RETHINKING THE HDI: A MORE THEORETICALLY CONSISTENT ALTERNATIVE

Seth Omondi Gor^a and Ciliaka Millicent Wanjiru Gitau^a ^aUniversity of Nairobi, Kenya ^aCorresponding author: gor_seth@uonbi.ac.ke or asegogor@yahoo.com

© Ontario International Development Agency. ISSN 1923-6654 (print), ISSN 1923-6662 (online). Available at http://www.ssrn.com/link/OIDA-Intl-Journal-Sustainable-Dev.html

Abstract: HDI has been roundly faulted on many different accounts. Of very notable interest to us is the mathematical treatment of the variables used. The Average Lack Indicator (ALI) used to construct the HDI is a simple arithmetic mean which gives the same weight to all the variables used. In response to the many flaws detected, considerable effort has been directed at formulating alternatives to the HDI. To the best of our knowledge no alternative has adequately resolved the problem of arbitrary and atheoretical weighting of the variables used in HDI, as well as in the other multi-dimensional measures of wellbeing.

This paper responds to this challenge. Using underfive child survivorship to proxy household welfare, we use a probit model to estimate parameters of an abbreviated social welfare function. IV probit is then used to estimate an underlying response variable which is also the probit index. The result is a subjective welfare index that a household attaches to child survival. The parameters are weights to each of the arguments of the welfare function that indicate the contribution of the various factors to household welfare. These weights are optimal since they maximize wellbeing of the household given its environment. They are also consistent and nonarbitrary because they reflect a household's preference orderings over the arguments of the welfare function.

Keywords: Abbreviated Social Welfare Index, Average Lack Indicator, Human Development Index, Instrumental Variable Probit, and Leontief Preference.

I. INTRODUCTION

The Human Development Approach emerged as an attempt to put people back in the centre of the discourse and actions related to economic and social policies. The HD paradigm was defined to cover all aspects of development, the core idea being that human well-being is central to the goal of development and that human beings

constitute the major economic resource. The HDI is the instrument that was used to operationalize the HD approach.

The following guiding principles were at the core of the formulation of the HDI [27]; that the new index would measure the basic concept of human development to enlarge people's choices; that the new index would include a limited number of variables to keep it simple and manageable; that a composite index would be constructed rather than a plethora of separate indices; that the HDI would cover both social and economic choices and that the coverage and methodology of HDI would be flexible to allow for gradual refinements over time.

Unveiled to the world in 1990 by the UNDP, HDI was developed as an alternative to the GNP and other income based measures. The purpose of the index [27] was to measure at least a few more choices besides income and to reflect them in a methodologically sound index. Initially, life expectancy was chosen as an index of longevity, adult literacy as an index of knowledge, and GNP per capita adjusted for Purchasing Power Parity as an index of access to a multiplicity of economic choices.

Mathematically, HDI is calculated as the complementary of the unit of the simple arithmetic mean of the Partial Lack Indicators (PLI) in the three domains. The PLI for a country in one domain indicates the distance between this country and the most advanced country as a percentage of the whole distance between the most developed country and the least developed country [24].

Successive improvements in the method of calculating the HDI has been witnessed over the

years by way of modifications in the methods of calculating the PLIs. The treatment of GDP per capita has been modified many times as has been the procedure of fixing limit values of variables. In addition, other variables have been introduced in the domain of education.

From inception, the HDI has elicited favourable and unfavourable reaction in equal measure from researchers, academicians and policy makers. Four major categories are discernible from the vast array of criticisms directed at HDI. The first line of criticism focuses on the variables used. Consideration here is given to the representativeness of the variables used, Noorbakhsh [22], Harkness [16], Dasgupta and Weale [10], Cotlear [9], Allen and Kelly [18], Hopkins [17], Manzoor et al [21] and to the overall coherence of the variables used [24]. Giovanni [14] suggested that there may exist a correlation between income and health. This position has been corroborated by Cahill [8] and Saha [24] and indeed confirmed by UNDP [29].

The second category of criticism focuses on distribution issues. The argument here is that the HDI does not indicate how resources are allocated throughout a country and does not take into account the levels of inequality that may exist [13], [27]. To illustrate, a country's HDI score will improve even if only a small minority of a country is able to earn more and better educate their children without any change in the earning power of the vast majority.

The third line deals with inter-temporal comparisons using HDI. A number of studies have confirmed that a country's HDI can vary through time for reasons other than its value [24], [26]. This suggests that the level and variations of the HDI over time, may not necessarily reflect the reality of living conditions of the population.

The final category concerns the mathematical treatment of variables used in calculating the HDI. In this regard, criticisms have centred on selection of the goal posts used in indexing the indicator of life expectancy, the choice of limits for the goal posts and the treatment of income over and above the poverty line. In reaction to the latter flaw, UNDP in 1991 used a simple version of Atkinson [4] formulae to modify computation of the HDI, an approach that has since been abandoned.

In this final category, the most significant criticism is reserved for the method of computing the Average Lack Indicator (ALI), from which the HDI is derived. The ALI is computed as the simple arithmetic mean of the PLI for each of the three domains. In this way it assigns the same weight to each of the respective indicators of the index. The present study, like others before it [5], [11] takes exception to this approach for the basic reason that the weights of the component indices should be directly derived from the data themselves. Despotis [11] and Gor [15] posit that this is one way of resolving the problem of arbitrary weighting of variables. Whereas the latter problem has been widely detected, it is instructive that even though the HDI does not have any sound theoretical basis in economics (problem of atheoretical weighting of component parts), this has gone unnoticed by critics.

Two strands of remedies have been proposed by critics of the HDI. One strand suggests modifications to the HDI itself to account for missing data or the addition of missing indicators deemed necessary by the critic. The other strand formulates an entirely new index with the aim of capturing a fuller understanding of human development and well-being. This paper presents a third strand that borrows from the existing domains of the HDI to formulate a more theoretically sound alternative with a view to resolving the problem of arbitrary and atheoretical weighting of variables used in HDI as well as in the other multidimensional measures of wellbeing.

II. THE ABBREVIATED SOCIAL WELFARE FUNCTION AS AN ALTERNATIVE TO THE HDI

One way of ranking alternate social states is through the formulation of a social welfare function (swf). In this paper, we construct an index of social welfare, using the concept of *abbreviated social welfare function*. A social welfare is abbreviated if it is expressed as a function of statistics calculated from the income distribution vector, controlling for other summary indicators of well-being. Following Fields [12], the general form of the abbreviated social welfare function can be expressed as:

W = f (PCI, GIN, POV, YCO)

Where

W = Abbreviated Social Welfare Index

PCI = Per Capita Income

GIN = Gini coefficient

POV = Poverty index or status

YCO = Control covariates, e.g key demographics, such as, family size, parents' education, age and area of residence.

Where:

$$\frac{\partial W}{\partial PCI} > 0; \frac{\partial W}{\partial GIN} < 0; \quad \frac{\partial W}{\partial POV} < 0; \text{ and } \frac{\partial W}{\partial YCO} > 0$$

$$P_{\alpha} = \frac{1}{N} \sum_{i=1}^{q} \left(1 - \frac{y_i}{z}\right)^{\alpha}$$

Where, P_{α} is a measure of absolute poverty, including food poverty; y_i is the total expenditure of household *i*, expressed in per adult equivalent terms (i = 1...N), *Z* is the poverty line expressed in per adult equivalent, *N* is the total number of households, *q* is the total number of poor households and α is the FGT parameter, interpreted as a measure of poverty aversion, $\alpha \ge 0$. For purposes of this study, we estimate and use only one of the three FGT measures, namely, the headcount ratio, for which $\alpha = 0$.

Since W is not observable, there is need to proxy social welfare with a measurable variable. We use child survival as a proxy for well-being at the household level. That is, a household with a surviving child is deemed to have a higher welfare than a household with a recent experience of a child death. There is ample evidence that within countries, as well as among countries, survival rates of individuals and their welfare levels are positively related. See for instance Adelman [1], Rodgers [23], Anker and Knowles [3].

A high level of welfare implies that people are well fed, have better sanitary conditions, and can live longer. If a household is facing a high risk of child death, its welfare level is deemed to be low. In this study, survival to age five is used to proxy wellbeing of a household, since one of the most striking features of African mortality is the heavy incidence of deaths in the second and third years of life relative to the normally high rates in the first year in other countries [6].

III. A DICHOTOMOUS MODEL OF CHILD SURVIVAL

This section presents a dichotomous model of determinants of child survival. The probability of a child surviving in a particular household is determined by an underlying response variable that captures the true socioeconomic and environmental conditions that the household faces. Since at a particular point in time, survival of a child is a binary variable (*i.e.*, a child is either alive or dead), let the underlying response variable y* be defined by the following regression relationship:

$$y_i^{\star} = \sum \mathbf{X}_i' \boldsymbol{\beta} + u_i \tag{1}$$

where

÷

$$\beta\beta = [\beta_1, \beta_2...\beta_k]$$
 and $\mathbf{x}_i' = [1, x_{i2}, x_{i3}...x_{ik}]$

In [1], y^* is not observable, as it is a latent variable. What is observable is an event represented by a dummy variable y defined by:

y = l if $y^* > 0$, if a child survived over a particular time period (2) and

$$y = 0$$
 otherwise.

۱

From (1) and (2) we can derive the following equation:

$$\operatorname{Pr} ob(y_{i} = 1) = \operatorname{Pr} ob(u_{i} > -\Sigma \mathbf{x}_{i}' \beta)$$
$$= 1 - F(-\Sigma \mathbf{x}_{i}' \beta)$$
(3)

where F is the cumulative distribution function for u_i , and

$$\operatorname{Pr}ob(y_i = 0 | \boldsymbol{\beta}, \mathbf{x}_i) = F(-\sum \mathbf{x}_i | \boldsymbol{\beta})$$
(4)

The observed values of y are the realization of the binomial variable with probabilities given by (3), which varies with \mathbf{X}_i . Thus, the likelihood function can be given by:

$$L = \prod_{y_i=0} \left[F\left(-\sum \mathbf{x}_i \, \beta\right) \right] \prod_{y_i=1} \left[1 - F\left(-\sum \mathbf{x}_i \, \beta\right) \right]$$
(5a)

Which can be written as:

$$L = \prod_{y_i=1} \left[F\left(-\sum \mathbf{x}_i \cdot \beta\right) \right]^{1-y_i} \left[1 - F\left(-\sum \mathbf{x}_i \cdot \beta\right) \right]^{y_i}$$
(5b)

The functional form imposed on F in (5)¹ depends on the assumptions made about u_i in (1).² The cumulative normal and logistic distributions are very close to each other. Thus, in certain circumstances, using one or the other will basically lead to the same result [20]. Moreover, following Amemiya [2], it is possible to derive the estimates of a probit model once we have parameters derived from the logit model.

The logit model assumes a logistic cumulative distribution of u_i in F (in (5a) and (5b)), so that the relevant logistic expressions are:

$$p(y_i = 1) = 1 - F\left(-\sum \mathbf{x}_i'\beta\right) = e^{\sum \mathbf{x}_i'\beta} / 1 + e^{\sum \mathbf{x}_i'\beta}$$
(6a)

 1 The log likelihood function for expressions [5a] and [5b] can be written as,

$$l(\beta) = \log L(\beta) = \sum_{i=0}^{n} y_i \log \left(1 - F(-\Sigma \mathbf{x}_i \beta_i) + (1 - y_i) \log F(-\Sigma \mathbf{x}_i \beta_i)\right)$$

² This basically forms the distinction between *logit* and *probit* (*normit*) models.

$$F(-\Sigma \mathbf{x}_{i} \boldsymbol{\beta}) = e^{-\sum \mathbf{x}_{i} \boldsymbol{\beta}} / 1 + e^{-\sum \mathbf{x}_{i} \boldsymbol{\beta}} = \frac{1}{1 + e^{\sum \mathbf{x}_{i} \boldsymbol{\beta}}} = \frac{1}{1 + e^{\sum \mathbf{x}_{i} \boldsymbol{\beta}}}$$
(6b)

As before, \mathbf{X}_i are the characteristics of the households/individuals, and $\boldsymbol{\beta}_i$ the coefficients for the respective variables in the logit regression. Having estimated (5) with maximum likelihood (ML) technique, (6a) basically gives us the probability of a child dying [Prob(y_i=1)] and (6b) the probability of a child surviving, i.e., Prob(y_i=0).

The underlying response variable for the probit model [see (1) for the logit model] can be expressed as:

$$y_i = \boldsymbol{\beta}' \boldsymbol{x}_i + u_i \tag{7}$$

where, the disturbance term in (7) follows a normal distribution and the dichotomous variables are defined as:

Zi = 1 if y_i is observed and Zi = 0 otherwise.

The cumulative probability distribution of the child survival status can now be written as:

$$\operatorname{Pr}ob(Z_{ij} = 1) = \Phi(\alpha_j - \beta' \mathbf{x}_i) - \Phi(\alpha_{j-1} - \beta' \mathbf{x}_i) \quad (8)$$

where, Φ is the cumulative distribution function³. The likelihood and log-likelihood functions for the model can be given by (9) and (10) respectively, as:

$$L = \prod_{i=1}^{n} \prod_{j=1}^{m} \left[\Phi(\alpha_{j} - \beta' x_{i}) - \Phi(\alpha_{j-1} - \beta' x_{i}) \right]^{Z_{i}}$$
(9)

In log-form, expression (9) becomes:

$${}^{*}_{L} = \log L = \sum_{i=1}^{n} \sum_{j=1}^{k} Z_{ij} \log \Phi \left[(\alpha_{j} - \beta' x_{i}) - \Phi(\alpha_{j-1} - \beta' x_{i}) \right]$$
(10)

Equation (10) can be maximized in the usual way, and can be solved iteratively by numerical methods, to yield maximum likelihood estimates of the probit model [20].

IV. PROBIT INDEX AS A WELFARE INDEX

The latent variable expression, $y_i = \boldsymbol{\beta}^{\mathsf{T}} \mathbf{X}_i + u_i$, depicted in (7) is the logit or probit index, depending

F= Φ (z) = (1/ $\sqrt{2} \pi$) exp (-z²/2). Moreover, the probit model marginal effects are;

 $\partial F_i / \partial x_i = \phi (x'_i \beta) \beta_i = \phi (\Phi^{-1} (F_i)) \beta_i$ where $F_i = \Phi (x'_i \beta)$

on whether it is the logistic or the normit model of child survival that is estimated. It shows the subjective welfare index that a household attaches to child survival. As is evident from (1) and (7), the subjective welfare index, y_i, depends on socioeconomic and environmental characteristics (X) of a household. In other words, the well-being of a household in any period depends on whether the household escaped child death in the previous period, and on other control variables such as household income, and education and health of household members. Sen [25] has argued that survival or death of a household member is the single most important summary measure of the wellbeing of a household at any particular time. Death of a family member, in this case a child, necessarily makes a household worse-off. That is, there is nothing that can replace the survival of a family member to keep the household at the same welfare level as before death.

This observation amounts to making a strong nonsubstitution assumption between survival of a family member *and* other goods that yield utility to the household. In other words, the household has Leontief preferences over *survival probabilities* of its members and other *goods*, e.g; real income and education. However, since death is eventually inevitable, this assumption applies only in cases of *premature* death. Without this assumption it is possible for a household to be made better-off by a monetary compensation after losing an elderly member, already at the natural end of a lifespan. We focus on child deaths because they are the prime examples of premature deaths in a society.

In (7), the parameters of interest, the β s are welfare weights. Once estimated, the total welfare that the household derives from child survival and from other "goods" can be computed. The weights indicate the contributions of the various factors to household welfare. That is, they are the ones used to weight the arguments of the welfare function. These weights are optimal, in the sense that they are the ones that maximize the well-being of the household given its environment. Moreover, the weights are consistent and non-arbitrary because they reflect a household's preference orderings of the arguments of the welfare function (i.e, the various determinants of the well-being). The weights here differ sharply from arbitrary welfare weights routinely reported in World Development Reports [28], [29]. Estimation of the welfare weights using (10) and computation of the welfare index via (7) enables calculation of the child survival probabilities using (6) or the normit formula.

³ The cumulative density is given by the following expression (see Wooldridge, 2002),

V. PROBIT INDEX AS AN INDIRECT UTILITY FUNCTION

The relationship between well-being and the price that a household must pay to avert a child death can be expressed as

 $v = v (p, y; a) = \max u (s; a)$ subject to s. p = y. (11)

where

p = price that households pay to increase, i.e, to improve child health.

y = exogenous household income;

a = economic, social and environmental conditions of the household

s = child survival, defined more precisely later

v = indirect utility, i.e, the maximum level of wellbeing the household can achieve if it must pay a price, *p*, to avert a child death when its income is equal to y

u = direct utility, i.e the maximum level of utility the household can obtain if it were to spend its income to avert a child death.

The demand for child survival, s(p, y; a) can be obtained via Roy's identity from the indirect utility function, v (.) or by maximizing the direct utility, u(.) subject to the budget constraint [30]. Letting s be a binary variable that takes a value of 1 if a child is alive and a value of 0 otherwise, the demand for s is clearly a probabilistic demand function for child survival as shown in (6). That is, the probability of child survival, s, decreases as p increases, holding constant household income, y, and its milieu, a. It is clear from (11), that by Roy's identity, the demand for child survival, s, is simply the negative of the marginal disutility of price divided by the marginal utility of income. The quantity demanded of child survival s, is expressed by a unitless probability measure that lies between zero and unity.

Thus, the welfare of a household conditional on child survival is W = W(s); where s, is empirically represented by child survival over a particular period. In this case, a household derives utility from child survival alone. Expanding the above expression we find that it can be written as W = W(s, a, y), where, a and y are other factors other than s that affect the well-being of a household. A poor household with a surviving child is considered better-off, in welfare terms, than a poor household without a surviving child. That is, if two households have the same income over a given time period, and at the end of the period, one has a surviving child, whereas the other has lost a child, the household with a surviving child is considered better-off; and likewise for rich households.

From (7), y_i is the logit index and β are parameters to be estimated. In this case, y_i is precisely the abbreviated social welfare index. It shows the level of wellbeing at the household level, conditional on child survival. If the error term ε is normally distributed, the probit model follows, and y_i becomes the probit index, which again, measures the wellbeing at the household level, conditional on survival. From the perspective child of macroeconomic theory, the probit/logit index associated with child survival is an indirect utility index, which at income, y, and price, p, (at which child survival can be secured) is equal to the direct utility the household derives from child survival.

IV. FROM UTILITY INDEX TO AN ABBREVIATED WELFARE INDEX

In order to make the idea of abbreviated social welfare empirically operational, it is necessary to specify a particular functional form for a probit index. The abbreviated social welfare function, as proxied by either a probit or a logit index, can be written in linear form as

$y_i = \alpha + \beta_1 Y + \beta_2 G + \beta_3 FGT + \delta W + \varepsilon$

Where

- y_i = Abbreviated social welfare index, the empirical value of a probit index
- *Y* = Household income per adult equivalent
- G = Distribution of income in a cluster

FGT = Poverty status in a cluster, which shows whether a household falls within a particular income distribution, or the proportion of poor households in that cluster.

W = Control variables at the household level, including maternal age, parental education, residence and household size.

As it happens, the probit or the logit index (y_i) , which can be aggregated at any level, is exactly the abbreviated social welfare index that is needed to rank regions according to the standard of living enjoyed by their populations.

This welfare indicator is a composite measure that captures the effects of income and its distribution, poverty status, parental education, mother's age and other socio-economic characteristics. It departs sharply from the UNDP's Human Development Index. While the UNDP estimates welfare by combining educational attainment, income and life expectancy in an arbitrary manner, our welfare index captures the welfare effects of each of the welfare determinants, in a well specified, and theoretically consistent manner. So, whereas the UNDP measure is arbitrary, the index proposed here is not. Whereas, UNDP index ignores income distribution and poverty status, the proposed index captures them both.

The index captures both the welfare and dis-welfare parts of the well-being. Unlike the income poverty index which is based only on income, and which captures only the bottom part of the household wellbeing, the abbreviated social welfare index includes both income and non-income measures of well-being over the entire distribution of well-being. Like the FGT poverty measure, the abbreviated social welfare index can be additively decomposed by regions or social groups using population shares of the groups or regions as welfare weights. There is also need to note that the ASWF takes into account economic deprivation of households. Thus, the abbreviated social welfare function incorporates effects of poverty in the well-being of the population.

REFERENCES

- Adelman, I. (1963). An Econometric Analysis of Population Growth, *American Economic Review*, 53, 314-39.
- [2] Amemiya, T. (1981). Qualitative Response Models: A Survey. Journal of Economic Literature, 19, 1483-1536. 1981.
- [3] Anker, R. & Knowles, J. (1980). An Empirical Analysis of Mortality Differentials in Kenya at the Macro and Micro Levels. *Economic Development and Cultural Change*, 29(1), 168-169.
- [4] Atkinson, A. (1983). *Social Justice and Public Policy*, Cambridge: The MIT Press.
- [5] Bhanojirao, V. (1991) Human Development Report 1990: Review and Assessment. World Development, 19(10), 14-55.
- [6] Brass, W. (1975). Introduction to Bio-Social Factors in African Demography. In R. P. Moss and R. J. A. Rathbone (Eds.) *The Population Factor in African Studies* (pp. 324-341). London: University of London Press.
- [7] Cahill, L. (2002). The Effects of Education on Farm Productivity. In K.J.K Griffin (Ed.) Human Development and the International Development Strategy for 1990s. London: Macmillan.
- [8] Cahill, L. (2002). Is the Human Development Index Redundant? Mimeo, Department of Economics, College of the Holy Cross.
- [9] Cotlear, D. (1990). The Effects of Education on Farm Productivity. In K.J.K. Griffin (Ed.) Human Development and the International Development Strategy for 1990s. London: Macmillan.

- [10] Dasgupta, P. & Weale, M. (1992). On Measuring the Quality of Life. World Development, 20(1), 1-20.
- [11] Despotis, D. (2004). A Reassessment of the Human Development Index Via Data Envelopment Analysis. Journal of the Operational Research Society, 1(12).
- [12] Fields, G. (2000). The Dynamics of Poverty, Inequality and Economic Well-Being, African Economic Growth in Comparative Perspective. *Journal of African Economies*, 9(1), 45-78.
- [13] Foster, J. (2005). Measuring the Distribution of Human Development: Methodology and an Application to Mexico. *Journal of Human Development*, 6(1), 5-29.
- [14] Giovanni, C. (1990). Investing in Human Resources: Health, Nutrition and Development for the 1990s. In K.J.K. Griffin, (Ed.) Human Development and International Development Strategy for the 1990s. London: Macmillan.
- [15] Gor, S. (2008). Regional Welfare and Income Distribution in Kenya: A New Approach to the Measurement of Wellbeing, Unpublished PhD Thesis: University of Nairobi.
- [16] Harkness, S. (2004). Social and Political Indicators of Human Well-Being, United Nations University: World Institute for Development Economics Research, 33, 1-22.
- [17] Hopkins, M. (1991). Human Development Revisited: a New UNDP Report. World Development, 19(10), 1469-1473.
- [18] Kelly, R. & Allen, C. (1991). The Human Development Index: Handle with Care. *Population and Development Review*, 17(2), 322-323.
- [19] Kimalu, P., Nafula, N., Manda, D., Mwabu, G., and Kimenyi, S. (2002). A Situational Analysis of Poverty in Kenya. *KIPPRA Working Paper 6*.
- [20] Maddala, G. (1983). Limited Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.
- [21] Manzoor, A. & Coomb, P. (1974). Attacking Rural Poverty: How Non-Formal Education can Help. Baltimore: The John Hopkins University Press.
- [22] Noorbakhsh, F. (1998). A Modified Human Development Index. World Development, 26(3), 517-528.
- [23] Rodgers, G. (1979). Income and Inequality as Determinants of Mortality: An International Cross- Section Analysis. *Population Studies*, 33(2), 343-51.
- [24] Saha, J. (2005). After Fifteen Year Use of the Human Development Index of the UNDP What Have We Learnt? Unpublished AERC Research Paper, 8-11.

- [25] Sen, A. (1988). Freedom of Choice: Concept and Content. *European Economic Review*, 32(2-3), 269-294.
- [26] Sudhir, A. & Sen, A. (1993). Human Development Index: Methodology and Measurement, New York, UNDP.
- [27] Ul Haq, M. (1995). *Reflections on Human Development*. United Nations Development Program 47. New York: Oxford University Press.
- [28] UNDP (2001), Human Development Report. New York: United Nations. Retrieved from: http://hdr.undp.org/
- [29] UNDP (2005), Human Development Report. New York: United Nations. Retrieved from: http://hdr.undp.org/
- [30] Varian, H. (1978). *Microeconomic Analysis* (2nd Ed.) New York: W.W. Norton & Company.
- [31] Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge Massachusetts: The MIT Press.