Human Development
Research Paper
2010/31

Explaining the Cross-National Time Series Variation in Life Expectancy: Income, Women's Education, Shifts, and What Else?

Lant Pritchett
and Martina Viarengo

United Nations Development Programme
Human Development Reports
Research Paper

October 2010

Human Development
Research Paper
2010/31

Explaining the Cross-National Time Series Variation in Life Expectancy: Income, Women's Education, Shifts, and What Else?

Lant Pritchett
and Martina Viarengo

United Nations Development Programme
Human Development Reports
Research Paper 2010/31
October 2010

Explaining the Cross-National Time Series Variation in Life Expectancy: Income, Women's Education, Shifts, and What Else?

Lant Pritchett
Martina Viarengo

Lant Pritchett (corresponding author) is Professor of the Practice of Economic Development at the Kennedy School of Government at Harvard University. Email: Lant_Pritchett@harvard.edu

Martina Viarengo is an Economist at the London School of Economics and a Postdoctoral Fellow at the Kennedy School of Government at Harvard University. Email: Martina_Viarengo@hks.harvard.edu

Comments should be addressed by email to the author(s).

Abstract

This paper examines the variation across countries and evolution over time of life expectancy. Using historical data going back to the beginning of the 20th century several basic facts about the relationship between national income and life expectancy are established. The paper shows that even by examining the augmented Preston curve there is no indication that the Preston curve is "breaking down" and no indication from over 100 years of data that a very strong relationship between national income and life expectancy will not persist, particularly over the ranges of income of primary interest to the Human Development Report.

Empirical findings show that there are actually fewer "puzzles" than might appear while trying to reconcile the strong cross-sectional association with the time evolution of life expectancy in specific countries and most of the existing "puzzles" come from using either very short time-horizons or very small moves in income per capita when the Preston curve is a long-run phenomena.

The paper also discusses the phenomena of the cross-national convergence, with the life expectancy of the poorer countries increasing, in absolute terms, faster than those of the rich countries and how the findings about the augmented Preston curve relate to discussions of health policy.

Keywords: economic development, economic growth, health, life expectancy, mortality.

JEL Classification: I10, O1, O40, J11

The Human Development Research Paper (HDRP) Series is a medium for sharing recent research commissioned to inform the global Human Development Report, which is published annually, and further research in the field of human development. The HDRP Series is a quick-disseminating, informal publication whose titles could subsequently be revised for publication as articles in professional journals or chapters in books. The authors include leading academics and practitioners from around the world, as well as UNDP researchers. The findings, interpretations and conclusions are strictly those of the authors and do not necessarily represent the views of UNDP or United Nations Member States. Moreover, the data may not be consistent with that presented in Human Development Reports.

Introduction[1]

This paper examines the variation across countries and evolution over time of life expectancy.

The opening section examines the impact of national income, measured as GDP per capita in PPP, in Preston and augmented Preston regressions. Rather than focus only on recent cross-sections since 1970 or so we use the available historical data going back to the beginning of the 20th century (the data are taken from the series created for the GAPMINDER application and are described in the data appendix). This long-run focus allows us to establish several basic facts about the relationship.

First, there has been a strong cross-national relationship between income and life expectancy for as far back as one can take the data. In the simple double natural log Preston curve (life expectancy regressed on GDP per capita) the R-squared for the 21 countries with data was as high as .8 as early as 1927 and was at that level through the pre-World War II period. The modern data sets with over 150 countries begin in 1952 and have availability every five years and in that data there has been a high and rising R-squared roughly ever since (once one controls for the AIDs affected countries).

Moreover, with the modern data one can create an augmented Preston relationship adding women's education and other variables (such as a variable to adjust for the onset of AIDS). In the simple double log augmented Preston regression the R-squared rises from .63 in 1952 to .79 in 2007. There is a stable cross-sectional relationship in which countries' life expectancy is now *more tightly* associated with the augmented Preston variables. There is no

[1] Many thanks to comments from Jeni Klugman, Fransisco Rodriqguez, and the HDR team that were helpful in revising the paper, with the usual caveats that they have no responsibility for the views and errors that the paper contains.

indication the Preston curve is "breaking down" and no indication from over 100 years of data that a very strong relationship between national income and life expectancy will not persist, particularly over the ranges of income or primary interest to the Human Development Report.

Second, the relationship between life expectancy and GDP per capita is not linear and, while the non-linearity is reasonably well approximated by a double natural log specification we also estimated augmented Preston regressions allowing for quite general non-linearity by allowing for up to a fourth power in YPC. In whatever functional form, per capita income has a strong and statistically robust association with life expectancy, especially over the range up to GDP per capita PPP$10,000. Interestingly, with a more relaxed functional form than the constant elasticity imposed by the ln-ln form the association with income gets *stronger* medium levels of income than at the lower levels so there is no sense that the income impact "tapers off" (in elasticity terms), at least over these ranges. The association appears to have been falling modestly over time, from an ln-ln elasticity of .10 in the early 1970s to around .08 today.

Third, much of the discussion is about trying to reconcile the strong cross-sectional association with the time evolution of LEX in specific countries. Here we argue that there are actually fewer "puzzles" than might appear. Many countries have seen improving LEX even with stagnant or slowly growth YPC but this is not a puzzle per se as the augmented Preston curve has shifted over time and so a country with stagnant per capita income, even if it stayed precisely on the predicted Preston curve, would improve over time.

Most of the other "puzzles" come from using either very short time-horizons or very small moves in YPC when the Preston curve is a long-run phenomena. So the answer to the question of "How should I expect a country's LEX to evolve over the long-run or as income

grows rapidly over an extended period?" is "Pretty much along the Preston curve." The Preston curve is one of the most amazingly stable and robustly estimated relationships in economics. Either countries' long-run trajectories or long-period panel regressions with fixed effects show income elasticities that are remarkably, if not *amazingly*, similar to the cross-sectional Preston curve. For instance, the average ln-ln income elasticity from the cross-sectional augmented Preston relationships from 1952 to 2002 is .112. If one uses the 51 countries for which there are observations before 1950 and which have at least 15 years of data the panel (changes on changes) estimate of the income elasticity is .110. If one uses the 27 countries with more than 50 years of observations to estimate the country specific income elasticity the median elasticity is .095 and 20 of the 27 country elasticities are within the range of .082 to .12. We illustrate that a persistently high R-squared is consistent with a very narrow range of long-run trajectories for countries—if countries own LEX-YPC trajectories differed one would expect the association to weaken over time.

However, if one attempts to use the Preston curve to guess what will happen over a year or two or over extended periods of stagnant growth then the answer is that, statistically, pretty much anything can happen, but without creating much of a puzzle. That is the augmented Preston curve is a quite amazingly stable relationship but which is identified off of the massive cross-national differences in YPC. But *doubling* per capita GDP per capita is predicted to increase LEX (in 2002) at low levels of income by about 3.5 to 5 years. So any given country could have extended periods of moderate growth and either rapid or stagnant growth in LEX and not wander outside the confidence intervals simply because this would not trace out very much distance along the Preston curve. Confusion can arise about reconciling the cross-sectional and time series evidence if one does not pay attention to the basics of statistical power and measurement error, which can lead to any relationship becoming "statistically insignificant."

The second section of the paper discusses the phenomena of the cross-national convergence, with the LEX of the poorer countries increasing, in absolute terms, faster than those of the rich countries. This might seem to create something of a puzzle as it is well known that YPC has been diverging in natural log units. However, our decomposition analysis shows that there is actually not a tremendous puzzle as there has actually been almost no convergence in natural log of LEX—the variance of LEX was roughly the same in 1982 and 2007.

The third section discusses how the findings about the augmented Preston curve relate to discussions of health policy. Basically, however these debates about the relationship of income and mortality are resolved, the case for health policy and the formulation of health policy is not much affected as they much depend on much deeper and more sophisticated analytics and data than can be produced from cross-national comparisons.

I) Basics of the relationship of national income and life expectancy

While there are a number of complications of moving from the household to the aggregated data that represent national averages, we find it conceptually easiest to frame expectations of what we would expect to find in the cross-national data by thinking about a single household[2].

There are two ways of going about thinking about the causes of death. One is to trace out specific medical causes of death as diseases and decompose that into probability of experiencing a given disease condition times fatality risk per disease condition. The other is

[2] For simplicity a single unitary household even though there are obviously gender differences.

to start from a model of household choice and see how risks of death are associated with household characteristics (without necessarily tracing it through specific disease conditions). Obviously eventually both should be reconciled (as disease incidence and fatality per incidence are both affected by conditions household face beyond their control plus household resources and decision making), but in this note we are focused mainly on the association of life expectancy with distal conditions not proximate causes.

It is easy enough to outline four broad categories of household characteristics that will be associated with health outcomes and hence risk of pre-mature death.

Command over material resources. Command over material resources is going to matter for at least three reasons. First, a variety of decisions about consumption are going to affect a households' health status, even if those decisions are not taken primarily for health reasons. The quantity and quality of food, the quality of housing, the quality of water and sanitation facilities, the quality of clothing, are all associated with higher total consumption expenditures and potentially have positive health effects, even if health effects are not the primary reason—e.g. households might be better water connections for convenience with a consequence of lower cost of keeping things clean. Second, people with more command over material resources can make purchases that *avoid* disease conditions. Third, people with greater command over material resources are more easily able to seek treatment that improve outcomes, conditional on a disease—more effective curative care. While sometimes the income is associated only with this latter, richer people get better (more prompt, more effective) curative care, this may only be a very small part of the explanation of an observed health-income link at the household level.

Ability to utilize resources effectively. A second component of a household's health status is an ability to utilize resources effectively to promote health. This includes a variety of

things from knowing good health practices (e.g. sanitation), knowing effective responses to health conditions (e.g. knowing of ORS), the ability to know when to seek medical expertise and follow their guidance.

These first two features are mainly features of the household itself. Then there are two broad factors external to the household.

Global technological capability. In health, as in many other domains, the application of science and technology has made technologically possible a variety of interventions of both preventive (e.g. effective immunizations), low-cost curative (e.g. ORS, penicillin and antibiotics), and high cost curative (e.g. heart surgeries).

Domestic Health Public Policy. The government can affect the ability of households in a huge variety of ways, from health promotion campaigns to disseminate information and encourage some behaviors (immunization, contraception, breastfeeding, sanitation) and discourage others (smoking, risky sex), to making preventive care services available (e.g. immunizations, targeted child nutritional services), to subsidizing curative care services either through direct production or through financing.

These four are obviously just simple headings under which a great deal of complexity is hidden, but even this simple structure can be a helpful way of asking what we might expect from an exploration of the cross-national data.

First, people demonstrably care a great deal about their health and the health of their children. They also have a variety of other needs, wishes, wants, desires but avoiding death and remaining healthy is clearly an objective of nearly everyone. This would mean we should expect people with more command over resources to have better health status and less pre-mature deaths. In fact, it would be truly stunning of it were *not* true.

Second, education, and in particular, women's education one would expect to be strongly associated with health, of both adults and children, for a huge variety of reasons. Again, it would be stunning if a measure of women's capability to process information and utilize resources effectively to pursue their objectives were not strongly related to health status.

Third, a president of the United States, Calvin Coolidge, an unquestionably rich and powerful man of his day, had a son die at 16 years old from an infected blister on his toe in 1924. There is no question that science and technology continuously expand the frontiers of the possible. It would be stunning if the health status on average did not improve over time, even for people of the same income and personal capability. How this translates into cross-national differences depends on how technological frontier expands and how knowledge and practices diffuse across countries.

These first three factors are well known and are the basics of a cross-national decomposition of a measure of average health status of a population, such as infant or child mortality or life expectancy. Below we will use these to explore the plain and augmented "Preston curve" (the cross-sectional relationship between national income and life expectancy) and its shifts over time in order to understand the evolution over time of health status cross-nationally (and why it has converged while income has diverged).

The more complex question is what we would expect to be the relationship between "public policy" broadly taken at the national level and measures of pre-mature death—like child mortality or life expectancy. We return to that question only in Section III.

I.A) Standard factors explain high proportion of variance, increasing over time

We start examining the relationship between life expectancy and real per capita income by estimating the Preston curve. In his pioneering study, Samuel Preston (1975) investigated

the influence of economic conditions on life expectancy for a cross-section of countries and found the relationship to hold at different points in time (1900s, 1930, and 1960s). There are two basic facts that became well-known as a result of Preston's analysis. First, the relationship between income and life expectancy is non-linear and concave. Second, Preston curve shifts upward (e.g. life expectancy improves at a given income) at all income levels over time. These specific features of the Preston curve reveal the importance of improvements in income for the poorest countries where small changes in income per capita are associated with larger improvements in life expectancy as well as the importance of technological progress (and effective domestic health public policy).

The importance of income per capita for life expectancy in Preston's and later studies has challenged the conventional view on the role of health policy and specific health interventions as being the only factors that have improved life expectancy in low income countries (Deaton 2007). Filmer and Pritchett (1999) find that these have extraordinarily high ability to explain health differences across countries: only six economic and social factors (average income, income inequality, women's education, ethno-linguistic fractionalization, the predominance of Muslim population and regional dummies) produce an R-squared of .95 and hence explain essentially *all* of the cross-country variation in infant and child mortality[3].

We start by estimating this basic Preston relationship going back into history and estimating a Preston curve every five years for the 1902- 2007 time period. We estimate the

[3] This is a data set with only "high quality" observations on child mortality, which eliminated from the standard data sets observations based on extrapolations. Obviously these add noise and lower potential explanatory power. Even in this data set Filmer and Pritchett (1999) use data on repeated household survey estimates of child mortality for the same period to suggest that pure measurement error in child mortality likely accounts for between 2.5 and 5 percent of total cross-national variance. This means the available 'true' variance unexplained by these factors was between 2.5 and 0 percent of the total.

following equations, double natural log and quartic, both of which allow for non-linearity in the relationship in the absolute levels of the variables:

$$\ln(LEX)_{it} = \alpha + \beta_{\ln} \ln(YPC_{it}) + \varepsilon_{it} \quad (1)$$

Where:

$\ln(LEX_{it})$ is the natural logarithm of life expectancy for country i^{th} at time t^{th}

$\ln(YPC_{it})$ is the natural logarithm of GDP per capita (PPP) for country i^{th} at time t^{th}

ε_{it} is the error term with standard errors clustered at country-level

α is a constant, and β_1 is the coefficient on income.

Or

$$LEX_{i,t} = \alpha_q + \sum_{k=1}^{4} \beta_k * YPC_{i,t}^{k} + \eta_{ii.t}$$

We are going to being with the ln-ln functional form as existing studies show that the life expectancy/income per capita relationship is non-linear and that the double log transformation appears to be roughly appropriate. Moreover, coefficients expressed in logs facilitate interpretation and comparisons (Filmer and Pritchett 1999). But the quartic form also allows for considerable flexibility in the non-linearity and we can compute elasticities at various levels of GDP per capita to compare with the log-log without imposing the constant elasticity assumption of ln-ln.

To estimate the long run relationship between income and life expectancy we rely on data sources for which a description is provided in the appendix. Data on life expectancy at birth is available for a limited number of countries in the long run, and for a larger sample starting in 1952 when the data of the United Nations Department of Economic and Social Affairs start

becoming available. On the other hand, income per capita data are based on Penn World Tables 6.2 and World Development Indicators and Angus Maddison's historical statistics for the long run. These collections of demographic and income statistics have the advantage of being comparable across countries and available over longer than a century. We acknowledge that data on life expectancy is not always high quality data particularly for the early years of the twentieth century and for the poorest countries, and this may be an issue when examining changes over time[4].

Regression results are presented in table 1. We show coefficients of ln-ln and quartic estimations as well as elasticities of the quartic at different income levels. It is possible to observe the existence of a stable relationship over the century. By looking at the estimated coefficient of the ln-ln regression we can observe that the elasticity has been stable and also that life expectancy has increased at all income levels over the century. While it is not really possible to compare the pre and post 1952 results due to the very different samples, it is that case that, among the countries for which there are data, the R2 was around .8 from 1927 to 1947, and rose from .54 in 1952 to around .6 from 1977 to 2007.

Even if we allow for a more flexible functional form and we look at the quartic regressions we can observe quite stable results. Income elasticity has been declining over time. The quartic specification provides a only slightly better fit than the ln-ln functional form, the R^2 is higher for the polynomial in every year.

[4] Pritchett and Summers (1996) provide a discussion on the life expectancy variable as well as on the advantages and possible limitations of using international datasets. Deaton (2007) describes how an accurate measurement of adult life expectancy requires the existence of a vital registration system which is still missing in many low-income countries.

Table 1: Basic Preston curve in double-log or quartic shows a high R2 since at least 1922

		ln-ln specification		Elasticity at various levels of Income per capita using the quartic functional form				R2 of quartic in
Year	N	β	R2	1250	2500	5000	10000	YPC
1902	15	0.239	0.319	0.708	0.488	-0.286	5.434	0.466
1907	15	0.320	0.515	0.078	0.975	-0.687	6.857	0.683
1912	15	0.244	0.462	0.238	0.362	0.093	2.010	0.489
1917	14	0.233	0.303	0.510	0.217	0.350	11.755	0.412
1922	20	0.351	0.606	-0.207	0.645	0.014	6.438	0.694
1927	21	0.305	0.805	0.234	0.433	0.286	1.045	0.826
1932	21	0.316	0.862	0.210	0.437	0.289	1.014	0.887
1937	25	0.263	0.765	0.066	0.316	0.391	0.196	0.837
1942	22	0.314	0.804	0.331	0.438	0.357	-0.196	0.886
1947	24	0.251	0.793	0.314	0.344	0.233	-0.006	0.846
1952	151	0.177	0.541	0.182	0.246	0.235	0.074	0.615
1957	152	0.165	0.554	0.164	0.226	0.221	0.067	0.621
1962	153	0.157	0.565	0.151	0.211	0.207	0.060	0.620
1967	158	0.140	0.553	0.135	0.193	0.198	0.049	0.603
1972	171	0.124	0.559	0.079	0.131	0.180	0.158	0.585
1977	174	0.120	0.594	0.096	0.152	0.191	0.128	0.666
1982	174	0.118	0.663	0.089	0.142	0.180	0.127	0.705
1987	174	0.115	0.715	0.094	0.146	0.175	0.105	0.748
1992	188	0.115	0.622	0.094	0.143	0.170	0.098	0.687
1997	188	0.112	0.657	0.089	0.138	0.166	0.102	0.678
2002	191	0.113	0.611	0.076	0.123	0.162	0.131	0.617
2007	185	0.106	0.590	0.062	0.103	0.142	0.129	0.593

A strong association between income per capita and health mortality outcomes (i.e., life expectancy, infant and child mortality) is now widely accepted. Yet, there is no universal agreement on the mechanisms through which income may affect life expectancy and on the relative importance of different factors that affect life expectancy (Bloom and Canning 2007). Pritchett and Summers (1996) suggest that the relationship between income per capita and health outcomes is causal by using as instruments variables that have been shown to be associated with growth but which are arguably not associated with health except insofar as they affect income (i.e., terms of trade shocks, the ratio of investment to GDP, black market premium for foreign exchange and the deviation of the official exchange rate from its

purchasing power parity level), although it should be pointed out that it is not exactly clear what instrumented cross-national regressions identify[5]. They estimate a medium run income elasticity from long period panels for infant and child mortality in developing countries between around -0.4 (smaller the shorter the period).

In the existing literature other factors apart from income have been found to have a positive effect on health outcomes. We are going add to the basic regression some non-income components to estimate what we call an augmented-Preston curve. Specifically, we are going to include the following explanatory variables:

Women's Education: higher levels of women's education appear to be positively related with different positive health outcomes such as decrease in infant mortality and increase in life expectancy (Behrman and Deolalikar 1988; Pritchett and Filmer 1999); beneficial effect on a child's health, schooling and adult productivity (Behrman 1997; Strauss and Thomas 1995); larger beneficial effect than would adding to a father's schooling of the same amount (King and Hill 1993). We are going to use women's average years of schooling from Barro and Lee (2001). Years of schooling are preferred to other education variables that are either bounded or measure the flows such as literacy and enrollments (for a discussion on the advantages/shortcomings of the flow/stock education variables see Woessmann (2003)).

Predominantly Muslim: following Caldwell (1986), Filmer and Pritchett (1999) we include an indicator that takes into account the fact that Islamic countries exhibit high infant mortality and high death in childbirth rates. We use a dummy variable equal to one if the

[5] We are not assessing the evidence that health might have some impact on economic performance as some regressions have suggested that lagged indicators of health status are associated with future growth. In contrast, Acemoglu and Johnson (2007) use the international epidemiological transition in the 1940s which led to a variation in changes in mortality rates across countries as an instrument and show that there is no evidence that improvements in life expectancy led to faster growth in income per capita. Our point is only that the relationship between health and income cannot by fully accounted for by "reverse causation" as estimates that account for this potential channel show equally strong associations.

share of Muslim population is greater than 90 percent in 1900 from Barro and McCleary (2005).

HIV/AIDS Prevalence: we need to control for the Human Immunodeficiency Virus (HIV) and the Acquired Immuno-Deficiency Syndrome (AIDS) given the fact that in spite of income and health improvements, starting from the 1980s the spread of HIV/AIDS has significantly reduced life expectancy in some Sub-Saharan states (Wilson 2001; Neumayer 2003). Moreover, some scholars have attributed the divergence in life expectancy in the 1990s largely to the effect of HIV/AIDS (e.g., Goesling and Firebaugh 2004; Neumayer 2003; Ram 2003). We therefore include a variable that measures the share of the population in the 15-49 age-group infected with HIV over 1979-2007. Before 1979 we are going to assume that HIV/AIDS was equal to zero given the fact that the first recognized cases of HIV/AIDS occurred in the early 1980s[6].

The sources and methodology used to construct these variables are discussed in the appendix. We are going to estimate the following equation over 1952-2007[7]:

$$\ln(LEX_{it}) = \alpha + \beta_{\ln} \ln(YPC_{it}) + \delta X_{it} + \varepsilon_{it} \qquad (2)$$

or

$$LEX_{i,t} = \alpha_q + \sum_{k=1}^{4} \beta_k * YPC_{i,t}^{k} + \delta * X_{i,t} + \eta_{ii.t}$$

Where:

$\ln(LEX_{it})$ is the natural logarithm of life expectancy for country i^{th} at time t^{th}

[6] Even if scientists were able to isolate and discover what is believed to be the first case of AIDS in 1959 it seems reasonable to assume that a significant spread of the disease has not occurred until the late 1970s.

[7] We can only estimate the augmented Preston curve for a shorter time period as data on women's education is only available from 1950 onwards. Also, we have to extrapolate women's education for 2007.

$\ln(YPC_{it})$ is the natural logarithm of GDP per capita (PPP) for country i^{th} at time t^{th}

X_{it} is the vector of control variables (Women's Education; Predominantly Muslim; AIDS Prevalence)

ε_{it} is the error term with standard errors clustered at country-level

α is a constant, β_{ln}, β_{k}, δ are coefficients to be estimated

We show the regression results of the estimation of the augmented-Preston curve in Table 2. The inclusion of the additional variables lowers the estimated income elasticities slightly (as to be expected from adding positively correlated variables).

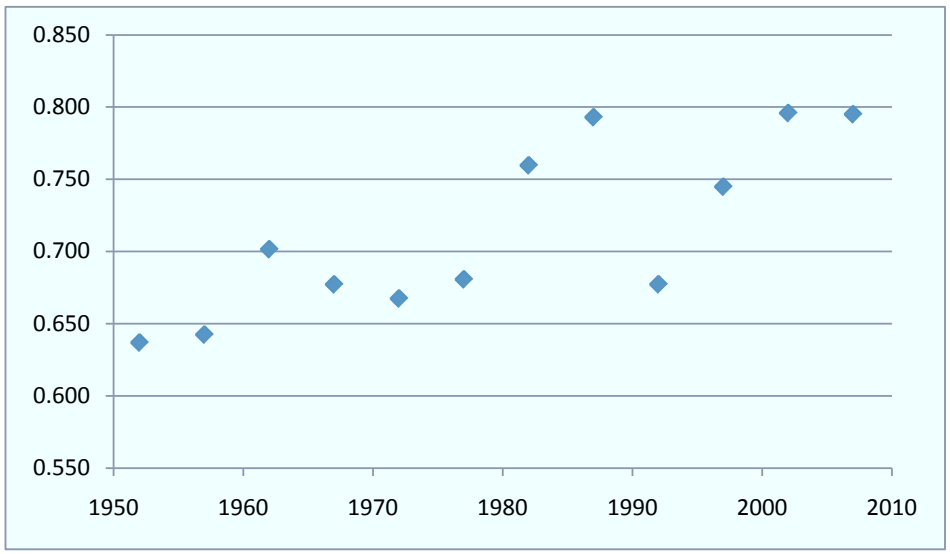
The comparison of the estimated elasticities from the regressions that include the fourth power of YPC and hence allow income elasticities to vary rather than imposing the constant elasticity of the ln-ln specification reveal the pattern that the income elasticity is lower at very low levels of income (in the 2002 results, .051 at income of 1250 versus .087) then higher (.116 at YPC of 5000) then declining again. But at YPC of 10,000 the elasticity is almost equal to the constant elasticity estimate. The quartic elasticities decline at a slower pace than the ln-ln elasticity. With respect to the previous regressions we can observe that life expectancy elasticities in the ln-ln regressions with controls decline faster starting from 1962 whereas after the predicted life expectancy is higher in the regression with controls.

Table 2: The Augmented-Preston Curve shows a high and rising R-Square since 1950								
		Ln-ln specification		Income elasticities at various levels of YPC in the quartic specification				R2 Quartic in
Year	N	β	R2	1250	2500	5000	10000	YPC
1952	151	0.180	0.637	0.158	0.218	0.212	0.066	0.675
1957	152	0.169	0.642	0.142	0.200	0.200	0.062	0.678
1962	153	0.122	0.702	0.111	0.154	0.138	-0.011	0.734
1967	158	0.113	0.677	0.104	0.146	0.136	-0.015	0.717
1972	171	0.102	0.667	0.058	0.097	0.134	0.112	0.684
1977	174	0.100	0.681	0.075	0.112	0.148	0.085	0.734
1982	174	0.097	0.760	0.075	0.118	0.146	0.087	0.798
1987	174	0.098	0.793	0.079	0.121	0.143	0.074	0.822
1992	188	0.091	0.677	0.070	0.110	0.134	0.081	0.740
1997	188	0.086	0.745	0.064	0.101	0.125	0.076	0.777
2002	191	0.082	0.796	0.051	0.083	0.112	0.087	0.806
2007	185	0.087	0.795	0.050	0.084	0.116	0.102	0.806

Note: N: number of countries included in the regression; β ln-ln spec: YPC coefficient in the ln-ln specification; R2: R squared in the ln-ln specification; elasticity of the quartic regression at different income levels ($1,250, $2,500, $5,000, $10,000); R2: R squared in the quartic specification

Figure 1 shows the R2 of the augmented Preston regression. There is no evidence of a "weakening" of the augmented Preston relationship as in the last two years of estimates, 2002 and 2007 the relationship is right at its all time peak .80.

Figure 1: Fit (R^2) of the Augmented Preston Curve Has Risen Over the Last 50 Years



Note: regressions include controls (HIV/AIDS share, predominantly Muslim share, Women's education); Results extracted from Table 2

There are three points to be made about the goodness of fit in the basic and "augmented" Preston regression.

First, apparent weakening of the LEX/YPC relationship in the simple Preston since 1982 in table 1 is entirey due to the impact of HIV/AIDS. If we just reproduce the basic Preston regression excluding the countries with large HIV epidemics (greater than 2 percent of the population affected) then instead of the basic Preston R2 falling from .715 to .590 from 1987 to 2007, the R2 remains much more steady, falling only from .718 to .706 (with the initial fall from 1987 to 1992 perhaps due to the inclusion as new countries the former Soviet Union countries). So there is no particular mystery here, if one looks at the basic scatter-plots of LEX and YPC it is obvious that countries like Botswana are massive outliers. But a region specific negative shock to health lowering the overall cross-national association does not create any fundamental puzzles about the income/health relationship itself.

Second, the incremental explanatory power of the added variables, particularly in the pre-HIV era is quite modest. The main variable that we are able to add is women's education (that is, we could not add income inequality with sufficient coverage). One might ask therefore whether the explanatory power is coming from YPC itself or the correlation of YPC with women's education. On this there has actually been an interesting shift over time, as seen in table 2b. Essentially in the early years there was near zero correlation of women's education and YPC and hence nearly all of the augmented Preston explanatory power came from income alone as the augmented Preston variables without income gave almost nothing. Over time there were two changes. One, the HIV/AIDS variable had more explanatory power (and that of income fell due to that exogenous shock as explained above). Two, the correlation with women's education and YPC rose steadily. This implies, given the simple mechanics of multivariate regression imply that the ability to disentangle the relative contributions falls (as when YPC is omitted OLS will naturally load the omitted variable on a correlated variable). But the data is consistent with YPC alone carrying most of the explanatory load and of roughly equal magnitude (correcting for AIDS).

Table 2b: The Augmented-Preston Curve, fit (WITHOUT and WITH income in the regressions)

Year	N	R2 of Augmented Preston specification (WITHOUT Income)	R2 of Ln-ln specification with YPC (same as table 2)	R2 of basic Preston (table 1)	Correlation between GDP per capita and Women's Education
1	2	3	4	5	6
1952	151	0.058	0.637	0.541	0.042
1962	153	0.365	0.702	0.565	0.218
1972	171	0.365	0.667	0.559	0.263
1982	174	0.463	0.760	0.663	0.418
1992	188	0.488	0.677	0.622	0.420
2002	191	0.660	0.796	0.611	0.448

Note: N: number of countries included in the regression; R2: R squared in the ln-ln specification. We need to interpret with caution results for 2007 as the women's education variable has been extrapolated (an explanation on the method used is provided in the appendix)

Third, to us the real puzzle is not that the R2 is high (as we explain below, one would expect from standard theories of choice that this fit be very good) but that it is so modest compared to the value of say, .95 found by Pritchett and Filmer (1999) for child mortality. Before assuming that if these factors explain 80 percent of the variation perhaps "health policy" even potentially explains the rest there are several considerations. One, the coverage of this data is massive, which means all manner of extrapolation has been made to achieve this coverage as reliable, death reporting based life expectancy estimates exist for relatively few developing countries. Moreover the income variable of GDP per capita is also measured with considerable error as a proxy for what perhaps ideally should be present the purchasing power of the typical consumer. Both of these types of measurement error will reduce the R2 (even if the "true" fit were perfect). Experimentation with different samples shows that excluding for instance just the very small countries (which may be particularly prone to measurement error) can increase the Basic Preston cross-section results by as much as .07 points. Second, there is a simple mathematical relationship that if the relationship between health status is ln-ln at the individual level then the correct aggregate specification is not the log of the average but the average of the logs (but the latter requires household data). The magnitude of the gap between the log of the averages (log of GDP per capita) and the average of the logs is itself a measure of inequality. So we know we are excluding a factor that should be present, inequality (even if inequality has no direct causal effect on mortality). This also lowers the achievable R2. All in all, .8 must be considered a *lower* bound on how much the simple measures of command over resources and women's education explain and this level is not in and of itself evidence that any other factor even potentially explains the rest, it really could just be measurement error and excluded non-health policy related factors (e.g. susceptibility to disease conditions form geography).

I.B) Functional form of the LEX association with per capita income

This section shows that the relationship with LEX and YPC is in an augmented Preston regression is (roughly) double natural log (ln-ln) and is definitely not linear.

Allowing for the YPC to enter with up to the fourth power allows for a quite flexible non-linear relationship. We calculated the elasticities implied by the quartic across four doublings of YPC over roughly the range of the "developing" countries (PPP 1,250 to 10,000 which spans from roughly Haiti, Nepal, Tanzania at the low end to Uruguay, Serbia, Costa Rica at the upper). Compared to constant elasticity formulation imposed by the ln-ln form the elasticities are *lower* at the lowest levels, then rise to be higher than the ln-ln at roughly the median of the sample ($5000) and then decline again until at $10,000 the elasticity is roughly at the ln-ln value (slightly lower in some years, slightly higher in others).

As seen above from the results of the R2 this additional flexibility in the specification of YPC does not gain a tremendous amount in overall goodness of fit.

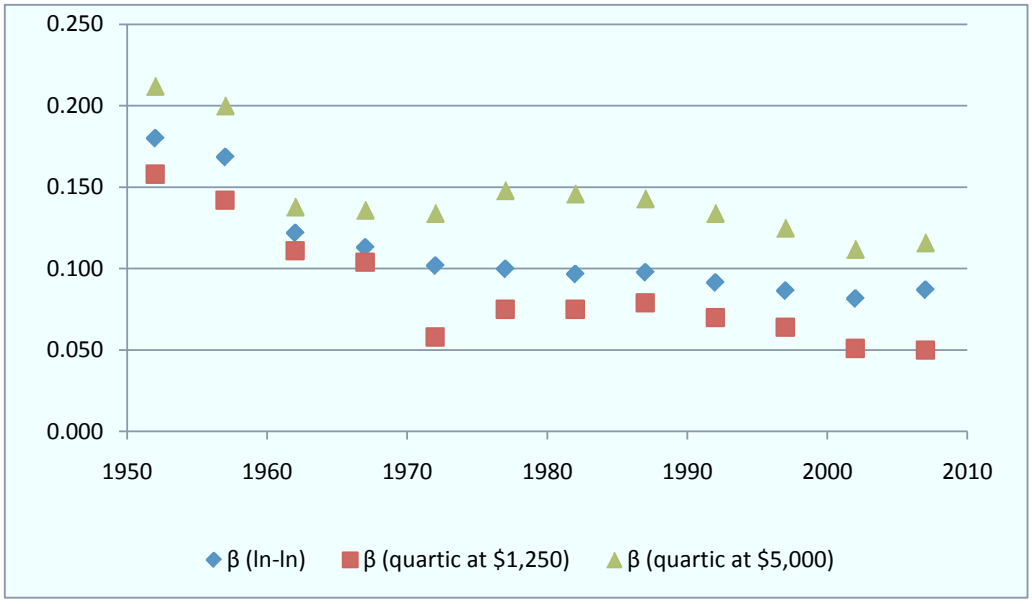
Table 3: augmented Preston curve: elasticities

Year	N	β ln-ln spec	1250	2500	5000	10000
1952	151	0.180	0.158	0.218	0.212	0.066
1957	152	0.169	0.142	0.200	0.200	0.062
1962	153	0.122	0.111	0.154	0.138	-0.011
1967	158	0.113	0.104	0.146	0.136	-0.015
1972	171	0.102	0.058	0.097	0.134	0.112
1977	174	0.100	0.075	0.112	0.148	0.085
1982	174	0.097	0.075	0.118	0.146	0.087
1987	174	0.098	0.079	0.121	0.143	0.074
1992	188	0.091	0.070	0.110	0.134	0.081
1997	188	0.086	0.064	0.101	0.125	0.076
2002	191	0.082	0.051	0.083	0.112	0.087
2007	185	0.087	0.050	0.084	0.116	0.102

Note: N: number of countries included in the regression; β ln-ln spec: YPC coefficient in the ln-ln specification; : elasticity of the quartic regression at different income levels ($1,250, $2,500, $5,000, $10,000); table extracted from table 2.

Figure 3 shows both the evolution over time of the estimated income elasticities for the ln-ln form and calculated from the quartic at $1250 and $5000. Interestingly, these trended downward until 1972 (from .18 to .102) and then have roughly stabilized (falling only from .102 to .087 in the subsequent 35 years). The elasticities have been consistently higher at $5000 and lower at $1250, though the gap has narrowed modestly.

Figure 3: Elasticity of the augmented- Preston curve



The main point of the non-linearity of the functional form between LEX and YPC is that it implies that *absolutely nothing* can be inferred about the relationship or its reliability from simple comparisons of the two variables. In particular, comparing the changes of the two variables in *levels* one can expect a small correlations even if the trajectories of both variables follow *exactly* the path the non-linear functional form over time would predict. This is important, as it is easy to imagine that the scatter-plot of changes on changes is informative about the relationship and that a difference between

the relationship in levels and the relationship in changes might suggest something of a puzzle—but this intuition is only reliable if in fact the underling relationship is linear.

To illustrate this point a simulation exercise will be helpful as a simulation helps make clear exactly what is going on as every single aspect of the evolution of the simulated variables is exactly understood. With a simulation we don't have to worry what about the complex things going on in the world that might explain the behavior of the data as we control the "data" exactly so we also understand it exactly. For the simulation we use the HDR 2005 data for life expectancy and for GDP per capita. We estimate both a quartic functional form and an ln-ln form, both of which produce results very similar to those reported above. As discussed the quartic produces a different pattern of non-linearity as the quartic allows for a much "flatter" slope in countries with high YPC.

Then we simulate new "data" by assuming that countries have economic growth over the next N (where N=10 or 25) years with the growth rate of each country drawn from a normal distribution with mean of 2 ppa and a standard deviation of 2 ppa (which is roughly the actual historical growth rate distribution). Then we predict LEX N years from now by assuming either:

(a) the predicted LEX follows *exactly* the cross-sectional relationship estimated today, that is, the cross-sectional Preston curve is *exactly* stable and countries fall *exactly* on this predicted relationship (so that the simulated R-squared is 1) or

(b) we simulate a random error such that the cross-sectional R2 N years ahead is roughly constant.

In either case it is clear that in the *simulated* data there has been no "break-down" in the Preston curve or "puzzle" about countries trajectories of YPC and LEX—they have moved in ways consistent with a stable unchanging cross-sectional relationship between YPC and LEX.

In this simulated data with a perfectly stable and even with a perfectly predictive Preston relationship over 25 years ***the correlation of changes in the absolute levels of LEX and changes in* levels of *YPC is small.*** Even for a long horizon of 25 years and even assuming the fit is perfect 25 years ahead the R-Squared of regressing the simulated *changes* in the level of LEX on the simulated *changes* in the level of YPC gives an R-Squared of .061. Some might think that this is a "puzzle" as the R-squared in changes is *an order of magnitude* smaller than the R-Squared in levels. But there is no puzzle as this is exactly what one would expect from countries moving along the existing cross-sectional relationship because it is non-linear and countries tend to the regression line.

Even if one allows for percentage changes, there is still the issue that even with a stable, constant and common parameter cross-national relationship that predicts the future LEX based on future YPC *perfectly* the R-squared in changes is only .136 for the quartic model and only .177 if one uses the ln-ln estimated model. Variations in the simulation like making the horizon shorter (shown for 10 year horizons) or allowing for a constant cross-national R-Squared (shown in rows below) naturally make the "changes" R-Squared even lower.

That even with a perfect fit in the correct ln-ln specification the R2 would be so low is perhaps counter-intuitive, but has a simple explanation. Suppose the current R2 is only .65 which means some countries are above the regression line and others below it. If the R-squared is going to be perfect in the future then those above the line have to move

down to the line and those below have to move up. Therefore the *changes* in the ln(LEX) are dominated not by the changes in ln(YPC) but by the return towards the regression line (as in a simple error correction model).

Table 4: Simulations show that very low associations in *changes* (either in levels or in percentage changes) are consistent with stable cross-national relationships with high explanatory power

Regression functional form	Assumption about residuals in t+N (no auto-correlation)	N=25 years				N=10 years			
		Level in T and T+N		Change T to T+N		Level in T and T+N		Change T to T+N	
		(actual data)	Level	Absolute	Percent	(actual data)	Level	Absolute	Percent
			Based on simulation of future data				Based on simulation of future data		
Quartic	Perfect fit	.675	1	.061	.136	.675	1	.009	.033
	Equal R-2	.675	.560	.040	.096	.675	.674	.008	.022
Ln-ln	Perfect fit	.651	1	.177	.166	.651	1	.061	.040
	Equal R-2	.651	.65	.090	.097	.651	.65	.031	.023

Notes: R-squared values from simulation are averages over 1000 iterations. The simulation assumes country future growth rates are drawn from a normal distribution. Future LEX is predicted based on the existing estimated coefficients (plus an error term when the R-squared is not 1). The results in the quartic are based on truncating the predicted values for only those countries under $50,000 YPC so as to not force the model to predict far out of sample at the top range.

The beauty of simple simulations of data is that there are no "puzzles"—so we demonstrate that small correlations in "changes" and high cross-national associations in

levels based on constant coefficients with a non-linear functional form are perfectly consistent and are not a "puzzle" at all. This is of course not *proof* of anything, there may well be another model, in which the Preston Curve has broken down, or is strikingly heterogeneous across countries that does explain a lack of correlation, but this does illustrate that a low changes on changes R2, in and of itself, implies *nothing* about the long-run stability of the relationship. In this simple example, once countries were exactly on the line then it could be that in the next period the changes on changes R2 would be a perfect fit even though the R2 in the previous period was low (as countries in the previous period were just getting back to the line).

I.C) Shift of augmented Preston curve over time

One of the well established facts is that Preston curve has shifted upwards over time so that, at any given income (or given levels of all characteristics in the augmented Preston) life expectancy is higher over time. One conventional interpretation of the shift of the Preston curve is that improvements in medical practices, new vaccines, immunization and overall technological progress which are not related to per-capita income in particular countries have increased people's life expectancy.

Our work is consistent with previous evidence about the Preston shift and we only make two points. First, the pace of the Preston shift does not appear to have accelerated over time, but rather is roughly constant, or, in the quartic specification, accelerated, then decelerated in the 1970s to 1990s and then accelerated again. Figures 4a and 4b show that there has been an upward shift since 1952. However, there is no evidence that, in either level or log terms this has been accelerating, if anything decelerating.

Figure 4a: Shift in the augmented-Preston curve (intercept in ln-ln)

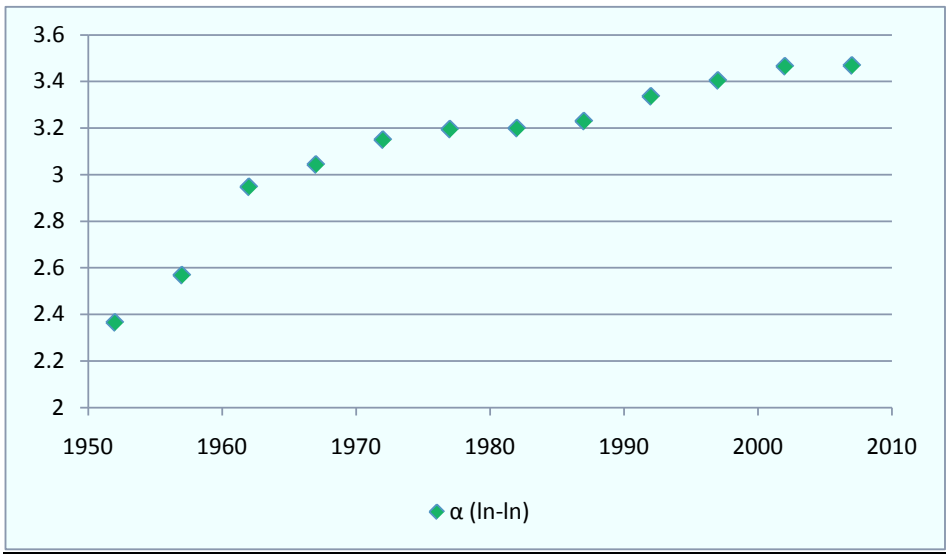
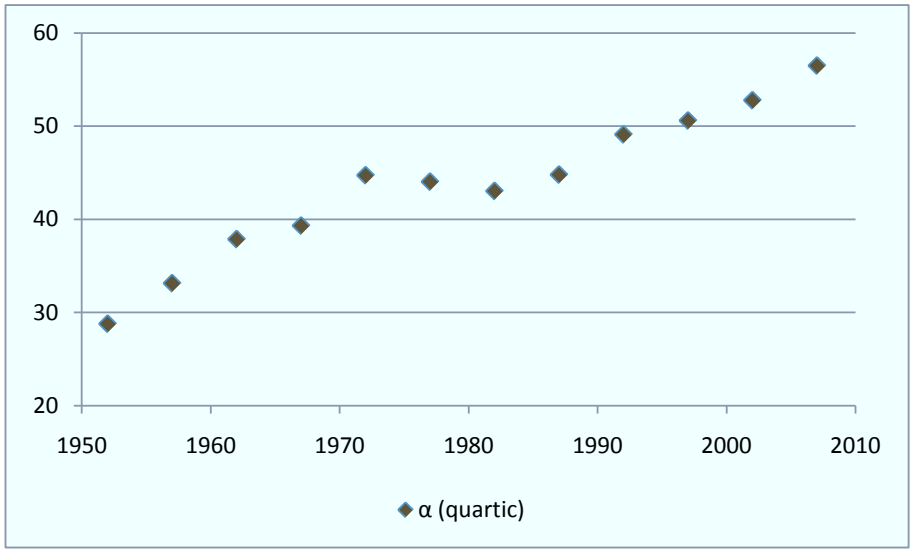


Figure 4b: Shift in the augmented-Preston curve (intercept in quartic)



With the augmented Preston curve, there have been increases life expectancy at all income levels, but more in the quartic than in the ln-ln specification and more at higher than at lower income levels and more in early periods that later periods (Table 4b). That is, when the augmented Preston curve is not constrained to follow exactly a ln-ln formulation by using

the quartic specification one can compare whether there the overall progress was "biased" in a particular direction.

Table 5: Evolution of the predicted values from the Augmented Preston curve at various levels of income is a measure of the improvement in the possible

	Predicted values, in years, from Augmented Preston (at constant values of all other variables) at various levels of YPC			
Year	$1250	$2500	$5000	$10000
1952	43.8	50.6	60.8	67.6
1957	45.2	51.5	61.2	68.1
1962	46.7	51.6	58.5	61.3
1967	48.4	53.2	59.9	63.2
1972	51.1	53.9	58.8	65.5
1977	51.9	55.5	61.6	68.5
1982	51.4	54.9	60.6	66.8
1987	55.5	59.4	65.6	71.8
1992	57.4	61.2	67.4	73.7
1997	59.5	63.2	69.2	75.4
2002	61.5	64.6	69.8	75.9
2007	61.4	64.3	69.0	74.8

Based on the regressions presented in table 2.

Table 6 shows the interesting pattern that progress at the levels of the poorest countries was more rapid in the earlier period than more recently (comparing 25 year periods) when compared to the richer, but still developing level income, countries. So in the period from 1952 to 1977 the predicted LEX at low levels of income grew by 8.1 years while at the middle levels of income, 5,000 and 10,000 it grew hardly at all, while in the more recent period the curve shift was higher in absolute terms at low levels (9.7 versus 8.1) but accelerated massively at the higher levels of income to be roughly equal in the absolute gains across levels of income.

Table 6: Early improvements came primarily at low incomes whereas more recent progress was more uniform

	Absolute change in predicted level of LEX at various levels of income		Percentage change in predicted LEX at various levels of income	
	1952-1977	1977-2002	1952-1977	1977-2002
1250	8.1	9.7	18.5%	18.6%
2500	4.9	9.1	9.7%	16.4%
5000	0.9	8.1	1.4%	13.2%
10000	0.9	7.4	1.3%	10.8%

Table 5 can also be used to examine the question of decomposing the overall observed progress in LEX between that which could be attributed to pure "expansion of the frontier" (the shift of the Preston curve) and that component which could be attributed to income growth (movement along the frontier). The obvious point is that this decomposition will vary across time (as the degree of progress has shifted) and level of income and by the pace of growth—so there is not answer to the question of the relative importance of income versus generalized health gains. For instance, a country which followed exactly the predicted pattern from 1952 to 1977 and doubled their income from $1250 to $2500 (which doubling requires a 2.85 percent rate of growth) would see the life expectancy increase from 43.8 to 55.5, an 11.7 year gain. Had their income been constant it would have improved to 51.8 (reading down the column of table 5), a gain of 7 years, whereas had the Preston curve not shifted the income gain would have been 7.6 years to 50.6 (reading across the column). In this case "most" of the gain, even with growth was due to global "technological progress" in health. In contrast, a country that had improved by exactly that same proportionate income gain (doubling) but from 5,000 to 10,000 would have attributed nearly all the gain (from to 61.7 to 68.5) to income growth because overall progress at those levels of income was slow.

Decomposing the actually historical progress—the question of "how much was due to technological progress?" is not a well-posed question because it is so contingent on growth. If a country has slow growth then nearly all of its gain will be attributed to "exogenous" improvements while another country starting from the same position will have nearly all the gains attributed to income if it grows very rapidly.

I.D) Reconciling cross section and time series in relationship of income and LEX

One of the big questions about a cross-sectional relationship is whether it really represents the trajectory a country should expect to follow if that particular country experiences income growth. There certainly are situations in which there appears to be a cross sectional relationship across countries but individual countries do not move along the cross-sectional relationship—a widely cited example is the oft cited Easterlin paradox that the cross-national relationship between GDP per capita and happiness is not reflected in countries improvement in happiness as individual countries improve their income (although this stylized fact is now contested by Stephenson and Wolpers 2008).

The relationship between LEX and YPC is completely different than the Easterlin paradox. We present four pieces of evidence to argue that countries should expect to trace out almost exactly the Preston curve as their YPC increases and that there is strikingly little cross-national homogeneity in the estimated relationship of LEX and YPC. We argue most of the confusion about the instability or unreliability of the Preston curve comes from using estimates with low statistical power, that is, estimating the Preston curve only over very short periods or with countries with low growth so that there is not sufficient variation in YPC to adequately identify the long-run relationship.

Stability of estimates of cross-section versus (long-period) first differences. If the LEX-YPC relationship is treated as linear in natural logs (ln-ln) then one can estimate the

cross-sectional relationship in first differences and expect to recover the same coefficient. However, it has been known at least since Pritchett and Summers (1996) that moving to shorter and shorter first differences leads to two problems. First, low statistical power as the distance traced out in the YPC space of the first differences for any given country gets shorter and shorter. Second, measurement error increases, with its usual effects of attenuation bias. That is, no one really believes that the relationship is between LEX and GDP per capita exactly and instantaneously period to period. Rather, GDP per capita is a crude proxy for the overall purchasing power of households and governments, so for instance changes in GDP over short periods that are driven by changes in investment but not consumption would not be expected to have the same impact as changes in consumption. Similarly, there are almost certainly long and complex lags in the relationship between income and health so one need not expect the same impact of cyclical and trend changes. Also, there are reasons to not expect exact symmetry, so that a one percent fall and one percent rise might not have exactly equal effects, especially in the short run (over which households might actively smooth consumption of other goods to protect health status). Pritchett and Summers (1996) show that fixed effects regressions using ln first differences over different periods have the expected result that the estimated elasticity is lower the shorter the period (even though arguably the fixed effect regressions of all lengths "solve" the bias of unobserved country specific effects that might plague the cross section).

Here we use the long-run historical data to estimate fixed effects regressions but only using countries with very long time series. That is, we allow for country specific effects by using the first differences, but only use the first differences of countries where we have a sufficiently long period of observations we are confident the results are not predominated by measurement error. Table 7 shows that using the long-period first differences recovers *exactly the same coefficient* as the cross-sectional estimates. The elasticity in the long-period

differences with controls (the augmented Preston) is .11 and the average elasticity from the cross-sectional estimates in the augmented Preston from 1952 to 2007 (reported in column 3 of table 2) is also .11. That is, if we ignore the cross-sectional variation entirely and estimate the augmented Preston curve using *only* the data of how individual countries move, we recover *exactly* the same coefficient (to the third digit, though this must be a coincidence)[8].

Table 7: Estimates of the LEX-YPC coefficient in ln-ln functional form with long-period fixed effects are very close to those in the cross-section

	Without	With Controls[a]	Without	With Controls[a]	Average from 1952 to 2007 of the ln-ln cross section (col. 3 of table 2)
Income coefficient	.115	.118	.103	.110	.110
N	27[b]	27[b]	51[c]	51[c]	151-185

a) controls include: aids share, Muslim dummy, life table (south) dummy [women's education has been excluded from the long-run regression as it is only available from 1950]; a time trend is included, b) at least 50 years of observations, sample includes: Australia, Austria, Belgium, Bulgaria, Canada, Czech Rep., Denmark, Finland, France, Hungary, Iceland, Italy, Japan, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Taiwan, United Kingdom, United States, Yugoslavia, c) at least 15 observations (at five year intervals) sample includes: Argentina, Australia, Austria, Belarus, Belgium, Brazil, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Cuba, Czech Rep., Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, India, Indonesia, Ireland, Italy, Jamaica, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, New Zealand, Norway, Pakistan, Poland, Portugal, Russia, Slovak Republic, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Turkey, Ukraine, United Kingdom, United States, Venezuela, Yugoslavia.

Estimates from long time series of individual countries. The long-period panels still use cross-national variation and so the overall coefficient might be hiding individual country heterogeneity. An alternative would be to use countries with long-times series on both indicators and estimate the relationship country by country using only the time series data. In

[8] Filmer and Pritchett (1999) noted the same result with child mortality. Using data covering a 100 year span recovered exactly the same ln-ln coefficient as the cross-section.

figure 5 we show the coefficient on YPC in LEX for the countries with more than 50 years of data. It is striking both that the average individual country coefficient looks very similar to the overall cross-sectional estimates—the median coefficient is .095 versus the cross-sectional average of .11. Moreover, there is very little variation across the countries. Other than Mexico all of the countries fall in a narrow bound between .05 and .118.

Figure 5: Long-run Country by Country augmented -Preston curve

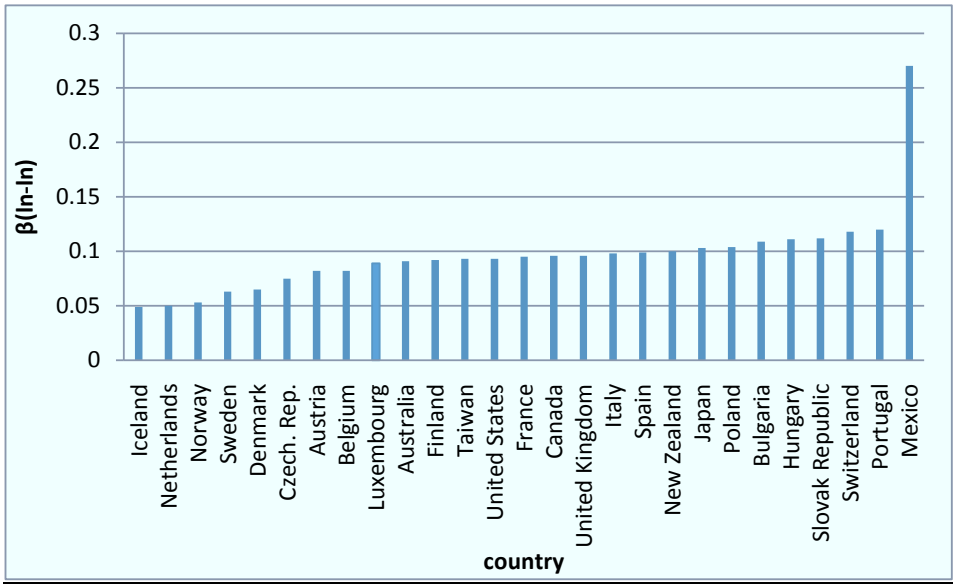


Figure 6 shows the cross-sectional Preston curve estimated from long-period panels with the actual time series evolution of the data from Canada since 1900. The slope of the evolution of Canada mirrors almost exactly the cross-sectional relationship. Of course, the difficult with the use of a single country is that one cannot separately identify the time series shift of the relationship and the income elasticity (as they both have a strong trend).

Figure 6: Comparing the Preston curve with the long period time series evolution of Canada

Left: Preston curve for the sample; Right: Preston curve for the sample and Preston curve for Canada

Note: sample of 51 countries (countries with at least 15 years of data). Graphs have been truncated at the 2007 average level of GDP per capita in the United States ($42,859).

Stability of the long-run R2 and country heterogeneity. We start from a basic point that we observed in an earlier section: an increasing R^2 in the cross-country regressions of life expectancy on income per capita, at least over the last 50 years of data for a common set of countries, and a high level of R-squared even 80 years ago. This fact itself has implications about the country trajectories. If countries had very different long-run trajectories we would expect the cross section relationship to become weaker over time and then eventually the long-run relationship to fall apart. Therefore, it is actually difficult to square any very enormous heterogeneity in the LEX/YPC relationship with an increasing R^2.

Again, a simple simulation exercise helps clarify the point as it makes the data completely under control to examine the implications of controllable features of the data. We

take a distribution of 100 units ("countries") by YPC simulated to reflect the actual mean and standard deviation of YPC. Then we assume that in the first period there is an exact fit of YPC and LEX (this is the opposite of the simulation above where we assumed perfect fit in the second period). Then assume that each individual country's own income elasticity is drawn on a normal distribution around .10 but with different standard deviations. Obviously if the standard deviation is exactly zero then every country moves exactly along the same trajectory and the R-Squared is always 1. Assume that there is some country heterogeneity in the income elasticity, then as countries growth (even at the same rate) the R-Squared will decline as countries with the same YPC have increasingly different LEX (and hence var(LEX|YPC) grows). Table 8 shows the evolution of the simple R-Squared if all countries grew at 2 percent per annum. Any substantial heterogeneity in the income elasticity produces a substantial decline in the R-Squared over time. For instance, if the distribution were such that even 10 percent of countries had a truly zero or negative income elasticity, this requires a standard deviation of .08 around a mean of .10 but in that case over fifty years of two percent growth the R-Squared would decline by .32 points.

Table 8: Simulation showing that heterogeneity in the YPC coefficient causes secular decreases in the cross-national goodness of fit							
	R-Squared in year						Change in R-Squared
Simulated Standard Deviation of Income Elasticity Across Countries	1	10	20	30	40	50	(Actual is a .12 *increase* from 1952 to 2002 in the Basic Preston (excluding HIV/AIDS countries) and a .161 *increase* in the augmented Preston R2)
Assuming perfect fit in period 1 and deterministic evolution of LEX country by country							
0	1.00	1.00	1.00	1.00	1.00	1.00	0.00
0.02	1.00	1.00	1.00	0.99	0.98	0.97	-0.03
0.04	1.00	1.00	0.98	0.95	0.92	0.88	-0.12
0.06	1.00	0.99	0.96	0.91	0.84	0.77	-0.23
0.08	1.00	0.98	0.93	0.86	0.77	0.68	-0.32
0.1	1.00	0.97	0.88	0.76	0.63	0.52	-0.48

Of course a simulation does not imply anything about the underlying data, but it does help to make clear the simple consequences of various assumptions. Any very substantial amount of heterogeneity in the YPC-LEX relationship across countries would create tendencies for the fit between the two in cross-section to fall. (As opposed to the quite substantial rise actually observed).

Shorter period first differences. What is clear from the above discussions is that the key question about using the differences over time to estimate the long-run relationship is how much of the dynamics are a return to the regression line and how much are a possible break down in the relationship. With the actual HDR data (not the long-run data set from GAPMINDER we have been using for consistency) we illustrate how important the dynamics are. Table 5 shows the regressions of the natural log first difference of LEX on the natural

log first difference of YPC in three different specifications. The first just runs the ln changes on natural ln changes—and one sees essentially no correlation. The second column (II) simply adds the residual of the regression from 1970 ln(LEX) on 1970 ln(YPC)—how far a country was from its predicted value. If there are strong "error correction" dynamics then this should be negative and suggest a return to the regression line. Indeed, already change ln(YPC) has a positive and statistically significant coefficient—and the scatter plot Figure 4a reveals that there are several obvious outliers, each of which has an obvious explanation: GNQ has had massive increase in GDP per capita from oil, but which has hardly affect actual available command over material resources of the typical citizen, Liberia's LEX has clearly had a massive recovery from an extremely low value due to conflict, and Botswana (BWA), South Africa (ZAF), Lesotho (LSO) and SWZ (Swaziland) have been affected by HIV/AIDS. If we just drop those six then column III and figure 4b show the results in which the coefficient on the *change* in ln(YPC) is *exactly* what the *level* regressions produce: .0938 from the changes regression 1970 to 2005 and .0928 from the average of the income elasticity coefficients from 1972 to 2007 reported in table 3.

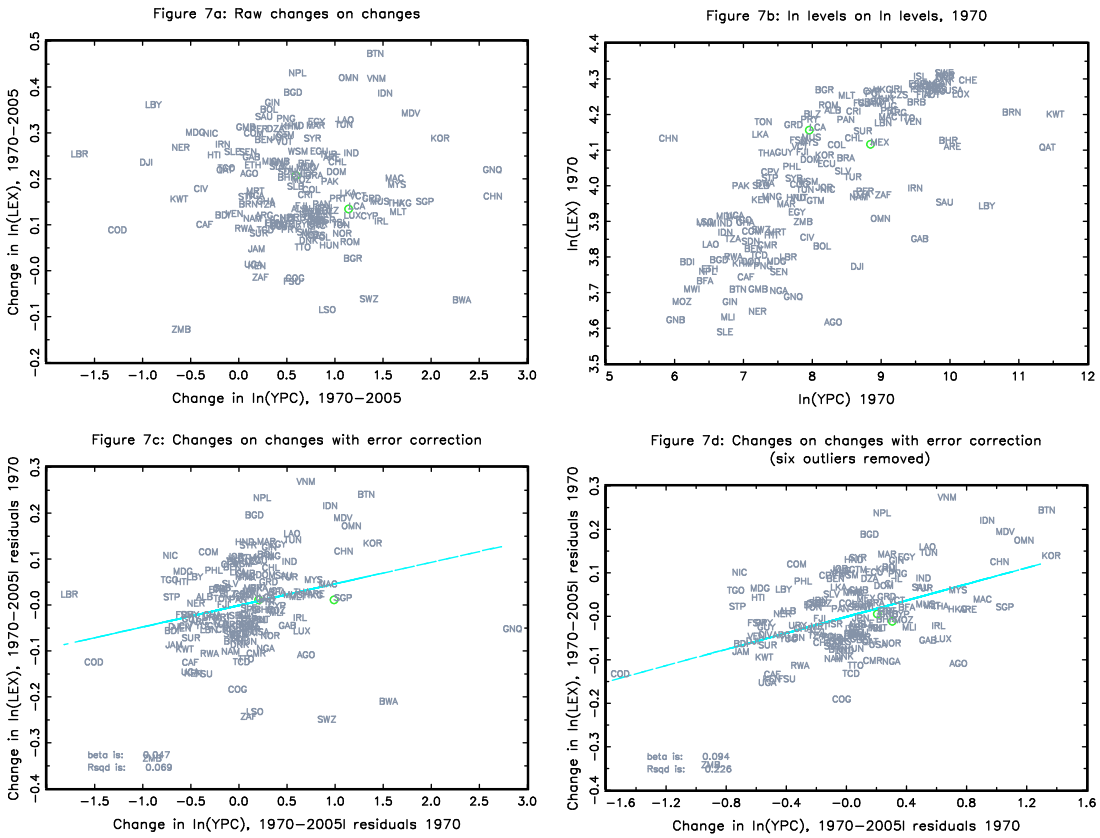
Table 9: Regressions of the change in natural log of LEX on the changes in YPC produce nearly exactly the results of the cross-section, once one allows for "error correction" dynamics in countries returning to the steady state relationship (adjusting for a few outliers)

	Raw First Differences	First Differences including the 1970 residual	First Differences including the 1970 residual and excluding six outliers
Constant	0.179 (15.68)	0.149 (13.79)	0.133 (13.3)
$\ln(YPC)_{2005} - \ln(YPC)_{1970}$	-0.008 (-0.609)	0.047 (3.29)	0.0938 (6.42)
$\ln(LEX)_{1970} - \ln(\hat{LEX})_{1970}$		-0.468 (-6.91)	-0.616 (-9.64)
N	150	150	144
R2	.003	.247	.398
Countries deleted	None	None	GNQ,BWA,ZAF, LBR,LSO,SWZ
Note: Using HDR data set.			

The four panels of figure 7 show the mechanics of the regression results very clearly. Figure 7a shows the raw scatter plot of the two variables. Notice on the far right hand side is China with very high economic growth and middle of the pack improvement in life expectancy. Figure 7b shows the regression results from the ln levels in 1970. Notice that China is very far about its predicted value (good LEX with low income). Figure 7c shows the

partial scatter plot of ln changes of LEX on ln changes of YPC, when the lagged residual is included in the regression—for instance China is closer to the regression line because one would have expected lower progress because it was so far above the predicted value at the beginning. In Figure 4c the outliers are all very clearly far off the regression line, which is of course no reason to excise them, but once identified each has a clear reason to believe that their particularly trajectory should not be expected to generalize to other countries. Figure 7d shows the regression with error correction allowed and the outliers deleted, in which the regression line is exactly the average of the cross-section regression averages over the same period.

Figure 7: With error correction first differences give exactly the levels results



What about "exceptions"? While there is just truly amazing consistency and homogeneity in the *long-run* income elasticity estimated from many repeated cross-sections, from long-period changes, or from individual countries evolution over time, there are always "exceptions"—countries that experience growth without LEX improvements or countries with LEX improvements but no growth. However, imagining that these constitute "contradictions" or "puzzles" to a prediction that countries will, over the long-run, experience improvements with income growth of roughly the magnitude predicted by the augmented Preston curve is usually just a misunderstanding of statistical power and the standard errors of prediction (or, perfectly understandable reasons why YPC would fail as a proxy for the typical household's command over resources).

The Preston curve is actually quite flat. Let us suppose the Preston curve in the relevant developing country range is well approximated by an ln-ln relationship with an elasticity of .10. Then a 10 percent rise in per capita income implies a one percent rise in life expectancy, so for a country at the predicted value of LEX at $1250 of 61 years this is a gain of only .61 years. If countries are growing at slow pace, suppose 10 percent growth took 10 years, then this will be much smaller than the upward shift in the Preston and it could easily appear that "growth" had not impact. That is, if countries grow at very low pace or experience fluctuations in GDP per capita growth the Preston curve is no longer a good predictor, in spite of being a good long-run predictor.

This means that even at high R^2 levels and even if the Preston curve is exactly stable and exactly homogenous across countries so that every countries is "expected" to converge onto exactly the cross-sectional Preston relationships, there is a standard error around the regression predictions. So think of the following thought experiment: a country on the regression line has improvement in income per capita but no improvement in LEX (so in the Preston growth is not moving "northeast" along the Preston curve but is heading due west).

How far would the country have to go before it became even a minor *statistical* anomaly (say was more than a single standard deviation from the regression prediction) and how long would that take. The change income times the elasticity has to be greater than the standard error of the predicted life expectancy:

β_{ln}*(%ΔYPC) > std err ($\hat{LEX}$)

If countries are growing slowly this means that the long-run relationship can be a good long-run predictor but, especially at slow growth rates not really be very predictive at all and it would take a very long time, around ten years, to detect this and in case of a very fast GDP per capita growth. Our calculations are only meant to convey the general idea about this point. We use the predicted results from our sample of 51 countries with long time series availability to calculate the elasticities and standard errors[9]. As this simple calculation shows, at slow rates of economic growth such as 1 percent it would take essentially forever to statistically detect a shift off of the Preston curve because the movement in the YPC is so absolutely small.

More interestingly, even at growth rates of say, 5 percent per annum one could observe growth for 14 years with not LEX improvement before a country were an additional Std. Err. off of its projected trajectory.

[9] These regressions actually give us our most precise estimates as the R-squared is so high (over .9) and hence the standard error of prediction are quite low.

Table 10: Years of income growth with no LEX improvement needed to detect a statistically significant anomaly, at different growth rates				
LEX	**cut-off point** **(std error)**	**Annual YPC Growth**		
		1%	**5%**	**10%**
		Years of growth (t) at growth rate *r* with zero change in LEX before $\beta*(1+r)^t$>Std. Err. Prediction		
72	0.075	72	14.5	7.31

Note: cut-off point indicates the point at a given life expectancy level beyond which any increase in income with no related increase in life expectancy would be detected as anomalous by being a standard error off of the prediction.

The empirical point to understand is that the Preston curve is precisely identified relationship because there is such enormous variation in income per capita—countries vary in income by over an order of magnitude and countries growing rapidly grow by an order of magnitude in a couple of decades and countries with steady growth change by an order of magnitude over longer periods. This huge variation allows the precise identification of a quite stable and homogenous relationship that is accurately predictive *over the range* so that if one predicts LEX for Tanzania at $1000 (the 10th percentile of countries) and Netherlands at $35000 (the 90th percentile) who differ by a factor of 35 in YPC then the predicted difference is likely to be quite close to the actual.

If one thinks of the Preston curve as a path leading Northeast (higher LEX with higher YPC) that also shifts North over time then as countries make *substantial* progress East (higher YPC) relative to the scale of the map one can reliably predict their trajectory. However, if one takes annual changes in growth (and throws away the cross-sectional variation) this is like trying to identify the course of the Mississippi river by using only a few randomly selected stretches of 100 yards each. Similarly, if countries have slow growth then even very long-periods will not trace out much distance and hence one will not be able to distinguish whether the country is, or is not, on the trail.

I.E) A small aside about causality

Nothing we have said has anything to do with whether health might also cause higher levels of income. There is a wide variety of empirical evidence from historical comparisons (e.g. Fogel), econometrics, randomized experiments and just common sense that suggests that healthier people will also be more productive. For the present purposes that is not the question. The question might be whether the cross-national regressions we have reported are reported a causal impact of income or health or are the result of "reverse causality" and merely reflect the impact of health on income. While we cannot go into this question in any depth, we think that the existing evidence is consistent with the view that the *magnitude* of the problem of reverse causality on the existing estimates is quite small, for two reasons.

First, we do not have to contradict any of the existing evidence about a positive relationship between health and income to believe that the *magnitude* of such impacts do not present a major econometric issue of reverse causality in the existing cross-national estimates. We are identifying the income impacts off of differences of up to two *orders of magnitude* so that even if one parsed out the reverse effects, even in the magnitudes the current micro and macro literatures might suggest exist (and which may be of sufficient magnitude to be of policy concern), the resulting estimates would be roughly similar.

Second, all existing efforts to "correct" the cross-national estimates for reverse causality using instrumental variables for identification are supportive of the view that the income estimates are roughly unaffected (both Pritchett and Summers 1996 and Pritchett and Filmer 1999) use the then standard approaches to IV and found results larger, not smaller, than the OLS results. However in the current state of play we understand that people are not convinced by the available instruments (as meeting the conditions for an appropriate instrument in a cross-national setting is difficult if not impossible and "passing" the over-

identification tests is hardly compelling given their potentially low power with so few observations).

Therefore, by not "correcting" our estimates for impacts of health on income we are not denying such exist, but this is a separate issue as to whether their magnitude and direction is sufficient to cause a massive re-interpretation of the results (and while there is evidence of health impacts there is absolutely no evidence on this latter issue and what weak evidence there is suggests it is not major econometric issue).

II) Convergence

The "convergence" puzzle, which is related to the fact that in the long run we observe convergence in life expectancy but not in income per capita, remains one of the biggest puzzles in the economic growth literature. The lack of absolute convergence of income per capita has been extensively documented in the literature; evidence from the existing studies suggests that there is no convergence across countries, whether interpreted as absolute convergence or growth rates conditional on initial levels (Barro and Sala-i-Martin 1995; Quah 1996; Boyle and McCarthy 1999).

However, it is not so puzzling as it may first appear. First, there is absolute convergence (reduction in the standard deviation) in the levels of life expectancy and the logs only until 1987 (table 11). So since the specification is clearly non-linear there is not necessarily any puzzle as a shift up in the X with constant non-linear relationship would lead to a decline in the standard deviation of Y in absolute but not necessarily ln terms.

While there has been some absolute divergence in some of the augmented Preston controls—particularly YPC, there has been convergence so it is not the case the Var(ln(LÊX)),

the variance of the predictions, has gone up until 2002. On the other hand, we can observe that Var(Life Expectancy) has declined until 1987 and then increased. We previously discussed how the spread of HIV/AIDS in low income countries may have caused this pattern as well as the increased mortality in countries of the former Soviet Union. Interestingly, Deaton (2006) shows how standard deviation of infant and child mortality continued to decline until 2004, being HIV/AIDS a disease that for the most part affects the adult population. On the other hand, Deaton (2006) shows that if the logarithm of infant and child mortality is considered, we can observe that the two variables started diverging in 2000.

Becker et al. (2005) decompose the convergence in life expectancy in order to estimate the contribution of each factor and find that by limiting the analysis to the change in life expectancy explained by the change in income there is no evidence of convergence.

Since the income elasticity has been declining a bit, the impact of diverging YPC on predicting divergence has moderated, so the predicted divergence using the ln-ln beta from the 1960s (i.e., which is equal to .122 in 1962) would give more divergence than the .082 we see in 2002.

Table 11: Decomposition of the shifts in the Preston Curve to examine convergence

Year	N	Life Expectancy	Var(Life Expectancy)	Var(ln(LEX))	Var(ln(LÊX))	Var(residual)	Var((ln(LEX)) predicted at β_{1962}	Var((ln(LEX)) predicted at β_{2002}
1952	151	50.648	147.558	0.057	0.069	0.032	0.001025	0.000464
1957	152	53.127	145.932	0.052	0.057	0.022	0.000847	0.000383
1962	153	55.289	141.788	0.048	0.033	0.016	0.000490	0.000222
1967	158	57.349	131.236	0.043	0.026	0.015	0.000386	0.000175
1972	171	59.169	121.811	0.038	0.021	0.015	0.000312	0.000141
1977	174	60.953	115.713	0.037	0.020	0.015	0.000297	0.000134
1982	174	62.840	106.627	0.030	0.019	0.017	0.000282	0.000128
1987	174	64.437	102.210	0.028	0.018	0.018	0.000267	0.000121
1992	188	65.338	106.319	0.032	0.015	0.018	0.000223	0.000101
1997	188	66.281	110.787	0.031	0.014	0.018	0.000208	0.000094
2002	191	67.103	123.784	0.034	0.014	0.019	0.000208	0.000094
2007	185	68.175	117.776	0.031	0.015	0.019	0.000223	0.000101

Note: var (ln(LEX)): variance of ln of life expectancy; β_{1962}=0.122; β_{2002}=0.082. We need to interpret with caution results for 2007 as the women's education variable has been extrapolated (an explanation on the method used is provided in the appendix)

We now examine the variance of the explanatory variables included in the regressions in order to understand what is happening to the evolution of the distribution of the variance. Part of the explanation of the puzzle may be that some of the variables we are using as controls are diverging. Summary statistics are shown in table 12.

We can observe that in addition to income per capita, also other non-income components included in the regressions are not converging. Specifically, the variance of the variable women's years of schooling is increasing over time (in the first two years the variable on women's education is only available for a limited number of countries: 1952: 17 countries; 1957: 7 countries; 1962: 100 countries). The variance of the share of people affected by HIV/AIDS is also increasing over time. The sharp increase since 1992 is due to the fact that from 1992 onwards some African states experience a sharp increase in the share of people affected by HIV/AIDS (e.g., in 1992 only Zimbabwe had the share of population affected by HIV/AIDS greater than 15 percent; in 1997: Botswana, Lesotho, Swaziland, Zambia and

Zimbabwe; in 2002: Botswana, Lesotho, Namibia, South Africa, Swaziland, Zambia and Zimbabwe). The reason why HIV/AIDS may have a strong impact on convergence is due to the fact that the disease is most widespread in low-income countries with relatively low life expectancies and survival rates. The evolution of the variance of the predominantly Muslim variable is related to the change in the sample composition.

Table 12: Convergence and divergence of Explanatory Variables

Year	N	Life Expectancy	Var(ln(YPC))	Var(Women Edu)	Var(HIV/AIDS)	Var(Muslim)
1952	151	49.386	1.056	0.699	…	0.084
1957	152	51.855	1.130	0.446	…	0.084
1962	153	53.988	1.164	4.710	…	0.083
1967	158	56.216	1.258	4.520	…	0.081
1972	171	58.463	1.444	4.604	…	0.090
1977	174	60.425	1.513	4.925	…	0.088
1982	174	62.365	1.499	5.231	0.481	0.088
1987	174	63.984	1.547	5.220	0.491	0.088
1992	188	65.033	1.530	5.960	5.507	0.103
1997	188	65.888	1.610	5.440	15.812	0.103
2002	191	66.852	1.621	5.582	18.001	0.101
2007	185	67.788	1.667	3.906*	15.894	0.105

Notes *as described in an earlier section, where observations related to control variables (other than income) are missing the average value of the variable over the period has been imputed. We need to interpret with caution results for the variance of women's education in 2007 as the variable has been extrapolated (an explanation on the method used is provided in the appendix).

III) What Role for Public Policy in Pre-mature Mortality?

The examination of cross-national differences in a measure of pre-mature mortality (either infant, child, or life expectancy) via an augmented Preston curve that includes aggregate per capita income, women's education, allows for technological shocks (as the obvious and ubiquitous health correlates) as well as particular health shock (e.g. AIDS) or

social characteristics then raises two important questions. First, what is the role of a mortality based health status indicator in comparing well-being across countries? Second, what is the role of cross-national comparisons in policy analysis or advocacy?

We are not going to treat any of these questions at length, but we did want to make some points that are not quite intuitive and hence are sometimes confusing, if not confused, in policy discussions.

III.A) Is pre-mature death price elastic?

The perhaps counter-intuitive point is that *because* economists believe that people *do in fact care a great deal* about their health, including pre-mature mortality as a separate indicator of well-being is unlikely to reveal the important differences in human well-being caused by differences in public policy with regard to health. Conversely, the view that one would expect large variations in pre-mature mortality across countries due to health policy is an implicit assertion that either (a) death is price elastic or (b) there are no other ways of avoiding pre-mature mortality other than through the public sector or public action.

The natural translation of "care a lot about" into economic terms is that capabilities or functionings about which people "care a lot" would have *low* price elasticities of demand. This means that price variations would be expected to cause large variations in the quantity consumed of goods about which people *do not* care a lot, perhaps for instance, goods for which there are many close substitutes or which, at the margin, are not so valuable. While price variations in achieving functionings about which people do care a lot would be expect to change the total expenditures, perhaps by a large amount, but not actually change the observed *quantity* by very much.

So, just take a very simplistic notion of a demand curve for pre-mature mortality (not for "medical care" which is a very different thing). This would have an income expansion path—people with higher total resources—would achieve longer life expectancies. What about price variation? What if it were cheaper to reduce mortality (relative to consumption of all else as an aggregate commodity)? The amount by which mortality was reduced would depend on the price elasticity, which is dependent on that extent to which people were willing to substitute consumption of other goods for the extension of life. In the extreme case of lexicographic preferences—which is an assertion that people "care a lot" about health--there would be zero price elasticity. Table 13 just does some simple calculations (just multiplication) of how much one would expect a 50 percent reduction in the price of reducing pre-mature mortality to lower observed mortality.

Table 13: Does anyone believe that pre-mature death is price elastic?

Assumed price elasticity of demand for extending life (avoiding pre-mature death)	Effect of a simulated reduction in "price" of 50 percent		Simulated R-squared of "price" variation in simulated LEX data (mean=65, std. dev.=10), uniform distribution of price variation with a 90^{th}-10^{th} spread (in ln units) of:	
	Percent reduction in LEX	Years, at LEX=65	.5	1
0	0.0%	0.0	0	0
.05	2.5%	1.6	.009	.020
.10	5.0%	3.3	.034	.11
,15	7.5%	4.9	.11	.361
.20	10.0%	6.5	.249	.562

An only slightly more complex calculation is to ask, what if there were notion of an aggregate price elasticity to an overall, country-wide, difference in the *relative* price of avoiding pre-mature mortality (which is not necessarily the price of health care services) then, one could ask, how much explanatory power should we expect price variations to have? This is just a simple combination of how much price variability across countries and the price elasticity relative to the underlying variability (driven perhaps by other factors such as income and female education). The final columns of Table 12 show the R-squared of "price" in a simulated cross-national regression. Again, if the price elasticity is very low then even substantial cross-country differentials (say a distribution with a 90^{th}-10^{th} percentile spread of a one log-unit difference) across countries in relative prices of avoiding pre-mature mortality (again, not "medical care") are going to produce very low cross-national explanatory power.

This is consistent with empirical findings. Filmer and Pritchett (1999) show that for child mortality the standard augmented Preston covariates (income, women's education, income inequality, a set of regional binary variables and a binary indicator of predominantly Muslim countries) explained essentially *all* of the explainable cross-national variation—the R-squared was around .95 while the pure measurement error in the child mortality measures was around .025--without any inclusion of *any* indicator of health policy at all. This is perfectly consistent with roughly lexicographic preference orderings over child mortality that would produce a very low elasticity of demand for avoiding child death.

This simple illustration is just to avoid the confusion that somehow arguing for a very high explanatory power of income on health somehow implies that "health is not important" to people or measures of well-being whereas it is in fact *exactly* the opposite. To expect that "price" (via health policy for instance) or other factors are going to make large differences in pre-mature death across countries given people's incomes (or more broadly command over resources) and capabilities of using those resources, which some health advocate seem to do,

one has to argue either that pre-mature death is price elastic—which is the *opposite* of arguing that health has enormous intrinsic important to people in assessing their capabilities and functionings—or that people ability to translate resources into health varies enormously across countries. We argue that one expects high (and causal) associations of income and education with health *because* it is so important to people and hence we do not expect the price elasticity of pre-mature death is high.

This is also not an argument against investments in health policy that would assist people in improving their health. Rather, we believe that because death is in fact very important to people we expect that most of the consequence of improving the price of avoiding death will be observed in people being able to consume *other* things and hence having higher consumption in other domains. I will spend what I have to spend to stay alive, but the less I spend on that the more I can spend on food, clothing, housing, education. Given the risk structure of life-threatening health conditions the consequences of better health policy will also be fewer people vulnerable to having to mortgage their futures to stay alive, and hence less overall vulnerability. So the point we are making is not that health policy is not potentially hugely important, but that looking for the effects of better health policy in a "deaths" dimension may be not the right approach.

III.B) Impact of public sector expenditures on health

There is no question that people place great value on their health status, as perhaps the single most important capability. There is therefore also no question that effective public policies to help people improve their health status are therefore important. However, it is very difficult to elicit from cross-national data that aggregate public expenditure on health has a large impact on observed aggregate health status. This is not to say that there are not, in theory, public sector interventions or policies that would improve health status, it is just that

these impacts of public spending are impossible to infer from cross-national data, for three fundamental reasons.

First, as studies of the cost effectiveness of various potential health policy interventions show, the impact on health status of interventions differs by several *orders of magnitude.* So, while some expenditures that provide cheap and effective treatment of common diseases, like ORS, might have very high health status gain per direct cost, others, like very advanced treatments with small net health gains are enormously larger. Filmer and Pritchett (1999) discuss that one possible explanation of the gap between the empirically demonstrated enormously cost-effective "micro" level interventions and the weak(ish) apparent impacts at the "macro" level may just be that countries differ substantially in the composition of their expenditures. In this case the "realized" health gain per dollar might be very low in practice (if a variety of organizational and political economy pressures lead to high spending on higher cost items) even if the "potential" cost effectiveness is very high.

Second, which is more subtle and complex, is that all existing studies of cost-effectiveness do not adequately differentiate between the cost-effectiveness of a *health* intervention (e.g. taking an aspirin, surgery, spraying for mosquitoes, iodizing the supply of salt) and the cost-effectiveness of a *public sector* intervention to promote health. The key difference is that, since people do care about their health status they will take many actions on their own to promote their health. Some government actions (those typically called *public health)* that provide *public goods* (in the economists sense of non-rival and non-excludable) the "with and without" health intervention and the "with and without public sector intervention" might look similar. However, when the government provides *private goods* (rival and excludable) that promote health (e.g. nutrition, curative care) then there is at least some displacement effect. The proper counter-factual for the impact of spending on a health

intervention is *not* the magnitude of the health intervention uptake but is the magnitude of the *net increase* in the *total* consumption/participation/ uptake of the health intervention.

As Filmer, Hammer, and Pritchett (2000) show this displacement effect will vary widely, depending on whether the intervention is targeted, the supply conditions in the private sector, the consumer demand, etc. Displacement can be anywhere between zero and full displacement. But *nothing* can be inferred from the cost effectiveness of a health intervention about the cost effectiveness of a public sector intervention without evidence about the counter-factual in the absence of the public sector intervention.

Third, governments of the world vary widely in their efficacy on average, and vary in their capability to undertake specific activities. Some governments are quite capable of standard logistical tasks (like immunizations) or concentrated professional tasks (like surveillance) but have difficulty with undertaking activities that require "discretionary, transaction intensive" (Pritchett and Woolcock 2004) decision making to be successful. Ambulatory curative care or individualized promotion/behavioral change (e.g. breastfeeding) are examples of complex, non-logistical tasks.

So the expected positive impact on health status of public sector spending in country *i* can be (crudely) approximated by the expenditure share weighted average across all public sector interventions in health *j* of the product of (a) the health status impact of an increase in the net uptake of health promoting activity j, (b) adjusted for the displacement effect to estimate the net increase in the health promoting activity from the *public sector* intervention creating an expansion in the effective supply and (c) the efficacy of country *i* in translating public sector resources into an expansion of the effective supply of activity *j*.

$$PSCE^i = \sum_{j=1}^{J} \alpha^{i,j} * \delta^{i,j} * \theta^{i,j} * \lambda^{i,j}$$

Where:

$\alpha^{i,j}$ is the share of the activity j in the total budget of country i,

$\delta^{i,j}$ is the efficacy of country i in translating spending on activity j into effective supply,

$\theta^{i,j}$ is the proportion of the expansion of supply via the public sector intervention that translates into an increase in the total effective increase in the health promoting activity,

$\lambda^{i,j}$ is the health status impact of an increase in activity j.

As Filmer, Hammer and Pritchett (2000,2002) articulate this creates a number of possible "weak links in the chain" of translating public sector spending into changes in health status.

Weak link (λ). The budget could be disproportionately allocated to activities that, even if they are done effectively and even if they expand utilization, do not have much impact. In many countries a good deal of spending is concentrated on hospital based care. While some hospital based care is "cost effective" in the narrow sense, much is not. (Moreover, the general *purpose* of hospital based care is typically not generalized improvement in health status but rather a mechanism for reducing vulnerability to the economic shocks from high cost per episode diseases).

Weak link (θ). The budget could be allocated into activities that, although they are effective the public sector spending largely displaces private sector spending. In many countries spending on facilities that provide curative care often dominate the budget and detract from public health and health promotion activities. But the public health activities are often those with low private displacement while curative care may largely substitute for private care and hence the net expansion in care may be small even for large public spending.

Weak link (δ). A third possibility is that government spends resources without creating any effective services. Deaton, Banerjee, Duflo 2003 documented in an extended study that tracked attendance in Rajasthan over a year that on any given day roughly half of the medical personnel being paid by the state were not at their post, confirming earlier estimates from a smaller number of visits (Chaudury et al 2005). Obviously even if these personnel would have engaged in cost-effective activities and even if these activities would have raised overall uptake this is all moot if they don't show up.

The point is not that governments *cannot* undertake health activities that lead to enormously improvements in health of quite cost effectively. Governments can, and do. But not all governments do. The efficacy of public spending on health status at any given time is going to be country specific depending on (at least) the four parameters above. Assessing the efficacy is going to be complex as if involves at least three major disciplines—the health efficacy is the domain of public health experts, the displacement effects the domain of economists, the public sector efficacy (by activity) is the domain of public administration/management and governance.

Moreover, if one were to know the efficacy of existing spending in a given country this would not be informative about most policy options, only the business as usual radial expansion in the budget. But even in countries in which the efficacy of *average* spending is low there might be possibly *very* high marginal spending possible. This does require realism however, as assuming that a government that has proved ineffective in a large range of existing activities can be effective in a new activity (without a specific causal account of why this would be so) can do what assumptions often do.

The final point is that it is not clear what any cross-national association between health status and public spending on health could even possibly reveal. The coefficient in

any regression is going to be, at best, some complex weighted average of the underlying parameters in the chain across countries. Finding that this coefficient is low (as do Filmer and Pritchett 1999) simply reveals that the some weighted average of those parameters over the relevant country set is low, which, given what we know about many developing country governments overall efficacy and about budget allocations (often to curative care services often utilized by the richer portion of the population) it should not be surprising. We would expect therefore that the coefficient is not identifying some underlying deep structural parameter and hence should vary, as some papers have shown, across measures of the efficacy of the government and shares of spending in various categories.

III.C) Policy impact and diffusion of health innovations

More broadly than the issue of the impact of public expenditures on health status is the question of how much cross-national variation in health status one should expect to be associated with cross national differences in health policy. Alternatively, is there any reason to be "surprised" by a very high and rising proportion of cross-national variation explained by an augmented Preston curve? This actually depends on the underlying "model" of policy adoption.

Suppose there is a innovation that unambiguously creates superior health status for countries with given resources and capacity. This innovation could either be medical—such as the invention of effective immunizations—or technological—such as improvements in cold chain—or new information about health practices and how to the effectively promote them—such as how to promote avoiding risky HIV/AIDS behaviors. Whatever the previous level of variation across countries, the early adopters would see improvements in health status relative to equivalently situated countries. This would therefore decrease the augmented Preston explanatory power.

However, the next stage would be that other countries would, perhaps on the basis of its demonstrated effectiveness in another country, adopt (perhaps adapt) the innovation themselves. If this adoption process followed common "S curve" dynamics, one would expect a few early adopters, followed by an increasingly rapid pace, followed by a few lagging countries that, for whatever reason (which could be temporary) are slow adopters. Then the explanatory power would be reduced temporarily and then revert (roughly) to its previously level as other countries adopted.

The point is that nothing about the potential effectiveness of health policy can be inferred from the cross-national variation in health status. The situation in which one would "expect" a high proportion of health status variation to be associated with health policy differences is a "model" in which there are health policy options capable of producing large difference in health status for countries with populations of equivalent income (and its distribution) and individual capability (e.g. women's education) *and* for some reason countries do not adopt those policies. Crudely put, the incremental R-squared is the *product* of "efficacy" of policy X (impact effect on health status, the "beta") and the variation in adoption of X across countries. With either zero or full adoption the incremental R-squared of policy could be zero even for very effective policies.

This means that, if one "expects" there to be large policy variation then one "expects" countries to not have adopted policies that would have led to large gains in well-being for their populations in a dimension of capability that people demonstrably do care a great deal about. Of course, this is not implausible, as governments around the world span the range from benign and effective to malign and chaotic.

But it does imply one should expect to see these more as a narrative of the *dynamics* of the change in health policy, as some countries will be early (effective) adopters and others late than necessarily long-term levels.

This is especially so if potential health interventions have different "sizes" in some combination of efficacy and cost so that some reforms are "no brainers" (big population effects at low cost) while others are effective, but even when fully implemented will have modest aggregate impact or others will have impact but will be expensive and hence require much tougher choices on sacrifices to implement. A common sense model of diffusion would suggest that the former, the high impact at low cost, will be adopted sooner and more widely than the latter. Again, even if there are dramatically effective health policies, the cross-sectional differences in the long-run could be very small as countries differ only in long-run adoption on the more marginal innovations.

This also potentially can explain why one could observe more rapid progress in health status than in economic performance. If the sets of policies and actions to achieve economic growth are less obvious, more particular, and face more difficult political economy for adoption than a progressing technological frontier in health then countries will stay near the world technological frontier (health status near their augmented Preston value) even though countries have very different trajectories in growth and some remain stagnant while others grow rapidly.

Conclusion

The augmented Preston curve is highly predictive of pre-mature death, as one would expect under any reasonable model of human behavior. The convergence puzzle is not so

puzzling given the non-linearity of the relationship between income and health. It is almost impossible to draw conclusions about health policy out of cross-national data.

References

Acemoglu, D., and S. Johnson (2007). 'Disease and Development: The Effect of Life Expectancy on Economic Growth', *Journal of Political Economy,* 115(6): 925-985.

Barro R. and Lee J. W. (2001), "International Data on Educational Attainment. Updates and Implications", *Oxford Economic Papers,* Vol. 53, No. 3, pp. 541-63.

Barro R. J. and R. M. McCleary (2005), "Which Countries Have State Religions?", *Quarterly Journal of Economics,* Vol. 120, No. 4, pp. 1331-1370.

Barro, Robert J. and Xavier Sala-i-Martin (1995). *Economic Growth*. New York, McGraw-Hill, Inc.

Barro and J. Ursua (2008), "Macroeconomic Crises since 1870", Brookings Papers on Economic Activity, Vol. 1, 2008: 255-350

Becker, Gary S., Tomas J. Philipson and Rodrigo R. Soares (2005), "The Quantity And Quality Of Life And The Evolution Of World Inequality", *American Economic Review*, 95(1): 277-291.

Bhaskaran K., O. Hamouda, M. Sannes, F. Boufassa, A.M. Johnson, P.C. Lambert, K. Porter (2008), "Changes in the Risk of Death After HIV Seroconversion Compared With Mortality in the General Population", *Journal of the American Medical Association.* 300(1):51-59.

Biraben, Jean-Noel. (1979). "Essai sur l'evolution du nombre des hommes," Population 34(1): 13-25.

Bloom D.E. and D. Canning (2007), "Commentary: The Preston Curve 30 years on: still sparking fires", *International Journal of Epidemiology,* 1-2.

Boyle, G. E. and T. G. McCarthy (1999). Simple measures of convergence in per capita GDP: A note on some further international evidence. *Applied Economic Letters*,6: 343-47.

Caldwell J.C. (1986) "Routes to Low Mortality in Poor Countries", *Population and Development Rev.* 12 (June): 171–220.

Deaton A. (2007), "Global Patterns of Income and Health: Facts, Interpretations, and Policies", *UNU-World Institute for Development Economics Research, Annual Lecture No.10.*

Case A. and A. Deaton (2008) "Health and well-being in Udaipur and South Africa," in *Developments in the Economics of Aging*, David Wise (ed.).

Filmer, D., and L. Pritchett (1999). 'The Impact of Public Spending on Health: Does Money Matter', *Social Science and Medicine* 49(10): 1309–23.

Preston S. H (1975), "The Changing Relation between Mortality and Level of Economic Development". *Population Studies* 29 (2): 231-248.

Livi-Bacci, Massimo. 1992. A Concise History of World Population. Trans. by Carl Ipsen. Cambridge, MA: Blackwell.

Maddison A. (2003) "The World Economy: Historical Statistics". Paris: Development Centre of the Organisation for Economic Co-operation and Development.

Pritchett L. and L.H. Summers (1996), "Wealthier is Healthier", *Journal of Human Resources*; 31:841–68.

Quah, D. T. (1996). Empirics for economic growth and convergence. *European Economic Review*, 40: 1353-75.

Riley J. C. (2005), "Estimates of Regional and Global Life Expectancy, 1800–2001", *Population and Development Review* 31(3): 537–543 (September 2005).

Woessmann L. (2003), "Specifying Human Capital: A Review and Some Extensions", *Journal of Economic Surveys,* Vol. 17, No. 3, pp. 293-70.

Appendix - DATA SOURCES

Life Expectancy: life expectancy at birth is defined as the number of years a newborn child would live if current mortality patterns were to stay the same

Life expectancy is available yearly but for most countries it is available at a 5-years interval. It is a combination of data from two main sources: the Human Mortality Database (HMD) and the World Population Prospect (WPP): The 2006 Revision of the United Nations Department of Economic and Social Affairs. The main source used is HMD as it provides data for single years. Where HMD data is missing for 2007, WPP data is included. The starting year from which data from the HMD is available varies across countries, ranging from 1871 for Sweden to 1970 for Taiwan, whereas the WPP data is available from 1952 onwards. Life tables for the HMD have been constructed by first collecting data on annual counts of live births and deaths over the longest period available; data on population size below 80. Then estimates of the annual exposure-to-risk of death and death rates are pooled for different time periods and cohorts. Finally, life tables are constructed by computing probabilities of death and exposure to risk of death. A detailed description of the sources and techniques used are available at: www.mortality.org. Data from the WPP rely on the most recently available data sources, such as censuses, demographic surveys and population registers. The underlying mortality trends are based on the HMD for those countries for which the data is available, whereas for the majority of the other countries mortality where both national sources and HMD data is not available life expectancy is estimated on the basis of infant and child mortality data extracted from various sources to which different life table models are applied.

For the remaining countries national statistical sources are used as well as the Human Life-Table Database. Riley's historical dataset represents the main source for the period 1902-1947. For the pre-1952 period Riley's (2005) uses a mix of pretransition and transition estimates. Life expectancy estimates have been drawn from more than 700 sources (listed at: www.lifetable.de/RileyBib.htm). Historical population estimates have been extracted from Biraben (1979), Livi-Bacci (1992) and Maddison (2003). Further information on the construction of the dataset is available in the GAPMINDER publications (GAPMINDER (2008), "Documentation for Life Expectancy at birth (years) for countries and territories")

Source: GAPMINDER Application

Income per Capita: The data is based on GDP per capita and is adjusted for Purchasing Power Parities (PPPs), as calculated in the 2005 round of the International Comparison Program (ICP) to adjust for price level differences across countries. The unit of measurement of GDP per capita is international dollars.

Data have been extracted from several sources: UNSTAT, the Penn World Tables 6.2, the World Development Indicators of the World Bank, the IMF World Economic Outlook and historical data from Maddison's (2008) dataset. Maddison's Historical Statistics for the World Economy represents the main source for the long run growth data, and are complemented by data from Barro and Ursua (2008) who have updated Maddison's data with a focus on improving GDP per capita historical data at time of major crises.

Further information on the construction of the dataset is available in the GAPMINDER publications (GAPMINDER (2008), "Documentation for GDP Per Capita by Purchasing Power Parities for countries and territories")

Source: GAPMINDER Application

Women's Education: number of years of schooling of the female population aged 15 and over.

Data is available for a large number of countries over 1950-2000. It relies on census and survey information as compiled by UNESCO and other sources. Missing information on school attainment is filled by using information on school-enrollment ratios and the structure of population by age groups. The perpetual inventory method is used to construct current flows of population that are added to the benchmark stocks. The variable is available at 5-year interval. Data for 2007 is not available. We therefore extrapolated women's education from its recent past. That is, we used the growth rate in women's years of schooling since 1992 to provide estimates for the year 2007. This implies that our results for the year 2007 in the regressions where women's education is used should be interpreted with caution.

Source: Barro and Lee (2001)

Predominantly Muslim: indicator variable equal to 1 if Muslim population is above or equal 90 percent

Barro and McCleary's dataset includes measures of religiosity and is constructed by combining the following international datasets: World Values Survey (waves 1981-84, 1990-93 and 1995-97), reports on religion by the International Social Survey Programme (1990-93 and 1998-2000) and the Millenium Gallup Survey (1999). Information is complemented by the World Christian Encyclopedia (1982).

Source: Barro and McCleary (2005)

HIV/AIDS Prevalence: percentage of HIV infected people among persons 15-49 year-olds

Data for the years 1990-2007 are extracted from the 2008 Global Report of the UNAIDS/WHO. This data cover most low and middle-income countries and provide comparable estimates of HIV prevalence. Data for high-income countries are complemented by national sources as UNAIDS/WHO mainly uses epidemic models that are less suited to measuring the spread of HIV/AIDS in high income countries. Consequently separate estimates for high income countries based on national longitudinal estimates and other sources have been used.

Data for the years 1979-1989 are based on extrapolations from the UNAIDS/WHO data.

National sources have been used for very small countries and for both countries with limited HIV assessment and countries with very low levels of HIV that are not included in the international statistics. Estimates are limited to countries with more than 100,000 inhabitants. Country-specific information on how the dataset has been constructed is available from GAPMINDER (GAPMINDER (2009), "Documentation for HIV indicators in Countries and Territories").

Source: GAPMINDER Application