Macroeconomic approaches to identifying the effects of health on output, growth and poverty

by

Robert Eastwood

University of Sussex[1]

June 2009

Framework paper for AERC project: Health, economic growth, and poverty reduction in Africa

[1] Email: r.k.eastwood@sussex.ac.uk I am grateful for the comments of participants in a workshop held in Accra in April 2009 for their comments on an earlier version of the paper.

1. Introduction

The research project as a whole is focused on causal relationships running from health to productivity and poverty reduction with the aim of informing health policy. This framework paper deals with research possibilities in the field of macroeconomics, and it is necessary to begin by asking why such research may be useful.

The most important limitation of microeconomic studies is that they cannot, by definition, investigate general equilibrium effects of changes in health, or health policy. For example, the likely economic impact of AIDS on Sub Saharan Africa has been much studied. At the microeconomic level, it is possible to estimate effects on household labour supply and consequently on household incomes and their distribution, *for given wage levels* (e.g. Cogneau and Grimm 2008). However, there will also be general equilibrium effects on wages, which will depend in the medium term partly on the link from AIDS to capital formation. That link is complex and unavoidably macroeconomic: it depends for example on what is assumed about the AIDS-related rise in public health expenditure and whether this is domestically or internationally financed, as well as on what is assumed about international capital mobility. In the longer term, there will also be general equilibrium effects on wages resulting from the aggregate effects on labour supply (including human capital) of fertility and mortality changes.

Therefore macroeconomic research is needed and a number of approaches are possible. Cross-country models have been used to try to uncover relationships among health variables, such as life expectancy (LE), and economic variables such as output per head, poverty or the rates of change of these variables. As will be discussed below, issues of data availability and endogeneity mean that the value of such models for quantifying causal relationships from health to productivity and other economic outcomes is questionable at best.

An alternative to cross-country modelling has been the use of a mixture of micro-estimation and macro simulation within a single country framework. Here there is considerable variation in the methods employed. One strand of literature employs computable general equilibrium (CGE) models. CGE models tend to be multisectoral and to rely on little if any econometric estimation. Functional forms are chosen for the underlying demand and supply relationships, together with a 'closure rule' specifying how the economy's savings-investment balance is to be maintained. Parameter values for the underlying functions are partly imposed and partly determined in such a way that the model can reproduce the initial state of the economy, as represented in a social

accounting matrix. Thus 'calibrated', the CGE model can then be used to simulate the effects of changes in exogenous variables. These models also have certain limitations, of both scope and credibility. As far as scope is concerned, such multisectoral models can be hard to dynamize, so they tend to be silent on some of the critical longer-term consequences of health changes referred to above. As regards credibility, sensitivity analysis is difficult when not only is there a large number of parameters which might be varied, but the underlying functional forms themselves may also be more or less arbitrary. That the results of CGE simulations may be extremely sensitive to some relatively arbitrary technical choices made in the modelling has been shown by McKitrick[1998].

A second, somewhat heterogeneous strand of literature within the single country framework is distinguished by attempts to use microeconometric estimates of certain key relationships within an aggregate macro simulation model. An example is a controversial paper by Young[2005], which suggests a positive impact of the AIDS epidemic in South Africa on future living standards, arising as a consequence of reduced population growth, thus a raised capital-labour ratio within a standard Solow growth model. Young's conclusions have been challenged by Bell et al [2004] who employ a calibration plus aggregate simulation framework, but reach quite different conclusions as a result of placing emphasis on different causal channels. Such methods have also been used to simulate the consequences of international health inequality for international income inequality by Weil[2007].

The plan of the rest of this paper is as follows. Cross-country models, CGE models and macro simulation models are considered in three separate sections. In each case I describe one or more papers in some detail, on the basis that getting 'inside' particular papers that are at the research frontier is a good way of establishing a foundation for the identification of new research directions.

2. Cross-country models

2.1 Aggregate measures of health and correlations with aggregate economic outcomes

To begin with, we must distinguish for an individual between *health inputs*, such as nutrition, exposure to disease and availability of medical care, and *health outcomes*, such as adult height, length of life and physical and mental capacities. Cross-country models aim to use cross-country data to identify causal links from health inputs or, more usually, health outcomes in the aggregate

to economic outcomes in the aggregate, such as GDP per capita. This research programme faces a number of fundamental challenges:

(a) Even at the individual level we do not have a reliable measure of what Weil[2007] calls *human capital in the form of health* (HCH). Outcomes such as adult height are known to be correlated with productivity, but while these capture childhood experience (e.g. malnutrition) they fail to capture the productivity reducing effects of adult illness.

(b) Aggregation creates severe difficulties. For example, *average* life expectancy (LE) is certainly a much worse indicator of *average* HCH than *individual* length of life is of *individual* HCH. LE equal to 56 could mean (i) everyone lives to 56, (ii) 20% die at birth and 80% live to 70, or (iii) 50% live to 42 and 50% live to 70. Average HCH among the living is presumably highest in (ii) and is likely to be lowest in (iii), especially if we suppose that the early decedents in this case have had low productivity for some time before death, as a result of – say – AIDS.

(c) Particularly for African countries, data is lacking, and even where data exists it must be approached with caution. For example, an accurate measure of LE requires that adult mortality be accurately measured, which in turn requires a complete vital registration system for births and deaths. In practice, for most poor countries, LE has historically been estimated using regression methods with infant and child mortality (IMR and CMR) as regressors, a method which is not reliable when adult mortality is changing. Deaton[2007], Fig. 2, illustrates the point using a graph which shows a perfect linear relationship for 1970-75 between (estimated) LE and IMR for high IMR countries. This relationship disappears for 1995-2000 as a result of corrections for adult AIDS mortality, but these more recent LE data are only as good as these corrections, about which Deaton is sceptical. So while comparative *levels* of LE across countries are surely informative, *changes* in LE are likely to be very noisy and uninformative.

(d) There are strong cross-country correlations between health outcomes and economic outcomes: – graphs of LE versus GDP/head (the Preston Curve) and the adult survival rate (ASR) versus GDP/head are given in the Appendix – but such correlations as they stand tell us nothing about causation. In trying to establish a causal link from a health variable, H, to an economic outcome, Y, we must pay attention both to *reverse causation* (from Y to H) and *incidental association* (where both Y and H are driven by some other, omitted, variable). One may reasonably believe either that there is some relatively direct causal link from higher income to higher health status, and/or that both higher income and higher health status can be attributed to common causal

variables such as female education or 'governance' (operating through the efficiency with which health care is delivered).

(e) Still considering simple bivariate correlations, one way of trying to control for incidental association is to first difference the variables, thus eliminating the influence of time-invariant country effects that cause both H and Y to be high, say. So we consider the correlation between _Y and _H. As noted above, one danger with this procedure is measurement error, particularly if first differences are taken over short (say five-year) periods, when the constructed first difference may be largely 'noise'.[2]

Another issue is the form in which a particular health variable is defined for the purposes of statistical analysis, as Deaton explains. Consider, as he does, the association between IMR and GDP/capita. The data over 1960-2000 tell us that richer countries have had (i) faster growth in GDP/capita, (ii) faster *proportionate* falls in IMR, (iii) smaller *absolute* falls in IMR. So in correlations or regressions involving changes in IMR and changes in GDP/capita, whichever direction of causation is posited, the *sign* of the coefficient will depend on whether IMR is included in level or logarithmic form. In other words, in this case one can obtain the result one desires by the choice of functional form.

The conclusion is not that we must deny any causal link, say, from GDP/capita to IMR (or the reverse), but that our degree of belief in this link, and its magnitude, will not be much affected by cross-country correlations or regressions involving changes in these variables (as in, for instance, Pritchett and Summers[1996] and Anand and Ravallion[1993]). We accordingly may give more weight to time series observations on individual countries, for instance the observation that in China the rapid fall in IMR, to about 50, occurred *before* the acceleration in economic growth in about 1980 (since which time IMR has remained roughly constant).

2.2 Cross-country regression analysis

It might be concluded from the above, and some have concluded, that cross-section regressions using countries as the units of analysis are unlikely to shed any useful light on causal links from health to income (or indeed on the reverse). For example, Weil[2007] comments:

[2] This may apply to Y as well as H. For example, since correlations (regressions) of (5-year, say) GDP/capita growth with (on) population growth use the *same* estimates of population (interpolated from decennial censuses at best) to construct both variables, the results are biased in favour of a negative association.

"Papers in this group suffer from severe problems of endogeneity and omitted variable bias. For example, Bloom, Canning, and Sevilla attempt to deal with the endogeneity of health and other inputs into production by using lagged values of these variables as instruments. The identifying assumption required for this strategy to work—that the error term in the equation generating health is serially correlated while the error term in the equation generating income is not—is not explicitly stated or defended. More generally, the problem with the aggregate regression approach is that, at the level of countries, it is difficult to find an empirically usable source of variation in health, either in cross section or time series, that is not correlated with the error term in the equation determining income."

In the remainder of this section I describe two papers, one by Acemoglu and Johnson (AJ), that has attempted to circumvent some of the key difficulties, and then the paper by Bloom et al (BCS) to which Weil refers. The reason for looking closely at AJ is that this is a widely-quoted 'state of the art' paper and its authors pay careful attention to the instrument validity issue.

Acemoglu-Johnson: Disease and development: the effect of life expectancy on economic growth

The basic idea in AJ is that improvements in health potentially affect output per head through three main channels. First is the *population effect*: by reducing mortality, health improvements raise population which, ceteris paribus, lowers output per head. Second is the *TFP effect*: if workers are both healthier and live longer, their average productivity will be raised, raising output. Third is the *capital accumulation effect*: a rise in the number of workers and in the efficiency of each worker will raise output and therefore savings and capital accumulation, giving a further boost to output. AJ aim to measure the population and output effects of health improvements as proxied by LE. They find, as expected, a positive population effect but the output effect is found to be smaller, so the effect on output per head is negative.

Let us examine the methods that AJ use to reach these conclusions. They start from a production function, reproduced below

$$Y_{it} = (A_{it} h_{it} N_{it})^{\alpha} K_{it}^{\beta} L_{it}^{1-\alpha-\beta} \qquad (1)$$

In this equation, output in country i at date t depends on effective labour (the bracketed term), capital, K and land, L, via a constant returns to scale production function. 'A' stands for total factor productivity and 'h' is human capital, equivalent to Weil's HCH, mentioned earlier. 'N' is

population (or labour force). As noted above, health, proxied by LE, is supposed to affect total factor productivity, A, population, N and human capital, h.

Two alternative analyses of the effects of an exogenous change in LE on output per worker, y=Y/N, are possible. In the first, K is held fixed. Then there are positive effects on y because of rises in A and h, and an offsetting effect because of the rise in N. In the second, equation (1) is embedded in a Solow growth model, so that feedbacks from income to savings and therefore to capital accumulation are allowed to occur. This gives a larger effect from LE to y.

In either case, the result is the relationship below, in which y depends on LE, fixed country effects, time effects common to all countries, a set of controls and an error term.

$$y_{it} = \pi LE_{it} + \varsigma_i + \mu_t + Z'_{it}\beta + \varepsilon_{it} \qquad (2)$$

To eliminate the country effects, AJ take 'long differences' of (2), between 1940 and 1980 in their central case to yield:

$$\Delta y_i = \pi \Delta LE_i + \Delta\mu_t + \Delta Z'_i\beta + \Delta\varepsilon_i \qquad (3)$$

OLS on (3) will generally yield inconsistent estimates of the effects of _LE on _y because of correlation between _LE (and $\Delta Z'_i$) and the error term, __. Two potentially important sources of such correlation are reverse causation and incidental association. It is entirely plausible that increased prosperity, Δy_i, will raise LE – this might easily account for all of the correlation between these variables - and that there will be other, omitted, variables that will raise both LE and y. Each effect will result in correlation between the *levels* of LE_t and $_t$ and also between *changes* in these two variables.

To address this problem AJ construct an instrument for LE, which they call 'predicted mortality' (PM). In 1940, PM is simply total mortality in the given country from a set of 15 identified diseases. Between 1940 and 1980, at defined dates, there have been global interventions against each of these diseases. At any date t, PM for country i is constructed by adding up actual mortality in that country for diseases for which no global intervention has yet taken place and world post-intervention mortality for the remaining diseases. PM is shown, decade-by-decade, to correlate

well with actual mortality. Since in 1980 PM is taken to be zero (there have been global interventions for all 15 diseases by then, and mortality at the health frontier for each of them is close enough to zero), _PM for each country is equal to ($-PM_{1940}$): total mortality in 1940 in that country from the 15 diseases.

Is the change in predicted mortality (_PM) a valid instrument for _LE? Two criteria must be satisfied: the instrument must be highly correlated with the variable to be instrumented while being independent of the error term in equation (3). The first criterion is directly testable and is fully satisfied: _PM and _LE are highly correlated.

As for the second, zero correlation between _PM and __, no direct test is available since the error terms in (3) are unobservable.[3] Nevertheless instrumentation here has removed many of the sources of correlation between the regressor and the error term. The *reverse causation* link is cut: exogenous innovations to output per head in a given country are likely to raise _LE, but cannot affect _PM. The same argument applies to omitted variables that affect both output per head and LE, thus creating *incidental association* between them.

In spite of this, there is a source of correlation which AJ explicitly recognise (ibid, p.947) but cannot exclude, namely a direct causal link from PM_{1940} to unexplained changes in output per head (__), i.e a link *other than* via LE. Especially given the limitations of LE as an indicator of HCH discussed in section 2.1, there seems every reason to believe that a challenging disease environment in a country in 1940 might affect growth over 1940-1980 in ways that are not fully captured by the effect on LE.

Summing up, this highly ingenious state-of-the-art paper brings the limitations of cross-country modelling of the links between health and demographic and economic aggregates into sharp focus. Fully-convincing instruments are very hard to find, a problem that besets all cross-country macroeconomic modelling. AJ are rightly cautious of short-period first-differencing, both because the noise-to-signal ratio in the thus-differenced variables tends to be high and because it is not plausible to imagine that the underlying hypothesized causal processes are played out within five or ten years. Therefore they elect to use long differences, but their need for data in 1940 on LE restricts their sample, and in particular means that they can include no data from African countries

[3] With more than one instrument an 'overidentification test' can be employed. For instance, with two instruments I1 and I2 for a variable X, it is possible to test whether the influence of the instruments on the dependent variable Y is fully accounted for by the linear combination of them generated in the first stage regression.

in their main analysis. In common with many others, they use LE as their health indicator, but as discussed in section 2.1 this is not only likely to be subject to considerable measurement error, but is highly imperfect as a measure of HCH.

Bloom-Canning-Sevilla: The effect of health on growth: a production function approach

BCS is representative of a large number of papers (helpfully tabulated by the authors) which use a panel estimation approach to try to identify the causal link from health to productivity. They start from a production function similar to (1), except that variables for schooling and work experience are included. Like AJ (and nearly all of the papers they tabulate) they use LE as the health proxy. Since, unlike AJ, they use a panel – of three decades 1960-90 – they retain changes in non-health inputs in their estimating equation. They also assume slow technological catch-up, which leads to the inclusion of lagged level variables as well. The resulting estimation equation has the following form:

$$\Delta Y_{it} = \pi \Delta LE_{it} + \Delta \mu_t + \Delta X'_{it}\alpha + [X'_{it-1}\gamma - \delta y_{it-1} + Z'_i\beta] + \varepsilon_{it} \qquad (4)$$

The dependent variable is the change in log output. The first two terms on the RHS are the same as in (3). The third term encompasses the effects of changes in labour, capital, schooling and experience inputs (represented by vector X), and the bracketed level terms are there because of the assumption of slow catch-up.[4] The Zs are governance and geography variables that get into the equation by affecting country equilibrium TFP (and therefore the rate of catch up, since this is assumed to depend on how far a country is from its equilibrium TFP level).

BCS conclude from their estimation of (4) for 104 countries that a one year increase in life expectancy increases output by 4% - in the decade in which the change occurs (slow catch-up only applies to technological change). This is a far bigger effect than the 0.3% found by AJ, notwithstanding the fact that AJ's estimate is over a much longer time interval and implicitly embodies the positive output effects would be expected from capital accumulation and increased labour supply. A key question, therefore, is whether the identification strategy used stands up to examination.

[4] The detailed derivation of this is clearly explained in Bloom, Canning and Sevilla [2002]

BCS's instruments are lagged levels and changes of inputs and output. This instrument set is very likely to pass the first test of validity (BCS provide no details on this): the instruments will be highly correlated with the variables being instrumented, so that the F-statistic from the first stage regression will be acceptably high (see Bound et al [1995]). The second test of validity is that the instruments should be uncorrelated with $\_{it}$. This requires the implausible assumption of no serial correlation in the $\_{it}$. If such serial correlation is present, then the lagged values of the variables contained in (4) are thereby automatically correlated with $\_{it}$, invalidating the instrumentation used.

Summing up, the difficulty of finding valid instruments is what has led to the widespread disenchantment – illustrated by the quotation from Weil cited earlier - with the use of cross country panel regressions to identify the determinants of productivity growth. Valid instruments are the sine qua non for separating out the effects of health on income from those of income on health.

3. Computable general equilibrium models

In order to gain some insight on what can, and what cannot be achieved via the CGE approach, I use a paper by Arndt and Lewis as an example.

Arndt-Lewis: The macro implication of HIV/AIDS in South Africa: a preliminary assessment(AL)

AL's objective is to simulate the evolution of the South African economy over 2001-10 with and without AIDS, in order to identify the effects of the AIDS epidemic. To this end they construct a CGE model with fourteen productive sectors, five primary inputs (four labour types plus capital), five types of household (income quintiles), seven types of government expenditure, and three types of government investment. On both the export and import sides there is imperfect substitutability between domestic and foreign goods. Output markets are cleared by flexible prices. Some but not all labour markets clear: the unskilled wage is exogenous so that there is involuntary unskilled unemployment in equilibrium. Crucially for the results, private investment is the variable which adjusts to achieve savings-investment balance: so if private savings fall and the government deficit rises (because of health spending, for instance), private investment falls –

the inflow of foreign capital is exogenous. Investment is distributed across productive sectors according to a rule based on relative profit rates: a fall in aggregate private investment, over time, impinges on sectoral capital stocks and therefore sectoral outputs.

CGE models contain a large number of parameters (in the sectoral production functions and household utility functions, for example) and these are usually pinned down via a *calibration* exercise. AL have a social accounting matrix for South Africa at their disposal, and the parameters in their model are chosen in such a way that the data for 1997-2000 are generated as equilibrium values in their model. As a simple illustration of how this works, consider the Cobb-Douglas household utility function used in AL. In this case the exponents on consumption of the various commodities are equal to the expenditure shares, so knowledge of the expenditure shares for a particular household category (quintile in this case) determines the unknown exponents.

Once calibrated, the CGE model can be used for *simulation*. AL represent the effects of the AIDS epidemic in four ways:

(a) Reduced labour supply by skill category, based on projected extra deaths.
(b) Reduced 'effective' labour supply per worker in each skill category – assumed proportional to AIDS mortality in that category. Overall TFP growth is halved.
(c) AIDS-affected households reduce savings to zero and reallocate 10-15% of consumption away from non-food to health.
(d) Health share of government recurrent spending rises from 15% in 1997 to 26% in 2010.

From (a)-(d) one can anticipate, qualitatively, what the simulations are going to generate. On the economy's supply side, (a) and (b) together will mean a fall in output: over time (c) and (d) will make matters worse as residually-determined capital accumulation is slowed. The effect on output per capita is ambiguous: population and labour supply are falling but so are capital and TFP growth. The effect on the unskilled unemployment rate is similarly ambiguous.

In the event, GDP in 2010 is projected 17% lower in the 'with-AIDS' scenario and GDP/capita is 8% lower. Unskilled unemployment is marginally lower. AL can also disaggregate the GDP loss among the causes (a)-(d) above. So, for example, 34% of the loss is attributable to the overall TFP growth slowdown, and 46% to the crowding-out of private investment by extra government health spending.

Discussion

How robust are the conclusions in AL? As is often the case in CGE models, the results are driven by key underlying assumptions for which empirical support may be relatively weak. For example, simple projections are used for AIDS mortality by skill group. Varying what is assumed about infection and death rates among the skilled and better educated, for instance, is likely to make a considerable difference, even within the confines of the model. Grimm[2008], for example, refers to microeconomic studies suggesting a reduction in risky behaviour among the better-educated. A second example is the assumed halving of TFP growth: one third of the output effect at 2010 in the model results from this assumption, which can only be described as arbitrary.

Another general concern as far as robustness is concerned is the sensitivity of the results to the functional forms used for the production and utility functions in the model. In general there is a tension in CGE modelling between the use of simple functional forms, which allow calibration (i.e. the determination of all needed parameters from one data point) and the use of more flexible functional forms, which have too many parameters for calibration to work so that these must be obtained in another way (e.g. by time series estimation).[5] It has been shown empirically that this choice matters: i.e. that, even for relatively small shocks, simulation results can be quite sensitive to the choice of functional form (McKitrick 1998).

The main conclusion in AL is that over the time horizon considered the economic effects are large but perhaps not catastrophic. However, as will be discussed below, starkly different conclusions (in either direction) can be obtained if longer-run feedbacks are taken into account, in particular those from AIDS mortality to fertility and human capital accumulation.

Despite these reservations, CGE models such as AL are not without value. At the very least, they can spell out the implications of consistent accounting at a medium-term horizon. For example, it is useful to know, approximately, how big a fall in private investment is likely to follow from the combined effects of reduced private and government savings under the assumption of exogenous foreign savings. AL do not try to trace out potential economic benefits from health interventions, but this would be possible in their model if the links from such interventions to mortality rates from AIDS could be estimated.

[5] AL use a flexible functional form for the modelling of production but do not provide details about how parameter values were set.

4. Micro estimation, macro simulation

This approach avoids the problems discussed above. I will outline three recent papers, by Young, Bell et al and Weil, which serve both to illustrate the approach and to indicate some of the key micro relationships on which further research may be fruitful. The inclusion of Bell et al in this section is a little arbitrary: this paper really straddles sections 3 and 4 as it is aggregative, so cannot properly be described as a CGE, while at the same time relying exclusively on calibration rather than estimation of the key model parameters.

Young: The gift of the dying: the tragedy of AIDS and the welfare of future African generations

Young's paper investigates the consequences of the AIDS epidemic in South Africa, finding that the epidemic is likely to raise future living standards. How?

The macro component of Young's model is simulated not estimated and consists of a standard Solow growth model, in which a constant savings rate and zero TFP growth are assumed. In order to be able to simulate future paths for output, consumption, wages and so forth, the model requires the path of aggregate *effective labour supply*.

Young gets to this by endogenizing fertility, education and participation (hours worked) using estimates from household data. The two most important micro relationships in the model are:

(a) From HIV prevalence to fertility
(b) From orphanhood to education

HIV prevalence is found to reduce fertility enough to compensate from the destruction of human capital as a result of premature death, so that wages and therefore living standards are raised. However, (presumed double) orphanhood is assumed to end education of the orphan, reducing human capital and potentially increasing inequality. This effect persists across generations, since poorly-educated adults have more children who, themselves, get relatively little education.

To estimate the micro relationships on which his model is founded, Young uses the 1995 SA 'October household survey' plus the 1998 DHS survey (for fertility) and a two-stage estimation procedure. This means first estimating a wage equation in which wages depend on age, gender

and education, assumed exogenous. Then relative fitted wages from this equation are used in equations for labour market participation, fertility and children's education, as follows:

Participation (M and F separate) – Poisson process (fitted relative wage and age)

Fertility – Poisson process (F's fitted relative wage, age, birth year, HIV prevalence)

Children's completed education – ordered probit (F's fitted relative wage, age, birth year)

Thus education is being used as the instrument which allows the causal relationship from relative wages to participation, fertility and children's education to be identified.
These building blocks allow Young to simulate aggregate effective labour supply over time, with relative wages acting as efficiency weights in this aggregation. Among the feedbacks over time that are encompassed by this procedure are those from HIV prevalence to fertility, and from orphanhood to children's education to future effective labour supply.

Young's results raise both microeconomic and macroeconomic questions. At the micro level, his projections for participation (identical in his model to employment), fertility and education all rest on foundations that can be contested and on which more research would be valuable.

On participation, the assumption of labour market clearing may seem to strain credulity in the South African case, given what is known about rural unemployment rates. It is not immediately obvious how much difference would be likely to result from different modelling of the labour market.

On fertility, the effect of HIV prevalence on fertility is central to Young's results. His method uses a time series of HIV prevalence by quinquennial age group to explain a woman's fertility at any given date. This has been criticised (Kalemli-Ozcan 2006) on the grounds that the variable is strongly time-trended (zero prevalence before 1990, data run from 1961-98). Separation of the trend effect from the HIV prevalence effect is therefore problematic. However, Young (2007) claims similar results from a panel of African countries.

On education, the feedback from parental death to children's education is important, as are the intergenerational effects feeding back to future fertility. Evidently, the assumption that orphanhood ends the education of the orphan is very strong – the link from orphanhood to

education both is critical to the longer term consequences of the AIDS epidemic and raises important public policy tradeoffs (Bell et al [2004]).[6]

At the macro level, a key assumption made by Young is that the proportion of income saved and therefore invested in fixed capital formation is unaffected by AIDS. The health care costs associated with the epidemic are assumed to fall on consumption. This means that the potentially important feedback from increased health care to lower investment and lower future income is blocked.

I now describe a model which, by taking account of this and other adverse feedbacks reaches quite different conclusions from Young's.

Bell-Devarajan-Gersbach(BDG): Thinking about the long run economic costs of AIDS

BDG use a completely different modelling strategy to Young's - for instructive reasons. They start from the proposition that human capital formation is the 'wellspring of economic growth' and that premature adult mortality associated with AIDS will reduce human capital in the present and in the future as a result of rational household behaviour, through a number of mechanisms. Aside from the direct destruction of human capital that the death of an adult worker entails, BDG emphasize effects on spending on children's education that arise (a) from reduced *expected* household income associated with higher mortality risk, (b) from the reduced value to children of higher human capital, because of *their* higher risk of premature death. Thus risk of premature death, of both parents and their children, will both serve to reduce household education expenditure. In addition to this, the costs of caring for the sick and for orphans will lead to higher taxation, which will further reduce current spending on education. In contrast to Young, BDG do not endogenize fertility.

Putting the above mechanisms at centre stage leads BDG to use an overlapping-generations framework. Any individual reaches adulthood with a particular amount of human capital, forms a union with a partner with the same amount of human capital, and the couple then makes rational

[6] Estimation of the effects of orphanhood on education is not straightforward. It might seem that a household cross-section would provide the answer, since it could allow comparison of orphans and non-orphans within a given household. But this is wrong, since (typically) the socio-economic status of an orphan's original household is higher than that of the new household. This is a good example of the benefits that can be achieved by using a panel, which allows a difference-in-difference analysis on enrolment, comparing households with and without an adult death (Evans-Miguel 2007, Case-Ardington 2006)

decisions on the allocation of income between the education of their children and consumption. The human capital acquired by a given child depends on this, and also on whether one or both parents die. Thus the distribution of human capital in the next generation is determined and so the process continues. BDG eventually are able to simulate the evolution of human capital and its distribution over time.

One spectre raised by this analysis is of a poverty trap affecting increasing numbers of people: once a child has been orphaned with low human capital, it pairs up in adulthood with a similarly deprived person, and from that point all descendants of the union will be mired in poverty. In every generation some unfortunate extra individuals will fall into this trap. Therefore either private altruism through extended family networks or state support for orphans becomes crucial if a progressive collapse is to be avoided.

For present purposes, with the emphasis on method, I say no more about BDG's results, except to note that they are able push their analysis far enough to reach useful policy conclusions. Let us see, though, how they are able to use their theoretical model to generate numerical results. Their approach is one of calibration, as discussed earlier in the context of CGE models. For this purpose, the functional forms in their model have to be chosen with sufficiently few parameters to allow them all to be extracted when the model is made to reproduce a baseline dataset as an equilibrium. Or, as BDG put it, the model has to be specified in an 'exactly identified' form.

In BDG the baseline dataset is of employment, average schooling and output in South Africa for a series of 6 quinquennnia over 1960-90. They are forced to assume that these aggregate data can be viewed as those of representative agents, and they take it that the households involved have no premature adult mortality. Human capital, _, is assumed to evolve according to equation (5) below[7]. _=1 corresponds to being unskilled, so that a child acquires greater human capital than that as a multiplicative result of e (education), z, (representing automatic transmission from parents in a two-parent household) and parental average human capital, λ_t.

$$\lambda_{t+1} = z.e.\lambda_t + 1 \qquad (5)$$

This formulation is just simple enough to allow average human capital at the 1960 baseline, the productivity of an unskilled worker (output is proportional to effective labour input) and the 'transmission' multiplier z, which in effect determines how a given amount of education translates

[7] Simplified from equation (2) in BDG

into productivity growth for the individual, all to be identified. In BDG's simulations, z varies according to the number of surviving parents, but that variation cannot be estimated or calibrated, so has to be imposed.

The second element in the calibration involves the parental utility function, which specifies the consumption-education choice. Here, the data for consumption and education expenditures in 1960 are taken to represent a household choice, and this allows the identification of a single weighting parameter in an additively-separable household utility function. This function can then be used to trace the causal chain from changes in (rationally anticipated) premature adult mortality both of adults and their children to optimal education expenditures and therefore the transmission of human capital across generations.

This model can be linked to policy choices in a number of ways. First, assuming that it is possible to estimate the costs of achieving given falls in premature adult mortality, it is possible to simulate the effects of such falls and to estimate welfare gains that would be attained via induced rises in both productivity and spending on the education of children. Therefore the elements of a narrowly economic cost-benefit analysis of increased health spending are present. Second, it is possible to simulate the effects of providing education to orphans, not only raising their human capital, but helping them and their descendants from falling into a permanent poverty trap.

Weil: Accounting for the effect of health on economic growth

Weil takes a development accounting approach to try to identify the direct contributions of health differences to productivity differences across countries, indexed i. He starts from a production function like equation (1) except that human capital in the form of education is included. Log-linearized this gives:

$$\ln(y_i) = \ln(A_i) + \alpha \ln(k_i) + (1-\alpha)\ln(h_i) + (1-\alpha)\ln(v_i) \qquad (6)$$

Output per head, y, depends on TFP (A), physical capital per head (k), human capital in the form of education per head (h) and human capital in the form of health per head (v). He takes an assumed value for _, and has data on (y, k, h). What he needs are country estimates for v, since this will allow the As to be calculated. After that it is possible to calculate, for example, what would happen to the distribution of y across countries if levels of v were equalized.

So the problem is reduced to that of estimating the v_i, or the $\ln(v_i)$. Weil's essential idea is that a microeconomic within-country estimate of the relation between $\ln(w_{i,j})$, the log wage of individual j in country i, and a health indicator such as adult height, $H_{i,j}$, provides an acceptable estimate of the relationship between $\ln(v_{i,j})$ and $H_{i,j}$. Differences in country averages of the health indicator, H_i, can then, assuming linearity, be used to estimate the required differences in the log of country per capita human capital in the form of health, $\ln(v_i)$.

More precisely, Weil starts from a structure in which *health inputs* (I) lead to *health outcomes*, one of which is 'human capital in the form of health' (HCH). The key assumption is that inputs lead to outcomes via a scalar unobserved quantity which Weil calls 'latent health'. The effect of this assumption, plus linearity, is that, error terms aside, differences across individuals in HCH are proportional to differences in any of the other health outcomes, with the constant of proportionality different in each case. Moreover log wages and productivity are assumed proportional to HCH, so that – again neglecting error terms – the log wage and productivity differences between individuals A and B is proportional to, say, the difference in their heights.

So we have: Health inputs _ (Latent health, L _ Health indicators, I
(Latent health, L _ HCH _ Wages, W

Error terms apart, and controlling for other wage determinants, we have:

$\ln(w_i)-\ln(w_j)$ proportional to (I_i-I_j)

Weil calls the constant of proportionality the return to the indicator – 'the return to height' for instance. This is best understood as $(d\ln(w)/dL)/(dI/dL)$ rather than $d\ln(w)/dI$. In other words, it is L via HCH, rather than I, that affects wages. I is an accurate indicator of L and proportional to it.

The more complicated part of the paper, and where micro estimations are needed, is in estimating the returns to various health indicators, such as height. I now discuss how these returns to health indicators are estimated. The indicators considered are: male height (H), male adult survival rate (ASR), and age at menarche (AM).

Consider first the return to height. There are two reasons why it is not adequate simply to estimate the return to (male) height by simply using OLS to fit W to H using data on individuals, measurement error and endogeneity. These produce opposing biases. H is a noisy indicator of

latent health, leading to the return being underestimated. On the other hand, H is positively correlated with unobservables that raise productivity and wages, lead to the return being overestimated

Weil gives three methods which cater for these difficulties.

(a) instrumental variables estimation: variables representing childhood health inputs are used, assuming that these affect health – and affect wages *only* via health (Schultz 2002)
(b) monozygotic twins studies, in which birthweight is used as an instrument for adult height (Behrman-Rosenzweig 2004)
(c) UK data over 100-200 years allowing point estimates of height and physical efficiency - calories available for physical labour (Fogel 1997)

For the return to the age of menarche, methods (a) and (c) are used. Since adequate cross-country data is also available on AM, this indicator can be used to give an answer to Weil's basic question – i.e. how much of the income variation across countries is caused by cross-country variation in health.

However, this cannot be done directly for height, since data on average height are not available for a sufficient number of countries. However Weil is able to get almost 200 country-time observations for both height and adult survival rate, so can use the correlation between these variables to derive an estimate for the return to ASR indirectly. Armed with the constructed estimates for $\ln(v_i)$, Weil is then able to undertake the development accounting exercise described above.

The punchline of Weil's paper is that his estimate of the effect of eliminating health gaps internationally is much smaller than the estimates that result from cross-country macro regressions. And this is true even if some indirect effects are taken into account, such as that which would result from healthier children acquiring more education. These conclusions are in line with those reached by AJ.

5. Conclusions

The main conclusions of the paper are as follows.

(a) There is an emerging and valid consensus in the literature that cross-country regression models are of rather little value in establishing causal relations between aggregate health variables and economic outcomes of interest.

(b) This means that as far as reaching macroeconomic conclusions about the effects of health policy interventions, which is presumably the ultimate aim of this research programme, an approach which inserts building blocks derived from microeconomic studies into macroeconomic simulation models is likely to be the most fruitful.

(c) Such simulation exercises can be conducted over different time horizons. The CGE approach is best adapted to medium-term horizons, perhaps 5-15 years. However CGE models have drawbacks, even in these circumstances. The principal risk is that a multisectoral structure is chosen perhaps because it happens to conform with the structure of an available social accounting matrix, but that in fact little is gained thereby. The disadvantage of a multisectoral CGE is that it becomes very difficult to conduct a sensitivity analysis: there are simply too many parameters and functional forms which could be varied. In the Arndt-Lewis paper considered here, it seems that the main results could have probably been obtained in a much less disaggregated model.

(d) For macroeconomic analysis at all horizons, there is much to be said for the aggregate simulation approach, exemplified here in the papers by Young, Bell et al and Weil. What comes out from a comparison of the first two of these papers is how much depends on the microeconomics that feeds into the analysis. Young's relatively optimistic analysis, as far as the prospects for output per head in South Africa are concerned, rests heavily on a link from HIV infection to fertility that seems far from certain in the light of later empirical work. The grim projections in Bell et al rest on a theoretically well worked out model of human capital over time, but the empirical basis for these projections are rather speculative as the authors are at pains to acknowledge.

(e) All of the work reviewed in this paper has faced a common challenge, namely the very noisy character of the indicators available for human capital in the form of health (compare schooling for educational input and qualification obtained for educational outcome in the case of for 'human capital in the form of education'). The contrast between education and health suggests that the returns to longitudinal studies on health inputs, health outcomes, and various behaviours (schooling, labour market participation) may be particularly high (Thomas 2001).

(f) For the purposes of influencing policymakers, the fact that many of the hypothesized effects of changes in health inputs occur in the medium to long term poses a difficulty. There is therefore scope for research which tries to identify short-term effects of relevance to policymakers. One example is the provision of increased access to contraception. Positive effects of reduced fertility on output per worker take many years to occur, since in a Solow framework they arise because of slower workforce growth (it is true that effects via a falling dependency ratio on output per head may be achieved rather quicker). But a short-run effect of lower fertility is lower demand for education, and therefore a saving in the education budget in situations where governments are committed to universal primary education for all. Relatively straightforward research into budgetary tradeoffs of this kind might accordingly be of value in the project.

References

Acemoglu, D and S. Johnson [2007], 'Disease and development: the effect of life expectancy on economic growth', Journal of Political Economy, 115, 6

Anand, S. and M. Ravallion [1993], 'Human development in poor countries, on the role of private incomes and public services', Journal of Economic Perspectives vol.7,1

Behrman, J, and M. Rosenzweig [2004], 'The returns to birthweight', Review of Economics and Statistics, LXXXVI, 586–601.

Bell, C, S. Devarajan, and H. Gersbach [2004] 'Thinking about the long-run economic costs of AIDS', in M. Haacker (ed.) The Macroeconomics of HIV/AIDS, processed; Washington DC: International Monetary Fund

Bloom, D. and D. Canning (2004), 'The health and wealth of Africa', World Economics, vol.5,2

Bloom, D., D. Canning and J. Sevilla (2004), 'The effect of health on economic growth: a production function approach, World Development 32(1): 1–13.

Case, A. and C. Ardington (2006), 'The impact of parental death on school outcomes: longitudinal data from South Africa', Demography 43:401–20.

Evans, D and E. Miguel (2007),'Orphans and schooling in Africa: a longitudinal analysis' Demography, 44, 1

Bound, J., D, Jaeger and R. Baker [1995], 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', Journal of the American Statistical Association, Vol. 90

Cogneau, D and M. Grimm [2008], 'The Impact of AIDS Mortality on the Distribution ofIncome in Côte d'Ivoire', Journal of African Economies, 17, 5, pp. 688–728

Deaton, A [2008], 'Global patterns of income and health: facts, interpretations, and policies', World Institute for Development Economics Research, Annual Lecture 10

Fogel, R. (1997), 'New Findings on Secular Trends in Nutrition and Mortality: Some Implications for Population Theory.' In M. R. Rosenzweig and O. Stark, eds., Handbook of Population and Family Economics, vol. 1a. Amsterdam: Elsevier Science.

Grimm, M. [2008], 'AIDS and economy: a biased relationship', Institute for Social Studies, Public Lecture Series, no.4, The Hague

Jamison, D., Jamison, E. and J Sachs (2003). 'Assessing the Determinants of Growth When Health Is Explicitly Included in the Measure of Economic Welfare', presented at the 4th World Congress of the International Health Economics Association, San Francisco, June.

Kalemli-Ozcan, S. (2006) 'AIDS, "reversal" of the demographic transition and economic development: evidence from Africa', NBER Working Paper 12181

McKitrick, R.R. [1998], 'The econometric critique of computable general equilibrium modeling: the role of functional forms', Economic Modelling, 15, 4

Pritchett, L. and L. Summers [1996], 'Wealthier is healthier', Journal of Human Resources, XXXI, 841–868.

Schultz, T. Paul, 'Wage Gains Associated with Height as a Form of Health Human Capital', American Economic Review, XCII (2002), 349–353.

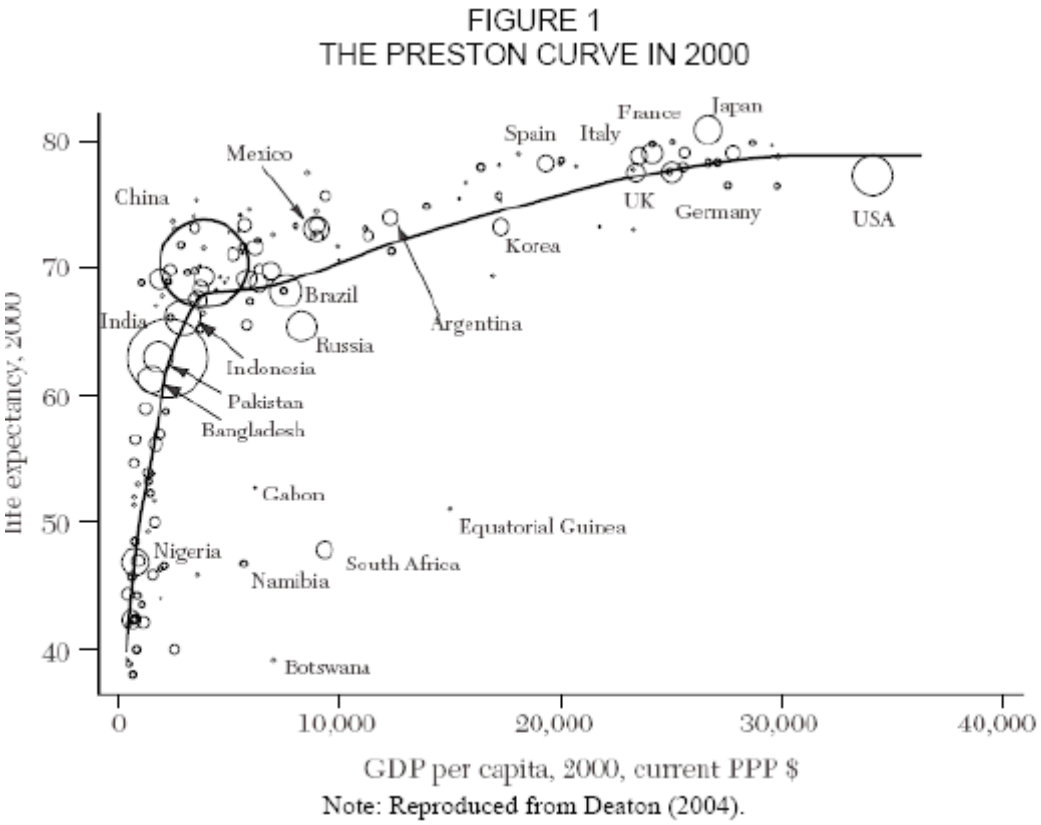
Thomas, D. (2001), 'Health, Nutrition, and Economic Prosperity: A Microeconomic Perspective', CMH Working Paper Series No.7, World Health Organization

Weil, D. [2007] 'Accounting for the effect of health on economic growth', Quarterly Journal of Economics, 122

Young, A. [2005], 'The gift of the dying: the tragedy of AIDS and the welfare of future African generations', Quarterly Journal of Economics 120:423–66.

Appendix

The Preston Curve (source: Deaton[2007])



Adult survival rate versus GDP/capita (source: Weil[2007])

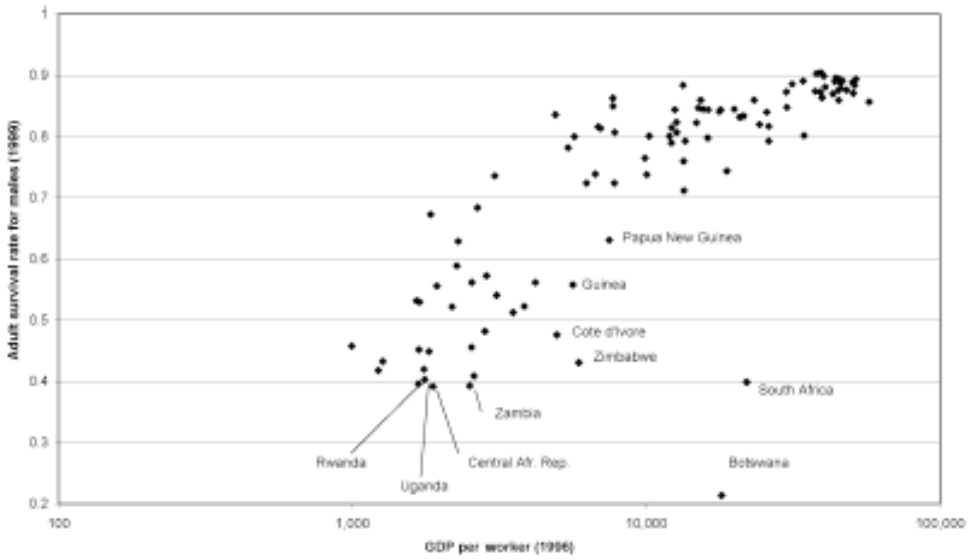


FIGURE I
GDP per Worker versus Adult Survival Rate