

Determinants of the Levels of Development Based on the Human Development Index: A Comparison of Regression Models for Limited Dependent Variables

Miraç Eren¹, Ali Kemal Çelik² & Arif Kubat³

¹ Department of Econometrics, Atatürk University, Erzurum, Turkey

² Department of Quantitative Methods, Atatürk University, Erzurum, Turkey

³ Department of Econometrics, İnönü University, Malatya, Turkey

Correspondence: Ali Kemal Çelik, Faculty of Economics and Administrative Sciences, Department of Quantitative Methods, Erzurum, Turkey. Tel: 90-442-231-2099. E-mail: akemal.celik@atauni.edu.tr

Received: December 2, 2013 Accepted: December 30, 2013 Online Published: January 16, 2014

doi:10.5539/res.v6n1p10

URL: <http://dx.doi.org/10.5539/res.v6n1p10>

Abstract

This paper aims to examine the factors affecting the level of development of countries using various regression models for limited dependent variables including binary logit, probit and Tobit analyses. In this manner, the paper may suggest a road map for high developed countries to achieve very high developed levels. For this purpose, Human Development Indices of 84 countries were involved in the analysis with respect to nine independent variables. The results of the all regression models indicate that determinants including life expectancy at birth, expected years of schooling, labour force participation rate (female-male ratio), and GDP per capita have statistically significant effects on the level of development.

Keywords: logit, probit, Tobit, human development index, development levels

1. Introduction

In the contemporary era, the concept of development has been in greater need of analysis and clarification and the word has come to be extraordinarily widely used in public discourse probably more so than ever before in its history (Payne & Phillips, 2010). As the Industrial Revolution brought forth the most rapid development of the productive forces and accumulation of wealth, it will be logical to examine both material and subjective indicators that may have been responsible for the economic expansion (Jaffee, 1998). Economists use the terms developing and transitional rather than growing to underline that the goals of these countries involve more than simply an increase in output. Thus, development refers to an increase in productive capacity and output by means of a change in the underlying institutions (Colander, 1998). Specifically, development should include increasingly effective institutions in all sectors of society, at economic, social, political, national, regional, and local levels. It should also comprise increasing participation of individuals and groups when making decisions which effect their lives economically and politically (Spalding, 1990). Development of a country is principally associated with both the improvement in basic human needs (health, sanitation, education, etc.) and the growth of per capita income (Mazumdar, 2003; Streeten, 1981), while the main objective of development is simply to create an enabling environment for better life (UNDP, 1990).

This paper employs to determine the factors affecting the development levels of selected countries using regression models for limited dependent variables. The paper is organized as follows. Section 2 deals with the concepts of the human development and the Human Development Index (HDI) in general. Section 3 reviews the existing literature to exhibit the studies that concentrate on the evaluation of the HDI data using several statistical approaches. Section 4 gives general information about the theoretical background of regression models for limited dependent variables. Section 5 introduces the methodology of the study and the data being used to perform the analysis. Section 6 presents the results of the logit, probit, and Tobit analyses and the interpretation of the underlying results in detail. Finally Section 7 involves the general discussion and comparison in terms of the present evidence.

2. Human Development and the HDI

Human development is evidently about enlarging people's choices on the basis of shared natural resources. Since freedoms and capabilities possess a more expansive notion than basic needs, human development can be adopted as the expansion of people's freedoms and capabilities to lead lives that they and have reason to value. In this sense, the human development approach is consistently concerned with making sense of the world and addressing challenges now and in the future (UNDP, 2011a). The human development approach involves two central theses about people and development which are concerned with evaluating improvements in human lives as a distinctive development objective and what human beings can do to achieve such improvements particularly policy and political changes (Fukuda-Parr, 2003).

The *Human Development Reports* are independent publications commissioned by the United Nations Development Programme (UNDP). They are recognized as independent intellectual exercises, important tools for raising awareness about human development around the world, and pioneers of methodological innovation and development thinking (UNDP, 2013a). UNDP also introduces regional, national, and local *Human Development Reports* to enable responsible people or countries making comparisons and evaluations with respect to the corresponding evidence. The first *Human Development Report* was launched in 1990 for the purpose of 'putting people back at the center of the development process with respect to economic policy', while the actual particular objective of the *Human Development Reports* was to investigate the progress of the conditions of human being, associated with the removal of disadvantages and the creation of opportunities to lead worthwhile lives (UNDP, 1990; Anand & Sen, 1997, 2000). So far, twenty two reports were annually published considering various similar or different components.

As human development shows characteristics of an abstract variable, it cannot be directly observed and measured. Many researches argue that GNP per capita cannot satisfy all aspects of development even if an affirmative relationship between GNP per capita and social and human welfare, so human development requires a range of socio-economic indicators. Therefore, a satisfactory measurement concept was developed by the UNDP (Nübler, 1995; McGillivray, 1991; Ivanova, Arcelus, & Srinivasan, 1999). Human Development Index (HDI) is a composite index that measures average achievement in three basic dimensions of human development such as along and healthy life, knowledge, and a decent standard of living (UNDP, 2011a). The HDI was initially developed to underline that people and their capabilities should be the ultimate criteria for evaluating the development level of a country along with economic growth (UNDP, 2013a). Despite a number of authors discuss some drawbacks of the HDI (Trabold-Nübler, 1991; Hicks, 1997; Ivanova, Arcelus, & Srinivasan, 1999; Chowdhury & Squire, 2006; Klugman, Rodríguez, & Choi, 2011), and even propose some modified indices (Paul, 1996; Sharma, 1997; Noorbakhsh, 1998; Chakravarty, 2003; Mazumdar, 2003; Grimm, et al., 2008; Hartgen & Klasen, 2012; Herrero, Martínez, & Villar, 2012; Bilbao-Ubillos, 2013; Blancard & Hoarau, 2013); one of the most significant superiorities of the HDI is regarded as its usefulness that is able to represent both social and economic development through a single value between 0 and 1 (UNDP, 2013b).

The HDI generally comprises three key components including longevity, knowledge, and income; where longevity is measured by life expectancy at birth, knowledge is measured by adult literacy and mean years of schooling, and the income in the HDI is a proxy for a bundle of goods and services needed for the best use of human capabilities (ul Haq, 2003). In order to keep simplicity and usefulness of the HDI, the *Human Development Reports* present a variety of relevant information in detail and they provide a summary for some of the major components of human development using the HDI to exhibit an alternative emphasis for several standard measures of economic development (Anand & Sen, 2000).

The HDI is the geometric mean of normalized indices measuring achievements in each three key components and there are two steps to calculating the HDI involving creating the dimension indices and aggregating the subindices to produce the HDI. First step principally sets minimum and maximum values to transform the indicators into indices between 0 and 1. The maximum values refer to the highest observed values in the time series (i.e. 1980-2011 for *Human Development Report 2011*). On the other hand, the minimum values are set at 20 years for life expectancy, at zero years for both education variables and at 100 dollars for per capita Gross National Income (GNI). In that way, the subindices are calculated as the following:

$$\text{Dimension index} = \frac{\text{actual value} - \text{minimum value}}{\text{maximum value} - \text{minimum value}} \quad (1)$$

This equation is applied to each of the two subcomponents for education, then a geometric mean of the resulting indices is created and the equation is employed again to the geometric mean of the indices using zero as the minimum and the highest geometric mean of the resulting indices for the time period under consideration as the

maximum. This will be equivalent to applying the corresponding equation directly to the geometric mean of the two subcomponents. Specifically, for income the natural logarithm of the actual minimum and maximum values is used to avoid concavity. Consequently, the HDI is the geometric mean of the three dimension indices as shown by the following formula (UNDP, 2011a):

$$(I_{life}^{1/3} \cdot I_{education}^{1/3} \cdot I_{income}^{1/3}) \quad (2)$$

Since 2010, in the knowledge component of the HDI is measured by combining the expected years of schooling for a school-age child in a country entering school today with the mean years of prior schooling for adults aged 25 and older. The income measurement has changed from purchasing-power-adjusted per-capita GDP to purchasing-power-adjusted-per-capita GNI, which provides a more accurate economic portrayal of many developing countries (UNDP, 2011b).

3. Literature Review

A number of studies concentrated on the statistical implementation of the HDI in the literature. Lai (2000) used the weighted principal component method on the human development indicators to measure and analyze the progress of human development in the world and employed the main principal component to quantify the temporal changes of the human development of several selected countries.

Lee, Lin, and Fang (2006) developed a fuzzy multiple objective data envelopment analysis model to assess the relative performance of the countries in terms of human development by using optimal weights for the component indices of the HDI.

Antony and Rao (2007) calculated the HDI and Human Poverty Index of Indian states and developed a composite index using several multivariate statistical methods that is able to explain variations in poverty, health, nutritional status, and standard of living.

Cherchye, Ooghe, and Van Puyenbroeck (2008) focused on robust human development rankings and showed that all proposed ranking procedures can be implemented through linear programming. Finally, they illustrated how their methodology could prove useful in the robustness of human development country ranking/classification in a statistical way.

Yang and Hu (2008) used both one- and multi-dimensional cluster analysis to analyze China's HDI data for 1982, 1995, 1999, and 2003 to classify China's provinces into four tiers based on the three basic development aspects embedded in the HDI.

Grimm et al. (2010) applied a simple approach to compute the three components and the overall HDI for quintiles of the income distribution to make an empirical assessment of 32 countries and they found that a strong overall negative correlation between the level of human development and inequality in human development.

Abayomi and Pizarro (2013) offered a straightforward methodology for measurement of progress, across many dimensions, using cross-national social indices, which they classified as linear combinations of multivariate country level data onto a univariate score and they suggested a Bayesian approach which yields probabilistic intervals for the point estimates of country scores.

Pinar, Stengos, and Topaloglou (2013) considered a stochastic dominance approach for measured human development such as the official equally-weighted HDI and they compared the official equally-weighted HDI to all possible indices constructed from a set of individual components to achieve the most optimistic scenario for development.

Terzi, Trezzini, and Moroni (2013) developed a partial least squares path model to investigate the causal effects between different types of institutions and human development.

Tofallis (2013) purposed a two-step automatic-democratic approach to weight setting for the new HDI, which has properties of non-subjectivity, fairness, and convenience. The first step aims to find the most advantageous set of weights for each nation in turn and the second step evaluates the associated optimal scores on the corresponding indicators to obtain a single weight set. According to the analysis results, the highest weight was placed on the life expectancy dimension.

Wu, Fan, and Pan (2013) employed a super efficiency model to evaluate the rationality of the HDI rankings of 19 evaluated OECD countries in 2009 and empirical results showed that approximately 75 % of the evaluated countries had rather different results in the efficiency rankings and the HDI rankings.

4. Regression Models for Limited Dependent Variables

One of the most important developments in econometrics is the increasing use of microeconomic data on individual economic units (Chow, 1983). Qualitative response models are adopted as regression models in which dependent variables take discrete values (Amemiya, 1985). Consideration of modeling behavior with discrete dependent variables can be divided into consideration of the functional form for the probabilities and consideration of the convenient estimation technique for alternative models and data sets (Hanushek & Jackson, 1977). Many social problems are naturally discrete or qualitative rather than continuous or quantitative such like an event occurs or it does not occur (i.e. employed vs unemployed, married vs unmarried, guilty vs unguilty, etc.), where binary discrete phenomena take the form of a dichotomous indicator or dummy variable (Pampel, 2000; Allison, 1999). In that circumstance, since empirical research in the social sciences increasingly encourages the use of probabilistic choice models involving the linear probability model, the logit model and the probit model, in recent years (Caudill & Jackson, 1989), many researchers in business and economics are frequently concerned with the analysis of a binary dependent variable, where 0 and 1 denote failure and success respectively. In principal, a sequence of success and failure is considered that the chance of that a particular trial is a 1 depends on the value of one or more independent variables. So, the appropriate tests and their estimates for these situations are related to in which the independent variable is preassigned and with independent variables that are functions of the corresponding sequence (Cox, 1958).

The binary response model has an S-shaped relationship between the probability of an event and the independent variables (Long, 1997). The following equation defines a univariate binary qualitative response model as

$$P(y_i = 1) = F(x_i' \beta_0), \quad i = 1, 2, \dots, n \quad (3)$$

where y_i denotes a sequence of independent binary random variables taking the value of 0 or 1, x_i denotes a K -vector of known constraints, β_0 denotes a K -vector unknown parameters, and finally F denotes a certain known function (Amemiya, 1985). The value of $F(x_i' \beta)$ must lie between 0 and 1, although $x_i' \beta$ can take any value on the real line (Davidson & MacKinnon, 2004). Since the most serious difficulty arises from the linear probability model predictions that may lie outside the (0, 1) interval, alternative distributional assumptions for which all predictions must lie within the appropriate interval were investigated (Pindyck & Rubinfeld, 1981). A formal probit or logit model provides estimation of probabilities, marginal effects, and other derivative results in regard to the normal or logistic distribution on the data (Greene, 2007). In other words, these models are actually interested in analyzing how a series of exogenous variables influences the underlying probability (Hanushek & Jackson, 1977).

4.1 Binary Logit Model

The outcome variable in logistic regression model is binary or dichotomous that distinguishes a logistic regression model from the linear regression model. However, the techniques employed in the linear regression analysis serve as a reference for the logistic regression approach. When the outcome variable is binary in a regression model, the conditional mean of the regression equation must be formulated to be bounded between 0 and 1, while the binomial and non-normal distribution describes the distribution of the errors (Hosmer & Lemeshow, 2000). Suppose Y denotes a binary variable and X denotes a quantitative explanatory variable, and let $\pi(x)$ denote the success probability for the binomial distribution, the logistic regression model possesses a linear form for the logit of the relevant possibility as the following:

$$\text{logit}[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \quad (4)$$

In this context, $\pi(x)$ can be described in terms of exponential function, $\exp(x)$ as follows (Agresti, 1996):

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \quad (5)$$

Many of the desirable characteristics of a linear regression model are able to be achieved through the logit transformation, where the logit is linear in its parameters, may be continuous, and may range from $-\infty$ to $+\infty$, depending on the range of x (Hosmer & Lemeshow). The errors in the logit model are assumed to have a standard logistic distribution with mean zero and variance $\pi^2/3$ (Long, 1997). A dummy variable f_{1i} such that

$$f_{1i} = \begin{cases} 1 & \text{if } i\text{th unit chose alternative 1} \\ 0 & \text{if } i\text{th unit did not choose alternative 1} \end{cases} \quad (6)$$

can be introduced to form the likelihood of observing the sample. The likelihood of the sample of n observations is as follows (Chow, 1983):

$$L = \prod_{i=1}^n P_{1i}^{f_{1i}} (1 - P_{1i})^{1-f_{1i}} \quad (7)$$

The logit model is particularly convenient to explain the odds of success or another substantive outcome or the odds of success for one group relative to another. Odds can be specifically defined as the ratio of the probability of one outcome to another, (Powers & Xie, 2000), that provides a more sensible scale for multiplicative comparisons. In addition, because odds ratios are less sensitive to changes in the marginal frequencies than other measures of association, they are widely-used measures of the relationship between two binary variables. Because the application of the linear probability model brings about some crucial problems such as heteroscedasticity, normality, biased predictions, etc., the logit model can be regarded as the most popular regression model for limited dependent variables by means of several advantages such as simpler interpretation and generalization, and desirable sampling properties (Allison, 1999).

4.2 Binary Probit Model

Suppose z_i denotes a vector of characteristics of the i th economic unit and of the object for choice which are explanatory variables for the probability P_1 that the i th unit will choose alternative 1. The probit model explains this probability as

$$P_1(z_i) = \int_{-\infty}^{\beta'z_i} (2\pi)^{-1/2} e^{-s^2/2} ds = \Phi(\beta'z_i) \quad (8)$$

where β is a vector of unknown coefficients (Chow, 1983). This equation can also be referred to the standardized cumulative normal function where s denotes a random variable which is normally distributed with mean zero and unit variance (Pindyck & Rubinfeld, 1981). Historically, probit analysis has been widely used whenever the analyst has individual or micro data and is considering a model with a discrete dependent variable (Hanushek & Jackson, 1977). The probit analysis assumes that there is an underlying response variable y_i^* defined by the regression relationship

$$y_i^* = \beta'x_i + u_i \quad (10)$$

where y_i^* is an unobservable latent variable in practice. The observed variable is a dummy variable y defined by

$$\begin{aligned} y &= 1 & \text{if } y_i^* > 0 \\ y &= 0 & \text{otherwise} \end{aligned} \quad (11)$$

The probability formulation can be achieved as

$$P(y_i = 1) = P(u_i > -\beta'x_i) = 1 - F(-\beta'x_i) \quad (12)$$

from the corresponding relations where F is the cumulative distribution function for u , and hence the likelihood function is as follows (Maddala, 1983):

$$L = \prod_{y_i=0} F(-\beta'x_i) \prod_{y_i=1} [1 - F(-\beta'x_i)] \quad (13)$$

The probit model precisely assumes that the underlying probability function is normal rather than logistic and the derivation of the maximum likelihood estimator can be obtained that exactly parallels the maximum likelihood logistic estimator (Hanushek & Jackson, 1977).

4.3 Tobit Model

Tobit model is a censored or truncated regression model in which the range of the dependent variable is constrained. In that case, the model is called truncated if the observations outside a specific range are totally lost and censored if at least the independent variables are observed (Amemiya, 1985). When the latent variable y_i^* is observed if it is more than zero and is not observed if it is less than or equal to zero, then the observed variable will be represented by the following formula:

$$y_i = \begin{cases} y_i^* = \beta x_i + u_i & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \quad (14)$$

$$u_i \sim \text{IN}(0, \sigma^2)$$

This condition is known as the Tobit model which was initially developed by Tobin (1958), that can be denominated as censored normal regression model because some observations on the latent variable are censored (Maddala, 2001). The term censoring refers to the only limited information about the dependent variable can be obtained, when the independent variables for the entire sample, but for specific observations are observed. The

Tobit model uses all of the information and provides consistent estimates of the parameters (Long, 1997). The following equation exhibits the likelihood function of the standard Tobit model as

$$L = \prod_0 [1 - \Phi(x_i'\beta/\sigma)] \prod_1 \sigma^{-1} \Phi[y_i - x_i'\beta]/\sigma \quad (15)$$

where Φ and ϕ denote the distribution and density function of the standard normal variable, respectively (Amemiya, 1985).

5. Methodology

The 2011 HDI is divided into four quartiles, from “very high” to “low” human development achievement, as introduced in the 2010 HDI, while the new classifications managed to improve functionality of earlier HDI cut-off points and also to reduce the amount of variation in each group (UNDP, 2011b). Within the scope of the relevant classification this study comparatively performs binary logit, probit, and Tobit analyses for 84 countries that have “very high” and “high” HDI values. Particularly, the HDI values under the censored point of 0.783 were involved in the Tobit analysis. Although there are many factors which may possibly affect the level of development of countries, this study classifies these factors in five main groups such as health, education, gender, environment, and economics-population. Furthermore, the study seeks to investigate the statistically significant factors through experiencing several combinations. As a result, four variables were statistically significant for all regression models: life expectancy at birth, expected years of schooling, labour participation rate, and GDP per capita (PPP\$).

Table 1. Dependent and Independent variables used in the study

Variables	Description ¹
HDI (dependent variable)	HDI values of 84 countries (0 = high HDI, 1 = very high HDI)
Life expectancy at birth	Number of years a newborn infant could expect to live if prevailing patterns of age-specific mortality rates at the time of birth stay the same throughout his/her life
Mean years of schooling	Average number of years of education received by people ages 25 and older, converted from education attainment levels using official durations of each level
Expected years of schooling	Number of years schooling that a child of school entrance age can expect to receive if prevailing patterns of age-specific enrolment rates persist throughout his/her life
Adult literacy rate	Percentage of the population ages 15 and older who can, with understanding, both read and write a short simple statement on their everyday life
Pupil-teacher ratio	Average number of students per teacher in primary education in a given school year
Seats in national parliament (% female)	Proportion of seats held by women in a lower or single house or an upper house or senate expressed as percentage of total seats
Labour participation rate (female-male ratio)	Ratio of female to male of the working age population (ages 15-64) that actively engages in the labour market, by either working or actively looking for work
Urban population	Particular matter concentrations in terms of fine suspended of particulates of human-made or natural origin less than 10 microns in diameter that are capable of penetrating deep into the respiratory tract
GDP per capita (PPP\$)	Gross domestic product (GDP) expressed in purchasing power parity (PPP) international dollar terms, divided by midyear population

¹Source: Adapted from “UNDP Human Development Report 2011-Sustainability and Equity: A Better Future for All”

6. Results

6.1 Binary Logit Analysis Results

Table 2 exhibits the results of the logit analysis for the selected variables. The log-likelihood value of the logit analysis was found as -14.504 for 84 observations. The likelihood ratio is widely-used, rather than *F*-test, for the

models which were solved by the maximum likelihood technique. The likelihood value of the logit model was 87.01 and it was statistically significant ($p < 0.05$) which means the four independent variables were jointly able to explain the logit model. The odds ratio (OR) values enable the analyst to interpret the relationship between dependent and the independent variables. For instance, the OR value which is more than 1 implies that the independent variable increases the probability of the occurrence of the event, in contrast, the OR value which is less than 1 decreases this probability. In this context, the logit analysis estimates that life expectancy at birth increases the probability of having a very high HDI for a country nearly 1.5 times more (OR = 1.57, 95% CI = 1.04 – 2.39) than having a high HDI value. Expected years of schooling had the most impact on the level of development for the selected countries, where this variable increases the likelihood of having a very high HDI approximately 3 times (OR = 3.21; 95 CI = 1.39 – 7.44) than high HDI value. Similarly, labour participation rate (OR = 1.08; 95% CI = 1.00 – 1.16) and GDP per capita (OR = 1.21; 95% CI = 1.05 – 1.41) also increase the likelihood of having a very high HDI value for a country.

Table 2. Binary logit analysis results

Log likelihood = -14.504 Number of observations = 84 LR $\chi^2(4) = 87.01$ Prob > $\chi^2 = 0.0000$ Pseudo R ² = 0.7500					
Dependent variable	OR	Std. Err.	Z	P > z	[95% CI]
Life expectancy at birth	1.57	0.335	2.12	0.034**	1.04 – 2.39
Expected years of schooling	3.21	1.376	2.72	0.006*	1.39 – 7.44
Labour participation rate (female-male ratio)	1.08	0.039	1.99	0.046**	1.00 – 1.16
GDP per capita	1.21	0.091	2.60	0.009**	1.05 – 1.41

* Statistically significant at the 0.01 level

** Statistically significant at the 0.05 level

Moreover, the goodness-of-fit of the logit model was also examined in Table 3. Pseudo- R² value of the logit model was equivalent to McFadden's R² value, which was found as 0.75. This value implies that independent variables are able to explain the relevant changes on the dependent variable with a percentage of 75. Because, pseudo- R² can take relatively lower values for regression models performing limited dependent variables when compared with classical regression models, the pseudo- R² value of the logit model in this study can be considered as quite acceptable. In addition, very low value of Akaike Information Criteria (AIC) and the negative value of Bayesian Information Criteria (BIC) ensure that the goodness-of-fit of the logit model was satisfactory.

Table 3. Goodness-of-fit test for binary logit model

Criteria	Value	Criteria	Value
Log-Lik Intercept Only	-58.010	Log-Lik Full Model	-14.504
D(79)	29.007	LR (4)	87.013
McFadden's R ²	0.750	Prob > LR	0.000
Maximum Likelihood R ²	0.645	McFadden's Adjusted R ²	0.664
McKelvey and Zavoina's R ²	0.910	Cragg & Uhler's R ²	0.862
Variance of y*	36.404	Efron's R ²	0.781
Count R ²	0.940	Variance of error	3.290
AIC	0.464	Adjusted Count R ²	0.872
BIC	-321.027	AIC*n	39.007
		BIC'	-69.289

A marginal effect explains the rate change in one quantity relative to another, or the change in the dependent variable per unit change in the independent variable. Since the logit and probit models are linear in the parameters, a unit change in x_i produces a β change in a latent variable y_i^* for the logit and probit models (Powers & Xie, 2000). Table 4 presents marginal effects of the four independent variables for the logit analysis.

More specifically, one unit change in life expectancy at birth increases the probability of having a very high HDI value with 6% ($dy/dx = 0.06$; 95% CI = -0.005 – 0.112). Similarly, expected years of schooling had the most impact ($dy/dx = 0.15$; 95% CI = -0.003 – 0.298) on the probability of the high HDI value, where one unit change in this independent variable increases the probability with 15%. Finally, one unit change on labour participation rate ($dx/dy = 0.09$; 95% CI = -0.002 – 0.020) and GDP per capita ($dx/dy = 0.03$; 95% CI = 0.005 – 0.044) increase the corresponding probability with 9% and 3%, respectively.

Table 4. Marginal effects for binary logit analysis

Dependent variable	dy/dx	Std. Err.	Z	P > z	[95% CI]	X
Life expectancy at birth	0.06	0.315	1.82	0.034**	(0.005) – 0.112	76.70
Expected years of schooling	0.15	0.077	1.93	0.006*	(0.003) – 0.298	14.64
Labour participation rate (female-male ratio)	0.09	0.006	1.67	0.046**	(0.002) – 0.020	69.54
GDP per capita	0.03	0.010	2.51	0.009*	0.005 – 0.044	22.42

* Statistically significant at the 0.01 level

** Statistically significant at the 0.05 level

6.2 Binary Probit Analysis Results

The impact of the four independent variables was analyzed using the binary probit model. As shown in Table 3, the log-likelihood of the probit model was -14.293 with 84 observations. The likelihood ratio of the probit analysis was 87.43, which was statistically significant ($p < 0.05$). Pseudo R^2 value of the probit model was found as 0.7536. This value implies that independent variables are able to explain the relevant changes on the dependent variable with an approximate percentage of 75. Table 5 also exhibits that all four independent variables have increasing impacts on the HDI values of the selected countries.

Table 5. Binary probit analysis results

Log likelihood = -14.293					
Number of observations = 84					
LR $\chi^2(4) = 87.43$					
Prob > $\chi^2 = 0.0000$					
Pseudo $R^2 = 0.7536$					
Dependent variable	Coeff.	Std. Err.	Z	P > z	[95% CI]
Life expectancy at birth	0.25	0.117	2.18	0.029**	1.04 – 2.39
Expected years of schooling	0.65	0.232	2.81	0.005*	1.39 – 7.44
Labour participation rate (female-male ratio)	0.04	0.039	2.04	0.041**	1.00 – 1.16
GDP per capita	0.11	0.091	3.12	0.002*	1.05 – 1.41
Constant	-33.4	9.859	-3.39	0.001*	(52.7) – (14.1)

* Statistically significant at the 0.01 level

** Statistically significant at the 0.05 level

Table 6 presents the goodness-of-fit test for the probit model. Pseudo- R^2 value of the logit model was very similar to McFadden's R^2 value, which was found as 0.754. In addition, very low value of Akaike Information Criteria (AIC) and the negative value of Bayesian Information Criteria (BIC) ensure that the goodness-of-fit of the probit model was satisfactory.

Table 6. Goodness-of-fit test for Binary Probit Model

Criteria	Value	Criteria	Value
Log-Lik Intercept Only	-58.010	Log-Lik Full Model	-14.293
D(79)	28.586	LR (4)	87.434
		Prob > LR	0.000
McFadden's R^2	0.754	McFadden's Adjusted R^2	0.667
Maximum Likelihood R^2	0.647	Cragg & Uhler's R^2	0.864
McKelvey and Zavoina's R^2	0.913	Efron's R^2	0.778
Variance of y^*	11.525	Variance of error	1.000
Count R^2	0.940	Adjusted Count R^2	0.872
AIC	0.459	AIC*n	39.586
BIC	-321.449	BIC'	-69.711

Table 7 presents marginal effects of the four independent variables for the probit analysis. In particular, one unit change in life expectancy at birth increases the probability of having a very high HDI value with 7% ($dy/dx = 0.07$; 95% CI = 0.009 – 0.121). Furthermore, expected years of schooling had the most impact ($dy/dx = 0.17$; 95% CI = 0.028 – 0.305) on the probability of the high HDI value, where one unit change in this independent variable increases the probability with 17%. Finally, one unit change on labour participation rate ($dx/dy = 0.01$; 95% CI = 0.01 – 0.021) and GDP per capita ($dx/dy = 0.03$; 95 % CI = 0.011 – 0.045) increase the corresponding probability with 1% and 3%, respectively.

Table 7. Marginal Effects for Binary Probit Analysis

Dependent variable	dy/dx	Std. Err.	Z	P > z	[95% CI]	X
Life expectancy at birth	0.07	0.029	2.27	0.023**	0.009 – 0.121	76.70
Expected years of schooling	0.17	0.070	2.36	0.018**	0.028 – 0.305	14.64
Labour participation rate (female-male ratio)	0.01	0.005	2.12	0.034**	0.001 – 0.021	69.54
GDP per capita	0.03	0.009	3.21	0.001*	0.011 – 0.045	22.42

* Statistically significant at the 0.01 level

** Statistically significant at the 0.05 level

6.3 Tobit Analysis Results

Table 8 represents the results of the Tobit analysis with respect to four independent variables and the HDI values. The countries which had lower and higher HDI values than 0.783 were considered as developed and very developed countries, respectively. In this circumstance, 39 countries were found as left-censored and 45 countries were uncensored and analyzed. The coefficients of the independent variables can be interpreted as the impact of the latent variable, y_i^* . As shown in Table 6, all independent variables were statistically significant ($p < 0.05$) and had increasing impacts on the dependent variable; however the constant term was not statistically significant on the contrary with the logit and probit models. The likelihood ratio (LR $\chi^2(4)=164.79$) was also statistically significant ($p < 0.05$), and all independent variables were able to explain the tobit model. In practice, pseudo R^2 values can be ignored for the Tobit models, which are generally inadequate for the interpretation.

Finally, Table 9 exhibits marginal effects of the four independent variables for the tobit analysis. In particular, one unit change in expected years of schooling increases the probability of having a very high HDI value with nearly 1.5% ($dy/dx = 0.0145$; 95% CI = 0.0103 – 0.0188). The other independent variables had little impacts which were under 1%.

Table 8. Tobit analysis results

Log likelihood = 100.10					
Number of observations = 84					
LR $\chi^2(4) = 164.79$					
Prob > $\chi^2 = 0.0000$					
Pseudo R ² = -4.6539					
Observation summary: 39 left-censored observations at HDI ≤ 0.783					
45 uncensored observations					
Dependent variable	Coeff.	Std. Err.	T	P > t	[95% CI]
Life expectancy at birth	0.0074	0.0013	5.72	0.000*	0.0005 – 0.0101
Expected years of schooling	0.0017	0.0026	6.46	0.000*	0.0115 – 0.0217
Labour participation rate (female-male ratio)	0.0008	0.0003	3.25	0.002*	0.0003 – 0.0013
GDP per capita	0.0014	0.0002	6.53	0.000*	0.0010 – 0.0018
Constant	-0.0969	0.8464	-1.14	0.256	(0.2653) – (0.071)
/sigma	0.0213	0.0846			0.0168 – 0.0251

* Statistically significant at the 0.01 level

Table 9. Marginal effects for Tobit analysis

Dependent variable	dy/dx	Std. Err.	Z	P > z	[95% CI]	X
Life expectancy at birth	0.0065	0.0011	5.72	0.000*	0.0043 – 0.0088	76.70
Expected years of schooling	0.0145	0.0002	6.68	0.000*	0.0103 – 0.0188	14.64
Labour participation rate (female-male ratio)	0.0007	0.0002	3.23	0.001*	0.0003 – 0.0012	69.54
GDP per capita	0.0012	0.0001	6.69	0.000*	0.0009 – 0.0016	22.42

* Statistically significant at the 0.01 level

7. Conclusion & Discussion

This paper performed the logit, probit, and Tobit models to determine factors affecting the level of development of 84 countries which had high and very high HDI values in the *Human Development Report 2011*. Marginal effects of all three models were also examined to illustrate the results in detail. The results suggested that life expectancy at birth, expected years of schooling, labour participation rate, and GDP per capita had statistically significant impact on the level of development of 84 countries being analyzed. For all three models, expected years of schooling had the most impact on the probability of having a very high HDI value for a country. Although the values were very similar, the logit model appeared to be the most convenient model to determine the impact of the four determinants through the interpretation of the odds ratios. Countries having high HDI values may concentrate on all four determinants and design their policies, but especially on the education dimension to reach the standards of top countries.

References

- Abayomi, K., & Pizarro, G. (2013). Monitoring human development goals: A straightforward (Bayesian) methodology for cross-national indices. *Social Indicators Research*, 110(2), 489-515. <http://dx.doi.org/10.1007/s11205-011-9946-y>
- Agresti, A. (1996). *An introduction to categorical data analysis*. New York: John Wiley & Sons, Inc.
- Allison, P. D. (1999). *Logistic regression using the SAS system: Theory and application*. Cary, NC: SAS Institute, Inc.
- Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.
- Anand, S., & Sen, A. (1997). Concepts of human development and poverty: A multidimensional perspective. In *Poverty and human development: Human development papers 1997* (pp. 1-20). New York, USA: United Nations Development Programme.
- Anand, S., & Sen, A. (2000). The income component of the human development index. *Journal of Human Development*, 1(1), 83-106. <http://dx.doi.org/10.1080/14649880050008782>

- Antony, G. M., & Visweswara Rao, K. (2007). A composite index to explain variations in poverty, health, nutritional status and standard of living: Use of multivariate statistical methods. *Public Health*, 121(8), 578-587. <http://dx.doi.org/10.1016/j.puhe.2006.10.018>
- Bilbao-Ubillos, J. (2013). Another approach to measuring human development: The composite dynamic human development index. *Social Indicators Research*, 111(2), 473-484. <http://dx.doi.org/10.1007/s11205-012-0015-y>
- Blancard, S., & Hoarau, J. (2013). A new sustainable human development indicator for small island developing states: A reappraisal from data envelopment analysis. *Economic Modelling*, 30, 623-635. <http://dx.doi.org/10.1016/j.econmod.2012.10.016>
- Caudill, S. B., & Jackson, J. D. (1989). Measuring marginal effects in limited dependent variable models. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 38(3), 203-206. <http://dx.doi.org/10.2307/2348876>
- Chakravarty, S. R. (2003). A generalized human development index. *Review of Development Economics*, 7(1), 99-104. <http://dx.doi.org/10.1111/1467-9361.00178>
- Cherchye, L., Ooghe, E., & Van Puyenbroeck, T. (2008). Robust human development rankings. *The Journal of Economic Inequality*, 6(4), 287-321. <http://dx.doi.org/10.1007/s10888-007-9058-8>
- Chow, G. C. (1983). *Econometrics*. New York: McGraw-Hill Book Company.
- Chowdhury, S., & Squire, L. (2006). Setting weights for aggregate indices: An application to the commitment to development index and human development index. *Journal of Development Studies*, 42(5), 761-771. <http://dx.doi.org/10.1080/00220380600741904>
- Colander, D. C. (1998). *Macroeconomics* (3rd ed.). Massachusetts: The McGraw-Hill Companies, Inc.
- Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2), 215-242. Retrieved from <http://www.jstor.org/stable/2983890>
- Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods*. New York: Oxford University Press.
- Fukuda-Parr, S. (2003). The human development paradigm: Operationalizing Sen's ideas on capabilities. *Feminist Economics*, 9(2-3), 301-317. <http://dx.doi.org/10.1080/1354570022000077980>
- Greene, W. H. (2007). *Econometric Analysis* (6th ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- Grimm, M., Harttgen, K., Klasen, S., & Misselhorn, M. (2008). A human development index by income groups. *World Development*, 36(12), 2527-2546. <http://dx.doi.org/10.1016/j.worlddev.2007.12.001>
- Grimm, M., Harttgen, K., Klasen, S., Misselhorn, M., Munzi, T., & Smeeding, T. (2010). Inequality in human development: An empirical assessment of 32 countries. *Social Indicators Research*, 97(2), 191-211. <http://dx.doi.org/10.1007/s11205-009-9497-7>
- Hanushek, E. A., & Jackson, J. E. (1977). *Statistical methods for social scientists*. New York: Academic Press.
- Harttgen, A., & Klasen, S. (2012). A household-based human development index. *World Development*, 40(5), 878-899. <http://dx.doi.org/10.1016/j.worlddev.2011.09.011>
- Herrero, C., Martínez, R., & Villar, A. (2012). A newer human development index. *Journal of Human Development and Capabilities: A Multi-Disciplined Journal for People-Centered Development*, 13(2), 247-268. <http://dx.doi.org/10.1080/19452829.2011.645027>
- Hicks, D. A. (1997). The inequality-adjusted human development index: A constructive proposal. *World Development*, 25(8), 1283-1298. [http://dx.doi.org/10.1016/S0305-750X\(97\)00034-X](http://dx.doi.org/10.1016/S0305-750X(97)00034-X)
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons, Inc.
- Ivanova, I., Arcelus, F. J., & Srinivasan, G. (1999). An assessment of the measurement properties of the human development index. *Social Indicators Research*, 46(2), 157-179. <http://dx.doi.org/10.1023/A:1006839208067>
- Jaffee, D. (1998). *Levels of socio-economic development theory* (2nd ed.). Westport, CT: Praeger Publishers.
- Klugman, J., Rodríguez, F., & Choi, H. (2011). The HDI 2010: New controversies, old critiques. *The Journal of Economic Inequality*, 9(2), 249-288. <http://dx.doi.org/10.1007/s10888-011-9178-z>

- Lai, D. (2000). Temporal analysis of human development indicators: Principal component approach. *Social Indicators Research*, 51(3), 331-366. <http://dx.doi.org/10.1023/A:1007065804509>
- Lee, H. S., Lin, K., & Fang, H. H. (2006). A fuzzy multiple objective DEA for the human development index. In B. Gabrys, & R. J. Howlett & L. C. Jain (Eds.), *Knowledge-Based Intelligent Information and Engineering Systems: Lecture Notes in Computer Science 4252* (pp. 922-928). Heidelberg, Berlin: Springer.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, California: Sage Publications, Inc.
- Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. New York: Cambridge University Press.
- Maddala, G. S. (2001). *Introduction to Econometrics* (3rd ed.). Chichester, UK: John Wiley & Sons, Ltd.
- Mazumdar, K. (2003). A new approach to human development index. *Review of Social Economy*, 61(4), 535-549. <http://dx.doi.org/10.1080/0034676032000160895>
- McGillivray, M. (1991). The human development index: Yet another redundant composite development indicator? *World Development*, 19(10), 1461-1468. [http://dx.doi.org/10.1016/0305-750X\(91\)90088-Y](http://dx.doi.org/10.1016/0305-750X(91)90088-Y)
- Nübler, I. (1995). The human development index revisited. *Intereconomics*, 30(4), 171-176. <http://dx.doi.org/10.1007/BF02928088>
- Noorbakhsh, F. (1998). A modified human development index. *World Development*, 26(3), 517-528. [http://dx.doi.org/10.1016/S0305-750X\(97\)10063-8](http://dx.doi.org/10.1016/S0305-750X(97)10063-8)
- Pampel, F. C. (2000). *Logistic regression: A primer*. California, Thousand Oaks: Sage Publications, Inc.
- Paul, S. (1996). A modified human development index and international comparison. *Applied Economics Letters*, 3(10), 677-682. <http://dx.doi.org/10.1080/135048596355925>
- Payne, A., & Phillips, N. (2010). *Development*. Cambridge, UK: Polity Press.
- Pinar, M., Stengos, T., & Topaloglou, N. (2013). Measuring human development: A stochastic dominance approach. *Journal of Economic Growth*, 18(1), 69-108. <http://dx.doi.org/10.1007/s10887-012-9083-8>
- Pindyck, R. S., & Rubinfeld, D. L. (1981). *Econometric models and economic forecasts* (2nd ed.). New York: McGraw-Hill Book Company.
- Powers, D. A., & Xie, Y. (2000). *Statistical methods for categorical data analysis*. London, Oval Road: Academic Press.
- Sharma, S. D. (1997). Making the human development index (HDI) gender-sensitive. *Gender & Development*, 5(1), 60-61. <http://dx.doi.org/10.1080/741922304>
- Spalding, N. L. (1990). The relevance of basic needs for political and economic development. *Studies in Comparative International Development*, 25(3), 90-115. <http://dx.doi.org/10.1007/BF02687181>
- Streeten, P. (1981). *First things first: Meeting basic human needs in developing Countries*. Oxford: Oxford University Press.
- Terzi, S., Trezzini, A., & Moroni, L. (2013). A PLS path model to investigate the relations between institutions and human development. *Quality & Quantity*, February 2013. <http://dx.doi.org/10.1007/s11135-013-9834-4>
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26(1), 24-36. Retrieved from <http://www.jstor.org/stable/1907382>
- Tofallis, C. (2013). An automatic-democratic approach to weight setting for the new human development index. *Journal of Population Economics*, 26(4), 1325-1345. <http://dx.doi.org/10.1007/s00148-012-0432-x>
- Trabold-Nübler, H. (1991). The human development index-A new development indicator? *Intereconomics*, 26(5), 236-243. <http://dx.doi.org/10.1007/BF02928996>
- ul Haq, M. (2003). The birth of the human development index. In S. Fukuda-Parr & A. K. Shiva Kuma (Eds.), *Readings in human development* (pp. 127-137). Oxford, UK: Oxford University Press.
- UNDP. (1990). *Human development report 1990*. Oxford: Oxford University Press. Retrieved from <http://hdr.undp.org/en/reports/global/hdr1990/chapters/>
- UNDP. (2011a). *Human development report 2011, Sustainability and equity: A better future for all*. New York: Palgrave Macmillan. Retrieved from <http://hdr.undp.org/en/reports/global/hdr2011/download/>

- UNDP. (2011b). *Frequently asked questions (FAQs) about the human development index (HDI)*. Retrieved September 12, 2013, from <http://hdr.undp.org/en/statistics/hdi/>
- UNDP. (2013a). *The human development report*. Retrieved September 25, 2013, from <http://hdr.undp.org/en/reports/about/>
- UNDP. (2013b). *Human development index*. Retrieved September 25, 2013, from <http://hdr.undp.org/en/statistics/hdi/>
- Wu, P., Fan, C., & Pan, S. (2013). Does human development index provide rational development rankings? Evidence from efficiency rankings in super efficiency. *Social Indicators Research*, March 2013. <http://dx.doi.org/10.1007/s11205-013-0285-z>
- Yang, Y., & Hu, A. (2008). Investigating regional disparities of China's human development with cluster analysis: A historical perspective. *Social Indicators Research*, 86(3), 417-432. <http://dx.doi.org/10.1007/s11205-007-9177-4>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).