Three Essays on the Health and Wealth of Nations

by

Weichun Chen

B.A., Sun Yat-sen University, P.R. China, 2002
M.A., University of Victoria, 2004

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

In the Department of Economics

Three Essays on the Health and Wealth of Nations

by

Weichun Chen

B.A., Sun Yat-sen University, P.R. China, 2002

M.A., University of Victoria, 2004

Supervisory Committee

Dr. Merwan H. Engineer, Co-Supervisor
(Department of Economics)

Dr. Nilanjana Roy, Co-Supervisor
(Department of Economics)

Dr. Judith A. Clarke, Department Member
(Department of Economics)

Dr. Zheng Wu, Outside Member
(Department of Sociology)

Supervisory Committee

Dr. Merwan H. Engineer, Co-Supervisor
(Department of Economics)

Dr. Nilanjana Roy, Co-Supervisor
(Department of Economics)

Dr. Judith A. Clarke, Department Member
(Department of Economics)

Dr. Zheng Wu, Outside Member
(Department of Sociology)

ABSTRACT

This dissertation both theoretically and empirically examines the relationship between health and wealth, using proxies for health and wealth that are standard in the economics literature. We first model the endogenous interactions between life expectancy and income by modifying a standard overlapping generation model to allow individuals to directly choose their own longevity. The model displays a positive feedback between life expectancy and income that generates multiple stable equilibria. The worse equilibrium is a "poverty-trap" in which poverty and low longevity reinforce each other. The second portion of the dissertation is empirical. We first show that income has statistically significant effects on various proxies for health. The results are robust to different ways of controlling for the endogeneity of income: both instrumental variable estimation with external instruments and also generalized method of moments estimation when internal instruments are applied. We next directly test for the causal relationship between income and various proxies for health using three panel Granger causality tests. Evidence is found to support the

existence of a bi-directional causal link. Sensitivity tests further suggest that middle-income countries play a more important role than low-income countries in explaining the overall wealth-health causality.

TABLE OF CONTENTS

LIST of TABLES

LIST of FIGURES

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisors, Dr. Merwan H. Engineer and Dr. Nilanjana Roy, for initializing the idea of this dissertation, for their encouragement, patience, and intensive support throughout my graduate student years. They always had useful advice whenever I was in doubt. My special thanks to Dr. Judith A. Clarke for her advice and thorough reading of the many versions of the papers from which this dissertation is compiled. Without Drs Clarke, Engineer, and Roy, this dissertation would never have become a reality. I also would like to extend my appreciation to Dr. Ian King for his valuable help and comments in the process of writing up Chapter 1, and to Dr. Zheng Wu and Dr. Thanasis Stengos for their helpful comments and suggestions for future work.

I am in debt to the Economics Department at the University of Victoria, for generous financial aid, high quality teaching, encouraging faculty members, and friendly support from the department staff Karen Crawford, Judy Nixon, Lori Cretney and Alma Osorio.

I would also like to thank my husband, Dr. Guangyu Fu, and my parents for their persistent love and encouragement. Last, but not least, I would like to thank my friends in Victoria for the many ways in which they have helped me throughout these years.

CHAPTER 1:

INTRODUCTION

1. Introduction

The United Nations Development Programme (UNDP) published the first Human Development Report in 1990. The report describes human development as a multi-faceted process but highlights health, standard of living, and knowledge as the three most important dimensions of human development.[1] In the report, proxy variables for each of these three dimensions of human development are used in designing the Human Development Index (HDI). In the HDI, life expectancy at birth and per capita gross domestic product (GDP) are the proxy variables for representing health and standard of living. The HDI is the leading measure of human development in the Human Development Reports and has become a widely reported statistic for comparing the level of human development across countries.

This dissertation examines the relationship between health and income/wealth, two of the key dimensions represented in the HDI.[2] We concentrate on examining the specific relationship between the proxy variables life expectancy and per capita GDP. Infant mortality as a proxy for health is also examined. The third dimension of the

[1] Human Development Report 1990, page 10-12. The definition of human development is also discussed by Sen (1985, 1987)

[2] In the Human Development Reports, standard of living is viewed as the ability to have access to personal resources. This ability to access resources is generally considered to be through income and/or wealth. Here the use of the language "income" and "wealth" is usually generic and interchangeable. In this dissertation, income and wealth are approximated by GDP per capita, which is consistent with the practice in the literature. For example, Pritchett and Summers' (1996) paper is entitled "Wealthier is Healthier", even though they only look at GDP per capita.

HDI, knowledge, is only touched upon insofar as education relates to health and income. The aim of the dissertation is to understand, both theoretically and empirically, the relationship between the variables that proxy health and wealth.

Since the first Human Development Report two decades ago, the literature on the relationship between income and health has grown substantially. For the most part, the older literature supports the view that income/wealth leads to improvements in health. In this literature, health is an output of the development process. In particular, inequalities in health outcomes are explained by resource-dependent personal behaviour as well as access to and utilization of health care services (Feinstein 1993). On one hand, income development in a country provides households extra financial resources to purchase non-medical goods and services (such as foods, clothes, and houses) and public goods (such as clean water) that ensure sanitary living conditions (Flegg 1982, Subbarao and Randy 1995, Zakir and Wunnava 1999). All these goods in turn produce good health. On the other hand, economic growth also allows households to spend more money on health care services, including pharmaceuticals and information about health (Anand and Ravallion 1993, Pritchett 1997, Filmer and Pritchett 1999). According to World Bank data, high-income countries spent, on average, 10% of GDP on health expenditure in 2000 while this number in low-income countries and least developed countries is only 4%.[3]

In the older literature, the estimation results of the impact of income on health are sometimes seriously biased by the endogeneity of income. The newer literature has examined several ways in which health affects income. First, healthy people expect to

[3] Data are from www.worldbank.org/data/onlinedatabases/onlinedatabases.html. We follow the World Bank classifications for grouping countries.

live longer and save more for their old age, which increases the amount of capital available to the domestic economy (Zhang et al. 2003, Chakraborty 2004, Chen et al. 2008). Secondly, better health can be regarded as a form of human capital that enters into the production process: healthy workers are physically and mentally more robust and hence more productive (Schultz 2002, Thomas and Strauss 1997), likely leading to higher income and wealth. Thirdly, health improvements indirectly affect income through education: people with greater longevity are willing to stay longer in schools; and parents are willing to invest more in their healthy children's education. Knowledge thus acts as a fundamental driver of economic growth (Ruger et al. 2006, Croix and Licandro 1999, Kalemli-Ozcan et al. 2000). Each of these different mechanisms leads to health affecting wealth and income.

Overall, the empirical literature finds a bi-directional association between income and health. However, the income-health *correlation* estimated by most of the empirical studies does not fully explain the *causal* link between these two variables. In econometrics, a correlation captures the co-movements of different variables but does not necessary reveal the direction of causation. In this dissertation, we go beyond the analysis of simple correlations and find evidence for bi-directional causality. This finding is noteworthy as it indicates that wealth and health can reinforce each other yielding a multiplier effect and forming virtuous or vicious cycles.

This dissertation focuses on modeling and estimating the causal relationship between income and health. We first model the causal linkages between income and health (life expectancy) in a general equilibrium overlapping generation (OLG) model. To examine endogenous life expectancy, we extend the model to allow agents to choose their own longevity. The model displays bi-directional causality so that impacts on

income and life expectancy contemporaneously reinforce each other.[4] These feedbacks may be sufficiently large to generate multiple equilibria. The equilibria can be Pareto ranked, and the worse equilibrium is a "poverty-trap equilibrium" where both income and life expectancy are low.

The dissertation then turns to empirically characterize the relationship between income and health (infant mortality and life expectancy). By controlling for the endogeneity of income using instrumental variables and generalized method of moments (GMM), we confirm the existence of positive income effects on health. We then formally test for a casual (in the Granger sense) relationship between income and health (infant mortality). We apply one popular method and two newly developed approaches (Holtz-Eakin et al. 1988, Hurlin and Venet 2003, Hurlin 2004, 2005, Weinhold 1999, Nair-Reichert and Weinhold 2001), which extend the classical Granger causality test to panel data models. Each of these panel data approaches allow for various forms of country-specific heterogeneity under the null and alternative hypotheses. We not only conclude that there is bi-directional Granger causality between wealth and health, but we also suggest that, on the basis of our empirical results, middle-income countries make more contribution to the observed overall wealth-health causality.

Overall, this dissertation provides theoretical and empirical evidence for positive bi-directional feedback between proxies for health and wealth. This opens up the possibility that there may be substantial multiplier effects in how these variables impact each other. This, in turn, suggests that policies aimed at health investment and

[4] This theoretical model does not model causality using the common empirical concept of Granger (1969) causality.

income redistribution, to promote human development, might be quite effective. However, investigations of such policies are beyond the scope of this dissertation.

2. Overview of the Dissertation Chapters

In Chapter 2, the objective is to develop a theoretical model that captures the interactions between life expectancy and income. We examine a simple OLG model with production, which is standard in every way except that agents in the model have the ability to choose their longevity. Agents make this decision based on a comparison between the utility of being alive for two periods (the life span) versus the utility of truncating their life, 'exiting', at the end of the first period of their life. The endogenous choice of longevity can result in multiple stable equilibria. In particular, if the initial value of capital is small enough, then some agents would choose to exit at the end of the first period of life. We find that the number of stable equilibria depends critically on the value of capital's share of income, α : when $\alpha > 0.5$, this model produces multiple equilibria and the steady state at the origin is also stable. The existence of multiple steady states arises because of the bi-directional causality between income and life expectancy: when individuals earn more money they choose to live longer, and when individuals choose to live longer they choose to save more, which in turn increases income.

In Chapter 3, we estimate the effect of income on health. First, we re-estimate the specification in Pritchett and Summers (1996) using a panel data set with longer time dimension. Controlling for the endogeneity of income with external instrumental variables (e.g. terms of trade shocks, investment ratio, black market premium, and

real exchange rate distortion), our estimated income impact is statistically significant but has a lower magnitude than found by Pritchett and Summers (1996). The lower magnitude implies that the overall income-health association from 1960 to 2000 is not as strong as the one over the sub-sample period (1960 – 1985), as some African countries' health status deteriorated while some developing countries' health growth stagnated in the post-1985 period. Secondly, other health achievement indices (Kakwani 1993, Anand and Ravallion 1993) are applied in our regressions to capture the possible nonlinearity of health improvements. We find that these indices are better at producing education impacts with intuitive signs but they do not change the role of income in having a positive impact on health. Thirdly, the contribution of public health expenditure in explaining health is also explored in this Chapter. Unfortunately, the role of public health expenditure is sensitive to the type of data used in the empirical study: results from cross-section regressions and panel regressions are quite different. Finally, to avoid the difficulty of finding appropriate external instrumental variables, we extend the model into a dynamic framework and control for the endogeneity of income using system GMM with internal instruments. Again, our results find a statistically significant income effect on health.

Chapter 4 focuses on testing for Granger causality between income and health with panel data. The classical Granger causality test is extended to panel models using three different methodologies (Holtz-Eakin et al. 1988, Hurlin and Venet 2003, Hurlin 2004, 2005, Weinhold 1999, Nair-Reichert and Weinhold 2001) that control for possible heterogeneity across countries or over time. The latter two approaches are relatively new, being less applied in empirical research than the older holtz-Eakin et al. (1988) approach. However, they have the advantage of being more flexible in the manner in which heterogeneity is modeled. Based on time series from developing

countries, all of these three panel causality tests suggest the existence of a bi-directional causal link between income and health. To ascertain whether this finding is sensitive to which countries are included in the panel, we repeat the analysis, splitting the countries into those that can be regarded as "middle-income" and those that are "low-income". Evidence is again found to support the bi-directional causality for middle income countries from all the tests. However, the two newly developed approaches (Hurlin and Venet 2003, Hurlin 2004, 2005, Weinhold 1999, Nair-Reichert and Weinhold 2001) failed to support causality for low income countries. This suggests that middle-income countries are driving the health-wealth causality result for the sample of all developing countries.

Chapter 5 concludes the dissertation by summarizing key results and suggesting some directions for further research.

CHAPTER 2:

CHOOSING LONGEVITY WITH OVERLAPPING GENERATIONS

1. Introduction

Life expectancy and income are considered to be key indicators of the "quantity" and "quality" of life (Becker et. al. 2005). To study the theoretical relationship between life expectancy and income, we employ an overlapping generations (OLG) model. Diamond's (1965) OLG model is a foundation framework in economics for analyzing how demography affects growth and welfare. An OLG model with agents who are born, live through a fixed-length lifecycle and then die in the model not only captures intergenerational heterogeneity but also provides a simple framework to study the accumulation of capital across generations. In this chapter, we extend the OLG model by enodogenizing agents' life expectancy and examine the interaction between longevity and income.

Most of the theoretical studies of life expectancy and income growth assume life expectancy is an exogenous variable and discuss its effects on human capital investment, labor supply, fertility, saving accumulation, and so on (Zhang et. al. 2003; Kalemli-Ozcan et. al. 2000; Croix and Licandro 1999; Chakraborty 2004; Ehrlinch and Liu 1991). These theoretical models only capture the causality from life

expectancy to income. However, empirical work (Pritchett and Summers 1996, Brinkley 2002, Lorentzen, Mcmillan and Waczairg 2005) supports the existence of a bi-directional causality between life expectancy and income. For theory to capture the bi-directonal causality, life expectancy must studied as an being endogenous. Only recently, have economists started to model the implications of endogenous longevity for growth.

Chakraborty (2004) develops an OLG model in which life expectancy is described as a function of public health expenditures. On one hand, the public health expenditures, and then the life expectancy, are funded by taxes and are increasing in wage income. On the other hand, agents' wage income depends on the society's capital accumulation, which is increasing in longevity. His model hence produces interactions between life expectancy and savings and multiple equilibria to explain the health-income trap: a short life expectancy slows down the capital accumulation and economic growth, while a lower income shortens the life expectancy. Batachaya and Qiao (2005) and Finlay (2006) extend Chakraborty's model by defining longevity as a function of private health expenditures, but discussions still focus on examining the indirect choice of longevity, that is, what an agent directly chooses is health expenditure rather than the length of her life. In the OLG model of Blackburn and Cipriani (2002), parents' choice of education of their children indirectly determines the children's longevity and their income. Again, agents in their model are not allowed to choose their own longevity.

Unlike these papers, we model each agent's longevity as a direct choice of that agent. We examine a simple two-period OLG model with production which is standard in every way except that the representative agent in this model has the ability to choose

the number of periods to live through. Agents make this decision based on the comparison between the utility of being alive in the second period and the utility of exiting at the end of the first period (we denote this as a parameter: x). Multiple equilibria can be produced by this endogenous choice of longevity. In particular, if the initial value of capital is small enough, some agents would choose to exit, and the number of steady state equilibria depends critically on the value of capital's share of income α : with α >0.5, besides one stable equilibrium with positive capital, the steady state at the origin is also stable.

In the equilibria of this model, higher values of capital and income are associated with longer average life expectancy. The multiple steady states also help to explain a bi-directional causality between these two variables: when individuals earn more money, they choose to live longer, and when individuals choose to live longer they choose to save more – which increases income. We also show the robustness of our model to two extensions: allowing agents to choose exit probabilities for the second period, and introducing public health in a way similar to Chakraborty (2004).

This chapter contributes to the theoretical literature which models the bi-directional causation between life expectancy and income. Unlike the previous literature we model the individual's direct choice of longevity and the implications of this choice on economic growth. We derive analytical solutions for the equilibrium transition function and characterize behavior in the steady states. Our analysis is based on a simple logarithmic period utility function, a Cobb-Douglas production function, and complete depreciation of capital. These featues of the model simplify our framework and allow clear interpretations of the roles of the parameters in the results mentioned above.

The chapter proceeds as follows. Section 2 describes the model. Section 3 characterizes the properties of the equilibrium, and analyses its comparative dynamics. Section 4 analyses two extensions of the base model. Section 5 concludes. Proofs of some of the propositions are presented in an Appendix 1, and the extension using arbitrary depreciation values is contained in Appendix 2.

2. The Model

Time is discrete, agents live for (at most) two periods, and generations overlap. In each time period, a constant number (normalized to unity) of young agents is born. Each agent within any generation is identical *ex ante*. As is standard, we refer to agents born in period t as "young agents" and those surviving through period t+1 as "old agents". All agents are endowed with one unit of labour when young, and none when old. Each young agent in period t chooses whether or not to exit life (terminate her life) at the end of period t. Apart from this decision to exit, the model is the same as Diamond's (1965) growth model.

The Young Agents' Problem

The novel feature of the analysis is the decision to exit life. Let $I_t \in \{0,1\}$ denote an indicator function, where $I_t = 1$ indicates a decision by agent t made in period t to exit life at the end of period t (i.e. at the end of the period of youth), and $I_t = 0$ indicates the decision not to exit. [5]

[5] Modelling the exit choice as discrete is not restrictive in this environment. In Section 4 we consider the problem of allowing agents to choose a probability of exit, from [0,1]. We show that the choice of a

We represent utility of agents born in period $t = 0,1,2,...$ as follows:

$$U_t = \begin{cases} u(c_{1t}) + \beta u(c_{2t+1}) & \text{if } I_t = 0 \\ \\ u(c_{1t}) + x & \text{if } I_t = 1 \end{cases} \tag{2.1}$$

where c_{1t} is consumption when young, and c_{2t+1} is consumption when old provided that the agent does not exit. Utility is time separable, where $u(c)$ is the utility from consuming in the period and $\beta \in (0,1)$ is the discount factor. Here the parameter $x \in \Re$ represents the agent's perception of the value, in utils, of exiting life early. Thus, x can be interpreted as the opportunity cost of living long. We introduce this parameter explicitly because we study circumstances where exiting may be the most palatable choice.[6] We assume that all agents have a common perception of this value.

Throughout the paper we also restrict attention to logarithmic utility:

$$u(c) = \ln c\,.$$

We choose this specification not only because it is standard, and easy to work with, but also because the decision to exit is plausible when agents are poor and not

---

corner is always superior to any interior solution.

[6] Blackburn and Cipriani (2002), Becsi (2003), Chakraborty (2004) and Finlay (2006) all set $x = 0$, either implicitly or explicitly. This is a harmless normalization when individual actions do not directly affect the probability of exit, but not here – where this choice is the central focus of the study. Hence, we prefer to keep this general. This also allows for an examination of the consequences of changes in x. In principle x can be inferred with knowledge of the other parameters in the model. The utility value of being alive in the second period of life, $\beta u(c_{2t+1}) - x$, can be used to derive the value of being alive in old age in terms of the numeraire good. This contrasts with the value of a statistical life which is derived from the willingness to pay for a marginal reduction in mortality.

infinitely averse to the prospect of an early death. With more general utility, it is straightforward to show that no agent would ever choose to exit if $\beta u(0) \geq x$. This is ruled out by the logarithmic specification.

The period constraints for youth and old age are, respectively, $c_{1t} + s_t = w_t$ and $c_{2t+1} = R_{t+1} s_t$, where $R_{t+1} = 1 + r_{t+1}$ is the gross interest rate and s_t is savings of the young. If an agent exits we assume for simplicity that any savings made at time t is discarded. [7]

Recall that exit occurs at the end of the period of youth. Agents that exit face no future decisions. Agents that do not exit trivially choose to consume $c_{2t+1} = R_{t+1} s_t$, in the usual way, in old age. Knowing this, agents born in period t choose in youth s_t, c_{1t}, and I_t, to maximize utility U_t subject to $c_{1t} + s_t = w_t$. The exit choice is discrete, so we consider the two possible cases.

Case $I_t = 0$

This is a standard life-cycle consumption problem, with the solutions:

$$s_t = \frac{\beta}{1+\beta} w_t \qquad c_{1t} = \frac{1}{1+\beta} w_t \qquad c_{2t+1} = \frac{\beta}{1+\beta} w_t R_{t+1} \tag{2.2}$$

Substitution of (2.2) into (2.1) yields the maximized value function:

[7] Of course, in this discrete case, no agent ever chooses both exit and positive savings. We get the same general results if we use an annuity market. Under either specification, the interest rate is exogenous to individuals and thus individual choice is generically similar. We introduce an annuity market in our extended model in Section 4.

$$V_{0t} = \ln(\frac{w_t}{1+\beta}) + \beta \ln(\frac{\beta w_t R_{t+1}}{1+\beta}) \tag{2.3}$$

Case $I_t = 1$

In this case, to maximize (2.1), the agent sets $s_t = 0$, and so $c_{1t} = w_t$. This implies a maximized value function:

$$V_{1t} = \ln w_t + x \tag{2.4}$$

This leads to the following lemma.

***Lemma* 1:** Given the wage rate $w_t > 0$ and the gross interest rate $R_{t+1} > 0$, young workers choose not to exit life, $I_t = 0$, at the end of the first period if and only if:

$$x \le \beta \ln w_t + \beta \ln R_{t+1} + \beta \ln \beta - (1+\beta)\ln(1+\beta) \tag{2.5}$$

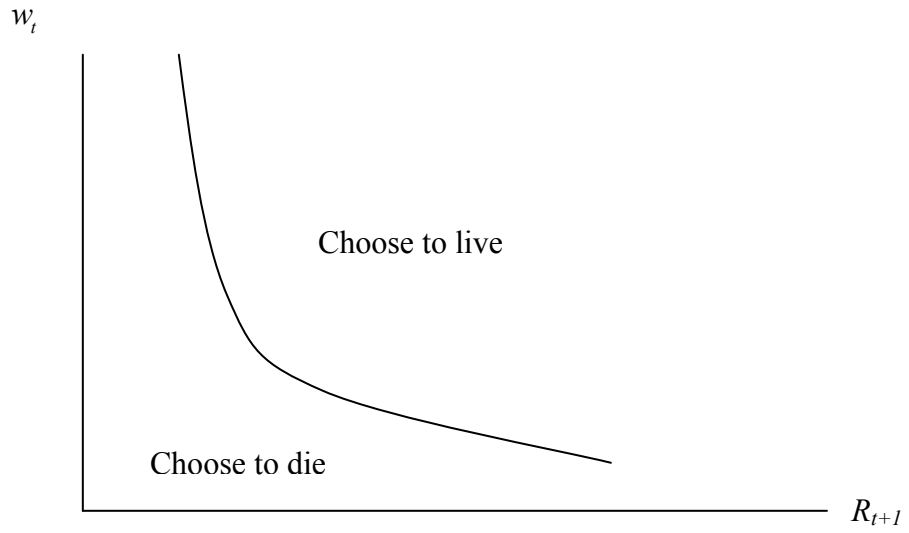
Proof: Agents will choose $I_t = 0$ if and only if $V_{0t} \ge V_{1t}$. Using equations (2.3) and (2.4) in this condition, one obtains (2.5). ■

Equation (2.5) identifies a critical value of x, in terms of the discount factor β and the rates of return w_t and R_{t+1}, beyond which agents would prefer not to live in the second period. Alternatively, given x and β, equation (2.5) identifies combinations

of w_t and R_{t+1}, associated with choosing to live in the second period. Agents are indifferent between exiting or not when (2.5) holds with equality.

Figure 1 identifies a locus of points in (w_t, R_{t+1}) where agents are indifferent about the two choices. It is easy to show that this locus is a rectangular hyperbola. Also, an increase in *x* (i.e. the value of exiting early) shifts the locus outwards indicating that individuals need higher wages or interest rates to choose not to exit.

Figure 1: A Locus of Indifferent Points



The Firms' Problem

The firms' problem in this model is entirely standard. There are many competitive firms in this economy, and the number is normalized to unity. In any period *t,* each firm takes $\{w_t, r_t\}$ as given and solves the following problem:

$$\underset{\{Y_t, K_t, L_t\}}{Max}\ \Pi_t = Y_t - (r_t + \delta)K_t - w_t L_t$$

subject to the production technology:

$$Y_t = AK_t^{\alpha} L_t^{1-\alpha}$$

where $A > 0$, and $\alpha \in (0,1)$. To derive simple closed form solutions we assume 100% depreciation, $\delta = 1$.[8] Defining capital per worker $k_t \equiv K_t / L_t$, the firm's problem leads to following standard first order-conditions in intensive form:

$$w_t = (1-\alpha) A k_t^{\alpha} \tag{2.6}$$

$$R_{t+1} = \alpha A k_{t+1}^{\alpha-1} \tag{2.7}$$

Equilibrium

Let p_t denote the fraction of young workers, in period *t*, who choose $I_t = 0$, i.e. not to exit. In any interior solution, where workers are indifferent about choosing to exit or not, p_t is determined where condition (2.5) holds with equality. The capital market equilibrium condition is influenced by p_t: since only those agents who do *not* exit save, p_t is also the proportion of savers. Savers provide the capital for period t+1. In equilibrium this supply must be equal to the demand for capital by firms:

$$k_{t+1} = p_t s_t \tag{2.8}$$

[8] The general case with $\delta \in [0,1]$ is developed in Appendix 2. Depreciation does not affect our generic results for the number and stability of equilibria. However, it does affect the shape of the implied Preston curve, as discussed below.

The labour market equilibrium condition in this model is perfectly standard:

$$L_t = 1 \tag{2.9}$$

Each young agent supplies one unit of labour inelastically, the number of agents in each generation is normalized to unity and, in equilibrium, this is equal to the demand for labour from firms.

Definition of a competitive equilibrium

A competitive equilibrium in this model, given k_0, is a set of wages, interest rates, and fractions of savers $\{w_t, r_t, p_t\}$ and a set of allocations $\{c_{1t}, c_{2t}, I_t, s_t, k_t\}$ such that

a) Individuals are maximizing utility (2.1) given the budget constraints, wages, and interest rates, with behaviour given in equations (2.2)-(2.5) and Lemma 1.

b) Firms are choosing capital and labor to maximize profits, subject to the production technology, wages and interest rate (equations (2.6) and (2.7) are satisfied).

c) Supply equals demand in the factor markets: equations (2.8) and (2.9) are satisfied.

3. Properties of the Equilibrium

For a given proportion of savers, p_t, we can derive the following equilibrium transition function from equations (2.2), (2.6), (2.8) and (2.9):

$$k_{t+1} = p_t \frac{\beta}{1+\beta}(1-\alpha)Ak_t^{\alpha} \tag{2.10}$$

The following lemma establishes that the equilibrium proportion of savers is strictly less than unity for small values of k_t, and increasing until a threshold level $\tilde{k}$ is achieved, at which point $p_t = 1$, and all agents choose not to exit the economy.

***Lemma 2*:** For any given configuration of parameters (α, β, A, x), the equilibrium relationship between the proportion of savers p_t (i.e. those that do not exit) and capital per worker k_t satisfies:

$$p_t = p_t^* \in (0,1) \quad \Leftrightarrow \quad 0 < k_t < \tilde{k} \tag{2.11a}$$

$$p_t = 1 \quad \Leftrightarrow \quad k_t \geq \tilde{k} \tag{2.11b}$$

$$p_t = 0 \quad \Leftrightarrow \quad k_t = 0 \tag{2.11c}$$

where $p_t^* = \left(\frac{\rho_2}{\rho_1}\right)\left[(1-\alpha)Ak_t^{\alpha}\right]^{\frac{\alpha}{1-\alpha}}$ is the internal proportion of savers,

$\tilde{k} = \left[\frac{1}{(1-\alpha)A}\left(\frac{\rho_1}{\rho_2}\right)^{\frac{1-\alpha}{\alpha}}\right]^{\frac{1}{\alpha}} > 0$ is a threshold level of per worker capital,

and coefficients

$$\rho_1 \equiv \frac{\beta}{1+\beta} > 0 \ , \qquad \rho_2 \equiv \left[\frac{\alpha\beta A}{e^{\frac{x}{\beta}} (1+\beta)^{\frac{1+\beta}{\beta}}} \right]^{\frac{1}{1-\alpha}} > 0 .$$

Proof: See Appendix 1.

Notice that this lemma tells us that the threshold value $\tilde{k}$ is positive for all $x \in \Re$. By (2.11a), this implies that $p_t < 1$ for some range $k_t \in (0, \tilde{k})$. The composite parameter ρ_2, importantly, has x as a key component. It is also easy to show that $\partial\rho_2 / \partial x < 0$ and so $\partial\tilde{k} / \partial x > 0$. Intuitively, as the value of exiting life early increases, the critical value of k for which all individuals choose not to exit early also increases: the economy must offer more to individuals if they are to choose not to exit.

We are now in a position to characterize the equilibrium transition function. Substitution of the equilibrium values of p_t described in Lemma 2 into equation (2.10) yields the following:

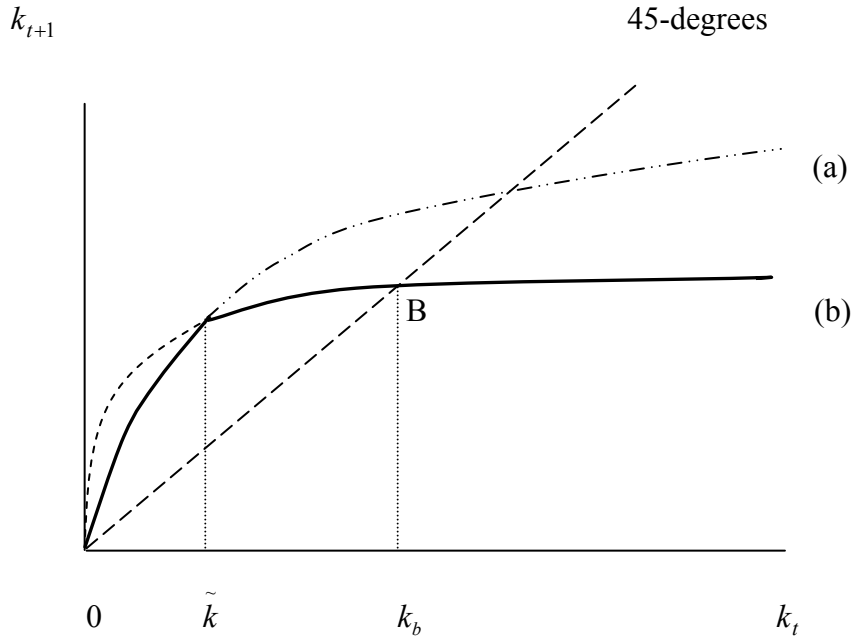
$$k_{t+1} = \begin{cases} \rho_2((1-\alpha)Ak_t^{\alpha})^{\frac{1}{1-\alpha}}, & \forall 0 \le k_t < \tilde{k} \quad (2.12a) \\ \rho_1(1-\alpha)Ak_t^{\alpha}, & \forall k_t \ge \tilde{k} \quad (2.12b) \end{cases}$$

Equation (2.12b) is the standard equilibrium transition function from Diamond's model, with logarithmic preferences and Cobb-Douglas production technologies. In our model it describes the equilibrium only for values of k greater than the threshold

$\tilde{k}$, beyond which no-one chooses to exit: p_t = 1. Equation (2.12a) is the transition function when agents are indifferent to exiting or not – derived directly by substitution of p_t^* from equation (2.11a) into equation (2.10).

Figures 2 and 3 illustrate the equilibrium transition function as the inner envelope of the loci (a) and (b) which corresponds to equations (2.12a) and (2.12b) respectively. These two diagrams are drawn under different assumptions about the value of α, as described in the next paragraph. In both diagrams, however, the transition function is represented by the solid line. Locus (a) applies along the range $k_t \in (0, \tilde{k})$, and locus (b) applies for all $k_t \geq \tilde{k}$. Along the transition path, the proportion of savers $p_t = p_t^* < 1$ increases with k_t until $k_t = \tilde{k}$ is reached, after which, p_t =1. [9]

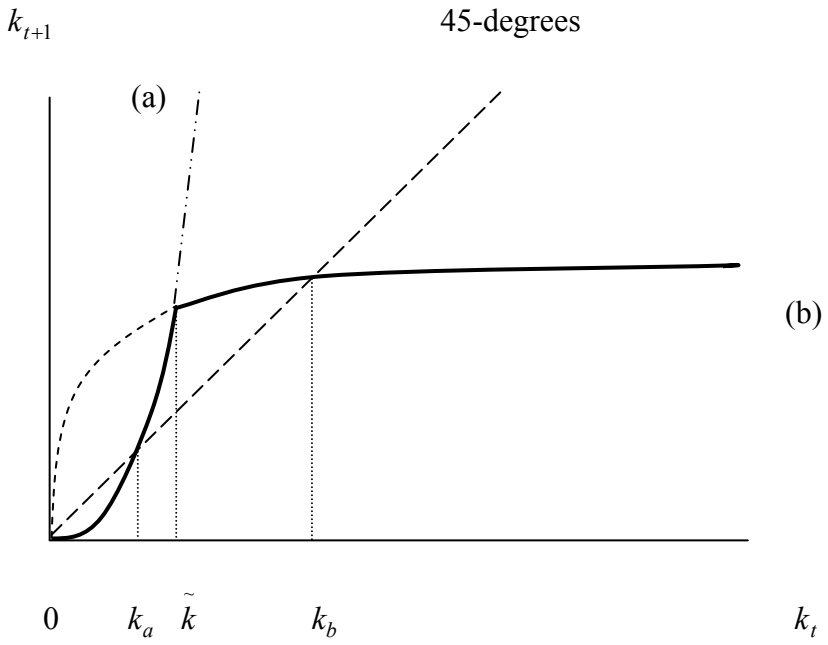
Figure 2:[10] Transition functions with $\alpha \leq 0.5$



[9] It is straightforward to show, from (2.12a) and (2.12b), that locus (b) lies above locus (a) for all $k_t < \tilde{k}$ and vice versa for all $k_t > \tilde{k}$.

[10] The solid black lines in Figures 2-4 and 6 represent the equilibrium transition functions.

Figure 3: Transition functions with $\alpha > 0.5$



The shape of the transition path depends on the value of α. Locus (b) is clearly strictly concave, but the curvature of locus (a) depends on α. If $\alpha < 0.5$, then the locus (a) is strictly concave, and thus the equilibrium transition function (2.12) is strictly concave. This is the case drawn in Figure 2. However, if $\alpha > 0.5$ then the locus (a) is strictly convex.[11] In this case, the transition function is strictly convex for all $k_t \in [0, \tilde{k})$, and concave for all $k_t \geq \tilde{k}$. This is the case drawn in Figure 3.[12]

[11] Intuitively, from equation (2.8), in any period *t*, two variables affect the capital stock in period *t* +1: s_t and p_t. From (2.2) and (2.6), s_t is increasing and concave in k_t in the usual way. From Lemma 2, when p_t is interior, it is also increasing in k_t, but the sign of it's second derivative with respect to k_t depends on α. When α is large, capital's share of income is large, and any increment in capital affects second period returns more than proportionately, increasing p_t more than proportionately. When α is large enough, the product $p_t s_t$ responds in the same way, producing a concave transition function.

[12] When $\alpha = 0.5$, locus (a) is a straight line. With a big *A* and $\rho_1(1-\alpha)A > 1$, the slope of this line is larger than 1, so the equilibrium transition function (the solid line) is concave and leads to a stable steady state with a positive capital. This is analogous to Figure 2. If $\rho_1(1-\alpha)A > 1$ (or *A* is too small), then the slope of this line is smaller than 1, and the equilibrium transition function is concave but only meets the 45 degree line once at the origin, that is, the only stable equilibrium of the economy is a degenerate equilibrium. When $\rho_1(1-\alpha)A = 1$, it is interesting to observe that locus (a) overlaps

Steady State Equilibria

A steady state equilibrium is a competitive equilibrium in which $k_{t+1} = k_t = k$ in the transition function (2.12). Proposition 1 summarizes key properties of these equilibria.

***Proposition* 1:** Steady state equilibria have the following properties.

(i) If $\alpha < .5$ then there exist two steady state equilibria, one of which is degenerate and the other which has positive income and longevity. The degenerate steady state is unstable and the non-degenerate one is stable. The stable (non-degenerate) equilibrium may have $0 < p < 1$ or $p = 1$, depending on parameter values.

(ii) If $\alpha > .5$ then the number of steady state equilibria depends on the value of *A*. The critical value $\overline{A}$ is given by:

$$\overline{A} = \frac{e^{\frac{x(1-\alpha)}{\beta}} (1+\beta)^{\frac{(1+\beta)(1-\alpha)}{\beta}}}{\beta^{\alpha}\alpha^{1-\alpha}(1-\alpha)^{\alpha}}$$

If $A < \overline{A}$ then the unique steady state equilibrium is degenerate. If $A \geq \overline{A}$ then three steady state equilibria exist, two of which are stable. Of the

---

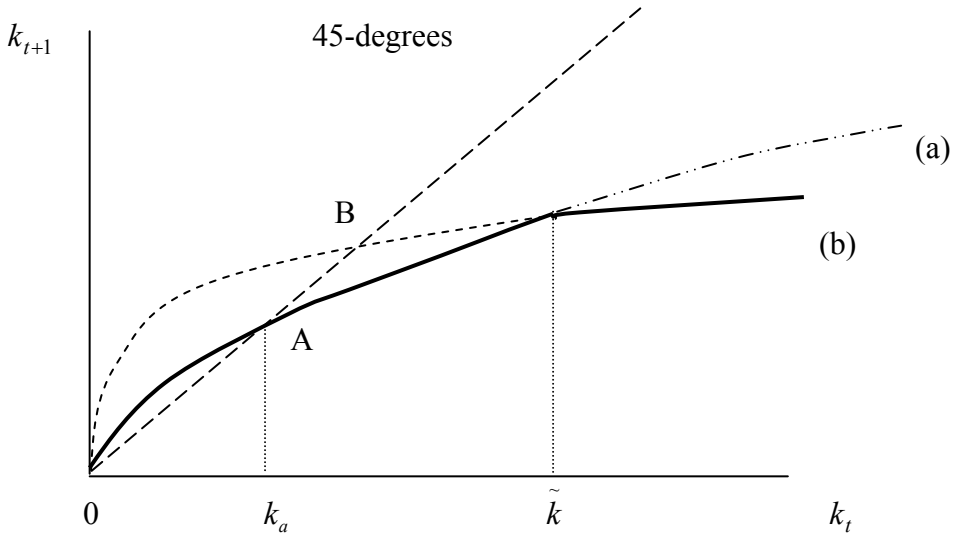
with the 45 degree line, so part of the equilibrium transition function stays on the 45 degree line. Hence, there are many stable steady states in this economy and each of them is Pareto-ranked. This is a generic case of Figure 3, where the economy produces multiple equilibria, but this case is "better" than the case in Figure 3 as the economy will not quickly deteriorate to the degenerate equilibrium. As this chapter discusses how saving and life expectancy reinforce each other and produce a virtuous or vicious cycle between these two variables, in the following sections we focus on comparing cases with $\alpha < 0.5$ and $\alpha > 0.5$

stable equilibria, one is degenerate and the other has positive income and longevity with $p = 1$.

Proof: See Appendix 1.

If $\alpha < 0.5$, then only one stable steady state equilibrium exists, and it has a strictly positive value of k. There are two cases. In Figure 2, this steady state is represented by the point *B*, where the locus (b) intersects the 45 degree line at k_b. This figure is drawn under the assumption that the critical value $\tilde{k}$ is smaller than k_b. Hence, in this steady state equilibrium, all agents choose not to exit: $p = 1$, as is implicitly assumed in Diamond's model. However, as illustrated in Figure 4, the model also allows for the possibility that the locus (a) intersects the 45 degree line (at point *A*) before it intersects locus (b). In this case, the steady state equilibrium is at point *A*, with capital stock $k_a < \tilde{k}$. In this equilibrium, in every period, some fraction of agents chooses to exit: $p < 1$. In both cases, however, the unique stable steady state equilibrium has a positive income level – the degenerate steady state equilibrium at $k = 0$ is unstable.

Figure 4: Transition functions with $\alpha \leq 0.5$ and some agents choose to exit at the stable steady state



If, however, $\alpha > 1/2$, then multiple stable steady state equilibria may exist. The locus (a) is convex, so the transition function in (2.12) is convex for $k < \tilde{k}$, and is concave for $k \geq \tilde{k}$, beyond the point where locus (a) intersects locus (b). Figure 3 illustrates how this can generate multiple steady states, when $A > \overline{A}$.[13] In this case, three steady state values of *k* exist: 0, k_a, and k_b, where $0 < k_a \leq k_b$. The intermediate steady state, k_a, is unstable and the other two are stable. Starting with any $k_0 > k_a$ yields a transition path to k_b; whereas starting with $k_0 < k_a$ yields a transition path to the origin. The steady state at the origin is stable and is the worst outcome with extreme poverty and minimum life expectancy of $1+p_t = 1$. This poverty trap is a *death trap* for poor economies. The poor exit after the period of youth and do not save, perpetuating extreme poverty.[14] Finally, when $A < \overline{A}$, the (a) locus is so low that it crosses the 45 degree line beyond the point k_b in Figure 3. In this case, the unique steady state equilibrium is at the origin – and, as before, this equilibrium is stable.

The basic model also has interesting implications for the path of wages and interest rates even ignoring the possibility of poverty traps. Notice, for example, that, when $\alpha < .5$, the transition path in Figures 2 and 4 initially lies on locus (a), which is below the transition path (b) for Diamond's model. Thus starting from a small k_0, the basic model displays higher interest rates, lower aggregate savings, and hence lower rates of wage growth on the transition path – as long as locus (a) lies below locus (b).

[13] Notice that this condition can be re-written as a restriction on *x*, given *A*, rather than the way we have expressed it here (and in Proposition 1). Written in this alternative way, multiple steady states exist if *x* is above a threshold value (identified in the formula in part (ii) of the Proposition). We chose to express this condition in terms of *A* because it is more straightforward to do.

[14] The model gives the stark result that in the poverty trap capital and wages are zero. In this dire situation, young agents, if they had the choice, might prefer to exit at the beginning of the period of youth. To avoid this possibility, one could extend the model by allowing young agents access to a sufficiently attractive primitive technology that needs only labour (e.g. hunting/gathering or simple agriculture). The poverty trap equilibrium in this extended model involves society de-industrializing so that there is a switch to the more primitive technology.

Even if the transition path eventually joins the Diamond transition path at $p = 1$, the overall time to converge to the steady state will be longer than in Diamond's model.

Income and Longevity

We now consider the implications that this model has for the relationship between per capita income and average longevity. In the model agents all live through youth and the proportion p_t live through old age. Thus, the average longevity of agents born at *t* (i.e., the average *cohort life expectancy*) in the model is $1 + p_t$. From Lemma 2 we know that there is a threshold level of capital, $\tilde{k}$, above which $p_t = 1$ and below which $p_t = p_t^* < 1$. Defining y_t as income per worker, we can then identify $\tilde{y} = A(\tilde{k})^{\alpha}$ as the threshold value of y corresponding to $\tilde{k}$. Using Lemma 2, we have the following new lemma.

Lemma 3. For any given configuration of parameters (α, β, A, x), the equilibrium relationship between income per worker y_t and average longevity of agents $1+ p_t$ born in period *t*, is:

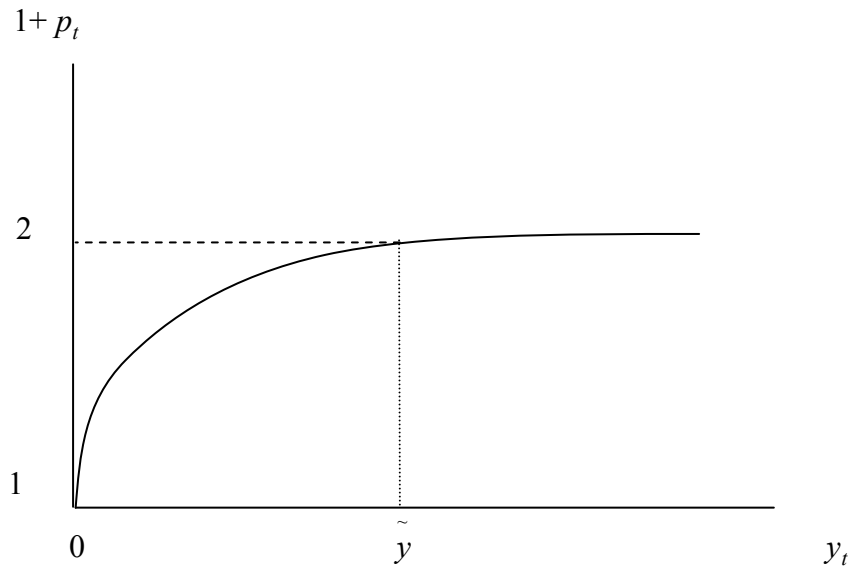
$$0 \le y_t < \tilde{y} \quad \Rightarrow \quad 1 + p_t = 1 + p_t^* = 1 + \left(\frac{\rho_2}{\rho_1}\right)((1-\alpha)y_t)^{\frac{\alpha}{1-\alpha}} < 2$$

$$y_t \ge \tilde{y} \quad \Rightarrow \quad 1 + p_t = 2.$$

For $y_t < \tilde{y}$, average longevity is increasing in y_t and is strictly concave if $\alpha < 0.5$, strictly convex if $\alpha > 0.5$, and linear if $\alpha = 0.5$.

Figure 5 illustrates the case when $\alpha < 0.5$; here, longevity is increasing and strictly concave in income per worker below the threshold $\tilde{y} = A(\tilde{k})^{\alpha}$, and constant, at the maximum longevity of 2 periods, for all $y_t > \tilde{y}$.

Figure 5: Longevity and income per worker with $\alpha < 0.5$



The concave relationship in Figure 5 resembles the "Preston curve": the empirical relationship named after Preston (1975), and studied somewhat extensively. However, the comparison is only suggestive because, empirically, the Preston curve is usually expressed with different variables on the axes. The vertical axis measures life expectancy, but does so using current survivorship data. That is, in the data, life expectancy in period t is represented by $1+p_{t-1}$. Also, the horizontal axis typically measures income *per capita*, rather than income *per worker*. Moreover, income per capita averages output over the young and the old in period t: $y_t/(1+p_{t-1})$.

Lemma 4. For any given configuration of parameters (α, β, A, x), the equilibrium relationship at time t between the average per capita income, $y_t/(1+p_{t-1})$,

and life expectancy, $1+p_{t-1}$, is increasing and strictly convex until the lifespan of $1+p_{t-1}=2$ is reached.

Proof: See Appendix 1.

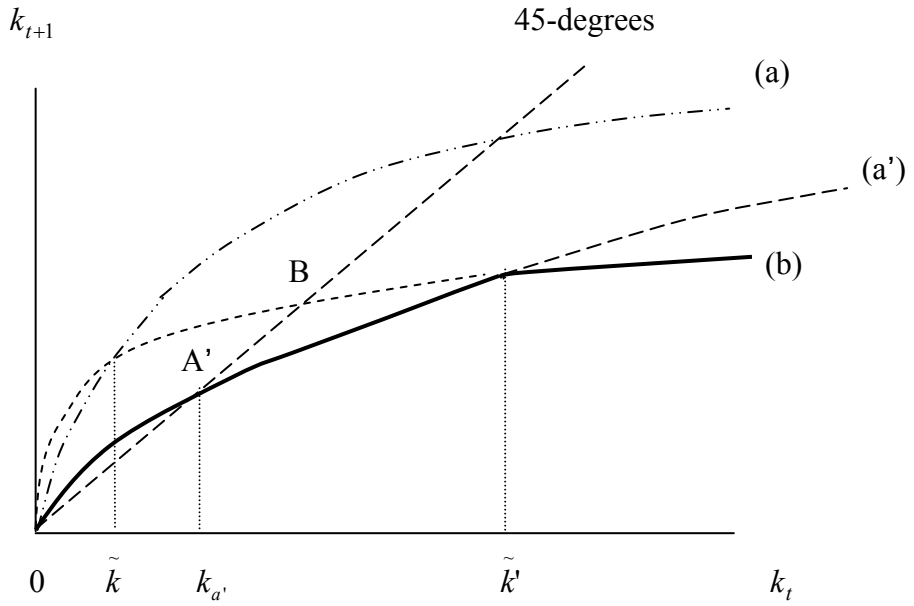
Along the transition path in this model, as in the Preston curve, life expectancy and per capita income move together until a threshold income level is reached – beyond which life expectancy is constant. However, the model predicts a strictly convex relationship in these measured variables whereas the Preston curve is usually described as concave. Recall, though, that our basic model assumes 100% depreciation. In Appendix 2, we show that, with incomplete depreciation, the Preston curve starts initially convex but may become concave as income increases.

Comparative Dynamics: The Effects of Changing x

The parameter x affects the equilibria in intuitive ways. First consider the case (i), with $\alpha < 0.5$, in Proposition 1, where there is a unique stable steady state. Figure 6 illustrates the effect of an increase in the utility value to exiting from x to x'. Suppose, initially, with x, the economy is in a steady state at point B on locus (b) and therefore p = 1 (the same as in Figure 2). An increase in x shifts the (a) locus downwards, but will leave the (b) locus unaffected. This will lead to an increase in $\tilde{k}$. If the increase is large enough, as in Figure 6, with the increase to x', the economy will move to a qualitatively different equilibrium, similar to that in Figure 4, where a fraction of agents chooses to exit after the change. As agents exit this lowers the

capital stock, which reduces the wage and induces more exit until a lower steady state capital is achieved at point A', where $p_t < 1$ and $k_{a'} < \tilde{k}'$.

Figure 6[15]: Transition functions with $\alpha < 0.5$ and the utility value to exiting increases from *x* to x'.



Alternatively, in case (ii), when $\alpha > 0.5$, raising x sufficiently high can result in $A < \bar{A}$ so that the only steady state equilibrium is the degenerate one. This could cause a catastrophic decline in an economy which, for example, was originally in a steady state with positive capital.

Viewing this somewhat differently, if an economy is currently in an equilibrium such as point A in Figure 4, where some fraction of the population chooses not to live out their whole lives, then a change in peoples' beliefs – reducing the value of x – could lead to growth in the medium run and increase per capita income and life expectancy in the long run.[16]

[15] The curve shifts from (a) to (a') with an increase in the exit utility $x' > x$.

[16] Thus, for example, St. Augustine's decision to make suicide a sin, and Thomas Aquinas' decision to

Thus, it is possible in this model to generate a positive relationship between per capita income and life expectancy, in steady states, by considering different values of *x*. In principle, this could be thought of as another interpretation of the Preston curve. However, this equilibrium relationship between per capita income and life expectancy clearly reaches a critical point when $p = 1$ and no further reductions in *x* will have any impact on either life expectancy or per capita income in the steady state. Moreover, it seems unrealistic to think of more advanced economies as being those with more pessimism about the payoff from exit. [17]

4. Extending the Model

This section generalizes the model to consider public health care and to allow agents to choose savings and exit probabilistically. We relate the very similar choices of exiting and not investing in health and show two main results. The first is that individuals would not choose interior probabilities of exit (i.e., the discrete choice analysed above is not restrictive). The second result is that allowing for investment in health can reinforce the exit mechanism and generate multiple equilibria with a lower threshold value of α than in the base model.

We define a biological survival probability $\phi \in [0,1]$ which is realized at the beginning of the second period of life, and where $\phi(h_t)$ is increasing and concave in

---

re-emphasize this, (making it illegal) may have had stimulative effects on the Christian economies of the times.

[17] Alternatively, one could consider the Preston curve to be generated by different values of *A* in the steady state equilibrium of this model. This interpretation is more plausible, perhaps, and has the added benefit that, beyond a threshold value of *A*, further increments increase per capita output without affecting life expectancy (which is at its maximum of 2). Once again, though, this generates a convex Preston curve, up to the point of the maximum *p*. Details are available from the authors upon request.

public health care expenditures h_t : $\phi'(h_t) > 0$, $\phi''(h_t) < 0$, $\phi(0) \geq 0$. This specification is the same as in Chakraborty (2004) except that he restricts $\phi(0) = 0$. The survival probability enters consumer expected utility as follows:

$$U_t = \begin{cases} u(c_{1t}) + [\phi(h_t)\beta u(c_{2t+1}) + (1-\phi(h_t)) \cdot d] & \text{if } I_t = 0 \\ u(c_{1t}) + x & \text{if } I_t = 1 \end{cases}$$

where *d* is the present value associated with exiting life by illness. In general, we can think of the value of *d* as being distinct from *x*. For example, $d < x$ could represent a situation where death from poor health comes through a painful illness that the agent would prefer not to experience. Alternatively, values of $d > x$ might describe situations where wilful exit is seen as sinful, so that death through natural causes is preferred to death by one's own hand. In any event, for wilful exit to be an optimal choice, *d* must be small enough so that $x > \phi(h_t)\beta u(c_{2t+1}) + (1-\phi(h_t))d$.

Notice that, with $\phi(h_t) < 1$, we have a contingency that does not arise in the basic model. An agent may choose to *not* exit, and so choose positive savings, but then die from illness anyway at the end of period 1. To allow for this contingency we assume there is an annuity market.[18]

[18] We get the same general results if agents' savings are discarded. We use an annuity market in order to compare our results with Chakraborty (2004) who also uses an annuity market.

Probabilistic Choices and Equilibrium

Here, we model agents as choosing the probability of exiting. From the perspective of an individual, h_t (and, hence, $\phi(h_t)$) is given exogenously. Let e_t denote the probability of an agent exiting (I_t =1) and 1- e_t is the probability of not exiting (I_t =0). Then expected utility can be written:

$$U_t = u(c_{1,t}) + e_t \cdot x + (1 - e_t)[\phi(h_t)\beta u(c_{2,t+1}) + (1 - \phi(h_t)) \cdot d]$$

Recall, exit is at the end of the period of youth after the savings decision. If an agent exits there are no further decisions. Otherwise, if the agent doesn't exit, the agent trivially spends all his savings in his second period of life: $c_{2t+1} = \hat{R}_{t+1} s_t$, where $\hat{R}_{t+1} = \frac{R_{t+1}}{\phi_t}$ is the gross rate of interest paid by the annuity. Knowing this, agents born in period $t = 0,1,2,...$ choose savings, (s_t), (c_{1t}), and e_t, to maximize utility U_t subject to $c_{1t} + s_t = w_t - h_t$ and $c_{2t+1} = \hat{R}_{t+1} s_t$. The following proposition characterizes the solution of this optimization problem.

***Proposition* 2**. Given h_t , w_t and $\hat{R}_{t+1}$, it is individually optimal to either choose to:

(i) exit and not save, $e_t = 1$ and $s_t = 0$, or

(ii) not exit and save, $e_t = 0$ and $s_t = \frac{\beta(w_t - h_t)}{(1+\beta)}$.

Choosing $e_t \in (0,1)$ is strictly inferior.

Proof. See Appendix 1.

Choosing any probability of $e_t \in (0,1)$ is dominated by certainty, due to the joint decision about savings.[19] [20] This proposition is proven assuming that the exit and savings decisions are made simultaneously. However, it also goes through if the exit decision is either before or after the savings decision. If the probabilistic exit decision is before the end of the period, the analysis is implicitly assuming a mechanism for precommitment. This might corresponds to lifestyle choices made early in the first period.[21]

Public Health Expenditure

We now examine what happens in equilibrium when h_t is determined, at the aggregate level, by the government's budget constraint. As in Chakraborty (2004), health expenditure here is financed by a proportional wage tax so that $h_t = \tau w_t$, where $\tau > 0$ is the wage tax rate. The following specific function nests Chakraborty's example.

$$\phi(h_t) = \begin{cases} \phi^o + \dfrac{\sigma h_t}{1+h_t}, & \forall \ 0 \le h_t < \bar{h} \\ 1, & \forall \ h_t \ge \bar{h} \end{cases}$$

[19] When we allow that $h_t = 0$ yields certain death, $\phi_t(0) = 0$, the results change slightly. Given $h_t = 0$, the optimal choice requires $s_t = 0$ and (i) $e_t = 1$ if $x > d$, (ii) $e_t = 0$ if $x < d$, or (iii) $e_t \in [0,1]$ Still, in all cases, restricting e_t to be discrete is nonbinding on optimizing behaviour. This is equivalent to choosing $I_t = 0$ or 1 non-probabilistically.

[20] This proposition is robust to small errors in picking exit. If an agent chose $e_t = 1$ but knew that there was a small probability that exit wouldn't happen, then they would save a small amount as insurance against the ghastly prospect of zero consumption. Still, they would choose the corner solution.

[21] A simple way to proxy lifestyle choices would be to include the choice of probability, e_t, as increasing first period utility; i.e. $u(c_{1t}, e_t)$, where $u_e > 0$. Such a specification captures the possibility that exit is associate with an enjoyable "carefree" risky lifestyle.

where $\bar{h} = \frac{1-\phi^o}{\sigma-(1-\phi^o)}$, $0 \le \phi^o < 1$ and $\sigma > (1-\phi^o)$. The function $\phi(h_t)$ is strictly concave in up to the maximum value $\phi_t = 1$ corresponding to $h_t = \bar{h}$. The constant ϕ^o is the probability of living when h_t = 0; i.e. $\phi(0) = \phi^o$. Chakraborty restricts $\phi(0) = 0$.

Lemma 2 can now be readily generalized, using an analogous proof.

***Lemma* 5:** For any given configuration of parameters (α, β, A, x, d), and given value of *h*, a threshold value of capital stock $\tilde{k}_h$ exists, and the equilibrium relationship between the proportion of savers p_t (i.e. those that do not exit) and capital per worker k_t satisfies:

$$p_t = p_t^*(h) \quad \Leftrightarrow \quad 0 \le k_t < \tilde{k}(h) \tag{2.13a}$$

$$p_t = 1 \quad \Leftrightarrow \quad k_t \ge \tilde{k}(h), \tag{2.13b}$$

where $p_t^*(h) = \left(\frac{\rho_2(h)}{\rho_1(h)}\right)\left[(1-\alpha)(1-\tau)Ak_t^{\alpha}\right]^{\alpha/1-\alpha}$ is the internal proportion of savers, and the coefficients $\rho_1(h)$ and $\rho_2(h)$ are defined by:

$$\rho_1(h) \equiv \frac{\beta\phi(h)}{1+\beta\phi(h)} > 0 \text{ and } \rho_2(h) \equiv \left[\frac{\alpha\beta A}{e^{\frac{x-(1-\phi(h))d}{\beta\phi(h)}}(1+\beta\phi(h))^{\frac{1+\beta\phi(h)}{\beta\phi(h)}}}\right]^{\frac{1}{1-\alpha}} > 0.$$

Including health expenditure in the model changes its solution in three ways. First, the effective discount rate changes from β to $\beta\phi$; second, disposable income is now

$w_t - h = w_t(1-\tau)$; third, the difference between the two exit utilities, $x-(1-\phi)d$, affects the equilibrium proportion of savers. From Lemma 5 the transition function can be written as:

$$k_{t+1} = \begin{cases} \rho_2(h_t)((1-\alpha)Ak_t^{\alpha})^{\frac{1}{1-\alpha}}, & \forall\ 0 \le k_t < \tilde{k}(\tilde{h}) \quad (2.14a) \\ \rho_1(h_t)(1-\alpha)(1-\tau)Ak_t^{\alpha}, & \forall\ k_t \ge \tilde{k}(\tilde{h}) \quad (2.14b) \end{cases}$$

where $h_t = \tau(1-\alpha)Ak_t^{\alpha}$. Here $\tilde{k}(\tilde{h})$ is implicitly defined by

$$\tilde{k}(\tilde{h}) = \left[\frac{1}{(1-\alpha)(1-\tau)A} \left(\frac{\rho_1(\tilde{h})}{\rho_2(\tilde{h})} \right)^{\frac{1-\alpha}{\alpha}} \right]^{\frac{1}{\alpha}} \text{ where } \tilde{h} = \tau(1-\alpha)A\tilde{k}(\tilde{h})^{\alpha}.$$

The introduction of a concave biological survival function in the analysis yields somewhat different results from Proposition 1. When $\alpha > .5$, we get analogous results to Proposition 1b; i.e. the degenerate equilibrium is stable, and if A is sufficiently large there may exist a stable steady state with positive income and longevity.[22] When $\alpha < .5$, the results depart from Proposition 1a (where there is a unique stable positive steady state) in that it is still possible to have multiple equilibria. Simulations reveal that (2.14a) can be strictly convex when $\alpha < 0.5$ (e.g. $\alpha = 0.35$, $\beta = 0.1$, $\tau = 0.1$, $A = 35$, $x = -0.4$, $d = -1.5$, $\phi^0 = 0$, $\sigma = 1$). Therefore, starting from a low level of the capital stock the economy moves up along a convex locus (2.14a), and when capital stock grows sufficiently, the transition function switches to the concave locus (2.14b). This convex-concave transition curve yields multiple equilibria as illustrated in Figure 6. The fact that the threshold value of α for multiple equilibria is now reduced

[22] The existence of multiple equilibria is perhaps not surprising as Chakraborty (2004) shows that the survival function on its own generates multiple equilibria when $\alpha > .5$. Our result is more general in the sense that we allow $\phi(0) \ge 0$ whereas Chakraborty restricts $\phi(0) = 0$.

indicates that public health and the individual's choice to exit can interact in a way to reinforce each other.

5. Conclusion

This paper examines how an endogenous life expectancy affects a country's economic development within an extended two-period OLG model, which allows agents to choose whether or not to live out the second period of their lives. We first show that given a logarithmic period utility function, there always exists a range of (low) capital stock values where some agents would choose not to live out their whole lives. We second discuss the transition function and the properties of stable equilibria, which are mainly determined by the income capital's share of income. Specifically, with α <0.5, the transition function is concave and the economy has a unique stable steady state equilibrium with a strictly positive value of income per capita, but this unique equilibrium is not identical to Diamond's as at this steady state some agents may still choose to exit. When α >0.5, this model has a convex equilibrium transition function. It introduces the possibility of multiple steady states: the degenerate steady state equilibrium is stable.

Some empirical implications for longevity are obtained as well. The model predicts that average longevity is strictly increasing and concave in wage income as long as capital's share of output is less than one half. If one is willing to interpret international income and life expectancy data as observations along this path, then the model generates a relationship that is comparable to the Preston curve.

We believe that this model sheds some light on the relationship between life expectancy and growth, but further works are needed to take this model more seriously empirically. For example, it is well-known that life expectancy figures are influenced significantly by infant mortality numbers. To take this into account, this model would need to be extended by introducing a third period of life: childhood. Secondly, individuals also face trade-offs between private health expenditure sand human capital investments, which can be also important. Finally, we haven't considered the heterogeneity of the value of *x* across individual agents. All of these extensions are left to future research.

Appendix 1: Proofs of Lemmas 2 and 4 and Propositions 1 and 3

Proof of Lemma 2.

We first establish that, in any time *t,* there exists at least one positive value of k_t, call it $\tilde{k}$, for which the equality in (2.5) holds with $p_t = 1$. Substitution of the firms' first order conditions (2.6) and (2.7) into the right hand side of (2.5) yields the following expression in k_t and k_{t+1}:

$$\beta \ln((1-\alpha)Ak_t^{\alpha}) + \beta \ln(\alpha A k_{t+1}^{\alpha-1}) + \beta \ln \beta - (1+\beta)\ln(1+\beta)$$

Setting $p_t = 1$ in (2.10) and substituting the result into the above expression yields the following function $g(k_t)$:

$$g(k_t) = \beta \ln((1-\alpha)Ak_t^{\alpha}) + \beta \ln(\alpha A((1-\alpha)(\frac{\beta}{1+\beta})Ak_t^{\alpha})^{\alpha-1}) + \beta \ln \beta - (1+\beta)\ln(1+\beta)$$

Differentiating:

$$g'(k_t) = \frac{\beta\alpha^2}{k_t} > 0$$

Thus, since the right hand side of (2.5) is strictly increasing in k_t and the left hand side is constant, there exists a unique value of k_t ($\widetilde{k}$) beyond which further increments in k_t lead to strict inequality in (2.5). Hence, if $k_t > \widetilde{k}$ then $p_t = 1$, as in (2.11b). Similarly, if $p_t = 1$, then, by Lemma 1, the right hand side of (2.5) must be no less than the left hand side and, since $g'(k_t) > 0$, by the definition of $\widetilde{k}$, it must be that $k_t \geq \widetilde{k}$. Hence, if $p_t = 1$ then $k_t \geq \widetilde{k}$, as in (2.11b). The value of $\widetilde{k}$, given in the Lemma, is found by solving the equation $g(\tilde{k}) = x$.

Now consider an equilibrium $k_t \in (0, \tilde{k})$. We first show that p_t cannot equal either 1 or 0 – it must be interior. We then show that an interior p_t implies (2.11a). Suppose $p_t = 1$, i.e. all young agents choose to stay in the model, since $g(k_t)$ is an increasing function of k_t, with $k_t < \widetilde{k}$ and the definition of $\tilde{k}$, the left-hand side of (2.5) is always bigger than the right-hand side, so the optimal strategy is to exit not to stay (a contradiction). Similarly, suppose $p_t = 0$, i.e., all young agents choose to exit the model, from (2.10) this case leads to a zero intensive capital k_{t+1} in period t+1. From the firm's first order condition (2.7), this implies (in the limit) unbounded R_{t+1} and thus the inequality in (2.5) is strict for any $x \in \Re$. Therefore agents would rather stay than exit the model in this case (a contradiction). Thus, for any given $k_t < \widetilde{k}$, in equilibrium, $p_t \in (0,1)$. If $p_t \in (0,1)$, then agents must be indifferent about staying or exiting, hence, (2.5) must hold with equality. Substitution of (2.10) into (2.5), with equality, yields (2.11a).

Finally, if $k_t = 0$, then, from (2.6), $w_t = 0$, and no income is earned in period *t*. Thus, no capital is available for period *t*+1. In this case, due to the term $\beta \ln w_t$, the right

hand side of (2.5) is unbounded from below, and condition (2.5) is violated for any $x \in \Re$. The optimal strategy is to exit: $p_t = 0$. Similarly, if $p_t = 0$ then it must be that $k_t = 0$. Suppose not, so that $p_t = 0$ and $k_t > 0$. Any deviator who saves would receive unbounded returns, so that (2.5) would hold as a strict inequality, for any $x \in \Re$. Hence, the deviator would not exit and $p_t > 0$ (a contradiction). This implies (2.11c). ■

Proof of Proposition 1

Consider the following functions, from the right hand side of equations (2.12a) and (2.12b) respectively:

$$l_a(k_t) = \rho_2 \left((1-\alpha) A\right)^{\frac{1}{1-\alpha}} k_t^{\frac{\alpha}{1-\alpha}}$$

$$l_b(k_t) = \rho_1 (1-\alpha) A k_t^{\alpha}$$

It is straightforward to show that $l_b(k_t)$ lies above $l_a(k_t)$ for all $k_t < \tilde{k}$ and vice versa for all $k_t > \tilde{k}$. Also, $l_a(k_t) = l_b(k_t)$ at two values of k_t: 0 and $\tilde{k}$. The following properties of $l_b(k_t)$ are well known: $l_b(0) = 0$, $l_b'(k_t) > 0$, $l_b''(k_t) < 0$, $\lim_{k_t \to 0} l_b'(k_t) = \infty$, and $\lim_{k_t \to \infty} l_b'(k_t) = 0$. It is also straightforward to show that $l_a(k_t)$ has the following properties: $l_a(0) = 0$, $l_a'(k_t) > 0$. Moreover, if $\alpha < 0.5$ then $l_a''(k_t) < 0$, $\lim_{k_t \to 0} l_a'(k_t) = \infty$, and $\lim_{k_t \to \infty} l_a'(k_t) = 0$. Also, if $\alpha > 0.5$ then $l_a''(k_t) > 0$, $\lim_{k_t \to 0} l_a'(k_t) = 0$, and $\lim_{k_t \to \infty} l_a'(k_t) = \infty$.

To prove part (i) of the Proposition, we first consider the degenerate equilibrium. By (2.12a), we know that the relevant transition function for all $k_t \in [0, \tilde{k})$ is given by

$k_{t+1} = l_a(k_t)$. Since $l_a(0) = 0$, then $k = 0$ is a steady state equilibrium. To prove that this is unstable, consider now a deviation from this equilibrium $k_t = \varepsilon > 0$ and $\varepsilon < \tilde{k}$, then k_{t+1} is given by $l_a(k_t)$. Since, when $\alpha < 0.5$, $\lim_{k_t \to 0} l_a'(k_t) = \infty$ then there exists a small $\varepsilon > 0$ such that $l_a'(\varepsilon) > 1$. Thus, using (2.12a), $k_{t+1} > k_t = \varepsilon$. Hence, the degenerate equilibrium is unstable.

We now consider two cases of non-degenerate steady state equilibria: when $p = 1$, and when $p = p^* < 1$. If $p = 1$ then, by (2.11b) and (2.12b), the steady state equilibrium value k_b is given by $k_b = l_b(k_b)$. Solving this equation, we obtain:

$$k_b = \left(\rho_1 (1-\alpha) A\right)^{\frac{1}{1-\alpha}}$$

It is easy to show that there exists an A such that $k_b > \tilde{k}$. Hence, if k_b is a steady state equilibrium then $p = 1$. To prove that this equilibrium is stable, note that $\lim_{k_t \to 0} l_b'(k_t) = \infty$, and that $l_b(k_t)$ is continuous and strictly concave. Hence, $l_b'(k_b) < 1$. That is, $l_b(k_t)$ crosses the 45^0 line from above. Hence, $k_t \in (0, k_b) \Rightarrow k_{t+1} > k_t$ and $k_t > k_b \Rightarrow k_{t+1} < k_t$. Hence, the steady state equilibrium at k_b is stable.

If $p = p^* < 1$ then, by (2.11a) and (2.12a), the steady state equilibrium value k_a is given by $k_a = l_a(k_a)$. Solving this equation, we obtain:

$$k_a = \left(\rho_2 \left((1-\alpha) A \right)^{\frac{1}{1-\alpha}} \right)^{\frac{1-\alpha}{1-2\alpha}}$$

It is easy to show that there exists an A such that $k_a < \tilde{k}$. Hence, if k_a is a steady state equilibrium then $p = p^* < 1$. To prove that this equilibrium is stable, note that, when

$\alpha < 0.5$, $\lim_{k_t \to 0} l_a^{'}(k_t) = \infty$, and that $l_a(k_t)$ is continuous and strictly concave. Hence, $l_a^{'}(k_a) < 1$. That is, $l_a(k_t)$ crosses the 45^0 line from above. Hence, $k_t \in (0, k_a) \Rightarrow k_{t+1} > k_t$ and $k_t > k_a \Rightarrow k_{t+1} < k_t$. Hence, the steady state equilibrium at k_a is stable.

To prove part (ii) of the Proposition, we first consider the degenerate equilibrium. By (2.12a), we know that the relevant transition function for all $k_t \in [0, \tilde{k})$ is given by $k_{t+1} = l_a(k_t)$. Since $l_a(0) = 0$, then $k = 0$ is a steady state equilibrium. To prove that this is stable, consider now a deviation from this equilibrium $k_t = \varepsilon > 0$ and $\varepsilon < \tilde{k}$, then k_{t+1} is given by $l_a(k_t)$. Since, when $\alpha > 0.5$, $\lim_{k_t \to 0} l_a^{'}(k_t) = 0$ then there exists a small $\varepsilon > 0$ such that $l_a^{'}(\varepsilon) < 1$. Thus, using (2.12a), $k_{t+1} < k_t$, and the degenerate equilibrium is stable.

We now consider two cases of non-degenerate steady state equilibria: when $p = 1$, and when $p = p^* < 1$. If $p = 1$ then, by (2.11b) and (2.12b), the steady state equilibrium value k_b is given by $k_b = l_b(k_b)$. This is entirely equivalent to the steady state equilibrium, analysed above, in case (i) when $\alpha < 0.5$, and $p = 1$. Hence, there exists an A such that $k_b = \left(\rho_1 (1-\alpha) A\right)^{\frac{1}{1-\alpha}}$ is a stable steady state equilibrium with $p = 1$.

If $p = p^* < 1$ then, by (2.11a) and (2.12a), the steady state equilibrium value k_a is given by $k_a = l_a(k_a)$. Solving this equation, as before, we obtain:

$$k_a = \left(\rho_2 \left((1-\alpha) A \right)^{\frac{1}{1-\alpha}} \right)^{\frac{1-\alpha}{1-2\alpha}}$$

The following condition identifies values of A for which $k_a < \tilde{k}$:

$$k_a < \tilde{k} \Leftrightarrow A > \bar{A} = \frac{e^{\frac{x(1-\alpha)}{\beta}} (1+\beta)^{\frac{(1+\beta)(1-\alpha)}{\beta}}}{\beta^{\alpha} \alpha^{1-\alpha} (1-\alpha)^{\alpha}}$$

Hence, if k_a is a steady state equilibrium then $p = p^* < 1$. To prove that this equilibrium is unstable, note that, when $\alpha > 0.5$, $\lim_{k_t \to 0} l_a'(k_t) = 0$, and that $l_a(k_t)$ is continuous and strictly convex. Hence, $l_a'(k_a) > 1$. That is, $l_a(k_t)$ crosses the 45^0 line from below. Hence, $k_t \in (0, k_a) \Rightarrow k_{t+1} < k_t$ and $k_t > k_a \Rightarrow k_{t+1} > k_t$. Hence, the steady state equilibrium at k_a is unstable. Finally, notice that, if $A < \bar{A}$, so that $k_a > \tilde{k}$, then (by (2.12a) and (2.12b)) k_a is not a steady state equilibrium. Hence, in this case, the only steady state equilibrium is the degenerate one. ■

Proof of Lemma 4

Output per worker is $y_t = A k_t^{\alpha}$. The relationship between k_t and k_{t-1} is given by (2.10). Using this relationship and $p_{t-1}^* = \left(\frac{\rho_2}{\rho_1} \right) \left[(1-\alpha) A k_{t-1}^{\alpha} \right]^{\frac{\alpha}{1-\alpha}}$, from (2.11a), yields:

$$y_t = A\left[p^*_{t-1} \frac{\beta}{1+\beta}(1-\alpha)Ak^{\alpha}_{t-1} \right]^{\alpha} = A\left[p^*_{t-1} \frac{\beta}{1+\beta}\left(p^*_{t-1}\right)^{\frac{1-\alpha}{\alpha}}\left(\frac{\rho_1}{\rho_2}\right)^{\frac{1-\alpha}{\alpha}} \right]^{\alpha}$$

$$= A\left(\frac{\beta}{1+\beta}\right)^{\alpha}\left(\frac{\rho_1}{\rho_2}\right)^{1-\alpha} p^*_{t-1} = A\rho_1(\rho_2)^{\alpha-1} p^*_{t-1} = \mu p^*_{t-1}$$

where $\mu = A\rho_1(\rho_2)^{\alpha-1} > 0$.

Now, define $z_t \equiv \frac{y_t}{1+p^*_{t-1}}$ as per capita income. Thus, we have:

$$z_t(p^*_{t-1}) = \frac{\mu p^*_{t-1}}{1+p^*_{t-1}} = \mu p^*_{t-1}\left(1+p^*_{t-1}\right)^{-1}$$

Simplifying notation:

$$z(p) = \mu p\left(1+p\right)^{-1}$$

Totally differentiating:

$$dz = [\mu\left(1+p\right)^{-1} - \mu p\left(1+p\right)^{-2}]dp$$

Hence:

$$\frac{dp}{dz} = \frac{1}{\mu\left(1+p\right)^{-1} - \mu p\left(1+p\right)^{-2}} > 0$$

Hence:

$$\frac{d^2 p}{dz^2} = \frac{d\left(\frac{dp}{dz}\right)}{dp}\cdot\frac{dp}{dz}$$

$$= -\left(\mu(1+p)^{-1} - \mu p(1+p)^{-2}\right)^{-2}\left(-\mu(1+p)^{-2} - \mu(1+p)^{-2} + 2\mu p(1+p)^{-3}\right)\left(\mu(1+p)^{-1} - \mu p(1+p)^{-2}\right)^{-1}$$

$$= -\left(\mu(1+p)^{-1} - \mu p(1+p)^{-2}\right)^{-3}\left(-\mu(1+p)^{-2} - \mu(1+p)^{-2} + 2\mu p(1+p)^{-3}\right) > 0$$

■

Proof of Proposition 2

The Kuhn-Tucker conditions to the household problem corresponding to the maximization problem yield three nontrivial possible solutions, two corner solutions and one interior solution. As before we use $u(c) = \ln(c)$.

Case 0: $e_t = 0$ and $s_t = \dfrac{\beta\phi(h_t)(w_t - h_t)}{1 + \beta\phi(h_t)}$.

Case 1: $e_t = 1$ and $s_t = 0$. If $s_t = 0$, then $c_{2t+1} = 0$ and $U \to -\infty$ unless $e_t = 1$. Conversely, if $I_t = 1$, then $s_t = 0$, maximizes $U = ln(w_t - h_t - s_t) + x$.

Case 2: $e_t = 1 - \dfrac{e^{\frac{x-(1-\phi)d}{\beta\phi}}}{\beta\phi[(w_t - h_t)\hat{R}_{t+1} - e^{\frac{x-(1-\phi)d}{\beta\phi}}]}$ and $s_t = \dfrac{e^{\frac{x-(1-\phi)d}{\beta\phi}}}{\hat{R}_{t+1}}$.

This interior solution is feasible when:

$$0 < \frac{e^{\frac{x-(1-\phi)d}{\beta\phi}}}{\beta\phi[(w_t - h_t)\hat{R}_{t+1} - e^{\frac{x-(1-\phi)d}{\beta\phi}}]} < 1 \qquad \text{and} \qquad 0 < \frac{e^{\frac{x-(1-\phi)d}{\beta\phi}}}{\hat{R}_{t+1}} < w_t - h_t \quad .$$

(2.15)

To find the optimum we compare the utility levels:

$$U_{0t} = \ln[\frac{w_t - h_t}{1 + \beta\phi(h_t)}] + \beta\phi(h_t)\ln[\frac{\beta\phi(h_t)(w_t - h_t)\hat{R}_{t+1}}{1 + \beta\phi(h_t)}] + (1-\phi)d;$$

$$U_{1t} = \ln(w_t - h_t) + x;$$

$$U_{2t} = \ln[(w_t - h_t) - \frac{e^{\frac{x-(1-\phi)d}{\beta\phi}}}{\hat{R}_{t+1}}] + x, \quad \text{if (2.15) is satisfied.}$$

Clearly, when Case 2 is feasible $U_{1t} > U_{2t}$. The interior solution never optimal, and the agent's optimal decision on her longevity reduces down to $e_t = 0$ or $e_t = 1$. ■

Appendix 2: Model Extension with an Arbitrary Depreciation Rate

The body of the paper assumes 100% depreciation. This appendix develops the analysis for a general depreciation rate $\delta \in [0,1]$. Depreciation does not affect our generic results when $\alpha > .5$. With $\alpha < .5$, we were unable to prove Proposition 1 (i) analytically. However, extensive numerical simulations uncovered no exceptions. Lowering δ can change the previous strictly concave relationship between wage and longevity when $\alpha < .5$ described in Lemma 3 to one that is first concave and then convex.

The depreciation rate affects the user cost of capital and enters the first-order condition for capital in the firm's problem yielding real interest rate $R_{t+1} = 1 - \delta + \alpha A k_{t+1}^{\alpha-1}$. The equilibrium condition is unchanged $k_{t+1} = p_t s_t$, and after substituting for savings (2.2) and the wage rate (2.6), the gross interest rate is:

$$R_{t+1} = 1 - \delta + \alpha A \left(p_t \frac{\beta}{1+\beta} (1-\alpha) A k_t^{\alpha} \right)^{\alpha-1}.$$

Recall that (2.5) describes the condition where agents prefer not to exit. Setting $p_t = 1$ and *s*ubstituting R_{t+1} into the inequality (2.5) and taking the antilog yields:

$$(1-\alpha)(1-\delta) A k_t^{\alpha} + (1-\alpha)^{\alpha} \alpha \left(\frac{\beta}{1+\beta} \right)^{\alpha-1} A^{\alpha+1} k_t^{\alpha^2} \geq \frac{e^{\frac{x}{\beta}} (1+\beta)^{\frac{1+\beta}{\beta}}}{\beta}.$$

The left-hand side of this inequality is strictly increasing in k_t. Since the right-hand side is a positive number, there exists a unique threshold value $\tilde{k}$ where the expression holds with equality. The above inequality is therefore satisfied for any

$k_t \geq \tilde{k}$. In another words, when $k_t > \tilde{k}$, $p_t = 1$ and the transition function is given by $k_{t+1} = \frac{\beta}{1+\beta}(1-\alpha)Ak_t^{\alpha}$.

On the other hand if $0 < k_t < \tilde{k}$, the equilibrium will settle down at $p_t \in (0,1)$ and (2.5) holds with strict equality. Substituting the gross interest rate into (2.5) for $p_t < 1$ yields:

$$(1-\alpha)(1-\delta)Ak_t^{\alpha} + (1-\alpha)^{\alpha}\alpha\left(\frac{\beta}{1+\beta}\right)^{\alpha-1} A^{\alpha+1}k_t^{\alpha^2} p_t^{\alpha-1} = \frac{e^{\frac{x}{\beta}}(1+\beta)^{\frac{1+\beta}{\beta}}}{\beta}.$$

This condition gives us the proportion of savers:

$$p_t = \left(\frac{\alpha A(\frac{\beta}{1+\beta})^{\alpha-1}\left((1-\alpha)Ak_t^{\alpha}\right)^{\alpha}}{\frac{e^{\frac{x}{\beta}}(1+\beta)^{\frac{1+\beta}{\beta}}}{\beta} - (1-\delta)(1-\alpha)Ak_t^{\alpha}} \right)^{\frac{1}{1-\alpha}}.$$

The equilibrium transition path of the economy with a general depreciation rate now can be summarized as the following:

$$k_{t+1} = \begin{cases} \left(\dfrac{\alpha A(1-\alpha)Ak_t^{\alpha}}{\frac{e^{\frac{x}{\beta}}(1+\beta)^{\frac{1+\beta}{\beta}}}{\beta} - (1-\delta)(1-\alpha)Ak_t^{\alpha}} \right)^{\frac{1}{1-\alpha}}, & \forall\ k_t < \tilde{k} \qquad \text{(B1a)} \\ \rho_1(1-\alpha)Ak_t^{\alpha}, & \forall k_t \geq \tilde{k} \qquad \text{(B1b)} \end{cases}$$

With $\alpha > 0.5$, the transition function is first described by the convex locus (B1a) and then switches to a concave locus (B1b). For a sufficiently large value of A, we get the

situation depicted in Figure 3. Thus, we get the same generic results as in the body of the paper.

With $\alpha < 0.5$, the curvature of (B1a) can be ambiguous when $\delta < 1$ whereas it is strictly concave when $\delta = 1$. This implies that the economy may have multiple equilibria in contrast to the unique equilibrium when $\delta = 1$. We were unable to prove uniqueness in for $\delta < 1$. However, extensive numerical simulations failed to generate multiple equilibria. Thus, we conjecture that there is a unique equilibrium, in which case, Proposition 1 generalizes to $\delta \in [0,1]$.

The depreciation rate can affect the empirical implications. For a general depreciation rate, the equilibrium relationship between the wage w_t and average longevity of agents $1+p_t$ born at t, is:

$$0 \le w_t < \tilde{w} \;\Rightarrow\; 1+p_t = 1+p_t^* = 1+\left(\frac{\alpha A\left(\frac{\beta}{1+\beta}\right)^{\alpha-1} w_t^{\alpha}}{\frac{e^{\frac{x}{\beta}}(1+\beta)^{\frac{1+\beta}{\beta}}}{\beta} - (1-\delta)w_t}\right)^{\frac{1}{1-\alpha}} < 2$$

$$w_t \ge \tilde{w} \;\Rightarrow\; 1+p_t = 2.$$

Below the threshold wage, $\tilde{w}$, the average longevity is increasing in w_t and is strictly convex for $\alpha > 0.5$. For $\alpha < 0.5$, the relationship is strictly concave when $\delta = 1$, the result in Lemma 3. However, when $\delta < 1$, the relationship may be concave and then convex (for example: $\delta = 0.1$, $\alpha = 0.3$, $\beta = 1$, $x = 0.6$, and $A = 1$).

The equilibrium relationship at time t between the average per capita income, $y_t/(1+p_{t-1})$, and life expectancy, $1+p_{t-1}$, is increasing until the lifespan of $1+p_{t-1}=2$ is reached for all depreciation rates. As described in Lemma 4, this relationship is strictly convex when $\delta = 1$. When $\delta < 1$ this relationship's shape is initially strictly

convex but then its shape appears ambiguous. With a systematic search across the parameter space, we have been unable to find an example where the curve becomes concave, consistent with the usual description of the Preston curve.

CHAPTER 3:

WEALTHIER IS HEALTHIER? RE-EXAMINING THE INCOME-HEALTH RELATION

1. Introduction

Human achievements can be measured from different angles; for example, by focusing on income, equality, freedom, or environmental conditions. Since the first Human Development Report published in 1990, the United Nations Development Programme has annually reported a composite index for each member country's average achievements. This index, the Human Development Index, covers three basic dimensions of human development: health, education, and standard of living. While the concept of human development is much broader than any composite index can measure, many people believe that health and education should be two primary components. A consequential important question for policy-makers is how to improve health and education, especially in developing countries. Researchers (e.g., Flegg 1982, Caldwell 1986, World Bank 1993 and Musgrove 1996) believe that development should concentrate on income growth, since higher incomes, it is argued, leads to better health through a variety of indirect channels (e.g., better nutrition, housing and sanitation). Some other researchers (e.g., Sen 1987, Anand and Ravallion 1993, Bidani and Ravalion 1997 and Jamison et al., 1996) argue that only focusing on income growth is not enough because the value of living is measured intrinsically by

people's "beings" and "doings" – or their "capabilities" to function, rather than income or the possessing of commodities.[23] This chapter contributes to this debate by examining the relationship between income and health using recent aggregated country-level data.

Over the last century the world has experienced amazing improvements in income and health compared to previous centuries. According to recent World Bank data[24], the world's real GDP per capita has increased by 117% over the past 40 years, although there exists significant variation across countries. From 1960 to 2000 real income per capita rose by 159% for the richest country quartile whereas for the poorest country quartile this number was only 51%. Extreme poverty (the share of the population living on less than 1$ a day) in developing countries has fallen by 19% over the last 10 years, especially in East Asia and South Asia, where accelerating growth in China and India puts these regions on track to meet one of the Millennium Development Goals (MDGs)[25] - halve the proportion of people whose income is less than $1 a day by 2015 . In fact, China's growth already helped East Asia reach this target in 2000. On the health side, from 1960 through 2000, average life expectancy has increased by more than 15 years and infant mortality has fallen by 53% around the world. In addition, child mortality by the age of five has also reduced by 54%.[26] Before 1990, rapid improvement in health gave hope to people that the infant and child mortality rates could be cut by two-thirds in the following 25 years, but progress has slowed

[23] The works of Sen (1985, 1987) defined human development as what people can do and be. As Sen (1987 p.25) wrote: "Value of the living standard lies in the living, and not in the possessing of commodities, which has derivative and varying relevance." Underdevelopment was viewed as the lack of certain basic capabilities, rather than lack of income per se. Public policies that directly expand people's capability are more important than increasing the income or equalling the income distribution.

[24] The following income data are available from the World Bank website at www.worldbank.org.

[25] Eight Millennium Development Goals were adopted by the 189 member countries of the United Nations in 2000. A complete list of the goals can be found at www.undp.org/mdg.

[26] Data of life expectancy increase and infant mortality reduction are calculated by myself based on the country-level data from the World Development Index.

down almost everywhere since 1990. Only two regions, Latin America and the Caribbean, and Central Asia, are close to achieving this target on average. But even in these regions, more than half the countries are off track. The gap between the health goal and reality is greatest in Sub-Saharan Africa, where civil disturbances and the HIV/AIDS epidemic have obstructed the reduction of infant and child mortality. According to recent data, only 35 countries around the world are making enough progress to achieve the health target of the Millennium Development Goals: reducing the infant and child mortality rates to one-third of their 1990 level by 2015.[27]

Such patterns in the improvement of health over the past 40 years raise a meaningful question: what are the drivers of health improvement? They could be food, shelter, investments in sanitation facilities, access to health services, health-related research, or desirable individual behaviour. Most of the empirical studies (e.g., Filmer and Pritchett 1999, Musgrove 1996 and Thomas, Strauss, and Henriques 1990) agree or even take it for granted that health improvements are related to income: a higher income certainly implies larger real consumption of foods or goods that reduce malnutrition, cut mortality, and prolong life expectancy; countries with higher income may free resources for their medical and public health services, which could broadly change the health situation. I explore for the presence of this assumed relationship. To ascertain whether this relationship between income and health exists, cross-section and panel data are used.[28] Most of the studies, summarized in this section, use country-level data from 1960 to 1990 and conclude that the income elasticity of health is statistically significant. However, as health progress has slowed down since 1990,

[27] Data are cited from the World Bank. See footnote 24.

[28] This chapter focuses on the relationship at the macro-level as (i) the data on aggregated health measures across different countries are more readily available; (ii) we can examine the impact of income on health as opposed to nutrition intake or other indirect measures of health; (iii) it also allows us to study what role public health expenditure and poverty plays.

while income still maintains an upward trend, today's income-health relationship may be different from the one that existed before 1990. We believe it is worthwhile to use more recent data and re-examine the income-health association.

Discussions on income, health and public services can be traced back to biblical times (Ezekiel 34: 1-4), but an extensive literature on the income-health relationship using cross-section data appeared after 1960, the year the World Bank began to publish world-wide country-level data. One of the benchmark income-health studies is Preston (1975), who compares different countries' life expectancies and incomes per capita in 1900, 1930, and 1960, and proposes the "Preston curve", a non-linear and concave empirical relationship between income per capita and life expectancy.[29] A 2000 version of the Preston curve with 72 countries is plotted in Figure 7. This income-life expectancy relation is significantly positive initially and dies down after income reaches a relatively high level. Compared to the situation in 1960 with 46 countries, the Preston curve in 2000 shifts up and there are more countries that are on the flat portion of the Preston curve (i.e., the flat tail of this curve is longer), which indicates that the income-health correlation observed currently may be different from the one in 1960. The story is quite similar when using infant mortality rate to plot the income-health relation (Figure 8) except that the relationship is negative initially.

[29] Preston's data covered 10 countries in 1900, 38 countries in 1930, and 57 countries in 1960.

Figure 7: GDP per capita vs. life expectancy in 1960 and 2000

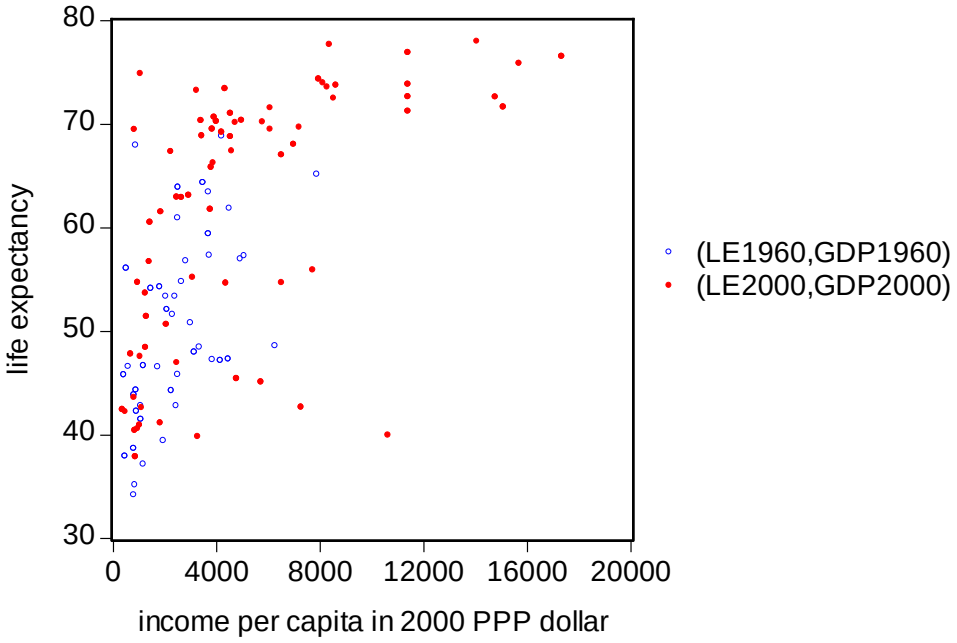
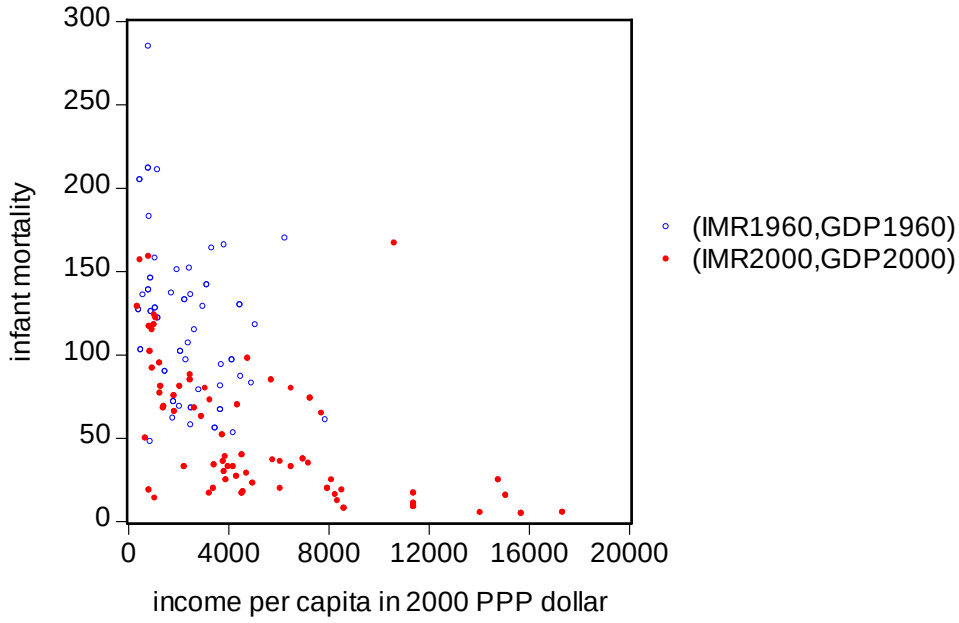


Figure 8: GDP per capita vs. infant mortality in 1970 and 2000



The concave Preston curve also becomes a rationale for the model specifications in other empirical research. After Preston (1975), a log-linear regression was widely

applied to study the income-health association since this model specification captures the non-linear property of the Preston curve and conveniently results in a constant elasticity. Flegg (1982) studies 46 underdeveloped countries, each country providing one year's observation on income and infant mortality over the period 1968-1972, and shows that the income elasticity of the infant mortality is -0.29. Subbarao and Raney (1995) test this income effect on a larger sample with 72 developing countries in 1985 and obtain a similar number of -0.25. Using more recent data of 117 countries, Zakir and Wunnava (1999) conclude that the elasticity should be smaller (-0.14) if the heteroskedasticity is adjusted with Generalized Least Squares (GLS). All these income effects are statistically significant.

In recent years, given the availability of panel data with longer time dimension and further developments in panel data econometrics, scholars are turning to studying the income-health relationship within a panel framework. For example, the statistically significant income elasticity of child mortality is -0.5 in Latin American countries (Jamison et al. 1996) and is -0.16 for sub-Saharan nations (Cornia and Mwabu 1997); using fixed-effect estimations, Pritchett (1997) reports an income elasticity of infant mortality of -0.59 for 22 developing countries, and a similar number of -0.5 is found from a larger panel sample of 106 countries in Issa and Ouattara (2005). Except for the work of Cornia and Mwabu (1997), estimates of the income elasticity are higher when based on a panel rather than on a single year of cross-sectional data.

Besides these researches, studies that use non-linear model specifications other than a log-linear model also provide evidence that income can statistically significantly reduce the mortality rate: Rodgers (1979) explains 54 percent of the variation in infant mortality of 56 countries by the reciprocal of income per head and the Gini index, a

measure of income inequality; a semi-log specification is considered by Boehmer and Williamson (1996), who report a negative income impact on infant mortality for 96 less developed countries.

No matter whether we use a simple cross-section regression or a more complex panel regression, one difficulty in obtaining consistent estimators is how to remove the reverse causality or endogeneity running from the explanatory variables to the dependent variables. In a regression where income explains health, causality running from the dependent variable (health) to the explanatory variable (income) may exist. Especially for countries on the flat portion of the Preston curve, health performance is already so good that it is difficult for income to improve it further. Then policy markers and economists are more likely to observe impacts coming from health to income. Some microeconomic studies, including Thomas and Strauss (1997), Strauss and Thomas (1998) and Miguel and Kremer (2004), carefully demonstrate how better health status can raise education level, adult labour quantity, and labour productivity; all of these can influence an individual's income. From a macroeconomic perspective, health status (such as life expectancy) is explained to be a cohort's "patience" or time discount rate; those who have poor health are usually impatient and are reluctant to invest in human capital/schooling when they are young, or save capital for their old age. As human capital and savings are two main factors for a country's long-term economic growth, this poor health reduces the hope for income growth (see, e.g., Croix and Licandro 1999, Kalemli-Ozcan, Ryder, and Weil 2000, Zhang, Zhang and Lee 2003, Blackburn and Cioriani 2002 and Chakraborty 2004). Hence, to study the income effect on health, we have to account for possible reverse causality.

Pritchett and Summers (PS hereafter) (1996) seminal paper addresses this problem. The authors control for reverse causality by using two-stage least squares estimators, with terms of trade shocks (changes in the relative price of a country's exports to imports), the investment ratio (the ratio of investment to GDP), black market premium (the difference between the foreign exchange rate in the black market and the official exchange rate), and price distortion (the deviation of the official exchange rate from its purchasing power parity level) being chosen as instruments for income growth. Based on an unbalanced panel dataset with 58 developing countries (each having at most five years' observations over the period 1960 through 1985), they report that the income elasticity of infant mortality ranges between –0.98 to –0.28. Without controlling for the potential endogeneity of income, this elasticity is found to be –0.19.

This understatement of the income effect brings to question the results of some studies (e.g., Anand and Ravallion 1993) that claim that income is less important than other variables such as public/private health expenditure. The importance of income policies on health improvement can be masked by the reverse causality that is ignored in the estimation. Similar work is undertaken by Easterly (1999), who uses black market premium, financial depth, and inflation as instrumental variables (IV). With cross country data for 1960, 1970, 1980, and 1990, Easterly's findings suggest that a one percent increase in income could reduce infant mortality and child mortality by 0.58 percent and 0.78 percent, respectively.[30]

[30] Easterly (1999) only examines one explanatory variable - income.

In reality, appropriate external instrumental variables that can control for the endogeneity of income are rare.[31] However, with panel models we can use internal instrumental variables instead. One contribution of this paper is to extend PS's model to a dynamic setup and to use a system Generalized Methods of Moments (GMM) estimation method with internal instruments to allow for the possible endogeneity of income. Results suggest that internal instruments can control for the endogeneity of income and the resulting estimated income effect in a dynamic model is much smaller than those of PS.

Most of the literature reviewed above uses country-level health data from standard sources such as the World Bank and the World Health Organization. Kakwani (1993) argues that using such data without any transformation did not capture the non-linear relationship between achievements of health and values of the health indicator. For example, an increase in life expectancy from 40 to 50 should be judged to be a smaller achievement than an increase from 70 to 75 since it is more difficult to increase life expectancy for older people. He proposes a new health achievement index that adopts a logistic transformation of the raw health data. Using cross-section regressions with 80 developing countries' data over 1970s and 1980s, his findings suggest the presence of a statistically significant relationship between (log) income and his health indices. A similar logistic transformation is applied by Anand and Ravallion (1993) for a sample of 22 developing countries. PS argue that using raw health data and health achievement indices should not result in substantially different results, but they did not undertake any empirical work to support this claim. In this chapter we add to this debate on the specification of the health data by reporting

[31] A valid instrument for GDP per capita is hard to find, but the macroeconomic growth literature provides various instruments for income growth (see, e.g., Easterly et al. 1993, Fischer 1993, Levine and Renelt, 1992). PS in their paper used such instruments.

results based on Anand and Ravallion's health achievement index along with those using the raw health data.

Our discussion so far only considers the impact of income on health. Other factors are also important. Preston (1975 p250), for example, finds that income only explains 20 percent of the growth of life expectancy from 1930 to 1960 with the other 80 percentage being attributed to changes from other exogenous variables. Although Preston claims that this finding does not refute the existence of an income-health relationship, health improvements have become increasingly dissociated from income changes because of a diffusion of health technologies, facilities, and personnel that occur independently of economic level. For instance, access to medical services is more economically and statistically significant than is GDP per capita (Flegg 1982); support is furthered by Anand and Ravallion's (1993) research, where income matters for health only through its effect on public health expenditure per capita and the share of the population in poverty.

However, not all researchers agree with this view of the ineffectiveness of income. McKeown et al. (1975) make another point when they examine causes of death over the second half of the nineteenth century. They show that the reduction in mortality, achieved since 1850 in western countries, is primarily a by-product of improved standards of living while improved medical treatments have played only a minor role; see also Filmer and Pritchett (1999), Zakir and Wunnava (1999) and Issa and Ouattara (2005). Since no consistent conclusion on the importance of other variables can be drawn from this body of work, we examine the role of public health expenditure in our models.

According to the above review, the existing literature, using data from before 1990, reports an income elasticity of infant mortality in the order of -0.14 to -0.59 when using OLS estimation. Correction for endogeneity of income raises the estimates of this elasticity to an interval of -0.28 and -0.98. These estimates are further complicated by the question of the functional form of the health index and the debate about the role of public health expenditure. In the following sections, Section 2 replicates PS and compares our OLS and IV estimates of the infant mortality function and the life expectancy function with those reported by PS; Section 3 provides sensitivity tests, including updating with recent country-level data, using other health indices, and introducing additional explanatory variables such as public health expenditure and poverty; Section 4 reports results when estimation uses internal IV instruments with GMM and system GMM methods to control for the endogeneity of income; Section 5 concludes.

2. Replication of Pritchett and Summers' (1996) Work

The first objective of this chapter is to re-examine the income-health relationship using recent data and compare the results to those derived using earlier data. Based on our discussion in the previous section, we have reasons to believe that the income-health relationship has changed over time. We choose Pritchett and Summers (1996) as a benchmark since this is one of the few papers in this literature that accounts for reverse causality running from health to income. It is also often cited as an important paper as it adjusts for cross-country heteroskedasticity in the panel data. We first replicate their work in ways as close to the original paper as possible.

As we will show later, we cannot exactly replicate their results. Here we discuss some of the reasons that might be causing these discrepancies. The first reason may be related to the health data. PS excluded countries that did not have reliable health data, based on a working paper by King and Rosenzweig (1991), and ended up with only 64 developing countries with data from some of the following years: 1965, 1970, 1975, 1980, and 1985. As we could not obtain a copy of King and Rosenzweig (1991), despite our best efforts, we do not know exactly which countries formed PS's sample. We obtain health data (life expectancy and infant mortality rate) on 105 developing countries from the World Development Indicators (WDI) database.[32] However, infant mortality data in 1965 and 1975 are not available from the WDI, so we use simple averages of the data from 1962 and 1967, and 1972 and 1977, respectively, to estimate these two years' observations. For example, 0.5*(infant mortality in 1962 + infant mortality in 1967) = infant mortality in 1965.[33]

The second reason could be the data used to construct the instrumental variables; PS do not explicitly provide data sources or details on how they constructed these variables. We use data from the World Bank's Global Development Network (GDN). Consistent with the macroeconomic growth literature (e.g., Easterly et al. 1993; Fischer 1993), we define the terms of trade shock as the difference in the logarithm of

[32] There are 105 developing countries in our sample: Afghanistan, Algeria, Angola, Argentina, Bangladesh, Benin, Bolivia, Botswana, Brazil, Burkina Faso, Burundi, Cameroon, Cape Verdi, Central African Rep, Chad, Chile, China, Columbia, Comoros, Congo Dem. Rep., Costa Rica, Cyprus, Dominica, Dominica Rep, Ecuador, Egypt, El Salvador, Ethiopia, Fiji, Gabon, Gambia, Ghana, Greece, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hungary, India, Indonesia, Iran, Iraq, Ivory Coast, Jamaica, Jordan, Kenya, Korea Republic of, Lesotho, Liberia, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mauritius, Mexico, Morocco, Mozambique, Myanmar, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Panama, P. New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Rwanda, Senegal, Seychelles, Sierra Leone, Solomon Islands, South Africa, Sri Lanka, St. Lucia, St. Vincent, Sudan, Suriname, Swaziland, Syria, Taiwan, Tanzania, Thailand, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Uganda, Uruguay, Vanuatu, Venezuela, Western Samoa, Yemen, Yugoslavia, Zaire, Zambia, and Zimbabwe.

[33] Another data source of life expectancy and infant mortality is the Global Development Network (GDN) database, which provides annual health data including for 1965 and 1975. We find that our results are not sensitive to whether we use this GDN data or our WDI averaged data.

terms of trade, and calculate the black market premium as the logarithm of (1+black market premium/100). Perhaps PS's definitions are different from ours; we are unable to judge the discrepancies. In the following section, we provide a detailed discussion of the definitions of the variables and data sources used in this paper, and report our empirical results.

2.1 Variables and Data Sources

There are three commonly used aggregate health measures: infant mortality, life expectancy at birth, and child mortality.[34] Child mortality rate, which is the number of children who die before reaching age five out of 1000 live births, is highly correlated with the infant mortality rate. Some papers mentioned in the previous section find their results for child mortality to be close to those when using infant mortality (e.g., Cornia and Mwabu 1997, Easterly 1999, Filmer and Pritchett 1999, Hanmer et al. 2003), or at least income effects on child mortality and infant mortality have consistent signs. Hence, we focus on reporting results using infant mortality. The second health variable we examine is life expectancy at birth. As we show, results for life expectancy at birth are more tenuous than those for the infant mortality. A few possible reasons for this are discussed in Section 2.3. Infant mortality and life expectancy data are taken from the WDI at five-year frequencies from 1960 to 1985 (but infant mortality in 1965 and 1975 are adjusted by using simple averages as previously mentioned).

[34] Based on the WDI online dataset, life expectancy is defined as the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life; infant mortality is defined as the number of infant deaths per 1000 live births with infants being one year of age or younger.

The income variable is taken to be GDP per capita, following most of the studies in the literature. It comprises the value of final products produced over a period, and covers a wide range of products that can be expected to influence health. At the same time, as the leading index of level of economic development, income per capita is the focus of growth models from which policy measures are derived. Like PS, we use RGDPCH, annual GDP per capita at 1985 real purchasing power (PPP) adjusted dollars, from Summers and Heston's Penn World Tables 5 (PWT 5).[35] To enable us to focus on developing countries, we follow PS by using I$6000 as a cut off point. This excludes 33 countries that have GDP per capita in 1985 higher than this number.

We also include education as another control variable in our models. Studies have shown that education, especially maternal education and the initial education level at the beginning of the study period (e.g., Flegg 1982, Subbarao and Raney 1995, Cornia and Mwabu 1997, Filmer and Pritchett 1999, Zakir and Wunnava 1999) have important effects on health. Indeed, some of this research suggests that education is more important than income in reducing infant mortality (e.g., Subbarao and Raney 1995). Following PS, we define education as the average years of schooling of the total population over age 25, taken from Barro and Lee (1992). The data are available at five-year frequencies from 1960 to 1985.[36]

There are four external instrumental variables in PS: terms of trade shock, investment rate, black market premium, and real exchange rate distortion. Annual data for these variables are available from the GDN database, but we average these annual data over

[35] This variable is measured in international dollars (I$), which adjusts for price level differences across countries.
[36] In some literatures, education is measured as average years of schooling of the total population over the age of 15. However, to be consistent with PS, we use schooling of population over the age of 25.

five-year intervals from 1960 to 1985 to explain income growth during those intervals, which is the approach primarily used in the growth literature.

2.2 Models

Consider the basic model:

$$\log y_{it} = \alpha + \beta \log x_{it} + \delta_t + \varepsilon_{it} . \tag{3.1}$$

where y_{it} is the health variable - infant mortality rate or life expectancy at birth, x_{it} is income per capita, and δ_t is a time-specific effect. PS extends this basic model to control for education:

$$\log y_{it} = \alpha + \beta \log x_{it} + \gamma \log z_{it} + \delta_t + \varepsilon_{it} , \tag{3.2}$$

where z_{it} is the education attainment from Barro and Lee's database. This is a simple pooled model as it ignores any heterogeneity existing across individual sectors, i.e. coefficients α and β are set to be identical across the countries. It further assumes $\varepsilon_{it} \sim N(0, \sigma^2)$, $E(\varepsilon_{it} \mid x_{it}, z_{it}) = 0$, and $E(\varepsilon_{it}\varepsilon_{js} \mid x_{it}, z_{it}) = 0$, if $t \neq s$ or $i \neq j$.

These error term assumptions will not hold if the data generation process in each country is different. Accordingly, the second model in PS controls for time-invariant effects across individual countries by considering the model:

$$\log y_{it} = \alpha_i + \beta \log x_{it} + \delta_t + \varepsilon_{it} , \tag{3.3}$$

where α_i represents country specific effects such as culture, geography, politics, etc. Taking the first-order difference prior to estimation removes the α_i terms:

$$\log y_{it} - \log y_{it-1} = \beta(\log x_{it} - \log x_{it-1}) + \delta_t^{'} + (\varepsilon_{it} - \varepsilon_{it-1}) . \tag{3.4}$$

Here $\delta_t^{'}$ is the time effect after the transformation. First-order differencing is not the

only way to handle country specific effects in model (3.3); a within-transformation is another popular approach. Using a within-transformation, model (3.3) is:

$$\log y_{it} - \overline{\log y_i} = \beta(\log x_{it} - \overline{\log x_i}) + \delta_t^{''} + (\varepsilon_{it} - \overline{\varepsilon_i}), \qquad (3.5)$$

where $\overline{y_i} = (\sum_{t=1}^{T} y_{it}) / T$ with $\overline{x_i}$ and $\overline{\varepsilon_i}$ similarly defined. There are two advantages to first-order differencing. First, the instrumental variables used to control for possible causality running from health to income are classical instruments for income growth, the right-hand-side variable of model (3.4), but they may not be valid to instrument the right-hand-side of (3.5). Secondly, we are able to address the possible endogeneity of income by applying internal instrumental variables in model (3.4). For example, the second lag of the control variable, x_{it-2}, is an appropriate internal instrumental variable as it may be closely related to x_{it} but does not relate to the error term in (3.4), given that x_{it} is contemporaneously correlated with y_{it}. However, x_{it} is correlated with the error term in (3.5) so it is not a valid instrumental variable in model (3.5). Accordingly, we proceed via this route.

To control for the impact of education, PS added initial education level into the differenced model:

$$\log y_{it} - \log y_{it-1} = \beta(\log x_{it} - \log x_{it-1}) + \gamma \log z_{it-1} + \delta_t^{'} + (\varepsilon_{it} - \varepsilon_{it-1}).^{37,38} \qquad (3.6)$$

We estimate models (3.1), (3.2), (3.4) and (3.6) using OLS and IV estimation techniques, using data from 1960 to 1985 in five-year frequencies.

[37] We find model (3.6) to be somewhat problematic. If country effect and education simultaneously exist in the health function: $\log y_{it} = \alpha_i + \beta \log x_{it} + \gamma \log z_{it} + \delta_t + \varepsilon_{it}$, then the differenced model should be $\log y_{it} - \log y_{it-1} = \beta(\log x_{it} - \log x_{it-1}) + \gamma(\log z_{it} - \log z_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1})$. However, PS note that they add the "level" of education, and in the IV estimation they use the instrument "initial education", which leads us to believe that model (3.6) is consistent with their estimated model. (See Pritchett and Summers (1996) Table 2 on page 850 and the second last paragraph on page 852.)

[38] We also estimate the model with current year's education (log) level rather than the lag of the education (log). Results from these two cases are very similar.

2.3 Results and Discussion

OLS for Infant Mortality

Table 1 reports results from estimating models (3.1) and (3.2) using least squares, with columns 2 and 3 providing our estimates whereas columns 4 and 5 present those reported by PS. Our results qualitatively accord with PS: our income term has a statistically significant negative effect on infant mortality but with a smaller magnitude than found by PS. Without controlling for education, a one percent increase in income per capita decreases infant mortality by 0.47 percent. Compared to the results of PS, this smaller income effect is perhaps not surprising because we used all available health data, with no attempt at distinguishing "reliable" from those with "bad" quality. This can result in a smaller income effect in two ways. First, there are more countries in our panel dataset when we use all available data. As OLS can be regarded as a pooled estimator that does not consider heterogeneity across sectors, the introduction of additional countries likely obscures the averaged income-health association. Secondly, PS argue that the extrapolation procedure used in projecting those bad quality health data induces a systematic bias, although theoretically the pure measurement error coming from the dependent variable should not induce biases in the income effect estimates.[39]

[39] For countries that do not have complete vital registration, the World Bank uses a procedure to extrapolate up-to-date estimates of infant mortality from their past estimates/records. This procedure builds in a non-linear downward trend: countries with higher initial mortality are assumed to have larger reductions over time while countries with lower initial mortality are assumed to have smaller reductions over time. This procedure attributes declines in infant mortality for countries that actually do not experience such large reductions and will cause downward bias in the income coefficient. (See Pritchett and Summer, 1996, footnote 15, p 850.)

Table 1 : The effect of per capita income on infant mortality, 1960-1985, simple OLS

Column #	1	2	3	4
	Our estimates – levels		PS estimates – levels	
	without edu	with edu	without edu	with edu
GDP per capita	-0.47**	-0.27**	-0.71**	-0.42**
	(15.45)	(8.38)	(16.71)	(8.13)
Education(level)		-0.29**		-0.136**
		(10.40)		(8.02)
1965	8.11**	6.77**	9.66	8.03
1970	8.06**	6.74**	9.70	8.07
1975	7.97**	6.68**	9.67	8.03
1980	7.85**	6.61**	9.69	8.05
1985	7.63**	6.44**	9.45	7.89
R^2	0.55	0.67	0.63	0.73
# Countries	66	66	58	58
# Observations	294	294	184	184

Note: Absolute t-statistics are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance respectively.

We find that the education variable is statistically significant, with the introduction of education resulting in income per capita still negatively affecting infant mortality but with a smaller impact. From column 3, education elasticity on infant mortality is –0.3; income elasticity reduces from –0.47 to –0.27 but it is still statistically significant. (In PS the income elasticity reduces from –0.71 to –0.42.) This is expected; there is a large literature that shows that income and education are positively related and hence leaving out education inflates the estimated coefficient on income. At the same time, all time effects are statistically significant, generally suggesting a slow decline in infant mortality over the sample period.

To control for potential country effects, we report OLS results of models (3.4) and

(3.6) in Table 2. Similar to the results in Table 1, including education in the model reduces the income effect on infant mortality from –0.10 to –0.08 but it is still statistically significant; also, our income elasticity is still smaller than that reported by PS. As with the findings of PS, we find that considering country effects reduces the income elasticity and the impact of education on infant mortality. For example, the coefficients of income and education in the differenced model are only one third and one tenth of those in the level model. These changes suggest that excluding country specific fixed effects leads to overestimating the impact of income and education on infant mortality.

Table 2 also reports heteroskedasticity consistent *t*-statistics (presented in angeled brackets) using White's (1980) cross-section method as provided by EViews.[40] The statistical significance of the income/education terms are not changed by the heteroskedasticity correction. Another approach to take care of the country effects is via a within-transformation (fixed effect estimation), as detailed in model (3.5). The second last column of Table 2 reports these results, from which we see that the income elasticity from the fixed effect estimation is much higher than the one from the first differencing model and education is not statistically significant. These findings were shown in PS as well. One possible explanation is that in the first-differenced estimation the assumption of uncorrelated errors is violated leading to a substantial loss in precision.

[40] In EViews, the White cross-section method is robust to contemporaneous correlation as well as different error variances in each cross-section.

Table 2: The effect of per capita income on infant mortality, 1960-1985

Column #	1	2	3	4	5	6
	Our five-year		PS' five-year differencing		Our fixed	PS' fixed
	without	with edu	without edu	with edu	With edu	with edu
GDP per capita	-0.10**	-0.08**	-0.24**	-0.19**	-0.29**	-0.313**
	(3.54)	(2.82) <2.87>	(5.03)	(4.34) <4.28>	(7.98)	(6.51)
Education(level)		-0.027**		-0.019**	0.033	-0.010
		(5.09) <3.15>		(6.22) <5.47>	(1.05)	(0.51)
1965	-0.074**	-0.069**	-0.097	-0.030	-0.067**	-0.09
1970	-0.086**	-0.080**	-0.078	-0.016	-0.131**	-0.17
1975	-0.117**	-0.109**	-0.124	-0.048	-0.239**	-0.28
1980	-0.159**	-0.145**	-0.158	-0.074	-0.397**	-0.42
1985	-0.169**	-0.143**	-0.192	-0.091	-0.568**	-0.61
R^2	0.176	0.24	0.18	0.33	0.96[41]	0.80
# Countries	66	66	58	58	75	64
# Observations	294	294	184	184	389	248

Note: angled brackets indicate heteroskedasticity consistent *t*-statistics. Asterisks ** and * represent 5% and 10% statistical significance respectively.

IV Estimation for Infant Mortality

None of the above methods - simple OLS, first-differencing or fixed effect estimation - takes into account the potential reverse causality from health to income. If income is endogenous, then these estimators are biased and inconsistent. In this section, we address this problem by using instrumental variables to estimate a "pure" income effect, which is not mixed with any health influence running from the other side. Following PS, we use terms of trade shock, investment rate, black market premium,

[41] The R squares in the first-order differenced model are much lower than those in the fixed effect model because the definition of R-square being used is not the same for both types of models. In the first-differenced model the R square is the goodness of fit to explain the difference of health by the difference of income, while in the fixed-effect model it is the goodness of fit to explain the level of health by the level of income and a country-effect. Usually the country-effect term in the fixed-effect model adds substantial explanatory power.

and real exchange rate distortion as instrumental variables for income growth. We apply Davidson and MacKinnon's (1989, 1993) version of the Hausman (1978) test and the Sargan (1988) test to ascertain the merit of moving to instrumental variables estimation. The Hausman test examines for the equivalence of the IV and OLS estimators. Under the null hypothesis, both estimators are consistent but OLS is more efficient. In our situation this null hypothesis implies that income is not an endogenous variable. Otherwise, under the alternative hypothesis, (income is endogenous) only the IV estimator is consistent. Hausman's approach tests the difference between the IV and OLS estimates, rejecting the null hypothesis if this difference is statistically significantly different from zero. The Sargan test is a test of the validity of the instrumental variables. The null hypothesis being tested is that the IVs are uncorrelated with the residuals. It should be noted that this test performs poorly in models with serially correlated errors. Statistics of the Hausman test and the Sargan test are both asymptotically Chi-square distributed under their appropriate null hypotheses with the degrees of freedom being given by J and $Z–J$, where J is the number of problematic explanatory variables and Z is the number of instrumental variables.

Results for estimating model (3.6) are provided in Table 3, where the country effect is controlled for by taking first-order differences. Using, in turn, the investment ratio, black market premium, and real exchange rate distortion as instrumental variables, we find that the income elasticity of infant mortality increases from –0.08 to –0.40, -0.20, -0.43, respectively. PS also report higher income effects with the IV estimates; for example, the income elasticity changes from –0.19 to –0.35 when the investment ratio is used as the instrumental variable. If a positive reverse causality from health to income exists, that is, higher income improves health but better health also increases

income, ignoring this reverse causality (OLS estimation) will overestimate the income elasticity on the infant mortality. IV estimation, on the other hand, should remove this upward-bias by driving down the income effect. However, both PS' and our IV estimations give consistently larger income elasticity estimates. This is likely because the IV estimation not only corrects for the reverse causality but also reduces the bias and inconsistency from the presence of measurement errors; PS suggest that the measurement error embedded in the five-year differences in health, income, education and IVs data is more serious than the reverse causality problem. As our sample includes those countries that do not have good-quality data, we are likely observing a dominance of the measurement error effects over the reverse causality effects.

The first-stage R^2 values of these four IV estimations are only between 18 percent and 23 percent, suggesting that the instruments are weak in explaining the five-year growth of income. This may lead us to have little faith in the IV results. Although PS (p854) argue that these low first-stage R^2 values might be suggesting that the instruments are exogenous and orthogonal to the error term in model (3.6), we are unconvinced: it is our belief that a low first stage R^2 only suggests that the IV is not closely related to the problematic explanatory variable – it does not imply that the IV is exogenous with the error term. Whether the IV is exogenous should be formally tested.

However, we agree with PS that it is inappropriate to reject the IV approach based on these low first-stage R^2 values. The highest correlation among these four instrumental variables is only 0.27, suggesting that each instrument brings independent information, so that it may be worthwhile to use them all as an instrument set. We provide this in column 6 of Table 3. Then the income elasticity is close to the OLS estimate (-0.116).

From this finding it seems that the OLS estimator does not seriously bias the income impact. But it is also possible that this lower income elasticity from the IV estimation is caused by the problematic instrumental variable, terms of trade: when this instrument is applied, the income impact on infant mortality is statistically insignificant. We report estimates using only investment ratio, black market premium, and real exchange rate distortion as instruments in column 7 of Table 3. In this case the income elasticity is larger and around three times that from the OLS regression. Therefore, we believe that OLS leads to underestimating the income elasticity.

From these comparisons, the IV estimates suggest that a robust income-health relationship exists but that the magnitude of the income effect may be understated if we do not allow for reverse causality from health to income. This finding is robust when we use heteroskedasticity-consistent standard errors in calculating our *t*-statistics. However, the IV results are sensitive, to some extent, across the choice of instruments: using investment ratio and real exchange distortion as the instruments results in an income elasticity around -0.4; when the instrument is black market premium, the income elasticity is only about -0.2; if all the instrumental variables are applied together the income impact is substantially reduced, but when the terms of trade is removed the effect goes up again. These sensitivity results suggest that we should further test the validity of the instrumental variables, and consider other approaches to control the endogeneity of income, for example, using internal instruments.

The effect of education on infant mortality remains negative throughout in Table 3, but its magnitude, similar to the income elasticity, is different across the choices of instrumental variables: in the OLS regression the coefficient of education is –0.027

while in IV estimation this coefficient is between –0.017 and –0.056, depending on the choices of IVs.

We apply two statistical tests to examine the appropriateness of our use of instrumental variables: the Hausman test and the Sargan test. At a 10% significance level, the Hausman test rejects the (asymptotic) equivalence of the OLS and IV estimates when investment ratio, black market premium, and real exchange distortion are used as individual instruments, but the null hypothesis is not rejected when multiple instruments are applied together. In other words, while the Hausman test results using single instrumental variables suggest that there is an endogeneity problem with income, the result from using multiple instruments together indicates otherwise. Hence, we cannot come to a consistent conclusion about the endogeneity problem based on the instruments that we have chosen so far. This problem again highlights the need to consider other approaches, such as the use of internal instruments.

The Sargan test examines for the orthogonality of the over-identifying instruments. Results are reported in the final two columns of Table 3 when multiple instruments are applied, because in cases with a single instrument there are no over-identifying restrictions. PS report that the Sargan test rejects the orthogonality of the instrument set at the 5 percent level when the black market premium is included, while it is far from rejecting when the black market premium is dropped from the instrument set. Neither of our Sargan tests rejects the hypothesis that the instrument sets are orthogonal to the error term. This suggests the instruments are valid to control for the endogeneity of income. Although a major drawback of this test of over-identifying restrictions is low power in finite samples with weak instruments (Blundell and Bond

2000), the high *p*-values of our Sargan tests still lend some credibility to the validity of our instrument sets (Greene 2002, page 893).

Table 3: The effect of per capita income on infant mortality, IV estimation, 1960-1985, five-year differencing

Column	1	2	3	4	5	6	7
Approach	OLS	IV	IV	IV	IV	IV	IV
IV	-	Terms of trade	Investment ratio	Black market	Real exchange	All four	All three
			Panel A: our estimation				
GDP per capita	-0.08**	***0.165***	-0.402**	-0.197**	-0.425**	-0.116	-0.253**
	(2.82)	***(0.79)***	(2.52)	(2.16)	(2.32)	(1.08)	(2.85)
	<2.87>	***<1.07>***	<3.08>	<2.00>	<2.21>	<1.02>	<2.64>
Education(level)	-0.027**	-0.049**	-0.017**	-0.026**	-0.027**	-0.056**	-0.031**
	(5.09)	(5.04)	(2.09)	(4.09)	(2.94)	(6.39)	(4.18)
	<3.15>	<4.71>	<1.84>	<3.88>	<2.34>	<5.20>	<2.86>
1965	-0.069**	-0.085**	-0.031	-0.056**	-0.027	-0.056**	-0.052**
1970	-0.080**	-0.084**	-0.045**	-0.066**	-0.034**	-0.051**	-0.057**
1975	-0.109**	-0.123**	-0.075**	-0.094**	-0.066**	-0.080**	-0.085**
1980	-0.145**	-0.162**	-0.120**	-0.135**	-0.109**	-0.125**	-0.123**
1985	-0.143**	-0.118**	-0.163**	-0.150**	-0.156**	-0.124**	-0.148**
First stage R^2	-	0.211	0.175	0.23	0.18	0.318	0.29
Davidson & MacKinnon[42] (p-value)	-	0.23	0.026	0.068	0.015	0.34	0.34
Sargan (p-value)	-	-	-	-	-	0.31	0.81
# countries	66	58	66	65	55	51	55
# observations	294	218	294	275	217	165	208
			Panel B: PS' estimation				
GDP per capita	-0.19**	-0.98	-0.35**	-0.234	-0.755	-0.28**	-0.29**
	(4.34)	(1.28)	(3.37)	(1.25)	(0.67)	(2.15)	(2.51)
Education(level)	-0.019**	-0.018**	0.017**	-0.019**	-0.012	-0.019**	-0.021**
	(6.22)	(2.64)	(5.05)	(4.53)	(0.821)	(4.70)	(5.32)
1965	-0.030	-0.127	-0.015	-0.022	-0.024	-0.030	-0.007
1970	-0.016	-0.118	-0.004	-0.015	-0.057	-0.011	-0.005
1975	-0.048	-0.065	-0.036	-0.046	-0.002	-0.041	-0.028
1980	-0.074	-0.035	-0.061	-0.069	-0.027	-0.065	-0.055
1985	-0.091	-0.213	-0.109	-0.094	-0.153	-0.095	-0.087
First stage R^2	-	0.243	0.353	0.285	0.175	0.436	0.412
Hausman (p-value)	-	0.102	0.077	0.979	0.488	0.573	0.319
Sargan (p-value)	-	-	-	-	-	0.051	0.343

[42] EViews uses Davidson and MacKinnon's (1989, 1993) version of the Hausman test.

# countries	58	51	58	50	58	48	51
# observations	184	143	184	150	184	134	143

Note: angled brackets indicate heteroskedasticity consistent *t*-statistics. Asterisks ** and * represent 5% and 10% statistical significance respectively.

Our replication of PS's work suggests that the income impact of the infant mortality is underestimated in the OLS regressions. Investment ratio, black market premium, and real exchange distortion are appropriate instrumental variables, according to the results of the Sargan tests, to correct for this downward bias. At the same time, our replication also highlights that good external instrumental variables are not easy to find. All the first-stage R-squares in the IV estimations are lower than 0.25, even when the four instruments are used together they only increase the first-stage R-square to 0.32. This motivates us to look for other instruments that are more closely related to the endogenous explanatory variable. In Section 4, we discuss the application of internal instrumental variables.

Estimations for Life Expectancy

In this section, similar work is undertaken for the life expectancy equation. Although life expectancy is also an aggregated health variable, our results for the life expectancy equations are not as robust as those reported when infant mortality is adopted as the health status measure. As discussed by PS, first, because life expectancy data are usually derived from infant mortality data and model life tables, rather than from real death registrations, they often do not correctly measure the patterns of mortality in developing countries, where records of death are generally incomplete. The second reason is that the relation between income and infant mortality and the one between income and life expectancy can be different. Infant mortality is usually determined by parents' income, education and technological improvements that help cure communicable diseases. Causes of life expectancy and

adult mortality are more complex. They relate to an individual's behavior, along with factors such as income, so, for example, those people participating in risky sports may face shorter life expectancies. These adult behaviors will finally be reflected in the calculation of the life expectancy at birth, so when we only focus on the income-health association, it is difficult to obtain a clear picture. Despite these concerns, Table 4 reports OLS and IV estimates for model (3.6) with life expectancy at birth as the dependent variable.

We find that income is positively associated with this health measure, with the OLS estimate of the income elasticity for life expectancy at birth being 0.015, while the IV estimates are much higher. These features are consistent with those from the infant mortality equation. We also find that the estimates of the income effect are not statistically significant in some cases, dependent on the instrument set. PS also find this feature (see Panel B of Table 4), where the income coefficients from all IV estimations are statistically insignificant.

Table 4: The effect of per capita income on life expectancy at birth, 1960-1985, five-year differencing

Column #	1	2	3	4	5	6	7
	OLS	IV	IV	IV	IV	IV	IV
IV	-	Terms of trade	Investment ratio	Black market premium	Real exchange distortion	All four	All three
			Panel A: our estimation				
GDP per capita	0.015**	***-0.014***[43]	0.029	0.054**	0.094*	0.023	0.049**
	(2.48)	***(0.30)***	(1.36)	(2.51)	(2.00)	(1.03)	(2.57)
	<2.61>	***<0.64>***	<1.77>	<3.28>	<4.95>	<2.74>	<3.38>
Education(level)	-0.006**	-0.006**	-0.007**	-0.007**	-0.007**	-0.005**	-0.005**
	(5.09)	(2.45)	(4.20)	(4.63)	(3.15)	(2.34)	(3.25)
	<2.79>	<3.85>	<9.44>	<4.78>	<3.32>	<2.76>	<3.70>
1965	0.042**	0.047**	0.039**	0.038**	0.031**	0.041**	0.037**
1970	0.050**	0.053**	0.046**	0.046**	0.039**	0.046**	0.045**
1975	0.039**	0.042**	0.036**	0.035**	0.030**	0.037**	0.035**
1980	0.044**	0.047**	0.042**	0.041**	0.038**	0.043**	0.041**
1985	0.037**	0.037**	0.038**	0.039**	0.040**	0.037**	0.037**
First stage R^2		0.222	0.182	0.237	0.184	0.343	0.31
Davidson&Mackinnon (p-value)	-	0.56	0.26	0.11	0.015	0.23	0.47
Sargan(p-value)	-	-	-	-	-	0.7	0.58
# countries	75	66	75	73	63	58	62
# observations	352	260	352	322	262	201	248
			Panel B: PS' estimation				
GDP per capita	0.015	0.075	0.007	-0.018	0.709	0.012	0.24
	(1.77)	(0.660)	(0.36)	(0.50)	(0.39)	(0.55)	(1.18)
Education(level)	-0.0015	-0.002	0.002	0.007	-0.046	0.004	0.001
	(0.64)	(0.27)	(0.06)	(1.51)	(0.38)	(1.05)	(0.27)
1965	0.037	0.016	0.038	-0.038	-0.058	0.026	0.026
1970	0.037	0.027	0.038	-0.042	-0.070	0.037	
1975	0.035	0.028	0.036	-0.038	-0.036	0.035	0.034
1980	0.036	0.029	0.037	-0.038	-0.039	0.035	0.035
1985	0.033	0.036	0.032	-0.029	-0.085	0.031	0.35
First stage R^2	-	0.243	0.353	0.285	0.175	0.478	0.466
Hausman (p-value)	-	0.618	0.66	0.288	0.021	0.34	0.933
Sargan(p-value)	-	-	-	-	-	0.219	0.153
# countries	58	51	58	50	58	48	51
# observations	184	143	184	150	184	134	143

Note: angled brackets indicate heteroskedasticity consistent *t*-statistics. Asterisks ** and * represent 5% and 10% statistical significance respectively.

[43] As with the results reported in Table 3, when terms of trade is the instrumental variable the IV estimates are not intuitive. We expect a positive impact of income on life expectancy but here the estimate is negative (although statistically insignificant).

One puzzle in panel A of Table 4 is that the impact of education on life expectancy is negative. PS also find that education is negatively associated with health in some cases but these estimates are not statistically significant. In our results, most of them are statistically significant. This suspicious result brings us back to the question raised in footnote 7: Should level of education or the difference of education be used in the differencing model? In the next section we compare these two specifications, but here we hypothesize that the negative education effect on life expectancy is *not* anti-intuitive. Since model (3.6) uses the *level* of (log) education to explain the *change* of (log) life expectancy, the resulting negative parameter of education indicates that education negatively affects the change (growth) of life expectancy. That is, as education increases, the speed of life expectancy growth slows down. This result can be observed from the data: life expectancy in most countries increases between 1960 and 1985 but the growth rate of life expectancy decreases over this period. If education does have a positive impact on life expectancy, this impact should be diminishing at the margin because it becomes more and more difficult to increase life expectancy as life expectancy reaches higher and higher levels. To confirm that the level of education still has a positive effect on the *level* of life expectancy, we regressed (log) life expectancy on (log) income and (log) education; then the education variable's coefficient is significantly positive, being 0.109 with a *t*-statistic of 16.38.[44]

[44] In the infant mortality equation, the coefficient of the education variable is negative, which indicates that as education increases, the change in infant mortality decreases. As the change in infant mortality is usually a negative number, the decreasing change in infant mortality indicates that health improves faster. This can be confirmed by the infant mortality data – for many countries' infant mortality decreased from 1960 to 1985 with the rate of this decrease accelerating over time. This finding does not necessarily imply that education has a differing impact on infant mortality and life expectancy; for example, the marginal effect of education can be diminishing for life expectancy but increasing for infant mortality. That is, we are suggesting that life expectancy arrives at a saddle point, where the education effect becomes diminishing, earlier than does infant mortality.

Regression with Difference of Education for Infant Mortality and Life Expectancy

This section discusses regression results when the difference in education, rather than the *level* of education, is used as the regressor. The estimated model is:

$$\log y_{it} - \log y_{it-1} = \beta(\log x_{it} - \log x_{it-1}) + \gamma(\log z_{it} - \log z_{it-1}) + \delta_t^{'} + (\varepsilon_{it} - \varepsilon_{it-1}) \quad (3.7)$$

To save space estimation results are not reported here. We find: the IV estimates are larger than the OLS estimates; first-stage R^2 values are low; based on the Sargan test, the external instruments are valid to control for the endogeneity of income. When life expectancy is used as the measure of health status, estimation of model (3.7) confirms that the impact of education is positive, producing the intuitively positive association between the *level* of education and the *level* of life expectancy. Despite these results, in the following sections, we base our analysis on model (3.6) instead of model (3.7) for two reasons. First, following most of the income growth literature where *level* of education is often used to explain the income growth, the health literature (e.g. Hill and King, 1992; Bhalla and Gill, 1993) also use *lagged* school enrollment as a control variable for health improvement, supporting the use of the *level*. Secondly, as one objective of this paper is to replicate PS, we wish to be consistent with their statement that they use the "level" of education in the differenced model.

3. Sensitivity Analyses

Without any formal tests, it is difficult to conclude that the income-health relationship observed recently is similar to that detected over 1960 through 1985. On the one hand, with social development, more developing countries' health indices have reached or

have been close to their respective upper bounds. Therefore, health may become less responsive to further increases in income. In our dataset, infant mortality in 34 out of our 105 countries decreased from year to year between 1960 and 1985 at an increasing rate, but after 1985, out of those 34 countries, only 10 maintained this trend.

On the other hand, deterioration of health in Africa in the post-1985 period weakens the global health-income association. From 1960 to 1985 there were 97 countries experiencing infant mortality decreases from year to year, but after 1985 only 82 of them continued experiencing the same trend. There were 92 countries that had increasing life expectancy before 1985 but only 62 of them maintained this trend after 1985. Health in those unfortunate countries, most of them being African countries, is dragged down by their unstable political environment, frequent civil wars, and HIV/AIDS. For instance, the genocide in Rwanda (1993-1994) killed 1,000,000 people, which is reflected as a dramatic drop in their life expectancy (43.74 to 31.17) and an increase in infant mortality (103 to 124). Given these features, in this section we extend the analysis to a larger sample covering data from 1960 to 2000. From the previous discussion (Tables 2-4) heteroskedasticity-corrected standard errors actually do not bring substantially different conclusions from those with non-heteroskedasticity-corrected standard errors, so below we reports results with non-heteroskedasticity-corrected standard errors.

3.1 OLS and IV Estimations for Infant Mortality (1960-2000)

The GDP per capita data from 1960 to 2000 are obtained from the Penn World Table 6.2, measured in 2000 real purchasing power (PPP) adjusted dollars. For consistency, we use the same 105 countries listed in footnote 3. Health data (infant mortality and

life expectancy) are still from the WDI. The education variable is defined as before, obtained from Barro and Lee (2000), as provided by the Center for International Development at Harvard University (www.cid.harvard.edu).

OLS and IV estimates of the infant mortality regression are reported in Table 5, which is comparable to Panel A of Table 3. Typically, there is little change in the IV income elasticity estimates - income still has a negative impact on infant mortality except when the terms-of-trade shock is used as the instrumental variable. But when the investment ratio and real exchange rate distortion are chosen as instruments, this elasticity is slightly smaller than presented in Table 3. Another finding is smaller first-stage R^2 values. As previously raised, we found terms of trade shock, investment ratio, black market premium and real exchange rate distortion to be weak instrumental variables for the 1960-1985 sample, so it is not surprising to see it is more difficult for them to explain the income variation over a longer sample window (1960-2000).

Table 5: The effect of per capita income on infant mortality, 1960-2000, five-year differencing

Column #	1	2	3	4	5	6
	OLS	IV	IV	IV	IV	IV
IV	-	Terms of trade	Investment ratio	Black market premium	Price level distortion	All four
GDP per capita	-0.081**	0.193	-0.297*	-1.13**	-0.091	-0.144
	(2.87)	(0.47)	(1.94)	(2.23)	(0.49)	(1.12)
education(level)	-0.041**	-0.058**	-0.037**	-0.028*	-0.048**	-0.061**
	(6.30)	(4.57)	(4.95)	(1.89)	(6.22)	(6.68)
1965	-0.064**	-0.082*	-0.042	0.058	-0.058**	-0.052*
1970	-0.069**	-0.081*	-0.045	0.055	-0.061**	-0.045
1975	-0.103**	-0.120**	-0.081	0.013	-0.098**	-0.075**
1980	-0.137**	-0.162**	-0.115**	-0.021	-0.127**	-0.118**
1985	-0.129**	-0.114**	-0.132**	-0.134**	-0.127**	-0.116**
1990	-0.113**	-0.107**	-0.106**	-0.067*	-0.107**	-0.09**
1995	-0.061**	-0.061**	-0.065**	-0.056**	-0.063**	-0.044**
2000	-0.066**	-0.081**	-0.045	-0.043	-0.062**	-0.044**
First stage R^2	-	0.089	0.11	0.087	0.092	0.14
# countries	78	72	78	75	66	64
# observations	473	397	473	351	382	328

Note: Absolute t-statistics (not heteroskedasticity-corrected) are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance respectively.

Comparing the estimates in Table 3 and Table 5, we find that the qualitative results from the first-differenced model are almost the same. The magnitudes of the IV estimates have changed but the signs continue to be the same, leading us to conclude that the qualitative income-health relation is not sensitive to the updating of the data. Therefore, although over recent years health of some countries have moved closer to the physical upper bound while in some underdeveloped countries health keeps deteriorating, these changes do not seem to substantially affect the world-wide income-health relationship. As our panel data model restricts the income impact on health to be homogenous across countries, that is, the income coefficient in model (3.6) is identical for all countries, intuitively, what we estimate is a world-wide averaged income-health association. Good news from fortunate countries cancels out

the bad news from those African countries, so that the averaged result only quantitatively changes a little but still remains qualitatively the same as found over the 1960-1985 period.

3.2 Estimating with Other Health Achievement Indices

So far we have used raw data on infant mortality and life expectancy as our health measures. This implies that an increase in the health status of a country when it is already at a high level is equal to an achievement of another country with an equal increase but from a lower base. Kakwani (1993) and Anand and Ravallion (AR hereafter) (1993), for example, have argued that this is not appropriate. Increasing life expectancy from 75 to 80 is much more difficult than increasing it from 50 to 55. To capture the nonlinearity of the health improvement, Kakwani defines a health achievement index:

$$f(x,m,M)=(1-\log(x-m)/\log(M-m))*100$$

for infant mortality and

$$f(x,m,M)=(1-\log(M-x)/\log(M-m))*100$$

for life expectancy. Anand and Ravallion define infant mortality and life expectancy indices as $f(x,m)=\log(x-m)$ and $f(x,M)=-\log(M-x)$, respectively. Here x refers to the health variable (infant mortality or life expectancy); m=5 (30) is the lower bound of the infant mortality (life expectancy at birth); and M=300 (85) is the upper bound of the infant mortality (life expectancy at birth). These achievement indices ensure that a further improvement in health at a higher level is regarded as a greater achievement than an equal improvement at lower levels of health. Using life expectancy as an example, both achievement indices are convex functions of the raw life expectancy

data (x), i.e. the indices increase faster than life expectancy.[45]

Without any formal testing, PS (page847, footnote 4) state that use of Kakwani's achievement indices do not change the estimates of the income elasticity. Filmer and Pritchett (1999, page1310) also believe that it is not an issue to ignore the nonlinear property of health indices, since "although these indices allow for varying income elasticity and consider the physiologically determined lower/upper bound on mortality/life expectancy from unavoidable natal deaths, for many developing countries this lower/upper bound is never approached". However, our sample covers 105 developing countries, so even if most of them are far away from the bounds, some of them have reached or are close to the lower/upper bounds. For example, infant mortality in South Korea hit the lower bound (3.5) in 2000, and the life expectancy in Costa Rica expectancy increased from 62 in 1960 to 78 in 2000. We report the results of estimating our models with Kakwani's and AR's health achievement indices in Table 6, to examine whether the choice of the functional form of the health measure affects the income elasticity estimates. Kakwani (1993, p324) states that "Anand and Ravallion's transformation function is quite similar to the achievement function used in the present paper". Accordingly, for space reasons, we only report results using AR's health indices.

Panel A of Table 6 shows that AR's achievement index leads to a similar conclusion for income's effect on infant mortality as we found using the raw data. This is expected as AR's achievement index for infant mortality $\log(x\text{-}m)$ is essentially the same as log infant mortality $\log(x)$: they both are concave transformations of the raw

[45] It is easy to prove that the health indices of life expectancy have finite positive first-order and second-order derivatives if $0 < x < M$.

data. In panel B the income effect on life expectancy is positive and statistically significant. However, the impact of education is positive while we found it to be negative when using the raw health data. In Section 2.3 we explained that this negative educational impact suggests that the growth of (log) life expectancy increases slower than does (log) education. The positive impact of education using AR's index does not conflict with this explanation. AR's achievement index of life expectancy, $-\log(M\text{-}x)$, is a convex transformation of the raw data, which is different from a simple log transformation. The following simplified proof explains why the impact of education on Anand and Ravallion's life expectancy achievement index should be positive and that this is consistent with the negative impact of education reported in Table 4.

Given model (3.6), simply denote the right-hand-side as LE_{PS} if it is the difference of log life expectancy (index function i), and denote it as LE_{AR} if it is the difference of AR's achievement index (index function j). LE_{PS} and LE_{AR} are estimated as functions, f and g, respectively, of log education. Hence,

$$\begin{aligned} LE_{PS} &= i(\text{life expectancy}) = d(\log(\text{life expectancy})) \\ &= f(\log(\text{edu})) \\ LE_{AR} &= j(\text{life expectancy}) = d(-\log(M - \text{life expectancy})) \\ &= g(\log(\text{edu})) \end{aligned}$$

We are interested in the impact of education on the difference of life expectancy achievement, or LE_{AR}:

$$\begin{aligned} \frac{d(LE_{AR})}{d(\log(\text{edu}))} &= \frac{d(LE_{AR})}{d(LE_{PS})}\frac{d(LE_{PS})}{d(\log(\text{edu}))} \\ &= \frac{d(LE_{AR})}{d(\text{life expectancy})}\frac{d(\text{life expectancy})}{d(LE_{PS})}\frac{d(LE_{PS})}{d(\log(\text{edu}))} \end{aligned}$$

$$\because \frac{d(LE_{AR})}{d(\text{life expectancy})} = \frac{dj}{d(\text{life expectancy})} = \frac{1}{(M - \text{life expectancy})^2} > 0;$$

$$\frac{d(\text{life expectancy})}{d(LE_{PS})} = \frac{1}{\frac{di}{d(\text{life expectancy})}} = -(\text{life expectancy})^2 < 0;$$

and $\frac{d(LE_{PS})}{d(\log(\text{edu}))} < 0$, according to the results in Table 4,

▪

Table 6: The effect of per capita income on health, IV estimation, 1960-2000, five-year differencing

Column #	1	2	3	4
Dep. Var.	Infant mortality		Life expectancy	
Health measure	PS	Anand &	PS	Anand &
GDP per capita	-0.144	-0.48**	0.079*	0.166**
	(1.12)	(2.12)	(1.89)	(2.12)
education(level)	-0.061**	-0.089**	-0.003	0.025**
	(6.68)	(5.5)	(0.96)	(3.85)
1965	-0.052	-0.008	0.033	0.040**
1970	-0.045	0.013	0.035	0.047**
1975	-0.075	-0.031	0.031	0.039**
1980	-0.118	-0.078**	0.036	0.058**
1985	-0.116	-0.115**	0.035	0.060**
1990	-0.09	-0.074**	0.015	0.035**
1995	-0.044	-0.049	0.002	-0.003
2000	-0.044	-0.038	0.004	-0.008
IV	All four	All four	All four	All four
# countries	64	64	64	64
# observations	328	328	365	365

Note: Absolute t-statistics (not heteroskedasticity-corrected) are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance respectively.

3.3 Estimates with Public Health Expenditure as a Regressor

PS state that model (3.6), for example, is estimating a "total" income effect on health, where all the other variables that may also affect health are approximated by income per capita. For example, an increase in income may lead to higher public or private

health expenditure, which essentially influences health status, but model (3.6) considers this direct effect of public/private health expenditure as part of the "total" income effect rather than as a separate effect to be estimated. For policy makers, a more interesting question may be to address the "net" income effect. For instance, when all the income impacts on health go through public/private health expenditures, the "net" income effect is zero. In this case it is more efficient for policy makers to focus on the public health system than to work on wealth redistribution.[46]

Controlling for public health expenditure per capita as well as poverty (proportion of the population consuming less than PPP $1 per day in 1985), AR (1993) conclude that the "net" income effect on health is statistically insignificant. However, Filmer and Pritchett (FP) (1999) find that the income effect is still important even after controlling for public health expenditure. Both studies use cross-section regressions but the latter considers public health expenditure to be endogenous and uses IV estimation to control for this endogeneity.

These studies differ in approach from ours. First, they use cross-sectional data rather than panel data and hence cannot control for unobserved country specific effects. Second, neither study accounts for the endogeneity of income.[47] To be consistent with these works, we first estimated a cross-section regression using data on AR's 22 developing countries for the year 1990, and then we extend the estimation to a panel

[46] These are defined as "support-led" (public/private health supports lead to better health) and "growth-mediated" (income/economy developments improve health) strategies. Dreze and Sen (1989) make a distinction between these two strategies of capability expansion. Examples of "support-led" countries include China, Jamaica, Costa Rica and Sri Lanka, where the governments directly invest resources to expand social capability. Examples of "growth-mediated" countries include Hong Kong, South Korea and Singapore, where capability expansion mainly resulted from wide distribution of the fruits of growth.

[47] FP considers the endogeneity of the variable public health expenditure as a percent of GDP and instruments it by the average defence spending share of GDP of a country's geographic neighbours, but they ignore the endogeneity of income.

data framework. Table 7 reports the OLS results, which are at odds with those reported by AR. After controlling for health expenditure per capita and poverty, income, health expenditure per capita and poverty become statistically insignificant.[48]

Table 7: The effect of income per capita on health in cross-country data (1990)

Column #	1	2	3	4
	Infant mortality		Life expectancy	
GDP per capita	-0.771**	-0.552	0.491**	0.261
	(2.96)	(0.78)	(2.81)	(0.53)
Health exp (log)		0.179		0.197
		(0.19)		(0.30)
Poverty (log)		0.194		-0.067
		(0.96)		(0.48)
Constant	9.954	7.424	-6.655	-4.915
R^2	0.32	0.32	0.29	0.28
# observations	21	17	21	17

Note: Absolute t-statistics (not heteroskedasticity-corrected) are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance respectively.

Data differences may be an issue here: AR use GNP per capita instead of GDP per capita as the income variable; their variable of health expenditure is public health expenditure per capita rather than total health expenditure per capita, and their poverty data are from 1985, not 1990.[49] When we replace total health expenditure per capita by public health expenditure as a percent of GDP, which is one of the controlling variables in FP, our results change.[50] Table 8 summarizes these results: only the significance of income vanishes but public health expenditure and poverty still have explanatory power. This finding is consistent with AR but different from FP. As AR suggests (p142), the insignificance of income "does not imply that economic

[48] Data on health expenditure per capita are obtained from the World Health Report 1995; data on poverty (the proportion of population consuming less than PPP $1 per day in 1990) are obtained from the World Bank Millennium Development Goals (MDG) database.

[49] In fact, AR don't explicitly state whether their estimations are based on data in 1990 or in 1985. We believe their estimates come from an income-health relationship for a year before 1990 because their health data (life expectancy) came from UNDP 1990, in which the data are for years prior to 1988.

[50] Data of public health expenditure as a percent GDP are from the WHO database.

growth is unimportant in expanding life expectancy (reducing infant mortality); rather it says that the importance of growth lies in the way that its benefits are distributed between people, and the extent to which growth supports public health services." However, limited sample size likely limits the interpretability of the findings reported by AR and ours contained in Tables 7 and 8. This is also recognized by AR (p142): "(results) are based solely on the patterns observed in this sample of 22 countries." The lack of robustness of their findings is also reflected when we compare Tables 7 and 8: estimates are sensitive to the choice of the health expenditure variable.

Table 8: The effect of income per capita on health in cross-country data (1990)

Column #	1	2
	Infant mortality	Life expectancy
GDP per capita	0.152	-0.125
	(0.37)	(0.5)
Pub health exp % (log)	-0.933**	0.823**
	(2.60)	(3.72)
Poverty (log)	0.321*	-0.198*
	(1.81)	(1.86)
Constant	2.715	-1.584
R^2	0.54	0.65
# observations	16	16

Note: Absolute t-statistics (not heteroskedasticity-corrected) are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance, respectively,

To further study the roles of public health expenditure and poverty on the income-health relationship, we move to a panel framework with data in 1990, 1995, and 2000.[51] Results for the first differencing and fixed effect methods are reported, respectively, in Tables 9 and 10. To save space, we do not report the estimates of the time/year effects.

[51] Restricted by the availability of poverty data, panel regressions can only use data for 1990, 1995, and 2000.

With the first differencing approach, we find that all the regressors - income, public health expenditure (%) and poverty - are statistically insignificant. Similar results are obtained when instrumental variables estimation is adopted to control for the endogeneity of income.

Table 9: The effect of income per capita on health in 22 countries' panel data (first differencing approach)

Column #	1	2	3	4
	Infant mortality		Life expectancy	
IV estimation	no	All four	No	All four
GDP per capita	-0.040	-0.166	0.028	0.103
	(0.25)	(0.54)	(0.33)	(0.62)
Pub health exp % (log)	-0.036	-0.031	0.015	0.018
	(0.53)	(0.42)	(0.41)	(0.45)
Poverty (log)	0.002	-0.0007	0.011	0.015
	(0.10)	(0.03)	(0.92)	(1.06)
# observations	34	27	34	27

Note: Absolute t-statistics (not hesteroskedasticity-corrected) are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance.

The fixed effects estimation results, reported in Table 10, are provided for comparison. Although public health expenditure (%) and poverty are statistically significant in the infant mortality equation, results from the life expectancy regression still suggest that our panel estimation neither reproduces the results of AR nor replicates the findings in FP.

One issue could be the assumption of a constant cross-country income coefficient. That is, we assume each country has the same income-health relationship or public health expenditure-health relationship. If this assumption is not valid, then we are estimating the mean of the slopes. Plots of public health expenditure (%) versus life expectancy and income versus life expectancy for 22 developing countries over 1990

through 2000 suggest that only half of them have positive associations between health

Table 10: The effect of income per capita on health in 22 countries' panel data (fixed effect approach)

Column #	1	2
	Infant mortality	Life expectancy
GDP per capita	-0.102	0.023
	(0.67)	(0.27)
Pub health exp % (log)	-0.147**	0.032
	(2.21)	(0.83)
Poverty (log)	-0.044*	0.023
	(1.97)	(1.77)
# observations	55	55

Note: Absolute t-statistics (not heteroskedasticity-corrected) are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance.

expenditure/income and life expectancy, whereas the remaining countries show negative associations or ambiguous relations.[52] [53] Hence, if we are estimating the "average" relationship with the panel model it is perhaps not surprising that we detect insignificant results.

We next include other developing countries used in our analysis reported in Section 3.1. However, as the data on public health expenditure and poverty are only available after 1990, the time dimension of this panel dataset is still restricted to be three years (T=3). If most countries have a statistically significant income-health or health expenditure-health relationship, even if methods estimate "average" coefficients, we should still expect to obtain a significant impact. Results are reported in Tables 11 and 12. We find that income is statistically significant whereas public health expenditure

[52] These 22 developing countries are: Bangladesh, Botswana, Brazil, China, Columbia, Costa Rica, Ghana, Guatemala, India, Indonesia, Cote d'Ivoire, Jamaica, Malaysia, Morocco, Pakistan, Peru, Philippine, Poland, Sri Lanka, Thailand, Venezuela, and Yugoslavia.

[53] Eight countries show positive public health-life expectancy association. They are: Bangladesh, Brazil, China, Columbia, Guatemala, India, Malaysia, Pakistan, Peru, Philippines, and Thailand. Twelve countries have positive income-health association: Bangladesh, Brazil, China, Columbia, Guatemala, India, Malaysia, Pakistan, Peru, Philippines, and Thailand.

and poverty are insignificant. This finding is also fund by FP, who conclude that income is still important after controlling for public health expenditure and other variables. However, the methodology used to generate the estimates provided in Tables 11 and 12 differs from FP, as FP use cross sectional data rather than panel data, and they control for the endogeneity of public health expenditure rather than for the endogeneity of income.

Table 11: The effect of income per capita on health in panel dataset (first-differencing approach)

Column #	1	2	3	4
	Infant mortality		Life expectancy	
IV	no	All four	No	All four
GDP per capita	-0.337**	-0.604**	0.168**	0.178
	(2.39)	(2.21)	(2.34)	(0.91)
Pub health exp % (log)	0.004	-0.007	0.013	0.015
	(0.08)	(0.16)	(0.50)	(0.46)
Poverty (log)	-0.016	-0.015	0.0005	-0.006
	(0.55)	(0.57)	(0.04)	(0.35)
# observations	83	67	83	67

Note: Absolute t-statistics (not heteroskedasticity-corrected) are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance.

For comparisons, results using fixed-effect estimation are reported in the table below.

Table 12: The effect of income per capita on health in panel dataset (fixed-effect approach)

Column #	1	2
	Infant mortality	Life expectancy
GDP per capita	-0.432**	0.179**
	(2.91)	(2.54)
Pub health exp % (log)	0.045	-0.001
	(0.73)	(0.02)
Poverty (log)	-0.015	0.014
	(0.47)	(0.9)
# observations	153	154

Note: Absolute t-statistics (not heteroskedasticity-corrected) are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance.

Our evidence, in Tables 7-12, suggests that the income effect on health is ambiguous

after controlling for public health expenditure and poverty. The estimates highlight the sensitivity of findings to the choice of health expenditure variable, the data type (cross-section versus panel), and the choice of countries.

4. Internal Instrumental Variables Estimates

In this section, we provide estimates of parameters for our models when using internal instruments. We choose to adopt these instruments given the difficulty of finding valid external instruments. The issue is that an instrumental variable needs to satisfy two conditions: it has to be correlated with the problematic regressor and simultaneously uncorrelated with the error term. Although PS suggest four different instrumental variables for the income growth, which we used to generate the two-stage least squared estimates provided in Tables 3-5, the first-stage R^2 values are low. Even when we use the four instrumental variables together, no more than 40 percent of the variation of income growth is explained. In this section, we consider an alternative route of using internal instruments that are highly correlated to income per capita. In order to maintain comparability with PS, regressions in this section use our 1960-1985 panel dataset.

In order to be able to generate internal instruments, we move to IV estimation with dynamic panels models; see, e.g., Amemiya and Macurdy (1986), Breusch et al. (1989) and Arellano and Bond (1991). Consider:

$$y_{it} = \alpha_i + \gamma y_{it-1} + \beta x_{it} + \varepsilon_{it}, \tag{3.8}$$

where x is an endogenous regressor. To remove the country-effect, consider the first-differencing model:

$$y_{it} - y_{it-1} = \gamma(y_{it-1} - y_{it-2}) + \beta(x_{it} - x_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1}). \quad (3.9)$$

There exist two difficulties with estimating equation (3.9): the differenced variable $(y_{it-1} - y_{it-2})$ is correlated with the differenced error term $(\varepsilon_{it} - \varepsilon_{it-1})$ and the differenced variable $(x_{it} - x_{it-1})$ is correlated with the error term $(\varepsilon_{it} - \varepsilon_{it-1})$, as x_{it} is endogenous. To proceed, we assume:

(i) $\alpha_i + \varepsilon_{it}$ has the standard error components structure: $E(\alpha_i) = 0$, $E(\varepsilon_{it}) = 0$, $E(\alpha_i \varepsilon_{it}) = 0$ for $i = 1,\ldots,N$ and $t = 2,..,T$.

(ii) The errors are serially uncorrelated: $E(\varepsilon_{it}\varepsilon_{is}) = 0$, for $i = 1,\ldots,N$ and $s \neq t$.

(iii) The initial conditions y_{i1} are predetermined: $E(y_{i1}\varepsilon_{it}) = 0$ for $i = 1,\ldots,N$ and $t = 2,\ldots,T$.

(iv) The explanatory variable x is strictly exogenous and satisfies $E(y_{it} \mid y_{it-1}, x_{i1},..., x_{iT}, \alpha_i) = E(y_{it} \mid y_{it-1}, x_{it}, \alpha_i)$.

These assumptions imply $m = 0.5(T\text{-}1)(T\text{-}2)$ moment restrictions: $E(y_{it-s}\Delta\varepsilon_{it}) = 0$ and $E(y_{it-s}\Delta y_{it-1}) \neq 0$, for $t = 3,\ldots,T$ and $s \geq 2$. Further, once we control for x_{it} and the country effect α_i, no past values of x affect the expected value of y_{it}. In addition, such a framework allows for both contemporaneous correlation between the current shock ε_{it} and x_{it}, and feedbacks from past shocks ε_{it-s} onto the current value of x_{it} in the sense that $E(x_{it}\varepsilon_{is}) \neq 0$ for $i = 1,\ldots,N$ and $s \leq t$. Taking first-differences to eliminate the country effect α_i, we obtain moment conditions $E(x_{it-s}\Delta\varepsilon_{it}) \neq 0$ for $t = 3,\ldots,T$ and $s \geq 2$. Therefore, lagged values of the dependent variable y_{it} and the endogenous variable x_{it} dated t-2 and earlier can then be used as instruments for the first-differenced equation.

To move to this framework, we must ask: can our income-health model (3.6) be extended to be a dynamic model? Although most of the empirical papers discussed in the introduction used static models, other researchers have adopted dynamic

specifications for the relationship. One of the seminal works on modeling the demand for health is Grossman (1972). A central tenet of this paper is that health can be considered as a *durable capital stock,* from a microeconomic perspective, which produces an output of healthy times. His work has encouraged empirical studies along this line. For example, to allow for the possibility that income and other variables may have lagged effects on the infant mortality, Anand and Ravallion (1993) include the lagged value of infant mortality when they analyze GDP and public health spending impacts on health for Sri Lanka;[54] McDonald and Roberts (2006) use a dynamic specification for their health function; Gan and Gong (2007) study the association between health and education within a dynamic setup; Bishai et al.'s (2007) dynamic model, with 21 European and American countries', suggests that infant mortality is often a random walk property.[55] Following these papers, we extend our model to a dynamic set up:

$$\log y_{it} = \alpha_i + \delta \log y_{it-1} + \beta \log x_{it} + \gamma \log z_{it-1} + \varepsilon_{it} \ . \tag{3.10}$$

Removing the unobserved country specific term α_i using first-order differencing (Arellano and Bond, 1991), we have:

$$\log y_{it} - \log y_{it-1} = \delta(\log y_{it-1} - \log y_{it-2}) + \beta(\log x_{it} - \log x_{it-1}) + \gamma(\log z_{it-1} - \log z_{it-2}) + (\varepsilon_{it} - \varepsilon_{it-1}) \ . \tag{3.11}$$

Based on previous discussions, lagged terms (y_{it-2-j} and x_{it-2-j}, j = 0,1,…,) are used as internal instruments. As the number of instruments varies across time periods, we use Generalized Method of Moments (GMM) estimation (Hansen 1982, Holtz-Eakin et al. 1988 and Arellano and Bond 1991) to estimate equation (3.11).

[54] The regressions in Anand and Ravallion (1993) are static, apart from their dynamic analysis for Sri Lanka.

[55] That is, infant mortality in period *t* is determined by its level in period *t-1* plus a random noise component.

Table 13 reports the GMM estimates of the infant mortality equation with one-step and two-step covariance matrices.[56] Although the standard two-step estimator is asymptotically more efficient, it can lead to severe downward bias in finite samples (e.g., Arellano and Bond 1991, Blundell and Bond 2000). To compensate, we use a finite-sample correction to the two-step covariance matrix proposed by Windmeijer (2005), which resolves some of the bias issues. It also leads to an estimator that is more efficient and robust than the one-step GMM estimator.

We also explore whether education should be treated as endogenous. This contrasts with our previous results, which treated education as an exogenous regressor. Others have explored this assumption. For example, Bobba and Coviello (2007) consider education as an endogenous variable when they test the education-democracy relationship. Hence, to reduce the noise from the potential endogeneity of education, we begin with regressions without education. The pooled OLS and fixed effect estimators are also reported in Table 13 for comparison with the GMM estimates.

[56] The one-step GMM estimator uses a weight matrix that does not depend on unknown parameters or residuals. The two-step GMM estimator is based on a weight matrix that depends on the calculated one-step GMM estimator. See Blundell and Bond (2000).

Table 13: GMM estimation of the infant mortality equation 1960-1985

	GMM-One Step	GMM-Two Step	OLS	Fixed Effect
Lag of IMR	0.927**	0.963**	1.042**	1.004**
	(11.66)	(9.94)	(97.94)	(29.42)
GDP per capita	-0.122*	-0.061	-0.038**	-0.065**
	(1.90)	(1.17)	(5.85)	(3.62)
1970	-0.002	-0.006	-0.020	-0.0003
1975	-0.027	-0.033	-0.045	-0.025
1980	-0.059	-0.062	-0.069	-0.053
1985	-0.062	-0.056	-0.047	-0.042
Sargan (P-value)	0.091	0.091	-	-

Note: Robust t-statistics are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance, respectively.

Values in parentheses are robust *t*-statistics using a standard error estimator that is consistent in the presence of any heteroskedasticity and autocorrelation within panels. One finding from our results is that the lag of infant mortality is statistically significant and its coefficient is close to one, which confirms that the infant mortality series has strong persistence and suggests the merits in adopting the dynamic specification. We also see that the coefficient on the lag of infant mortality is smaller when using GMM than compared to that from the OLS and the fixed-effect estimators. With regard to comparing these estimates, Bond (2002, page 5) note: "The fact that these two estimators are likely to be biased in opposite directions is useful. Thus we might hope that a candidate consistent estimator will lie between the OLS and Within Groups estimates, or at least not be significantly higher than the former or significantly lower than the latter." Although we find that the GMM estimate is smaller than the OLS estimate, it is not larger than the fixed-effect estimator. For the two-step GMM estimator, the coefficient on the lagged infant mortality increases from 0.927 to 0.963, a move towards the fixed-effect estimator. However, the impact of income is sensitive to the choice of the GMM estimation procedure – the income elasticity is halved and becomes statistically insignificant with the two-step

approach.[57]

One difficulty with the approach we have taken is that the correlation between the levels and the subsequent differences reduces with the persistence present in the instruments. That is, high persistence may lead to weak instruments.[58] This suggests that we need to determine the persistence present in our series. One method is to ascertain whether they have unit roots. To this end, Table 14 reports standard panel unit root tests for infant mortality and GDP per capita. The tests support the belief that infant mortality has a unit root, whereas there is some uncertainty about the trending properties of GDP per capita. This suggests that our internal instruments may be weak so that the GMM estimates may not be as good as we would like. Specifically, simulations in Blundell and Bond (1998) illustrate a huge downward bias in the GMM estimator of the dynamic term (δ in (3.11)) with weak instruments. This could explain why the impact of lagged infant mortality in our GMM regressions is smaller than that using the fixed effects estimator.

[57] The magnitude of the income effects reported in Table 13 are not strictly comparable to those provided in Table 3, as the former are from a dynamic model whereas the latter are obtained from a static model.

[58] To illustrate, consider a simple AR(1) process: $x_t = \alpha x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$ so $x_t - x_{t-1} = (\alpha - 1)x_{t-1} + \varepsilon_t$. If $\alpha \to 1$, that is, x is persistent, then x_{t-1} has difficulty in explaining $x_t - x_{t-1}$ as $(\alpha - 1) \to 0$.

Table 14: Unit Root Tests of infant mortality and GDP per capita[59]

Column #	1	2	3	4
	Infant Mortality		GDP per capita	
Method	Test Statistic	P-Value	Test Statistic	P-Value
Null: unit root (assumes common persistence parameter across countries)				
Levin, Lin & Chu t	25.391	1.00	-13.789	0.00
Null: unit root (allows persistence parameters to vary across countries)				
Im, Pesaran and Shin W-stat	17.144	1.00	0.188	0.58
Fisher-type ADF Chi-square	71.276	1.00	192.043	0.41
Fisher-type PP Chi-square	144.816	0.69	277.30	0.00
Null: no unit root (assumes common persistence parameter across countries)				
Hadri Z-stat	14.858	0.00	14.561	0.00

Probabilities for Fisher tests are computed using an asymptotic Chi-square distribution. Other tests assume asymptotic normality.

When there are weak internal instruments, Blundell and Bond (2000) propose further restrictions on the model to yield additional moment conditions. They propose a system GMM method that stacks the equations in first differences and the equations in levels together and employs both lagged levels and differences as internal instruments. Specifically, they use lagged levels variables as instruments in the first differenced equation and use first differenced variables as instruments in the levels equation. This may reduce finite sample bias by exploiting additional moment conditions. It may also improve precision. The validity of the additional over-identifying moment conditions can be tested using a standard Sargan (1988) test. Moreover, since the moment conditions used in the non-system GMM are a strict subset of those used in the system GMM, a Sargan test based on the difference between the two standard Sargan statistics provides a more specific test of the additional moment conditions exploited by the system GMM estimator. Table 15 reports results from using the system GMM approach.

[59] More details on these five panel unit root tests are available in Levin, Lin, and Chu (2002), Im, Pesaran and Shin (2003), Maddala and Wu (1999), Choi (2001), and Hadri (2000). Also, see Greene (2008, p767) for a summary of some of these tests.

Table 15: System GMM estimation of the infant mortality equation 1960-1985

Column #	1	2
	System GMM-One Step	System GMM-Two Step
Lag of IMR	1.087**	1.089**
	(103.61)	(93.01)
GDP per capita	-0.073**	-0.074**
	(9.71)	(8.59)
1970	0.007**	0.002**
1975	-0.009**	-0.012**
1980	-0.023**	-0.016**
1985	0.011**	0.015**
Sargan (P-value)	0.058	0.058
Dif Sargan (P-value)	0.034	0.034

Note: Robust t-statistics are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance, respectively.

One advantage of the system GMM is reflected by the income effect: it is less sensitive than the results in Table 13 with respect to the one-step and two-step approaches. The Sargan test that examines the validity of lagged levels dated *t*-3 (and earlier) as instruments in the first-differnced equations, combined with lagged first differences dated *t*-2 as instruments in the levels equations, and the differenced Sargan test, which examines the validity of the additional moment conditions in the levels equations, appear to be accepted marginally in the system GMM. Although the lower p-value of the Sargan test can partly reflect the increased power of the test to reject the instruments used in the first-differenced equations, the low p-value of the differenced Sargan test suggests that our internal instruments are still correlated with the error term. For example, the presence of measurement error or serial correlated error term causes $E(y_{it-s}\Delta\varepsilon_{it}) \neq 0$ for $s \geq 2$. Hence, internal instruments may need to be lagged by one more period, that is, appropriate instruments are y_{it-3-j} and x_{t-3-j}, j = 0,1,… in the differenced equations, and Δy_{it-2} and Δy_{it-2} are instruments in the levels equations. GMM estimates with these new internal instruments are reported in Table

16. Results are similar to those reported in Tables 13 and 15: system GMM increases the coefficient of the dynamic term, and produces a reasonable and statistically significant income effect. Furthermore, using instruments dated t-3 (and earlier), the Sargan and differenced Sargan tests both easily accept the validity of the over-identifying moment conditions in the differenced and level equations.

Table 16: GMM and System GMM of the infant mortality with instruments dated t-3 (and earlier)

Column #	1	2
	GMM-Two Step (t-3)	System GMM-Two Step (t-3)
Lag of IMR	0.960**	1.072**
	(6.01)	(96.39)
GDP per capita	-0.072	-0.062**
	(0.62)	(7.82)
1970	-0.005**	0.001**
1975	-0.028**	-0.012**
1980	-0.049**	-0.014**
1985	-0.050**	-0.007**
Sargan (P-value)	0.15	0.16
Dif Sargan (P-value)		0.14

Note: Robust t-statistics are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance, respectively.

We have three conclusions from the above discussions. First, a dynamic model set-up is appropriate as the unit-root tests suggest accepting the nonstationarity of infant mortality, and the coefficient of the lagged infant mortality is statistically significant. Second, with internal instruments, system GMM is better able to correct the bias caused by weak internal instruments, and income becomes statistically significant in the infant mortality equation. Third, Sargan tests and differenced Sargan tests suggest that instruments dated *t*-3 (and earlier) are more appropriate than those dated *t*-2.

Education is not included in the above GMM and system GMM regressions. Pritchett

and Summers (1996) point out that ignoring education can overestimate the income impact on health, which suggests the merit in a sensitivity analysis. We do so here by including current education and estimating with system GMM.[60] We first assume education is an exogenous variable (Column 1 in Table 17) and then allow it to be endogenous with its lagged levels and differences as internal instruments (Column 2 in Table 17). We find that the impact of income on infant mortality is still statistically significant, but the educational impact is positive, which is opposite to our expectation that education reduces or has a negative impact on the infant mortality. This may be a result of assuming that education is exogenous. We see that controlling for the potential endogeneity of education barely changes the income effect but education's coefficient becomes statistically insignificant. We interpret this as evidence of the primacy of the direct effect of past infant mortality and income on current infant mortality and supportive of the interpretation that, at least in a dynamic model setup, the indirect effect working from contemporaneous education to infant mortality is negligible once income and past infant mortality are taken into account.

[60] In Section 2, to replicate PS, we use lagged level of (log) education as an explanatory variable, but in this GMM case Stata automatically uses first-order differencing to remove the country-effect, so in the end what we have in the reduced-form equation is the difference of education rather than the level of education, even when we put the lagged level of (log) education in the GMM specification. Given this, our model specification, after removing the country-effects, is not exactly comparable to the one in PS. At the same time, we also try the specification with lagged education in the GMM regression, finding that the results are very similar to those reported in Table 17.

Table 17: System GMM of infant mortality with education

Column #	1	2
	System GMM-Two Step exogenous education	System GMM-Two Step endogenous education
Lag of IMR	1.126**	1.097**
	(57.79)	(61.28)
GDP per capita	-0.101**	-0.079**
	(7.30)	(5.98)
Education (no lag)	0.072**	0.026
	(3.45)	(1.16)
1970	0.001**	0.0003**
1975	-0.020**	-0.021**
1980	-0.043**	-0.041**
1985	-0.020**	-0.020**
Sargan (P-value)	0.11	0.31
Dif Sargan (P-value)	0.38	0.45

Note: Robust t-statistics are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance, respectively.

To complete this analysis, Table 18 provides system GMM estimates for the life expectancy equation. There we see more ambiguity than found when using infant mortality as the dependent variable. First, although the impact of income has a positive sign, its magnitude and statistical significance are sensitive to the inclusion of education. Secondly, educational impact has a negative sign. These findings suggest that more consideration is required for the analysis of the income-life expectancy association. As discussed in Section 2.3, this may be because the data on life expectancy are not as reliable as that on infant mortality, and the income-life expectancy relationship, which can be influenced by other factors such as adult behaviour, is not as straight forward as the income-infant mortality relationship. In the system GMM framework, no matter whether we take education to be exogenous or endogenous, results with life expectancy as the dependent variable are less robust than those when modelling infant mortality.

Table 18: System GMM of life expectancy 1960-1985

Column #	1	2
	System GMM-Two Step	System GMM-Two Step *
Lag of IMR	1.001**	0.950**
	(79.37)	(52.7)
GDP per capita	0.006	0.034**
	(0.88)	(3.43)
Education		-0.027**
		(4.66)
1970	0.005**	0.006**
1975	-0.007**	-0.004**
1980	-0.004**	0.004**
1985	-0.015**	0.002**
Sargan (P-value)	0.00	0.02
Dif Sargan (P-value)	0.19	0.13

Note: Robust t-statistics are reported in parenthesis. Asterisks ** and * represent 5% and 10% statistical significance,.
* Education is considered to be endogenous and its lagged levels and differences are used as internal instruments.

5. Conclusion

The relationship between income and health is of interest both to economists and social scientists. The interaction of these two variables has been observed and been intensively studied based on post-1960 data. Since 1990, the world-wide improvement of health has slowed down as some African countries' health status deteriorated while some developing countries' health growth stagnated. This chapter re-examines the long-run relationship between income and health by updating Pritchett and Summers' (1996) study using data from 1960 to 2000. We find that the positive income effect on health still exists, but the magnitude of the income impact is lower. For instance, in the regression of the infant mortality, based on data from 1960 through 1985, we estimate an income elasticity in the interval [-0.402,-0.116], whereas when the sample is from 1960 through 2000, the range narrows to [-0.297,-0.091].

We also find that Kakwani's achievement index and Anand and Ravallion's health achievement indices, that aim to capture the nonlinearity of health improvement, are better at producing education impacts with intuitive signs in the income-health relationship.

In this chapter, we also explored the role of public health expenditure in explaining health. We found that the importance of public health expenditure is sensitive to the type of data used in the regressions. Specifically, when we estimate cross-sectional regressions, public health expenditure dominates income in explaining health, leading to income becoming statistically insignificant. In contrast, in regressions with a panel dataset, public expenditure is statistically insignificant. Even though the test results suggest insignificant impact of public health expenditure, this likely arises because differences across countries in the health expenditure-health relationship are smoothed out by our panel model estimation methods. Hence, our results with panels suggested the importance of estimating relationship for countries that share similar characteristics.

A final contribution of this chapter is to extend the models to a dynamic framework, which enabled us to estimate them using system GMM with internal instruments. This overcomes the difficulties in obtaining external instruments. Our results indicate that lagged infant mortality is important in explaining current infant mortality, system GMM is better than GMM to correct the bias arising from weak internal instruments, and a positive income-health association is statistically significant in our system GMM equations. One direction for future work is to extend the dynamic framework to potentially allow for cointegration among income, health and education.

CHAPTER 4:

GRANGER CAUSALITY BETWEEN HEALTH AND WEALTH: ANALYSING DATA IN SHORT PANELS

1. Introduction

A positive relationship between socio-economic status and health, the so-called "health-wealth gradient" (Adler et al. 1994), is an interesting topic for economists and sociologists. One reason for the interest in this gradient is that income and health are two dimensions of the human development index, a measure of human well-being used by the United Nations Development Programme. If there is feedback between health and wealth (i.e. a bi-directional association), then the simultaneous use of income generation and health enhancement programs and policies, such as direct income subsidies to the poor, school feeding programs, would have a larger impact on human development, as they will mutually impact each other.

Income and health have both improved dramatically worldwide over the past century. According to the World Bank, the world's real GDP per capita has increased by 117% over the past 40 years; the average life expectancy has increased by more than 15 years; infant mortality has fallen by 53% around the world; and child mortality at the age of five has also reduced by 54%. However, there are large disparities in the health status of rich and poor countries and the goal of worldwide health, declared at the

Alma-Ata Conference in 1978 as "by 2000, everyone should enjoy a level of health that will permit them to lead a socially and economically productive life" (as cited in Bloom Canning 2003, page 48), has not been achieved.[61] For example, Bloom and Canning (2003) show that a child born in Japan in 1999 could look forward to 74.5 years of healthy life (measured in disability-adjusted life expectancy (DALE)); children in another 23 countries expect to enjoy longer than 70 years of healthy life; in another 51 countries, however, children only expect less than 50 years of healthy life; in the least healthy countries DALEs are of less than 30 years. Health deterioration is also observed in some republics of the former Soviet Union, where life expectancies have been in long-term decline since the 1950s (Becker and Bloom 1998). At the same time, the income gap has also been widening: from 1960 to 2000 real income per capita rose by 159% for the richest country quartile whereas for the poorest country quartile this number was only 51%. With these disparities in income and health, there are still many challenges to improving the health and wealth of millions of people around the world.

Despite some discrepancies shown by worldwide data, most of the microeconomic and macroeconomic studies have found a positive association between health and wealth. This is good news as this means that any policy aimed at improving one of the variables, will have a positive impact on the other as well. Below we summarize some mechanisms through which the link between health and wealth can work.

[61] In 1978, at the Alma-Ata Conference, ministers from 124 countries met together with representatives of the World Health Organization (WHO) and UNICEF and made a declaration calling for "health for all by the year 2000". This detailed declaration is available in the website of the World Health Organization: http://www.who.int/hpr/NPH/docs/declaration_almaata.pdf.

Health leads to wealth

At the country level, health is a major input in the generation of income or wealth according to the human capital approach, while at the individual level, health is essential to ensuring that people achieve a more productive life. The link from health to wealth appears to operate through a number of mechanisms, including via investment, demography, education, and the labour market. Healthy people can expect to live longer and save more for their old age, which increases the amount of investment available to the domestic economy (Chen et al. 2008, Bloom et al. 2002, Zhang et al. 2003, Chakraborty 2004). Improvements in health also set off a demographic transition with lower fertility and mortality, but the lag between declines in mortality and fertility results in a "baby-boom" generation, which raises economic growth when they enter the workforce (Bloom and Canning 2003). On the other hand, as fertility falls, parents are likely to invest more in their children's health and education. Healthy children attend more school and are better able to learn with fewer nutritional deficiencies, fewer infectious diseases and fewer disabilities (Ruger et al. 2001, Croix and Licandro 1999, Kalenli-Ozcan et al. 2000). With better health, education thus acts as a fundamental driver of human development and economic growth. From the viewpoint of the labour market, healthier workers are physically and mentally more robust, and likely to be more productive with higher wages (Schultz 2002, Thomas and Strauss 1997).

Wealth leads to health

The classical view of the relationship between health and economic development is that wealth leads to health, with health an output of the development process. By plotting multiple countries' income per capita and life expectancy in 1900, 1930, and 1960, Preston (1975) proposed a concave, positive empirical relationship between

wealth and health. Empirical studies, including regressions using cross-section data, (e.g., Flegg 1982, Subbarao and Randy 1995, Zakir and Wunnava 1999) and estimations using panel data (e.g., Pritchett and Summers 1996, Easterly 1999, Issa and Ouattara 2005) find positive impacts of wealth on aggregated health status (e.g., infant mortality and life expectancy). Other variables relating to income/wealth are also shown to be important in explaining health. For example, Anand and Ravallion (1993), Pritchett (1997) and Filmer and Pritchett (1999) believe that public/private health expenditure, income equality, and gender equality are also significant in improving health.

However, the empirical health-wealth association estimated by regressions using cross-sectional, time series, or panel data may not necessarily fully explain the causal links running between these two variables. Specifically, there is little indication of causality between health and wealth using country-level data. For example, there is little evidence that periods of rapid health improvement follow periods of high income growth (Bloom and Canning 2003).

To assist in exploring the causal relationships between variables, Granger (1969, 1980) proposed a causality test based on the precept that for a time series, "the cause preceded the effect and a causal series had information about the effect that was not contained in any other series according to the conditional distributions." (Granger 2003, page 69) The empirical association found between health and wealth suggests the existence of co-movement between these variables but it need not lead to us finding that using wealth/health produces a superior forecast of health/wealth. Recently, researchers have been studying the causal relationships between wealth and health within models that have a similar set-up to applying Granger's causality test.

Mayer (2001) used a health indicator (probability of survival) and income data in 18 Latin American countries, from 1950 to 1990, to confirm the health impact on economic growth. Further support for uni-directional causality running from health to wealth was found for elderly couples in the US (Michaud and Soest 2004). A statistically significant income effect on health was found in the duration model developed by Frijters et al. (2005) using panel data for post reunified East German households between 1991 and 1999. We aim to contribute to this literature.

This chapter applies three panel causality tests, which are extensions of the classical Granger causality test to panel data, to examine for health-wealth causality using 105 developing countries' data from 1960 to 2000 at five-year frequencies.[62] We compare test results from these methodologies. The first method proposed by Holtz-Eakin et al. (1988) has been adopted by many studies. Although the authors allow the causal relationship to vary over time, they do not allow for heterogeneity in the causality across countries. Approaches by Hurlin and Venet (2003), Hurlin (2004, 2005), Weinhold (1999) and Nair-Reichert and Weinhold (2001) that have yet to be extensively used in empirical research, allow for unit or country level heterogeneity. Hurlin and Venet (2003) and Hurlin (2004, 2005) assume that any present causality is heterogeneous but *fixed* across individual units, while Weinhold (1999) and Nair-

[62] We cut off countries that have real GDP per capita larger than I$6000 in 1985. This is consistent with the definition of developing countries that we used in Chapter 3. There are 105 developing countries in our sample: Afghanistan, Algeria, Angola, Argentina, Bangladesh, Benin, Bolivia, Botswana, Brazil, Burkina Faso, Burundi, Cameroon, Cape Verdi, Central African Rep, Chad, Chile, China, Columbia, Comoros, Congo Dem. Rep., Costa Rica, Cyprus, Dominica, Dominica Rep, Ecuador, Egypt, El Salvador, Ethiopia, Fiji, Gabon, Gambia, Ghana, Greece, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hungary, India, Indonesia, Iran, Iraq, Ivory Coast, Jamaica, Jordan, Kenya, Korea Republic of, Lesotho, Liberia, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mauritius, Mexico, Morocco, Mozambique, Myanmar, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Panama, P. New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Rwanda, Senegal, Seychelles, Sierra Leone, Solomon Islands, South Africa, Sri Lanka, St. Lucia, St. Vincent, Sudan, Suriname, Swaziland, Syria, Taiwan, Tanzania, Thailand, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Uganda, Uruguay, Vanuatu, Venezuela, Western Samoa, Yemen, Yugoslavia, Zaire, Zambia, and Zimbabwe.

Reichert and Weinhold (2001) allow the coefficients of the causal variables to be random. This latter approach introduces extra flexibility.

One issue with the last two methods is that they involve inferences from estimations in individual countries. With a short panel (few observations in the time series dimension), there might not be enough observations to allow estimations to proceed, or to produce "good" estimates of test statistics. Hence, to enable our models to be estimated, we simplify some of the specifications when we apply these two approaches in our work. For example, in some of the following tests we are limited to using only one lag in the models.

In addition to undertaking the three panel causality tests discussed above, we undertake sensitivity analyses using sub-groups of data for 57 middle-income countries and 43 low-income countries.[63] The objective of these sensitivity tests is not only to study whether the health-wealth causality in middle-income countries or low-income countries is different from that observed for all developing countries, but also to evaluate the robustness of the above three panel causality tests with short panels.

The rest of this chapter is organized as follows. Section 2 discusses the three different panel causality tests that are adopted in this chapter and outlines the data used in this work. Section 3 reports and compares results from these three methods. Section 4 presents the outcomes from our sensitivity tests. Section 5 concludes.

[63] The World Bank uses gross national income (GNI) per capita as criteria to classify economies. Economies are divided according to 2006 GNI per capita, calculated using the World Bank Atlas method. The groups are: low income, $905 or less; lower middle income, $906 - $3,595; upper middle income, $3,596 - $11,115; and high income, $11,116 or more. Full details are available from www.worldbank.org.

2. Methodology and Data

Of the relationships between socioeconomic variables discovered by empirical regressions, some may seem to be deeper in some sense than "mere" correlation. Granger (1969) specified a causality concept based on the precept that for a time series, the *cause* preceded the *effect* and a causal series had information about the *effect* that was not contained in any other series according to their conditional distributions. In another words, the *cause* can produce a superior forecast of the *effect*. This interpretation has become the common definition of "Granger Causality" in many studies. More specifically, the Granger causality test is usually undertaken via a vector autoregression (VAR) system. To illustrate, consider two series X_t and Y_t , assumed to be stationary with zero means. A simple VAR model between these two variables is:

$$\begin{cases} Y_t = \sum_{j=1}^{m} a_j Y_{t-j} + \sum_{j=1}^{n} b_j X_{t-j} + \varepsilon_t \\ X_t = \sum_{j=1}^{o} c_j Y_{t-j} + \sum_{j=1}^{p} d_j X_{t-j} + \eta_t \end{cases},$$

where ε_t and η_t are uncorrelated white-noise; i.e. $E(\varepsilon_t \varepsilon_s) = E(\eta_t \eta_s) = 0$ for $s \neq t$. If all b_j, j = 1,…,n are zero, then we say X is not a cause of Y and does not help in predicting Y in the Granger sense. Similarly, Y is causing X if some c_j, j = 1,…p are not zero. These coefficient restrictions can be tested by classical F and Wald tests. When $E(\varepsilon_t \eta_s) = 0$ for all t and s, and there are no contemporaneous terms on the right-hand-side of the equations, that is, X_t does not show up in the equation of Y_t and Y_t is not included in the equation of X_t, the above VAR system can be estimated by two individual Ordinary Least Squared (OLS) regressions. Otherwise, one has to

estimate the equations as a system. To simplify our work, we also restrict $m = o = n = p$, a common assumption in the VAR literature.

Recently, Granger's causality test within the VAR framework has been extended to applications with panel data sets. Panel data provide more observations, increasing the degrees of freedom and reducing the collinearity among explanatory variables (Hsiao 1995, 2000); they help to analyze a number of important economic questions that cannot be addressed using cross-sectional or time-series data alone; they also generate more accurate predictions of individual outcomes than time-series data alone, because in a panel set-up "an individual's behavior can be learned by observing the behavior of others, in addition to the information on that individual's behavior" (Hsiao 2003 page 7). However, conventional estimation approaches with time-series data, from which we undertake the Granger causality tests, encounters a challenge with panel data: how to deal with heterogeneity among the cross-sectional units. This heterogeneity not only refers to the heterogeneous country-effects that can be captured by fixed effects or random effects, but also the heterogeneous causal relationship between X and Y that may be different across the countries. In this paper, we allow for such heterogeneity by applying three methodologies, proposed by Holtz-Eakin, Newey, and Rosen (1988), Hurlin and Venet (2003), Hurlin (2004, 2005), Weinhold (1999) and Nair-Reichert and Weinhold (2001). These three approaches allow for heterogeneity across countries, but the latter two methods attempt to allow each country to have differing causal relationships.

2.1 Holtz-Eakin, Newey and Rosen (1988)

A simple bivariate panel Granger causality test can be undertaken using the model:

$$y_{it} = \alpha_i + \sum_{k=1}^{K} \gamma_{kt} y_{i,t-k} + \sum_{k=1}^{K} \beta_{kt} x_{i,t-k} + \varepsilon_{it} \qquad i = 1,...,N; t = K+1,...T \qquad (4.1)$$

The error term ε_{it} satisfies the orthogonality conditions:

$$E(y_{is}\varepsilon_{it}) = E(x_{is}\varepsilon_{it}) = E(\alpha_i \varepsilon_{it}) = 0 \text{ , for } s < t. \qquad (4.2)$$

This specification allows the time series relationship of *x* and *y* to vary across each country and over time: first, it introduces a unit (here, country) effect, α_i, which translates in practice into a unit specific intercept in equation (4.1); secondly, it relaxes the assumption of "time stationarity" (Holtz-Eakin et al. 1988, page 1373) that the lag coefficients γ_{kt} and β_{kt} are constant for each *t*. The lag length *K* can be arbitrarily large but it cannot be too big as there might not be enough observations to estimate all of the coefficients γ_{kt} and β_{kt}. In this paper, we restrict *K* to be a finite number and take a value that enables us to estimate equation (4.1).

A first-order differencing can remove the unobserved individual effect α_i from (4.1) to give:[64]

$$y_{it} = \sum_{k=1}^{K+1} c_{kt} y_{i,t-k} + \sum_{k=1}^{K+1} d_{kt} x_{i,t-k} + v_{it} \qquad i = 1,...,N; t = K+2,...T \qquad (4.3)$$

Here

$$\begin{aligned}
c_{1t} &= 1 + \gamma_{1t}, \\
c_{kt} &= \gamma_{kt} - \gamma_{k-1,t}, \\
c_{K+1,t} &= -a_{Kt}, \\
d_{1t} &= \beta_{1t}, \\
d_{kt} &= \beta_{kt} - \beta_{k-1,t}, \\
d_{K+1,t} &= -\beta_{Kt} \\
v_{it} &= \varepsilon_{it} - \varepsilon_{it-1}.
\end{aligned}$$

[64] This is a simplified case of Holtz-Eakin et al.'s approach. In the general case, they suggest the use of a quasi-differencing transformation rather than first differencing.

Given the orthogonality conditions, the error term of the transformed equation (4.3) satisfies:

$$E(y_{is}v_{it}) = E(x_{is}v_{it}) = 0\,, \quad s < t-1. \tag{4.4}$$

It is easy to see that this first-order differencing also causes a problem: the error term, v_{it}, in equation (4.3) is correlated with the right-hand-side explanatory variable, y_{it-1}. Anderson and Hsiao (1982) suggest using instrumental variables (IVs) to estimate the autoregressive equation (4.3). The vector of instrumental variables that is available to identify the parameters of equation (4.3) is: $Z_{it} = [y_{it-2}, \ldots, y_{i1}, x_{it-2}, \ldots, x_{i1}]$.

At the same time, a necessary condition for identification is that there are at least as many instrumental variables as right-hand-side variables of (4.3). Because there are a total of $2(K+1)$ right-hand variables and the dimension of Z_{it} is $2t-4$, this order condition reduces to $t > K+2$.

For convenience, we rewrite the approach of Holtz-Eakin et al. (1988) by stacking up equation (4.3) in one time period for all cross-sections, with $Y_t = [y_{1t} \ldots y_{Nt}]'$, $X_t = [x_{1t} \ldots x_{Nt}]'$, and $V_t=[v_{1t}, \ldots, v_{Nt}]'$. Then, let $W_t = [Y_{t-1}\, Y_{t-2} \ldots Y_{t-K-1}\, X_{t-1}\, X_{t-2} \ldots X_{t-K-1}]$ be the $N \times 2(K+1)$ matrix of right-hand-side variables of equation (4.3), and let $B_t = [c_{1t} \ldots c_{K+1,t}\, d_{1t} \ldots d_{K+1,t}]'$ be the $2(K+1) \times 1$ vector of coefficients for the equation. Then equation (4.3) can be written as:

$$Y_t = W_t B_t + V_t, \quad t = K+3, \ldots, T.$$

To further combine all data points in our matrix form, we stack up the above t-period equations ($t = K+3,\ldots,T$):

$$Y = WB + V, \tag{4.5}$$

where

$$Y = [Y_{K+3}' \ \ldots \ Y_T']',$$

$$((T\text{-}K\text{-}2)N \times 1)$$

$$B = [\mathrm{B}'_{\mathrm{K}+3}, \ldots, B_T']'$$

$$((T\text{-}K\text{-}2)(2K+2) \times 1)$$

$$W = \mathrm{diag}\,[\mathrm{W}'_{\mathrm{K}+3} \ \ldots \ \mathrm{W}'_{\mathrm{T}}]'$$

$$((T\text{-}K\text{-}2)N \times (T\text{-}K\text{-}2)(2K+2))$$

$$V = [\mathrm{V}'_{\mathrm{K}+3} \ \ldots \ \mathrm{V}'_{\mathrm{T}}]'$$

$$((T\text{-}K\text{-}2)N \times 1).$$

The matrix of variables which qualify for use as instrumental variables is defined as:

$$Z = \mathrm{diag}\,[Z_{K+3} \ \ldots \ Z_T],$$

where $Z_t = [Y_{t-2} \ \ldots \ Y_1 \ X_{t-2} \ \ldots \ X_1]$, $t = K+3, \ldots, T,$ and the dimension of Z_t increases as t increases.

The orthogonality conditions (4.4) ensure that:

$$\operatorname*{plim}_{N\to\infty}(Z'V)/N = \operatorname*{plim}_{N\to\infty}\begin{bmatrix} (Z'_{K+3}V_{K+3})/N \\ \vdots \\ (Z'_T V_T)/N \end{bmatrix} = 0 .$$

So we can pre-multiply the equation (4.5) by Z' and the IV estimates can be obtained by estimating the following model:

$$Z'Y = Z'WB + Z'V. \qquad (4.6)$$

As the error term in equation (4.6) may be heteroskedastic and autocorrelated, Holtz-Eakin et al. (1988) suggest that we form a consistent and (asymptotically) efficient instrumental variables estimator by applying generalized least squares (GLS) to the equation. Such an estimator requires knowledge of the covariance matrix of the

(transformed) disturbances, ZV. To estimate this covariance matrix, two-stage least squares (2SLS) regressions are undertaken for each time period $t = K+3,\ldots T$, and the residuals from these 2SLS regressions are then used to construct a consistent estimator of the covariance matrix of ZV.

An alternative approach is via GMM estimation. For two reasons we adopt this methodology over the two-stage approach suggested by Holtz-Eakin et al.: first, GMM estimation is now the common estimation principle with dynamic panel models with autoregressive terms and endogenous explanatory variables (e.g., Arellano and Bond 1991, Blundell and Bond 2000, Bond et al. 2001, Bond 2002); secondly, the GMM estimator tends to be more robust than the GLS estimator as it does not require information on the exact distribution of the disturbances and so is able to deal with unknown forms of heteroskedasticity and serial correlation.

Testing for causality[65] between X and Y is equivalent to testing zero constraints on the coefficients in equation (4.3). Prior to undertaking this, we first check whether the coefficients on the lagged explanatory variables are "time stationary", that is, whether c_{kt} and d_{kt} are constant for each t. If they are, then the number of coefficients can be reduced to be $2K+2$ and the equation in each time period is:

$$y_{it} = \sum_{k=1}^{K+1} c_k y_{i,t-k} + \sum_{k=1}^{K+1} d_k x_{i,t-k} + v_{it} \qquad i = 1,\ldots,N; t = K+2,\ldots T\,. \tag{4.7}$$

The hypothesis of whether x is non-causal for y then corresponds to testing whether all the d_k are simultaneously zero. If the coefficients on the lagged right-side explanatory variables are not time stationary but vary across periods, the causality test is undertaken on a longer list of coefficients, d_{kt}. To test for "time-stationarity", we

[65] The null hypothesis corresponds to non-causality, but we follow the terminology popular in the literature of referring to the test as one for causality.

examine whether in equation (4.3) $c_{kt} = c_{ks}$ and $d_{kt} = d_{ks}$, for any $k = 1,\ldots,K$ and $t \neq s$. As N grows, the test statistics for "time stationarity" and non-causality have chi-squared distributions, under their respective null hypotheses, with degrees of freedom equal to the number of restrictions.[66]

2.2 Hurlin and Venet (2003), Hurlin (2004, 2005)

To deal with possible heterogeneity in the causal patterns over time, Holtz-Eakin et al. (1988) introduce varying coefficients on the lagged explanatory variables within a system of equations, with the data prepared in a way with N as the faster index.[67] However, with panel data we are usually more interested in allowing for heterogeneity across countries rather than over time. This can be accommodated using the same framework, by reshaping the dataset into N sub-stacks where T is the faster index so that the first-stage regressions are based on estimations for each country across all time periods. The estimation methods, which we consider in this section and the next section, have been proposed to allow for country specific heterogeneity but not time-specific heterogeneity.

Hurlin and Venet (2003) and Hurlin (2004, 2005) propose a panel Granger non-causality test with heterogeneous panel data models with fixed coefficients. Under the null hypothesis, there is no causal relationship for all countries in the panel. The authors term this the Homogenous Non-Causality (HNC) hypothesis. An alternative hypothesis is for Heterogeneous Non-Causality (HENC): there are assumed to be two

[66] These two tests can be easily undertaken in EViews 5.1.
[67] If panel data are rearranged as T stacks and each stack has N observations for N individual countries, we say this panel dataset has N as the faster index and T as the slower index. If panel data are reshaped as N stacks and each stack has T observations for T years' data, we say this panel dataset has T as the faster index and N as the slower index.

subgroups of units, one with causal relationships from *x* to *y*, but not necessarily with the same data generation process (DGP), and the other subgroup where there is no causal relationship from *x* to *y*. An example of HENC is that in equation (4.8) the coefficient $\beta_i^{(k)}$ is equal to be zero for some *i* but is different from zero for other values of *i*. Hence, this HENC null hypothesis further distinguishes the heterogeneity in the DGP and the heterogeneity of causal relationships; while in Holtz-Eakin et al. (1988) under the alternative hypothesis causality may exist in all countries (the assumed alternative hypothesis) but may also be only present for some countries – their approach cannot distinguish between such possibilities.

Hurlin (2004) considers the following autoregressive model:

$$y_{it} = \alpha_i + \sum_{k=1}^{K} \gamma_i^{(k)} y_{i,t-k} + \sum_{k=1}^{K} \beta_i^{(k)} x_{i,t-k} + \varepsilon_{it} \tag{4.8}$$

with $K \in \mathrm{N}$, $\gamma_i = [\gamma_i^{(1)},...,\gamma_i^{(K)}]'$, and $\beta_i = [\beta_i^{(1)},...,\beta_i^{(K)}]'$. Individual effects α_i are assumed to be fixed effects rather than random effects. Similar to Holtz-Eakin et al. (1988), Hurlin (2004) assumes that the lag orders *K* are identical for all cross-sectional units in the panel, but the fixed autoregressive parameters $\gamma_i^{(k)}$ and regression coefficients slopes $\beta_i^{(k)}$ are allowed to be different across countries. However, if we restrict $\beta_i = \beta_j, \forall i \neq j$, then the estimators of β_i are biased (Pesaran and Smith 1995) and so Hurlin's specification does not suffer from this bias.

As this approach allows for differing causality across countries, Hurlin and Venet (2003), and Hurlin (2004, 2005) propose a test statistic that is the average of the individual non-causality Wald statistics across all the *N* countries. Intuitively, this averaged Wald statistic considers the averaged/mean non-causality between *x* and *y*

across all countries. This idea is analogous to the panel unit root tests (e.g., Im, Pesaran, and Shin 2003), where inferences from the country-specific unit root tests are incorporated together to draw a conclusion on the unit root properties of the panel. The averaged Wald statistic associated with the HNC null hypothesis is:

$$W_{N,T} = \frac{1}{N}\sum_{i=1}^{N} W_{i,T} \tag{4.9}$$

where $W_{i,T}$ denotes the individual Wald statistics for the i^{th} country from testing $H_0 : \beta_i^{(1)} = \beta_i^{(2)} = ... = \beta_i^{(K)} = 0$. The general form of this Wald test for the i^{th} country, based on model (4.8), can be expressed by using the following "stack-up" matrices. Define

$$y_i^{(k)} = \underset{(T-t+1)\times 1}{\begin{bmatrix} y_{i,t-k} \\ \vdots \\ y_{i,T-k} \end{bmatrix}}, \quad x_i^{(k)} = \underset{(T-t+1)\times 1}{\begin{bmatrix} x_{i,t-k} \\ \vdots \\ x_{i,T-k} \end{bmatrix}}, \quad \varepsilon_i = \underset{(T-t+1)\times 1}{\begin{bmatrix} \varepsilon_{i,t-k} \\ \vdots \\ \varepsilon_{i,T-k} \end{bmatrix}},$$

$$Z_i = \underset{(T-t+1)\times(2K+1)}{\left[e\, y_i^{(1)}\, y_i^{(2)} \ldots y_i^{(K)}\, x_i^{(1)}\, x_i^{(2)} \ldots x_i^{(K)}\right]},$$

where e is a vector of ones, and $\theta_i = [\alpha_i\ \gamma_i'\ \beta_i']'$ to be the vector of parameters for the model. The non-causality null hypothesis for the country-specific Wald test can then be expressed in the classical form of $R\theta_i = 0$ where R is a $K\times(2K+1)$ matrix with $R = [0{:}I_K]$. The Wald statistic is then:

$$W_{i,T} = \frac{\hat{\theta}_i' R'[R(Z_i'Z_i)^{-1}R']^{-1} R\hat{\theta}_i}{\hat{\varepsilon}_i'\hat{\varepsilon}_i/(T-2K-1)} \tag{4.10}$$

where $\hat{\theta}_i$ is the estimator of the parameter vector θ_i and $\hat{\varepsilon}_i$ is the vector of associated residuals.

Under the null hypothesis of non-causality, the asymptotic distribution of each country-specific Wald statistic is chi-square with *K* degrees of freedom:

$$W_{i,T} \underset{T\to\infty}{\overset{d}{\to}} \chi^2(K) \qquad \forall i = 1,...,N$$

These individual Wald statistics are independent if the residuals across the countries are independent. Then the null distribution of the averaged Wald statistic can be deduced from a standard Lindberg-Levy central limit theorem (Grimmett and Stirzaker 2001):

$$Z_{N,T} = \sqrt{\frac{N}{2K}}(W_{N,T} - K) \underset{T,N\to\infty}{\overset{d}{\to}} N(0,1) \text{ , where } W_{N,T} = 1/N \sum_{i=1}^{N} W_{i,T} \text{ .}$$

If the panel has finite *T*, the individual Wald statistic $W_{i,T}$, i = 1, …, *N*, are independently but not identically distributed with finite second order moments. Based on the Lyapunov central limit theorem and results of Magnus (1986), Hurlin and Venet (2003), and Hurlin (2004, 2005) prove that when *T>5+2K,* the first and second moments of the individual Wald statistics exist and a finite-sample standardized statistic $\tilde{Z}_{N,T}$, which still has an asymptotic normal distribution is given by:

$$\tilde{Z}_{N,T} = \sqrt{\frac{N}{2K} \times \frac{T-2K-5}{T-K-3}} \times \left(\frac{T-2K-3}{T-2K-1} W_{N,T} - K\right) \underset{N\to\infty}{\overset{d}{\to}} N(0,1) \text{ .}$$

However, in a finite sample with small *T* and *N*, the average statistic $W_{N,T}$ and the standardized statistic $\tilde{Z}_{N,T}$ have finite sample null distributions that may markedly differ from their respective asymptotic distributions. Applying the approach proposed by Im, Pesaran and Shi (2003), Hurlin (2004) suggests computing an approximate critical value for $\tilde{Z}_{N,T}$, for a finite panel with fixed *T* and *N,* using:

$$c_{N,T}(\alpha) = z_\alpha \times \frac{T-2K-1}{T-2K-3} \times \sqrt{\frac{2K(T-K-3)}{N(T-2K-5)}} + \frac{K(T-2K-1)}{T-2K-3},$$

where α is the significance level and z_α is the critical value of a standard normal variable corresponding to the α significance level. If this methodology is applied with an unbalanced panel[68], following Hurlin and Venet (2003) and Hurlin (2004), we then revise the finite-sample standardized statistic and approximate critical value as:

$$\tilde{Z}_{N,T} = \frac{\sqrt{J}\left[W_{N,T} - \frac{1}{J}\sum_{i=1}^{J} K \times \frac{T_i - 2K - 1}{T_i - 2k - 3} \right]}{\sqrt{\frac{1}{J}\sum_{i=1}^{J} 2K \times \frac{(T_i - 2K - 1)^2 \times (T_i - K - 3)}{(T_i - 2K - 3)^2 \times (T_i - 2K - 5)}}},$$

$$c_{N,T}(\alpha) = z_\alpha \times \frac{\sum_{i=1}^{J}(T_i - 2K - 1)}{\sum_{i=1}^{J}(T_i - 2K - 3)} \times \sqrt{\frac{2K / J\sum_{i=1}^{J}(T_i - K - 3)}{\sum_{i=1}^{J}(T_i - 2K - 5)} + \frac{K\sum_{i=1}^{J}(T_i - 2K - 1)}{\sum_{i=1}^{J}(T_i - 2K - 3)}},$$

where T_i is the number of observations in country i, and J is the total number of countries that satisfy $T_i > 2K+5$. As our panel dataset covers 105 countries, and each country's time series is no longer than 9 years, we rely on the finite-sample standardized statistic and its approximate critical values.

2.3 Weinhold (1999), Nair-Reichert and Weinhold (2001)

The panel Granger causality test proposed by Weinhold (1999) and Nair-Reichert and Weinhold (2001) is similar to that considered by Hurlin (2004), whereby they both consider inference for each country and use a statistic that is an "average". The key difference is that in Weinhold (1999) and Nair-Reichert and Weinhold (2001) the coefficients on the lagged causal variables are assumed to be random and normally distributed across countries while they are assumed to be fixed in Hurlin and Venet

[68] With an unbalanced panel, the number of observations differs across countries. That is, T_i varies across i.

(2003) and Hurlin (2004, 2005). This "mixed fixed-random" (MFR) approach was originally developed by Hsiao (1989) in the context of a non-dynamic model.

The model specification is identical to equation (4.8) except we assume that $\beta_i^{(k)} = \bar{\beta}^{(k)} + \eta_i^{(k)}$ with $\beta_i^{(k)} \sim N(\bar{\beta}^{(k)}, v^{(k)^2})$ for all $i = 1,\ldots, N$ and $k = 1,\ldots, K$. That is, the coefficients on the lagged causal variables are assumed to be random, following a normal distribution with mean $\bar{\beta}^{(k)}$ and variance $v^{(k)^2}$. There are several reasons for this random-coefficient assumption: first, random coefficients for the causal variable allows for heterogeneity to exist in the causal relationships across countries; secondly, the use of random coefficients is more appropriate if β_i is "viewed as random draws from a common population rather than as coming from a heterogeneous population" (Hsiao 2003, page 149)[69]; thirdly, compared to a model with fixed (but varying across countries) coefficients, the use of random coefficients reduces the number of parameters to be estimated; fourthly, the random coefficients introduce extra flexibility to interpreting the probability of finding causality for any particular country in the panel (Granger 2003), and the estimate of the variance of the random coefficients across countries can be used as a diagnostic tool to judge the range of the relationship between the economic variables *x* and *y*.

Weinhold (1999) extends the Bayesian method of Hsiao (1989) to obtain an estimator of the mean of the random coefficients $\bar{\beta} = [\bar{\beta}^{(1)}, \ldots, \bar{\beta}^{(K)}]'$. Specifically, she proposes the estimator:

[69] This is analogous to the difference between fixed and random effects in a classical panel model.

$$\hat{\bar{\beta}} = \left\{ \sum_i X_i' \Phi_i^{-1} X_i - \sum_i X_i' \Phi_i^{-1} Y_i (Y_i' \Phi_i^{-1} Y_i)^{-1} Y_i' \Phi_i^{-1} X_i \right\}^{-1} \times \left\{ \sum_i X_i' \Phi_i^{-1} y_i - \sum_i X_i' \Phi_i^{-1} Y_i (Y_i' \Phi_i^{-1} Y_i)^{-1} Y_i' \Phi_i^{-1} y_i \right\},$$

where $X_i = \begin{bmatrix} x_{i,t-1} & & x_{i,t-K} \\ \vdots & \vdots & \vdots \\ x_{i,T-1} & & x_{i,T-K} \end{bmatrix}$, $Y_i = \begin{bmatrix} y_{i,t-1} & & y_{i,t-K} \\ \vdots & \vdots & \vdots \\ y_{i,T-1} & & y_{i,T-K} \end{bmatrix}$, $\Phi_i = X_i \Delta_{11} X_i' + \sigma_i^2 I_{T-K}$. In the formula for Φ_i, σ_i^2 is the sample variance calculated from the OLS residuals $\hat{\varepsilon}_i$ of individual countries: $\sigma_i^2 = \hat{\varepsilon}_i' \hat{\varepsilon}_i / ((T-K)-(1+2K))$, and Δ_{11} is a sub-matrix of the estimated variance-covariance matrix corresponding to the random coefficients: $\Delta_{11} = \Delta[(K+2):(2K+1),(K+2):(2K+1)]$, where Δ is a consistent estimate of the variance-covariance matrix using Swamy's (1970) approach: $\Delta = \sum_i \frac{1}{N-1} [(\hat{\theta}_i - \bar{\hat{\theta}})(\hat{\theta}_i - \bar{\hat{\theta}})']$, and $\hat{\theta}_i$ is the OLS estimator of equation (4.8) for the i^{th} country. The variances for each $\beta_i^{(k)}$, $v^{(k)2}$, is estimated by the diagonal elements of the sub-matrix Δ_{11}. This variance, as explained by Weinhold (1999, page6), can be used as a diagnostic tool to determine the extent of heterogeneity in the causal relationship: "If the estimated variance is quite large relative to the coefficient estimates, this is a signal of significant heterogeneity in the panel. Rather than continuing with the panel data analysis at that point it may be prudent to rather investigate the heterogeneity in greater detail to assess the appropriateness of the specification and/or the pooling of the data." Individual random effects and individual fixed effects, $\beta_i = [\beta_i^{(1)},...,\beta_i^{(K)}]'$ and $\gamma_i = [\gamma_i^{(1)},...,\gamma_i^{(K)}]'$, can also be retrieved from the mixed fixed-random model by:

$$\hat{\beta}_i = \left\{ \frac{1}{\sigma_i^2}[X_i'X_i - X_i'Y_i(Y_i'Y_i)^{-1}Y_i'X_i] + \Delta_{11}^{-1} \right\}^{-1} \times \left\{ \frac{1}{\sigma_i^2}[X_i'X_i - X_i'Y_i(Y_i'Y_i)^{-1}Y_i'X_i\beta_i^* + \Delta_{11}^{-1}\hat{\bar{\beta}}] \right\}$$

$$\hat{\gamma}_i = (Y_i'Y_i)^{-1}Y_i'(y_i - X_i\hat{\beta}_i),$$

where β_i^* is the individual OLS estimator of the random coefficients for country *i* .

Besides estimating the mean $\bar{\beta}^{(k)}$ and the variance $v^{(k)2}$, the standard error for the estimator $\hat{\bar{\beta}}$ can also be generated. The way to calculate the standard error is analogous to the calculation of coefficient standard errors in a classical OLS regression: define $W = \begin{bmatrix} X_i & Y_i \\ ... & \\ X_N & Y_N \end{bmatrix}$ and $\hat{\varepsilon} = [\hat{\varepsilon}_1, ..., \hat{\varepsilon}_N]'$, then a $K \times 1$ vector that is formed by the first K diagonal elements of $\sigma^2(W'W)^{-1}$, where $\sigma^2 = \hat{\varepsilon}'\hat{\varepsilon}/((T-K)N - ((1+K)N + K))$, is the estimated standard error of $\hat{\bar{\beta}}$. Weinhold also proposes a "rule of thumb" for interpreting the estimated standard error of $\hat{\bar{\beta}}$ from the MFR model in the context of causality testing. She proposes that the following interval be interpreted as corresponding to observations on $\beta_i^{(k)}$, $\forall i = 1,..., N$ and $k = 1,\ldots, K$, that are not statistically significantly different from zero:

$$prob(\text{non - causality}) = prob(0 - 2\sqrt{N}se_{\hat{\bar{\beta}}^{(k)}} \le \beta_i^{(k)} \le 0 + 2\sqrt{N}se_{\hat{\bar{\beta}}^{(k)}}).$$

Because the approach assumes $\beta_i^{(k)} \sim N(\bar{\beta}^{(k)}, v^{(k)2})$, this interval can be standardized into a standard normal interval:

$$prob\left(\frac{-2\sqrt{N}se_{\hat{\bar{\beta}}^{(k)}} - \bar{\beta}^{(k)}}{v^{(k)}} \le z \le \frac{2\sqrt{N}se_{\hat{\bar{\beta}}^{(k)}} - \bar{\beta}^{(k)}}{v^{(k)}}\right).$$

However, as this "rule of thumb" is quite inexact, in this chapter we use the information from the estimated variance $v^{(k)2}$ and the standard error of the estimated mean coefficient $se_{\hat{\bar{\beta}}^{(k)}}$ to gain a general idea of the degree of heterogeneity in the causal relationship: if these two numbers are large compared to the estimated mean coefficient, then we take that as evidence of heterogeneity across countries.

Simulations in Weinhold (1999) show that the estimate of the variance $v^{(k)2}$ in MFR has a serious upward bias in a short panel, while the estimate of the mean coefficient $\bar{\beta}^{(k)}$ is more robust. In particular, in three finite samples, one with $T = 5$ and $N = 80$, one with $T = 5$ and $N = 40$, the other with $T = 5$ and $N = 20$, the estimates of $v^{(k)2}$ are 10, 9.4, and 7.7 times larger than the one for a sample with $T = 50$ and $N = 80$. Further, the standard error of the estimated mean coefficient, $se_{\hat{\bar{\beta}}^{(k)}}$, also seems to suffer from an upward bias in panels with short time series . For example, for the previously cited three finite samples, the estimated standard errors are 2, 3, and 4 times that from the sample with $T = 50$ and $N = 80$. Weinhold (1999, page 8) states "… researchers using the MFR model with short time series should take into account that the variance estimate is highly biased upwards. This would affect the diagnostic ability of the model if the bias is not taken into account." Therefore, later when we implement this causality test, we calculate an adjusted variance by dividing the estimated variance by the square root of 10 to remove the potential upward bias in a short panel, and adjust the estimated standard error of the mean coefficient by dividing it by 2.[70] The adjusted standard error is used to test for non-causality.

[70] The maximum time span of our panel data for using the approach of Weinhold (1999) and Nair-Reichert and Weinhold (2001) is $T = 8$ (the first observation is dropped as there is a lagged term in the model specification). The number of countries that are covered by these tests vary (See Appendix 2 and

2.4 Data

Infant mortality data on 105 developing countries are taken from the WDI at five-year frequencies from 1960 to 1985.[71] However, as infant mortality data for 1965 and 1975 are not available, we use simple averages of the data from 1962 and 1967, and 1972 and 1977, respectively, to estimate these two years' observations. For example, 0.5*(infant mortality in 1962 + infant mortality in 1967) = infant mortality in 1965. We take the income/wealth variable as real GDP per capital (RGDPCH) at 2000 real purchasing power (PPP) adjusted dollars from Summers and Heston's Penn World Tables 6.2 (PWT 6.2). This variable is measured in international dollars (I$), which adjusts for price level differences across countries. Another control variable, education, is obtained from Barro and Lee (2000). As health data (infant mortality) are only available at five-year frequencies, we also use the income and education data at a five-year frequency. In addition, in this chapter countries that have missing data in some years may not satisfy the order conditions to implement the three panel causality tests discussed in Section 2.1 - 2.3. Therefore, although our initial dataset covers 105 developing countries, only some of them are used in the following tests. We list these countries, corresponding to each test in Section 3 and Section 4, in the appendices. Countries that are in our initial dataset but are not covered by any of the following tests include: Afghanistan, Angola, Columbia, Comoros, Congo Dem. Rep., Cyprus, Dominica, Fiji, Gabon, Gambia, Greece, Grenada, Guinea-Bissau, Guyana,

---

Appendix 4), depending on the model specification, as well as the country group under consideration. The maximum number of countries available for this approach is 74. Therefore, we use results from Weinhold's (1999) simulation case with $T = 5$ and $N = 80$ and correct the bias in the variance and the standard error by dividing the non-adjusted variance by the square root of 10 and dividing the non-adjusted standard error by 2. We recognize that such adjustments are extremely ad-hoc, but they provide some notion of the sensitivity of the causality outcomes to these statistics.

[71] These countries are listed in footnote 2.

Haiti, Korea Republic of, Liberia, Nepal, Myanmar, P. New Guinea, Poland, Portugal, Seychelles, Solomon Islands, Sri Lanka, St. Lucia, Sudan, Syria, Western Samoa, and Yugoslavia.

3. Results and Comparisons

This section presents our results of Granger causality tests using the modified Holtz-Eakin et al. (1988), Hurlin and Venet (2003), Hurlin (2004, 2005), and Weinhold (1999) and Nair-Reichert and Weinhold (2001) frameworks. For each approach, results are expected to vary because the three methods have different model setups and the appropriateness of the null distribution approximations in finite samples likely varies across the three methods.

3.1 Results Using the Holtz-Eakin et al. (1988) Framework

Following Holtz-Eakin et al. (1988), we prepared our panel data (N=105, T=9) as T stacks with each stack containing N records. This setup implies that we have $2T(K+1)$ coefficients to be estimated. Because T is much smaller than N for our panel and given that the setup has N being the faster index, there should be enough data points to enable identification of the required coefficients for any reasonable choice of K. However, as we wish to compare the outcomes with those that we obtain using the methods proposed by Hurlin and Venet (2003), Hurlin (2004, 2005), Weinhold (1999) and Nair-Reichert and Weinhold (2001), for which T is the faster index so that there are observation numbers resulting in fewer available observations for identification of the parameters, we limit the lag order in our models to $K = 1$, $K = 2$, and $K = 3$. We

also consider models with and without education as an explanatory variable to examine the robustness of our results.

We modify Holtz-Eakin's approach by replacing their IV and GLS estimations with GMM estimation. As the data are prepared as T stacks of N countries in each stack, the GMM estimation can be straightforwardly undertaken as system estimation in EViews of the T equations. We use an identity weighting matrix as the weight matrix for the GMM criterion function and allow for unknown heteroskedasticity by using White's (1980) heteroskedasticity consistent estimator of the variance-covaraince matrix of the parameter vector's GMM estimator. EViews uses the consistent two-stage-least-square estimator to provide the initial estimates of coefficients to enable the GMM estimation to proceed. Having estimated the models, we undertake the panel Granger causality tests outlined in the previous sub-sections.

After removing missing data from our panel dataset, there are 58 countries with $T = 9$ left in the model without education and 41 countries with $T = 9$ in the model with education (see Appendix 1). Table 19 reports results from these various tests, with and without education as an additional explanatory variable. The first finding is that those coefficients on the lagged variables are not "time stationary"; that is, we cannot assume that the coefficients are constant over time. Because our estimations are based on data at five-year frequencies, that is, two consecutive equations in the system do not represent the wealth-health relationship for two consecutive years, but for two five-year apart periods, it is very likely that this relationship changes over time.

Given these "time non-stationarity" results, we then test for causality between wealth and health. The second finding in Table 19 is that there exists bi-directional Granger

causality between health and income, except for the case where we test the causality running from health to income in the model without education. Bi-directional causality implies feedback between the two variables: health causes wealth and wealth simultaneously causes health. This finding can be interpreted as an empirical explanation of the multiple-equilibrium model in Chen et al. (2008) and Chakraborty (2004). Theoretical models in these papers suggest that a shorter life expectancy discourages national savings (and hence wealth) and, simultaneously, poor countries with lower initial capitals are experiencing a "health-trap" where agents expect to live for a shorter period. This mechanism demonstrates a reinforcement of the key relationship between wealth (via income) and health.

The third finding from the results reported in Table 19 is that the introduction of education does not change the qualitative outcomes of the causality tests. Including education in the model specification, the Wald statistics of the "time stationarity" tests are much higher than those from the models without education.[72] This is not surprising: when education and its lagged terms are added into each period's equation there are more coefficients involved in the "time stationarity" test, so it is likely that it is more difficult to support the null hypothesis. The test outcomes reported in Table 19 also suggest that the causality between health and income is robust to the lag order K.

Despite these robust results, we have two potential concerns. First, the approach taken by Holtz-Eakin et al. (1988) does not allow for the coefficients of the lagged variables to be different across countries. Although the coefficients can vary across time, forcing them to be homogenous across countries, when they are heterogenous, may

[72] Except for the case when we are testing for causality from health to income with $K = 1$.

lead to substantial bias (Pesaran and Smith 1995). Tests in the next sections focus on this potential issue by using two different approaches to accommodate country-specific heterogeneity. Secondly, given that the assumption of "time stationarity" is rejected, there are quite a few coefficients to be tested in the causality test. For example, with education in the model and K = 3, the non-causality hypothesis involves testing 16 simultaneous zero restrictions. The reliability of the test conclusions depend on our panel data being of sufficient size to allow all these constraints to be tested.[73]

Table 19: Causality between health and wealth using Holtz-Eakin et al.'s (1988) method

Column #	1	2	3	4
	wealth to health		health to wealth	
	no education	with education	no education	with education
		K = 1		
Time stationarity	48.59	130.92	84.96	82.44
p value	0.00	0.00	0.00	0.00
Non-Causality	35.48	23.63	10.34	26.27
p value	0.00**	0.02**	0.59	0.01**
		K = 2		
Time stationarity	80.12	359.32	51.84	964.48
P value	0.00	0.00	0.00	0.00
Non-Causality	27.95	70.20	26.31	50.39
P value	0.02**	0.00**	0.03**	0.00**
		K = 3		
Time stationarity	34.09	750.25	58.62	992.58
P value	0.08	0.00	0.00	0.00
Non-Causality	24.53	70.76	31.05	44.01
P value	0.08*	0.00**	0.01**	0.00**

Note: Asterisk * and ** denote statistical significance at 10% and 5% levels respectively.

[73] Even if the null hypothesis of "time stationarity" is accepted, and the number of our coefficients to be tested is reduced, we still need to pay attention to the significance level of the causality test. As Holtz-Eakin (1988) considers the "time stationarity" test and the non-causality test as a sequence of tests, the significance of the overall testing strategy is: $a_1+a_2-a_1a_2$. (Holtz-Eakin et al. 1988, page 1383), where a_1 is the significance level of the "time stationarity" test and a_2 is the significance level of the causality test.

3.2 Results Using the Hurlin and Venet (2003), Hurlin (2004, 2005) Framework

To allow for possible heterogeneity of causality across individual countries, Hurlin and Venet (2003), and Hurlin (2004, 2005) construct an "average" Wald statistic formed from the statistics for non-causality for each individual country. Hence, our sample size needs to be large enough to undertake these individual causality tests. At the same time, in a finite sample, the first and second moments of the Wald statistics for the individual causality tests only exist when $T > 5K+2$ (Magnus 1986, Hurlin 2004). As our panel has $T = 9$, the lag order K has to be set to 1. This may be the most problematic practical shortcoming of Hurlin and Venet's panel causality test.

Countries that have enough observations to allow us to undertake the test are listed in Appendix 1. Applying Hurlin and Venet's panel causality test to these data leads to the results reported in Table 20. In each case, we compute the finite-sample test statistic $\tilde{Z}_{N,T}$ for models with and without education. The approximate critical values of this statistic at the 5% and 10% significance levels are reported in the fifth and sixth rows of Table 20. For comparison purposes, the standardized statistic $Z_{N,T}$ is also reported, but it is worthwhile to note that in our finite samples, it is doubtful that we can rely on the test outcomes from the $Z_{N,T}$ statistic as the asymptotic approximation is likely unreliable.

Columns 1 and 3 report results without education in the model and columns 2 and 4 report outcomes when education is included. The outcomes reported in columns 3 and 4 both support the existence of causality running from health to wealth/income. This

is consistent with our results reported in Table 19: better health improves labour productivity, lengthens longevity and encourages savings and capital accumulation. In the health equation, income has causal effects on health when we control for education, but there is some uncertainty when education is omitted. This finding suggests that the exclusion of education potentially mis-estimates the impact of income.

The results in Table 20 also show that the values of the finite-sample standardized statistic, $\tilde{Z}_{N,T}$ and the asymptotic Wald statistic, $Z_{N,T}$,described previously, are quite different. This implies that in this case the finite sample correction for small T brings about a big change in the calculated test statistics. However, the qualitative conclusions obtained using the two statistics are similar: using the asymptotic statistic, at a 5% significance level (corresponding critical value is 1.96) the non-causality null hypothesis is rejected in all four cases in Table 20; we also reject the non-causality hypothesis in three out of four cases using the finite sample statistic and the corresponding approximate critical values. All in all, there is evidence suggesting causality between wealth and health. We now turn our attention to the method suggested by Weinhold and her co-author Nair-Reichert.

So far we have evidence in favour of bi-directional wealth-health causality for most of the specifications using Holtz-Eakin et al. (1988), Hurlin and Venet (2003) and Hurlin (2004, 2005). Next we use the approach in Weinhold (1999) and Nair-Reichert and Weinhold (2001) to see whether the results hold up under a third approach.

Table 20: Causality between health and income using Hurlin and Venet (2003) and Hurlin (2004, 2005)'s approach with $K = 1$

Column #	1	2	3	4
	wealth to health		health to wealth	
	no education	with education	no education	with education
$\tilde{Z}_{N,T}$	1.43	3.21**	3.55**	2.78*
$c_{N,T}(0.1)$	2.46	2.61	2.46	2.61
$c_{N,T}(0.05)$	2.87	3.15	2.87	3.15
$Z_{N,T}$	8.34**	13.73**	15.38**	12.30**

Note: Asterisk * and ** denotes statistical significance at the 10% and 5% levels respectively.

3.3 Results Using Weinhold (1999), Nair-Reichert and Weinhold (2001) Framework

Weinhold's mixed fixed-random (MFR) approach allows the coefficients of the causal variable to be random and focuses on estimating the mean of these random coefficients. This mean is calculated using estimates from individual country regressions. Similar to Hurlin and Venet's (2003) approach, the use of individual regressions restricts the lag order K that can be modelled. For example, with $K = 1$, in the equation with education as an extra control variable, the first observation of each individual sector is lost due to the lagging on the right-hand side of each equation, , and as there are four parameters to be estimated (a constant, lag of infant mortality, lag of income, and lag of education), those countries with $T < 5$ have to be removed; with $K = 2$, the first two observations are lost, and we have seven parameters to be estimated in the model with education, so T has to be larger than nine. This is infeasible as our data set has a maximum of (T=) 9 periods. Therefore, as with Section 3.2, we only estimate the models with $K = 1$. After removing the missing data, we have 74 countries left for the model without education, and 53 countries for the

model with education (see Appendix 2). These cleaned up panels are unbalanced as some countries have data for the 9 years while others have data missing for some years.[74] In the model with education, to be consistent with Weinhold (1999), we also treat the coefficient of lagged education as random.

The proposed approach to estimating the MFR model provides estimates of the mean and variance of $\beta_i^{(k)}$ in equation (4.8), and these estimates of the moments provide general information about the heterogeneity of the causal relationship across countries. All of these results are provided in Table 21. Because coefficients of the lagged dependent variable are assumed to be fixed but vary across the N countries, to save space, we do not provide their estimates. Several interesting findings arise. First, the estimated mean coefficients for lagged income and lagged infant mortality have intuitive signs – we expect to see a negative relationship between these two variables. However, using the estimated standard errors, all of the mean estimates are statistically insignificant at a conventional significance level as their standard errors are larger than or equal to the magnitudes of the estimates. Secondly, the estimates of the variances of the mean coefficients suggest that there exists heterogeneity of causality across individual countries, especially in the income equation (columns 3 and 4 in Table 21), because these variances are relatively large compared to their corresponding mean coefficients.

In the health equation (columns 1 and 2 in Table 21), the degree of heterogeneity is smaller and might not be enough to imply that the coefficients are completely

[74] In Section 3.1 and Section 3.2, the cleaned up panels are all balanced with $N = 58$ or $N = 41$. Here we have more countries left, because the approach of Weinhold (1999) and Nair-Reichert Weinhold (2001) does not impose extra order restrictions, such as $T > 5K+2$ as in Hurlin (2004).

idiosyncratic as these estimated variances are not more than two times the estimated mean coefficient.[75] Thirdly, as Weinhold's simulations suggest, our short panels with small *T* may produce misleading test conclusions if we ignore the upward bias in the estimated variance and the standard error of the estimated mean coefficient. In Table 21, we provide the results when we adjust the variance and standard errors in the manner described in Section 2.3, and use the adjusted standard error to construct the adjusted p-value (last row in Table 21) of the non-causality test that test for the statistical significance of the mean coefficient. Compared with the non-adjusted p-value (second last row of Table 21), we reject the null of non-causality from income to health and the null of non-causality from health to income in the model without education. However, the accuracy of these adjusted p-values may be questioned as the way we calculate the adjusted estimates of $v^{(k)^2}$ and $se_{\hat{\bar{\beta}}^{(k)}}$, based on the simulations undertaken by Weinhold (1999), may be inexact. Therefore, in the following sections, we carry out some sensitivity analyses on country sub-groups to double check the causality finding between income and health.

[75] In Nair-Reichert and Weinhold (2001, pages 165 and 166), the variance is about five times larger than the estimated mean coefficient, but the authors believe "Fortunately, … the heterogeneity in the relationship between GFDI and growth does not seem to imply that the coefficients are completely idiosyncratic across countries." In Table 21 our estimated variances are no more than 2 times the estimated mean coefficient in the equation of health, so we do not think our coefficients are idiosyncratic.

Table 21: Causality between health and wealth using the method suggest by Weinhold (1999) and Nair-Reichert and Weinhold (2001) with $K = 1$

Column #	1	2	3	4
	wealth to health		health to wealth	
Regressor	no education	with education	no education	with education
Income(*t*-1)	-0.095	-0.118		
Infant mortality (*t*-1)			-0.162	-0.197
Education(*t*-1)		0.005		0.032
Variance of individual coef. of the cause	0.115	0.174	0.557	1.663
Adj. variance of individual coef. of the cause	0.036	0.055	0.176	0.526
Standard error of the estimated mean coef.	0.090	0.135	0.139	0.388
Adj. standard error of the estimated mean coef	0.045	0.068	0.070	0.194
p-value	0.29	0.38	0.25	0.61
Adj. p-value	0.04**	0.08*	0.02**	0.31

Note: Asterisk * and ** denotes statistical significance at the 10% and 5% levels respectively. P-values are for the non-causality test, which examine for the statistically significance of the estimated mean coefficient.

4. Wealth-Health Causality for Middle-Income Countries and Low-Income Countries

It is of interest to explore the sensitivity of the results to the sample of countries covered in the analysis. Causality from income to health might be statistically significant in the whole panel dataset but be statistically insignificant in a smaller country sub-group. We explore such a question in this section by classifying the countries into low-income, middle-income and high-income groups.[76] Using the World Bank's country classification, our 105 developing countries can be grouped as 43 low-income countries, 57 middle-income countries, and 5 high-income countries.[77]

[76] We use the World Bank's country classification criteria, which can be found from the website www.worldbank.org.

[77] The low-income countries are: Afghanistan, Bangladesh, Benin, Burkina Faso, Burundi, Central African Rep., Chad, Comoros, Congo, Dem. Rep., Ethiopia, Gambia, Ghana, Guinea, Guinea-Bissau, Haiti, India, Ivory Coast, Kenya, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique,

The following sensitivity tests focus on the middle-income countries and low-income countries. We ignore the high-income countries as there are only 5 of them in our dataset and the existence of a bi-directional causality between income and health may be more meaningful for countries that are still suffering poor health conditions and struggling for stable economic growth. If bi-directional causality is confirmed then policy makers can explore ways in which health and wealth policies can be used to improve each other.

As the results in Table 19 show that the inclusion of education and varying the lag order K do not affect the test conclusions for causality between income and health, we only focus on the cases without education with $K = 1$. This simplification is also appropriate because we have limited number of available countries. The small N causes difficulties when the information set expands, either by introducing more variables or increasing the lag order.

Results from applying the Holtz-Eakin et al. (1988) methodology with GMM estimation are reported in Table 22. Dropping countries that do not have enough observations to satisfy the GMM order condition, which was discussed in Section 2.1, leaves 33 middle-income countries and 23 low-income countries in the sample for the analysis when education is excluded (see Appendix 3). The test results are similar to those in Table 19: we, again, confirm the existence of bi-directional causality between

---

Myanmar, Nepal, Niger, Nigeria, Pakistan, Papua, Rwanda, Senegal, Sierra, Solomon, Somalia, Sudan, Tanzania, Togo, Uganda, Yemen, Yugoslavia, Zambia, and Zimbabwe. Middle-income countries include Algeria, Angola, Argentina, Bolivia, Botswana, Brazil, Cameroon, Cape Verdi, Chile, China, Colombia, Costa Rica, Dominica, Dominican Rep., Ecuador, Egypt, El Salvador, Fiji, Gabon, Grenada, Guatemala, Guyana, Honduras, Hungary, Indonesia, Iran, Iraq, Jamaica, Jordan, Lesotho, Malaysia, Mauritius, Mexico, Morocco, Nicaragua, Panama, Paraguay, Peru, Philippine, Poland, Samoa, Seychelles, South Korea, Sri Lanka, St. Lucia, St. Vincent, Suriname, Swaziland, Syria, Thailand, Tonga, Tunisia, Turkey, Uruguay, Vanuatu, Venezuela, and Zaire. See footnote 5 as well.

wealth and health for both the middle-income countries and the low-income countries. Furthermore, these causal relationships are heterogeneous across time periods.

Table 22: Causality between health and wealth for middle-income countries and low-income countries using the Holtz-Eakin et al. (1988) approach with *K* = 1 and without education

Column #	1	2	3	4
	wealth to health		health to wealth	
Countries	Middle-income	Low-income	Middle-income	Low-income
Time stationarity	40.46	1003.70	350.60	1812.00
P value	0.00**	0.00**	0.00**	0.00**
Non-Causality	42.17	91.90	34.80	120
P value	0.00**	0.00**	0.00**	0.00**

Note: Asterisk * and ** denote statistical significance at 10% and 5% levels respectively.

Interesting results are also found when we implement the method considered by Hurlin and Venet (2003) and Hurlin (2004, 2005). We report these in Table 23. First, for middle-income countries, bi-directional causality is found in the model with education, while in the model excluding education there is no causality in either direction. Secondly, for the group of low-income countries, there is no causality in either direction, using the finite-sample statistic rather than the asymptotic statistic. Such outcomes suggest that it matters which countries are used in the analysis.

The middle-income countries make the main contribution to the test results that are found in Table 20: out of four cases, three indicate the existence of causality between wealth and health. In contrast to the results reported in Table 19 and Table 22, the outcomes provided in Table 23 further stress that ignoring education can lead to a different conclusion for the middle-income countries. This is similar to the findings of Pritchett and Summers (1996) and those in Chapter 3, where the exclusion of education biases the estimate of income effects on health. The findings given in Table

23 also suggest that a health-wealth causal pattern may require a certain level of development (income) before there is feedback between them.

Although for the poor countries we still observe a positive association between wealth and health, we may not be able to detect a causal relationship. For low-income countries, policy-makers may need to focus on economic development first before they use income policies as an instrument to improve health (Bloom and Canning 2003), or rely on direct health reforms to help to eliminate illness before the virtuous wealth-health cycle is established.

Table 23: Causality between health and wealth for low-income countries using Hurlin and Venet (2003) and Hurlin (2004, 2005) with $K = 1$

Column #	1	2	3	4	5	6	7	8
	wealth to health				Health to wealth			
	Middle-income		Low-income		Middle-income		Low-income	
Regressor	no education	with education	no education	with education	no education	with education	no education	with education
$\tilde{Z}_{N,T}$	1.60	3.34*	0.06	0.96	0.39	3.68**	1.27	0.24
$c_{N,T}(0.1)$	3.01	3.13	2.93	3.41	3.01	3.13	2.93	3.41
$c_{N,T}(0.05)$	3.28	3.41	3.28	3.90	3.28	3.41	3.28	3.89
$Z_{N,T}$	8.05**	13.63**	2.47	4.83**	3.79**	14.77**	5.85**	0.82

Note: Asterisk * and ** denote statistical significance at 10% and 5% levels respectively.

When interpreting these results in this manner, we also need to remember that there might be issues with this methodology when N is small (as it is here). As explained in Section 2.2, in a finite sample the statistics associated with panel causality rely on the exact first and second moment estimates of the individual Wald statistics. These first and second moment estimators can be severely biased if N is too small. In particular, to apply Magnus' (1986) theorem to propose a simple test statistic that is not coming

from simulation or bootstrap experiments, Hurlin and Venet impose an approximation which may not be valid if N is too small.[78]

We now turn to the results when using the method examined by Weinhold (1999) and Nair-Reichert and Weinhold (2001). Countries explored for this approach are listed in Appendix 4. Comparing the results with those in Table 21, the Weinhold (1999) and Nair-Reichert and Weinhold (2001) approach produces quite robust results for both country groups: signs of the lagged income, lagged infant mortality are all intuitive; and the magnitudes of the estimated variance of individual coefficients (β_i) do not change substantially. The variance suggests that there still exists heterogeneity in the causal relationships between wealth and health across these middle-income countries or low-income countries. In other words, using income as a criterion to classify countries may not be ideal to remove the heterogeneity.

As the numbers of countries in these sensitivity analyses are smaller than those in the tests using the sample of developing countries, we expect that the standard errors of the estimated mean coefficient here to be larger than the one from the full sample test, as is suggested by simulations in Weinhold (1999). This indeed is what we observe from the results given in Table 24: standard errors are larger than those reported in each corresponding element in Table 21. At the same time, as we discussed in Section 2.3, simulations in Weinhold (1999) show that the standard errors in short panels can be two times, three times, or four times of the one from a larger panel. To adjust this upward biasness, we decide to divide the non-adjusted standard error by 3 rather than 2, because the numbers of countries in these sensitivity tests are smaller than those in

[78] See Hurlin (2004 pages 12 and 13).

the tests on the full sample.[79,80] Causality tests detailed in Table 24 suggest the existence of bi-directional causality for the middle-income countries, while none of the cases for the low-income countries provide evidence of causality. This finding is consistent with those in Table 23: middle-income countries are more important in explaining the income-health causal relationship that is observed in Tables 19 – 21.

Table 24: Causality between health and wealth for low-income countries using the approach of Weinhold (1999) and Nair-Reichert and Weinhold (2001) with *K* = 1

Column #	1	2	3					4
	wealth to health				Health to wealth			
	Middle-income		Low-income		Middle-income		Low-income	
Regressor	no education	with education	no education	with education	no education	with education	no education	with education
Income(*t*-1)	-0.132	-0.129	-0.027	-0.086				
Infant mortality (*t*-1)					-0.202	-0.168	-0.098	-0.283
Education(*t*-1)		0.018		-0.001		0.060		-0.01
Variance of individual coef. of the cause	0.118	0.183	0.107	0.188	0.602	1.696	0.566	2.310
Adj. variance of individual coef. of the cause	0.037	0.058	0.033	0.059	0.190	0.536	0.179	0.674
Standard error of the estimated mean coef.	0.173	0.233	0.190	0.335	0.192	0.528	0.273	1.006
Adj. standard error of the estimated mean	0.058	0.078	0.063	0.112	0.064	0.176	0.091	0.335
p-value	0.45	0.60	0.30	0.80	0.30	0.75	0.72	0.81
Adj. p-value	0.04*	0.05**	0.66	0.44	0.002**	0.34	0.28	0.40

Note: Asterisk * and ** denote statistical significance at 10% and 5% levels respectively.

[79] Test results in Table 24 cover no more than 43 countries (Appendix 6). The maximal time span is 8. Hence, we use the results from Weinhold's (1999) simulation case with $N = 40$ and $T = 5$ and correct the standard error of the estimated mean coefficient by dividing the non-corrected standard error by 3.

[80] The variance and the adjusted variance reported in Table 21 and Table 24 only provide information on the heterogeneity in causality across countries. But unlike the standard errors, they are not applied to test the significance of the estimated mean coefficient or produce any p-value. Therefore, even though the number of countries covered by the test results in Table 24 is less then the one in Table 21, we still calculate the adjusted variance by dividing the non-adjusted one by square root of 10, same as what we suggest in footnote 10.

5. Conclusion

Testing for the aggregated health-wealth relationship has become more feasible for researchers these days with the availability of more country-level panel data. The empirical relationship between these two variables has been extensively studied in regressions based on post-1960 data, typically using correlation concepts. Rather than focusing on the linear correlation between health and wealth, we contribute to the literature by examining for a deeper relationship, causality, using three panel causality methodologies. These three tests control for individual country effects and potential heterogeneity in causality across countries and/or across time periods. Using data on developing countries from 1960 to 2000 at five-year frequencies, we find evidence for the existence of bi-directional causality between health and wealth. In particular, using a variant of the method proposed by Holtz-Eakin et al. (1988), we find that the causal relationship varies over time. In addition, the panel causality approach suggested by Weinhold (1999) and Nair-Reichert and Weinhold (2001) shows that the causal impacts from wealth to health (and from health to wealth) are heterogeneous across countries.

Sensitivity tests are completed with data on middle-income countries and low-income countries. The Holtz-Eakin et al. (1988) method produces robust results across our country sub-groups and suggests that bi-directional causality exists for both middle-income countries and low-income countries. However, the approaches proposed by Hurlin and Venet (2003) and Hurlin (2004, 2005), Weinhold (1999) and Nair-Reichert and Weinhold (2001) – methods that control for the potential heterogeneity of the causal relationship across individual countries - support the conclusion that the middle-income countries play a more significant role than the low-income countries

in explaining the wealth-health causality observed for the full sample. At the same time, because these sensitivity analyses involve the use of limited number of countries, we should be careful when interpreting these results as there may be some potential finite sample issues with the short panels.

Our test results provide policy implications from several perspectives. First, the existence of bi-directional wealth-health causality indicates that wealth and health (infant mortality) are not only correlated but also have the ability to reinforce each other. Improvements in either raise the possibility of improvements in the other. In the long-run, sustained economic growth is unlikely without sustained health improvements, and *vice versa*. Secondly, as the wealth-health causality is heterogeneous across countries, the efficiency and the effectiveness of health/income policies may also differ across countries. Thirdly, additional policy implications arise from the findings from our sensitivity analysis. If middle-income countries make a bigger contribution to the overall wealth-health causality, as there is little evidence of causality for low-income countries, then the implication is that the same health/income policies are not going to be suitable for both groups of countries. For low-income countries, the lack of causality suggests that we cannot rely on projects and programs that aim to improve income to simultaneously affect health (and *vice versa*). Rather, policy makers need to make sure that polices and programs are adopted that directly eliminate poverty and others that directly affect health. After income reaches a level that allows a virtuous cycle between wealth and health, health policy can be implemented to improve health and also economic development. Fourthly, using the approach of Holtz-Eakin et al. (1988), the wealth-health causality is heterogeneous over time so that policy makers need to dynamically monitor the impact of health and income policies on their target variables.

Much remains for possible future research. First, we have ignored any consequences of non-stationarity of the data on the methods, including adapting models for possible cointegrating relationships. Secondly, we might study impulse response functions that trace out how income (health) responds to a health (income) shock. Thirdly, the exact moments of the individual Wald statistics used in the Hurlin and Venet (2003) and Hurlin (2004, 2005) method can be estimated using bootstrapping. This may reduce the bias and lead to more accurate test results.

Appendix 1: Countries in Table 19 and Table 20

In the model without education:

Tests cover 58 developing countries. They are: Algeria, Argentina, Benin, Bolivia, Brazil, Burkina Faso, Burundi, Cameroon, Chile, China, Costa Rica, Dominican Rep., Ecuador, Egypt, El Salvador, Ethiopia, Ghana, Guatemala, Guinea, Honduras, India, Indonesia, Iran, Ivory Coast, Jamaica, Jordan, Kenya, Lesotho, Madagascar, Malawi, Malaysia, Mali, Mauritius, Mexico, Morocco, Mozambique, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Panama, Paraguay, Peru, Philippines, Rwanda, Senegal, South Africa, Taiwan, Tanzania, Thailand, Togo, Trinidad, Turkey, Uruguay, Venezuela, Zambia, and Zimbabwe.

In the model with education:

Tests cover 41 developing countries. They are: Algeria, Argentina, Bolivia, Brazil, Cameroon, Chile, Costa Rica, Dominican Rep., Ecuador, El Salvador, Ghana,

Guatemala, Honduras, India, Indonesia, Iran, Jamaica, Jordan, Kenya, Lesotho, Malawi, Malaysia, Mali, Mauritius, Mexico, Mozambique, Nepal, Nicaragua, Niger, Pakistan, Panama, Paraguay, Peru, Philippines, South Africa, Tanzania, Thailand, Trinidad, Turkey, Zambia, and Zimbabwe.

Appendix 2: Countries in Table 21

In the model without education:

Tests cover 74 developing countries. They are: Algeria, Argentina, Bangladesh, Benin, Bolivia, Botswana, Brazil, Burkina Faso, Burundi, Cameroon, Central Africa, Chad, Chile, China, Costa Rica, Dominican Rep., Ecuador, Egypt, El Salvador, Ethiopia, Guatemala, Guinea, Honduras, Hungary, India, Indonesia, Iran, Iraq, Ivory Coast, Jamaica, Jordan, Kenya, Lesotho, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mauritius, Mexico, Morocco, Mozambique, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Panama, Paraguay, Peru, Philippines, Rwanda, Samoa, Senegal, Sierra, Somalia, South Africa, St. Vincent, Suriname, Swaziland, Taiwan, Tanzania, Thailand, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Uruguay, Vanuatu, Venezuela, Zaire, Zambia, and Zimbabwe.

In the model with education:

Tests cover 53 developing countries are: Algeria, Argentina, Benin, Bolivia, Botswana, Brazil, Cameroon, Central Africa, Chile, Costa, Dominican Rep., Ecuador, El Salvador, Ghana, Guatemala, Honduras, Hungary, India, Indonesia, Iran, Jamaica, Jordan, Kenya, Lesotho, Malawi, Malaysia, Mali, Mauritius, Mexico, Mozambique, Nepal, Nicaragua, Niger, Pakistan, Panama, Paraguay, Peru, Philippines, Rwanda,

Samoa, South Africa, St. Vincent, Suriname, Swaziland, Tanzania, Thailand, Tonga, Trinidad, Tunisia, Turkey, Zaire, Zambia, and Zimbabwe.

Appendix 3: Countries in Table 22 and Table 23

In the model without education: tests cover 33 middle-income countries and 23 low-income countries.

Middle-income countries are: Algeria, Argentina, Bolivia, Brazil, Cameroon, Chile, China, Costa Rica, Dominican Rep., Ecuador, Egypt, El Salvador, Guatemala, Honduras, Indonesia, Iran, Jamaica, Jordan, Lesotho, Malaysia, Mauritius, Mexico, Morocco, Nicaragua, Panama, Paraguay, Peru, Philippines, South Africa, Thailand, Turkey, Uruguay, and Venezuela.

Low-income countries are: Benin, Burkina Faso, Burundi, Ethiopia, Ghana, Guinea, India, Ivory Coast, Kenya, Madagascar, Malawi, Mali, Mozambique, Nepal, Niger, Nigeria, Pakistan, Rwanda, Senegal, Tanzania, Togo, Zambia, and Zimbabwe.

In the model with education, tests cover 28 middle-income countries and 12 low-income countries.

Middle-income countries are: Algeria, Argentina, Bolivia, Brazil, Cameroon, Chile, Costa Rica, Dominican Rep., Ecuador, El Salvador, Guatemala, Honduras, Indonesia, Iran, Jamaica, Jordan, Lesotho, Malaysia, Mauritius, Mexico, Nicaragua, Panama, Paraguay, Peru, Philippines, South Africa, Thailand, and Turkey.

Low-income countries are: Ghana, India, Kenya, Malawi, Mali, Mozambique, Nepal, Niger, Pakistan, Tanzania, Zambia, and Zimbabwe.

Appendix 4: Countries in Table 24

In the model without education, tests cover 43 middle-income countries and 29 low-income countries.

Middle-income countries are: Algeria, Argentina, Bolivia, Botswana, Brazil, Cameroon, Chile, China, Costa Rica, Dominican Rep., Ecuador, Egypt, El Salvador, Guatemala, Honduras, Hungary, Indonesia, Iran, Iraq, Jamaica, Jordan, Lesotho, Malaysia, Mauritius, Mexico, Morocco, Nicaragua, Panama, Paraguay, Peru, Philippine, Samoa, South Korea, St. Vincent, Suriname, Swaziland, Thailand, Tonga, Tunisia, Turkey, Uruguay, Venezuela, and Zaire.

Low-income countries are: Bangladesh, Benin, Burkina Faso, Burundi, Central African Rep., Chad, Ethiopia, Ghana, Guinea, India, Ivory Coast, Kenya, Madagascar, Malawi, Mali, Mauritania, Mozambique, Nepal, Niger, Nigeria, Pakistan, Rwanda, Senegal, Sierra, Somalia, Tanzania, Togo, Zambia, and Zimbabwe.

In the model with education, tests cover 37 middle-income countries and 15 low-income countries.

Middle-income countries are: Algeria, Argentina, Bolivia, Botswana, Brazil, Cameroon, Chile, Costa Rica, Dominican Rep., Ecuador, El Salvador, Guatemala, Honduras, Hungary, Indonesia, Iran, Jamaica, Jordan, Lesotho, Malaysia, Mauritius, Mexico, Nicaragua, Panama, Paraguay, Peru, Philippine, Samoa, South Korea, St. Vincent, Suriname, Swaziland, Thailand, Tonga, Tunisia, Turkey, and Uruguay.

Low-income countries are: Benin, Central African Rep., Ghana, India, Kenya, Malawi, Mali, Mozambique, Nepal, Niger, Pakistan, Rwanda, and Tanzania,

CHAPTER 5:

CONCLUSION AND FUTURE RESEARCH

1. Conclusion

Health and wealth are two key dimensions of human development. Beyond modeling and testing for correlation between income and health, this dissertation examines a deeper association, causality, between these two variables. Chapter 2 models the endogenous interactions between income and life expectancy by modifying a standard overlapping generations model to allow individuals to directly choose their own longevity. This Chapter contributes to the literature on the individual's direct choice of longevity and the implications of this direct choice on income growth. By specifying a Cobb-Douglas utility and production functions, we are able to derive analytical solutions for the equilibrium transition function and characterize behavior in the steady states. In particular, multiple equilibria can be produced by the direct choice of endogenous longevity: if the initial value of capital is small enough, some agents would choose to live shorter lives, and with a value of capital's share of income which is larger than 0.5, we have multiple stable steady states. When there are multiple equilibria they can be Pareto ranked, and the worse equilibrium is a "poverty-trap equilibrium" in which there is pervasive poverty and low life expectancy. The existence of a poverty-trap equilibrium establishes the possibility of a vicious cycle between income and life expectancy.

In Chapter 3, we provide some empirical evidence that income positively affects health and that the endogeneity of health can bias the estimates. Using external instrumental variables for data over 1960 to 2000, we find that income has a positive and statistically significant impact on infant mortality and life expectancy, but that its magnitude is lower than over the sub-sample period (1960 and 1985) used by Pritchett and Summers (1996). Given that the available external instruments appear to be weak instruments, we estimate the relationship using system generalized method of moments approach with internal instruments in a dynamic panel framework. The results suggest that a dynamic model is more appropriate as it captures the importance of lagged health. We find that the impact of income on health, although now smaller in magnitude, continues to be positive and statistically significant.

The objective of Chapter 4 is to formally test for Granger causality between health (infant mortality) and income across developing countries, using three panel causality tests. We conclude that there exists a bi-directional causal relationship between income and infant mortality and the causality is heterogeneous across countries. Sensitivity tests are completed with data on middle-income countries and low-income countries. In most cases, our evidence supports the bi-directional causal relationship found for the full sample of countries. However, there is some evidence to suggest that middle-income countries play a more significant role compared to low-income countries in explaining the overall wealth-health causality.

2. Future Research

The analysis in the thesis suggests a number of interesting avenues for future research to improve our understanding of the relationship between health and income. The model in Chapter 2 was highly stylized. Extending the model in various directions would make it more empirically relevant. First, it is well known that the level of infant mortality substantially influences life expectancy at birth. To take this into account, we need to extend the model to include a third period of life: childhood. Secondly, individuals also face trade-offs between private health expenditures and human capital investments. Introducing human capital into the model would allow for the incorporation of the knowledge dimension of human development, found to be important in the HDI, into the framework. Such a model would be a more complete picture of human development. Finally, the model could be extended to allow agents to have different utilities of exiting life early.

The analysis in Chapter 3 ignored the various possibilities. First, there is the possibility that there is a non-linear relation between health and wealth. This might be modeled using threshold regression methods (Hansen 2000, Caner and Hansen 2004). Such methods are able to endogenously separate countries into different income groups based on the response of health to income. Also, this approach might offer a way to test the model in Chapter 2 where behavior depended on the capital share of income (parameter α in Chapter 2). Second, the analysis in Chapter 3 ignored the possibility of income and health being non-stationary and co-integrated. Future research might apply the recently developed panel unit root tests and panel co-integration tests to see if the time series properties of the data affect the estimated income-health association. Further, it would be interesting to examine whether our

key qualitative conclusions remain robust to the inclusion of additional explanatory variables, such as the Gini coefficient of income inequality. Data constraints currently limit such an extension but this may become more feasible with the passing of time and better data collection in developing countries.

The findings reported in Chapter 4 suggest that a challenge remains of finding ways to improve the finite sample performance of panel causality tests. One way to improve performance would be to obtain health data with longer time dimension, as the applied tests may have issues in short panels. Whether such data is available from other sources for infant mortality remains to be seen. Alternatively, perhaps other proxies for health might prove a fruitful avenue for research. Also, rather than adopting Magnus' (1986) theorem, it might be preferable to use bootstrapping to approximate the finite sample distribution of the individual Wald statistics used in the method suggested by Hurlin and Venet (2003), and Hurlin (2004, 2005). The use of bootstrap procedures may also be useful for the other examined causality tests. Secondly, we have ignored any consequences of non-stationarity of the data on the methods, including adapting the dynamic models for any possible co-integrating relationships. Finally, we have limited our attention to exploring for the Granger causal links between health and wealth. Other tools are also popular in the empirical macroeconomic literature; e.g., impulse response functions and forecast error variance decompositions. It would be interesting to also use these tools to investigate the dynamic interactions between health and income.

Although the analysis in this dissertation falls well short of a full description of the relationship between health and wealth, this dissertation has begun the work of systematically analyzing for causal linkages between these important variables. In

particular, this dissertation provides theoretical and empirical evidence for positive bi-directional feedback between income and proxies for health. This opens up the possibility that there may be substantial multiplier effects in how these variables impact each other. This, in turn, suggests that policies aimed at health investment and income redistribution, to promote human development, might be quite effective. Despite our contributions, much remains for future research.

REFERENCES

Adler, N., Boyce T., Chesney, M., Cohen S., Folkman S., Kahn R., and L. Syme. (1994) "Socioeconomic Status and Health, The Challenge of the Gradient." *American Psychologist* 49: 15-24.

Amemiya, T. and T. Macurdy. (1986) "Instrumental-Variable Estimation of an Error-Component Model." *Econometrica* 54: 869-80.

Anand, S. and M. Ravallion. (1993) "Human Development in Poor Countries: On the Role of Private Incomes and Public Services." *Journal of Economic Perspectives* 7: 133-50.

Anderson, T. and C. Hsiao. (1982) "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics* 18: 47-82.

Arellano, M. and S. Bond. (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58: 277-97.

Barro, R. and J. Lee. (1992) "International Comparisons of Educational Attainment: 1960-1985." *Journal of Monetary Economics* 32: 363-94.

Barro, R. and J. Lee. (2000) "International Data on Education Attainment: Updates and Implications." CID Working Paper #42, Center for International Development, Harvard University.

Becker, C. and D. Bloom. (1998) "The Demographic Crisis in the Former Soviet Union: Introduction." *World Development* 26: 1913-19.

Becker, G., Philipson, T. and R. Soares. (2005) "The Quantity and Quality of Life and the Evolution of World Inequality." *American Economic Review* 95: 277-291.

Becsi, Z. (2003) "Longevity and the Life Cycle." Southern Illinois University manuscript.

Bhalla, S. and I. Gill. (1993) "Social Expenditure Policies and Welfare Achievement in Developing Countries." Washington, D.C: World bank. Mimeo.

Bhattacharya, J. and X. Qiao. (2005) "Public and Private Expenditures on Health in a Growth Model." Iowa State University Department of Economics Working Paper #05020.

Bidani, B. and M. Ravallion. (1997) "Decomposing social indicators using distributional data." *Journal of Econometrics* 77: 125-39.

Bishai, D., Marjorie O., and P. Andrew. (2007) "Does the Level of Infant Mortality Affect the Rate of Decline? Time Series Data from 21 Countries." *Economics and Human Biology* 5: 74-81.

Blackburn, K. and G. P. Cipriani. (2002) "A Model of Longevity, Fertility and Growth." *Journal of Economic Dynamics and Control* 26: 187-204.

Blackburn, K. and H. Issa. (2002) "Endogenous Life Expectancy in a Simple Model of Growth." University of Manchester CGBCR Discussion Paper #013.

Bloom, D., Canning, D. and B. Graham. (2002) "Longevity and Life Cycle Savings", NBER Working Paper # 8808.

Bloom, D. and D. Canning. (2003) "The Health and Poverty of Nations: from Theory to Proactice." *Journal of Human Development* 4: 47-71.

Blundell, R. and S. Bond. (2000) "GMM Estimation with Persistent Panel Data: an Application to Production Functions." *Econometric Reviews* 19: 321-40.

Bobba, M. and D. Colleilo. (2007) "Weak Instruments and Weak Identification, in Estimating the Effects of Education, on Democracy." *Economics Letters* 96: 301-06.

Boehmer, U. and J. Williamson. (1996) "The Impact of Women's Status on Infant Mortality Rate: A Cross-Country Analysis." *Social Indicators Research* 37(3): 333-60.

Bond, S. (2002) “Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice.” Centre for Microdata Methods and Practice, Institute for Fiscal Studies, CeMMAP working paper #CWP09-02.

Bond. S., Hoeffler, A. and J. Temple. (2001) “GMM Estimation of Empirical Growth Models.” Economics Paper # 2001-W21, University of Oxford.

Breusch, T., Mizon, G., and P. Schmidt. (1989) “Efficient Estimation Using Panel Data.” *Econometrica* 57: 695-700.

Brinkley, G. (2000) “The Macroeconomic Impact of Improving Health, Investigating the Causal Direction.” Working paper, University of California, Davis.

Caldwell, J. C. (1986) “Routes to low mortality in poor countries.” *Population and Development Review* 12: 171-220.

Canner M. and B. Hansen. (2004) “Instrumental Variable Estimation of a Threshold Regression.” *Econometric Theory* 20: 813-843.

Chakraborty S. (2004) “Endogenous Lifetime and Economic Growth.” *Journal of Economic Theory* 116: 119-37.

Chen, W., Engineer, M. and I. King. (2008) "Choosing Longevity with Overlapping Generations: To Be or Not to Be in Diamond's Model." *The B.E. Journal of Macroeconomics (Contributions)* 8 : Article 6.

Choi, I. (2001) “Unit Root Tests for Panel Data.” *Journal of International Money and Finance* 20: 249-72.

Cornia, A., and G. Mwabu. (1997) “Health Status and Health Policy in Sub-Saharan Africa: A Long-Term Perspective.” UNU/WIDER, Working Paper #141, Helsinki, Finland.

Croix, D. and O. Licandro. (1999) “Life Expectancy and Endogenous Growth.” *Economics Letter* 65: 255-63.

Davidson, R. and J. MacKinnon. (1989) “Testing Consistency Using Artificial Regressions.” *Econometric Theory* 5: 363-84.

Davidson, R. and J. MacKinnon. (1993) *Estimation and Inference in Econometrics*, Oxford University Press: New York.

Deaton, A. (2003) "Health, Inequality, and Economic Development." *Journal of Economic Literature* 41: 113-158.

Diamond, P. (1965) "National Debt in a Neoclassical Growth Model." *American Economic Review* 55: 1126-1150.

Dreze, P., and A. Sen. (1989) *Hunger and Public Action*, Oxford: Clarendon Press.

Easterly, W. (1999) "Life During Growth." *Journal of Economic Growth* 4: 239-76.

Easterly, W., Kremer M., Pritchett L., and L. Summers. (1993) "Good Policy or Good Luck? Country Growth Performance and Temporary Shocks." *Journal of Monetary Economics* 32: 459-84.

Ehrlich, I. and H. Chuma. (1990) "A Model of the Demand for Longevity and the Value of Life Extension." *Journal of Political Economy* 98: 761-782.

Ehrlich, I. and F. Lui. (1991) "Intergenerational Trade, Longevity and Economic Growth." *Journal of Political Economy* 99: 1029-1059.

Filmer D. and L. Pritchett. (1999) "The Impact of Public Spending on Health: Does Money Matter?" *Social Science and Medicine* 49: 1309-23.

Finlay, J. (2006) "Endogenous Longevity and Economic Growth." PhD thesis, Australian National University.

Fischer, S. (1993) "The Role of Macroeconomic Factors in Growth." *Journal of Monetary Economics* 32: 485-512.

Flegg, A. T. (1982) "Inequality of Income, Illiteracy and Medical Care as Determinants of Infant Mortality in Underdeveloped Countries." *Population Studies* 36: 441-58.

Frijters, P., Haisken-DeNew, J. and M., Shields. (2005) "Estimating the Causal Effect of Income on Health: Evidence from Post Reunification East Germany." Working paper # 465, Centre for Economic Policy Research, Research School of Social Sciences, Australian National University.

Gan, L. and G. Gong. (2007) "Estimating Independence Between Health and Education in a Dynamic Model." NBER Working Paper #12830.

Granger, C. (1969) "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37: 424-38.

Granger, C. (1980) "Testing for Causality." *Journal of Economic Dynamics and Control* 2: 329-52.

Granger, C. (2003) "Commentary: Some Aspects of Causal Relationships." *Journal of Econometrics* 112: 69-71.

Greene, W. (2002) *Econometric Analysis*, fifth edition, Prentice Hall: NJ.

Greene, W. (2008) *Econometric Analysis*, sixth edition, Prentice Hall: NJ.

Grimmett, G. and D. Stirzaker. (2001) *Probability and Random Processes*, third edition, Oxford University Press: NY.

Grossman, M. (1972) "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80: 223-255.

Hadri, K. (2000) "Testing for Stationarity in Heterogeneous Panel Data." *Econometric Journal* 3: 148–161.

Hanmer L., Lensink, R. and H. White. (2003) "Infant and Child Mortality in Developing Countries: Analysing the Data for Robust Determinants." *Journal of Development Studies* 40: 101-18.

Hansen, B. (2000) "Sample Splitting and Threshold Estimation." *Econometrica* 68: 575-603.

Hansen, P. (1982) “Large Sample Properties of Generalized Method of Moments Estimators.” *Econometrica* 50: 1029-54.

Hausman, J. (1978) “Specification Test in Econometrics.” *Econometrica* 46: 1251-71.

Hill, M. and E. King. (1992) “Women’s Education in the Third World: An Overview.” In *Women’s Education in Developing Countries: Barriers, Benefits and Policy,* eds. Elizabeth King and M. Anne Hill, 1050. Johns Hopkins University Press for the World Bank: Baltimore.

Holtz-Eakin, D., Newey, W., and R. Harvey. (1988) “Estimating Vector Autoregressions with Panel Data.” *Econometrica* 56: 1371-95.

Hsiao, C. (1989) “Modeling Ontario Regional Electricity System Demand Using a Mixed Fixed and Random Coefficients Approach.” *Regional Science and Urban Economics* 19: 565-87.

Hsiao, C. (1995) “Panel Analysis for Metric Data.” In Handbook of Statistical Modelling in the Social and Behavioral Sciences, edited by Arminger, G., Clogg, C., and M. Dobel, 361-400. Plenum.

Hsiao, C. (2000) “Economic Panel Data Methodology.” in *International Encyclopedia of the Social and Behavioral Sciences*, edited by Snelser, N. and P. Bates, Elsevier: Oxford.

Hsiao, C. (2003) *Analysis of Panel Data*, second edition, Cambridge University Press: Cambridge.

Hurlin, C. (2004) “Testing Granger Causality in Heterogeneous Panel Data Models with Fixed Coefficients.” Working Paper # 2004-05, Laboratoire d’Economie d’Orleans.

Hurlin, C. (2005) “Granger Causality Tests in Panel Data Models with Fixed Coefficients.” *Revue Economique* 56: 1-11.

Hurlin, C. and B. Venet. (2003) “Granger Causality Tests in Panel Data Models with Fixed Coefficients.” Mimeograph: Universitie Paris I.

Im, S., Pesaran, H., and Y. Shin. (2003) “Testing for Unit Roots in Heterogeneous Panels.” *Journal of Econometrics* 115: 53–74.

Issa, H. and B. Ouattara. (2005) “The Effect of Private and Public Health Expenditure on Infant Mortality Rates: Does the Level of Development Matter?” Department of Economics Discussion Paper, University of Wales Swansea.

Jamison, D.T., Wang, J., Hill, K., and J-L. Londoño. (1996) “Income, Mortality and Fertility Control in Latin America: Country-Level Performance 1960–90.” Mimeo, LAC Technical Department, World Bank, Washington, DC.

Kakwani, N. (1993) “Performance in Living Standards: An International Comparison.” *Journal of Development Economics* 41: 307-36.

Kalemli-Ozcan, S. (2002) “Does Mortality Decline Promote Economic Growth?” *Journal of Economic Growth* 7: 411-439.

Kalemli-Ozcan, S., Ryder, H. and D. Weil. (2000) “Mortality Decline, Human Capital Investment and Economic Growth.” *Journal of Development Economics* 62: 1-23.

King, E. and M. Rosenzweig. (1991) “Do Public Expenditures Explain Improvements in Human Development? Results from Fixed Effects Analysis of Country Level Data.” Background for World Development Report 1991. Mimeo.

Levin, A., Lin, C. and C. Chu. (2002) “Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties.” *Journal of Econometrics* 108: 1-24.

Levine, R. and D. Renelt. (1992) “A Sensitivity Analysis of Cross-Country Growth Regressions.” *American Economic Review* 82: 942-63.

Lorentzen, P., McMillan, J. and R. Wacziarg. (2005) "Death and Development." NBER Working Paper # W11620.

Maddala, S. and S. Wu. (1999) “A Comparative Study of Unit Root Tests with Panel Data and A New Simple Test.” *Oxford Bulletin of Economics and Statistics* 61: 631–52.

Magnus, J. (1986) “The Exact Moments of a Ratio of Quadratic Forms in Normal Variables.” *Annales d’Economie et de Statistiques*, 4: 96-109.

Mayer, D. (2001) "The Long-Term Impact of Health on Economic Growth in Latin America." *World Development* 29: 1025-33.

McDonald, S. and J. Roberts. (2006) "AIDS and Economic Growth: A Human Capital Approach." *Journal of Development Economics* 80: 228-50.

McKeown, T., Record, G., and R. Turner. (1975) "An Interpretation of the Decline of Mortality in England and Wales during the Twentieth Century." *Population Studies* 29: 391-422.

Michaud, P. and A. Soest. (2004) "Health and Wealth of Elderly Couples: Causality Tests Using Dynamic Panel Data Models." Discussion Paper # 81, Center for Economic Research, Tilberg University.

Miguel, E. and M. Kremer. (2004) "Worms: Identifying Impacts on Health and Education in the Presence of Treatment Externality." *Econometrica* 72: 159-217.

Musgrove, P. (1996) "Public and Private Roles in Health." World Bank Discussion Paper #339.

Nair-Reichert U. and D. Weinhold. (2001) "Causality Tests for Cross-Country Panels: A New Look at FDI and Economic Growth in Developing Countries." *Oxford Bulletin of Economics and Statistics* 63:153-171.

Pesaran, H. and R. Smith. (1995) "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels." *Journal of Econometrics* 68: 79-113.

Pincus, T., Esther, R., DeWalt, D. and L. Callahan. (1998) "Social conditions and Self-Management Are More Powerful Determinants of Health than Access to Care." *Annals of Internal Medicine* 129: 406-411.

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

Pritchett, L. (1997) "Divergence, Big Time." *Journal of Economic Perspectives* 11: 3-17.

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

Rodgers, G. B. (1979) "Income and inequality as determinants of mortality: an international cross-section analysis." *Population Studies* 33: 343-51.

Ruger, P., Jamison, D., Bloom, D. and D. Canning (2001). "Health and the Economy", in *International Public Health: Diseases, Programs, Systems and Policies*, eds. Merson, M., Black, B., and A. Mills, 601-47, Jones and Bartlett Publishers: MA.

Sargan, J.D. (1988) "Testing for Misspecification After Estimating Using Instrumental Variables." In *Contributions to Econometrics: J.D. Sargan*, Cambridge University Press: Cambridge.

Schultz, P. (2002) "Wage Gains Associated with Height as a Form of Human Capital." *American Economic Review* 92: 349-353.

Sen, A. (1985) *Commodities and Capabilities*, North-Holland: Amsterdam.

Sen, A. (1987) *The Standard of Living: The Tanner Lectures*, Cambridge University Press: Cambridge.

Smith , V., Taylor D. and F. Sloan. (2001) "Longevity Expectations and Death: Can People Predict Their Own Demise?" *American Economic Review* 91: 1126-1134.

Soares, R. (2007) "On the Determinants of Mortality Reductions in the Developing World." *Population and Development Review* 33: 247-287.

Strauss J. and D. Thomas. (1998) "Health, Nutrition, and Economic Development." *Journal of Economic Literature* 36: 766-817.

Subbarao, K. and L. Raney. (1995) "Social Gains from Female Education: A Cross-National Study." *Economic Development and Cultural Change* 44: 105-28.

Summers, R. and A. Heston. (1991) "The Penn World Tables (mark 5): An Expanded Set of International Comparisons. 1982-88." *Quarterly Journal of Economics* 106: 327-68.

Swamy, P. (1970) “Efficient Inference in a Random Coefficient Regression Model.” *Econometrica* 38: 311-23.

Thomas, D. and J. Strauss. (1997) “Health and Wages: Evidences on Men and Women in Urban Brazil.” *Journal of Econometrics* 77: 159-85.

Thomas, D., Strauss J., and M. Henriques. (1990) “Child Survival, Height for Age and Household Characteristics in Brazil.” *Journal of Development Economics* 33: 197-234.

United Nations Development Programme. (1990) *Human Development Report* 1990, Oxford University Press: NY.

Usher, D. (1980) *The Measurement of Economic Growth*, Basil Blackwell: Oxford.

Weinhold, D. (1999) “A Dynamic “Fixed Effects” Model for Heterogeneous Panel Data.” Working paper, London School of Economics.

Weinhold, D. (2001) “Model Evaluation and Causality Testing in Short Panels: the Case of Infrastructure Provision and Population Growth in the Brazilian Amazon.” *Journal of Regional Science* 41: 639-58.

White, H. (1980) “A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity.” *Econometrica* 48: 817-38.

Windmeijer, F. (2005) “A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators.” *Journal of Econometrics* 126: 25-51.

World Bank. (1990) *World Development Report*. Oxford University Press: New York.

World Bank. (1993) *World Development Report*. Oxford University Press: New York.

Yaari, M. (1965) “Uncertain Lifetime, Life Insurance, and the Theory of the Consumer.” *Review of Economic Studies* 32: 137-150.

Zakir, M. and P. Wunnava. (1999) “Factors Affecting Infant Mortality Rate: Evidences from Cross-Section Data.” *Applied Economics Letter* 6: 271-73.

Zhang, J., Zhang, J., and R. Lee. (2003) “Rising Longevity, Education, Savings, and Growth.” *Journal of Development Economics* 70: 83-101.