3.10}$$

Table 3.9 presents the results from our estimation. In column (1) we estimate the model using the Between Estimator.[23] In a recent paper Hauk and Wacziarg (2004) argue that using an OLS estimator applied to a single cross-section of variables averaged over time (BE) performs best in terms of the extent of bias on each of the estimated coefficients. The estimate on $GiniF_{it}$ is positive and

[22] Temple (1999) discusses several advantages of using panel data analysis. First, it allows one to control for omitted variables that are persistent over time. For example, variations in technology across countries are likely to be correlated with the regressors. By using the panel data technique, the unobserved heterogeneity in the initial level of efficiency is controlled for. Second, it allows several lags of the regressors to be used as instruments. A commonly used approach in the literature is GMM to estimate dynamic panel data models. Despite these advantages, panel data techniques leave some uncertainty about the time intervals. Most researchers find it useful to use five or ten year averages to avoid business cycle effects.

[23] We refer the interested reader to Greene (2000, Ch.14, pp. 562-565) for further information on the Between Estimator.

Table 3.9: Panel infant regressions

Dependent variable: ln(Infant)$_{it}$		
	(1) Full sample with Between Estimator	(2) Full sample with time effects
Constant	3.1421***	3.5689***
	(0.4864)	(0.2326)
GiniF$_{i,t}$	1.5415**	1.3296***
	(0.6345)	(0.2986)
Growth$_{i,t}$	−1.2625*	−0.5482***
	(0.6984)	(0.1842)
lh(phys)$_{i,t}$	−0.0898	−0.0646**
	(0.0594)	(0.0270)
ln(schoolf)	−0.2117	−0.2280***
	(0.1632)	(0.0837)
malaria	0.6925***	0.8389***
	(0.1590)	(0.0684)
d65		−0.1654**
		(0.0846)
d70		−0.2638***
		(0.0845)
d75		−0.4060***
		(0.0995)
d80		−0.6248***
		(0.1037)
d85		−0.6238***
		(0.1079)
d90		−0.8056***
		(0.0992)
d95		−0.9129***
		(0.1137)
Adj. R^2	0.81	0.37
Obs.	396	530

Notes: d65-d95 denote time dummies for 1965-1995, respectively. It is assumed that g+δ = 0.05. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10 % level.

significant at the 5% level, the estimate on growth is negative and significant at the 10% level, and the estimate on $\ln(schoolf)_{it}$ is insignificant as the coefficient on $\ln(phys)_{it}$. The coefficient on *malaria* is significant. To allow for the possibility of time effects, we have also estimated the model by adding $(T-1)$ time dummies, using OLS. These dummies are meant to capture exogenous shocks specific to each five-year period. The coefficient of $GiniF_{it}$ continues to be positive and highly significant, and that of $Growth_{it}$ and $\ln(phys)_{it}$ are negative and significant. Finally, the estimate on *malaria* is positive and highly significant.[24] In summary, our panel estimation is generally supportive of our cross-sectional results. In particular, the impact of $GiniF_{it}$ on infant mortality obtained from the panel estimation is shown to be positive and similar in magnitude to that obtained from our cross-sectional estimation. Our results also provide evidence on the decline of infant mortality for the period 1960-2000.[25]

We also present a panel-data analysis of our growth equation (3.8). Following much of the literature on cross-country panel estimation, we average the data in five-year time intervals and include $\ln(Y/L)_{i,t}$ at the beginning of each the five-year periods. Our panel dataset is unbalanced with a total of 530 observations. Our growth panel regression equation is:

$$\begin{aligned}\ln(Y/L)_{it} - \ln(Y/L)_{i,t-5} &= \beta_0 + \beta_1 \ln(Y/L)_{i,t} + \beta_2 \ln(s_{itk}) + \beta_3 \ln(n_{it} + g + \delta) \\ &\quad + \beta_4 \ln(s_{ith}) + \beta_5 \ln(\inf ant)_{it} + \beta_6 \ln(Gov)_{it} + \beta_7 GiniF_{it} \\ &\quad + \beta_8 Inter_{it} + \varepsilon_{it,} \qquad (3.11)\end{aligned}$$

where $\ln(Y/L)_{it} - \ln(Y/L)_{i,t-5}$ is growth for each of the five-year periods, $\ln(Infant)_{it}$ is average infant mortality for each five-year period, $\ln(Gov)_{it}$, $GiniF_{it}, Inter_{it}$ are five-year averages as well.

Table 3.10 presents results from the panel data analysis for the full sample under different specifi-

[24] Furthermore, to account for the possibility of country-specific effects as well as time effects, we estimate a two-way fixed-effect specification that involves the addition of 73 country-specific dummy variables and 7 time dummy variables. However, as there are more coefficients to estimate, we lose a large number of degrees of freedom which clearly biases our estimates. As Griliches and Hausman (1986) note, in regressions using panel data with fixed effects specifications, measurement error in the explanatory variables can lead to coefficient estimates that are "too low" and therefore insignificant; in controlling for the various fixed effects, the relative importance of measurement errors in the explanatory variables becomes greatly exacerbated, biasing coefficient estimates.

[25] Accompanying the decline in the mortality rates, there has been a sharp decline in the fertility rates (Sebnem Kalemli-Ozcan, 2002). Demographers view these declines in mortality and fertility as part of a single "demographic transition." There are different theories trying to explain the reasons why fertility declined. One theory suggests that fertility decline is due to mortality decline. Another theory, supported from Galor and Weil (1999, 2000), Galor and Moav (2002) suggests that the demographic transition was caused by the increase in the return to education which led to a quantity-quality trade-off and demographic transition. A different theory attributes the demographic transition to the decline in gender wage gap (Galor and Weil, 1996).

Table 3.10: Panel growth regressions

Dependent variable: $ln(Y/L)_{it} - \ln(Y/L)_{i,t-5}$		
	(1) Full Sample with Between Estimator	(2) Full Sample with time effects
Constant	0.3376^{***}	0.3078^{***}
	(0.0699)	(0.0460)
$\ln(\mathrm{Y/L})_{i,in}$	-0.0022	-0.0030
	(0.0025)	(0.0020)
$\ln(s_{itk})$	0.0523^{**}	0.0826^{***}
	(0.0233)	(0.0185)
$\ln(\mathrm{n}_{it}+\mathrm{g}+\delta)$	-0.0164	-0.0154^{**}
	(0.0114)	(0.0065)
$\ln(s_{ith})$	0.0155	0.0227^{**}
	(0.0173)	(0.0111)
$\ln(\mathrm{infant})_{it}$	-0.0167	0.0016
	(0.0200)	(0.0112)
$\ln(\mathrm{gov})_{it}$	-0.0015	-0.0014
	(0.0010)	(0.0009)
GiniF_{it}	0.0883	0.0970^{**}
	(0.0689)	(0.0446)
d65		0.0874^{***}
		(0.0194)
d70		0.1228^{***}
		(0.0209)
d75		0.1595^{***}
		(0.0206)
d80		0.0529^{**}
		(0.0222)
d85		-0.0008
		(0.0238)
d90		-0.0553^{**}
		(0.0250)
d95		-0.0679^{***}
		(0.0241)
Adj. R^2	0.36	0.45
Obs.	530	530

cations. Column (1) presents results from estimation of the full sample with the Between Estimator. The estimate on $\ln(Y/L)_{i,in}$ is negative, but insignificant, the estimate on $\ln(\mathrm{s}_{itk})$ is positive and significant, the estimates on $\ln(n_{it}+g+\delta)$ and $\ln(s_{ith})$ are insignificant. The estimates on $\ln(\inf ant)_{it}$ and $GiniF_{it}$ are both insignificant. To allow for the possibility of time effects, we have also estimated the model by adding (T-1) time dummies. The results are presented in column 2. The estimates on $\ln(\inf ant)_{it}$ is negative but insignificant. The estimate on $\ln(Y/L)_{i,in}$ continues to be negative and insignificant, the estimate on $\ln(\mathrm{s}_{itk})$ is positive and highly significant. The coefficients on $\ln(n_{it}+g+\delta)$ and $\ln(s_{ith})$ are both significant and have the expected signs. In summary, our results from the growth panel regression are weaker than the cross-sectional regression. This is consistent with other work in growth literature, where results under panel estimation are noiser than cross-sectional estimation due to variations in growth picking up cycle effects rather than long-run effects.

3.4 Conclusion

This essay provides new evidence on the effect of female human capital inequality on infant mortality and the effect of the latter on economic growth. First, this paper considers the relationship between infant mortality and female human capital inequality measured by the Gini coefficient in both cross-sectional and panel estimations. It is shown that higher female human capital inequality leads to higher infant mortally rates. Second, following Mankiw, Romer and Weil (1992) we estimate a second equation where our dependent variable is the growth rate of income per worker following. We add infant morality as another regressor and show that it has a negative and significant effect on income. Third, in order to address the problem of reverse causality between infant mortality and growth, we correct for endogeneity of infant mortality using common instruments. Our results suggest a positive effect of female human capital inequality on infant mortality and a negative effect of the latter on economic growth.

This study contributes to the literature not only because it constructs gender-specific human capital inequality measures using the Gini coefficient, but also because it considers a new channel through which infant mortality affects economic growth-female human capital inequality. It is in-

equality in education among women, that affects infant mortality and the latter affects economic growth and development. We think that this is important since it has valuable policy implications. Specifically, our analysis suggests diverting general education subsidy money directly into the education of the least educated women, especially in less-developed countries. This can have large payoffs in economic development and, consequently the welfare of future generations.

Chapter 4

Conclusion

This dissertation aims to investigate the following: (i) the potential effect of AIDS on cross-country income; (ii) the effect of human capital inequality on infant mortality and the effect of the latter on economic growth. The first essay, *What Do We Know About the Impact of AIDS on Cross-Country Income Level So Far?*, investigated the impact of AIDS on cross-country income levels. Contrary to previous work on AIDS, we make use of the officially reported AIDS incidents from UNAIDS/WHO on 89 countries for the period 1979-2000, during which the AIDS epidemic has spread across the world.

Using the extended Solow model as the basis of our empirical analysis we first showed that in the full sample and non-OECD subsample, the coefficient estimate for AIDS is negative and marginally significant. For the OECD countries, we obtained an insignificant coefficient estimate, which implies that AIDS has no quantifiable effect on the income level for these countries. We also utilized the time dimension of our data and employed panel-data techniques on the extended Solow model with AIDS as a regressor. AIDS enters negative and highly significant in all of the specifications considered except from the specification with country and time effects, where the estimate is insignificant. Regression analysis using AIDS by age group reveals that only the coefficient on AIDS between the ages 16-34 is significant with a negative sign.

The second essay, *Female Human Capital Inequality, Infant Mortality and Growth*, provided new evidence on the effect of female human capital inequality on infant mortality. It also showed a negative effect of infant mortality on economic growth.

First, this essay considered the relationship between infant mortality and female human capital inequality measured by the Gini coefficient in both cross-sectional and panel estimations. Our

results confirmed the positive effect of female human capital inequality on infant mortality. Second, we estimated a second equation where our dependent variable is the growth rate of income per worker following Mankiw, Romer and Weil (1992). We add infant morality as another regressor and show that it has a negative and significant effect on income. Third, in order to address the problem of reverse causality between infant mortality and growth, we correct for endogeneity of infant mortality using common instruments. Our results suggest a positive effect of female human capital inequality on infant mortality and a negative effect of the latter on economic growth.

This essay contributes to the literature not only because it constructs gender-specific human capital inequality measures using the Gini coefficient, but also because it considers a new channel through which infant mortality affects economic growth-female human capital inequality. We provided evidence on the positive effect of female human capital inequality on infant mortality across countries. In addition to that this essay showed a negative effect of infant mortality on economic growth and development. We think that this is important since it has valuable policy implications, especially in less-developed countries.

Bibliography

[1] Acemoglu, D., S. Johnson and J. Robinson. (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* **91***, 1369-1401.*

[2] Auld, M.C. (2003). "Choices, Beliefs, and Infectious Disease Dynamics," *Journal of Health Economics* **22**, 361-377.

[3] Barro, R.J. (1991). "Economic Growth in a Cross Section of Countries," *The Quarterly Journal of Economics* **106***,* 407-443.

[4] Barro, R.J. and J-W Lee. (2001). "International Data on Educational Attainment: Updates and Implications," *Oxford Economic Papers* **53**, 541-563.

[5] Bernanke, B.S. and R.S. Gürkaynak (2001). "Is Growth Exogenous? Taking Mankiw, Romer and Weil Seriously," *NBER Macroeconomics Annual* **16**, eds. Bernanke B. and K. Rogoff, MIT Press.

[6] Bloom, D.E. and Ajay S. Mahal. (1997). "Does the AIDS Epidemic Threaten Economic Growth?," *Journal of Econometrics* **77**, 105-124.

[7] Breierova L., E. Duflo. (2004). "The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers?," *NBER Working Paper*, no. **10513**.

[8] Burger, R. (1996). "The Macroeconomic Impact of HIV/AIDS in South Africa: A Supply-side Analysis," working paper, University of Stellenbosch.

[9] Caselli, F., G. Esquivel and F. Lefort. (1996). "Reopening the Convergence Debate: A New Look at Cross Country Growth Empirics," *Journal of Economic Growth* **1**, 363-389.

[10] Caselli, F., and D.J. Wilson. (2004). "Importing Technology," *Journal of Monetary Economics* 51, 1-32.

[11] Castello, A. and R. Domenech (2002). "Human Capital Inequality and Economic Growth: Some New Evidence," *Economic Journal vol.* **112(127)**, C187-C200.

[12] Chakraborty, Shankha (2003). "Endogenous Lifetime and Economic Growth," *Joural of Economic Theory*, *vol.* **116**, 2004.

[13] Chakraborty, S., M. Das (2004). "Mortality, Human Capital and Persistent Inequality," , *working paper,* University of Oregon.

[14] Chakraborty, S., C. Papageorgiou and F.Sebastian (2005). "Disease and Development," , *working paper,* University of Oregon.

[15] Clark, R. and D. Vencatachellum. (2003). "Economic Development and HIV/AIDS Prevalence," working paper, Cirano, Montreal.

[16] Corrigan, P., G. Glomm and F. Mendez. (forthcoming). "AIDS Crisis and Growth," *Journal of Development Economics.*

[17] Cuddington, J.T. (1993). "Modeling the Macroeconomic Effects of AIDS, with Application to Tanzania," *The World Bank Economic Review* **7(2)**, 173-189.

[18] Cuddington, J.T. and J.D. Hancock. (1994). "Assessing the Impact of AIDS on the Growth Path of the Malawian Economy," *Journal of Development Economics* **43**, 363-368.

[19] de Walque, D. (2004). "How Does the Impact of an HIV/AIDS Information Campaign Vary with Educational Attainment? Evidence from Rural Uganda," working paper, World Bank.

[20] Duffy, J. and C. Papageorgiou. (2000). "A Cross-Country Empirical Investigation of the Aggregate Production Function Specification," *Journal of Economic Growth* **5**, 87-120.

[21] Durlauf, S. and P. Johnson. (1995). "Multiple Regimes and Cross-Country Growth Behavior," *Journal of Applied Econometrics* **10**, 365-84.

[22] Durlauf, S. and D. Quah. (1999). "The New Empirics of Economic Growth," *Handbook of Macroeconomics*, eds. Taylor J.B. and M. Woodford, Vol. 1, Ch.4, 235-308.

[23] Durlauf, S., A. Kourtellos and A. Minkin. (2001). "The Local Solow Growth Model," *European Economic Review* **45**, 928-940.

[24] Durlauf, S., P. Johnson and J. Temple. (forthcoming). "Growth Econometrics," *Handbook of Economic Growth*, eds. Aghion P. and S. Durlauf, Part II, Ch.7.

[25] Easterly, W. and R. Levine (2001). "It's Not Factor Accumulation: Stylized Facts and Growth Models," *World Bank Economic Review* **13**, 129-158.

[26] Ferreira, C.P, and S. Pessoa. (2003). "The Long-Run Economic Impact of AIDS," working paper, Graduate School fo Economics, Fundacao Getulio Vargas.

[27] Frankel, J.A. and D. Romer. (1999). "Does Trade Cause Growth?," *American Economic Review* **89**, 379-399.

[28] Gallup, J.L. and Jeffrey D. Sachs (2000). "The Economic Burden of Malaria," *CID Working Paper No.* **52**.

[29] Galor, O., D. Weil (1999). "From Malthusian Stagnation to Modern Growth," *American Economic Review* no.**89**, 374-387.

[30] Galor, O., D. Weil (2000). "Population , Technology and Growth: From the Multhusian Regime to the Demographic Transition and Beyond," *American Economic Review* no.**90**, 806-828.

[31] Galor, O., O. Moav (2002). "Natural Selection and the Origin of Economic Growth," *Quaterly Journal of Economics* no.**117**.

[32] Glied, Sh., A. Lleras-Muney (2003). "Health Inequality, Education and Medical Innovation," *NBER Working Paper* no.**9738**.

[33] Greene, W.H. (2000). *Econometric Analysis*, Fourth Edition, Prantice Hall, New Jersey.

[34] Griliches, Z. and J.A. Hausman. (1986). "Errors in Variables in Panel Data," *Journal of Econometrics* **31**, 93-118.

[35] Hall, R.E. and C.I. Jones, (1999). "Why Do Some Countries Produce so Much More Output Per Worker than Others," *Quarterly Journal of Economics* **114**, 83-116.

[36] Hammer J.S. and M. Gersovitz. (2004). "The Economical Control of Infectious Diseases," *Economic Journal* **114**, 1-27.

[37] Hansen, B.E. (2000). "Sample Splitting and Threshold Estimation," *Econometrica* **68**, 575-603.

[38] Hauk, W. and R. Wacziarg. (2004). "A Monte Carlo Study of Growth Regressions," *NBER Technical Working Paper No.* 296.

[39] Islam, N. (1995). "Growth Empirics: A Panel Data Approach," *The Quarterly Journal of Economics* **110**, 1127-1170.

[40] Jamison, D. T, M. Sandbu, J. Wang (2001). "Cross-Country Variation in Mortality Decline, 1962-87: The Role of Country-Specific Technical Progress," *CMH Working Paper Series.*

[41] Kalaitzidakis, P., T.P. Mamuneas, A. Savvides and T. Stengos. (2001). "Measures of Human Capital and Nonlinearities in Economic Growth," *Journal of Economic Growth* **6**, 229-254.

[42] Kalemli-Ozcan, S., N. Ryder, D. Weil (2000). "Mortality Decline, Human Capital Investment, and Economic Growth," *Journal of Development Economics*, **62**.

[43] Kalemli-Ozcan, S. (2002). "Does the Mortality Decline Promote Economic Growth?," *Journal of Economic Growth,* **7**.

[44] Kalemli-Ozcan, S. (2004). "AIDS and Economic Development: Evidence on the Reversal of the Fertility Transition," working paper, University of Houston.

[45] Kalipeni, E., S. Craddock, J.R. Oppong and J. Ghosh. (2004). *HIV & AIDS in Africa: Beyond Epidemiology*, Blackwell Publishing.

[46] Kelly, K. and D. Vencatachellum. (2004). "HIV/AIDS Behavioral Prevention: Evidence from South Africa, working paper, HEC Montréal.

[47] Kremer, M. (2002). "Pharmaceuticals and the Developing World," *Journal of Economic Perspectives* **16**, 67-90.

[48] Levine, R., N. Loyaza and T. Beck (2000). "Financial Intermediation and Growth: Causality and Causes," *Journal of Monetary Economics* **46**, 31-77.

[49] Leigh, A and C. Jencks (2005). "Inequality and Health: Long-Run Evidence from a Panel of Countries", *working paper, NBER Summer Institute.*

[50] Levine, R., N. Loyaza and Th. Beck (2000). "Financial intermediation and growth: Causality and causes," *Journal of Monetary Economics* **46**, 31-77.

[51] Levy, A. (2002). "A Lifetime Portfolio of Risky and Risk-Free Sexual Behaviour and the Prevalence of AIDS," *Journal of Health Economics* **21**, 993-1007.

[52] Liu, C. and T. Stengos. (1999). "Non-Linearities in Cross-Country Growth Regressions: A Semiparametric Approach," *Journal of Applied Econometrics* **14**, 527-538.

[53] Lorentzen P., J. McMillan, R. Wacziarg (2004). "Death and Development", *working paper*, Stanford University.

[54] Lorgelly P. (2000). "Are there Gender-Separate Human Capital Effects on Growth? A Review of the Recent Empirical Literature," *Centre for Research in Economics Development and International Trade*, University of Nottingham.

[55] Mankiw, N.G., D. Romer and D.N. Weil. (1992). "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* **107**, 407-37.

[56] Masanjala, W. and C. Papageorgiou. (2004). "The Solow Model with CES Technology: Nonlinearities and Parameter Heterogeneity," *Journal of Applied Econometrics* **17**, 171-201.

[57] Masanjala, W. and C. Papageorgiou (2004). "Rough and Lonely Road to Prosperity: A reexamination of the sources of growth in Africa using Bayesian Model Averaging," *working paper*, LSU.

[58] McCarthy, F., D.H. Wolf and Y. Wu. (2002). "The Growth Costs of Malaria," working paper, World Bank.

[59] Oster, E. (2004). "Sexually Transmitted Infections, Sexual Behavior Change and the HIV/AIDS Epidemic," working paper, Harvard University.

[60] Papageorgiou, C. (2002). "Trade as a Threshold Variable for Multiple Regimes," *Economics Letters* **77,** 85-91.

[61] Ressler, R., M.S. Waters, J.K. Watson, and R.C. Hill. (2002). "Welfare Payments and the Spread of AIDS in the United States," working paper, LSU.

[62] Schultz, P. (1993). "Mortality Decline in the Low-Income World: Causes and Consequences," *AEA Ppaers and Procedings* **83**, 336-342.

[63] Shastry G., D. N. Weil (2002). "How Much of Cross-Country Income Variation is Explained by Health?," *working paper*, Brown University.

[64] Temple, J. (1998). "Robustness Tests of the Augmented Solow Model," *Journal of Applied Econometrics* **13**, 361-375.

[65] Temple, J. (1999). "The New growth Evidence," *Journal of Economic Literature* **37**, 112-156.

[66] UNAIDS/WHO. (2002). *AIDS Epidemic Update*.

[67] UNAIDS/WHO. (2002). *Epidemiological Fact Sheets on HIV/AIDS and Sexually Transmitted Infections*.

[68] Wachter, K.W., J.E. Knodel, and M. VanLandingham. (2003). "Parental Bereavement: Heterogenous Impacts of AIDS in Thailand," *Journal of Econometrics* **112**, 193-206.

[69] Waldmann, R. (1992). "Income Distribution and Infant Mortality," *The Quarterly Journal of Economics* **107**, 1283-1302.

[70] World Development Indicators (WDI). (2002).

[71] Woolridge, J.M. (2002). "Econometric Analysis of Cross Section and Panel Data," *Cambridge, MA: MIT Press*

[72] Young, A. (2004). "The Gift of the Dying: The Tragedy of AIDS and the Welfare of Future African Generations," working paper, Graduate School of Business, University of Chicago.

Appendix A

Data Used in the Extended Solow Model

Table A1: Data used in the extended Solow model

Country	PWT Code	Mean values for relevant variables				
		Y/L	I/Y	SCHOOL	$n+g+\delta$	AIDS
Algeria*	DZA	10005.4	13.65	0.0825	0.0811	0.1165
Angola	AGO	4360.1	6.35	0.0241	0.0759	3.5434
Argentina*	ARG	18742.5	15.89	0.0859	0.0647	2.6654
Australia	AUS	40452.0	23.98	0.1108	0.0633	2.8723
Austria*	AUT	36615.7	25.61	0.1075	0.0556	1.4293
Bangladesh	BGD	3046.7	10.30	0.0381	0.0709	0.0009
Belgium*	BEL	38061.8	23.13	0.1094	0.0515	1.6902
Benin	BEN	2406.2	7.19	0.0252	0.0795	5.4167
Bolivia*	BOL	5205.1	9.01	0.0646	0.0739	0.2169
Botswana	BWA	14769.7	17.38	0.0635	0.0790	57.0842
Brazil*	BRA	11723.9	17.34	0.0587	0.0716	7.4395
Burkina Faso*	BFA	2051.0	11.25	0.0073	0.0725	11.2315
Burundi	BDI	1248.1	6.07	0.0066	0.0698	27.4842
Cameroon	CMR	4321.1	6.64	0.0345	0.0772	10.8619

Notes: The sources for these data are Bernanke and Gürkaynak (2001), UNAIDS/WHO and PWT 6.1.

* denotes the 63 nations included in the sample used to carry out age-sepcific AIDS estimation.

Country	PWT Code	Mean values for relevant variables				
		Y/L	I/Y	SCHOOL	$n+g+\delta$	AIDS
Canada*	CAN	42080.2	24.97	0.1155	0.0614	3.0637
C.African Rep.	CAF	2357.0	5.11	0.0191	0.0708	20.3963
Chad*	TCD	1903.4	6.63	0.0108	0.0745	12.7695
Chile*	CHL	16137.4	18.79	0.0941	0.0657	1.7143
Colombia*	COL	9276.3	12.14	0.0834	0.0733	1.5264
Congo	COG	5024.4	7.48	0.1059	0.0771	168.5997
Costa Rica*	CRI	9391.8	16.04	0.0806	0.0776	3.4051
Denmark*	DNK	42759.9	22.52	0.1151	0.0532	2.4675
Dom. Rep.*	DOM	9089.1	13.43	0.0764	0.0731	4.2897
Ecuador*	ECU	6051.4	15.90	0.0917	0.0785	0.7835
Egypt*	EGY	7282.9	6.06	0.1082	0.0756	0.0295
El Salvador*	SLU	7778.1	7.85	0.0525	0.0732	3.2685
Ethiopia	ETH	1388.1	4.27	0.0179	0.0733	7.1639
Finland*	FIN	36433.6	24.42	0.1164	0.0525	0.3876
France*	FRA	36165.8	24.60	0.1065	0.0549	4.8720
Ghana*	GHA	2464.5	6.08	0.0678	0.0826	16.6795
Greece*	GRC	23087.6	21.53	0.0968	0.0556	1.2263
Guatemala*	GTM	8202.7	7.40	0.0350	0.0768	2.2228
Haiti	HTI	6235.0	5.31	0.0256	0.0724	8.1973
Honduras*	HND	3947.2	14.48	0.0503	0.0820	13.2563
Hong Kong*	HKG	38179.1	25.05	0.0859	0.0674	0.4939
India	IND	4360.6	12.35	0.0609	0.0710	0.0734

Notes: The sources for these data are Bernanke and Gürkaynak (2001), UNAIDS/WHO and PWT 6.1.

* denotes the 63 nations included in the sample used to carry out age-specific AIDS estimation.

Country	PWT Code	Mean values for relevant variables				
		Y/L	I/Y	SCHOOL	$n+g+\delta$	AIDS
Indonesia*	IDN	6263.5	17.76	0.0629	0.0717	0.0159
Ireland*	IRL	40520.7	19.79	0.1453	0.0616	1.0947
Israel*	ISR	30942.5	26.60	0.1163	0.0794	0.8832
Italy*	ITA	33816.6	22.27	0.0836	0.0528	4.5305
Jamaica*	JAM	5648.5	17.72	0.1233	0.0660	11.1127
Japan*	JPN	38057.5	32.56	0.1038	0.0531	0.0950
Jordan*	JOR	7490.8	15.15	0.1548	0.0998	0.1469
Kenya	KEN	2451.1	8.07	0.0417	0.0853	24.9535
Korea*	KOR	20719.5	36.29	0.1261	0.0644	0.0306
Madagascar*	MDG	1677.6	3.03	0.0383	0.0769	0.0211
Malawi	MWI	1591.9	7.92	0.0147	0.0735	40.9708
Malaysia	MYS	15251.6	26.56	0.0906	0.0777	1.6425
Mali	MLI	1995.9	8.23	0.0162	0.0730	3.7066
Mauritania	MRT	2984.3	8.70	0.0201	0.0779	2.0821
Mauritius*	MUS	21132.0	12.52	0.0808	0.0643	0.4024
Mexico*	MEX	15629.6	17.49	0.0953	0.0759	2.9271
Morocco*	MAR	7024.9	11.95	0.0547	0.0746	0.2073
Mozambique	MOZ	2107.5	3.41	0.0112	0.0672	9.8234
Netherlands*	NLD	37847.2	22.58	0.1226	0.0564	1.8466
New Zealand*	NZL	30608.2	22.20	0.1223	0.0605	1.1704
Nicaragua*	NIC	3584.3	12.41	0.0775	0.0810	0.4314

Notes: The sources for these data are Bernanke and Gürkaynak (2001), UNAIDS/WHO and PWT 6.1.

* denotes the 63 nations included in the sample used to carry out age-specific AIDS estimation.

Country	PWT Code	Mean values for relevant variables				
		Y/L	I/Y	SCHOOL	$n+g+\delta$	AIDS
Niger*	NER	1875.0	4.61	0.0091	0.0816	4.2395
Nigeria	NGA	1592.5	9.39	0.0330	0.0778	3.1480
Norway*	NOR	49423.1	28.65	0.1129	0.0555	0.9070
Pakistan*	PAK	3956.5	11.14	0.0359	0.0736	0.0112
Panama*	PAN	10528.0	18.78	0.1079	0.0736	7.7935
Papua N.G.*	PNG	5778.8	10.35	0.0218	0.0762	1.5274
Paraguay*	PRY	8423.9	12.70	0.0558	0.0800	0.6948
Peru*	PER	7767.1	17.62	0.1068	0.0747	2.3352
Philippines	PHL	6896.7	14.36	0.1239	0.0754	0.0420
Portugal*	PRT	25241.1	23.10	0.0836	0.0538	4.8888
Rwanda	RWA	1839.0	4.64	0.0101	0.0773	18.5401
Senegal	SEN	3161.3	6.71	0.0258	0.0766	2.5547
Sierra Leone	SLE	1388.0	4.85	0.0258	0.0701	0.5959
Singapore*	SGP	40393.7	42.45	0.0971	0.0741	1.3665
Spain*	ESP	27861.2	24.47	0.1157	0.0553	8.4116
Sri Lanka*	LKA	5695.3	12.34	0.1030	0.0677	0.0467
Sweden*	SWE	38254.8	21.12	0.0960	0.0535	1.1200
Switzerland*	CHE	41885.1	27.79	0.0946	0.0562	5.6556
Syria*	SYR	7742.7	9.17	0.1052	0.0875	0.0360
Tanzania*	TZA	932.4	16.46	0.0079	0.0815	26.0605
Thailand*	THA	9858.3	32.98	0.0570	0.0685	17.0469

Notes: The sources for these data are Bernanke and Gürkaynak (2001), UNAIDS/WHO and PWT 6.1.

* denotes the 63 nations included in the sample used to carry out age-specific AIDS estimation.

Country	PWT Code	Mean values for relevant variables				
		Y/L	I/Y	SCHOOL	$n+g+\delta$	AIDS
Togo*	TGO	1760.4	8.12	0.0425	0.0782	21.9104
Tri.&Tobago*	TTO	20072.5	9.39	0.1175	0.0642	21.9104
Tunisia	TUN	11064.1	13.26	0.0695	0.0758	0.4423
Turkey*	TUR	11548.5	18.80	0.0740	0.0716	0.0376
Uganda	UGA	2132.7	13.65	0.0172	0.0753	19.1190
UK*	GBR	37153.1	18.77	0.0998	0.0531	1.6040
Uruguay*	URY	16503.9	10.76	0.0907	0.0565	2.8308
USA*	USA	53979.1	21.29	0.1163	0.0603	14.8092
Venezuela	VEN	11757.8	14.30	0.0686	0.0771	2.6470
Zambia	ZMB	1664.6	8.94	0.0367	0.0774	39.7673
Zimbabwe	ZWE	5053.0	13.49	0.0577	0.0752	55.4721

Notes: The sources for these data are Bernanke and Gürkaynak (2001), UNAIDS/WHO and PWT 6.1.

* denotes the 63 nations included in the sample used to carry out age-specific AIDS estimation.

Appendix B

AIDS Definition

In a meeting convened in Geneva by the WHO Global Programme on AIDS (1994) was suggested the following: the 1985 provisional WHO clinical case definition for AIDS ("Bangui definition") to be referred to as the WHO AIDS surveillance case definition and it was introduced an expanded WHO AIDS surveillance case definition. (Weekly Epidemiological Record, 1994, issue 69, pp. 273-280).

1. WHO case definition for AIDS surveillance

For the purposes of AIDS surveillance an adult or adolescent ($>$ 12 years of age) is considered to have AIDS if at least 2 of the following major signs are present in combination wit hat least 1 of the minor signs listed below, and if these signs are not known to be due to a condition unrelated to HIV infection.

Major signs

- weight loss 10% of body weight
- chronic diarrhoea for more than 1 month
- prolonged fever for more than 1 month (intermittent or constant)

Minor signs

- persistent cough for more than 1 month
- generalized pruritic dermatitis
- history of herpes zoster
- oropharyngeal candidiasis
- chronic progressive or disseminated herpes simplex infection generalized lymphadenopathy

The presence of either generalized Kaposi sarcoma or cryptococcal meningitis is sufficient for the diagnosis of AIDS for surveillance purposes.

2. Expended WHO case definition for AIDS surveillance

For the purposes of AIDS surveillance an adult or adolescent (> 12 years of age) is considered to have AIDS if a test for HIV antibody gives a positive result, and 1 or more of the following conditions are present:

- 10% body weight loss or cachexia, with diarrhoea or fever, or both, intermittent or constant, for at least 1 month, not known to be due to a condition unrelated to HIV infection
- cryptococcal meningitis
- pulmonary or extra-pulmonary tuberculoses
- Kaposi sarcoma
- neurological impairment that is sufficient to prevent independent daily activities, not known to be due to a condition unrelated to HIV infection (for example, trauma or cerebrovascular accident)
- candidiasis of the oesophagus (which may be presumptively diagnosed based on the presence of oral candidiasis accompanied by dysphagia)
- clinically diagnosed life-threatening or recurrent episodes of pneumonia, with or without etiological confirmation
- invasive cervical cancer

Appendix C

Constructing the AIDS Cases by Age

- The officially reported AIDS cases for the different age groups are reported as a total before 1997 and annually for 1997, 1998, 1999 and 2000.

- In addition to the officially reported cases per age group, UNAIDS/WHO also reports "Not specified/unknown cases" (NS).

- Although the data in the OECD countries have very few NS cases, the data in many low-income countries like sub-Saharan Africa countries contain a lot of NS cases.

- We can not use NS cases in our calculation of the age groups and recognize that this is a source of measurement error due to aggregation.

- We chose aggregate AIDS cases into four age-group samples as follows: AIDS[0-4] (*infancy period*), AIDS[5-15] (*schooling period*), AIDS[16-34] (*productive period*) and AIDS[35-60+] (*less productive period*).

- We divide the total number of reported AIDS cases in each age group by the number of years cases are reported and multiply by 100,000 and divide by average population. This the mean AIDS cases reported per 100,000 people by each of the four age groups.

- Data on population are taken for the WDI (2002). We start from the year, during which an AIDS case was reported till 2000.

Appendix D

Excluding Potential Outliers

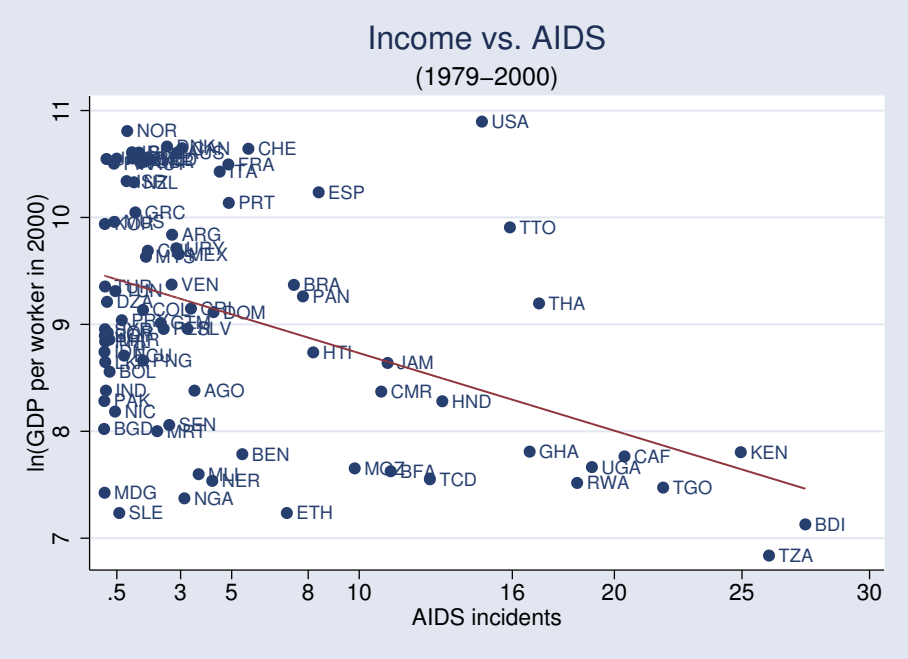

Figure D.1: Cross-country correlation between income and AIDS. The plot includes 84 countries. We exclude Botswana, Congo, Malawi, Zimbabwe, Zambia with very high AIDS incidents.

Table D1:Cross-country regressions

Dependent variable: ln(GDP per worker in 2000)			
Specification	Extended Solow model with AIDS (PWT 6.1 – WHO 2002)		
	Non-oil	OECD	Non-OECD
Constant	4.5334***	10.0434***	5.8110***
	(0.9542)	(2.0069)	(1.5857)
$\ln s_{ik}$	0.6092***	0.5142	0.5874***
	(0.1267)	(0.3173)	(0.1386)
$\ln(n_i + g + \delta)$	−2.7933***	−1.3294**	−2.2245***
	(0.3017)	(0.5799)	(0.6147)
$\ln s_{ih}$	0.5575***	1.2162***	0.5078***
	(0.0945)	(0.2401)	(0.0991)
AIDS	−.0141	0.0247	−0.0188**
	(0.0094)	(0.0174)	(0.0094)
Adj. R^2	0.86	0.66	0.75
Obs.	84	21	63

Notes: Standard errors are in parentheses. It is assumed that $g + \delta = 0.05$ as in MRW. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level.

Appendix E

Countries in Three Regimes

Table E1: Countries in three regimes

Regime 1		Regime 2		Regime 3
Angola	Kenya	Algeria	Mauritania	Bangladesh
Benin	Malawi	Argentina	Mauritius	Egypt
Botswana	Mali	Australia	Mexico	India
Brazil	Mozambique	Austria	Morocco	Indonesia
Burkina Faso	Niger	Belgium	Netherlands	Korea
Burundi	Nigeria	Bolivia	New Zealand	Madagascar
Cameroon	Panama	Chile	Nicaragua	Pakistan
C. Afr. Rep.	Portugal	Canada	Norway	Philippines
Chad	Rwanda	Columbia	Papua N.G.	Sri Lanka
Congo	Spain	Denmark	Paraguya	Syria
Costa Rica	Switzerland	Ecuador	Peru	Turkey
Dom. Rep.	Tanzania	Finland	Senegal	
El Salvador	Thailand	Greece	Sierra Leone	
Ethiopia	Togo	Guatemala	Singapore	
France	Tri.&Tobago	Hong Kong	Sweden	
(39)		(39)		(11)

Regime 1		Regime 2		Regime 3
Ghana	Uganda	Ireland	Tunisia	
Haiti	USA	Israel	UK	
Honduras	Zambia	Japan	Uruguay	
Italy	Zimbabwe	Jordan	Venezuela	
Jamaica		Malaysia		
(39)		(39)		(11)

Appendix F

Growth Regressions

Table F1: Growth regressions for the full sample and OECD and non-OECD subsamples

Dependent variable: Growth GDP per worker (initial-2000)						
Specification	Extended Solow Model (PWT 6.1)			Extended Solow Model with AIDS ($PWT6.1 - WHO$ 2000)		
	Non-oil	OECD	Non-OECD	Non-oil	OECD	Non-OECD
Constant	1.8918	2.7079	1.9609	1.9513	2.9759	2.0686
	(1.6168)	(1.6568)	(2.3318)	(1.6465)	(1.8789)	(2.3836)
$\ln y_{i0}$	-0.4544^{**}	-0.1285	-0.4748^{**}	-0.4600^{**}	-0.1758	-0.4823^{**}
	(0.1976)	(0.1565)	(0.2111)	(0.2016)	(0.1596)	(0.2165)
$\ln s_{ik}$	0.4606^{***}	-0.2290	0.4677^{***}	0.4585^{***}	-0.1776	0.4649^{***}
	(0.1568)	(0.2180)	(0.1546)	(0.1570)	(0.2261)	(0.1547)
$\ln(n_i + g + \delta)$	-1.6132^{***}	-0.1404	-1.6480^{***}	-1.6133^{***}	-0.2232	-1.6373^{***}
	(0.3948)	(0.4732)	(0.4601)	(0.3972)	(0.4031)	(0.4602)
$\ln s_{ih}$	0.3058^{**}	0.6203^{***}	0.3010^{**}	0.3092^{**}	0.6333^{**}	0.3056^{**}
	(0.1337)	(0.2013)	(0.1394)	(0.1365)	(0.2395)	(0.1430)
AIDS				-0.0008	0.0176^{**}	-0.0009
				(0.0013)	(0.0073)	(0.0015)
Adj. R^2	0.50	0.36	0.45	0.50	0.52	0.45
Obs.	89	21	68	89	21	68

Notes: Standard errors are in parentheses. It is assumed that $g + \delta = 0.05$ as in MRW. All regressions are estimated using OLS. White's heteroskedasticity correction was used. *** Significantly different from 0 at the 1% level. ** Significantly different from 0 at the 5% level. * Significantly different from 0 at the 10% level. Investment and population growth rates are averages for the period 1979-2000. s_h is the average percentage of the working-age population in secondary school for the period 1970-1995.

Appendix G

Data Used in the Infant Mortality-Female Human Capital Inequality Equation

Table G1: Data used in the infant mortality-female human capital inequality equation

Country	Code	Values for relevant variables						
		Infant	GiniF	Y/L_{1960}	Y/L_{2000}	Phys.	SchoolF	Malaria
Argentina	ARG	35	0.25	8711.3	12790.55	3	6.96	0.09
Australia	AUS	12	0.16	12593.15	28479.77	2	9.93	0
Austria	AUT	17	0.26	8249.95	25820.15	2	6.59	0
Bagladesh	BGD	115	0.80	1329.38	2174.65	0	0.91	1
Belgium	BEL	15	0.21	8815.76	25233.67	3	8.27	0
Bolivia	BOL	116	0.57	2995.62	3360.68	0	4.23	0.34
Botswana	BWA	77	0.49	1257.05	4391.11	0	3.54	0.76
Brazil	BRA	70	0.45	3032.1	8609.03	1	3.49	0.89
Cameroon	CMR	107	0.69	2107.24	2592.84	0	1.86	1
Canada	CAN	13	0.29	12475.1	29408.37	2	10.14	0
C. Afri. Rep.	CAF	122	0.87	2697.12	2230.58	0	0.84	1
Chile	CHL	35	0.28	4798.46	11531.51	1	6.18	0

Notes: The sources for these data are WDI(2000), Barro&Lee(2000), PWT 6.1, and Gallup, Mellinger and Sachs' Geography dataset.

Country	Code	Values for relevant variables						
		Infant	GiniF	Y/L_{1960}	Y/L_{2000}	Phys.	SchoolF	Malaria
Colombia	COL	49	0.40	3291.72	7028.28	1	4.27	0.74
Costa Rica	CRI	34	0.33	4556.99	7382.88	1	5.02	0.21
Denmark	DNK	11	0.25	12576.14	29214.81	3	8.83	0
Dom. Rep.	DOM	77	0.49	2213.66	6269.87	1	3.71	1
El Salvador	SLV	79	0.50	4272.25	5655.84	0	3.18	0.98
Finland	FIN	8	0.21	8833.28	26137.43	2	7.51	0
France	FRA	13	0.26	9012.38	24837.32	3	6.34	0
Ghana	GHA	92	0.77	1114.3	1743.42	0	1.72	1
Greece	GRC	20	0.34	4805.43	16211.37	3	5.74	0
Guatemala	GTM	79	0.65	3044.46	4686.98	0	2.16	0.83
Honduras	HND	78	0.51	2202.64	2619.72	0	2.80	0.27
Hong Kong	HKG	12	0.44	3885.03	28985.27	1	6.53	0.5
India	IND	102	0.78	1057.29	3029.63	0	2.03	0.38
Indonesia	IDN	90	0.56	1170.83	4309.68	0	2.78	0.91
Ireland	IRL	14	0.21	6077.69	29673.53	2	7.78	0
Israel	ISR	17	0.29	6757.7	19731.21	3	8.38	0
Italy	ITA	19	0.32	7870.53	23409.35	4	5.55	0
Jamaica	JAM	38	0.25	3466.06	4398.9	0	4.26	0
Japan	JPN	10	0.20	5352.21	26607.24	2	8.07	0
Jordan	JOR	39	0.65	2938.15	4764.41	1	3.55	0
Kenya	KEN	86	0.63	1057.9	1660.26	0	2.20	1

Notes: The sources for these data are WDI(2000), Barro&Lee(2000), PWT 6.1, and Gallup, Mellinger and Sachs' Geography dataset.

Country	Code	Values for relevant variables						
		Infant	GiniF	Y/L_{1960}	Y/L_{2000}	Phys.	SchoolF	Malaria
Korea, Rep.	KOR	32	0.37	1890.55	17871.16	1	6.66	-
Malawi	MWI	161	0.61	543.02	1051.85	0	1.80	1
Malaysia	MYS	30	0.53	2732.36	11881.36	0	3.89	0.88
Mali	MLI	169	0.94	1254.45	1266.79	0	0.31	0.80
Mauritania	MRT	130	0.67	1335.74	1980.26	0	1.85	0.78
Mexico	MEX	56	0.42	5157.89	10517.05	1	4.51	0.13
Mozambique	MOZ	156	0.77	1982.94	1220.98	0	0.40	1
Nepal	NPL	129	0.92	962.16	1916.18	0	0.47	0.58
Netherlands	NLD	10	0.18	10876.95	26779.49	2	7.62	0
N. Zealand	NZL	14	0.19	13810.97	21675.12	2	10.70	0
Nicaragua	NIC	82	0.56	3783.31	2262.5	0	3.19	0.13
Niger	NER	152	0.93	2054.86	1147.25	0	0.31	0.77
Norway	NOR	10	0.17	9463.86	30064.78	2	8.48	0
Pakistan	PAK	123	0.85	810.79	2373.3	0	1.33	0.80
Panama	PAN	37	0.34	2972.48	7183.22	1	6.37	0.89
Papua N.G.	PNG	92	0.69	2728.78	3911.93	0	1.42	0.95
Paraguay	PRY	45	0.31	3148.7	5870.3	1	4.73	1
Peru	PER	81	0.48	4118.79	5509.87	1	4.89	0.53

Notes: The sources for these data are WDI(2000), Barro&Lee(2000), PWT 6.1, and Gallup, Mellinger and Sachs' Geography dataset.

Country	Code	Values for relevant variables						
		Infant	GiniF	Y/L_{1960}	Y/L_{2000}	Phys.	SchoolF	Malaria
Philippines	PHL	63	0.36	2633.35	4290.72	0	6.14	0.79
Portugal	PRT	33	0.51	4014.21	17372.31	2	3.38	0
Senegal	SEN	114	0.75	1338.46	1555.28	0	1.50	1
Sierra Leone	SLE	191	0.87	2756.36	12319.64	0	1.06	1
Singapore	SGP	14	0.49	6205.21	9009.19	1	4.88	0
Spain	ESP	18	0.28	5374.52	19526.76	3	5.28	0
Sri Lanka	LKA	34	0.39	1696.02	4135.5	0	4.90	0.19
Sweden	SWE	8	0.24	11425.35	25994.72	3	9.19	0
Switzerland	CHE	11	0.24	16985.64	28795.71	2	8.81	0
Syria	SYR	66	0.70	1803.3	5126.3	1	2.52	0.23
Thailand	THA	54	0.37	1412.79	7888.54	0	4.41	0.90
Togo	TGO	113	0.76	1140.31	1121.38	0	1.08	1
Tr.&Tobago	TTO	34	0.25	5569.74	12713.71	1	6.31	0
Tunisia	TUN	69	0.70	2546.42	8021.32	1	2.04	0.76
Turkey	TUR	93	0.62	3385.51	8031.86	1	2.54	0.31
Uganda	UGA	110	0.64	729.18	1233.64	0	1.45	1
U. K.	GBR	13	0.19	10947.38	24535.04	2	8.35	0
USA	USA	15	0.23	14527.6	37255.59	2	10.64	0
Uruguay	URY	31	0.31	6823.21	10989.42	2	6.54	0
Venezuela	VEN	38	0.41	10188.71	7726.34	1	4.65	0.28
Zambia	ZMB	108	0.53	1557.93	1152.75	0	2.95	1
Zimbabwe	ZWE	79	0.49	1595.46	3191.33	0	2.60	1

Notes: The sources for these data are WDI(2000), Barro&Lee(2000), PWT 6.1, and Gallup, Mellinger and Sachs' Geography dataset.

Appendix H

Countries in Geographic Regions

SouthEast Asia & Pacific

Afghanistan, Australia, Bangladesh, China, Fiji , Hong Kong , India, Indonesia, Japan, Korea, Malaysia, Myanmar (Burma), Nepal, New Zealand, Pakistan, Papua New Guin., Singapore, Sri Lanka, Thailand

Europe

Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, West, Greece, Hungary, Iceland, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom, Yugoslavia

Latin America&Caribbean

Argentina, Barbados, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Rep., Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Trinidad & Tob., Uruguay, Venezuela

Middle East & North America

Algeria, Bahrain, Egypt, Iran, I.R. of Iraq, Israel, Jordan, Kuwait, Syria, Tunisia

North America

Canada, U.S.

Sub-Suharan Africa

Benin, Botswana, Cameroon, Central Afr. R., Congo, Gambia, Guinea-Bissau, Kenya, Lesotho, Liberia, Malawi, Mali, Mauritania, Mauritius, Mozambique, Niger, Rwanda, Senegal, Sierra Leone, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zaire, Zambia, Zimbabwe

Vita

Petia Stoianova Stoytcheva was born on September 17, 1977 in Yambol, Bulgaria. She attended Mathematical high school in Yambol, and studied finance at the University of National and World Economy, Sofia, Bulgaria. She completed a Bachelor of Science degree in finance in June of 2000 and obtained her Master of Science degree in economics at Louisiana State University in December of 2002. Her Doctor of Philosophy degree in economics will be awarded in December of 2005.