HIV/AIDS and Economic Growth: An Econometric Analysis with Particular Consideration of the Role of Education Capital Accumulation

Duncan Maclaine

School of Economics, University of Queensland, Australia
Australian Centre for Economic Research on Health (ACERH)

Abstract: This paper investigates, using a panel data approach, the effects of the HIV/AIDS epidemic on economic growth using a neoclassical growth model that incorporates human capital in the form of both health capital and a Mincerian specification of education capital that encompasses both quantitative and qualitative components. Two important innovations of my analysis are that: (i) it represents the first cross-country econometric assessment that pays particular attention to human capital accumulation through education, and (ii) it implements the dynamic panel system-GMM estimator that is known to be superior to alternative estimators that underlie previous econometric assessments of the epidemic. Based on a sample spanning 45 years and 142 countries, results indicate that the epidemic’s effects on growth have been large and that a material component of this effect is due to its detrimental impact on the accumulation of education capital. For a full world sample and a sub-sample of developing world countries, the impact of a 1% increase in adult HIV prevalence is estimated to be a reduction in income per capita of around 0.14% and 0.12% on average respectively, which is substantially larger than that found in previous econometric assessments of the epidemic.

JEL Classification: O11; O15; O47; I12; I20.
Keywords: HIV; AIDS; Growth; Human Capital; Health; Education

1 School of Economics, University of Queensland, Brisbane, QLD 4072, Australia; Telephone +61 7 33654028. Email: d.maclaine@uq.edu.au
2 Financial support from a University of Queensland Joint Research Scholarship and an ACERH research scholarship are gratefully acknowledged.
Preface

Title of thesis: Essays in applied health economics

Supervisors: Assoc. Prof. Luke Connelly (40%)
Dr. Alicia Rambaldi (40%)
Dr. Kam Ki Tang (20%)

The following paper represents a shortened version of the first of three essays that will comprise my thesis dissertation.
1. Introduction

It is now twenty-seven years since the first clinical evidence of AIDS emerged. The death and suffering caused by the epidemic during these years represents a monumental human tragedy. This paper seeks to determine if we are also witnessing the unfolding of an economic tragedy.

It is estimated that between 34 and 47 million people worldwide were living with HIV in 2006 (UNAIDS, 2006). Prevalence rates of HIV worldwide appear, on average, to be stabilizing, with some of the most affected regions in Sub-Saharan Africa experiencing some declines, although this trend is neither widespread nor particularly strong (UNAIDS, 2006). Table 1, included in Appendix A, details prevalence estimates for the years 2005 and 2003. Swaziland continues to experience the worst epidemic, with one-third of 15 to 49 year-olds infected with HIV. Outside Africa, Haiti has the highest prevalence rate of 3.8%. Although not featured in Table 1, India has the largest absolute number of adult HIV cases in the world, approximately 5.6 million, despite a prevalence rate of less than 1%, due to the sheer size of its population. Readers who are interested in a detailed assessment of the epidemic are referred to the comprehensive report published by UNAIDS in conjunction with the WHO entitled ‘Report on the Global AIDS Epidemic 2006.’

A small number of recent econometric studies have investigated the role HIV/AIDS has on determining cross-country differences in per capita income and / or its role in the evolution of a country’s economic performance over time. The analysis presented in this paper represents the first cross-country econometric analysis that pays particular attention to the role HIV/AIDS plays in human capital accumulation through education. This innovation is important because, to date, this impact channels has been investigated largely through calibration and simulation of theoretical overlapping generations (OLG) or computable general equilibrium (CGE) modelling, with such techniques open to the criticism that they rely heavily on assumptions that can often be difficult to justify. A second innovation of this paper is the adoption of a Mincerian approach to specifying and incorporating human capital into empirical growth models, an approach that is becoming increasingly popular in the growth literature. Careful consideration of how schooling participation, education quality and health translate into human capital, and in turn how human capital should enter the macroeconomic production function, is considered critical to isolating the economic impact of

---

3 Prevalence estimates for 2005 and all previous years have been substantially revised downwards by UNAIDS/WHO as a result of improved estimation methods. For example, in Swaziland, HIV prevalence was previously estimated to be 38.8% in 2003, whilst latest estimates for 2003 are 32.4%. For Cambodia, previous 2003 prevalence estimates were 2.6%, whilst the updated estimate for 2003 is 2.0%.

4 This report is available at www.unaids.org
the HIV/AIDS epidemic. A third innovation of this paper is the use of the dynamic panel system-GMM estimator which is considered a-priori superior to the more traditional difference-GMM estimator under the precise circumstances here, casting doubt over the accuracy of earlier estimates of the macroeconomic impact of HIV/AIDS obtained from growth regressions that have employed the difference-GMM estimator.

The role HIV/AIDS plays in the accumulation of education capital has been the focus of much of the very recent economic literature. For example, Ferreira and Pessoa (2003), Corrigan et al. (2005) and Bell et al. (2006) all highlight, through the use of calibrated overlapping-generations models, the potentially large negative impact HIV/AIDS may have on human capital accumulation through education and in turn on future income per capita. One might expect that as the body of literature assessing the economic growth implications of HIV/AIDS grows we might be converging towards a consensus view, but in fact, the opposite appears to be occurring, with some studies predicting virtual economic collapse (e.g. Bell et al., 2006), whilst others suggesting that future income per capita could in fact rise as a consequence of the epidemic (Young, 2005). In the light of such contrasting assessments, it is considered a worthy exercise to conduct a rigorous empirically investigation of the impact HIV/AIDS has on welfare by carefully considering the role of human capital accumulation.

2 Literature

A summary of various literatures, relevant to the task at hand, is presented here.

2.1 The concept of human capital and role in economic growth

Modern human capital theory, to which Shultz and Becker were the main contributors, sees education and skills as capital goods for which acquisition is costly and where individuals invest in it to obtain productive skills, enabling them to increase productivity and income. In the context of a neoclassical macroeconomic production function, this could be considered a ‘direct effect’, where education is included as an additional input into the production process.\(^5\) Causality may in fact run in

\(^5\) A number of researchers also stress indirect effects of education on growth. Education can lead to improvements in technology as a result of innovation and research, improved processes and products or an increased ability for countries to adopt technological innovations from elsewhere (Romer, 1990; Aghion and Howitt, 1998; Nelson and Phelps, 1966; Hall and Jones, 1999). Macroeconomic models emphasizing this indirect role are typically referred to as endogenous growth models.
both directions, with higher incomes providing greater resources for the individual to invest in their own or their children’s education.

Health is another important component of human capital. Grossman (1972) was the first to construct a theoretical model of the demand for health capital. Grossman described the demand for health and health care via the theory of human capital. Healthy workers are more physically and mentally able and take less time off work due to illness and as a consequence – are more productive. Health is therefore viewed as a durable capital stock that generates production benefits (increased incomes). Again, causality may not just be in one direction, with higher incomes increasing the resources available for individuals to invest in their own or children’s health.

Health and education should not be viewed as mutually exclusive components of human capital – they are interlinked in many ways. Grossman (1972) suggests that individuals with higher levels of education are more efficient in addressing their health needs, whilst Schultz (1999) stresses that better health enables an individual to utilize better the knowledge and skills they acquired through education. Improved health, by increasing life expectancy, can increase the incentive to invest in one’s own education due to the greater possible returns from the investment (Barro, 1996; Ferreira and Pessoa, 2003). Corrigan et al. (2005) highlight generational linkages between health and education by arguing that the children of adults, who drop out of the workforce due to ill-health, may enter the labour-force prematurely, thereby forgoing their education to maintain the household’s income.

### 2.2 Human Capital and Neoclassical Empirical Growth Models

In the traditional neoclassical growth model (Solow, 1956), income is modelled as a function of physical capital and labour only, and human capital variables are not considered. This model did not do a very good job at explaining income disparities between countries. In their seminal 1992 paper, Mankiw, Romer and Weil, henceforth MRW (1992), proposed adding human capital in the form of education (henceforth referred to as education capital) as an additional separate factor of production. MRW (1992) concluded that this modification removed the bias in the coefficient estimates of physical capital and population growth that would otherwise be present and that education capital helps to explain cross-country income differences. This ‘Augmented Solow model’ allows for the ‘direct effect’ of education on growth discussed above.
Islam (1995) extended the MRW (1992) framework by introducing a panel data approach. The approach did not reveal a significant relationship between education capital and growth. This was a worrying result, one that continues to plague growth models estimated with panel data, given the wealth of microeconomic evidence of the significant returns to education capital. Knowles and Owen (1995) extended the MRW (1992) model by including health capital as well as education capital. Their results confirm the importance of health for economic growth, however the relationship between education capital and growth is found to be insignificant. This result is common in studies that incorporate both health and education capital in the production function, including those investigating the role of HIV/AIDS on economic growth discussed in Section 2.4.

In a comprehensive critical summary of the empirical literature on the macroeconomic returns to education, Sianesi and van Reenen (2003) highlight a number of factors that may have contributed to the often disappointing results when a proxy for education capital is incorporated into empirical neoclassical growth models of the type highlighted above. These include; the selection of poor or inappropriately narrow choice of proxy, data measurement error, and other methodological problems such as not controlling for potential reverse causality.\(^6\)

With regards to the selection of proxies for education capital, often a single measure such as enrolment ratios is used. For example, two of the studies investigating the economic impact of HIV/AIDS, MR (2006) and Tandon (2005), use the secondary school enrolment ratio. This proxy potentially misses the growth impact of improved primary education enrolment rates for less developed countries that have yet to approximate universal primary school attendance, and also the growth effects of higher education for advanced countries that approximate universal secondary school attendance. There is also increasing evidence that one needs to account for the quality, as well as the quantity of schooling.\(^7\) A discussion of my selection of, and justification for, proxies for education and health capital is provided in section 3.1.

### 2.3 HIV/AIDS and the Economy

A review of the literature involving the assessment of the macroeconomic impact of HIV/AIDS highlights that any estimated effect of HIV/AIDS differs according to (i) the economic framework and modelling approach utilised, (ii) the country or countries of focus, (iii) the impact channels that

---

\(^6\) Soto (2006) also provides an excellent updated summary of the many empirical studies that have failed to detect significant impacts of education capital on the level or growth rate of GDP per capita.

\(^7\) See for example, Hanushek and Kimko (2000), Barro (2001) and Gundlach et. al. (2002).
are accommodated and the assumptions underlying the magnitude of these, (iv) the time frame of the analysis, and (v) the epidemiological estimates underlying the economic projections.

Impact channels of HIV/AIDS

A summary of the individual channels in which HIV/AIDS can affect economic outcomes is presented here along with a brief discussion of some of the more influential research in the macroeconomic literature. The purpose of this section is not to report in extensive detail specific results from studies, but rather to highlight in general the impact channels considered by them and their broad conclusions. It is, however, worthwhile to present a more detailed discussion of the recent wave of studies that consider the role HIV/AIDS plays in education capital accumulation.

A - Productivity of HIV/AIDS workers: Workers with HIV/AIDS, and especially those without access to antiretroviral therapies, are likely to be less productive in the workplace and take more days off work due to illness. Depending on the structure of the labour market, this may result in lower wages for those individuals.

B - Savings and consumption behaviour of households directly affected by HIV/AIDS: The saving rates of households affected by HIV/AIDS may fall due to the need to spend more of their income on health care. Additionally, savings will decline if household wages have fallen due to the individual being less productive at work or dropping out of the workforce completely.

C - Average skill-set of the workforce: As HIV/AIDS victims drop out of the workforce, their replacements, assuming the existence of available labour, are likely to bring less experience and skills into the role.

D - Population and labour supply: Deaths due to HIV/AIDS will lead to a lower population than in the absence of the epidemic. The size of the labour force will also decline due to morbidity and mortality of working age individuals. Additionally, individuals may drop out of the labour force to care for sick spouses or relatives. There may also be absenteeism due to employee attendance of funerals (Liu et al., 2004). The proportional reduction in the size of the labour force as a result of HIV/AIDS may be larger than the reduction in overall population due to the epidemic disproportionately affected young adults.

Those studies involving cross-country econometric analysis are discussed separately in Section 2.4
The above four linkages between HIV/AIDS and macroeconomic outcomes were the main impact channels addressed in the first wave of studies that emerged in the early 1990s. Early influential studies by Cuddington (1993a, 1993b), Cuddington and Hancock (1994, 1995) and Cuddington (1993b) utilised neoclassical Solow models to analyse the effect of HIV/AIDS on growth in Tanzania and Malawi. Assumptions were made about the share of medical expenses paid out of household savings, demographic projections, how less productive an HIV/AIDS worker is relative to a healthy worker, the existence of surplus labour, and the reduction in average experience embodied in the workforce. The models were calibrated and simulated over a range of plausible assumptions. The general conclusion was that there would be a modest negative impact on GDP growth per capita.\textsuperscript{9} Another early study by Over (1992), considered a dual-sector economy, with the labour market disaggregated by productivity and between urban and rural sectors. Applying this framework to thirty African countries over the period 1990-2025, a modest negative impact on GDP per capita was found. Kambou et. al. (1992), introducing HIV/AIDS as an exogenous shock into a detailed CGE model for Cameroon, estimated that GDP growth rates between 1987 and 1991 were halved as a consequence.

In retrospect, these early studies suffer from deficiencies in relation to the severe underestimates of the scale of the epidemic that were employed and that they were arguably narrow in the range of economic impact channels that were incorporated.

\textbf{E - Public sector}: Public sector health care costs will most likely rise, placing pressure on the government budget. A comprehensive analysis of other pressures on the public sector as highlighted by Haacker (2002a) identified increased social security costs, increased death benefits and sick leave in the public service, and the need to increase the number of school teachers recruited and trained due to deaths in the profession. Because HIV/AIDS disproportionately affects the working age population, the government’s tax base will decline and increase the dependency ratio as mortality and morbidity in the workforce increases.

\textbf{F - Business operating environment, total factor productivity and technological progress}: In a review of published microeconomic studies, Liu et al. (2004) concluded that although the literature is

\textsuperscript{9} Where it was assumed that HIV/AIDS would have no impact on household savings rates, the models produced a small positive effect on GDP per capita.
far from satisfactory, there is a consensus view that HIV/AIDS has increased business costs due to, among others things, increased sick leave benefits, higher life and health insurance premiums (if borne by the employer), pensions, funeral benefits, lower productivity, and the higher costs associated with recruiting and training new staff to replace sick employees.

Arndt and Lewis (2000) utilise a disaggregated dynamic CGE model to estimate the impact of HIV/AIDS on economic growth prospects in South Africa between 2001 and 2010. In addition to impact channels A through D, the analysis incorporates the role the epidemic plays in the determination of total factor productivity (TFP) in the form of increased hiring and training costs, staff absenteeism, and a slowdown in technological adoption. Arndt and Lewis (2000) also undertake a more detailed treatment of the impact HIV/AIDS may have on the government sector. Incorporation of a time dimension, absent from the CGE modeling of Kambou et al. (1992), allowed for cumulative effects of lower investment and TFP growth to be captured. A significantly greater negative economic impact of HIV/AIDS on GDP and GDP per capita is found relative to earlier studies due to a combination of more up-to-date demographic projections and also the significant role TFP and the government sector plays in driving the results.

_G - International competitiveness and investment:_ In the event that HIV/AIDS leads to reductions in worker productivity, a slowing in the rate technological innovation, or rises in business operating costs, a country that experiences a more severe epidemic than its competitors could lose any comparative advantage it held, resulting in a reduction in exports and foreign direct investment (FDI). For an open economy, any increase in the capital to-labour ratio as a result of the epidemic could lead to a decline in the return to capital, discouraging domestic investment, increasing capital outflows, and reducing FDI (Haacker, 2002a).

Two influential studies focusing on Botswana by MacFarlan and Sgherri (2001) and BIDPA (2000) use calibrated dual-sector variants of neoclassical growth models to estimate medium to long-term growth prospects. Both papers consider impact channels A through E as well as F and G in the form of TFP growth and the role of international capital flows. MacFarlan and Sgherri (2001) undertake a range of simulations with varying assumptions that produce negative impacts which are broadly similar in magnitude to the South African study by Arndt and Lewis (2000), with the largest impacts occurring due to reduced labour productivity and the cumulative effects of slower capital
accumulation due to greater HIV/AIDS-related health care expenditure. The analysis undertaken by BIDPA (2000) produces significantly smaller estimates of the epidemic’s impact on GDP growth and virtually no effect on GDP per capita. This difference is mainly driven by the assumption in BIDPA (2000) that any reduction in domestic savings will be offset by increased capital inflows into the country. This is quite a strong assumption as it is quite likely that although some of the domestic shortfall in savings may be offset by inflows, capital flight should be considered a distinct possibility especially if the epidemic severely impacts the availability and reliability of the skilled labour pool.

I - Incentives to invest in skills and training: An employer facing high levels of staff turnover due to the prevalence of HIV/AIDS may have reduced incentives to invest in additional training and skills for its staff. Individuals with HIV/AIDS would also face reduced incentives to invest in advancing their own skills.

Haacker (2002a) undertakes a detailed analysis of the impact of HIV/AIDS on the economies of Southern Africa. The role of HIV/AIDS on the public sector, formal education sector, the workplace, and the training, experience and productivity of the labour force, are explored in detail. In the formal modelling, based on simulations using a simple open economy neoclassical growth framework, a decline in GDP per capita of 2% is predicted, on average, in the long-term across Southern Africa. Although Haacker (2002a) presents a detailed discussion of the impact HIV/AIDS may have on the public sector, formal schooling, and the incentives for employers to invest in on-the-job skills training, these impact channels are not explicitly incorporated into the formal modelling. In a closed-economy version of the model, Haacker (2002a) finds that despite declines in TFP, labour force productivity, and private savings, GDP per capita may in fact increase in the long-term in many Southern African countries due to the effects of lower population growth and an increased capital-to-labour ratio.

J - Life expectancy and savings: Separate from the effect HIV/AIDS may have on savings rates due to increased health care expenditure, the near certain premature death of individuals with HIV/AIDS, may itself lead to lower rates of savings. It is also possible, that people who do not currently have HIV/AIDS will save less in the face of widespread prevalence of HIV/AIDS due to the lower average life expectancy of the population.
**K - Accumulation of education capital:** Widespread prevalence of HIV/AIDS, by reducing life expectancy, can reduce the incentive to invest in one’s own education due to the lower possible returns from the investment (Ferreira and Pessoa, 2003; Bell et al., 2006). Additionally, the incomes of those households where an adult is directly impacted by HIV/AIDS are likely to decline, therefore reducing the ability to invest in their children’s education. This could lead to children entering the labour market prematurely to supplement household income. The result is a vicious cycle where these children, because of their lower levels of education find themselves in lower-skilled, low-paid jobs, and therefore have limited resources to invest in their own children’s education (Corrigan et al., 2005; Bell et al., 2006). Children who become orphans are also likely to struggle to continue in formal education relative to non-orphans, at least beyond the level in which the state will provide resources.

**L – Fertility:** Young (2005) suggests two ways in which HIV/AIDS could affect the decision to have children. Firstly, fertility rates might decline due to the unwillingness of individuals to engage in sexual activity because of the risks of contracting HIV/AIDS. Secondly, deaths of workers due to HIV/AIDS can lead to labour scarcity and therefore higher wages. This increase in wages could tempt more women into the labour force and reduce women’s demand for children due to the higher opportunity cost of time.

The economic impact of HIV/AIDS through its ability to slow the accumulation of human capital through education has been ignored until very recently. One recent study by Ferreira and Pessoa (2003) that investigates this impact channel, develops a theoretical OLG model whereby the long-run economic costs of HIV/AIDS are driven by a reduction in the incentive to invest in one’s own education due to the epidemic’s impact on life expectancy. Simulations with their model produce significantly larger estimates than earlier studies of the macroeconomic impact of HIV/AIDS. GDP per capita in the long-run steady state is estimated to be 26.3% lower than in a no-AIDS counterfactual scenario for the nine countries with the worst epidemics. Corrigan et al. (2005), also utilise an OLG framework to consider the effect of the creation of orphans on the accumulation of education capital and its transmission across generations. The incomes of those households impacted by HIV/AIDS are negatively affected by morbidity and eventually mortality, reducing the ability to invest in a child’s education. This can lead children into entering the labour market prematurely. A
vicious cycle ensues as those children obtain lower levels of education and hence lower-skilled, low-paid jobs, and therefore have limited resources to invest in their own children’s education. The life expectancy effect on human capital accumulation proposed by Ferreira and Pessoa (2003) is not considered by Corrigan et al. (2005), however its effect on saving and physical capital accumulation is captured. Calibrating the model for a typical sub-Saharan African country and assuming prevalence rates of 20 per cent for one generation, their baseline scenario suggests that HIV/AIDS will lower per capita income by 6.25 per cent in the long-run.\(^\text{10}\) Similarly, Bell et al. (2006) focus on these intergenerational effects utilising an OLG framework, but in addition consider the ‘life expectancy’ effect proposed by Ferreira and Pessoa (2003). The model generates much larger negative long-run predictions than any previous studies of its kind. Applying the model to South Africa, if the epidemic continues unabated, in the absence of government policy to tackle the epidemic, virtual economic collapse arises within two generations due to the cumulative effects of a breakdown in transmission of human capital across generations. By contrast, their modelling also suggests that appropriate government policies can prevent such a collapse, albeit with a huge fiscal burden in the order of four per cent of GDP.

Not all recent research points to large negative impacts of the HIV/AIDS epidemic on economic growth. In the provocatively-titled paper “The gift of the dying: The tragedy of AIDS and the welfare of future African Generations”, Young (2005) concludes that the epidemic will raise future per capita GDP, with a positive ‘fertility effect’ more than offsetting the negative impact of reduced human capital accumulation.

### 2.4 The Econometric Analysis of HIV/AIDS and Economic Growth

A small number of cross-country econometric studies have attempted to identify what effect HIV/AIDS has on economic growth. The first such study by Bloom and Mahal (1997) is based on cross-sectional regressions estimated using a variant of the MRW (1992) growth model. Estimation was based on a sample of 51 countries for the period 1980 to 1992. No statistically significant effect of adult HIV/AIDS prevalence or incidence on economic growth was detected. Many have subsequently questioned these results on a number of grounds; they were based on serious underestimates of HIV/AIDS prevalence, it may have been too early for the impact of the epidemic on morbidity and mortality and in turn economic activity to be detected, and that many countries

\(^{10}\) The negative impact becomes 14 per cent when HIV/AIDS is assumed not to be eradicated for two generations.
experiencing severe epidemics were excluded from the sample (Dixon et al., 2001; McDonald and Roberts, 2006). Additionally, the methodology used has been questioned. In the subsequent research of Dixon et al. (2001), McDonald and Roberts (2006) and Tandon (2005), a theoretical derived relationship between HIV prevalence and health capital is determined through a separate model, rather than HIV/AIDS variables being added in an *ad hoc* way directly into a growth equation as is the case in Bloom and Mahal (1997).

The two most recent studies that most closely resemble the analysis in this paper are those of McDonald and Roberts (2006), henceforth MR (2006), and Tandon (2005). Both papers conclude that a 1% rise in HIV prevalence leads to between 0.05% and 0.08% decline in income per capita based on full world or developing world samples. Both papers utilize a version of the augmented Solow model with the dependent variable in the growth equation being income per capita and health and education capital included as regressors. Health capital is treated as endogenous and a separate reduced form equation is specified for health capital, proxied by either infant mortality rates or life expectancy. In MR (2006), the health capital equation is estimated using dynamic panel difference-GMM estimation of Arellano and Bond (1991). In both studies, the predicted values for health capital are then used as instruments for health capital in the growth equation, which in turn is specified with two-way fixed effects and estimated using the difference-GMM estimator. Statistically significant and material negative effects of increased HIV prevalence are detected in both studies for a full world sample and in MR (2006) for a sub-sample of developing and African countries. Disappointing results obtained for the Latin American & Caribbean sub-sample in MR (2006) and for the Asian sub-sample in Tandon (2005) are most likely due to small sample size. In this study I attempt to overcome these problems associated with small samples, and improve the efficiency of the estimation through the use of intercept and interaction dummy variables as a means of sub-sample analysis.

A general concern with these econometric studies is that, in each of them, the chosen proxy variable for education capital, when included in the growth equation is found to be statistically insignificant or the opposite sign to what would be expected.

\[\text{As discussed in section 4.4, the paper adopts the system-GMM estimator of Blundell and Bond (1998) which is considered *a priori* superior to the difference-GMM estimator of Arellano and Bond (1991).}\]
3 Specifying Human Capital

3.1 Specifying and measuring education and health capital

Since human capital is a latent variable, researchers are compelled to find ‘indicators’ of or proxy variables for, the stock of the latent variable, the selection of which is a difficult task. There is often a trade-off between a measure that is conceptually appealing but has a limited time-series or country availability and a measure that is less appealing but is widely available over time and across countries. In empirical studies many different proxies for the education component of human capital have been utilized. Wossmann (2003) provides an excellent critical analysis of the many proxies used in the growth literature. Historically, the most common proxies for the stock of education capital included, school enrolment ratios (even though this is conceptually a flow measure), average years of schooling attained, and indicators of cognitive ability (e.g. adult literacy rates). In recent years, measures of average years of schooling in the adult population have become the most popular indicators of the stock of education capital for empirical work.\(^\text{12}\) The use of adult literacy rates has largely fallen out of favour as it become clear that it represents a very noisy measure: adult literacy rates disregard the level and type of literacy and the acquisition of skills beyond basic literacy (Wossmann, 2003). The school enrolment ratio, as a proxy for the stock of education capital, has also come under criticism recently. Enrolment ratios represent a flow variable and pertain mostly to people who are not currently in the labour force.\(^\text{13}\)

Although measures of ‘average years of schooling in the adult population’ are regarded by many as an improvement on the use of literacy rates and enrolment ratios as a proxy, this measure still has some major shortcomings: it does not allow for the decreasing private returns to schooling as detected by Psacharopoulos and Patrinos (2004) and others, or for the quality of education received (Hanushek and Kimko, 2000; Gundlach et. al., 2002; Wossmann, 2003). Additionally, the most widely used measure of educational attainment, that of Barro and Lee (1993, 2001), has been identified by Krueger and Lindahl (2001), de la Fuente and Domenech (2006), Cohen and Soto (2007), and others, as being subject, potentially, to a high degree of measurement error.

In the studies concerning the impact of HIV/AIDS on cross-country economic growth, education capital has been proxied with either the secondary enrolment ratio or educational attainment in the adult population.

\(^\text{12}\) The most common dataset that has been used in empirical work was constructed by Barro and Lee (1993, 2001).
\(^\text{13}\) Hanushek and Kimko (2000) and Wossmann (2003) are also critical of its use a proxy or indicator for the flow of human capital investment.
In this paper I adopt a Mincer specification of education capital that has recently gained popularity in the macroeconomics literature, driven by the dissatisfaction with the way in which measures of educational attainment had traditionally been incorporated into economic growth models (Cohen and Soto, 2007). Wossmann (2003) argues that the functional form of education capital is misspecified if one simply incorporates a proxy such as average years of schooling linearly into a macroeconomic production function. Bils and Klenow (1998, 2000) were the first to suggest that one should allow for the effect of human capital terms on income per capita to be powers of an exponential. Such a representation is consistent with the long established and widely used Mincer equation (Mincer, 1974) from the labour economics literature. Derived from human capital theory, the Mincer equation in its most simple form postulates a log-linear relationship between wages and years of schooling. A simple representation of a macroeconomic equivalent of the Mincer equation is given by

\[ \ln ec = b \cdot sch \] or \[ ec = e^{b \cdot sch} \] (1)

where \( ec \) is per capita education capital, \( sch \) is years of schooling attained per capita, and \( b \) is the private returns to a year of schooling.\(^{14}\) One can then combine estimates of the private returns to a year of schooling from the microeconomic literature with available measures of average years of schooling attainment to produce a macroeconomic proxy for education capital per capita that is consistent with human capital theory.

Hanushek and Kimko (2000) and Barro and Lee (2001b) and others have recently demonstrated the importance of considering the quality of education received, in addition to quantity, when estimating empirical growth models. I therefore follow Wossmann (2003) and Gundlach et. al. (2002) and adjust equation (1) by incorporating an indicator of education quality derived from international tests of students cognitive skills. A proxy for education capital can then be expressed as

\[ ec_{it} = e^{b \cdot Q \cdot sch_t} \] (2)

where subscripts denote country (i) and time (t), and \( Q \) is an indicator of education quality. To construct this cross-country proxy, the following data sources are drawn upon:\(^{15}\)

\(^{14}\) Hall and Jones (1999), Topel (1999), Gundlach et. al. (2002) and Wossmann (2003), among others, adopt a ‘Macro-Mincer’ specification of education capital in their cross-country growth analysis. To enable this simple jump from micro to macro, one must assume equal distribution of education capital within a countries population.

\(^{15}\) Further details of the data sources drawn upon are provided in section 4.3 and Appendix B.
1. Psacharopoulos and Patrinos (2004). The best available estimates of the private returns to a year of schooling are produced by Psacharopoulos and Patrinos (2004), henceforth PP (2004). Although PP (2004) provides country-specific estimates, they are based on the assumption of perfectly competitive labour markets with perfectly mobile labour. The questionable assumptions underlying the production of these country specific estimates, along with problems of measurement error, have typically lead researchers to resort to using PP’s world-average estimates (Wossmann, 2003; Lim and Tang, 2007). Another limitation is that time varying estimates are not available. The world-average return to a year of schooling (\(b\)) is estimated to be 0.097.

2. Altinok and Murseli (2007). Based on international surveys of student’s learning achievement, Altinok and Murseli (2007) produce a general index of education quality (QIHC-G) for a cross-section of 105 countries. Although a panel database is available covering the years 1960 to 2005, I do not choose to use this because there are very few countries with more than two observations. Most studies that have incorporated an indicator of education quality into economic growth models have used the database of Hanushek and Kimko (2000). Relative to Hanushek and Kimko (2000), the recently constructed database of Altinok and Murseli (2007) has a larger country coverage and its construction is based on a broader selection of international surveys and tests. I normalise the Altinok and Murseli (2007) QIHC-G index for each country relative to the measure for the United States to produce \(Q_i\).

3. Barro and Lee (2001). I have chosen the Barro and Lee (2001) series - average years of schooling in the population aged 15 years and over - as my predominant source for \(sch_i\).\(^{16}\)

Turning to health as a component of human capital, a number of alternative indicators or proxies for the health status of the workforce are available. The most commonly-used in cross-country empirical work are life expectancy at birth and infant mortality rates. Life expectancy and mortality rate data are widely available and both are considered in this study. Both indicators do, however, have some shortcomings.\(^{17}\) Notwithstanding these shortcomings, both are considered for inclusion in this study as they are the only ones that are widely available over time and across countries.

\(^{16}\) Justification for the use of the Barro and Lee (2001) dataset, notwithstanding recent criticisms of it, is provided in Appendix B.

\(^{17}\) See MR (2006) for a discussion of the relative merits of the two proxies.
4 Model, Data and Methodology

The innovations in this study with regards to modelling, data and methodology, relative to the cross-country empirical investigations into the HIV/AIDS epidemic of MR (2006) and Tandon (2005), are threefold:

1. A quality adjusted Mincerian specification of education capital is adopted and schooling attainment is treated as endogenous with the schooling equation estimated guided by a theoretical model of optimal schooling choice. The channels through which HIV/AIDS can impact on schooling investments, as suggested by Ferreira and Pessoa (2003), Corrigan et al. (2005), Bell et al. (2006) and others, are allowed for.

2. Inspired by recent developments in the empirical growth literature, rather than treating education and health capital as separate inputs from ‘raw labour’ in the production function, they are assumed embodied directly within the labour supply.\(^\text{18}\)

3. Methodological advances in the econometric literature are drawn upon by employing a system-GMM estimator, representing an improvement over the traditional differenced-GMM estimator.\(^\text{19}\)

4.1 The production of human capital

Here I do not feel it necessary for the purposes of this paper - that is to identify empirically the cross-country economic growth consequences of HIV/AIDS - to trace out a full general equilibrium model of the economic growth process. Instead, the chosen health, education and growth equations to be estimated, are guided by three well developed, widely accepted, ‘partial equilibrium’ frameworks from the literature, with potentially endogeneous regressors in each estimated equation dealt with by employing the method of instrumental variables.

4.1.1 An education production function

Human capital theory suggests that individuals maximise the present value of lifetime income by investing in education up to the point where marginal benefits equal marginal costs. The marginal benefits can be thought of the additional earnings that an individual would obtain from acquiring

\(^\text{18}\) See Section 4.2 for a detailed discussion.
\(^\text{19}\) See Section 4.4 for a detailed discussion.
further education relative to what could be obtained otherwise, whilst marginal costs include the 
foregone earnings whilst attending school, plus any other direct costs of schooling.

Here I present a model of schooling choice similar to that developed by Card (1999) and Kling 
(2000), itself influenced by the seminal works of Becker and Mincer. The model will be developed 
from the perspective of a country rather than the individual by simply introducing variables in per 
capita terms.20

Education capital per capita is assumed to be a function of the average number of schooling years 
obtained, \( ec = f(s) \). At each point in time, schooling decisions are made to maximise a utility 
function \( U \) with arguments \( ec \) and \( s \)

\[
U(ec, s) = \ln ec - \Phi(s) 
\]  
(3)

Where \( \Phi \) is an increasing convex function. The function generalises the discounted present value 
(DPV) objective function

\[
\int_s^\infty f(s)e^{-rt} \, dt = \frac{f(s)e^{-rs}}{r} 
\]  
(4)

where earnings are discounted at a rate \( r \), schooling is measured in years and it is assumed that 
earnings are zero whilst in school and equal to \( f(s) \) per year subsequently. The DPV function sets 
\( \Phi(s) = rs \), however here I follow Card (1999) and allow \( \Phi(s) \) to be a strictly convex function by 
assuming that the marginal cost of schooling rises by more than the foregone earnings of that year of 
schooling. Introducing country (i) and time (t) subscripts, the marginal benefit of an additional year 
of schooling is given by

\[
\frac{f_u'(s)}{f_u(s)} = b 
\]  
(5)

where \( b \) is a positive constant.21

Integrating equation (5) and substituting for \( \ln ec \), gives

---

20 To enable this simple jump from the micro to macro level, I assume equal distribution of education amongst a countries population.

21 Conceivably, the marginal benefit of schooling could differ between countries and time due to, for example, differences in labour market conditions 
or technology (Card, 1999), as well as differences in the quality of institutions and property rights. Card (1999) allows for decreasing marginal benefit 
by specifying the marginal benefit as \( b - k_s \), where \( k_s \) is a non-negative constant. Marginal benefits are assumed here to be constant, due to the 
cross-country data availability and quality issues already discussed in section 3.1.
\[
\int_0^{s_i} f_u(s) \frac{d}{ds} = \ln ec_i = b \cdot s_i \quad \text{or} \quad ec_i = e^{b \cdot s_i}
\]  
(6)

which is consistent with equation (2) in section 3.1, except for the education quality adjustment made.

Following Kling (2000), the marginal cost of schooling is assumed to be of the form

\[
\Phi_i'(s) = r_i + ks ; \quad k \geq 0
\]  
(7)

The marginal cost of schooling is allowed to vary across countries and time due to, for example, differences in average income levels, where lower incomes, in the absence of complete credit markets, represent a reduced capacity for households to pay for schooling (Card, 1999; Kling, 2000). However, the inclusion of a component that increases at a constant rate, k, with schooling, would capture, for example, the different ways in which education is financed as one moves from primary to secondary to tertiary education (Kling, 2000).

Equating (5) and (7), the optimal amount of per capita schooling, \( s_i^* \), is given by

\[
s_i^* = \frac{b - r_i}{k}
\]  
(8)

Following Card (1999) and Kling (2000), I allow the heterogeneous component of marginal cost, \( r_i \), to be linearly related to a vector of instrumenting variables, \( J_i \)

\[
r_i = \pi_1 J_i + \mu_i^i + \eta_i^i + \nu_i^i
\]  
(9)

where \( \pi_1 \) is a vertical vector of parameters, \( \mu_i^i \) and \( \eta_i^i \) are unobservable country and time specific terms respectively, and \( \nu_i^i \) a standard residual term. The schooling equation now becomes

\[
s_i = \pi_2 + \pi_3 J_i + \mu_i^2 + \eta_i^2 + \nu_i^2
\]  
(10)

where \( \pi_2 = \frac{b}{k} \), \( \pi_3 = -\frac{\pi_1}{k} \), \( \mu_i^2 = \frac{\mu_i^i}{k} \), \( \eta_i^2 = \frac{\eta_i^i}{k} \), and \( \nu_i^2 \) is a standard residual terms. Variables that can influence the marginal cost of schooling (those contained in \( J_i \)) are those that can be conceptualised as impacting on the discount rate, such as the capacity to pay for schooling (Kling, 2000). Although the model of optimal schooling outlined above does not exhibit an overlapping generations structure, the schooling equation to be estimated attempts to capture the reality that schooling investment

\[\text{Kling (2000) conceptualises variables that reduce the capacity to pay for schooling as being equivalent to an increase in the discount rate.}\]
decisions are typically made early in life, rather than throughout one's life. Inclusion of a lagged schooling term \( s_{i,t-1} \) will go some way towards capturing this reality.

I have chosen to draw predominantly on the Barro and Lee (2001) series - average years of schooling in the population aged 15 years and over \( (sch_i) \) – as my proxy for \( s_{i,t} \).

The schooling equation to be estimated is given by

\[
sch_i = \theta_1 sch_{i,t-1} + \theta_2 \ln YC_{it} + \theta_3 ED\$_{it} + \theta_4 \ln LE_{it} + \theta_5 POP15_{it} + \theta_6 URB_{it} + \eta_i + \mu_i + \nu_i 
\]

(11)

where

- \( \ln YC_{it} \) is the natural log of income per capita. Average incomes, in the absence of complete credit markets, are hypothesised to impact on one’s capacity to pay for schooling (Card, 1999; Kling, 2000). Its coefficient is therefore expected to be positive;

- \( ED\$_{it} \) is government expenditure on education as a proportion of GDP. Greater government resources directed towards education are hypothesised to improve access to education institutions and also to reduce the direct private costs of schooling that face a country's population. Its coefficient is therefore expected to be positive. Baldacci et. al. (2004), in a cross-country panel analysis of the role of public spending on health, education and economic growth, identify a strong positive relationship between ED$ and schooling attainment;

- \( \ln LE_{it} \) is the natural log of average life expectancy at birth. Life expectancy at birth, a proxy for health capital per capita, is expected to have a positive relationship with schooling attainment. Because declining life expectancy is hypothesised to reduce the incentive to invest in schooling by raising one’s discount rate;

- \( POP15_{it} \) is the share of population below the age of 15. Mingat and Tan (2003, 1999) argue that a lighter demographic burden reduces pressures on the education system and allows for more money to be spent per school-age child. The expected sign of the coefficient on this variable is therefore negative;

---

11 The model that has been outlined treats \( S_{i,t} \) as a per capita measure. What is important in terms of economic growth, however, is the education embodied in the labour force, not the total population. The population aged over 15 is more likely to be reflective of the size of a country’s labour force than the total population. Barro and Lee (2001) use the population over 15 years of age as a substitute for using a measure of the actual labour force, due to the unavailability of accurate labour force data for many countries. In addition to the Barro and Lee (2001) data, I also utilize data from some other sources for a small number of countries that are not adequately covered by the Barro and Lee (2001) database – see Appendix B.
• \( URB_{it} \) is the proportion of the population living in urban areas. This variable is included because a higher degree of urbanisation is hypothesised to be associated with easier access to education institutions and increased urbanization may allow for greater cost efficiency in the provision of education institutions;

• \( sch_{i,t-1} \), is the lagged average years of schooling in the population aged over 15 years and is included to capture potentially important omitted variables that exhibit persistence over time as well as reflecting the reality that schooling investment decisions are typically made early in life, rather than continually made at every stage of the life cycle in response to changes in the variables contained in \( J_{it} \);

• \( \mu_i^3 \) and \( \eta_t^3 \) denote the country-specific and period-specific fixed effects respectively. Fixed effects are included to allow for unobserved heterogeneity in schooling attainment across countries and time. Time-specific fixed effects are captured by including a full set of time dummy variables;

• \( v_{it}^3 \) is the error term.

### 4.1.2 A health production function

Grossman (1972) proposed a model of health production inspired by the human capital model of Becker (1964). In this model individuals are assumed to derive utility from health due to both its ‘production’ and ‘consumption’ benefits. An empirical formulation of Grossman’s (1972) pure investment version of the demand for health results in a reduced form equation that relates the log of health status (\( \ln H_{it} \)) to individual log wages (\( \ln W_{it} \)), the log price of medical inputs (\( \ln P_{it}^M \)), the log rate of depreciation of health (\( \ln d_{it} \)) and one’s education (\( EC_{it} \)), where subscripts refer to individuals (\( i \)) and time (\( t \)).\(^{24}\) I do not derive the equation here, as it has been derived many times elsewhere, see for example, Grossman (2000). The reduced form pure investment health equation is given by

\[
\ln H_{it} = \kappa_1 + \kappa_2 \ln W_{it} - \kappa_3 \ln P_{it}^M - \kappa_4 \ln d_{it} + \kappa_5 EC_{it}
\]  

\(^{24}\) Grossman (1972) views education as raising the efficiency with which one addresses their health needs.
where $\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5$ are positive coefficients. To switch to country level analysis, I make the simplifying assumption of equal distribution of health and simply restate the variables in the model in per capita terms. I replace $W_{it}$ with income per capita ($y_{it}$), $H_{it}$ with health capital per capita ($h_{it}$), and $EC_{it}$ with per capita education capital ($ec_{it}$). According to Grossman (1972), depreciation is assumed to be an increasing function of an individual’s age. Cropper (1981) also allows depreciation to be a function of a set of environmental variables. Here, I specify the log of the per capita depreciation rate as

$$\ln d_{it} = \ln d_{i0} + C \cdot X_{it}$$

(13)

where $X_{it}$ is a vector of per capita environmental variables, and $C$ is a vector of parameters. I assume that the age structure of a country remains constant over time, thereby excluding the role of age in determining the depreciation rate. Substituting (13) into (12) yields

$$\ln h_{it} = \kappa_1 + \kappa_2 \ln y_{it} - \kappa_3 \ln P_{it}^M - \kappa_4 \ln d_{i0} - \kappa_4 CX_{it} + \kappa_5 ec_{it}$$

(14)

and the health equation to be estimated is given by

$$\ln LE_{it} = \theta_1 \ln LE_{it-1} + \theta_2 \ln YC_{it} + \theta_3 \ln URB_{it} + \theta_4 \ln CPC_{it} + \theta_1 MEAL_{it}$$

$$+ \theta_2 HIV_{it} + \theta_3 ec_{it} + \eta_i + \mu_i + \nu_i$$

(15)

where

- $\ln LE_{it}$ refers to the log of per capita health capital, proxied by average life expectancy at birth.
- In the robustness testing in section 6, sensitivity of results to an alternative proxy – the infant mortality rate is assessed;
- $\ln LE_{it-1}$ refers to the lagged log of average life expectancy at birth. A lagged dependent variable is included to capture persistence over time of any omitted variables;\(^{25}\)
- $HIV_{it}$ is the adult HIV prevalence rate and $MEAL_{it}$ is the proportion of the population at risk of malaria. According to Cropper (1981), environmental factors impact on observed health through their role in determining the rate of depreciation of health. I therefore include $HIV_{it}$ and $MEAL_{it}$ as

---

\(^{25}\) A number of studies have extended the Grossman (1972) framework, providing a theoretical justification for the importance of lagged health status as a determinant of current health status. These extensions typically involve assuming either diminishing returns to health investment rather than constant returns (Arendt and Lauridsen, 2006), or assuming that health follows some form of partial adjustment process (Wagstaff, 1993; Lopez-Nicolas, 1998; Salas, 2002).
indicators of a country’s disease environment. The expected sign on both these variables is negative;

- In $CPC_t$ is the natural log of calorie intake per capita. This variable is included due to its role in determining the rate of depreciation of health. Maintaining good health requires the availability of an adequate nutritious food supply. Calorie intake per capita is commonly used as a proxy of the average nutritional status of a country’s population. A positive relationship is expected with health capital, at least up to a particular threshold. Beyond some level, a level most likely observed in all high-income countries, the health benefits of increased calorie intake are likely to be negligible;

- $ec_t$ is per capita education capital. According to Grossman (1972), better educated individuals are more efficient at producing health. I therefore expect a positive coefficient on $ec_t$.

- A higher degree of urbanization ($\ln URB_t$) is hypothesized to be associated with easier access to health care services and information (Fayissa and Gutema, 2005). Additionally, for a given dollar amount of public funding for the health care system, greater urbanization may allow for increased cost-efficiency in service provision. The degree of urbanisation in a country is therefore included in my estimated equation and conceptualised as a factor that can influence the average ‘price’ of medical inputs faced by a country’s population. Baldacci et al. (2004) identifies a strong positive effect of the degree of urbanization on health status in their cross-country panel regressions. An alternative hypothesis is that a greater degree of urbanization may be associated with increased pollution and congestion and have a negative effect on health status (Fayissa and Gutema, 2005). It is also possible that a rapid increase in urbanization may have a negative impact on health status if urban infrastructure such as sanitation is unable to keep pace. Based on this second interpretation, urbanization reduces health by raising the rate of depreciation. The overall effect of urbanization on health could therefore be positive or negative.

- Higher average income levels ($\ln YC_t$) raise the productive benefits of good health. Or alternatively, higher average income levels reflect a higher opportunity cost of poor health. Income per capita is therefore hypothesized to have a positive effect on health capital. In addition, a positive relationship is also expected because, in the absence of complete credit markets, low average income levels may represent a constraint on the resources available to
investment in health. Numerous past studies indicate that income levels are an important determinant of health capital.\textsuperscript{26}

- Country-specific fixed effects ($\mu_i^t$) and time-specific fixed effects ($\eta_i^t$) are included to capture unobserved heterogeneity in health capital across countries and time. For example, the initial rate of depreciation of health ($d_{i0}$) cannot be observed and would be captured in $\mu_i^t$.

Inclusion of a measure of public expenditures on health was considered as one proxy for the average price of medical inputs facing a country's population ($P^M$). In equation (14), higher prices for medical inputs are predicted to reduce the demand for health. Unfortunately, cross-country comparable health expenditure data could not be obtained for a large enough time series and cross-section of countries. Apart from the potential role of urbanization in influencing the ‘price’ of medical inputs, further widely available country-level determinants of the prices of medical inputs could not be obtained. I will make the assumption that the unobserved variation in medical input prices is fully captured in the country-specific and time-specific fixed effects ($\eta_i^t + \mu_i^t$).

### 4.2 An empirical growth model

Here I depart from the standard MRW (1992) growth framework and adopt an augmented Solow production function that incorporates a Mincerian approach to specifying education capital. Additionally, rather than treating human capital as a separate input from labour in the production function, as in MRW (1992), it is assumed embodied directly in the labour force.\textsuperscript{27}

I propose a Cobb-Douglas production function given by

$$Y_{it} = K_{it}^\alpha \left( A_{it} L_{it} (ec_{it}^\beta h_{it}^{1-\beta}) \right)^{1-\alpha}$$  \hspace{1cm} (16)

where $Y$ represents output, $K$ is physical capital and $L$ is labour. The inner bracketed term, $(ec_{it}^\beta h_{it}^{1-\beta})$, represents a production function for total human capital exhibiting constant returns to scale, comprising education capital ($ec_{it}$) and health capital ($h_{it}$) per worker, where $\beta$ is the partial

\textsuperscript{26} See, for example, Baldacci \textit{et al.} (2004) and references within.

elasticity of human capital with respect to education capital. Multiplication of $L_{it}$, and $\left( ec_{it}^{\beta} h_{it}^{1-\beta} \right)$ can be interpreted as ‘productive labour’ input. $A_{it}$ is the level of productive labour augmenting technology and $\alpha$ is the partial elasticity of output with respect to physical capital.

Following MRW (1992), Islam (1995), McDonald and Roberts (2002) and others, I allow the initial state of technology ($A_{i0}$) to differ across countries. Technology is assumed to encompass not just technology in the typical sense, but also natural resource endowments and institutions, etc. Initial technology takes the form:

$$\ln A_{i0} = a + \varepsilon_i$$  \hspace{1cm} (17)

where ‘$a$’ represents a constant and ‘$\varepsilon_i$’ a country-specific shock term. It is assumed that an exogenously given fraction of output is saved and invested ($w^K$) in physical capital in each period and that the capital stock depreciates over time at an exogenously given rate ($\delta$). It is also assumed that labour and technology grow at exogenously given rates $n$ and $g$ respectively. The evolution of physical capital per technology augmented worker (where $k = \frac{K}{A \cdot L}$) can be expressed as (temporarily dropping country (i) and time (t) subscripts):

$$\dot{k} = w^K \left( k^\alpha \left( ec_{it}^{\beta} h_{it}^{1-\beta} \right) \right)^{1-\alpha} \left( n + g + \delta \right) k$$ \hspace{1cm} (18)

where () represents the time derivative. In a steady-state, physical capital per technology augmented worker must be constant. Setting the above equation of motion equal to zero, the steady state value for $k$, denoted with an asterisk, is defined by

$$k^* = \left[ \frac{w^K}{n + g + \delta} \right]^{\frac{1}{1-\alpha}} ec_{it}^{\beta} h_{it}^{1-\beta}$$ \hspace{1cm} (19)

Time and country subscripts will now be reintroduced. The share of output saved and invested in physical capital is assumed to be country-specific but constant over time. Labour is assumed to grow at a country-specific constant rate and technology grows at a period-specific constant rate. Steady-state output per worker is given by:

$$y_{it}^* = A_{it} \left[ \frac{w_{it}^K}{n_i + g + \delta} \right]^{\frac{\alpha}{1-\alpha}} \left( ec_{it}^{\beta} h_{it}^{1-\beta} \right)$$ \hspace{1cm} (20)
Recalling that initial technology is denoted by $A_{i0}$ and that it is assumed to grow at a period-specific rate $g_t$, then an expression for $A_t$ is given by

$$A_t = A_{i0}e^{g_t}$$

(21)

Substituting (21) into (20) and taking logs yields

$$y^*_{it} = A_{i0} + g_t t + \frac{\alpha}{1-\alpha} \ln w^\kappa_t + \beta \ln ec_{it} + \left(1 - \beta\right) \ln h_{it} - \frac{\alpha}{1-\alpha} \ln \left(n_t + g_t + \delta\right)$$

(22)

and linearising equation (22) around the steady state level of income, following the method outlined by Romer (2001, p.24) yields

$$\ln y^*_{it} = Z \ln A_{i0} + g_t t + \left(\frac{Z\alpha}{1-\alpha}\right) \ln w^\kappa_t - \left(\frac{Z\alpha}{1-\alpha}\right) \ln \left(n_t + g_t + \delta\right) + \beta Z \ln ec_{it}$$

$$+ \left(1 - \beta\right) Z \ln h_{it} - (Z - 1) \ln y^*_{i0}$$

(23)

Where $y^*_{i0}$ is the initial output per worker in country $i$, $Z = \left(1 - e^{-\lambda t}\right)$ and $\lambda$, which represents the speed of convergence to the steady-state is given by:

$$\lambda = \left(n_t + g_t + \delta\right)(1-\alpha)$$

(24)

Equation (23) can be represented as a dynamic panel data model by using the conventional notation of the panel data literature as follows:

$$\ln y^*_{it} = \Psi_1 \ln y^*_{i0} + \Psi_2 \ln w^\kappa_t + \Psi_3 \ln x_{it} + \Psi_4 \ln ec_{it} + \Psi_5 \ln h_{it} + \eta^s_t + \mu^s_t + \nu^s_t$$

(24)

where, from equation (2), $\ln ec_{it} = b \cdot Q_t \cdot sch_{it}$ and

<table>
<thead>
<tr>
<th>$Z = \left(1 - e^{-\lambda t}\right)$</th>
<th>$\Psi_5 = \left(1 - \beta\right)Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Psi_1 = 1 - Z$</td>
<td>$x_{it} = n_t + g_t + \delta$</td>
</tr>
<tr>
<td>$\Psi_2 = \left(\frac{Z\alpha}{1-\alpha}\right)$</td>
<td>$\mu^s_t = ZA_{i0}$</td>
</tr>
<tr>
<td>$\Psi_3 = -\Phi_{q_t}$</td>
<td>$\eta^s_t = g_t t$</td>
</tr>
<tr>
<td>$\Psi_4 = \beta Z$</td>
<td></td>
</tr>
</tbody>
</table>
The growth equation to be estimated is given by

\[ \ln YC_{it} = \theta_{14} \ln YC_{i,t-1} + \theta_{15} \ln \left( n + g + \delta \right)_{it} + \theta_{16} \ln INV_i + \theta_{17} \ln LE_{it} + \theta_{18} \left( 0.097 \cdot sch_i \cdot Q_i \right) \]

\[ + \eta_i^6 + \mu_i^6 + \nu_i^6 \] (26)

where

• the lagged log of income per capita (\( \ln YC_{i,t-1} \)) is included to capture conditional convergence.

The augmented Solow model predicts ‘conditional convergence’, meaning that countries with the same steady-state determinants (i.e. investment in physical and human capital, population growth, and technological growth) are expected to converge to the same long-run steady-state level of income per capita, with those countries starting at a lower levels of income per capita predicted to grow faster. The coefficient on \( \ln YC_{i,t-1} \) is therefore expected to be positive;

• \( \ln \left( n + g + \delta \right)_{it} \) is the logarithm of the sum of the rate of population growth, technological growth, and the physical capital depreciation rate.\(^{29}\) It is often referred to as the ‘capital widening’ term. Following a convention in the literature, the sum of technological growth and the depreciation rate is here assumed to be uniform at 5%. Higher population growth is expected to reduce income per capita as it results in a dilution of capital. This is because, although an increase in population, absent any change in technology or capital, will raise total income, income per capita will fall in the presence of diminishing returns;

• \( \ln INV_i \) denotes the log investment ratio, measured as the average from 1970 to 2004 of the log of domestic investment in physical capital divided by GDP. The coefficient is expected to be positive because an increased stock of physical capital is expected, \textit{ceteris paribus}, to raise the productivity of human capital augmented labour and its returns.

• Both the health and education components of per capita human capital, \( \ln LE_{it} \) and \( \left( 0.097 \cdot sch_i \cdot Q_i \right) \) respectively, are expected to have a positive effect on income per capita because increases in the average level of human capital each are hypothesised to raise workers productivity and the productivity and returns to physical capital. The education component of human capital enters in a linear form due to the assumed Mincer specification. Both \( \ln LE_{it} \) and

---

\(^{28}\) Ideally, the dependent variable would be income per worker, not income per capita. Unfortunately, an income per worker series has not been updated in the Penn World Tables since 1990 due to a lack of good quality data regarding labour force size, particularly for developing countries.

\(^{29}\) Ideally, \( Q \) would measure the rate of growth of the labour force, but due to data restrictions, I have used the rate of population growth.
are treated as endogenous and instrumented for using their predicted values from the health and schooling equations.

- Country-specific fixed effects ($\mu_i^6$) and time-specific fixed effects ($\eta_t^6$) are included to capture unobserved heterogeneity in income per capita across countries and time. Specifically, $\eta_t^6$ captures time-specific improvements in technology that are assumed to be the same for all countries, whilst $\mu_i^6$ captures differences in initial technology ($A_{i0}$) amongst countries. Time-specific fixed effects are captured by including a full set of time dummy variables. The error term is given by $\nu_i^6$.

4.4 The Data

This study uses data from a variety of sources. The World Development Indicators (WDI, 2006) database is drawn upon for various demographic, health and education indicators. The WDI (2006) database contains data for 208 countries or territories, representing the widest coverage of all the data sources utilized. GDP and investment data are drawn from Penn World Tables (PWT) version 6.2. Data regarding schooling achievement and returns to schooling are drawn from a number of sources including; Barro and Lee (2001), WDI (2006), UNESCO Institute of Statistics (2006), Cohen and Soto (2007) and PP (2004). I draw on Altinok and Murseli (2007) to produce an indicator of education quality. Proxies for nutritional status are drawn from the United Nations Food and Agriculture Organization database (FAO; 2006, 1998). Measures of the population at risk of malaria are obtained from data compiled by Gallup et al. (2001) and adult HIV prevalence rates are obtained from UNAIDS (2006). HIV prevalence rates are assumed to be zero for all periods prior to 1980. Table 2 lists all variables, their acronyms (in parentheses) and their sources. Appendix B describes the variables used in more detail.

The data-set is grouped into panels of five-year periods from 1960 through 2004, although some of the variables are only available from 1970 onwards. All the variables used are simple averaged data for each five-year period, with the exception of the malaria, investment in physical capital, and education quality data (see Appendix B). The available sample consists of 142 countries. This sample contains within it all of the countries included in the studies by Tandon (2005) and MR
There are nine 5-year time periods, i.e., 1960-1964, 1965-1969 etc. Since observations are not available for all countries and time periods for every variable, the sample represents an unbalanced panel. Appendix B describes, in more detail, the process via which the sample was chosen and contains a list of countries included. Summary statistics are provided in Table 3 in Appendix B.

### Table 2: Variable Definitions and Sources

<table>
<thead>
<tr>
<th>Variable</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real GDP per capita (YC)</td>
<td>Penn World Tables 6.2 (2006)</td>
</tr>
<tr>
<td>Investment share of GDP (INV)</td>
<td>Penn World Tables 6.2 (2006)</td>
</tr>
<tr>
<td>Average years of schooling achieved in the</td>
<td>Barro and Lee (2001); Cohen and Soto (2007); World Development</td>
</tr>
<tr>
<td>population aged over 15 years (SCH)</td>
<td>Indicators (WDI) (World Bank 2006); UNESCO (2006)</td>
</tr>
<tr>
<td>Education quality index (Q)</td>
<td>Altinok and Murseli (2007)</td>
</tr>
<tr>
<td>Government expenditure on education as a</td>
<td>WDI (World Bank 2006); UNESCO (2006)</td>
</tr>
<tr>
<td>proportion of GDP (ED$)</td>
<td></td>
</tr>
<tr>
<td>Proportion of population aged 15 or under</td>
<td>WDI (World Bank 2006)</td>
</tr>
<tr>
<td>(POP15)</td>
<td></td>
</tr>
<tr>
<td>Population growth rate (n)</td>
<td>WDI (World Bank 2006)</td>
</tr>
<tr>
<td>Proportion of population living in urban</td>
<td>WDI (World Bank 2006)</td>
</tr>
<tr>
<td>areas (URB)</td>
<td></td>
</tr>
<tr>
<td>Life expectancy at birth (LE)</td>
<td>WDI (World Bank 2006)</td>
</tr>
<tr>
<td>Infant mortality rate (INF)</td>
<td>WDI (World Bank 2006)</td>
</tr>
<tr>
<td>Proportion of the population at risk of</td>
<td>Gallup <em>et. al</em> (1999)</td>
</tr>
<tr>
<td>malaria (MAL)</td>
<td></td>
</tr>
<tr>
<td>Calorie intake per capita (CPC)</td>
<td>FAO (2006, 1998)</td>
</tr>
<tr>
<td>Adult prevalence of HIV (HIV)</td>
<td>UNAIDS (2006)</td>
</tr>
<tr>
<td>Technological growth rate (g)</td>
<td>N/A</td>
</tr>
<tr>
<td>Capital depreciation rate (δ)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### 4.5 Estimation methodology

The estimation methodology employed addresses the problem of endogenous regressors in two respects. Firstly, in the growth equation, health and education capital are considered endogenous for reasons previously highlighted. This means that the health and education capital terms in equation (26) are potentially correlated with the residuals \( ν_o i \), rendering Ordinary Least Squares (OLS) estimation biased and inconsistent (Baltagi, 2005). Secondly, in models with unobserved country-specific fixed effects, \( μ_i \), and a lagged dependent variables, as in equations (11), (15) and (26), OLS estimation will produce upwardly biased coefficient estimates on the lagged dependent variable (Baltagi, 2005).
Here, two forms of Instrumental Variable (IV) estimation are used to solve the problem of endogenous regressors, two-stage least squares estimation (2SLS) and System Generalised Method of Moments estimation (Sys-GMM).

**Two-stage least squares estimator**

Potential endogeneity of health and education capital in equation (26) is addressed through the use of 2SLS. The 2SLS approach adopted here is the approach used by MR (2006) and Tandon (2005) in their analysis of the HIV/AIDS epidemic, except here, education capital, as well as health capital, is treated as endogenous. Predicted values of $sch_{it}$ and $LE_{it}$ are obtained from equations (11) and (15) and used as instruments for $sch_{it}$ and $LE_{it}$ in the second-stage regression – the growth regression. These ‘instruments’ will be uncorrelated with the residuals and the slope estimators obtained in this second stage will be consistent (Hill et al., 2001). Separate estimation of health and education capital production functions also allows for a theoretically sound approach to identifying, more precisely, the macroeconomic effects of the HIV/AIDS epidemic.

**Generalised Method of Moments Estimator**

The second econometric issue that needs to be addressed involves the dynamic nature of the estimated equations. As previously mentioned, in dynamic panel models with fixed effects, OLS is an inconsistent estimator. One standard approach for panel models with unobserved country-specific effects is to use the ‘fixed-effects estimator’ (FE estimator). However, in the presence of lagged dependent variables, FE estimation will also produce biased estimates of the coefficient (downwardly biased), except in the situation where the number of time periods is very large (Baltagi, 2005). An alternative to FE estimation, particularly suited to situations where the time dimension of the dataset is short, is proposed by Arellano and Bond (1991). This method, commonly known as difference-GMM, effectively takes first-differences of the equation to remove fixed effects and then uses appropriate instruments for the lagged differenced dependent variable. Under particular assumptions, a set of moment conditions can be exploited allowing two-period lagged levels and deeper as instruments in the first-differenced equation. This will produce consistent estimates of the coefficient on the lagged dependent variable when the time dimension is fixed and $i \to \infty$. MR

---

31 Please refer to Arellano and Bond (1991) for a detailed treatment.
(2006) and Tandon (2005) employ this difference-GMM estimator in their analysis of the HIV/AIDS epidemic and this is where the current analysis makes a significant methodological departure.

This study is the first to apply the Sys-GMM approach, an approach initially proposed by Arellano and Bover (1995) and further developed by Blundell and Bond (1998), to the analysis of the cross-country growth effects of the HIV/AIDS epidemic. The choice of the Sys-GMM estimator has been motivated by recent developments in the econometric literature.

In addition to the assumptions underlying the method of Arellano and Bond (1991), under certain additional assumptions, another set of moment conditions will become available. These additional moment conditions allow for lagged first differences to be used as instruments for the equation in levels. The approach is then to estimate, using GMM, a system of equations; a set of first-differenced equations with lagged levels used as instruments, and a set of equations in levels with lagged differences used as instruments. The Sys-GMM has been shown to be superior to difference-GMM in certain circumstances, ones which are present in this study. Furthermore, it has been found to significantly outperform difference-GMM in the context of the estimation of productions functions – in particular empirical growth models. These recent developments cast doubt on the accuracy of earlier estimates of the macroeconomic impact of HIV/AIDS obtained from growth regressions employing the difference-GMM estimator.

The validity of the instruments underlying system-GMM can be checked using a range of specification tests including the Arellano and Bond (1991) test for autocorrelation in the first-differenced residuals, the Hansen J-test and the ‘Difference in Sargan’ C-test, each of which are reported in the results section. In section 5, only the results from Sys-GMM estimation are presented as these are a-priori considered superior to OLS, FE and difference-GMM estimation. Knowledge of the direction of the theoretical bias resulting from these alternative estimation techniques is useful, because confidence in the consistency of the Sys-GMM estimator can be gained if it is observed to lie somewhere between the OLS and FE estimates and above the difference-GMM.

---

32 Please refer to Arellano and Bover (1995) for a detailed treatment.
33 The calculation of the Sys-GMM estimator is discussed in detail in Blundell and Bond (1998).
34 See, for example, Blundell and Bond (1998), Blundell and Bond (2000) and Bond et al. (2001).
35 With regards to the Arellano and Bond (1991) test for autocorrelation in the first-differenced residuals, mathematically, the process of first-differencing the data will induce first-order autocorrelation of the differenced residuals. The absence of second-order autocorrelation in the first-differenced equation is equivalent to there being no autocorrelation in the levels residuals \( V_{q} \). The moment conditions underlying both difference-GMM and sys-GMM estimates rely on the absence of autocorrelation in the levels residuals. When we have more instruments than are needed to identify an equation, the Hansen (1982) J-test of overidentifying restrictions is available. This J-test assesses the independence of the instruments from the residuals (see Arellano and Bond (1991) for a discussion). Another important specification test when using sys-GMM is the ‘difference in Sargan’ test, or the ‘C’ test. This test allows for an assessment of the validity of the additional instruments used in sys-GMM relative to difference-GMM. The statistic underlying this test is the difference between two Hansen J-test statistics (see Arellano and Bover (1995) for a discussion).
estimates (Bond, 2002). Results from these alternative estimation techniques are therefore presented as part of the robustness tests in Section 6.

5 Results

The schooling and health equations as described by equations (11) and (15) are initially expanded by adding a set of squared terms for each variable, and each variable interacted with the income per capita and HIV prevalence term (except for the lagged dependent variable). Variables that were individually insignificant at the 10% level, and where inclusion added little explanatory value to the model, were removed. The final model specifications are reported in Tables 4 through 6.

5.1 Results for the health equation

Table 4 presents the Sys-GMM estimates for equation (15). Results are reported for both the full sample and a sub-sample of developing countries. Sub-sample analysis is achieved by incorporating a full set of intercept and slope-shifting dummy variables into the equation. The results for both samples are very similar.

The coefficient on lnYC has the expected positive sign and is significant at the 5% level for both the full and developing world sample. The coefficient on HIV is negative and strongly significant in both cases, whilst the other indicator of a country’s disease environment, MAL, although exhibiting a negative sign as expected, is not statistically significant at any conventional level in either sample. The degree of urbanization was not found to be significant and was thus dropped from both estimated equations. It is possible that the insignificance of urbanization could be due to the theorized competing positive and negative effects outlined in Section 4.1 offsetting one another. The quantitative indicator of education, sch, rather than the quality adjusted education capital proxy, ec, was included in the final estimated equation as it had significantly greater explanatory power.

---

36 One-step sys-GMM estimates are reported as opposed to two-step. The distinction between one and two-step estimation is discussed in Baltagi (2001) page 134. The one-step estimator has been chosen because, although sys-GMM two-step estimators are theoretically more efficient, Monte Carlo simulations have highlighted two problems with two-step estimators, including; (i) the estimator converges to its asymptotic distribution relatively slowly and can be biased in finite samples, and (ii) the usual asymptotically valid standard errors associated with the estimator exhibit downward bias in finite samples (Bond et al., 2001). Although a solution to the second problem has been proposed by Windmeijer (2005), in the form of a finite sample correction to the downwardly biased standard errors, the first problem remains, and, in any case, Monte Carlo simulations show that the efficiency gains of two-step estimators are often only small (Bond et al., 2001).

37 This is also the approach used in the estimation of the education and the growth equations.

38 Both current and lagged income were considered and current income provided marginally better results for both samples.
One period lagged average years of schooling in the population aged over 15 exhibits the expected positive sign and is statistically significant in both samples at the 5% level. Calorie intake per capita, $CPC_t$, has a positive coefficient, as expected, and is significant at the 5% level both samples.

### Table 4: Results for the health equation (World and Developing world sample)

<table>
<thead>
<tr>
<th>Dependent variable: $\ln(LE)$</th>
<th>Estimation technique</th>
<th>System-GMM (one-step robust estimates)</th>
<th>Full sample</th>
<th>Developing World sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln(LE)$</td>
<td></td>
<td></td>
<td>0.6238 (0.000)***</td>
<td>0.6247 (0.000)***</td>
</tr>
<tr>
<td>$\ln(YC)$</td>
<td></td>
<td></td>
<td>0.0209 (0.018)**</td>
<td>0.0238 (0.018)**</td>
</tr>
<tr>
<td>$\ln(CPC)$</td>
<td></td>
<td></td>
<td>0.0560 (0.026)**</td>
<td>0.0580 (0.027)**</td>
</tr>
<tr>
<td>$L.\ln(SCH)$</td>
<td></td>
<td></td>
<td>0.0314 (0.032)**</td>
<td>0.0312 (0.037)**</td>
</tr>
<tr>
<td>MAL</td>
<td></td>
<td></td>
<td>-0.0170 (0.128)</td>
<td>-0.0161 (0.135)</td>
</tr>
<tr>
<td>HIV</td>
<td></td>
<td></td>
<td>-0.0121 (0.000)***</td>
<td>-0.0122 (0.000)***</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
<td>0.9288 (0.000)***</td>
<td>0.8898 (0.009)***</td>
</tr>
<tr>
<td>Time dummies</td>
<td></td>
<td></td>
<td>74.64 (0.000)***</td>
<td>77.19 (0.000)***</td>
</tr>
<tr>
<td>Wald</td>
<td></td>
<td></td>
<td>8824.43 (0.000)***</td>
<td>28139 (0.000)***</td>
</tr>
<tr>
<td>Hansen (J-test)</td>
<td></td>
<td></td>
<td>Chi2(26)=58.60 (0.000)***</td>
<td>Chi2(42)=57.06 (0.060)*</td>
</tr>
<tr>
<td>Difference-in-Sargan (C-test)</td>
<td></td>
<td></td>
<td>Chi2 (6)=17.97 (0.006)***</td>
<td>Chi2(11)=22.21 (0.023)**</td>
</tr>
<tr>
<td>AR(2)</td>
<td></td>
<td></td>
<td>$z =-1.63 (0.103)$</td>
<td>$z =-1.66 (0.097)*$</td>
</tr>
<tr>
<td>AR(3)</td>
<td></td>
<td></td>
<td>$z = 0.57 (0.570)$</td>
<td>$z = 0.540 (0.592)$</td>
</tr>
<tr>
<td>Instrument count</td>
<td></td>
<td></td>
<td>39</td>
<td>62</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td></td>
<td>657</td>
<td>657</td>
</tr>
<tr>
<td>Countries</td>
<td></td>
<td></td>
<td>125</td>
<td>125</td>
</tr>
</tbody>
</table>

**Notes:**
1. *Significant at 10%; **Significant at 5%; ***Significant at 1%. Standard errors are robust to arbitrary autocorrelation and heteroskedasticity within countries.
2. L. denotes lagged value.
3. AR(2) and AR(3) denotes the Arellano and Bond (1991) test of autocorrelation of the first-differenced residuals of order two and order three respectively.
4. Hansen denotes the Hansen (1982) J-test for over-identifying restrictions, whilst the difference-in-Sargan C-test is a test for the validity of the additional instruments used in system-GMM relative to difference-GMM.
5. Values in parenthesis indicate p-values, that is, the probability of incorrectly rejecting the null hypothesis. The p-value corresponding to the time dummies is for a null hypothesis of all time dummies being jointly zero. The p-value corresponding to the ‘Wald’ is for a null hypothesis of all the parameters in the model, excluding the constant, being jointly zero.
6. L.\ln(LE) is instrumented for with system-GMM style instruments. Lags two and earlier of the instrumenting variable is used for the equation in differences, and lags one and earlier of the instrumenting variable in differences for the levels equation.

For both the full and developing world, we would fail to reject the null hypothesis of no autocorrelation at the conventional 5% level of significance - the Arellano and Bond (1991) test for second-order autocorrelation in the estimated residuals in first-differences. However, observing p-values close to 0.1 for the AR(2) test suggests autocorrelation may be a problem. Rejection of the null hypothesis underlying the AR(2) would render the GMM instruments invalid. I have therefore restricted the instrument set underlying the predetermined but endogenous variable, $\ln LE_{i,t-1}$ to lags two and earlier of the instrumenting variable for the equation in differences, and lags one and earlier...
of the instrumenting variable in differences for the levels equation.\textsuperscript{39} With regards to the other specification tests, failure to reject the null underlying the J and C-test, reported in Table 4, is also necessary for the validity of the GMM instruments. Unfortunately the null hypothesis for the J-test, for both samples, is strongly rejected. Notwithstanding the failure of the Hansen J-test, the specification is retained for now. The implications of the failure of the J-test for the reliability of the estimated coefficients, is assessed in the robustness testing in section 6.

5.2 Results for the schooling equation

Table 5 presents the Sys-GMM estimates for equation (11). In the full sample, both $\ln YC_{i,t-1}$ and $EDS_{i,t-1}$, are statistically significant at the 5% level with the expected positive sign.\textsuperscript{40} For the developing world sample, both $\ln YC_{i,t-1}$ and $EDS_{i,t-1}$ exhibit the expected positive sign, however only $\ln YC_{i,t-1}$ is statistically significant at any conventional significance level. Lagged schooling attainment in the population aged 15 and over ($SCH_{i,t-1}$) exhibits a highly significant positive sign in both samples, indicating, as expected, a high degree of persistence in average schooling attainment in the population aged over 15 years. Life expectancy at birth ($LE_{i,t-1}$) exhibits, as expected, a positive impact on $SCH_{i}$ and is highly significant in both samples. This result is consistent with the proposition that rising life expectancy increases the incentive to invest in schooling, by increasing the expected returns from the investment. $POP1_{i,t-1}$ exhibits a negative coefficient and is highly statistically in both samples, reflecting that a lighter demographic burden, by reducing pressure on the ability of governments to maintain a properly resourced education system, can lead to increased participation. Urbanization ($URB_{i,t-1}$) was found to be individually statistically insignificant in both samples, however it is retained as it improves the overall fit of the model.

The null hypothesis underlying the AR(2) test, J-test and C-test cannot be rejected for either sample, supporting the validity of the instrument set.

\textsuperscript{39} Typically, if one failed to reject the null hypothesis underlying the AR(2) test, for predetermined endogenous variables, lags one and earlier of the instrumenting variable for the equation in differences, and lags zero and earlier of the instrumenting variable in differences for the levels equation would be used.

\textsuperscript{40} For all explanatory variables, both current and lagged levels were considered. Lagged values were found to produce better results in all cases.
Table 5: Results for the schooling equation (World and Developing world sample)

<table>
<thead>
<tr>
<th>Dependent variable: SCH</th>
<th>Estimation technique</th>
<th>Full sample</th>
<th>Developing World sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>System-GMM (one-step robust estimates)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L.SCH</td>
<td>0.6309 (0.000)*****</td>
<td>0.6747 (0.000)*****</td>
<td></td>
</tr>
<tr>
<td>L.in(YC)</td>
<td>0.1917 (0.024)**</td>
<td>0.1633 (0.038)**</td>
<td></td>
</tr>
<tr>
<td>L.EDS</td>
<td>0.0492 (0.018)**</td>
<td>0.0266 (0.125)</td>
<td></td>
</tr>
<tr>
<td>L.in(LE)</td>
<td>1.9476 (0.000)*****</td>
<td>1.7606 (0.000)*****</td>
<td></td>
</tr>
<tr>
<td>L.POP15</td>
<td>-0.0299 (0.004)*****</td>
<td>-0.0281 (0.011)**</td>
<td></td>
</tr>
<tr>
<td>L.URB</td>
<td>0.0032 (0.212)</td>
<td>0.0027 (0.217)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-6.6565 (0.000)*****</td>
<td>-5.8358 (0.000)*****</td>
<td></td>
</tr>
<tr>
<td>Time dummies</td>
<td>Chi2[5]=24.77 (0.000)*****</td>
<td>Chi2[5]=20.64 (0.000)*****</td>
<td></td>
</tr>
<tr>
<td>Wald</td>
<td>Chi2[11]=7936 (0.000)*****</td>
<td>Chi2[16]= 24826 (0.000)*****</td>
<td></td>
</tr>
<tr>
<td>Hansen (J-test)</td>
<td>Chi2[30]=29.90 (0.471)</td>
<td>Chi2[60]=60.56 (0.455)</td>
<td></td>
</tr>
<tr>
<td>Difference-in-Sargan (C-test)</td>
<td>Chi2[6]=6.23 (0.398)</td>
<td>Chi2[12]=17.35 (0.137)</td>
<td></td>
</tr>
<tr>
<td>AR(2)</td>
<td>Z=-0.26 (0.792)</td>
<td>Z=0.33 (0.741)</td>
<td></td>
</tr>
<tr>
<td>Instrument count</td>
<td>42</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>634</td>
<td>634</td>
<td></td>
</tr>
<tr>
<td>Countries</td>
<td>132</td>
<td>132</td>
<td></td>
</tr>
</tbody>
</table>

Notes: See Table 4 notes (1) through (5)
6. L.(SCH) is instrumented for with system-GMM style instruments. Lags one and earlier of the instrumenting variable is used for the equation in differences, and lags zero and earlier of the instrumenting variable in differences for the levels equation.

5.3 Results for the growth equation

Table 6 presents the Sys-GMM estimates from equation (26) with predicted values from the health equation used as instruments for the endogenous health regressor. Predicted values from the schooling equation are multiplied by 0.097, being the world average private rates of return to schooling as reported by PP (2004), and by an indicator of education quality constructed from the database of Altinok and Murseli (2007). The resulting variable is then used as an instrument for \( (0.097 \cdot Q_t \cdot sch_{it}) \) in equation (26).

The results for the growth equation are generally very pleasing. For both samples there is strong evidence of conditional convergence. The sign on the investment term is positive, as expected, and significant at the 1% and 5% level in the world and developing world samples respectively. For both samples, the coefficient for ‘capital widening’, \( \ln(n + g + \delta) \), has the expected negative sign, however it is only statistically significant in the developing world sample. This is in contrast to the findings of MR (2006) and Tandon (2005) who detect in their full and sub-sample analysis either an insignificant effect or an unexpected positive effect.
### Table 6: Results for the growth equation (World and Developing world sample)

<table>
<thead>
<tr>
<th>Dependent variable: ln(YC)</th>
<th>Estimation technique</th>
<th>Full sample</th>
<th>Developing World sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GMM-Sys (one-step robust estimates)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convergence coefficient (L.ln(YC))</td>
<td>0.7644 (0.000)***</td>
<td>0.6597 (0.000)***</td>
<td></td>
</tr>
<tr>
<td>ln(INV)</td>
<td>0.0801 (0.000)***</td>
<td>0.0635 (0.012)**</td>
<td></td>
</tr>
<tr>
<td>ln(n + g + δ)</td>
<td>-0.0601 (0.376)</td>
<td>-0.2368 (0.018)**</td>
<td></td>
</tr>
<tr>
<td>EC</td>
<td>0.3532 (0.071)*</td>
<td>0.2830 (0.068)*</td>
<td></td>
</tr>
<tr>
<td>ln(LE)</td>
<td>0.6799 (0.019)**</td>
<td>0.9588 (0.000)***</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.2071 (0.014)**</td>
<td>-1.9575 (0.009)***</td>
<td></td>
</tr>
<tr>
<td>Time dummies</td>
<td>Chi2[6]= 144.21 (0.000)***</td>
<td>Chi2[6]=151.53 (0.000)***</td>
<td></td>
</tr>
<tr>
<td>Wald</td>
<td>Chi2[11]= 23776 (0.000)***</td>
<td>Chi2[17]=32081 (0.000)***</td>
<td></td>
</tr>
<tr>
<td>Hansen (J-test)</td>
<td>Chi2[26]=21.66 (0.707)</td>
<td>Chi2[52]=57.51 (0.279)</td>
<td></td>
</tr>
<tr>
<td>Difference-in-Sargan (C-test)</td>
<td>Chi2[6]=5.62 (0.467)</td>
<td>Chi2[12]=14.89 (0.247)</td>
<td></td>
</tr>
<tr>
<td>AR(2)</td>
<td>z =-2.41 (0.016)**</td>
<td>z = -2.15 (0.031)**</td>
<td></td>
</tr>
<tr>
<td>AR(3)</td>
<td>z = 0.67 (0.670)</td>
<td>z = -0.17 (0.865)</td>
<td></td>
</tr>
<tr>
<td>Instrument count</td>
<td>38</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>771</td>
<td>771</td>
<td></td>
</tr>
<tr>
<td>Countries</td>
<td>131</td>
<td>131</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** See Table 4 notes (1) through (5)
6. L.ln(YC) is instrumented for with system-GMM style instruments. Lags two and earlier of the instrumenting variable is used for the equation in differences, and lags one and earlier of the instrumenting variable in differences for the levels equation. ln(LE) and EC are instrumented for with their corresponding predicted values from the first stage regressions.

Life expectancy is found to have a positive and significant effect on income per capita at the 5% and 1% level in the full and developing world samples respectively. The coefficient on the education capital term exhibits the expected positive sign and, although not significant at the conventional 5% level, we would easily reject the null under a less conservative 10% significance level.

Although we cannot conclude that the effects of education capital on income per capita are significant at the 5% level, observing the expected positive sign, with p-values of around 0.07, represents a promising result relative to many other empirical growth studies – including those investigating the HIV/AIDS epidemic. As in the case of the health equation, the instrument set underlying the lagged dependent variable is appropriately restricted due to the rejection of the AR(2) test. Finally, for both the full and developing world equations, the results of the J and C-test do not invalidate the instrument set.

One observation that can be made regarding the results is in relation to the relative importance of investment in physical and human capital in developed versus developing countries. There is a commonly held view in the development literature that, in developing countries, relative to developed countries, investment in human capital is comparatively more important than investment.
in physical capital. The differing magnitudes of the coefficients on investment and the combined coefficients on the human capital terms in the full and developing world sample are consistent with this view.\textsuperscript{41} Another observation that can be made regards the much larger magnitude of the coefficient on the capital widening term for the developing world sample relative to the full sample. This result is consistent with the development literature where it is often suggested that high rates of fertility and population growth are a major determinant of the poor growth performance of some developing countries relative to others with lower fertility rates and population growth.

### 5.4 The economic impact of HIV prevalence

The impact of HIV prevalence on income per capita can be derived from the coefficients of the estimated equations. The elasticities of income per capita with respect to adult HIV prevalence, calculated at the relevant sample means, are reported in Table 7. Elasticities are also reported for a sub-sample of Sub-Saharan African countries. These should be regarded as preliminary because the estimated equations underlying these remain in their earlier stages and have not yet been subject to robustness testing.

For the full world sample, the total effect of a 1% rise in adult HIV prevalence is, on average, a decrease in income per capita of 0.135%, more than twice as large as the estimates of other similar studies. This corresponds to a semi-elasticity of -13.75, implying that a full 1 percentage point rise in the HIV prevalence rate will result in a 13.75% reduction in income per capita, \textit{ceteris paribus}. These estimates are significantly larger than those found by MR (2006) and Tandon (2005), even if the role the epidemic is having on education capital accumulation is ignored. For the developing world, the elasticity is estimated to be -0.123% and the semi-elasticity -10.87%. The estimated impact due solely to health capital accumulation is slightly higher than that of MR (2006) but, taking into account the role of education capital accumulation, much larger.\textsuperscript{42} For Sub-Saharan Africa, a 1% rise in HIV prevalence is estimated to lead to a fall in income per capita of 0.54% on average. A full one percentage point increase in prevalence is estimated to lead to a 19.9% decrease in income in per capita. The preliminary results for Sub-Saharan Africa are similar to those estimated by MR (2006).

\textsuperscript{41} In further support of this commonly held view, the impact of investment in physical capital for a sub-sample of high-income countries, has a magnitude nearly twice that found in the full-sample analysis and is highly significant.

\textsuperscript{42} The impacts on income per capita that occur through education capital, can only be regarded as statistically significant at the 10% level.
6 Robustness testing

Robustness of the results presented in Section 5 were investigated in the following two ways; (i) the sensitivity of the results to using infant mortality rates as an alternative proxy for health capital were analysed, and (ii) a comparison of the results obtained using alternative estimators such as OLS, FE and difference-GMM estimator were made.

Overall, results were not found to change significantly when using infant mortality rates as an alternative proxy for health capital.

In section 4.4 it was stated that Sys-GMM estimation was considered *a-priori* superior to alternative estimation techniques. When estimating dynamic panels, the coefficient on the lagged dependent variable will be upwardly biased under OLS. When the time dimension is small, the coefficient on the lagged dependent variable will be downwardly biased when applying the FE estimator. The coefficient on the lagged dependent variable will be downwardly biased when using the difference-GMM estimator if the time dimension is small and the dependent variable is highly persistent (Blundell and Bond, 1998). According to Bond (2002), confidence in the consistency of the system-GMM estimator can be gained if estimates are observed to lie somewhere between the OLS and FE estimates and above the difference-GMM estimates, therefore I undertook the task of re-estimating each equation using OLS, FE and difference-GMM. In all cases, the Sys-GMM estimates were found to lie between the OLS and FE estimates and above the difference-GMM estimates, as theory suggests they should. This result is important particularly for the health equation, where the null hypothesis underlying the Hansen J-test was rejected. I interpret this result as suggesting that we can
have some confidence in the results of the health equation notwithstanding the inability to determine the validity of the instrument set.

7 Summary and Conclusions

The objective of this paper was to examine empirically the impact of HIV prevalence on economic growth. The analysis was motivated by an emerging literature that considers the growth impact of the HIV/AIDS epidemic due to its potential to slow human capital accumulation through education. Based on a sample of 142 countries over 45 years, a system of equations for economic growth, health capital, and education capital are estimated. The economic growth model represents a variation of the augmented Solow model that includes human capital in the form of both health capital and a Mincerian specification of education capital. Both forms of human capital are treated as endogenous and inter-linkages between them are explicitly captured.

Two important innovations of my analysis are that: (i) it represents the first econometric assessment that pays particular attention to human capital accumulation through education within an augmented neoclassical growth framework, and (ii) it implements dynamic panel system-GMM estimator that is know to be superior to the more traditional difference-GMM estimator under particular circumstances.

Increased adult HIV prevalence is found to reduce life expectancy, a proxy for health capital, which in turn slows economic growth. Reduced economic growth then reduces the resources available for future investment in health and education. This in turn feeds back into economic growth process. Additionally, an HIV/AIDS induced decline in average life expectancy has a direct and statistically significant impact on investment in education.

For the full world sample, the estimated impact of a 1% rise in adult HIV prevalence is a reduction in income per capita of between -0.135%, an impact that is substantially higher than that identified in previous cross-country empirical research. In a sub-sample of developing countries, the estimated impact is -0.123%. A material component of this estimated impact, is through the epidemics capacity to slow education capital accumulation. These results are considered to be robust. Preliminary results suggest that the impact of HIV/AIDS on income per capita in a sub-sample of Sub-Saharan African countries is considerably large, although work is continuing regarding the final selection of specifications of the estimated equations and robustness testing of results.
The actual economic damage resulting from the HIV/AIDS epidemic could in fact be even larger than that identified in this study as the analysis does not consider the potential for the epidemic to slow physical capital accumulation and technological growth – impact channels that are potentially important. A shortcoming of our analysis, one that would also suggest that the estimated impacts are conservative, is that the chosen proxy for education capital may not be capturing the detrimental impact HIV/AIDS could have on informal human capital accumulation, that is, the skills of the workforce obtained either through job experience or on-the-job training.

In conclusion, this study finds that the HIV/AIDS epidemic involves not only devastating human and social costs, but also large economic costs. The magnitude and sources of the estimated impacts are consistent with the broad conclusions arising from an emerging body of literature that identifies as an important impact channel the potential for HIV prevalence to reduce economic growth by undermining the accumulation of education capital. This literature had to date investigated this impact channel through simulations of theoretical overlapping generations or CGE models based on individual or small groups of African countries. Here, the impact of the HIV/AIDS epidemic on cross-country economic growth, incorporating this previously overlooked impact channel, was directly estimated using data for 142 countries spanning 45 years.

The broad policy implications that have arisen from this emerging literature are supported by our results. These policy implications include; (i) that in countries considered at particular risk of increased HIV prevalence in the future, from a purely economic cost-benefit perspective, devotion of substantial preventative resources is justified to avoid large long-run economic damage, and (ii) for those countries already in the midst of a severe epidemic, not only should resources be devoted to tackling the epidemics spread and improving the quality and productivity of life for HIV sufferers, they should also be directed to maintaining participation in, and quality of, the education system that in turn would prevent a breakdown in the transmission of human capital across generations.
Appendix A

Table 1: HIV prevalence in the 15-49 years age group as a percentage of all persons 15-49 years of age

<table>
<thead>
<tr>
<th></th>
<th>2005 %</th>
<th>2003 %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Africa (Worst Eight)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swaziland</td>
<td>33.4</td>
<td>32.4</td>
</tr>
<tr>
<td>Botswana</td>
<td>24.1</td>
<td>24.0</td>
</tr>
<tr>
<td>Lesotho</td>
<td>23.2</td>
<td>23.7</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>20.1</td>
<td>22.1</td>
</tr>
<tr>
<td>Namibia</td>
<td>19.6</td>
<td>19.5</td>
</tr>
<tr>
<td>South Africa</td>
<td>18.8</td>
<td>18.6</td>
</tr>
<tr>
<td>Zambia</td>
<td>17.0</td>
<td>16.9</td>
</tr>
<tr>
<td>Mozambique</td>
<td>16.1</td>
<td>16.0</td>
</tr>
<tr>
<td><strong>Rest of the World (Worst Eight)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haiti</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>Bahamas</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Trinidad &amp; Tobago</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Belize</td>
<td>2.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Guyana</td>
<td>2.4</td>
<td>2.4</td>
</tr>
<tr>
<td>Suriname</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Papua New Guinea</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Cambodia</td>
<td>1.6</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>Weighted averages by region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>6.1</td>
<td>6.2</td>
</tr>
<tr>
<td>Caribbean</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>North America</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Eastern Europe &amp; Central Asia</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>South &amp; South-East Asia</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Latin America</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Oceania</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Western &amp; Central Europe</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>North Africa &amp; Middle East</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>East Asia</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Source: UNAIDS (2006)

Appendix B

B1 Variable Description

This section provides a brief description of the variables considered in this study

Economic indicators

- **Real GDP per capita (YC):** is equal to GDP adjusted for Purchasing Power Parity (PPP) divided by the average population during a year. Adjusting for PPP results in a common set of prices in a common currency allowing for real quantity comparisons. These quantities are expressed in current prices. This variable was sourced from PWT 6.2 (2006).

43 See Heston et al. (2006) for a detailed description of all data series obtained from PWT 6.2.
• **Investment share of GDP (INV):** is equal to domestic investment (in physical capital) divided by GDP (in current prices) and was sourced from PWT 6.2 (2006). In the model’s specification, investment is assumed exogenously given and constant over time for each country. As such, the average value for each country over the entire period 1970 to 2004 is used.

• **Capital depreciation rate (δ):** is a time invariant, exogenously given rate at which all forms of capital depreciate.

• **Technological growth rate (g):** is equal to the annual improvement in Total Factor Productivity (TFP) and is assumed to be exogenously given. Economists define TFP as the actual measured growth of output minus the growth rate expected from increases in all forms of capital and labour.

*Health indicators*

• **Life expectancy at birth (LE):** is equal to the number of years a newborn would live if prevailing patterns of mortality were to stay in place throughout their life. This variable was sourced from the WDI (2006) database.

• **Infant mortality rate (INF):** Proportion of infant deaths (deaths of individuals under one year of age) per 1000 live births. This variable was sourced from the WDI (2006) database.

• **Calorie intake per capita (CPC):** is equal to the total amount of calories from all food sources consumed by the population during the year. It is expressed on a per person, per day basis. This variable is derived from national food balance sheets produced by FAO (Food and Agricultural Organisation of the United Nations). Food available for consumption is estimated as a balancing item by considering total national supply and deducting the amounts used for feed, seed, industrial uses and waste. The new online database released in 2006 utilises an improved methodology to compute food consumption. FAO have restated the time-series back to 1990 utilising this new methodology and these figures were used for the post-1990 period. Pre-1990 data was sourced from the FAO CD-ROM (1998). For the period from 1990-1998 estimates based on the old methodology were observed to occasionally be materially different to that under the new methodology. As such, pre-1990 data was generated by taking the 1990 estimate from the new time series and applying the percentage changes based on the old series to generate a series dating back to 1970.

• **Adult prevalence of HIV (HIV):** is equal to the estimated number of people aged 15–49 with HIV divided by the total population of 15-49 year olds. The time series estimates of HIV prevalence are sourced from UNAIDS (2006) and represent an improvement on previously available data due to better estimation methods and improved data availability. UNAIDS have only made available to the general public the 2005 and 2003 estimates based on this new methodology. UNAIDS were kind enough to provide a full time series back to 1980 using the new methodology for use in this study on the understanding that individual country time series estimates were not published. The time series provided did not include some Western European and North American high income countries and as such only the observation in the publicly available UNAIDS (2006) report are used.

• **Proportion of the population at risk of Malaria (MAL):** is equal to the percentage of the population living in areas of high malaria risk in a country in a particular year. The figures are based on estimates of the world distribution of malaria and a 1995 estimate of the world population distribution. This variable was sourced from Gallup et al. (1999). Data is only available at four points in time - 1946, 1966, 1982 and 1994. Between 1966 and 1982 straight-line interpolation is applied to produce estimates for the 1970-74 and 1975-79 time intervals. The 1982 observation is used as the 1980-84 observation. Similarly, straight-line interpolation is
applied between 1982 and 1994. The observation for the 2000-04 period is produced by extending the ten-year trend to 1994 forward. If this process resulted in a number greater than one or less than zero, then it was changed to one or zero respectively. The 1995-1999 observation was generated by applying straight-line interpolation between the 1990-1994 and 2000-2004 figures.

**Education indicators**

- **Average years of schooling in the population aged 15 and over (SCH):** The Barro and Lee (2001) database is the most widely used measure of average years of schooling in cross-country empirical work. Data is available for 105 countries at 5 yearly intervals from 1960 through to 2000. The database is constructed by drawing on UNESCO’s national censuses and surveys. The UNESCO data identifies whether individuals have completed no schooling, partly or fully completed primary, partly or fully completed secondary, or partly or fully completed tertiary. Based on these censuses and surveys, the average years of schooling in the population aged 15 and over is calculated as: \[ SCH = \sum_a \left( \frac{N_a}{P_{15}} \sum_c D_c \right) \], where \( N_a \) is the number of individuals for whom \( a \) is the highest level of schooling achieved (i.e. no schooling, primary, secondary, or tertiary), \( P_{15} \) is the total population of individuals aged 15 years or over, and \( D_a \) is the duration of schooling at level \( a \). Barro and Lee (2001) take into account variations in the duration of each level of schooling over time within a country. For years where census or survey data is not available, Barro and Lee (2001) estimate missing observations by drawing on other UNESCO data including literacy rates, school enrollment ratios at each level of education, dropout rates and repetition rates. Cohen and Soto (2007) have constructed an alternative database of average years of schooling that addresses some of the measurement error concerns surrounding the Barro and Lee (2001) dataset as raised by Krueger and Lindahl (2001), de la Fuente and Domenech (2006), and others. Cohen and Soto (2007) draw upon a wider range of survey and census information than does Barro and Lee (2001) and apply an improved methodology for extrapolating missing data. Notwithstanding this, the Barro and Lee (2001) was chosen as my primary source because the Cohen and Soto (2007) database includes 10 fewer countries overall, and data is only provided at 10 yearly intervals. Although the overall coverage of the Cohen and Soto (2007) data is narrower, it does include data for twelve countries that is not covered by Barro and Lee (2001). I therefore draw on Cohen and Soto (2007) for these twelve countries, and to generate 5-yearly observations, I simply assume that half the difference between the 10-yearly data points is closed within five years.

- **Index of education quality (QL):** Altinok and Murseli (2007) have constructed a cross-country index of education quality based on international surveys of student’s learning achievement. Results of seven different groups of surveys covering mathematics, science and reading are utilised. By first matching countries that took part in multiple surveys simultaneously, an anchoring of survey results relative to each other is achieved. Survey results were then re-adjusted to a common scale of 0 to 100. Two different databases are provided by Altinok and Murseli (2007), one is simply a cross-sectional database based on the most recently conducted mathematics, science and reading surveys. An index is calculated separately for mathematics, science and reading, with an overall quality index being the arithmetic average of these (QIHC-G). This is the index utilised in my study. This cross-sectional database covers 105 countries. Although a panel database is available covering the years 1960 to 2005, I do not choose to use
this because there are very few countries with more than two observations. I normalize the QIHC-G index by setting the value for the United States to one.

- **Government expenditure on education as a proportion of GDP (EDS):** is equal to total public expenditure on education (current and capital expenses) divided by GDP for a given year. Data from 1998 onwards was sourced from the WDI database, whilst pre-1998 data was sourced from the UNESCO databases.

**Demographic indicators**

- **Population growth rate (n):** is equal to the annual percentage change in population and is sourced from the WDI database.
- **Proportion of population aged 15 or under (POP15):** is equal to the population aged 15 or under divided by the total population and is sourced from the WDI database.
- **Proportion of population living in urban areas (URB):** is equal to the population living in urban areas, divided by the total population and is sourced from the WDI database.

**B2 The Sample**
The WDI database contains data for 208 countries or territories and represents the widest coverage of all the data sources utilized. Inclusion in this database does not necessarily mean that data is available for each of the indicators for each time period. Using the coverage in this database as a starting point, countries were progressively removed from the sample if they met one of two conditions:

1. The country did not exist in its current form for the whole of the time period (for example, members of the former Soviet Union, Germany, Yugoslavia, East Timor etc.)
2. Data was missing entirely for the majority of variables or for the majority of time periods for a majority of variables

This process resulted in the inclusion of 147 countries. A further five small oil producing nations were removed from the sample. This was due to the well-known unreliability of income per capita measures for these countries from PWT 6.2. The resulting final sample consists of 142 countries. The countries included are: Algeria, Angola, Argentina, Armenia, Australia, Austria, The Bahamas, Bahrain, Bangladesh, Barbados, Belgium, Belize, Benin, Bhutan, Bolivia, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Central African Republic, Chad, Chile, China, Colombia, Comoros, Democratic Republic of Congo, Republic of Congo, Costa Rica, Côte d’Ivoire, Cuba, Cyprus, Denmark, Djibouti, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Fiji, Finland, France, Gabon, The Gambia, Ghana, Greece, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Islamic Republic of Iran, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Korea Republic, Kuwait, Lao People’s Democratic Republic, Latvia, Lebanon, Lesotho, Liberia, Libya, Luxembourg, Madagascar, Malawi, Malaysia, Mali, Malta, Mauritania, Mauritius, Mexico, Mongolia, Morocco, Mozambique, Myanmar, Namibia, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Rwanda, Samoa, Saudi Arabia, Senegal, Sierra Leone, Singapore, Somalia, South Africa, Spain, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syrian Arab Republic, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, United Kingdom, United States, Uruguay, Uzbekistan, Vanuatu, Venezuela, Vietnam, Zambia, and Zimbabwe.
**B2 Summary Statistics**

**Table 3: Summary Descriptive Statistics – Full and Sub-Samples**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>Mean</td>
<td>St. Dev.</td>
</tr>
<tr>
<td>YC</td>
<td>142</td>
<td>9864</td>
<td>10358</td>
</tr>
<tr>
<td>INV</td>
<td>142</td>
<td>13.7</td>
<td>7.6</td>
</tr>
<tr>
<td>EDS</td>
<td>108</td>
<td>4.6</td>
<td>1.9</td>
</tr>
<tr>
<td>N</td>
<td>143</td>
<td>1.55</td>
<td>1.02</td>
</tr>
<tr>
<td>POP15</td>
<td>142</td>
<td>32.7</td>
<td>10.4</td>
</tr>
<tr>
<td>URB</td>
<td>142</td>
<td>54.5</td>
<td>24.7</td>
</tr>
<tr>
<td>HIV</td>
<td>137</td>
<td>2.47</td>
<td>5.32</td>
</tr>
<tr>
<td>LE</td>
<td>142</td>
<td>64.6</td>
<td>13.1</td>
</tr>
<tr>
<td>MAL</td>
<td>135</td>
<td>0.38</td>
<td>0.45</td>
</tr>
<tr>
<td>CPC</td>
<td>127</td>
<td>2806</td>
<td>752</td>
</tr>
<tr>
<td>SCH</td>
<td>131</td>
<td>5.91</td>
<td>2.79</td>
</tr>
</tbody>
</table>

**Developing and Least Developed Countries (i=110)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Obs</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Obs</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>YC</td>
<td>109</td>
<td>4626</td>
<td>3899</td>
<td>388</td>
<td>17010</td>
<td>99</td>
<td>2551</td>
<td>2024</td>
<td>295</td>
<td>10769</td>
<td>97</td>
<td>887</td>
<td>714</td>
<td>72</td>
<td>3667</td>
</tr>
<tr>
<td>INV</td>
<td>109</td>
<td>11.5</td>
<td>6.2</td>
<td>2.3</td>
<td>30.7</td>
<td>99</td>
<td>11.9</td>
<td>6.2</td>
<td>2.6</td>
<td>34.0</td>
<td>97</td>
<td>13.7</td>
<td>8.3</td>
<td>1.7</td>
<td>42.4</td>
</tr>
<tr>
<td>EDS</td>
<td>81</td>
<td>4.29</td>
<td>1.89</td>
<td>0.61</td>
<td>9.59</td>
<td>88</td>
<td>3.78</td>
<td>2.12</td>
<td>0.40</td>
<td>10.90</td>
<td>75</td>
<td>3.57</td>
<td>1.40</td>
<td>0.60</td>
<td>7.70</td>
</tr>
<tr>
<td>N</td>
<td>110</td>
<td>1.66</td>
<td>0.98</td>
<td>-0.92</td>
<td>3.41</td>
<td>110</td>
<td>2.27</td>
<td>1.09</td>
<td>-0.71</td>
<td>4.97</td>
<td>110</td>
<td>2.41</td>
<td>1.03</td>
<td>-1.24</td>
<td>6.14</td>
</tr>
<tr>
<td>POP15</td>
<td>110</td>
<td>36.5</td>
<td>8.6</td>
<td>14.9</td>
<td>50.4</td>
<td>110</td>
<td>40.7</td>
<td>6.9</td>
<td>21.0</td>
<td>49.7</td>
<td>109</td>
<td>42.5</td>
<td>6.0</td>
<td>20.4</td>
<td>49.4</td>
</tr>
<tr>
<td>URB</td>
<td>110</td>
<td>46.0</td>
<td>21.7</td>
<td>9.2</td>
<td>92.0</td>
<td>110</td>
<td>39.8</td>
<td>20.9</td>
<td>5.2</td>
<td>87.9</td>
<td>110</td>
<td>32.1</td>
<td>19.3</td>
<td>2.7</td>
<td>82.8</td>
</tr>
<tr>
<td>HIV</td>
<td>107</td>
<td>3.17</td>
<td>5.93</td>
<td>0.00</td>
<td>31.03</td>
<td>101</td>
<td>0.41</td>
<td>1.04</td>
<td>0.00</td>
<td>5.91</td>
<td>110</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>LE</td>
<td>110</td>
<td>60.3</td>
<td>12.3</td>
<td>37.3</td>
<td>78.3</td>
<td>110</td>
<td>58.8</td>
<td>9.7</td>
<td>39.3</td>
<td>74.6</td>
<td>108</td>
<td>53.0</td>
<td>10.0</td>
<td>35.1</td>
<td>71.1</td>
</tr>
<tr>
<td>MAL</td>
<td>110</td>
<td>0.50</td>
<td>0.45</td>
<td>0.00</td>
<td>1.00</td>
<td>110</td>
<td>0.57</td>
<td>0.41</td>
<td>0.00</td>
<td>1.00</td>
<td>110</td>
<td>0.63</td>
<td>0.39</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CPC</td>
<td>101</td>
<td>2634</td>
<td>754</td>
<td>1261</td>
<td>4819</td>
<td>101</td>
<td>2438</td>
<td>500</td>
<td>1409</td>
<td>3835</td>
<td>106</td>
<td>2531</td>
<td>447</td>
<td>1515</td>
<td>3562</td>
</tr>
<tr>
<td>SCH</td>
<td>101</td>
<td>4.92</td>
<td>2.24</td>
<td>0.84</td>
<td>10.59</td>
<td>95</td>
<td>3.84</td>
<td>2.13</td>
<td>0.34</td>
<td>9.26</td>
<td>79</td>
<td>2.86</td>
<td>2.11</td>
<td>0.10</td>
<td>9.71</td>
</tr>
</tbody>
</table>

**Source:** See Table 2

**Notes:** A full description of variables is provided in Appendix B1
References


